The Cross-Section of Expected Stock Returns

Eugene F. Fama; Kenneth R. French

The Journal of Finance, Vol. 47, No. 2 (Jun., 1992), 427-465.

Stable URL:
http://links.jstor.org/sici?sici=0022-1082%28199206%2947%3A2%3C427%3ATCOESR%3E2.0.CO%3B2-N

The Journal of Finance is currently published by American Finance Association.

THE JOURNAL OF FINANCE • VOL. XLVII, NO. 2 • JUNE 1992

The Cross-Section of Expected Stock Returns

EUGENE F. FAMA and KENNETH R. FRENCH*

ABSTRACT

Two easily measured variables, size and book-to-market equity, combine to capture the cross-sectional variation in average stock returns associated with market β, size, leverage, book-to-market equity, and earnings-price ratios. Moreover, when the tests allow for variation in β that is unrelated to size, the relation between market β and average return is flat, even when β is the only explanatory variable.

THE ASSET-PRICING MODEL OF Sharpe (1964), Lintner (1965), and Black (1972) has long shaped the way academics and practitioners think about average returns and risk. The central prediction of the model is that the market portfolio of invested wealth is mean-variance efficient in the sense of Markowitz (1959). The efficiency of the market portfolio implies that (a) expected returns on securities are a positive linear function of their market βs (the slope in the regression of a security's return on the market's return), and (b) market βs suffice to describe the cross-section of expected returns.

There are several empirical contradictions of the Sharpe-Lintner-Black (SLB) model. The most prominent is the size effect of Banz (1981). He finds that market equity, ME (a stock's price times shares outstanding), adds to the explanation of the cross-section of average returns provided by market βs. Average returns on small (low ME) stocks are too high given their β estimates, and average returns on large stocks are too low.

Another contradiction of the SLB model is the positive relation between leverage and average return documented by Bhandari (1988). It is plausible that leverage is associated with risk and expected return, but in the SLB model, leverage risk should be captured by market β. Bhandari finds, however, that leverage helps explain the cross-section of average stock returns in tests that include size (ME) as well as β.

Stattman (1980) and Rosenberg, Reid, and Lanstein (1985) find that average returns on U.S. stocks are positively related to the ratio of a firm's book value of common equity, BE, to its market value, ME. Chan, Hamao, and Lakonishok (1991) find that book-to-market equity, BE/ME, also has a strong role in explaining the cross-section of average returns on Japanese stocks.

* Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637. We acknowledge the helpful comments of David Booth, Nai-fu Chen, George Constantinides, Wayne Ferson, Edward George, Campbell Harvey, Josef Lakonishok, Rex Sinquefield, René Stulz, Mark Zmijewski, and an anonymous referee. This research is supported by the National Science Foundation (Fama) and the Center for Research in Security Prices (French).

Finally, Basu (1983) shows that earnings-price ratios (E/P) help explain the cross-section of average returns on U.S. stocks in tests that also include size and market β. Ball (1978) argues that E/P is a catch-all proxy for unnamed factors in expected returns; E/P is likely to be higher (prices are lower relative to earnings) for stocks with higher risks and expected returns, whatever the unnamed sources of risk.

Ball's proxy argument for E/P might also apply to size (ME), leverage, and book-to-market equity. All these variables can be regarded as different ways to scale stock prices, to extract the information in prices about risk and expected returns (Keim (1988)). Moreover, since E/P, ME, leverage, and BE/ME are all scaled versions of price, it is reasonable to expect that some of them are redundant for describing average returns. Our goal is to evaluate the joint roles of market β, size, E/P, leverage, and book-to-market equity in the cross-section of average returns on NYSE, AMEX, and NASDAQ stocks.

Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) find that, as predicted by the SLB model, there is a positive simple relation between average stock returns and β during the pre-1969 period. Like Reinganum (1981) and Lakonishok and Shapiro (1986), we find that the relation between β and average return disappears during the more recent 1963–1990 period, even when β is used alone to explain average returns. The appendix shows that the simple relation between β and average return is also weak in the 50-year 1941–1990 period. In short, our tests do not support the most basic prediction of the SLB model, that average stock returns are positively related to market βs.

Unlike the simple relation between β and average return, the univariate relations between average return and size, leverage, E/P, and book-to-market equity are strong. In multivariate tests, the negative relation between size and average return is robust to the inclusion of other variables. The positive relation between book-to-market equity and average return also persists in competition with other variables. Moreover, although the size effect has attracted more attention, book-to-market equity has a consistently stronger role in average returns. Our bottom-line results are: (a) β does not seem to help explain the cross-section of average stock returns, and (b) the combination of size and book-to-market equity seems to absorb the roles of leverage and E/P in average stock returns, at least during our 1963–1990 sample period.

If assets are priced rationally, our results suggest that stock risks are multidimensional. One dimension of risk is proxied by size, ME. Another dimension of risk is proxied by BE/ME, the ratio of the book value of common equity to its market value.

It is possible that the risk captured by BE/ME is the relative distress factor of Chan and Chen (1991). They postulate that the earning prospects of firms are associated with a risk factor in returns. Firms that the market judges to have poor prospects, signaled here by low stock prices and high ratios of book-to-market equity, have higher expected stock returns (they are penalized with higher costs of capital) than firms with strong prospects. It is also possible, however, that BE/ME just captures the unraveling (regression toward the mean) of irrational market whims about the prospects of firms.

Whatever the underlying economic causes, our main result is straightforward. Two easily measured variables, size (ME) and book-to-market equity (BE/ME), provide a simple and powerful characterization of the cross-section of average stock returns for the 1963–1990 period.

In the next section we discuss the data and our approach to estimating β. Section II examines the relations between average return and β and between average return and size. Section III examines the roles of E/P, leverage, and book-to-market equity in average returns. In sections IV and V, we summarize, interpret, and discuss applications of the results.

I. Preliminaries

A. Data

We use all nonfinancial firms in the intersection of (a) the NYSE, AMEX, and NASDAQ return files from the Center for Research in Security Prices (CRSP) and (b) the merged COMPUSTAT annual industrial files of income-statement and balance-sheet data, also maintained by CRSP. We exclude financial firms because the high leverage that is normal for these firms probably does not have the same meaning as for nonfinancial firms, where high leverage more likely indicates distress. The CRSP returns cover NYSE and AMEX stocks until 1973 when NASDAQ returns also come on line. The COMPUSTAT data are for 1962–1989. The 1962 start date reflects the fact that book value of common equity (COMPUSTAT item 60), is not generally available prior to 1962. More important, COMPUSTAT data for earlier years have a serious selection bias; the pre-1962 data are tilted toward big historically successful firms.

To ensure that the accounting variables are known before the returns they are used to explain, we match the accounting data for all fiscal yearends in calendar year $t - 1$ (1962–1989) with the returns for July of year t to June of $t + 1$. The 6-month (minimum) gap between fiscal yearend and the return tests is conservative. Earlier work (e.g., Basu (1983)) often assumes that accounting data are available within three months of fiscal yearends. Firms are indeed required to file their 10-K reports with the SEC within 90 days of their fiscal yearends, but on average 19.8% do not comply. In addition, more than 40% of the December fiscal yearend firms that do comply with the 90-day rule file on March 31, and their reports are not made public until April. (See Alford, Jones, and Zmijewski (1992).)

We use a firm's market equity at the end of December of year $t - 1$ to compute its book-to-market, leverage, and earnings-price ratios for $t - 1$, and we use its market equity for June of year t to measure its size. Thus, to be included in the return tests for July of year t, a firm must have a CRSP stock price for December of year $t - 1$ and June of year t. It must also have monthly returns for at least 24 of the 60 months preceding July of year t (for "pre-ranking" β estimates, discussed below). And the firm must have COMPUSTAT data on total book assets (A), book equity (BE), and earnings (E), for its fiscal year ending in (any month of) calendar year $t - 1$.

Our use of December market equity in the E/P, BE/ME, and leverage ratios is objectionable for firms that do not have December fiscal yearends because the accounting variable in the numerator of a ratio is not aligned with the market value in the denominator. Using ME at fiscal yearends is also problematic; then part of the cross-sectional variation of a ratio for a given year is due to market-wide variation in the ratio during the year. For example, if there is a general fall in stock prices during the year, ratios measured early in the year will tend to be lower than ratios measured later. We can report, however, that the use of fiscal-yearend MEs, rather than December MEs, in the accounting ratios has little impact on our return tests.

Finally, the tests mix firms with different fiscal yearends. Since we match accounting data for all fiscal yearends in calendar year $t - 1$ with returns for July of t to June of $t + 1$, the gap between the accounting data and the matching returns varies across firms. We have done the tests using the smaller sample of firms with December fiscal yearends with similar results.

B. Estimating Market βs

Our asset-pricing tests use the cross-sectional regression approach of Fama and MacBeth (1973). Each month the cross-section of returns on stocks is regressed on variables hypothesized to explain expected returns. The time-series means of the monthly regression slopes then provide standard tests of whether different explanatory variables are on average priced.

Since size, E/P, leverage, and BE/ME are measured precisely for individual stocks, there is no reason to smear the information in these variables by using portfolios in the Fama-MacBeth (FM) regressions. Most previous tests use portfolios because estimates of market βs are more precise for portfolios. Our approach is to estimate βs for portfolios and then assign a portfolio's β to each stock in the portfolio. This allows us to use individual stocks in the FM asset-pricing tests.

B.1. β Estimation: Details

In June of each year, all NYSE stocks on CRSP are sorted by size (ME) to determine the NYSE decile breakpoints for ME. NYSE, AMEX, and NASDAQ stocks that have the required CRSP-COMPUSTAT data are then allocated to 10 size portfolios based on the NYSE breakpoints. (If we used stocks from all three exchanges to determine the ME breakpoints, most portfolios would include only small stocks after 1973, when NASDAQ stocks are added to the sample.)

We form portfolios on size because of the evidence of Chan and Chen (1988) and others that size produces a wide spread of average returns and βs. Chan and Chen use only size portfolios. The problem this creates is that size and the βs of size portfolios are highly correlated (-0.988 in their data), so asset-pricing tests lack power to separate size from β effects in average returns.

To allow for variation in β that is unrelated to size, we subdivide each size decile into 10 portfolios on the basis of pre-ranking βs for individual stocks. The pre-ranking βs are estimated on 24 to 60 monthly returns (as available) in the 5 years before July of year t. We set the β breakpoints for each size decile using only NYSE stocks that satisfy our COMPUSTAT-CRSP data requirements for year $t - 1$. Using NYSE stocks ensures that the β breakpoints are not dominated after 1973 by the many small stocks on NASDAQ. Setting β breakpoints with stocks that satisfy our COMPUSTAT-CRSP data requirements guarantees that there are firms in each of the 100 size-β portfolios.

After assigning firms to the size-β portfolios in June, we calculate the equal-weighted monthly returns on the portfolios for the next 12 months, from July to June. In the end, we have post-ranking monthly returns for July 1963 to December 1990 on 100 portfolios formed on size and pre-ranking βs. We then estimate βs using the full sample (330 months) of post-ranking returns on each of the 100 portfolios, with the CRSP value-weighted portfolio of NYSE, AMEX, and (after 1972) NASDAQ stocks used as the proxy for the market. We have also estimated βs using the value-weighted or the equal-weighted portfolio of NYSE stocks as the proxy for the market. These βs produce inferences on the role of β in average returns like those reported below.

We estimate β as the sum of the slopes in the regression of the return on a portfolio on the current and prior month's market return. (An additional lead and lag of the market have little effect on these sum βs.) The sum βs are meant to adjust for nonsynchronous trading (Dimson (1979)). Fowler and Rorke (1983) show that sum βs are biased when the market return is autocorrelated. The 1st- and 2nd-order autocorrelations of the monthly market returns for July 1963 to December 1990 are 0.06 and -0.05, both about 1 standard error from 0. If the Fowler-Rorke corrections are used, they lead to trivial changes in the βs. We stick with the simpler sum βs. Appendix Table AI shows that using sum βs produces large increases in the βs of the smallest ME portfolios and small declines in the βs of the largest ME portfolios.

Chan and Chen (1988) show that full-period β estimates for portfolios can work well in tests of the SLB model, even if the true βs of the portfolios vary through time, if the variation in the βs is proportional,

$$\beta_{jt} - \beta_j = k_t(\beta_j - \beta), \tag{1}$$

where β_{jt} is the true β for portfolio j at time t, β_j is the mean of β_{jt} across t, and β is the mean of the β_j. The Appendix argues that (1) is a good approximation for the variation through time in the true βs of portfolios (j) formed on size and β. For diehard β fans, sure to be skeptical of our results on the weak role of β in average stock returns, we can also report that the results stand up to robustness checks that use 5-year pre-ranking βs, or 5-year post-ranking βs, instead of the full-period post-ranking βs.

We allocate the full-period post-ranking β of a size-β portfolio to each stock in the portfolio. These are the βs that will be used in the Fama-MacBeth cross-sectional regressions for individual stocks. We judge that the precision of the full-period post-ranking portfolio βs, relative to the imprecise β estimates that would be obtained for individual stocks, more than makes up for the fact that true βs are not the same for all stocks in a portfolio. And note that assigning full-period portfolio βs to stocks does not mean that a stock's β is constant. A stock can move across portfolios with year-to-year changes in the stock's size (ME) and in the estimates of its β for the preceding 5 years.

B.2. β Estimates

Table I shows that forming portfolios on size and pre-ranking βs, rather than on size alone, magnifies the range of full-period post-ranking βs. Sorted on size alone, the post-ranking βs range from 1.44 for the smallest ME portfolio to 0.92 for the largest. This spread of βs across the 10 size deciles is smaller than the spread of post-ranking βs produced by the β sort of *any* size decile. For example, the post-ranking βs for the 10 portfolios in the smallest size decile range from 1.05 to 1.79. Across all 100 size-β portfolios, the post-ranking βs range from 0.53 to 1.79, a spread 2.4 times the spread, 0.52, obtained with size portfolios alone.

Two other facts about the βs are important. First, in each size decile the post-ranking βs closely reproduce the ordering of the pre-ranking βs. We take this to be evidence that the pre-ranking β sort captures the ordering of true post-ranking βs. (The appendix gives more evidence on this important issue.) Second, the β sort is not a refined size sort. In any size decile, the average values of ln(ME) are similar across the β-sorted portfolios. Thus the pre-ranking β sort achieves its goal. It produces strong variation in post-ranking βs that is unrelated to size. This is important in allowing our tests to distinguish between β and size effects in average returns.

II. β and Size

The Sharpe-Lintner-Black (SLB) model plays an important role in the way academics and practitioners think about risk and the relation between risk and expected return. We show next that when common stock portfolios are formed on size alone, there seems to be evidence for the model's central prediction: average return is positively related to β. The βs of size portfolios are, however, almost perfectly correlated with size, so tests on size portfolios are unable to disentangle β and size effects in average returns. Allowing for variation in β that is unrelated to size breaks the logjam, but at the expense of β. Thus, when we subdivide size portfolios on the basis of pre-ranking βs, we find a strong relation between average return and size, but no relation between average return and β.

A. Informal Tests

Table II shows post-ranking average returns for July 1963 to December 1990 for portfolios formed from one-dimensional sorts of stocks on size or β. The portfolios are formed at the end of June each year and their equal-weighted returns are calculated for the next 12 months. We use returns for July to June to match the returns in later tests that use the accounting data. When we sort on just size or 5-year pre-ranking βs, we form 12 portfolios. The middle 8 cover deciles of size or β. The 4 extreme portfolios (1A, 1B, 10A, and 10B) split the bottom and top deciles in half.

Table II shows that when portfolios are formed on size alone, we observe the familiar strong negative relation between size and average return (Banz (1981)), and a strong positive relation between average return and β. Average returns fall from 1.64% per month for the smallest ME portfolio to 0.90% for the largest. Post-ranking βs also decline across the 12 size portfolios, from 1.44 for portfolio 1A to 0.90 for portfolio 10B. Thus, a simple size sort seems to support the SLB prediction of a positive relation between β and average return. But the evidence is muddied by the tight relation between size and the βs of size portfolios.

The portfolios formed on the basis of the ranked market βs of stocks in Table II produce a wider range of βs (from 0.81 for portfolio 1A to 1.73 for 10B) than the portfolios formed on size. Unlike the size portfolios, the β-sorted portfolios do not support the SLB model. There is little spread in average returns across the β portfolios, and there is no obvious relation between β and average returns. For example, although the two extreme portfolios, 1A and 10B, have much different βs, they have nearly identical average returns (1.20% and 1.18% per month). These results for 1963–1990 confirm Reinganum's (1981) evidence that for β-sorted portfolios, there is no relation between average return and β during the 1964–1979 period.

The 100 portfolios formed on size and then pre-ranking β in Table I clarify the contradictory evidence on the relation between β and average return produced by portfolios formed on size or β alone. Specifically, the two-pass sort gives a clearer picture of the separate roles of size and β in average returns. Contrary to the central prediction of the SLB model, the second-pass β sort produces little variation in average returns. Although the post-ranking βs in Table I increase strongly in each size decile, average returns are flat or show a slight tendency to decline. In contrast, within the columns of the average return and β matrices of Table I, average returns and βs decrease with increasing size.

The two-pass sort on size and β in Table I says that variation in β that is tied to size is positively related to average return, but variation in β unrelated to size is not compensated in the average returns of 1963–1990. The proper inference seems to be that there is a relation between size and average return, but controlling for size, there is no relation between β and average return. The regressions that follow confirm this conclusion, and they produce another that is stronger. The regressions show that when one allows

Table I

Average Returns, Post-Ranking βs and Average Size For Portfolios Formed on Size and then β: Stocks Sorted on ME (Down) then Pre-Ranking β (Across): July 1963 to December 1990

Portfolios are formed yearly. The breakpoints for the size (ME, price times shares outstanding) deciles are determined in June of year t (t = 1963–1990) using all NYSE stocks on CRSP. All NYSE, AMEX, and NASDAQ stocks that meet the CRSP-COMPUSTAT data requirements are allocated to the 10 size portfolios using the NYSE breakpoints. Each size decile is subdivided into 10 β portfolios using pre-ranking βs of individual stocks, estimated with 2 to 5 years of monthly returns (as available) ending in June of year t. We use only NYSE stocks that meet the CRSP-COMPUSTAT data requirements to establish the β breakpoints. The equal-weighted monthly returns on the resulting 100 portfolios are then calculated for July of year t to June of year $t + 1$.

The post-ranking βs use the full (July 1963 to December 1990) sample of post-ranking returns for each portfolio. The pre- and post-ranking βs (here and in all other tables) are the sum of the slopes from a regression of monthly returns on the current and prior month's returns on the value-weighted portfolio of NYSE, AMEX, and (after 1972) NASDAQ stocks. The average return is the time-series average of the monthly equal-weighted portfolio returns, in percent. The average size of a portfolio is the time-series average of monthly averages of ln(ME) for stocks in the portfolio at the end of June of each year, with ME denominated in millions of dollars.

The average number of stocks per month for the size-β portfolios in the smallest size decile varies from 70 to 177. The average number of stocks for the size-β portfolios in size deciles 2 and 3 is between 15 and 41, and the average number for the largest 7 size deciles is between 11 and 22.

The All column shows statistics for equal-weighted size-decile (ME) portfolios. The All row shows statistics for equal-weighted portfolios of the stocks in each β group.

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
Panel A: Average Monthly Returns (in Percent)											
All	1.25	1.34	1.29	1.36	1.31	1.33	1.28	1.24	1.21	1.25	1.14
Small-ME	1.52	1.71	1.57	1.79	1.61	1.50	1.50	1.37	1.63	1.50	1.42
ME-2	1.29	1.25	1.42	1.36	1.39	1.65	1.61	1.37	1.31	1.34	1.11
ME-3	1.24	1.12	1.31	1.17	1.70	1.29	1.10	1.31	1.36	1.26	0.76
ME-4	1.25	1.27	1.13	1.54	1.06	1.34	1.06	1.41	1.17	1.35	0.98
ME-5	1.29	1.34	1.42	1.39	1.48	1.42	1.18	1.13	1.27	1.18	1.08
ME-6	1.17	1.08	1.53	1.27	1.15	1.20	1.21	1.18	1.04	1.07	1.02
ME-7	1.07	0.95	1.21	1.26	1.09	1.18	1.11	1.24	0.62	1.32	0.76
ME-8	1.10	1.09	1.05	1.37	1.20	1.27	0.98	1.18	1.02	1.01	0.94
ME-9	0.95	0.98	0.88	1.02	1.14	1.07	1.23	0.94	0.82	0.88	0.59
Large-ME	0.89	1.01	0.93	1.10	0.94	0.93	0.89	1.03	0.71	0.74	0.56

Table I—*Continued*

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
						Panel B: Post-Ranking βs					
All		0.87	0.99	1.09	1.16	1.26	1.29	1.35	1.45	1.52	1.72
Small-ME	1.44	1.05	1.18	1.28	1.32	1.40	1.40	1.49	1.61	1.64	1.79
ME-2	1.39	0.91	1.15	1.17	1.24	1.36	1.41	1.43	1.50	1.66	1.76
ME-3	1.35	0.97	1.13	1.13	1.21	1.26	1.28	1.39	1.50	1.51	1.75
ME-4	1.34	0.78	1.03	1.17	1.16	1.29	1.37	1.46	1.51	1.64	1.71
ME-5	1.25	0.66	0.85	1.12	1.15	1.16	1.26	1.30	1.43	1.59	1.68
ME-6	1.23	0.61	0.78	1.05	1.16	1.22	1.28	1.36	1.46	1.49	1.70
ME-7	1.17	0.57	0.92	1.01	1.11	1.14	1.26	1.24	1.39	1.34	1.60
ME-8	1.09	0.53	0.74	0.94	1.02	1.13	1.12	1.18	1.26	1.35	1.52
ME-9	1.03	0.58	0.74	0.80	0.95	1.06	1.15	1.14	1.21	1.22	1.42
Large-ME	0.92	0.57	0.71	0.78	0.89	0.95	0.92	1.02	1.01	1.11	1.32
						Panel C: Average Size (ln(ME))					
All	4.11	3.86	4.26	4.33	4.41	4.27	4.32	4.26	4.19	4.03	3.77
Small-ME	2.24	2.12	2.27	2.30	2.30	2.28	2.29	2.30	2.32	2.25	2.15
ME-2	3.63	3.65	3.68	3.70	3.72	3.69	3.70	3.69	3.69	3.70	3.68
ME-3	4.10	4.14	4.18	4.12	4.15	4.16	4.16	4.18	4.14	4.15	4.15
ME-4	4.50	4.53	4.53	4.57	4.54	4.56	4.55	4.52	4.58	4.52	4.56
ME-5	4.89	4.91	4.91	4.93	4.95	4.93	4.92	4.93	4.92	4.92	4.95
ME-6	5.30	5.30	5.33	5.34	5.34	5.33	5.33	5.33	5.33	5.34	5.36
ME-7	5.73	5.73	5.75	5.77	5.76	5.73	5.77	5.77	5.76	5.72	5.76
ME-8	6.24	6.26	6.27	6.26	6.24	6.24	6.27	6.24	6.24	6.24	6.26
ME-9	6.82	6.82	6.84	6.82	6.82	6.81	6.81	6.81	6.81	6.80	6.83
Large-ME	7.93	7.94	8.04	8.10	8.04	8.02	8.02	7.94	7.80	7.75	7.62

Table II
Properties of Portfolios Formed on Size or Pre-Ranking β: July 1963 to December 1990

At the end of June of each year t, 12 portfolios are formed on the basis of ranked values of size (ME) or pre-ranking β. The pre-ranking βs use 2 to 5 years (as available) of monthly returns ending in June of t. Portfolios 2–9 cover deciles of the ranking variables. The bottom and top 2 portfolios (1A, 1B, 10A, and 10B) split the bottom and top deciles in half. The breakpoints for the ME portfolios are based on ranked values of ME for all NYSE stocks on CRSP. NYSE breakpoints for pre-ranking βs are also used to form the β portfolios. NYSE, AMEX, and NASDAQ stocks are then allocated to the size or β portfolios using the NYSE breakpoints. We calculate each portfolio's monthly equal-weighted return for July of year t to June of year $t + 1$, and then reform the portfolios in June of $t + 1$.

BE is the book value of common equity plus balance-sheet deferred taxes, A is total book assets, and E is earnings (income before extraordinary items, plus income-statement deferred taxes, minus preferred dividends). BE, A, and E are for each firm's latest fiscal year ending in calendar year $t - 1$. The accounting ratios are measured using market equity ME in December of year $t - 1$. Firm size ln(ME) is measured in June of year t, with ME denominated in millions of dollars.

The average return is the time-series average of the monthly equal-weighted portfolio returns, in percent. ln(ME), ln(BE/ME), ln(A/ME), ln(A/BE), E/P, and E/P dummy are the time-series averages of the monthly average values of these variables in each portfolio. Since the E/P dummy is 0 when earnings are positive, and 1 when earnings are negative, E/P dummy gives the average proportion of stocks with negative earnings in each portfolio.

β is the time-series average of the monthly portfolio βs. Stocks are assigned the post-ranking β of the size-β portfolio they are in at the end of June of year t (Table I). These individual-firm βs are averaged to compute the monthly βs for each portfolio for July of year t to June of year $t + 1$.

Firms is the average number of stocks in the portfolio each month.

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Panel A: Portfolios Formed on Size												
Return	1.64	1.16	1.29	1.24	1.25	1.29	1.17	1.07	1.10	0.95	0.88	0.90
β	1.44	1.44	1.39	1.34	1.33	1.24	1.22	1.16	1.08	1.02	0.95	0.90
ln(ME)	1.98	3.18	3.63	4.10	4.50	4.89	5.30	5.73	6.24	6.82	7.39	8.44
ln(BE/ME)	−0.01	−0.21	−0.23	−0.26	−0.32	−0.36	−0.36	−0.44	−0.40	−0.42	−0.51	−0.65
ln(A/ME)	0.73	0.50	0.46	0.43	0.37	0.32	0.32	0.24	0.29	0.27	0.17	−0.03
ln(A/BE)	0.75	0.71	0.69	0.69	0.68	0.67	0.68	0.67	0.69	0.70	0.68	0.62
E/P dummy	0.26	0.14	0.11	0.09	0.06	0.04	0.04	0.03	0.03	0.02	0.02	0.01
E(+)/P	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.09	0.09
Firms	772	189	236	170	144	140	128	125	119	114	60	64

Table II—*Continued*

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
	Panel B: Portfolios Formed on Pre-Ranking β											
Return	1.20	1.20	1.32	1.26	1.31	1.30	1.30	1.23	1.23	1.33	1.34	1.18
β	0.81	0.79	0.92	1.04	1.13	1.19	1.26	1.32	1.41	1.52	1.63	1.73
ln(ME)	4.21	4.86	4.75	4.68	4.59	4.48	4.36	4.25	3.97	3.78	3.52	3.15
ln(BE/ME)	−0.18	−0.13	−0.22	−0.21	−0.23	−0.22	−0.22	−0.25	−0.23	−0.27	−0.31	−0.50
ln(A/ME)	0.60	0.66	0.49	0.45	0.42	0.42	0.45	0.42	0.47	0.46	0.46	0.31
ln(A/BE)	0.78	0.79	0.71	0.66	0.64	0.65	0.67	0.67	0.70	0.73	0.77	0.81
E/P dummy	0.12	0.06	0.09	0.09	0.08	0.09	0.10	0.12	0.12	0.14	0.17	0.23
E(+)/P	0.11	0.12	0.10	0.10	0.10	0.10	0.10	0.09	0.10	0.09	0.09	0.08
Firms	116	80	185	181	179	182	185	205	227	267	165	291

for variation in β that is unrelated to size, the relation between β and average return is flat, even when β is the only explanatory variable.

B. Fama-MacBeth Regressions

Table III shows time-series averages of the slopes from the month-by-month Fama-MacBeth (FM) regressions of the cross-section of stock returns on size, β, and the other variables (leverage, E/P, and book-to-market equity) used to explain average returns. The average slopes provide standard FM tests for determining which explanatory variables on average have non-zero expected premiums during the July 1963 to December 1990 period.

Like the average returns in Tables I and II, the regressions in Table III say that size, ln(ME), helps explain the cross-section of average stock returns. The average slope from the monthly regressions of returns on size alone is -0.15%, with a t-statistic of -2.58. This reliable negative relation persists no matter which other explanatory variables are in the regressions; the average slopes on ln(ME) are always close to or more than 2 standard errors from 0. The size effect (smaller stocks have higher average returns) is thus robust in the 1963–1990 returns on NYSE, AMEX, and NASDAQ stocks.

In contrast to the consistent explanatory power of size, the FM regressions show that market β does not help explain average stock returns for 1963–1990. In a shot straight at the heart of the SLB model, the average slope from the regressions of returns on β alone in Table III is 0.15% per month and only 0.46 standard errors from 0. In the regressions of returns on size and β, size has explanatory power (an average slope -3.41 standard errors from 0), but the average slope for β is negative and only 1.21 standard errors from 0. Lakonishok and Shapiro (1986) get similar results for NYSE stocks for 1962–1981. We can also report that β shows no power to explain average returns (the average slopes are typically less than 1 standard error from 0) in FM regressions that use various combinations of β with size, book-to-market equity, leverage, and E/P.

C. Can β Be Saved?

What explains the poor results for β? One possibility is that other explanatory variables are correlated with true βs, and this obscures the relation between average returns and measured βs. But this line of attack cannot explain why β has no power when used alone to explain average returns. Moreover, leverage, book-to-market equity, and E/P do not seem to be good proxies for β. The averages of the monthly cross-sectional correlations between β and the values of these variables for individual stocks are all within 0.15 of 0.

Another hypothesis is that, as predicted by the SLB model, there is a positive relation between β and average return, but the relation is obscured by noise in the β estimates. However, our full-period post-ranking βs do not seem to be imprecise. Most of the standard errors of the βs (not shown) are

Table III

Average Slopes (t-Statistics) from Month-by-Month Regressions of Stock Returns on β, Size, Book-to-Market Equity, Leverage, and E/P: July 1963 to December 1990

Stocks are assigned the post-ranking β of the size-β portfolio they are in at the end of June of year t (Table I). BE is the book value of common equity plus balance-sheet deferred taxes, A is total book assets, and E is earnings (income before extraordinary items, plus income-statement deferred taxes, minus preferred dividends). BE, A, and E are for each firm's latest fiscal year ending in calendar year $t - 1$. The accounting ratios are measured using market equity ME in December of year $t - 1$. Firm size ln(ME) is measured in June of year t. In the regressions, these values of the explanatory variables for individual stocks are matched with CRSP returns for the months from July of year t to June of year $t + 1$. The gap between the accounting data and the returns ensures that the accounting data are available prior to the returns. If earnings are positive, E(+)/P is the ratio of total earnings to market equity and E/P dummy is 0. If earnings are negative, E(+)/P is 0 and E/P dummy is 1.

The average slope is the time-series average of the monthly regression slopes for July 1963 to December 1990, and the t-statistic is the average slope divided by its time-series standard error.

On average, there are 2267 stocks in the monthly regressions. To avoid giving extreme observations heavy weight in the regressions, the smallest and largest 0.5% of the observations on E(+)/P, BE/ME, A/ME, and A/BE are set equal to the next largest or smallest values of the ratios (the 0.005 and 0.995 fractiles). This has no effect on inferences.

β	ln(ME)	ln(BE/ME)	ln(A/ME)	ln(A/BE)	E/P Dummy	E(+)/P
0.15 (0.46)						
	−0.15 (−2.58)					
−0.37 (−1.21)	−0.17 (−3.41)					
		0.50 (5.71)				
			0.50 (5.69)	−0.57 (−5.34)		
					0.57 (2.28)	4.72 (4.57)
	−0.11 (−1.99)	0.35 (4.44)				
	−0.11 (−2.06)		0.35 (4.32)	−0.50 (−4.56)		
	−0.16 (−3.06)				0.06 (0.38)	2.99 (3.04)
	−0.13 (−2.47)	0.33 (4.46)			−0.14 (−0.90)	0.87 (1.23)
	−0.13 (−2.47)		0.32 (4.28)	−0.46 (−4.45)	−0.08 (−0.56)	1.15 (1.57)

0.05 or less, only 1 is greater than 0.1, and the standard errors are small relative to the range of the βs (0.53 to 1.79).

The β-sorted portfolios in Tables I and II also provide strong evidence against the β-measurement-error story. When portfolios are formed on pre-ranking βs alone (Table II), the post-ranking βs for the portfolios almost perfectly reproduce the ordering of the pre-ranking βs. Only the β for portfolio 1B is out of line, and only by 0.02. Similarly, when portfolios are formed on size and then pre-ranking βs (Table I), the post-ranking βs in each size decile closely reproduce the ordering of the pre-ranking βs.

The correspondence between the ordering of the pre-ranking and post-ranking βs for the β-sorted portfolios in Tables I and II is evidence that the post-ranking βs are informative about the ordering of the true βs. The problem for the SLB model is that there is no similar ordering in the average returns on the β-sorted portfolios. Whether one looks at portfolios sorted on β alone (Table II) or on size and then β (Table I), average returns are flat (Table II) or decline slightly (Table I) as the post-ranking βs increase.

Our evidence on the robustness of the size effect and the absence of a relation between β and average return is so contrary to the SLB model that it behooves us to examine whether the results are special to 1963–1990. The appendix shows that NYSE returns for 1941–1990 behave like the NYSE, AMEX, and NASDAQ returns for 1963–1990; there is a reliable size effect over the full 50-year period, but little relation between β and average return. Interestingly, there is a reliable simple relation between β and average return during the 1941–1965 period. These 25 years are a major part of the samples in the early studies of the SLB model of Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973). Even for the 1941–1965 period, however, the relation between β and average return disappears when we control for size.

III. Book-to-Market Equity, E/P, and Leverage

Tables I to III say that there is a strong relation between the average returns on stocks and size, but there is no reliable relation between average returns and β. In this section we show that there is also a strong cross-sectional relation between average returns and book-to-market equity. If anything, this book-to-market effect is more powerful than the size effect. We also find that the combination of size and book-to-market equity absorbs the apparent roles of leverage and E/P in average stock returns.

A. Average Returns

Table IV shows average returns for July 1963 to December 1990 for portfolios formed on ranked values of book-to-market equity (BE/ME) or earnings-price ratio (E/P). The BE/ME and E/P portfolios in Table IV are formed in the same general way (one-dimensional yearly sorts) as the size and β portfolios in Table II. (See the tables for details.)

The relation between average return and E/P has a familiar U-shape (e.g., Jaffe, Keim, and Westerfield (1989) for U.S. data, and Chan, Hamao, and Lakonishok (1991) for Japan). Average returns decline from 1.46% per month for the negative E/P portfolio to 0.93% for the firms in portfolio 1B that have low but positive E/P. Average returns then increase monotonically, reaching 1.72% per month for the highest E/P portfolio.

The more striking evidence in Table IV is the strong positive relation between average return and book-to-market equity. Average returns rise from 0.30% for the lowest BE/ME portfolio to 1.83% for the highest, a difference of 1.53% per month. This spread is twice as large as the difference of 0.74% between the average monthly returns on the smallest and largest size portfolios in Table II. Note also that the strong relation between book-to-market equity and average return is unlikely to be a β effect in disguise; Table IV shows that post-ranking market βs vary little across portfolios formed on ranked values of BE/ME.

On average, only about 50 (out of 2317) firms per year have negative book equity, BE. The negative BE firms are mostly concentrated in the last 14 years of the sample, 1976–1989, and we do not include them in the tests. We can report, however, that average returns for negative BE firms are high, like the average returns of high BE/ME firms. Negative BE (which results from persistently negative earnings) and high BE/ME (which typically means that stock prices have fallen) are both signals of poor earning prospects. The similar average returns of negative and high BE/ME firms are thus consistent with the hypothesis that book-to-market equity captures cross-sectional variation in average returns that is related to relative distress.

B. Fama-MacBeth Regressions

B.1. BE/ME

The FM regressions in Table III confirm the importance of book-to-market equity in explaining the cross-section of average stock returns. The average slope from the monthly regressions of returns on ln(BE/ME) alone is 0.50%, with a t-statistic of 5.71. This book-to-market relation is stronger than the size effect, which produces a t-statistic of -2.58 in the regressions of returns on ln(ME) alone. But book-to-market equity does not replace size in explaining average returns. When both ln(ME) and ln(BE/ME) are included in the regressions, the average size slope is still -1.99 standard errors from 0; the book-to-market slope is an impressive 4.44 standard errors from 0.

B.2. Leverage

The FM regressions that explain returns with leverage variables provide interesting insight into the relation between book-to-market equity and average return. We use two leverage variables, the ratio of book assets to market equity, A/ME, and the ratio of book assets to book equity, A/BE. We interpret A/ME as a measure of market leverage, while A/BE is a measure

Table IV

Properties of Portfolios Formed on Book-to-Market Equity (BE/ME) and Earnings-Price Ratio (E/P): July 1963 to December 1990

At the end of each year $t - 1$, 12 portfolios are formed on the basis of ranked values of BE/ME or E/P. Portfolios 2–9 cover deciles of the ranking variables. The bottom and top 2 portfolios (1A, 1B, 10A, and 10B) split the bottom and top deciles in half. For E/P, there are 13 portfolios; portfolio 0 is stocks with negative E/P. Since BE/ME and E/P are not strongly related to exchange listing, their portfolio breakpoints are determined on the basis of the ranked values of the variables for all stocks that satisfy the CRSP-COMPUSTAT data requirements. BE is the book value of common equity plus balance-sheet deferred taxes, A is total book assets, and E is earnings (income before extraordinary items, plus income-statement deferred taxes, minus preferred dividends). BE, A, and E are for each firm's latest fiscal year ending in calendar year $t - 1$. The accounting ratios are measured using market equity ME in December of year $t - 1$. Firm size ln(ME) is measured in June of year t, with ME denominated in millions of dollars. We calculate each portfolio's monthly equal-weighted return for July of year t to June of year $t + 1$, and then reform the portfolios at the end of year t.

Return is the time-series average of the monthly equal-weighted portfolio returns (in percent). ln(ME), ln(BE/ME), ln(A/ME), ln(A/BE), E(+)/P, and E/P dummy are the time-series averages of the monthly average values of these variables in each portfolio. Since the E/P dummy is 0 when earnings are positive, and 1 when earnings are negative, E/P dummy gives the average proportion of stocks with negative earnings in each portfolio.

β is the time-series average of the monthly portfolio βs. Stocks are assigned the post-ranking β of the size-β portfolio they are in at the end of June of year t (Table I). These individual-firm βs are averaged to compute the monthly βs for each portfolio for July of year t to June of year $t + 1$.

Firms is the average number of stocks in the portfolio each month.

Portfolio	0	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Panel A: Stocks Sorted on Book-to-Market Equity (BE/ME)													
Return		0.30	0.67	0.87	0.97	1.04	1.17	1.30	1.44	1.50	1.59	1.92	1.83
β		1.36	1.34	1.32	1.30	1.28	1.27	1.27	1.27	1.27	1.29	1.33	1.35
ln(ME)		4.53	4.67	4.69	4.56	4.47	4.38	4.23	4.06	3.85	3.51	3.06	2.65
ln(BE/ME)		−2.22	−1.51	−1.09	−0.75	−0.51	−0.32	−0.14	0.03	0.21	0.42	0.66	1.02
ln(A/ME)		−1.24	−0.79	−0.40	−0.05	0.20	0.40	0.56	0.71	0.91	1.12	1.35	1.75
ln(A/BE)		0.94	0.71	0.68	0.70	0.71	0.71	0.70	0.68	0.70	0.70	0.70	0.73
E/P dummy		0.29	0.15	0.10	0.08	0.08	0.08	0.09	0.09	0.11	0.15	0.22	0.36
E(+)/P		0.03	0.04	0.06	0.08	0.09	0.10	0.11	0.11	0.12	0.12	0.11	0.10
Firms		89	98	209	222	226	230	235	237	239	239	120	117

Table IV—*Continued*

Portfolio	0	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Panel B: Stocks Sorted on Earnings-Price Ratio (E/P)													
Return	1.46	1.04	0.93	0.94	1.03	1.18	1.22	1.33	1.42	1.46	1.57	1.74	1.72
β	1.47	1.40	1.35	1.31	1.28	1.26	1.25	1.26	1.24	1.23	1.24	1.28	1.31
ln(ME)	2.48	3.64	4.33	4.61	4.64	4.63	4.58	4.49	4.37	4.28	4.07	3.82	3.52
ln(BE/ME)	−0.10	−0.76	−0.91	−0.79	−0.61	−0.47	−0.33	−0.21	−0.08	0.02	0.15	0.26	0.40
ln(A/ME)	0.90	−0.05	−0.27	−0.16	0.03	0.18	0.31	0.44	0.58	0.70	0.85	1.01	1.25
ln(A/BE)	0.99	0.70	0.63	0.63	0.64	0.65	0.64	0.65	0.66	0.68	0.71	0.75	0.86
E/P dummy	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E(+)/P	0.00	0.01	0.03	0.05	0.06	0.08	0.09	0.11	0.12	0.14	0.16	0.20	0.28
Firms	355	88	90	182	190	193	196	194	197	195	195	95	91

of book leverage. The regressions use the natural logs of the leverage ratios, ln(A/ME) and ln(A/BE), because preliminary tests indicated that logs are a good functional form for capturing leverage effects in average returns. Using logs also leads to a simple interpretation of the relation between the roles of leverage and book-to-market equity in average returns.

The FM regressions of returns on the leverage variables (Table III) pose a bit of a puzzle. The two leverage variables are related to average returns, but with opposite signs. As in Bhandari (1988), higher market leverage is associated with higher average returns; the average slopes for ln(A/ME) are always positive and more than 4 standard errors from 0. But higher book leverage is associated with lower average returns; the average slopes for ln(A/BE) are always negative and more than 4 standard errors from 0.

The puzzle of the opposite slopes on ln(A/ME) and ln(A/BE) has a simple solution. The average slopes for the two leverage variables are opposite in sign but close in absolute value, e.g., 0.50 and -0.57. Thus it is the difference between market and book leverage that helps explain average returns. But the difference between market and book leverage is book-to-market equity, $\ln(BE/ME) = \ln(A/ME) - \ln(A/BE)$. Table III shows that the average book-to-market slopes in the FM regressions are indeed close in absolute value to the slopes for the two leverage variables.

The close links between the leverage and book-to-market results suggest that there are two equivalent ways to interpret the book-to-market effect in average returns. A high ratio of book equity to market equity (a low stock price relative to book value) says that the market judges the prospects of a firm to be poor relative to firms with low BE/ME. Thus BE/ME may capture the relative-distress effect postulated by Chan and Chen (1991). A high book-to-market ratio also says that a firm's market leverage is high relative to its book leverage; the firm has a large amount of market-imposed leverage because the market judges that its prospects are poor and discounts its stock price relative to book value. In short, our tests suggest that the relative-distress effect, captured by BE/ME, can also be interpreted as an involuntary leverage effect, which is captured by the difference between A/ME and A/BE.

B.3. E/P

Ball (1978) posits that the earnings-price ratio is a catch-all for omitted risk factors in expected returns. If current earnings proxy for expected future earnings, high-risk stocks with high expected returns will have low prices relative to their earnings. Thus, E/P should be related to expected returns, whatever the omitted sources of risk. This argument only makes sense, however, for firms with positive earnings. When current earnings are negative, they are not a proxy for the earnings forecasts embedded in the stock price, and E/P is not a proxy for expected returns. Thus, the slope for E/P in the FM regressions is based on positive values; we use a dummy variable for E/P when earnings are negative.

The U-shaped relation between average return and E/P observed in Table IV is also apparent when the E/P variables are used alone in the FM regressions in Table III. The average slope on the E/P dummy variable (0.57% per month, 2.28 standard errors from 0) confirms that firms with negative earnings have higher average returns. The average slope for stocks with positive E/P (4.72% per month, 4.57 standard errors from 0) shows that average returns increase with E/P when it is positive.

Adding size to the regressions kills the explanatory power of the E/P dummy. Thus the high average returns of negative E/P stocks are better captured by their size, which Table IV says is on average small. Adding both size and book-to-market equity to the E/P regressions kills the E/P dummy and lowers the average slope on E/P from 4.72 to 0.87 (t = 1.23). In contrast, the average slopes for ln(ME) and ln(BE/ME) in the regressions that include E/P are similar to those in the regressions that explain average returns with only size and book-to-market equity. The results suggest that most of the relation between (positive) E/P and average return is due to the positive correlation between E/P and ln(BE/ME), illustrated in Table IV; firms with high E/P tend to have high book-to-market equity ratios.

IV. A Parsimonious Model for Average Returns

The results to here are easily summarized:

(1) When we allow for variation in β that is unrelated to size, there is no reliable relation between β and average return.
(2) The opposite roles of market leverage and book leverage in average returns are captured well by book-to-market equity.
(3) The relation between E/P and average return seems to be absorbed by the combination of size and book-to-market equity.

In a nutshell, market β seems to have no role in explaining the average returns on NYSE, AMEX, and NASDAQ stocks for 1963–1990, while size and book-to-market equity capture the cross-sectional variation in average stock returns that is related to leverage and E/P.

A. Average Returns, Size and Book-to-Market Equity

The average return matrix in Table V gives a simple picture of the two-dimensional variation in average returns that results when the 10 size deciles are each subdivided into 10 portfolios based on ranked values of BE/ME for individual stocks. Within a size decile (across a row of the average return matrix), returns typically increase strongly with BE/ME: on average, the returns on the lowest and highest BE/ME portfolios in a size decile differ by 0.99% (1.63% – 0.64%) per month. Similarly, looking down the columns of the average return matrix shows that there is a negative relation between average return and size: on average, the spread of returns across the size portfolios in a BE/ME group is 0.58% per month. The average return matrix gives life to the conclusion from the regressions that,

Table V

Average Monthly Returns on Portfolios Formed on Size and Book-to-Market Equity; Stocks Sorted by ME (Down) and then BE/ME (Across): July 1963 to December 1990

In June of each year t, the NYSE, AMEX, and NASDAQ stocks that meet the CRSP-COMPUSTAT data requirements are allocated to 10 size portfolios using the NYSE size (ME) breakpoints. The NYSE, AMEX, and NASDAQ stocks in each size decile are then sorted into 10 BE/ME portfolios using the book-to-market ratios for year $t-1$. BE/ME is the book value of common equity plus balance-sheet deferred taxes for fiscal year $t-1$, over market equity for December of year $t-1$. The equal-weighted monthly portfolio returns are then calculated for July of year t to June of year $t+1$.

Average monthly return is the time-series average of the monthly equal-weighted portfolio returns (in percent).

The All column shows average returns for equal-weighted size decile portfolios. The All row shows average returns for equal-weighted portfolios of the stocks in each BE/ME group.

	Book-to-Market Portfolios										
	All	Low	2	3	4	5	6	7	8	9	High
All	1.23	0.64	0.98	1.06	1.17	1.24	1.26	1.39	1.40	1.50	1.63
Small-ME	1.47	0.70	1.14	1.20	1.43	1.56	1.51	1.70	1.71	1.82	1.92
ME-2	1.22	0.43	1.05	0.96	1.19	1.33	1.19	1.58	1.28	1.43	1.79
ME-3	1.22	0.56	0.88	1.23	0.95	1.36	1.30	1.30	1.40	1.54	1.60
ME-4	1.19	0.39	0.72	1.06	1.36	1.13	1.21	1.34	1.59	1.51	1.47
ME-5	1.24	0.88	0.65	1.08	1.47	1.13	1.43	1.44	1.26	1.52	1.49
ME-6	1.15	0.70	0.98	1.14	1.23	0.94	1.27	1.19	1.19	1.24	1.50
ME-7	1.07	0.95	1.00	0.99	0.83	0.99	1.13	0.99	1.16	1.10	1.47
ME-8	1.08	0.66	1.13	0.91	0.95	0.99	1.01	1.15	1.05	1.29	1.55
ME-9	0.95	0.44	0.89	0.92	1.00	1.05	0.93	0.82	1.11	1.04	1.22
Large-ME	0.89	0.93	0.88	0.84	0.71	0.79	0.83	0.81	0.96	0.97	1.18

controlling for size, book-to-market equity captures strong variation in average returns, and controlling for book-to-market equity leaves a size effect in average returns.

B. The Interaction between Size and Book-to-Market Equity

The average of the monthly correlations between the cross-sections of ln(ME) and ln(BE/ME) for individual stocks is -0.26. The negative correlation is also apparent in the average values of ln(ME) and ln(BE/ME) for the portfolios sorted on ME or BE/ME in Tables II and IV. Thus, firms with low market equity are more likely to have poor prospects, resulting in low stock prices and high book-to-market equity. Conversely, large stocks are more likely to be firms with stronger prospects, higher stock prices, lower book-to-market equity, and lower average stock returns.

The correlation between size and book-to-market equity affects the regressions in Table III. Including ln(BE/ME) moves the average slope on ln(ME) from -0.15 ($t = -2.58$) in the univariate regressions to -0.11 ($t = -1.99$) in the bivariate regressions. Similarly, including ln(ME) in the regressions lowers the average slope on ln(BE/ME) from 0.50 to 0.35 (still a healthy 4.44 standard errors from 0). Thus, part of the size effect in the simple regressions is due to the fact that small ME stocks are more likely to have high book-to-market ratios, and part of the simple book-to-market effect is due to the fact that high BE/ME stocks tend to be small (they have low ME).

We should not, however, exaggerate the links between size and book-to-market equity. The correlation (-0.26) between ln(ME) and ln(BE/ME) is not extreme, and the average slopes in the bivariate regressions in Table III show that ln(ME) and ln(BE/ME) are both needed to explain the cross-section of average returns. Finally, the 10×10 average return matrix in Table V provides concrete evidence that, (a) controlling for size, book-to-market equity captures substantial variation in the cross-section of average returns, and (b) within BE/ME groups average returns are related to size.

C. Subperiod Averages of the FM Slopes

The message from the average FM slopes for 1963–1990 (Table III) is that size on average has a negative premium in the cross-section of stock returns, book-to-market equity has a positive premium, and the average premium for market β is essentially 0. Table VI shows the average FM slopes for two roughly equal subperiods (July 1963–December 1976 and January 1977–December 1990) from two regressions: (a) the cross-section of stock returns on size, ln(ME), and book-to-market equity, ln(BE/ME), and (b) returns on β, ln(ME), and ln(BE/ME). For perspective, average returns on the value-weighted and equal-weighted (VW and EW) portfolios of NYSE stocks are also shown.

In FM regressions, the intercept is the return on a standard portfolio (the weights on stocks sum to 1) in which the weighted averages of the explanatory variables are 0 (Fama (1976), chapter 9). In our tests, the intercept is weighted toward small stocks (ME is in millions of dollars so $\ln(\text{ME}) = 0$ implies ME = \$1 million) and toward stocks with relatively high book-to-market ratios (Table IV says that ln(BE/ME) is negative for the typical firm, so $\ln(\text{BE/ME}) = 0$ is toward the high end of the sample ratios). Thus it is not surprising that the average intercepts are always large relative to their standard errors and relative to the returns on the NYSE VW and EW portfolios.

Like the overall period, the subperiods do not offer much hope that the average premium for β is economically important. The average FM slope for β is only slightly positive for 1963–1976 (0.10% per month, $t = 0.25$), and it is negative for 1977–1990 (-0.44% per month, $t = -1.17$). There is a hint that the size effect is weaker in the 1977–1990 period, but inferences about the average size slopes for the subperiods lack power.

Unlike the size effect, the relation between book-to-market equity and average return is so strong that it shows up reliably in both the 1963–1976 and the 1977–1990 subperiods. The average slopes for ln(BE/ME) are all more than 2.95 standard errors from 0, and the average slopes for the

Table VI

Subperiod Average Monthly Returns on the NYSE Equal-Weighted and Value-Weighted Portfolios and Subperiod Means of the Intercepts and Slopes from the Monthly FM Cross-Sectional Regressions of Returns on (a) Size (ln(ME)) and Book-to-Market Equity (ln(BE/ME)), and (b) β, ln(ME), and ln(BE/ME)

Mean is the time-series mean of a monthly return, Std is its time-series standard deviation, and t(Mn) is Mean divided by its time-series standard error.

	7/63–12/90 (330 Mos.)			7/63–12/76 (162 Mos.)			1/77–12/90 (168 Mos.)		
Variable	Mean	Std	t(Mn)	Mean	Std	t(Mn)	Mean	Std	t(Mn)
NYSE Value-Weighted (VW) and Equal-Weighted (EW) Portfolio Returns									
VW	0.81	4.47	3.27	0.56	4.26	1.67	1.04	4.66	2.89
EW	0.97	5.49	3.19	0.77	5.70	1.72	1.15	5.28	2.82
$R_{it} = \mathrm{a} + \mathrm{b}_{2t}\ln(\mathrm{ME}_{it}) + \mathrm{b}_{3t}\ln(\mathrm{BE/ME}_{it}) + e_{it}$									
a	1.77	8.51	3.77	1.86	10.10	2.33	1.69	6.67	3.27
b_2	−0.11	1.02	−1.99	−0.16	1.25	−1.62	−0.07	0.73	−1.16
b_3	0.35	1.45	4.43	0.36	1.53	2.96	0.35	1.37	3.30
$R_{it} = \mathrm{a} + \mathrm{b}_{1t}\beta_{it} + \mathrm{b}_{2t}\ln(\mathrm{ME}_{it}) + \mathrm{b}_{3t}\ln(\mathrm{BE/ME}_{it}) + e_{it}$									
a	2.07	5.75	6.55	1.73	6.22	3.54	2.40	5.25	5.92
b_1	−0.17	5.12	−0.62	0.10	5.33	0.25	−0.44	4.91	−1.17
b_2	−0.12	0.89	−2.52	−0.15	1.03	−1.91	−0.09	0.74	−1.64
b_3	0.33	1.24	4.80	0.34	1.36	3.17	0.31	1.10	3.67

subperiods (0.36 and 0.35) are close to the average slope (0.35) for the overall period. The subperiod results thus support the conclusion that, among the variables considered here, book-to-market equity is consistently the most powerful for explaining the cross-section of average stock returns.

Finally, Roll (1983) and Keim (1983) show that the size effect is stronger in January. We have examined the monthly slopes from the FM regressions in Table VI for evidence of a January seasonal in the relation between book-to-market equity and average return. The average January slopes for ln(BE/ME) are about twice those for February to December. Unlike the size effect, however, the strong relation between book-to-market equity and average return is not special to January. The average monthly February-to-December slopes for ln(BE/ME) are about 4 standard errors from 0, and they are close to (within 0.05 of) the average slopes for the whole year. Thus, there is a January seasonal in the book-to-market equity effect, but the positive relation between BE/ME and average return is strong throughout the year.

D. β and the Market Factor: Caveats

Some caveats about the negative evidence on the role of β in average returns are in order. The average premiums for β, size, and book-to-market equity depend on the definitions of the variables used in the regressions. For example, suppose we replace book-to-market equity (ln(BE/ME)) with book equity (ln(BE)). As long as size (ln(ME)) is also in the regression, this change will not affect the intercept, the fitted values or the R^2. But the change, in variables increases the average slope (and the t-statistic) on ln(ME). In other words, it increases the risk premium associated with size. Other redefinitions of the β, size, and book-to-market variables will produce different regression slopes and perhaps different inferences about average premiums, including possible resuscitation of a role for β. And, of course, at the moment, we have no theoretical basis for choosing among different versions of the variables.

Moreover, the tests here are restricted to stocks. It is possible that including other assets will change the inferences about the average premiums for β, size, and book-to-market equity. For example, the large average intercepts for the FM regressions in Table VI suggest that the regressions will not do a good job on Treasury bills, which have low average returns and are likely to have small loadings on the underlying market, size, and book-to-market factors in returns. Extending the tests to bills and other bonds may well change our inferences about average risk premiums, including the revival of a role for market β.

We emphasize, however, that different approaches to the tests are not likely to revive the Sharpe-Lintner-Black model. Resuscitation of the SLB model requires that a better proxy for the market portfolio (a) overturns our evidence that the simple relation between β and average stock returns is flat and (b) leaves β as the only variable relevant for explaining average returns. Such results seem unlikely, given Stambaugh's (1982) evidence that tests of the SLB model do not seem to be sensitive to the choice of a market proxy. Thus, if there is a role for β in average returns, it is likely to be found in a multi-factor model that transforms the flat simple relation between average return and β into a positively sloped conditional relation.

V. Conclusions and Implications

The Sharpe-Lintner-Black model has long shaped the way academics and practitioners think about average return and risk. Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) find that, as predicted by the model, there is a positive simple relation between average return and market β during the early years (1926–1968) of the CRSP NYSE returns file. Like Reinganum (1981) and Lakonishok and Shapiro (1986), we find that this simple relation between β and average return disappears during the more recent 1963–1990 period. The appendix that follows shows that the relation between β and average return is also weak in the last half century (1941–1990) of returns on NYSE stocks. In short, our tests do not support the central prediction of the SLB model, that average stock returns are positively related to market β.

Banz (1981) documents a strong negative relation between average return and firm size. Bhandari (1988) finds that average return is positively related to leverage, and Basu (1983) finds a positive relation between average return and E/P. Stattman (1980) and Rosenberg, Reid, and Lanstein (1985) document a positive relation between average return and book-to-market equity for U.S. stocks, and Chan, Hamao, and Lakonishok (1992) find that BE/ME is also a powerful variable for explaining average returns on Japanese stocks.

Variables like size, E/P, leverage, and book-to-market equity are all scaled versions of a firm's stock price. They can be regarded as different ways of extracting information from stock prices about the cross-section of expected stock returns (Ball (1978); Keim (1988)). Since all these variables are scaled versions of price, it is reasonable to expect that some of them are redundant for explaining average returns. Our main result is that for the 1963–1990 period, size and book-to-market equity capture the cross-sectional variation in average stock returns associated with size, E/P, book-to-market equity, and leverage.

A. *Rational Asset-Pricing Stories*

Are our results consistent with asset-pricing theory? Since the FM intercept is constrained to be the same for all stocks, FM regressions always impose a linear factor structure on returns and expected returns that is consistent with the multifactor asset-pricing models of Merton (1973) and Ross (1976). Thus our tests impose a rational asset-pricing framework on the relation between average return and size and book-to-market equity.

Even if our results are consistent with asset-pricing theory, they are not economically satisfying. What is the economic explanation for the roles of size and book-to-market equity in average returns? We suggest several paths of inquiry.

(a) The intercepts and slopes in the monthly FM regressions of returns on ln(ME) and ln(BE/ME) are returns on portfolios that mimic the underlying common risk factors in returns proxied by size and book-to-market equity (Fama (1976), chapter 9). Examining the relations between the returns on these portfolios and economic variables that measure variation in business conditions might help expose the nature of the economic risks captured by size and book-to-market equity.

(b) Chan, Chen, and Hsieh (1985) argue that the relation between size and average return proxies for a more fundamental relation between expected returns and economic risk factors. Their most powerful factor in explaining the size effect is the difference between the monthly returns on low- and high-grade corporate bonds, which in principle captures a kind of default risk in returns that is priced. It would be interesting to test whether loadings on this or other economic factors, such as those of Chen, Roll, and Ross (1986), can explain the roles of size and book-to-market equity in our tests.

(c) In a similar vein, Chan and Chen (1991) argue that the relation between size and average return is a relative-prospects effect. The earning prospects of distressed firms are more sensitive to economic conditions. This results in a distress factor in returns that is priced in expected returns. Chan and Chen construct two mimicking portfolios for the distress factor, based on dividend changes and leverage. It would be interesting to check whether loadings on their distress factors absorb the size and book-to-market equity effects in average returns that are documented here.

(d) In fact, if stock prices are rational, BE/ME, the ratio of the book value of a stock to the market's assessment of its value, should be a direct indicator of the relative prospects of firms. For example, we expect that high BE/ME firms have low earnings on assets relative to low BE/ME firms. Our work (in progress) suggests that there is indeed a clean separation between high and low BE/ME firms on various measures of economic fundamentals. Low BE/ME firms are persistently strong performers, while the economic performance of high BE/ME firms is persistently weak.

B. Irrational Asset-Pricing Stories

The discussion above assumes that the asset-pricing effects captured by size and book-to-market equity are rational. For BE/ME, our most powerful expected-return variable, there is an obvious alternative. The cross-section of book-to-market ratios might result from market overreaction to the relative prospects of firms. If overreaction tends to be corrected, BE/ME will predict the cross-section of stock returns.

Simple tests do not confirm that the size and book-to-market effects in average returns are due to market overreaction, at least of the type posited by DeBondt and Thaler (1985). One overreaction measure used by DeBondt and Thaler is a stock's most recent 3-year return. Their overreaction story predicts that 3-year losers have strong post-ranking returns relative to 3-year winners. In FM regressions (not shown) for individual stocks, the 3-year lagged return shows no power even when used alone to explain average returns. The univariate average slope for the lagged return is negative, -6 basis points per month, but less than 0.5 standard errors from 0.

C. Applications

Our main result is that two easily measured variables, size and book-to-market equity, seem to describe the cross-section of average stock returns. Prescriptions for using this evidence depend on (a) whether it will persist, and (b) whether it results from rational or irrational asset-pricing.

It is possible that, by chance, size and book-to-market equity happen to describe the cross-section of average returns in our sample, but they were and are unrelated to expected returns. We put little weight on this possibility, especially for book-to-market equity. First, although BE/ME has long been touted as a measure of the return prospects of stocks, there is no evidence that its explanatory power deteriorates through time. The 1963–1990 relation between BE/ME and average return is strong, and remarkably similar for the 1963–1976 and 1977–1990 subperiods. Second, our preliminary work on economic fundamentals suggests that high-BE/ME firms tend to be persistently poor earners relative to low-BE/ME firms. Similarly, small firms have a long period of poor earnings during the 1980s not shared with big firms. The systematic patterns in fundamentals give us some hope that size and book-to-market equity proxy for risk factors in returns, related to relative earning prospects, that are rationally priced in expected returns.

If our results are more than chance, they have practical implications for portfolio formation and performance evaluation by investors whose primary concern is long-term average returns. If asset-pricing is rational, size and BE/ME must proxy for risk. Our results then imply that the performance of managed portfolios (e.g., pension funds and mutual funds) can be evaluated by comparing their average returns with the average returns of benchmark portfolios with similar size and BE/ME characteristics. Likewise, the expected returns for different portfolio strategies can be estimated from the historical average returns of portfolios with matching size and BE/ME properties.

If asset-pricing is irrational and size and BE/ME do not proxy for risk, our results might still be used to evaluate portfolio performance and measure the expected returns from alternative investment strategies. If stock prices are irrational, however, the likely persistence of the results is more suspect.

Appendix
Size Versus β: 1941–1990

Our results on the absence of a relation between β and average stock returns for 1963–1990 are so contrary to the tests of the Sharpe-Lintner-Black model by Black, Jensen, and Scholes (1972), Fama and MacBeth (1973), and (more recently) Chan and Chen (1988), that further tests are appropriate. We examine the roles of size and β in the average returns on NYSE stocks for the half-century 1941–1990, the longest available period that avoids the high volatility of returns in the Great Depression. We do not include the accounting variables in the tests because of the strong selection bias (toward successful firms) in the COMPUSTAT data prior to 1962.

We first replicate the results of Chan and Chen (1988). Like them, we find that when portfolios are formed on size alone, there are strong relations between average return and either size or β; average return increases with β and decreases with size. For size portfolios, however, size (ln(ME)) and β are almost perfectly correlated (-0.98), so it is difficult to distinguish between the roles of size and β in average returns.

One way to generate strong variation in β that is unrelated to size is to form portfolios on size and then on β. As in Tables I to III, we find that the resulting independent variation in β just about washes out the positive simple relation between average return and β observed when portfolios are formed on size alone. The results for NYSE stocks for 1941–1990 are thus much like those for NYSE, AMEX, and NASDAQ stocks for 1963–1990.

This appendix also has methodological goals. For example, the FM regressions in Table III use returns on individual stocks as the dependent variable. Since we allocate portfolio βs to individual stocks but use firm-specific values of other variables like size, β may be at a disadvantage in the regressions for individual stocks. This appendix shows, however, that regressions for portfolios, which put β and size on equal footing, produce results comparable to those for individual stocks.

A. Size Portfolios

Table AI shows average monthly returns and market βs for 12 portfolios of NYSE stocks formed on the basis of size (ME) at the end of each year from 1940 to 1989. For these size portfolios, there is a strong positive relation between average return and β. Average returns fall from 1.96% per month for the smallest ME portfolio (1A) to 0.93% for the largest (10B) and β falls from 1.60 to 0.95. (Note also that, as claimed earlier, estimating β as the sum of the slopes in the regression of a portfolio's return on the current and prior month's NYSE value-weighted return produces much larger βs for the smallest ME portfolios and slightly smaller βs for the largest ME portfolios.)

The FM regressions in Table AI confirm the positive simple relation between average return and β for size portfolios. In the regressions of the size-portfolio returns on β alone, the average premium for a unit of β is 1.45% per month. In the regressions of individual stock returns on β (where stocks are assigned the β of their size portfolio), the premium for a unit of β is 1.39%. Both estimates are about 3 standard errors from 0. Moreover, the βs of size portfolios do not leave a residual size effect; the average residuals from the simple regressions of returns on β in Table AI show no relation to size. These positive SLB results for 1941–1990 are like those obtained by Chan and Chen (1988) in tests on size portfolios for 1954–1983.

There is, however, evidence in Table AI that all is not well with the βs of the size portfolios. They do a fine job on the relation between size and average return, but they do a lousy job on their main task, the relation between β and average return. When the residuals from the regressions of returns on β are grouped using the pre-ranking βs of individual stocks, the average residuals are strongly positive for low-β stocks (0.51% per month for group 1A) and negative for high-β stocks (−1.05% for 10B). Thus the market lines estimated with size-portfolio βs exaggerate the tradeoff of average return for β; they underestimate average returns on low-β stocks and overestimate average returns on high-β stocks. This pattern in the β-sorted average residuals for individual stocks suggests that (a) there is variation in β across stocks that is lost in the size portfolios, and (b) this variation in β is not rewarded as well as the variation in β that is related to size.

B. Two-Pass Size-β Portfolios

Like Table I, Table AII shows that subdividing size deciles using the (pre-ranking) βs of individual stocks results in strong variation in β that is

Table AI

Average Returns, Post-Ranking βs and Fama-MacBeth Regression Slopes for Size Portfolios of NYSE Stocks: 1941–1990

At the end of each year $t - 1$, stocks are assigned to 12 portfolios using ranked values of ME. Included are all NYSE stocks that have a CRSP price and shares for December of year $t - 1$ and returns for at least 24 of the 60 months ending in December of year $t - 1$ (for pre-ranking β estimates). The middle 8 portfolios cover size deciles 2 to 9. The 4 extreme portfolios (1A, 1B, 10A, and 10B) split the smallest and largest deciles in half. We compute equal-weighted returns on the portfolios for the 12 months of year t using all surviving stocks. Average Return is the time-series average of the monthly portfolio returns for 1941–1990, in percent. Average firms is the average number of stocks in the portfolios each month. The simple βs are estimated by regressing the 1941–1990 sample of post-ranking monthly returns for a size portfolio on the current month's value-weighted NYSE portfolio return. The sum βs are the sum of the slopes from a regression of the post-ranking monthly returns on the current and prior month's VW NYSE returns.

The independent variables in the Fama-MacBeth regressions are defined for each firm at the end of December of each year $t - 1$. Stocks are assigned the post-ranking (sum) β of the size portfolio they are in at the end of year $t - 1$. ME is price times shares outstanding at the end of year $t - 1$. In the individual-stock regressions, these values of the explanatory variables are matched with CRSP returns for each of the 12 months of year t. The portfolio regressions match the equal-weighted portfolio returns with the equal-weighted averages of β and ln(ME) for the surviving stocks in each month of year t. Slope is the average of the (600) monthly FM regression slopes and SE is the standard error of the average slope. The residuals from the monthly regressions for year t are grouped into 12 portfolios on the basis of size (ME) or pre-ranking β (estimated with 24 to 60 months of data, as available) at the end of year $t - 1$. The average residuals are the time-series averages of the monthly equal-weighted portfolio residuals, in percent. The average residuals for regressions (1) and (2) (not shown) are quite similar to those for regressions (4) and (5) (shown).

	Portfolios Formed on Size											
	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Ave. return	1.96	1.59	1.44	1.36	1.28	1.24	1.23	1.17	1.15	1.13	0.97	0.93
Ave. firms	57	56	110	107	107	108	111	113	115	118	59	59
Simple β	1.29	1.24	1.21	1.19	1.16	1.13	1.13	1.12	1.09	1.05	1.00	0.98
Standard error	0.07	0.05	0.04	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
Sum β	1.60	1.44	1.37	1.32	1.26	1.23	1.19	1.17	1.12	1.06	0.99	0.95
Standard error	0.10	0.06	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.01	0.01	0.01

Table AI—*Continued*

	Portfolio Regressions				Individual Stock Regressions			
	(1) β	(2) ln(ME)	(3) β and ln(ME)		(4) β	(5) ln(ME)	(6) β and ln(ME)	
Slope	1.45	−0.137	3.05	0.149	1.39	−0.133	0.71	−0.060
SE	0.47	0.044	1.51	0.115	0.46	0.043	0.81	0.062

Average Residuals for Stocks Grouped on Size

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Regression (4)	0.17	0.00	−0.04	−0.06	−0.05	−0.04	0.00	−0.03	0.03	0.08	0.01	0.04
Standard error	0.11	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.05	0.06
Regression (5)	0.30	0.02	−0.05	−0.06	−0.08	−0.07	−0.03	−0.04	0.02	0.08	0.01	0.13
Standard error	0.14	0.07	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.07
Regression (6)	0.20	0.02	−0.05	−0.07	−0.08	−0.06	−0.01	−0.02	0.04	0.09	0.00	0.06
Standard error	0.10	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.05	0.05

Average Residuals for Stocks Grouped on Pre-Ranking β

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Regression (4)	0.51	0.61	0.38	0.32	0.16	0.12	0.03	−0.10	−0.27	−0.31	−0.66	−1.05
Standard error	0.21	0.19	0.13	0.08	0.04	0.03	0.04	0.05	0.09	0.11	0.18	0.23
Regression (5)	−0.10	0.00	0.02	0.09	0.05	0.07	0.05	0.00	−0.03	−0.01	−0.11	−0.33
Standard error	0.11	0.10	0.07	0.05	0.04	0.03	0.03	0.04	0.05	0.07	0.10	0.13
Regression (6)	0.09	0.25	0.13	0.19	0.11	0.14	0.09	0.01	−0.11	−0.12	−0.38	−0.70
Standard error	0.41	0.37	0.24	0.14	0.07	0.04	0.04	0.09	0.16	0.21	0.34	0.43

Table AII

Properties of Portfolios Formed on Size and Pre-Ranking β: NYSE Stocks Sorted by ME (Down) then Pre-Ranking β (Across): 1941–1990

At the end of year $t-1$, the NYSE stocks on CRSP are assigned to 10 size (ME) portfolios. Each size decile is subdivided into 10 β portfolios using pre-ranking βs of individual stocks, estimated with 24 to 60 monthly returns (as available) ending in December of year $t-1$. The equal-weighted monthly returns on the resulting 100 portfolios are then calculated for year t. The average returns are the time-series averages of the monthly returns, in percent. The post-ranking βs use the full 1941–1990 sample of post-ranking returns for each portfolio. The pre- and post-ranking βs are the sum of the slopes from a regression of monthly returns on the current and prior month's NYSE value-weighted market return. The average size for a portfolio is the time-series average of each month's average value of ln(ME) for stocks in the portfolio. ME is denominated in millions of dollars. There are, on average, about 10 stocks in each size-β portfolio each month. The All column shows parameter values for equal-weighted size-decile (ME) portfolios. The All rows show parameter values for equal-weighted portfolios of the stocks in each β group.

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
Panel A: Average Monthly Return (in Percent)											
All		1.22	1.30	1.32	1.35	1.36	1.34	1.29	1.34	1.14	1.10
Small-ME	1.78	1.74	1.76	2.08	1.91	1.92	1.72	1.77	1.91	1.56	1.46
ME-2	1.44	1.41	1.35	1.33	1.61	1.72	1.59	1.40	1.62	1.24	1.11
ME-3	1.36	1.21	1.40	1.22	1.47	1.34	1.51	1.33	1.57	1.33	1.21
ME-4	1.28	1.26	1.29	1.19	1.27	1.51	1.30	1.19	1.56	1.18	1.00
ME-5	1.24	1.22	1.30	1.28	1.33	1.21	1.37	1.41	1.31	0.92	1.06
ME-6	1.23	1.21	1.32	1.37	1.09	1.34	1.10	1.40	1.21	1.22	1.08
ME-7	1.17	1.08	1.23	1.37	1.27	1.19	1.34	1.10	1.11	0.87	1.17
ME-8	1.15	1.06	1.18	1.26	1.25	1.26	1.17	1.16	1.05	1.08	1.04
ME-9	1.13	0.99	1.13	1.00	1.24	1.28	1.31	1.15	1.11	1.09	1.05
Large-ME	0.95	0.99	1.01	1.12	1.01	0.89	0.95	0.95	1.00	0.90	0.68

Table AII—*Continued*

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
Panel B: Post-Ranking β											
All		0.76	0.95	1.05	1.14	1.22	1.26	1.34	1.38	1.49	1.69
Small-ME	1.52	1.17	1.40	1.31	1.50	1.46	1.50	1.69	1.60	1.75	1.92
ME-2	1.37	0.86	1.09	1.12	1.24	1.39	1.42	1.48	1.60	1.69	1.91
ME-3	1.32	0.88	0.96	1.18	1.19	1.33	1.40	1.43	1.56	1.64	1.74
ME-4	1.26	0.69	0.95	1.06	1.15	1.24	1.29	1.46	1.43	1.64	1.83
ME-5	1.23	0.70	0.95	1.04	1.10	1.22	1.32	1.34	1.41	1.56	1.72
ME-6	1.19	0.68	0.86	1.04	1.13	1.20	1.20	1.35	1.36	1.48	1.70
ME-7	1.17	0.67	0.88	0.95	1.14	1.18	1.26	1.27	1.32	1.44	1.68
ME-8	1.12	0.64	0.83	0.99	1.06	1.14	1.14	1.21	1.26	1.39	1.58
ME-9	1.06	0.68	0.81	0.94	0.96	1.06	1.11	1.18	1.22	1.25	1.46
Large-ME	0.97	0.65	0.73	0.90	0.91	0.97	1.01	1.01	1.07	1.12	1.38
Panel C: Average Size (ln(ME))											
All		4.39	4.39	4.40	4.40	4.39	4.40	4.38	4.37	4.37	4.34
Small-ME	1.93	2.04	1.99	2.00	1.96	1.92	1.92	1.91	1.90	1.87	1.80
ME-2	2.80	2.81	2.79	2.81	2.83	2.80	2.79	2.80	2.80	2.79	2.79
ME-3	3.27	3.28	3.27	3.28	3.27	3.27	3.28	3.29	3.27	3.27	3.26
ME-4	3.67	3.67	3.67	3.67	3.68	3.68	3.67	3.68	3.66	3.67	3.67
ME-5	4.06	4.07	4.06	4.05	4.06	4.07	4.06	4.05	4.05	4.06	4.06
ME-6	4.45	4.45	4.44	4.46	4.45	4.45	4.45	4.45	4.44	4.45	4.45
ME-7	4.87	4.86	4.87	4.86	4.87	4.87	4.88	4.87	4.87	4.85	4.87
ME-8	5.36	5.38	5.38	5.38	5.35	5.36	5.37	5.37	5.36	5.35	5.34
ME-9	5.98	5.96	5.98	5.99	6.00	5.98	5.98	5.97	5.95	5.96	5.96
Large-ME	7.12	7.10	7.12	7.16	7.17	7.20	7.29	7.14	7.09	7.04	6.83

independent of size. The β sort of a size decile always produces portfolios with similar average ln(ME) but much different (post-ranking) βs. Table AII also shows, however, that investors are not compensated for the variation in β that is independent of size. Despite the wide range of βs in each size decile, average returns show no tendency to increase with β. AII

The FM regressions in Table AIII formalize the roles of size and β in NYSE average returns for 1941–1990. The regressions of returns on β alone show that using the βs of the portfolios formed on size and β, rather than size alone, causes the average slope on β to fall from about 1.4% per month (Table AI) to about 0.23% (about 1 standard error from 0). Thus, allowing for variation in β that is unrelated to size flattens the relation between average return and β, to the point where it is indistinguishable from no relation at all.

The flatter market lines in Table AIII succeed, however, in erasing the negative relation between β and average residuals observed in the regressions of returns on β alone in Table AI. Thus, forming portfolios on size and β (Table AIII) produces a better description of the simple relation between average return and β than forming portfolios on size alone (Table AI). This improved description of the relation between average return and β is evidence that the β estimates for the two-pass size-β portfolios capture variation in true βs that is missed when portfolios are formed on size alone.

Unfortunately, the flatter market lines in Table AIII have a cost, the emergence of a residual size effect. Grouped on the basis of ME for individual stocks, the average residuals from the univariate regressions of returns on the βs of the 100 size-β portfolios are strongly positive for small stocks and negative for large stocks (0.60% per month for the smallest ME group, 1A, and -0.27% for the largest, 10B). Thus, when we allow for variation in β that is independent of size, the resulting βs leave a large size effect in average returns. This residual size effect is much like that observed by Banz (1981) with the βs of portfolios formed on size and β.

The correlation between size and β is -0.98 for portfolios formed on size alone. The independent variation in β obtained with the second-pass sort on β lowers the correlation to -0.50. The lower correlation means that bivariate regressions of returns on β and ln(ME) are more likely to distinguish true size effects from true β effects in average returns.

The bivariate regressions (Table AIII) that use the βs of the size-β portfolios are more bad news for β. The average slopes for ln(ME) are close to the values in the univariate size regressions, and almost 4 standard errors from 0, but the average slopes for β are negative and less than 1 standard error from 0. The message from the bivariate regressions is that there is a strong relation between size and average return. But like the regressions in Table AIII that explain average returns with β alone, the bivariate regressions say that there is no reliable relation between β and average returns when the tests use βs that are not close substitutes for size. These uncomfortable SLB results for NYSE stocks for 1941–1990 are much like those for NYSE, AMEX, and NASDAQ stocks for 1963–1990 in Table III.

C. Subperiod Diagnostics

Our results for 1941–1990 seem to contradict the evidence in Black, Jensen, and Scholes (BJS) (1972) and Fama and MacBeth (FM) (1973) that there is a reliable positive relation between average return and β. The βs in BJS and FM are from portfolios formed on β alone, and the market proxy is the NYSE equal-weighted portfolio. We use the βs of portfolios formed on size and β, and our market is the value-weighted NYSE portfolio. We can report, however, that our inference that there isn't much relation between β and average return is unchanged when (a) the market proxy is the NYSE EW portfolio, (b) portfolios are formed on just (pre-ranking) βs, or (c) the order of forming the size-β portfolios is changed from size then β to β then size.

A more important difference between our results and the earlier studies is the sample periods. The tests in BJS and FM end in the 1960s. Table AIV shows that when we split the 50-year 1941–1990 period in half, the univariate FM regressions of returns on β produce an average slope for 1941–1965 (0.50% per month, $t = 1.82$) more like that of the earlier studies. In contrast, the average slope on β for 1966–1990 is close to 0 (-0.02, $t = 0.06$).

But Table AIV also shows that drawing a distinction between the results for 1941–1965 and 1966–1990 is misleading. The stronger tradeoff of average return for β in the simple regressions for 1941–1965 is due to the first 10 years, 1941–1950. This is the only period in Table AIV that produces an average premium for β (1.26% per month) that is both positive and more than 2 standard errors from 0. Conversely, the weak relation between β and average return for 1966–1990 is largely due to 1981–1990. The strong negative average slope in the univariate regressions of returns on β for 1981–1990 (-1.01, $t = -2.10$) offsets a positive slope for 1971–1980 (0.82, $t = 1.27$).

The subperiod variation in the average slopes from the FM regressions of returns on β alone seems moot, however, given the evidence in Table AIV that adding size always kills any positive tradeoff of average return for β in the subperiods. Adding size to the regressions for 1941–1965 causes the average slope for β to drop from 0.50 ($t = 1.82$) to 0.07 ($t = 0.28$). In contrast, the average slope on size in the bivariate regressions (-0.16, $t = -2.97$) is close to its value (-0.17, $t = -2.88$) in the regressions of returns on ln(ME) alone. Similar comments hold for 1941–1950. In short, any evidence of a positive average premium for β in the subperiods seems to be a size effect in disguise.

D. Can the SLB Model Be Saved?

Before concluding that β has no explanatory power, it is appropriate to consider other explanations for our results. One possibility is that the variation in β produced by the β sorts of size deciles in just sampling error. If so, it is not surprising that the variation in β within a size decile is unrelated to average return, or that size dominates β in bivariate tests. The standard errors of the βs suggest, however, that this explanation cannot save the SLB

Table AIII

Average Slopes, Their Standard Errors (SE), and Average Residuals from Monthly FM Regressions for Individual NYSE Stocks and for Portfolios Formed on Size and Pre-Ranking β: 1941-1990

Stocks are assigned the post-ranking β of the size-β portfolio they are in at the end of year $t-1$ (Table AII). ln(ME) is the natural log of price times shares outstanding at the end of year $t-1$. In the individual-stock regressions, these values of the explanatory variables are matched with CRSP returns for each of the 12 months in year t. The portfolio regressions match the equal-weighted portfolio returns for the size-β portfolios (Table AII) with the equal-weighted averages of β and ln(ME) for the surviving stocks in each month of year t. Slope is the time-series average of the monthly regression slopes from 1941-1990 (600 months); SE is the time-series standard error of the average slope.

The residuals from the monthly regressions in year t are grouped into 12 portfolios on the basis of size or pre-ranking β (estimated with 24 to 60 months of returns, as available) as of the end of year $t-1$. The average residuals are the time-series averages of the monthly equal-weighted averages of the residuals in percent. The average residuals (not shown) from the FM regressions (1) to (3) that use the returns on the 100 size-β portfolios as the dependent variable are always within 0.01 of those from the regressions for individual stock returns. This is not surprising given that the correlation between the time-series of 1941-1990 monthly FM slopes on β or ln(ME) for the comparable portfolio and individual stock regressions is always greater than 0.99.

	Portfolio Regressions				Individual Stock Regressions			
	(1) β	(2) ln(ME)	(3) β and ln(ME)		(4) β	(5) ln(ME)	(6) β and ln(ME)	
Slope	0.22	−0.128	−0.13	−0.143	0.24	−0.133	−0.14	−0.147
SE	0.24	0.043	0.21	0.039	0.23	0.043	0.21	0.039

	Average Residuals for Stocks Grouped on Size											
	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Regression (4)	0.60	0.26	0.13	0.06	−0.01	−0.03	−0.03	−0.09	−0.10	−0.11	−0.25	−0.27
Standard error	0.21	0.10	0.06	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.06	0.08
Regression (5)	0.30	0.02	−0.05	−0.06	−0.08	−0.07	−0.03	−0.04	0.02	0.08	0.01	0.13
Standard error	0.14	0.07	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.07
Regression (6)	0.31	0.02	−0.05	−0.06	−0.09	−0.07	−0.03	−0.04	0.02	0.08	0.01	0.13
Standard error	0.14	0.07	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.07

Table AIII—*Continued*

	Portfolio Regressions			Individual Stock Regressions		
	(1) β	(2) ln(ME)	(3) β and ln(ME)	(4) β	(5) ln(ME)	(6) β and ln(ME)

Average Residuals for Stocks Grouped on Pre-Ranking β

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Regression (4)	−0.08	0.03	−0.01	0.08	0.04	0.08	0.04	0.02	−0.03	0.02	−0.11	−0.32
Standard error	0.07	0.05	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.06	0.07
Regression (5)	−0.10	0.00	0.02	0.09	0.05	0.07	0.05	0.00	−0.03	−0.01	−0.11	−0.33
Standard error	0.11	0.10	0.07	0.05	0.04	0.03	0.03	0.04	0.05	0.07	0.10	0.13
Regression (6)	−0.17	−0.07	−0.02	0.07	0.04	0.06	0.05	0.03	0.00	0.04	−0.04	−0.23
Standard error	0.05	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.06	0.07

Table AIV

Subperiod Average Returns on the NYSE Value-Weighted and Equal-Weighted Portfolios and Average Values of the Intercepts and Slopes for the FM Cross-Sectional Regressions of Individual Stock Returns on β and Size (ln(ME))

Mean is the average VW or EW return or an average slope from the monthly cross-sectional regressions of individual stock returns on β and/or ln(ME). Std is the standard deviation of the time-series of returns or slopes, and t(Mn) is Mean over its time-series standard error. The average slopes (not shown) from the FM regressions that use the returns on the 100 size-β portfolios of Table AII as the dependent variable are quite close to those for individual stock returns. (The correlation between the 1941–1990 month-by-month slopes on β or ln(ME) for the comparable portfolio and individual stock regressions is always greater than 0.99.)

Panel A

	1941–1990 (600 Mos.)			1941–1965 (300 Mos.)			1966–1990 (300 Mos.)		
Variable	Mean	Std	t(Mn)	Mean	Std	t(Mn)	Mean	Std	t(Mn)
NYSE Value-Weighted (VW) and Equal-Weighted (EW) Portfolio Returns									
VW	0.93	4.15	5.49	1.10	3.58	5.30	0.76	4.64	2.85
EW	1.12	5.10	5.37	1.33	4.42	5.18	0.91	5.70	2.77
$R_{it} = a + b_{1t}\beta_{it} + e_{it}$									
a	0.98	3.93	6.11	0.84	3.18	4.56	1.13	4.57	4.26
b_1	0.24	5.52	1.07	0.50	4.75	1.82	−0.02	6.19	−0.06
$R_{it} = a + b_{2t}\ln(ME_{it}) + e_{it}$									
a	1.70	8.24	5.04	1.88	6.43	5.06	1.51	9.72	2.69
b_2	−0.13	1.06	−3.07	−0.17	1.01	−2.88	−0.10	1.11	−1.54
$R_{it} = a + b_{1t}\beta_{it} + b_{2t}\ln(ME_{it}) + e_{it}$									
a	1.97	6.16	7.84	1.80	4.77	6.52	2.14	7.29	5.09
b_1	−0.14	5.05	−0.66	0.07	4.15	0.28	−0.34	5.80	−1.01
b_2	−0.15	0.96	−3.75	−0.16	0.94	−2.97	−0.13	0.99	−2.34

Table AIV—*Continued*

Panel B:

Return	1941-1950 Mean	1941-1950 t(Mn)	1951-1960 Mean	1951-1960 t(Mn)	1961-1970 Mean	1961-1970 t(Mn)	1971-1980 Mean	1971-1980 t(Mn)	1981-1990 Mean	1981-1990 t(Mn)
NYSE Value-Weighted (VW) and Equal-Weighted (EW) Portfolio Returns										
VW	1.05	2.88	1.18	3.95	0.66	1.84	0.72	1.67	1.04	2.40
EW	1.59	3.16	1.13	3.76	0.88	1.96	1.04	1.82	0.95	2.01
$R_{it} = a + b_{1t}\beta_{it} + e_{it}$										
a	0.24	0.66	1.41	6.36	0.64	1.94	0.27	0.62	2.35	5.99
b_1	1.26	2.20	−0.19	−0.63	0.32	0.72	0.82	1.27	−1.01	−2.10
$R_{it} = a + b_{2t}\ln(ME_{it}) + e_{it}$										
a	2.63	3.47	1.08	2.73	1.78	2.50	2.18	2.03	0.82	1.20
b_2	−0.37	−2.90	0.03	0.53	−0.17	−2.19	−0.20	−1.57	0.04	0.57
$R_{it} = a + b_{1t}\beta_{it} + b_{2t}\ln(ME_{it}) + e_{it}$										
a	2.14	3.93	1.38	4.03	2.01	4.16	1.50	2.12	2.84	4.25
b_1	0.34	0.75	−0.17	−0.53	−0.11	−0.27	0.41	0.75	−1.14	−2.16
b_2	−0.34	−2.92	0.01	0.20	−0.18	−2.89	−0.16	−1.50	−0.07	−0.84

model. The standard errors for portfolios formed on size and β are only slightly larger (0.02 to 0.11) than those for portfolios formed on size alone (0.01 to 0.10, Table AI). And the range of the post-ranking βs within a size decile is always large relative to the standard errors of the βs.

Another possibility is that the proportionality condition (1) for the variation through time in true βs, that justifies the use of full-period post-ranking βs in the FM tests, does not work well for portfolios formed on size and β. If this is a problem, post-ranking βs for the size-β portfolios should not be highly correlated across subperiods. The correlation between the half-period (1941–1965 and 1966–1990) βs of the size-β portfolios is 0.91, which we take to be good evidence that the full-period β estimates for these portfolios are informative about true βs. We can also report that using 5-year βs (pre- or post-ranking) in the FM regressions does not change our negative conclusions about the role of β in average returns, as long as portfolios are formed on β as well as size, or on β alone.

Any attempt to salvage the simple positive relation between β and average return predicted by the SLB model runs into three damaging facts, clear in Table AII. (a) Forming portfolios on size and pre-ranking βs produces a wide range of post-ranking βs in every size decile. (b) The post-ranking βs closely reproduce (in deciles 2 to 10 they exactly reproduce) the ordering of the pre-ranking βs used to form the β-sorted portfolios. It seems safe to conclude that the increasing pattern of the post-ranking βs in every size decile captures the ordering of the true βs. (c) Contrary to the SLB model, the β sorts do not produce a similar ordering of average returns. Within the rows (size deciles) of the average return matrix in Table AII, the high-β portfolios have average returns that are close to or less than the low-β portfolios.

But the most damaging evidence against the SLB model comes from the univariate regressions of returns on β in Table AIII. They say that when the tests allow for variation in β that is unrelated to size, the relation between β and average return for 1941–1990 is weak, perhaps nonexistent, even when β is the only explanatory variable. We are forced to conclude that the SLB model does not describe the last 50 years of average stock returns.

REFERENCES

Alford, Andrew, Jennifer J. Jones, and Mark E. Zmijewski, 1992, Extensions and violations of the statutory SEC Form 10-K filing date, Unpublished manuscript, University of Chicago, Chicago, IL.

Ball, Ray, 1978, Anomalies in relationships between securities' yields and yield-surrogates, *Journal of Financial Economics* 6, 103–126.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Basu, Sanjoy, 1983, The relationship between earnings yield, market value, and return for NYSE common stocks: Further evidence, *Journal of Financial Economics* 12, 129–156.

Bhandari, Laxmi Chand, 1988, Debt/Equity ratio and expected common stock returns: Empirical evidence, *Journal of Finance* 43, 507–528.

Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, *Journal of Business* 45, 444–455.

———, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: some empirical tests, in M. Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger).

Chan, Louis K., Yasushi Hamao, and Josef Lakonishok, 1991, Fundamentals and stock returns in Japan, *Journal of Finance* 46, 1739-1789.

Chan, K. C. and Nai-fu Chen, 1988, An unconditional asset-pricing test and the role of firm size as an instrumental variable for risk, *Journal of Finance* 43, 309-325.

———, and Nai-fu Chen, 1991, Structural and return characteristics of small and large firms, *Journal of Finance* 46, 1467-1484.

———, Nai-fu Chen, and David A. Hsieh, 1985, An exploratory investigation of the firm size effect, *Journal of Financial Economics* 14, 451-471.

Chen, Nai-fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 56, 383-403.

DeBondt, Werner F. M., and Richard H. Thaler, 1985, Does the stock market overreact, *Journal of Finance* 40, 557-581.

Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 197-226.

Fama, Eugene F., 1976, *Foundations of Finance* (Basic Books, New York).

———, and James MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.

Fowler, David J. and C. Harvey Rorke, 1983, Risk measurement when shares are subject to infrequent trading: Comment, *Journal of Financial Economics* 12, 279-283.

Jaffe, Jeffrey, Donald B. Keim, and Randolph Westerfield, 1989, Earnings yields, market values, and stock returns, *Journal of Finance* 44, 135-148.

Keim, Donald B., 1983, Size-related anomalies and stock return seasonality, *Journal of Financial Economics* 12, 13-32.

———, 1988, Stock market regularities: A synthesis of the evidence and explanations, in Elroy Dimson, ed.: *Stock Market Anomalies* (Cambridge University Press, Cambridge).

Lakonishok, Josef, and Alan C. Shapiro, 1986, Systematic risk, total risk and size as determinants of stock market returns, *Journal of Banking and Finance* 10, 115-132.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13-37.

Markowitz, Harry, 1959, *Portfolio Selection: Efficient Diversification of Investments* (Wiley, New York).

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867-887.

Reinganum, Marc R., 1981, A new empirical perspective on the CAPM, *Journal of Financial and Quantitative Analysis* 16, 439-462.

Roll, Richard, 1983, Vas ist Das? The turn-of-the-year effect and the return premia of small firms, *Journal of Portfolio Management* 9, 18-28.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9-17.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341-360.

Sharpe, William F., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425-442.

Stambaugh, Robert F., 1982, On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis, *Journal of Financial Economics* 10, 237-268.

Stattman, Dennis, 1980, Book values and stock returns, *The Chicago MBA: A Journal of Selected Papers* 4, 25-45.

Journal of Financial Economics 33 (1993) 3–56. North-Holland

Common risk factors in the returns on stocks and bonds*

Eugene F. Fama and Kenneth R. French
University of Chicago, Chicago, IL 60637, USA

Received July 1992, final version received September 1992

This paper identifies five common risk factors in the returns on stocks and bonds. There are three stock-market factors: an overall market factor and factors related to firm size and book-to-market equity. There are two bond-market factors, related to maturity and default risks. Stock returns have shared variation due to the stock-market factors, and they are linked to bond returns through shared variation in the bond-market factors. Except for low-grade corporates, the bond-market factors capture the common variation in bond returns. Most important, the five factors seem to explain average returns on stocks and bonds.

1. Introduction

The cross-section of average returns on U.S. common stocks shows little relation to either the market βs of the Sharpe (1964)–Lintner (1965) asset-pricing model or the consumption βs of the intertemporal asset-pricing model of Breeden (1979) and others. [See, for example, Reinganum (1981) and Breeden, Gibbons, and Litzenberger (1989).] On the other hand, variables that have no special standing in asset-pricing theory show reliable power to explain the cross-section of average returns. The list of empirically determined average-return variables includes size (ME, stock price times number of shares), leverage, earnings/price (E/P), and book-to-market equity (the ratio of the book value of a firm's common stock, BE, to its market value, ME). [See Banz (1981), Bhandari (1988), Basu (1983), and Rosenberg, Reid, and Lanstein (1985).]

Correspondence to: Eugene F. Fama, Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637, USA.

*The comments of David Booth, John Cochrane, Nai-fu Chen, Wayne Ferson, Josef Lakonishok, Mark Mitchell, G. William Schwert, Jay Shanken, and Rex Sinquefield are gratefully acknowledged. This research is supported by the National Science Foundation (Fama) and the Center for Research in Securities Prices (French).

0304-405X/93/$05.00

Fama and French (1992a) study the joint roles of market β, size, E/P, leverage, and book-to-market equity in the cross-section of average stock returns. They find that used alone or in combination with other variables, β (the slope in the regression of a stock's return on a market return) has little information about average returns. Used alone, size, E/P, leverage, and book-to-market equity have explanatory power. In combinations, size (ME) and book-to-market equity (BE/ME) seem to absorb the apparent roles of leverage and E/P in average returns. The bottom-line result is that two empirically determined variables, size and book-to-market equity, do a good job explaining the cross-section of average returns on NYSE, Amex, and NASDAQ stocks for the 1963–1990 period.

This paper extends the asset-pricing tests in Fama and French (1992a) in three ways.

(a) We expand the set of asset returns to be explained. The only assets considered in Fama and French (1992a) are common stocks. If markets are integrated, a single model should also explain bond returns. The tests here include U.S. government and corporate bonds as well as stocks.

(b) We also expand the set of variables used to explain returns. The size and book-to-market variables in Fama and French (1992a) are directed at stocks. We extend the list to term-structure variables that are likely to play a role in bond returns. The goal is to examine whether variables that are important in bond returns help to explain stock returns, and vice versa. The notion is that if markets are integrated, there is probably some overlap between the return processes for bonds and stocks.

(c) Perhaps most important, the approach to testing asset-pricing models is different. Fama and French (1992a) use the cross-section regressions of Fama and MacBeth (1973): the cross-section of stock returns is regressed on variables hypothesized to explain average returns. It would be difficult to add bonds to the cross-section regressions since explanatory variables like size and book-to-market equity have no obvious meaning for government and corporate bonds.

This paper uses the time-series regression approach of Black, Jensen, and Scholes (1972). Monthly returns on stocks and bonds are regressed on the returns to a market portfolio of stocks and mimicking portfolios for size, book-to-market equity (BE/ME), and term-structure risk factors in returns. The time-series regression slopes are factor loadings that, unlike size or BE/ME, have a clear interpretation as risk-factor sensitivities for bonds as well as for stocks.

The time-series regressions are also convenient for studying two important asset-pricing issues.

(a) One of our central themes is that if assets are priced rationally, variables that are related to average returns, such as size and book-to-market equity, must proxy for sensitivity to common (shared and thus undiversifiable) risk factors in returns. The time-series regressions give direct evidence on this issue. In particular, the slopes and R^2 values show whether mimicking portfolios for risk factors related to size and BE/ME capture shared variation in stock and bond returns not explained by other factors.

(b) The time-series regressions use excess returns (monthly stock or bond returns minus the one-month Treasury bill rate) as dependent variables and either excess returns or returns on zero-investment portfolios as explanatory variables. In such regressions, a well-specified asset-pricing model produces intercepts that are indistinguishable from 0 [Merton (1973)]. The estimated intercepts provide a simple return metric and a formal test of how well different combinations of the common factors capture the cross-section of average returns. Moreover, judging asset-pricing models on the basis of the intercepts in excess-return regressions imposes a stringent standard. Competing models are asked to explain the one-month bill rate as well as the returns on longer-term bonds and stocks.

Our main results are easy to summarize. For stocks, portfolios constructed to mimic risk factors related to size and BE/ME capture strong common variation in returns, no matter what else is in the time-series regressions. This is evidence that size and book-to-market equity indeed proxy for sensitivity to common risk factors in stock returns. Moreover, for the stock portfolios we examine, the intercepts from three-factor regressions that include the excess market return and the mimicking returns for size and BE/ME factors are close to 0. Thus a market factor and our proxies for the risk factors related to size and book-to-market equity seem to do a good job explaining the cross-section of average stock returns.

The interpretation of the time-series regressions for stocks is interesting. Like the cross-section regressions of Fama and French (1992a), the time-series regressions say that the size and book-to-market factors can explain the differences in average returns across stocks. But these factors alone cannot explain the large difference between the average returns on stocks and one-month bills. This job is left to the market factor. In regressions that also include the size and book-to-market factors, all our stock portfolios produce slopes on the market factor that are close to 1. The risk premium for the market factor then links the average returns on stocks and bills.

For bonds, the mimicking portfolios for the two term-structure factors (a term premium and a default premium) capture most of the variation in the returns on our government and corporate bond portfolios. The term-structure factors also 'explain' the average returns on bonds, but the average premiums for the term-structure factors, like the average excess bond returns, are close to 0. Thus, the hypothesis that all the corporate and government bond portfolios have the same long-term expected returns also cannot be rejected.

The common variation in stock returns is largely captured by three stock-portfolio returns, and the common variation in bond returns is largely explained by two bond-portfolio returns. The stock and bond markets, however, are far from stochastically segmented. Used alone in the time-series regressions, the term-structure factors capture strong variation in stock returns; indeed, the slopes on the term-structure factors in the regressions for stocks are much like those for bonds. But interestingly, when stock-market factors are also included in the regressions, all of our stock portfolios load in about the same way on the two term-structure factors and on the market factor in returns. As a result, a market portfolio of stocks captures the common variation in stock returns associated with the market factor and the two term-structure factors.

The stochastic links between the bond and stock markets do, however, seem to come largely from the term-structure factors. Used alone, the excess market return and the mimicking returns for the size and book-to-market equity factors seem to capture common variation in bond returns. But when the two term-structure factors are included in the bond regressions, the explanatory power of the stock-market factors disappears for all but the low-grade corporate bonds.

In a nutshell, our results suggest that there are at least three stock-market factors and two term-structure factors in returns. Stock returns have shared variation due to the three stock-market factors, and they are linked to bond returns through shared variation in the two term-structure factors. Except for low-grade corporate bonds, only the two term-structure factors seem to produce common variation in the returns on government and corporate bonds.

The story proceeds as follows. We first introduce the inputs to the time-series regressions: the explanatory variables and the returns to be explained (sections 2 and 3). We then use the regressions to attack our two central asset-pricing issues: how do different combinations of variables capture (a) the common variation through time in the returns on bonds and stocks (section 4) and (b) the cross-section of average returns (section 5).

2. The inputs to the time-series regressions

The explanatory variables in the time-series regressions include the returns on a market portfolio of stocks and mimicking portfolios for the size, book-to-market, and term-structure factors in returns. The returns to be explained are for government bond portfolios in two maturity ranges, corporate bond portfolios in five rating groups, and 25 stock portfolios formed on the basis of size and book-to-market equity.

2.1. *The explanatory returns*

The explanatory variables fall into two sets, those likely to be important for capturing variation in bond returns and those likely to be important for stocks. Segmenting the explanatory variables in this way sets up interesting tests of whether factors important in stock returns help to explain bond returns and vice versa.

2.1.1. *Bond-market factors*

One common risk in bond returns arises from unexpected changes in interest rates. Our proxy for this factor, *TERM*, is the difference between the monthly long-term government bond return (from Ibbotson Associates) and the one-month Treasury bill rate measured at the end of the previous month (from the Center for Research in Security Prices, CRSP). The bill rate is meant to proxy for the general level of expected returns on bonds, so that *TERM* proxies for the deviation of long-term bond returns from expected returns due to shifts in interest rates.

For corporate bonds, shifts in economic conditions that change the likelihood of default give rise to another common factor in returns. Our proxy for this default factor, *DEF*, is the difference between the return on a market portfolio of long-term corporate bonds (the Composite portfolio on the corporate bond module of Ibbotson Associates) and the long-term government bond return.

Chen, Roll, and Ross (1986) use *TERM* and a variable like *DEF* to help explain the cross-section of average returns on NYSE stocks. They use the Fama and MacBeth (1973) cross-section regression approach; the cross-section of average stock returns is explained with the cross-section of slopes from time-series regressions of returns on *TERM*, a default factor, and other factors. In their tests, the default factor is the most powerful factor in average stock returns, and *TERM* sometimes has power. We confirm that the tracks of *TERM* and *DEF* show up clearly in the time-series variation of stock returns. We also find that the two variables dominate the common variation in government and corporate bond returns. In contrast to the cross-section regressions of Chen, Roll, and Ross, however, our time-series regressions say that the average premiums for *DEF* and *TERM* risks are too small to explain much variation in the cross-section of average stock returns. [Shanken and Weinstein (1990) make a similar point.]

2.1.2. *Stock-market factors*

Motivation – Although size and book-to-market equity seem like ad hoc variables for explaining average stock returns, we have reason to expect that they proxy for common risk factors in returns. In Fama and French (1992b) we document that size and book-to-market equity are related to economic fundamentals. Not surprisingly, firms that have high *BE/ME* (a low stock price relative to book value) tend to have low earnings on assets, and the low earnings persist for at least five years before and five years after book-to-market equity is measured. Conversely, low *BE/ME* (a high stock price relative to book value) is associated with persistently high earnings.

Size is also related to profitability. Controlling for book-to-market equity, small firms tend to have lower earnings on assets than big firms. The size effect in earnings, however, is largely due to the 1980s. Until 1981, controlling for *BE/ME*, small firms are only slightly less profitable than big firms. But for small firms, the 1980–1982 recession turns into a prolonged earnings depression. For some reason, small firms do not participate in the economic boom of the middle and late 1980s.

The fact that small firms can suffer a long earnings depression that bypasses big firms suggests that size is associated with a common risk factor that might explain the negative relation between size and average return. Similarly, the relation between book-to-market equity and earnings suggests that relative profitability is the source of a common risk factor in returns that might explain the positive relation between *BE/ME* and average return. Measuring the common variation in returns associated with size and *BE/ME* is a major task of this paper.

The Building Blocks – To study economic fundamentals, Fama and French (1992b) use six portfolios formed from sorts of stocks on *ME* and *BE/ME*. We use the same six portfolios here to form portfolios meant to mimic the underlying risk factors in returns related to size and book-to-market equity. This ensures a correspondence between the study of common risk factors in returns carried out here and our complementary study of economic fundamentals.

In June of each year t from 1963 to 1991, all NYSE stocks on CRSP are ranked on size (price times shares). The median NYSE size is then used to split NYSE, Amex, and (after 1972) NASDAQ stocks into two groups, small and big (*S* and *B*). Most Amex and NASDAQ stocks are smaller than the NYSE median, so the small group contains a disproportionate number of stocks (3,616 out of 4,797 in 1991). Despite its large number of stocks, the small group contains far less than half (about 8% in 1991) of the combined value of the two size groups.

We also break NYSE, Amex, and NASDAQ stocks into three book-to-market equity groups based on the breakpoints for the bottom 30% (*Low*), middle 40% (*Medium*), and top 30% (*High*) of the ranked values of *BE/ME* for NYSE stocks. We define book common equity, *BE*, as the COMPUSTAT book value of stockholders' equity, plus balance-sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock. Depending on availability, we use the redemption, liquidation, or par value (in that order) to estimate the value of preferred stock. Book-to-market equity, *BE/ME*, is then book common equity for the fiscal year ending in calendar year $t-1$, divided by market equity at the end of December of $t-1$. We do not use negative-*BE* firms, which are rare before 1980, when calculating the breakpoints for *BE/ME* or when forming the size–*BE/ME* portfolios. Also, only firms with ordinary common equity (as classified by CRSP) are included in the tests. This means that ADRs, REITs, and units of beneficial interest are excluded.

Our decision to sort firms into three groups on BE/ME and only two on ME follows the evidence in Fama and French (1992a) that book-to-market equity has a stronger role in average stock returns than size. The splits are arbitrary, however, and we have not searched over alternatives. The hope is that the tests here and in Fama and French (1992b) are not sensitive to these choices. We see no reason to argue that they are.

We construct six portfolios (S/L, S/M, S/H, B/L, B/M, B/H) from the intersections of the two ME and the three BE/ME groups. For example, the S/L portfolio contains the stocks in the small-ME group that are also in the low-BE/ME group, and the B/H portfolio contains the big-ME stocks that also have high BE/MEs. Monthly value-weighted returns on the six portfolios are calculated from July of year t to June of $t+1$, and the portfolios are reformed in June of $t+1$. We calculate returns beginning in July of year t to be sure that book equity for year $t-1$ is known.

To be included in the tests, a firm must have CRSP stock prices for December of year $t-1$ and June of t and COMPUSTAT book common equity for year $t-1$. Moreover, to avoid the survival bias inherent in the way COMPUSTAT adds firms to its tapes [Banz and Breen (1986)], we do not include firms until they have appeared on COMPUSTAT for two years. (COMPUSTAT says it rarely includes more than two years of historical data when it adds firms).

Size – Our portfolio SMB (small minus big), meant to mimic the risk factor in returns related to size, is the difference, each month, between the simple average of the returns on the three small-stock portfolios (S/L, S/M, and S/H) and the simple average of the returns on the three big-stock portfolios (B/L, B/M, and B/H). Thus, SMB is the difference between the returns on small- and big-stock portfolios with about the same weighted-average book-to-market equity. This difference should be largely free of the influence of BE/ME, focusing instead on the different return behaviors of small and big stocks.

BE/ME – The portfolio HML (high minus low), meant to mimic the risk factor in returns related to book-to-market equity, is defined similarly. HML is the difference, each month, between the simple average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low- BE/ME portfolios (S/L and B/L). The two components of HML are returns on high- and low-BE/ME portfolios with about the same weighted-average size. Thus the difference between the two returns should be largely free of the size factor in returns, focusing instead on the different return behaviors of high- and low-BE/ME firms. As testimony to the success of this simple procedure, the correlation between the 1963–1991 monthly mimicking returns for the size and book-to-market factors is only -0.08.

True mimicking portfolios for the common risk factors in returns minimize the variance of firm-specific factors. The six size–BE/ME portfolios in SMB and

HML are value-weighted. Using value-weighted components is in the spirit of minimizing variance, since return variances are negatively related to size (table 2, below). More important, using value-weighted components results in mimicking portfolios that capture the different return behaviors of small and big stocks, or high- and low-BE/ME stocks, in a way that corresponds to realistic investment opportunities.

Market – Finally, our proxy for the market factor in stock returns is the excess market return, $RM-RF$. RM is the return on the value-weighted portfolio of the stocks in the six size-BE/ME portfolios, plus the negative-BE stocks excluded from the portfolios. RF is the one-month bill rate.

2.2. *The returns to be explained*

Bonds – The set of dependent variables used in the time-series regressions includes the excess returns on two government and five corporate bond portfolios. The government bond portfolios (from CRSP) cover maturities from 1 to 5 years and 6 to 10 years. The five corporate bond portfolios, for Moody's rating groups Aaa, Aa, A, Baa, and LG (low-grade, that is, below Baa) are from the corporate bond module of Ibbotson Associates (provided to us by Dimensional Fund Advisors).

Stocks – For stocks, we use excess returns on 25 portfolios, formed on size and book-to-market equity, as dependent variables in the time-series regressions. We use portfolios formed on size and BE/ME because we seek to determine whether the mimicking portfolios SMB and HML capture common factors in stock returns related to size and book-to-market equity. Portfolios formed on size and BE/ME will also produce a wide range of average returns to be explained by competing asset-pricing equations [Fama and French (1992a)]. Later, however, we use portfolios formed on E/P (earnings/price) and D/P (dividend/price), variables that are also informative about average returns [e.g., Keim (1988)], to check the robustness of our results on the ability of our explanatory factors to capture the cross-section of average returns.

The 25 size-BE/ME portfolios are formed much like the six size-BE/ME portfolios discussed earlier. In June of each year t we sort NYSE stocks by size and (independently) by book-to-market equity. For the size sort, ME is measured at the end of June. For the book-to-market sort, ME is market equity at the end of December of $t-1$, and BE is book common equity for the fiscal year ending in calendar year $t-1$. We use NYSE breakpoints for ME and BE/ME to allocate NYSE, Amex, and (after 1972) NASDAQ stocks to five size quintiles and five book-to-market quintiles. We construct 25 portfolios from the intersections of the size and BE/ME quintiles and calculate value-weighted monthly returns on the portfolios from July of t to June of $t+1$. The excess returns on these 25 portfolios for July 1963 to December 1991 are the dependent variables for stocks in the time-series regressions.

Table 1

Descriptive statistics for 25 stock portfolios formed on size and book-to-market equity: 1963–1991, 29 years.[a]

Size quintile	Low	2	3	4	High	Low	2	3	4	High
	Book-to-market equity (*BE/ME*) quintiles									
	Average of annual averages of firm size					Average of annual *B/E* ratios for portfolio				
Small	20.6	20.8	20.2	19.4	15.1	0.30	0.62	0.84	1.09	1.80
2	89.7	89.3	89.3	89.9	88.5	0.31	0.60	0.83	1.09	1.71
3	209.3	211.9	210.8	214.8	210.7	0.31	0.60	0.84	1.08	1.66
4	535.1	537.4	545.4	551.6	538.7	0.31	0.61	0.84	1.09	1.67
Big	3583.7	2885.8	2819.5	2700.5	2337.9	0.29	0.59	0.83	1.08	1.56
	Average of annual percent of market value in portfolio					Average of annual number of firms in portfolio				
Small	0.69	0.49	0.46	0.48	0.64	428.0	276.6	263.8	291.5	512.7
2	0.92	0.71	0.65	0.61	0.55	121.6	94.0	86.7	79.8	71.3
3	1.78	1.36	1.26	1.14	0.82	102.7	78.3	73.0	64.5	45.9
4	3.95	3.01	2.71	2.41	1.50	90.1	68.9	60.7	53.1	33.4
Big	30.13	15.87	12.85	10.44	4.61	93.6	63.7	52.7	44.0	23.6
	Average of annual *E/P* ratios (in percent) for portfolio					Average of annual *D/P* ratios (in percent) for portfolio				
Small	2.42	7.24	8.26	9.06	2.66	1.00	1.94	2.60	3.13	2.82
2	5.20	8.61	10.16	10.95	9.28	1.59	2.45	3.45	4.25	4.53
3	5.91	8.72	10.43	11.62	10.78	1.56	3.03	4.04	4.68	4.64
4	5.85	8.94	10.45	11.64	11.39	1.80	3.09	4.22	5.01	4.94
Big	6.00	9.07	10.90	12.45	13.92	2.34	3.69	4.68	5.49	5.90

[a]The 25 size-*BE/ME* stock portfolios are formed as follows. Each year t from 1963 to 1991 NYSE quintile breakpoints for size (*ME*, stock price times shares outstanding), measured at the end of June, are used to allocate NYSE, Amex, and NASDAQ stocks to five size quintiles. Similarly, NYSE quintile breakpoints for *BE/ME* are used to allocate NYSE, Amex, and NASDAQ stocks to five book-to-market equity quintiles. The 25 size-*BE/ME* portfolios are formed as the intersections of the five size and the five *BE/ME* groups. Book equity, *BE*, is the COMPUSTAT book value of stockholders' equity, plus balance sheet deferred taxes and investment tax credits (if available), minus the book value of preferred stock. Depending on availability, we use the redemption, liquidation, or par value (in that order) to estimate the book value of preferred stock. Book-to-market equity, *BE/ME*, for a stock is *BE* for the fiscal year ending in calendar year $t-1$, divided by *ME* at the end of December of $t-1$.

A portfolio's book-to-market equity, *BE/ME*, for the portfolio formation year t is the sum of book equity, *BE*, for the firms in the portfolio for the fiscal year ending in calendar year $t-1$, divided by the sum of their market equity, *ME*, in December of $t-1$. A portfolio's earnings/price ratio (*E/P*) for year t is the sum of equity income for the firms in the portfolio for the fiscal year ending in calendar year $t-1$, divided by the sum of their market equity in December of $t-1$. Equity income is income before extraordinary items, plus income-statement deferred taxes, minus preferred dividends. A portfolio's dividend yield (*D/P*) for year t is the sum (across firms in the portfolio) of the dividends paid from July of $t-1$ to June of t, divided by the sum of market equity in June of $t-1$. We use the procedure described in Fama and French (1988) to estimate dividends.

The descriptive statistics are computed when the portfolio is formed in June of each year, 1963–1991, and are then averaged across the 29 years.

Table 1 shows that, because we use NYSE breakpoints to form the 25 size–BE/ME portfolios, the portfolios in the smallest size quintile have the most stocks (mostly small Amex and NASDAQ stocks). Although they contain many stocks, each of the five portfolios in the smallest size quintile is on average less than 0.70% of the combined value of stocks in the 25 portfolios. In contrast, the portfolios in the largest size quintile have the fewest stocks but the largest fractions of value. Together, the five portfolios in the largest ME quintile average about 74% of total value. The portfolio of stocks in both the largest size and lowest BE/ME quintiles (big successful firms) alone accounts for more than 30% of the combined value of the 25 portfolios. And note that using all stocks, rather than just NYSE stocks, to define the size quintiles would result in an even more skewed distribution of value toward the biggest size quintile.

Table 1 also shows that in every size quintile but the smallest, both the number of stocks and the proportion of total value accounted for by a portfolio decrease from lower- to higher-BE/ME portfolios. This pattern has two causes. First, using independent size and book-to-market sorts of NYSE stocks to form portfolios means that the highest-BE/ME quintile is tilted toward the smallest stocks. Second, Amex and NASDAQ stocks, mostly small, tend to have lower book-to-market equity ratios than NYSE stocks of similar size. In other words, NYSE stocks that are small in terms of ME are more likely to be fallen angels (big firms with low stock prices) than small Amex and NASDAQ stocks.

3. The playing field

Table 2 summarizes the dependent and explanatory returns in the time-series regressions. The average excess returns on the portfolios that serve as dependent variables give perspective on the range of average returns that competing sets of risk factors must explain. The average returns on the explanatory portfolios are the average premiums per unit of risk (regression slope) for the candidate common risk factors in returns.

3.1. The dependent returns

Stocks – The 25 stock portfolios formed on size and book-to-market equity produce a wide range of average excess returns, from 0.32% to 1.05% per month. The portfolios also confirm the Fama–French (1992a) evidence that there is a negative relation between size and average return, and there is a stronger positive relation between average return and book-to-market equity. In all but the lowest-BE/ME quintile, average returns tend to decrease from the small- to the big-size portfolios. The relation between average return and book-to-market equity is more consistent. In every size quintile, average returns tend to increase with BE/ME, and the differences between the average returns for the highest- and lowest-BE/ME portfolios range from 0.19% to 0.62% per month.

Our time-series regressions attempt to explain the cross-section of average returns with the premiums for the common risk factors in returns. The wide range of average returns on the 25 stock portfolios, and the size and book-to-market effects in average returns, present interesting challenges for competing sets of risk factors.

Most of the ten portfolios in the bottom two BE/ME quintiles produce average excess returns that are less than two standard errors from 0. This is an example of a well-known problem [Merton (1980)] : because stock returns have high standard deviations (around 6% per month for the size-BE/ME portfolios), large average returns often are not reliably different from 0. The high volatility of stock returns does not mean, however, that our asset-pricing tests will lack power. The common factors in returns will absorb most of the variation in stock returns, making the asset-pricing tests on the intercepts in the time-series regressions quite precise.

Bonds – In contrast to the stock portfolios, the average excess returns on the government and corporate bond portfolios in table 2 are puny. All the average excess bond returns are less than 0.15% per month, and only one of seven is more than 1.5 standard errors from 0. There is little evidence in table 2 that (a) average returns on government bonds increase with maturity, (b) long-term corporate bonds have higher average returns than government bonds, or (c) average returns on corporate bonds are higher for lower-rating groups.

The flat cross-section of average bond returns does not mean that bonds are uninteresting dependent variables in the asset-pricing tests. On the contrary, bonds are good candidates for rejecting asset-pricing equations that predict patterns in the cross-section of average returns based on different slopes on the common risk factors in returns.

3.2. The explanatory returns

In the time-series regression approach to asset-pricing tests, the average risk premiums for the common factors in returns are just the average values of the explanatory variables. The average value of $RM-RF$ (the average premium per unit of market β) is 0.43% per month. This is large from an investment perspective (about 5% per year), but it is a marginal 1.76 standard errors from 0. The average SMB return (the average premium for the size-related factor in returns) is only 0.27% per month ($t = 1.73$). We shall find, however, that the slopes on SMB for the 25 stock portfolios cover a range in excess of 1.7, so the estimated spread in expected returns due to the size factor is large, about 0.46% per month. The book-to-market factor HML produces an average premium of 0.40% per month ($t = 2.91$), that is large in both practical and statistical terms.

Table 2

Summary statistics for the monthly dependent and explanatory returns (in percent) in the regressions of tables 3 to 8: July 1963 to December 1991, 342 observations.[a]

Name	Mean	Std.	$t(mn)$	Autocorr. for lag 1	2	12	Correlations *RM-RF*	*RMO*	*SMB*	*HML*	*TERM*
				Explanatory returns							
RM	0.97	4.52	3.97	0.05	−0.05	0.03					
TB	0.54	0.22	45.97	0.94	0.90	0.65					
LTG	0.60	3.03	3.66	0.05	−0.00	0.00					
CB	0.62	2.24	5.10	0.20	−0.04	0.04					
RM-RF	0.43	4.54	1.76	0.05	−0.04	0.03					
RMO	0.50	3.55	2.61	−0.10	−0.05	0.02	0.78	1.00			
SMB	0.27	2.89	1.73	0.19	0.07	0.23	0.32	−0.00	1.00		
HML	0.40	2.54	2.91	0.18	0.06	0.07	−0.38	−0.00	−0.08	1.00	
TERM	0.06	3.02	0.38	0.05	−0.00	−0.00	0.34	0.00	−0.07	−0.05	1.00
DEF	0.02	1.60	0.21	−0.20	−0.04	−0.00	−0.07	−0.00	0.17	0.08	−0.69
				Dependent variables: Excess returns on government and corporate bonds							
1-5G	0.12	1.25	1.71	0.15	−0.08	0.01					
6-10G	0.14	2.03	1.24	0.12	−0.05	0.02					
AAA	0.06	2.34	0.44	0.16	−0.04	0.02					
AA	0.07	2.23	0.58	0.19	−0.04	0.03					
A	0.08	2.25	0.63	0.21	−0.03	0.04					
BAA	0.14	2.35	1.09	0.21	0.00	0.03					
LG	0.13	2.52	0.98	0.23	0.05	0.08					

Dependent variables: Excess returns on 25 stock portfolios formed on ME and BE/ME

Book-to-market equity (BE/ME) quintiles

Size quintile	Low	2	3	4	High	Low	2	3	4	High
			Means					Standard deviations		
Small	0.39	0.70	0.79	0.88	1.01	7.76	6.84	6.29	5.99	6.27
2	0.44	0.71	0.85	0.84	1.02	7.28	6.42	5.85	5.33	6.06
3	0.43	0.66	0.68	0.81	0.97	6.71	5.71	5.27	4.92	5.69
4	0.48	0.35	0.57	0.77	1.05	5.97	5.44	5.03	4.95	5.75
Big	0.40	0.36	0.32	0.56	0.59	4.95	4.70	4.38	4.27	4.85
			t-statistics for means							
Small	0.93	1.88	2.33	2.73	2.97					
2	1.11	2.05	2.69	2.91	3.11					
3	1.18	2.12	2.39	3.04	3.15					
4	1.49	1.19	2.08	2.88	3.36					
Big	1.50	1.42	1.34	2.43	2.26					

[a]RM is the value-weighted monthly percent return on the stocks in the 25 size-BE/ME portfolios, plus the negative-BE stocks excluded from the portfolios. RF is the one-month Treasury bill rate, observed at the beginning of the month. LTG is the long-term government bond return. CB is the return on a proxy for the market portfolio of long-term corporate bonds. $TERM$ is $LTG-RF$. DEF is $CB-LTG$. SMB (small minus big) is the difference between the returns on small-stock and big-stock portfolios with about the same weighted average book-to-market equity. HML (high minus low) is the difference between the returns on high and low book-to-market equity portfolios with about the same weighted average size. RMO is the sum of the intercept and residuals from the regression (1) of $RM-RF$ on $TERM$, DEF, SMB, and HML.

The seven bond portfolios used as dependent variables in the excess-return regressions are 1- to 5-year and 6- to 10-year governments (1–5G and 6–10G) and corporate bonds rated Aaa, Aa, A, Baa, and below Baa (LG) by Moody's. The 25 size-BE/ME stock portfolios are formed as follows. Each year t from 1963 to 1991 NYSE quintile breakpoints for size (ME, stock price times shares outstanding), measured at the end of June, are used to allocate NYSE, Amex, and NASDAQ stocks to five size quintiles. Similarly, NYSE quintile breakpoints for BE/ME are used to allocate NYSE, Amex, and NASDAQ stocks to five book-to-market equity quintiles. In BE/ME, BE is book common equity for the fiscal year ending in calendar year $t-1$, and ME is for the end of December of $t-1$. The 25 size-BE/ME portfolios are formed as the intersections of the five size and the five BE/ME groups. Value-weighted monthly percent returns on the portfolios are calculated from July of year t to June of $t+1$.

The average risk premiums for the term-structure factors are trivial relative to those of the stock-market factors. *TERM* (the term premium) and *DEF* (the default premium) are on average 0.06% and 0.02% per month; both are within 0.4 standard errors of 0. Note, though, that *TERM* and *DEF* are about as volatile as the stock-market returns *SMB* and *HML*. Low average premiums will prevent *TERM* and *DEF* from explaining much cross-sectional variation in average returns, but high volatility implies that the two factors can capture substantial common variation in returns. In fact, the low means and high volatilities of *TERM* and *DEF* will be advantageous for explaining bond returns. But the task of explaining the strong cross-sectional variation in average stock returns falls on the stock-market factors, *RM–RF*, *SMB*, and *HML*, which produce higher average premiums.

We turn now to the asset-pricing tests. In the time-series regression approach, the tests have two parts. In section 4 we establish that the two bond-market returns, *TERM* and *DEF*, and the three stock-market returns, *RM–RF*, *SMB*, and *HML*, are risk factors in the sense that they capture common (shared and thus undiversifiable) variation in stock and bond returns. In section 5 we use the intercepts from the time-series regressions to test whether the average premiums for the common risk factors in returns explain the cross-section of average returns on bonds and stocks.

4. Common variation in returns

In the time-series regressions, the slopes and R^2 values are direct evidence on whether different risk factors capture common variation in bond and stock returns. We first examine separately the explanatory power of bond-market and stock-market factors. The purpose is to test for overlap between the stochastic processes for stock and bond returns. Do bond-market factors that are important in bond returns capture common variation in stock returns and vice versa? We then examine the joint explanatory power of the bond- and stock-market factors, to develop an overall story for the common variation in returns.

4.1. Bond-market factors

Table 3 shows that, used alone as the explanatory variables in the time-series regressions, *TERM* and *DEF* capture common variation in stock and bond returns. The 25 stock portfolios produce slopes on *TERM* that are all more than five standard errors above 0; the smallest *TERM* slope for the seven bond portfolios is 18 standard errors from 0. The slopes on *DEF* are all more than 7.8 standard errors from 0 for bonds, and more than 3.5 standard errors from 0 for stocks.

Table 3

Regressions of excess stock and bond returns (in percent) on the bond-market returns, $TERM$ and DEF: July 1963 to December 1991, 342 months.[a]

$$R(t) - RF(t) = a + mTERM(t) + dDEF(t) + e(t)$$

Dependent variable: Excess returns on 25 stock portfolios formed on size and book-to-market equity

Size quintile	Book-to-market equity (BE/ME) quintiles: Low	2	3	4	High	Low	2	3	4	High
			m					$t(m)$		
Small	0.93	0.90	0.89	0.86	0.89	5.02	5.50	5.95	6.08	6.01
2	0.99	0.96	0.99	1.01	0.98	5.71	6.32	7.29	8.34	6.92
3	0.99	0.94	0.94	0.95	0.99	6.25	7.10	7.80	8.50	7.60
4	0.92	0.95	0.97	1.05	1.03	6.58	7.57	8.53	9.64	7.83
Big	0.82	0.82	0.80	0.80	0.77	7.14	7.60	8.09	8.26	6.84
			d					$t(d)$		
Small	1.39	1.31	1.33	1.45	1.52	3.96	4.27	4.73	5.45	5.45
2	1.26	1.28	1.35	1.38	1.41	3.84	4.47	5.28	6.05	5.29
3	1.21	1.19	1.25	1.24	1.21	4.05	4.74	5.49	5.89	4.88
4	0.96	1.01	1.13	1.21	1.22	3.65	4.28	5.25	5.89	4.92
Big	0.78	0.73	0.78	0.83	0.89	3.59	3.60	4.18	4.56	4.15
			R^2					$s(e)$		
Small	0.06	0.08	0.09	0.10	0.10	7.50	6.57	6.00	5.68	5.95
2	0.08	0.10	0.13	0.17	0.12	6.97	6.09	5.45	4.87	5.69
3	0.10	0.12	0.15	0.17	0.14	6.38	5.35	4.86	4.48	5.28
4	0.11	0.14	0.17	0.21	0.15	5.63	5.04	4.57	4.39	5.31
Big	0.13	0.15	0.16	0.17	0.12	4.61	4.33	4.00	3.89	4.55

Dependent variable: Excess returns on government and corporate bonds

	1-5G	6-10G	Aaa	Aa	A	Baa	LG
m	0.45	0.72	1.02	0.99	1.00	1.01	0.81
$t(m)$	31.73	38.80	99.94	130.44	139.80	56.24	18.05
d	0.25	0.27	0.94	0.96	1.02	1.10	1.01
$t(d)$	9.51	7.85	48.95	67.54	75.74	32.33	11.95
R^2	0.79	0.87	0.97	0.98	0.98	0.90	0.49
$s(e)$	0.57	0.75	0.41	0.30	0.29	0.72	1.80

[a] $TERM$ is $LTG-RF$, where LTG is the monthly percent long-term government bond return and RF is the one-month Treasury bill rate, observed at the beginning of the month. DEF is $CB-LTG$, where CB is the return on a proxy for the market portfolio of corporate bonds.

The seven bond portfolios used as dependent variables in the excess-return regressions are 1- to 5-year and 6- to 10-year governments (1–5G and 6–10G) and corporate bonds rated Aaa, Aa, A, Baa, and below Baa (LG) by Moody's. The 25 size-BE/ME stock portfolios are formed as follows. Each year t from 1963 to 1991 NYSE quintile breakpoints for size (ME, stock price times shares outstanding), measured at the end of June, are used to allocate NYSE, Amex, and NASDAQ stocks to five size quintiles. Similarly, NYSE quintile breakpoints for BE/ME are used to allocate NYSE, Amex, and NASDAQ stocks to five book-to-market equity quintiles. In BE/ME, BE is book common equity for the fiscal year ending in calendar year $t-1$, and ME is for the end of December of $t-1$. The 25 size-BE/ME portfolios are formed as the intersections of the five size and the five BE/ME groups. Value-weighted monthly percent returns on the portfolios are calculated from July of year t to June of $t+1$.

R^2 and the residual standard error, $s(e)$, are adjusted for degrees of freedom.

The slopes on *TERM* and *DEF* allow direct comparisons of the common variation in stock and bond returns tracked by the term-structure variables. Interestingly, the common variation captured by *TERM* and *DEF* is, if anything, stronger for stocks than for bonds. Most of the *DEF* slopes for stocks are bigger than those for bonds. The *TERM* slopes for stocks (all close to 1) are similar to the largest slopes produced by bonds.

As one might expect, however, the fractions of return variance explained by *TERM* and *DEF* are higher for bonds. In the bond regression, R^2 ranges from 0.49 for low-grade corporates to 0.97 and 0.98 for high-grade corporates. In contrast, R^2 ranges from 0.06 to 0.21 for stocks. Thus, *TERM* and *DEF* clearly identify shared variation in stock and bond returns, but for stocks and low-grade bonds, there is plenty of variation left to be explained by stock-market factors.

There is an interesting pattern in the slopes for *TERM*. The slopes increase from 0.45 to 0.72 for 1- to 5-year and 6- to 10-year governments, and then settle at values near 1 for four of the five long-term corporate bond portfolios. (The low-grade portfolio LG, with a slope of 0.81, is the exception.) As one would expect, long-term bonds are more sensitive than short-term bonds to the shifts in interest rates measured by *TERM*. What is striking, however, is that the 25 stock portfolios have *TERM* slopes like those for long-term bonds. This suggests that the risk captured by *TERM* results from shocks to discount rates that affect long-term securities, bonds and stocks, in about the same way.

There are interesting parallels between the *TERM* slopes observed here and our earlier evidence that yield spreads predict bond and stock returns. In Fama and French (1989), we find that a spread of long-term minus short-term bond yields (an ex ante version of *TERM*) predicts stock and bond returns, and captures about the same variation through time in the expected returns on long-term bonds and stocks. We conjectured that the yield spread captures variation in a term premium for discount-rate changes that affect all long-term securities in about the same way. The similar slopes on *TERM* for long-term bonds and stocks observed here seem consistent with that conjecture.

Our earlier work also finds that the return premium predicted by the long-term minus short-term yield spread wanders between positive and negative values, and is on average close to 0. This parallels the evidence here (table 2) that the average premium for the common risk associated with shifts in interest rates (the average value of *TERM*) is close to 0.

The pattern in the *DEF* slopes in table 3 is also interesting. The returns on small stocks are more sensitive to the risk captured by *DEF* than the returns on big stocks. The *DEF* slopes for stocks tend to be larger than those for corporate bonds, which are larger than those for governments. *DEF* thus seems to capture a common 'default' risk in returns that increases from government bonds to corporates, from bonds to stocks, and from big stocks to small stocks. Again, there is an interesting parallel between this pattern in the *DEF* slopes and the similar pattern observed in Fama and French (1989) in time-series regressions of stock and bond returns on an ex ante version of *DEF* (a spread of low-grade minus high-grade bond yields).

Using the Fama–Macbeth (1973) cross-section regression approach and stock portfolios formed on ranked values of size, Chan, Chen, and Hsieh (1985) and Chen, Roll, and Ross (1986) find that the cross-section of slopes on a variable like *DEF* goes a long way toward explaining the negative relation between size and average stock returns. Given the negative relation between size and the slopes on *DEF* in table 3, it is easy to see why the *DEF* slopes work well in cross-section return regressions for size portfolios.

Our time-series regressions suggest, however, that *DEF* cannot explain the size effect in average stock returns. In the time-series regressions, the average premium for a unit of *DEF* slope is the mean of *DEF*, a tiny 0.02% per month. Likewise, the average *TERM* return is only 0.06% per month. As a result, we shall see that the intercepts in the regressions of stock returns on *TERM* and *DEF* leave strong size and book-to-market effects in average returns. We shall also find that when the stock-market factors are added to the regressions, the negative relation between size and the *DEF* slopes in table 3 disappears.

4.2. *Stock-market factors*

The role of stock-market factors in returns is developed in three steps. We examine (a) regressions that use the excess market return, *RM–RF*, to explain excess bond and stock returns, (b) regressions that use *SMB* and *HML*, the mimicking returns for the size and book-to-market factors, as explanatory variables, and (c) regressions that use *RM–RF*, *SMB*, and *HML*. The three-factor regressions work well for stocks, but the one- and two-factor regressions help explain why.

The Market – Table 4 shows, not surprisingly, that the excess return on the market portfolio of stocks, *RM–RF*, captures more common variation in stock returns than the term-structure factors in table 3. For later purposes, however, the important fact is that the market leaves much variation in stock returns that might be explained by other factors. The only R^2 values near 0.9 are for the big-stock low-book-to-market portfolios. For small-stock and high-*BE*/*ME* portfolios, R^2 values less than 0.8 or 0.7 are the rule. These are the stock portfolios for which the size and book-to-market factors, *SMB* and *HML*, will have their best shot at showing marginal explanatory power.

The market portfolio of stocks also captures common variation in bond returns. Although the market βs are much smaller for bonds than for stocks, they are 5 to 12 standard errors from 0. Consistent with intuition, β is higher for corporate bonds than for governments and higher for low-grade than for high-grade bonds. The β for low-grade bonds (LG) is 0.30, and *RM–RF* explains a tidy 29% of the variance of the LG return.

Table 4

Regressions of excess stock and bond returns (in percent) on the excess stock-market return, $RM-RF$: July 1963 to December 1991, 342 months.[a]

$$R(t) - RF(t) = a + b[RM(t) - RF(t)] + e(t)$$

Dependent variable: Excess returns on 25 stock portfolios formed on size and book-to-market equity

Size quintile	Book-to-market equity (BE/ME) quintiles									
	Low	2	3	4	High	Low	2	3	4	High
	b					$t(b)$				
Small	1.40	1.26	1.14	1.06	1.08	26.33	28.12	27.01	25.03	23.01
2	1.42	1.25	1.12	1.02	1.13	35.76	35.56	33.12	33.14	29.04
3	1.36	1.15	1.04	0.96	1.08	42.98	42.52	37.50	35.81	31.16
4	1.24	1.14	1.03	0.98	1.10	51.67	55.12	46.96	37.00	32.76
Big	1.03	0.99	0.89	0.84	0.89	51.92	61.51	43.03	35.96	27.75
	R^2					$s(e)$				
Small	0.67	0.70	0.68	0.65	0.61	4.46	3.76	3.55	3.56	3.92
2	0.79	0.79	0.76	0.76	0.71	3.34	2.96	2.85	2.59	3.25
3	0.84	0.84	0.80	0.79	0.74	2.65	2.28	2.33	2.26	2.90
4	0.89	0.90	0.87	0.80	0.76	2.01	1.73	1.84	2.21	2.83
Big	0.89	0.92	0.84	0.79	0.69	1.66	1.35	1.73	1.95	2.69

Dependent variable: Excess returns on government and corporate bonds

	1–5G	6–10G	Aaa	Aa	A	Baa	LG
b	0.08	0.13	0.19	0.20	0.21	0.22	0.30
$t(b)$	5.24	5.57	7.53	8.14	8.42	8.73	11.90
R^2	0.07	0.08	0.14	0.16	0.17	0.18	0.29
$s(e)$	1.21	1.95	2.17	2.05	2.05	2.12	2.12

[a]RM is the value-weighted monthly percent return on all the stocks in the 25 size-BE/ME portfolios, plus the negative-BE stocks excluded from the 25 portfolios. RF is the one-month Treasury bill rate, observed at the beginning of the month.

The seven bond portfolios used as dependent variables in the excess-return regressions are 1- to 5-year and 6- to 10-year governments (1–5G and 6–10G) and corporate bonds rated Aaa, Aa, A, Baa, and below Baa (LG) by Moody's. The 25 size-BE/ME stock portfolios are formed as follows. Each year t from 1963 to 1991 NYSE quintile breakpoints for size (ME, stock price times shares outstanding), measured at the end of June, are used to allocate NYSE, Amex, and NASDAQ stocks to five size quintiles. Similarly, NYSE quintile breakpoints for BE/ME are used to allocate NYSE, Amex, and NASDAQ stocks to five book-to market equity quintiles. In BE/ME, BE is book common equity for the fiscal year ending in calendar year $t-1$, and ME is for the end of December of $t-1$. The 25 size-BE/ME portfolios are formed as the intersections of the five size and the five BE/ME groups. Value-weighted monthly percent returns on the portfolios are calculated from July of year t to June of $t+1$.

R^2 and the residual standard error, $s(e)$, are adjusted for degrees of freedom.

SMB and HML – Table 5 shows that in the absence of competition from the market portfolio, *SMB* and *HML* typically capture substantial time-series variation in stock returns; 20 of the 25 R^2 values are above 0.2 and eight are above 0.5. Especially for the portfolios in the larger-size quintile, however, *SMB* and *HML* leave common variation in stock returns that is picked up by the market portfolio in table 4.

The Market, SMB, and HML – Table 5 says that, used alone, *SMB* and *HML* have little power to explain bond returns. Table 6 shows that when the excess market return is also in the regressions, each of the three stock-market factors captures variation in bond returns. We shall find, however, that adding the term-structure factors to the bond regressions largely kills the explanatory power of the stock-market factors. Thus the apparent role of the stock-market factors in bond returns in table 6 probably results from covariation between the term-structure and stock-market factors.

The interesting regressions in table 6 are for stocks. Not surprisingly, the three stock-market factors capture strong common variation in stock returns. The market βs for stocks are all more than 38 standard errors from 0. With one exception, the t-statistics on the *SMB* slopes for stocks are greater than 4; most are greater than 10. *SMB*, the mimicking return for the size factor, clearly captures shared variation in stock returns that is missed by the market and by *HML*. Moreover, the slopes on *SMB* for stocks are related to size. In every book-to-market quintile, the slopes on *SMB* decrease monotonically from smaller- to bigger-size quintiles.

Similarly, the slopes on *HML*, the mimicking return for the book-to-market factor, are systematically related to *BE/ME*. In every size quintile of stocks, the *HML* slopes increase monotonically from strong negative values for the lowest-*BE/ME* quintile to strong positive values for the highest-*BE/ME* quintile. Except for the second *BE/ME* quintile, where the slopes pass from negative to positive, the *HML* slopes are more than five standard errors from 0. *HML* clearly captures shared variation in stock returns, related to book-to-market equity, that is missed by the market and by *SMB*.

Given the strong slopes on *SMB* and *HML* for stocks, it is not surprising that adding the two returns to the regressions results in large increases in R^2. For stocks, the market alone produces only two (of 25) R^2 values greater than 0.9 (table 4); in the three-factor regressions (table 6), R^2 values greater than 0.9 are routine (21 of 25). For the five portfolios in the smallest-size quintile, R^2 increases from values between 0.61 and 0.70 in table 4 to values between 0.94 and 0.97 in table 6. Even the lowest three-factor R^2 for stocks, 0.83 for the portfolio in the largest-size and highest-*BE/ME* quintiles, is much larger than the 0.69 generated by the market alone.

Adding *SMB* and *HML* to the regressions has an interesting effect on the market βs for stocks. In the one-factor regressions of table 4, the β for the portfolio of stocks in the smallest-size and lowest-*BE/ME* quintiles is 1.40. At

Table 5

Regressions of excess stock and bond returns (in percent) on the mimicking returns for the size (SMB) and book-to-market equity (HML) factors: July 1963 to December 1991, 342 months.[a]

$$R(t) - RF(t) = a + sSMB(t) + hHML(t) + e(t)$$

Dependent variable: Excess returns on 25 stock portfolios formed on size and book-to-market equity

Size quintile	Low	2	3	4	High	Low	2	3	4	High
	Book-to-market equity (BE/ME) quintiles									
			s					$t(s)$		
Small	1.93	1.73	1.63	1.59	1.67	22.52	21.38	21.88	22.30	22.16
2	1.52	1.46	1.35	1.18	1.40	17.23	17.68	17.08	15.47	16.42
3	1.28	1.12	1.05	0.93	1.16	14.43	13.89	13.42	12.13	13.45
4	0.86	0.82	0.77	0.72	0.95	10.16	9.64	9.29	8.57	10.02
Big	0.28	0.35	0.22	0.29	0.44	3.70	4.39	2.79	3.69	5.02
			h					$t(h)$		
Small	−0.95	−0.57	−0.35	−0.18	0.01	−9.72	−6.19	−4.10	−2.20	0.16
2	−1.23	−0.66	−0.38	−0.16	0.00	−12.25	−7.02	−4.20	−1.82	0.05
3	−1.09	−0.65	−0.31	−0.11	−0.01	−10.84	−7.07	−3.43	−1.23	−0.12
4	−1.11	−0.65	−0.36	−0.11	−0.01	−11.43	−6.69	−3.80	−1.12	−0.09
Big	−1.07	−0.65	−0.42	−0.06	0.08	−12.46	−7.07	−4.64	−0.66	0.81
			R^2					$s(e)$		
Small	0.65	0.60	0.60	0.60	0.59	4.57	4.31	3.98	3.79	4.01
2	0.59	0.53	0.49	0.42	0.44	4.68	4.41	4.20	4.06	4.53
3	0.51	0.43	0.37	0.31	0.35	4.71	4.31	4.19	4.10	4.60
4	0.43	0.30	0.24	0.18	0.23	4.53	4.55	4.40	4.48	5.06
Big	0.34	0.18	0.08	0.04	0.06	4.02	4.27	4.20	4.19	4.69

	Dependent variable: Excess returns on government and corporate bonds						
	1–5G	6–10G	Aaa	Aa	A	Baa	LG
s	−0.02	−0.06	−0.00	0.00	0.03	0.09	0.19
$t(s)$	−0.66	−1.50	−0.15	0.22	0.77	1.99	4.19
h	0.00	−0.03	−0.02	−0.01	−0.00	0.02	0.00
$t(h)$	0.24	−0.71	−0.45	−0.22	−0.05	0.46	0.15
R^2	−0.00	0.00	−0.00	−0.00	−0.00	0.00	0.04
$s(e)$	1.26	2.03	2.34	2.24	2.25	2.34	2.46

[a]*SMB* (small minus big), the return on the mimicking portfolio for the common size factor in stock returns, is the difference each month between the simple average of the percent returns on the three small-stock portfolios (S/L, S/M, and S/H) and the simple average of the returns on the three big-stock portfolios (B/L, B/M, and B/H). *HML* (high minus low), the return on the mimicking portfolio for the common book-to-market equity factor in returns, is the difference each month between the simple average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low-BE/ME portfolios (S/L and B/L).

The seven bond portfolios used as dependent variables in the excess-return regressions are 1- to 5-year and 6- to 10-year governments (1–5G and 6–10G) and corporate bonds rated Aaa, Aa, A, Baa, and below Baa (LG) by Moody's. The 25 size-BE/ME stock portfolios are formed as follows. Each year t from 1963 to 1991 NYSE quintile breakpoints for size (ME, stock price times shares outstanding), measured at the end of June, are used to allocate NYSE, Amex, and NASDAQ stocks to five size quintiles. Similarly, NYSE quintile breakpoints for BE/ME are used to allocate NYSE, Amex, and NASDAQ stocks to five book-to-market equity quintiles. In BE/ME, BE is book common equity for the fiscal year ending in calendar year $t-1$, and ME is for the end of December of $t-1$. The 25 size-BE/ME portfolios are formed as the intersections of the five size and the five BE/ME groups. Value-weighted percent monthly returns on the portfolios are calculated from July of year t to June of $t+1$.

R^2 and the residual standard error, $s(e)$, are adjusted for degrees of freedom.

Table 6

Regressions of excess stock and bond returns (in percent) on the excess market return ($RM-RF$) and the mimicking returns for the size (SMB) and book-to-market equity (HML) factors: July 1963 to December 1991, 342 months.[a]

$$R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + e(t)$$

Dependent variable: Excess returns on 25 stock portfolios formed on size and book-to-market equity

Size quintile	Book-to-market equity (BE/ME) quintiles: Low	2	3	4	High	Low	2	3	4	High
			b					$t(b)$		
Small	1.04	1.02	0.95	0.91	0.96	39.37	51.80	60.44	59.73	57.89
2	1.11	1.06	1.00	0.97	1.09	52.49	61.18	55.88	61.54	65.52
3	1.12	1.02	0.98	0.97	1.09	56.88	53.17	50.78	54.38	52.52
4	1.07	1.08	1.04	1.05	1.18	53.94	53.51	51.21	47.09	46.10
Big	0.96	1.02	0.98	0.99	1.06	60.93	56.76	46.57	53.87	38.61
			s					$t(s)$		
Small	1.46	1.26	1.19	1.17	1.23	37.92	44.11	52.03	52.85	50.97
2	1.00	0.98	0.88	0.73	0.89	32.73	38.79	34.03	31.66	36.78
3	0.76	0.65	0.60	0.48	0.66	26.40	23.39	21.23	18.62	21.91
4	0.37	0.33	0.29	0.24	0.41	12.73	11.11	9.81	7.38	11.01
Big	−0.17	−0.12	−0.23	−0.17	−0.05	−7.18	−4.51	−7.58	−6.27	−1.18
			h					$t(h)$		
Small	−0.29	0.08	0.26	0.40	0.62	−6.47	2.35	9.66	15.53	22.24
2	−0.52	0.01	0.26	0.46	0.70	−14.57	0.41	8.56	17.24	24.80
3	−0.38	−0.00	0.32	0.51	0.68	−11.26	−0.05	9.75	16.88	19.39
4	−0.42	0.04	0.30	0.56	0.74	−12.51	1.04	8.83	14.84	17.09
Big	−0.46	0.00	0.21	0.57	0.76	−17.03	0.09	5.80	18.34	16.24

	R^2					$s(e)$				
Small	0.94	0.96	0.97	0.97	0.96	1.94	1.44	1.16	1.12	1.22
2	0.95	0.96	0.95	0.95	0.96	1.55	1.27	1.31	1.16	1.23
3	0.95	0.94	0.93	0.93	0.93	1.45	1.41	1.43	1.32	1.52
4	0.94	0.93	0.91	0.89	0.89	1.46	1.48	1.49	1.63	1.88
Big	0.94	0.92	0.88	0.90	0.83	1.16	1.32	1.55	1.36	2.02

Dependent variable: Excess returns on government and corporate bonds

	1-5G	6-10G	Aaa	Aa	A	Baa	LG
b	0.10	0.18	0.25	0.25	0.26	0.27	0.34
$t(b)$	6.45	6.75	8.60	9.30	9.46	9.58	12.22
s	− 0.06	− 0.14	− 0.12	− 0.11	− 0.09	− 0.04	0.04
$t(s)$	− 2.70	− 3.65	− 2.89	− 2.72	− 2.18	− 0.91	0.89
h	0.07	0.08	0.14	0.15	0.16	0.20	0.23
$t(h)$	2.66	1.83	2.77	3.26	3.51	4.08	4.75
R^2	0.10	0.12	0.17	0.20	0.20	0.22	0.33
$s(e)$	1.19	1.91	2.13	2.00	2.01	2.08	2.06

[a] *RM* is the value-weighted percent monthly return on all the stocks in the 25 size–*BE*/*ME* portfolios, plus the negative-*BE* stocks excluded from the 25 portfolios. *RF* is the one-month Treasury bill rate, observed at the beginning of the month. *SMB* (small minus big) is the return on the mimicking portfolio for the size factor in stock returns. *HML* (high minus low) is the return on the mimicking portfolio for the book-to-market factor. (See table 5.)

The seven bond portfolios used as dependent variables are 1- to 5-year and 6- to 10-year governments (1–5G and 6–10G) and corporate bonds rated Aaa, Aa, A, Baa, and below Baa (LG) by Moody's. The 25 size–*BE*/*ME* stock portfolios are formed as follows. Each year t from 1963 to 1991 NYSE quintile breakpoints for size, *ME*, measured at the end of June, are used to allocate NYSE, Amex, and NASDAQ stocks to five size quintiles. Similarly, NYSE quintile breakpoints for *BE*/*ME* are used to allocate NYSE, Amex, and NASDAQ stocks to five book-to-market equity quintiles. In *BE*/*ME*, *BE* is book common equity for the fiscal year ending in calendar year $t - 1$, and *ME* is for the end of December of $t - 1$. The 25 size–*BE*/*ME* portfolios are the intersections of the five size and the five *BE*/*ME* groups. Value-weighted monthly percent returns on the 25 portfolios are calculated from July of t to June of $t + 1$.

R^2 and the residual standard error, $s(e)$, are adjusted for degrees of freedom.

the other extreme, the univariate β for the portfolio of stocks in the biggest-size and highest-BE/ME quintiles is 0.89. In the three-factor regressions of table 6, the βs for these two portfolios are 1.04 and 1.06. In general, adding SMB and HML to the regressions collapses the βs for stocks toward 1.0; low βs move up toward 1.0 and high βs move down. This behavior is due, of course, to correlation between the market and SMB or HML. Although SMB and HML are almost uncorrelated ($-$ 0.08), the correlations between $RM-RF$ and the SMB and HML returns are 0.32 and $-$ 0.38.

4.3. Stock-market and bond-market factors

Used alone, bond-market factors capture common variation in stock returns as well as bond returns (table 3). Used alone, stock-market factors capture shared variation in bond returns as well as stock returns (table 6). These results demonstrate that there is overlap between the stochastic processes for bond and stock returns. We emphasize this point because the joint tests on the stock- and bond-market factors that follow muddy the issue a bit.

First Pass – Table 7 shows that, used together to explain returns, the bond-market factors continue to have a strong role in bond returns and the stock-market factors have a strong role in stock returns. For stocks, adding $TERM$ and DEF to the regressions has little effect on the slopes on the stock-market factors; the slopes on $RM-RF$, SMB, and HML for stocks in table 7a are strong and much like those in table 6. Similarly, adding $RM-RF$, SMB, and HML to the regressions for bonds has little effect on the slopes on $TERM$ and DEF, which are strong and much like those in table 3.

The five-factor regressions in table 7 do, however, seem to contradict the evidence in tables 3 and 6 that there is strong overlap between the return processes for bonds and stocks. Adding the stock-market factors to the regressions for stocks kills the strong slopes on $TERM$ and DEF observed in the two-factor regressions of table 3. The evidence in table 6 that bond returns respond to stock-market factors also largely disappears in table 7b. In the five-factor regressions, only the low-grade bond portfolio, LG, continues to produce nontrivial slopes on the stock-market factors.

Table 7 seems to say that the only shared variation in bond and stock returns comes through low-grade bonds. But tables 3 and 6 say there is strong common variation in bond and stock returns when bond- and stock-market factors are used alone to explain returns. Can we reconcile these results? We argue next that the two term-structure factors are indeed common to bond and stock returns. In the five-factor regressions for stocks, however, the tracks of $TERM$ and DEF are buried in the excess market return, $RM-RF$. In contrast to the two term-structure factors, the three stock-market factors are generally confined to stock returns; except for low-grade bonds, these factors do not spill over into bond returns. In short, we argue that stock returns share three stock-market factors, and the links between stock and bond returns come largely from two shared term-structure factors.

Second Pass: An Orthogonalized Market Factor – If there are multiple common factors in stock returns, they are all in the market return, RM, which is just a value-weighted average of the returns on the stocks in the CRSP–COMPUSTAT sample. The regression of $RM-RF$ on SMB, HML, $TERM$, and DEF for monthly returns of July 1963 to December 1991 illustrates the point:

$$\begin{aligned} RM-RF &= \underset{(2.55)}{0.50} + \underset{(6.48)}{0.44}\,SMB - \underset{(-8.23)}{0.63}\,HML + \underset{(9.09)}{0.81}\,TERM \\ &\quad + \underset{(4.62)}{0.79}\,DEF + e. \end{aligned} \tag{1}$$

The t-statistics are in parentheses below the slopes; the R^2 is 0.38. This regression demonstrates that the market return is a hodgepodge of the common factors in returns. The strong slopes on $TERM$ and DEF produced by $RM-RF$ (the excess return on a proxy for the portfolio of stock-market wealth) are clear evidence that the two term-structure factors capture common variation in stock returns.

The sum of the intercept and the residuals in (1), call it RMO, is a zero-investment portfolio return that is uncorrelated with the four explanatory variables in (1). We can use RMO as an orthogonalized market factor that captures common variation in returns left by SMB, HML, $TERM$, and DEF. Since the stock-market returns, SMB and HML, are largely uncorrelated with the bond-market returns, $TERM$ and DEF (table 2), five-factor regressions that use RMO, SMB, HML, $TERM$, and DEF to explain bond and stock returns will provide a clean picture of the separate roles of bond- and stock-market factors in bond and stock returns. The regressions are in table 8.

The story for the common variation in bond returns in table 8b is like that in table 7b. The bond-market factors, $TERM$ and DEF, have strong roles in bond returns. Some bond portfolios produce slopes on the stock-market factors that are more than two standard errors from 0. But this is mostly because $TERM$ and DEF produce high R^2 values in the bond regressions, so trivial slopes can be reliably different from 0. As in table 7b, only the low-grade bond portfolio (LG) produces nontrivial slopes on the stock-market factors. Otherwise, the stock-market factors don't add much to the shared variation in bond returns captured by $TERM$ and DEF.

For the stock portfolios, the slopes on RMO in the five-factor regressions of table 8a are identical (by construction) to the large slopes on $RM-RF$ in table 7a. The slopes on the size and book-to-market returns in table 8a shift somewhat (up for SMB, down for HML) relative to the slopes in table 7a. But the spreads

Table 7a

Regressions of excess stock returns on 25 stock portfolios formed on size and book-to-market equity (in percent) on the stock-market returns, $RM-RF$, SMB, and HML, and the bond-market returns, $TERM$ and DEF: July 1963 to December 1991, 342 months.[a]

$$R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + mTERM(t) + dDEF(t) + e(t)$$

Size quintile	Book-to-market equity (BE/ME) quintiles: Low	2	3	4	High	Low	2	3	4	High
			b					$t(b)$		
Small	1.06	1.04	0.96	0.92	0.98	35.97	47.65	54.48	54.51	53.15
2	1.12	1.06	0.98	0.94	1.10	47.19	54.95	49.01	54.19	59.00
3	1.13	1.01	0.97	0.95	1.08	50.93	46.95	44.57	47.59	46.92
4	1.07	1.07	1.01	1.00	1.17	48.18	47.55	44.83	41.02	41.02
Big	0.96	1.02	0.98	1.00	1.10	53.87	51.01	41.35	48.29	35.96
			s					$t(s)$		
Small	1.45	1.26	1.20	1.15	1.21	37.02	43.42	50.89	51.36	49.55
2	1.01	0.98	0.89	0.74	0.89	32.06	38.10	33.68	32.12	35.79
3	0.76	0.66	0.60	0.49	0.68	25.82	22.97	20.83	18.54	22.32
4	0.38	0.34	0.30	0.26	0.42	12.71	11.36	9.99	8.05	11.07
Big	−0.17	−0.11	−0.23	−0.17	−0.06	−7.03	−4.07	−7.31	−6.07	−1.44
			h					$t(h)$		
Small	−0.27	0.10	0.27	0.40	0.63	−5.95	2.90	9.82	15.47	22.27
2	−0.51	0.02	0.25	0.44	0.71	−14.01	0.69	8.11	16.50	24.61
3	−0.37	−0.00	0.31	0.50	0.69	−10.81	−0.11	9.28	16.18	19.34
4	−0.42	0.04	0.29	0.53	0.75	−12.09	1.10	8.37	14.20	16.88
Big	−0.46	0.01	0.21	0.58	0.78	−16.85	0.38	5.70	18.16	16.59

	m					$t(m)$				
Small	−0.10	−0.11	−0.05	−0.04	−0.06	−1.93	−2.70	−1.49	−1.19	−1.87
2	−0.05	−0.04	0.07	0.14	−0.05	−1.16	−1.12	1.90	4.33	−1.48
3	−0.04	0.02	0.06	0.09	0.01	−0.91	0.53	1.48	2.48	0.25
4	−0.02	0.00	0.08	0.18	−0.01	−0.55	0.19	1.92	3.98	−0.19
Big	0.03	−0.04	−0.00	−0.04	−0.16	0.82	−0.98	−0.06	−0.98	−2.82
	d					$t(d)$				
Small	−0.17	−0.19	−0.10	0.06	0.02	−1.74	−2.70	−1.76	1.06	0.34
2	−0.12	−0.11	0.04	0.15	−0.07	−1.59	−1.83	0.61	2.64	−1.24
3	−0.09	−0.01	0.07	0.10	−0.16	−1.25	−0.17	1.00	1.51	−2.11
4	−0.11	−0.10	0.04	0.13	−0.12	−1.51	−1.44	0.59	1.64	−1.30
Big	0.06	−0.14	−0.02	−0.07	−0.18	0.97	−2.15	−0.25	−1.08	−1.84
	R^2					$s(e)$				
Small	0.94	0.96	0.97	0.97	0.96	1.93	1.43	1.16	1.11	1.20
2	0.95	0.96	0.95	0.95	0.96	1.55	1.27	1.31	1.13	1.23
3	0.95	0.94	0.93	0.93	0.93	1.45	1.41	1.43	1.31	1.50
4	0.94	0.93	0.91	0.90	0.89	1.46	1.47	1.48	1.59	1.88
Big	0.94	0.92	0.87	0.90	0.83	1.17	1.31	1.55	1.36	2.00

[a]See footnote under table 7b.

Table 7b

Regressions of excess stock returns on government and corporate bonds (in percent) on the stock-market returns, $RM-RF$, SMB, and HML, and the bond-market returns, $TERM$ and DEF: July 1963 to December 1991, 342 months.[a]

$$R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + mTERM(t) + dDEF(t) + e(t)$$

	Bond portfolio						
	1-5G	6–10G	Aaa	Aa	A	Baa	LG
b	− 0.02	− 0.04	− 0.02	0.00	0.00	0.02	0.18
$t(b)$	− 2.84	− 3.14	− 2.96	0.06	1.05	1.99	7.39
s	0.00	− 0.02	− 0.02	− 0.01	0.00	0.05	0.08
$t(s)$	0.30	− 1.12	− 2.28	− 2.42	0.40	3.20	2.34
h	0.00	− 0.02	− 0.02	− 0.00	0.00	0.04	0.12
$t(h)$	0.44	− 1.29	− 2.46	− 0.40	0.90	2.39	3.13
m	0.47	0.75	1.03	0.99	1.00	0.99	0.64
$t(m)$	30.01	36.84	93.30	117.30	124.19	50.50	14.25
d	0.27	0.32	0.97	0.97	1.02	1.05	0.80
$t(d)$	9.87	8.77	49.25	65.04	71.51	30.33	9.92
R^2	0.80	0.87	0.97	0.98	0.98	0.91	0.58
$s(e)$	0.56	0.73	0.40	0.30	0.29	0.70	1.63

[a] RM is the value-weighted monthly percent return on all stocks in the 25 size-BE/ME portfolios, plus the negative-BE stocks excluded from the portfolios. RF is the one-month Treasury bill rate, observed at the beginning of the month. SMB (small minus big) is the difference each month between the simple average of the returns on the three small-stock portfolios (S/L, S/M, and S/H) and the simple average of the returns on the three big-stock portfolios (B/L, B/M, and B/H). HML (high minus low) is the difference each month between the simple average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low-BE/ME portfolios (S/L and B/L). $TERM$ is $LTG-RF$, where LTG is the long-term government bond return. DEF is $CB-LTG$, where CB is the return on a proxy for the market portfolio of corporate bonds.

The seven bond portfolios used as dependent variables in the excess-return regressions are 1- to 5-year and 6- to 10-year governments (1–5G and 6–10G) and corporate bonds rated Aaa, Aa, A, Baa, and below Baa (LG) by Moody's. The 25 size-BE/ME stock portfolios are formed as follows. Each year t from 1963 to 1991 NYSE quintile breakpoints for size (ME, stock price times shares outstanding), measured at the end of June, are used to allocate NYSE, Amex, and NASDAQ stocks to five size quintiles. Similarly, NYSE quintile breakpoints for BE/ME are used to allocate NYSE, Amex, and NASDAQ stocks to five book-to-market quintiles. In BE/ME, BE is book common equity for the fiscal year ending in calendar year $t-1$, and ME is for the end of December of $t-1$. The 25 size-BE/ME portfolios are the intersections of the five size and the five BE/ME groups. Value-weighted monthly percent returns on the portfolios are calculated from July of year t to June of $t+1$.

R^2 and the residual standard, error, $s(e)$, are adjusted for degrees of freedom.

in the SMB and HML slopes across the stock portfolios in table 8a are like those in table 7a, and SMB and HML again capture strong shared variation in stock returns.

What changes dramatically in the five-factor regressions of table 8, relative to table 7, are the slopes on the term-structure factors for stocks. The slopes on

TERM are more than 14 standard errors from 0; the *DEF* slopes are more than seven standard errors from 0. The slopes on *TERM* and *DEF* for stocks are like those for bonds. Thus unlike table 7, the five-factor regressions in table 8 say that the term-structure factors capture strong common variation in stock and bond returns.

How do the tracks of the term-structure variables get buried in the five-factor regressions for stocks in table 7a? Table 8a says that stocks load strongly on *RMO*, *TERM*, and *DEF*, but there is little cross-sectional variation in the slopes on these factors. All the stock portfolios produce slopes on *TERM* and *DEF* close to 0.81 and 0.79, the slopes produced by the excess market return in (1). And the stock portfolios all produce slopes close to 1.0 on *RMO* in table 8a, and thus on *RM–RF* in table 7a. Tables 7a and 8a then say that because there is little cross-sectional variation in the slopes on *RM–RF*, *RMO*, *TERM*, and *DEF*, the excess market return in table 7a absorbs the common variation in stock returns associated with *RMO*, *TERM*, and *DEF*. In short, the common variation in stock returns related to the term-structure factors is buried in the excess market return in table 7a.

Is there any reason to prefer the five-factor regressions in table 8 over those in table 7? Only to show that, in addition to the three stock-market factors, there are two bond-market factors in stock returns. Otherwise, the two sets of regressions produce the same R^2 values and thus the same estimates of the total common variation in returns. And the two sets of regressions produce the same intercepts for testing the implications of five-factor models for the cross-section of average stock returns.

5. The cross-section of average returns

The regression slopes and R^2 values in tables 3 to 8 establish that the stock-market returns, *SMB*, *HML*, and *RM–RF* (or *RMO*), and the bond-market returns, *TERM* and *DEF*, proxy for risk factors. They capture common variation in bond and stock returns. Stock returns have shared variation related to three stock-market factors, and they are linked to bond returns through shared variation in two term-structure factors. We next test how well the average premiums for the five proxy risk factors explain the cross-section of average returns on bonds and stocks.

The average-return tests center on the intercepts in the time-series regressions. The dependent variables in the regressions are excess returns. The explanatory variables are excess returns (*RM–RF* and *TERM*) or returns on zero-investment portfolios (*RMO*, *SMB*, *HML*, and *DEF*). Suppose the explanatory returns have minimal variance due to firm-specific factors, so they are good mimicking returns for the underlying state variables or common risk factors of concern to investors. Then the multifactor asset-pricing models of Merton (1973) and Ross

Table 8a

Regressions of excess stock returns on 25 stock portfolios formed on size and book-to-market equity (in percent) on the stock-market returns, RMO, SMB, and HML, and the bond-market returns, $TERM$ and DEF: July 1963 to December 1991, 342 months.[a]

$$R(t) - RF(t) = a + bRMO(t) + sSMB(t) + hHML(t) + mTERM(t) + dDEF(t) + e(t)$$

Size quintile	Book-to-market equity (BE/ME) quintiles: Low	2	3	4	High	Low	2	3	4	High
			b					$t(b)$		
Small	1.06	1.04	0.96	0.92	0.98	35.97	47.65	54.48	54.51	53.15
2	1.12	1.06	0.98	0.94	1.10	47.19	54.95	49.01	54.19	59.00
3	1.13	1.01	0.97	0.95	1.08	50.93	46.95	44.57	47.59	46.92
4	1.07	1.07	1.01	1.00	1.17	48.18	47.55	44.83	41.02	41.02
Big	0.96	1.02	0.98	1.00	1.10	53.87	51.01	41.35	48.29	35.96
			s					$t(s)$		
Small	1.92	1.72	1.62	1.56	1.64	51.96	62.88	73.21	73.72	71.32
2	1.50	1.45	1.33	1.16	1.38	50.66	59.80	53.02	53.20	58.79
3	1.26	1.11	1.03	0.91	1.16	45.37	40.94	37.83	36.47	40.24
4	0.85	0.81	0.75	0.70	0.94	30.49	28.84	26.42	23.02	26.22
Big	0.26	0.34	0.20	0.28	0.43	11.56	13.69	6.85	10.62	11.17
			h					$t(h)$		
Small	−0.94	−0.56	−0.34	−0.18	0.01	−22.65	−18.19	−13.67	−7.49	0.57
2	−1.22	−0.65	−0.37	−0.15	0.01	−36.52	−23.89	−13.09	−6.22	0.51
3	−1.08	−0.64	−0.30	−0.10	0.00	−34.68	−21.18	−9.82	−3.61	0.16
4	−1.09	−0.64	−0.35	−0.10	0.00	−34.85	−20.12	−10.93	−2.83	0.10
Big	−1.07	−0.63	−0.41	−0.05	0.09	−42.62	−22.46	−12.30	−1.75	2.06

	m					$t(m)$				
Small	0.75	0.73	0.73	0.71	0.73	15.66	20.60	25.32	25.67	24.24
2	0.85	0.82	0.86	0.89	0.84	22.08	25.96	26.40	31.68	27.57
3	0.88	0.84	0.84	0.86	0.88	24.21	23.85	23.73	26.34	23.52
4	0.85	0.87	0.90	0.98	0.94	23.24	23.77	24.35	24.76	20.11
Big	0.80	0.79	0.79	0.77	0.73	27.60	24.17	20.42	22.83	14.66
	d					$t(d)$				
Small	0.67	0.63	0.66	0.78	0.79	7.25	9.20	11.90	14.81	13.73
2	0.76	0.72	0.81	0.89	0.79	10.23	11.94	12.96	16.36	13.57
3	0.80	0.78	0.83	0.84	0.69	11.53	11.64	12.25	13.53	9.63
4	0.74	0.74	0.84	0.91	0.80	10.56	10.48	11.88	12.01	8.98
Big	0.81	0.66	0.75	0.72	0.68	14.56	10.62	10.15	11.04	7.15
	R^2					$s(e)$				
Small	0.94	0.96	0.97	0.97	0.96	1.93	1.43	1.16	1.11	1.20
2	0.95	0.96	0.95	0.95	0.96	1.55	1.27	1.31	1.13	1.23
3	0.95	0.94	0.93	0.93	0.93	1.45	1.41	1.43	1.31	1.50
4	0.94	0.93	0.91	0.90	0.89	1.46	1.47	1.48	1.59	1.88
Big	0.94	0.92	0.87	0.90	0.83	1.17	1.31	1.55	1.36	2.00

[a]See footnote under table 8b.

Table 8b

Regressions of excess returns on government and corporate bonds (in percent) on the stock-market returns, RMO, SMB, and HML, and the bond-market returns, $TERM$ and DEF: July 1963 to December 1991, 342 months.[a]

$$R(t) - RF(t) = a + bRMO(t) + sSMB(t) + hHML(t) + mTERM(t) + dDEF(t) + e(t)$$

	Bond portfolio						
	1–5G	6–10G	Aaa	Aa	A	Baa	LG
b	−0.02	−0.04	−0.02	0.00	0.00	0.02	0.18
$t(b)$	−2.84	−3.14	−2.96	0.06	1.05	1.99	7.39
s	−0.00	−0.03	−0.03	−0.01	0.00	0.06	0.16
$t(s)$	−0.68	−2.30	−3.47	−2.55	0.80	4.09	5.09
h	0.02	−0.00	−0.01	−0.00	0.00	0.03	0.00
$t(h)$	1.76	−0.00	−1.36	−0.47	0.52	1.72	0.12
m	0.45	0.72	1.02	0.99	1.00	1.01	0.79
$t(m)$	32.09	39.55	102.65	130.93	139.11	57.34	19.56
d	0.25	0.29	0.95	0.97	1.02	1.07	0.94
$t(d)$	9.46	8.25	50.04	67.08	74.00	31.77	12.09
R^2	0.80	0.87	0.97	0.98	0.98	0.91	0.58
$s(e)$	0.56	0.73	0.40	0.30	0.29	0.70	1.63

[a]RMO, the orthogonalized market return, is the sum of intercept and residuals from the regression of $RM-RF$ on SMB, HML, $TERM$, and DEF. RM is the value-weighted monthly percent return on all stocks in the 25 size-BE/ME portfolios, plus the negative-BE stocks excluded from the portfolios. RF is the one-month Treasury bill rate, observed at the beginning of the month. SMB (small minus big), the return on the mimicking portfolio for the common size factor in stock returns, is the difference each month between the simple average of the returns on the three small-stock portfolios (S/L, S/M, and S/H) and the simple average of the returns on the three big-stock portfolios (B/L, B/M, and B/H). HML (high minus low), the return on the mimicking portfolio for the common book-to-market equity factor in returns, is the difference each month between the simple average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low-BE/ME portfolios (S/L and B/L). $TERM$ is $LTG-RF$, where LTG is the long-term government bond return. DEF is $CB-LTG$, where CB is the return on a proxy for the market portfolio of corporate bonds.

The seven bond portfolios used as dependent variables in the excess-return regressions are 1- to 5-year and 6- to 10-year governments (1–5G and 6–10G) and bonds rated Aaa, Aa, A, Baa, and below Baa (LG) by Moody's. The 25 size-BE/ME stock portfolios are formed as follows. Each year t from 1963 to 1991 NYSE quintile breakpoints for size (ME, stock price times shares outstanding), measured at the end of June, are used to allocate NYSE, Amex, and NASDAQ stocks to five size quintiles. NYSE quintile breakpoints for BE/ME are also used to allocate NYSE, Amex, and NASDAQ stocks to five book-to-market equity quintiles. In BE/ME, BE is book common equity for the fiscal year ending in calendar year $t-1$, and ME is for the end of December of $t-1$. The 25 size-BE/ME portfolios are the intersections of the five size and the five BE/ME groups. Value-weighted monthly percent returns on the portfolios are calculated from July of year t to June of $t+1$.

R^2 and the resdiual standard error, $s(e)$, are adjusted for degrees of freedom.

(1976) imply a simple test of whether the premiums associated with any set of explanatory returns suffice to describe the cross-section of average returns: the intercepts in the time-series regressions of excess returns on the mimicking portfolio returns should be indistinguishable from 0.[1]

Since the stock portfolios produce a wide range of average returns, we examine their intercepts first. We are especially interested in whether the mimicking returns *SMB* and *HML* absorb the size and book-to-market effects in average returns, illustrated in table 2. We then examine the intercepts for bonds. Here the issue is whether different factor models predict patterns in average returns that are rejected by the flat average bond returns in table 2.

5.1. The cross-section of average stock returns

RM–RF – When the excess market return is the only explanatory variable in the time-series regressions, the intercepts for stocks (table 9a) show the size effect of Banz (1981). Except in the lowest-*BE/ME* quintile, the intercepts for the smallest-size portfolios exceed those for the biggest by 0.22% to 0.37% per month. The intercepts are also related to book-to-market equity. In every size quintile, the intercepts increase with *BE/ME*; the intercepts for the highest-*BE/ME* quintile exceed those for the lowest by 0.25% to 0.76% per month. These results parallel the evidence in Fama and French (1992a) that, used alone, market βs leave the cross-sectional variation in average stock returns that is related to size and book-to-market equity.

In fact, as in Fama and French (1992a), the simple relation between average return and β for the 25 stock portfolios used here is flat. A regression of average return on β yields a slope of -0.22 with a standard error of 0.31. The Sharpe (1964)–Lintner (1965) model (β suffices to describe the cross-section of average returns and the simple relation between β and average return is positive) fares no better here than in our earlier paper.

SMB and HML – The two-factor time-series regressions of excess stock returns on *SMB* and *HML* produce similar intercepts for the 25 stock portfolios (table 9a). The two-factor regression intercepts are, however, large (around 0.5% per month) and close to or more than two standard errors from 0. Intercepts that are similar in size support the conclusion from the cross-section regressions in Fama and French (1992a) that size and book-to-market factors explain the strong differences in average returns across stocks. But the large intercepts also say that *SMB* and *HML* do not explain the average premium of stock returns over one-month bill returns.

RM–RF, SMB, and HML – Adding the excess market return to the time-series regressions pushes the strong positive intercepts for stocks observed in the

[1] This implication is only an approximation in the Ross (1976) model. See, for example, Shanken (1982).

Table 9a

Intercepts from excess stock return regressions for 25 stock portfolios formed on size and book-to-market equity: July 1963 to December 1991, 342 months.[a]

	Book-to-market equity (BE/ME) quintiles									
	a					$t(a)$				
Size quintile	Low	2	3	4	High	Low	2	3	4	High
	(i) $R(t) - RF(t) = a + mTERM(t) + dDEF(t) + e(t)$									
Small	0.31	0.62	0.71	0.80	0.92	0.75	1.73	2.20	2.61	2.87
2	0.35	0.63	0.77	0.75	0.93	0.93	1.91	2.60	2.85	3.03
3	0.34	0.58	0.60	0.73	0.89	1.00	1.99	2.28	3.01	3.11
4	0.41	0.27	0.49	0.69	0.96	1.34	1.01	1.96	2.88	3.35
Big	0.34	0.30	0.25	0.50	0.53	1.35	1.27	1.17	2.36	2.14
	(ii) $R(t) - RF(t) = a + b[RM(t) - RF(t)] + e(t)$									
Small	− 0.22	0.15	0.30	0.42	0.54	− 0.90	0.73	1.54	2.19	2.53
2	− 0.18	0.17	0.36	0.39	0.53	− 1.00	1.05	2.35	2.79	3.01
3	− 0.16	0.15	0.23	0.39	0.50	− 1.12	1.25	1.82	3.20	3.19
4	− 0.05	− 0.14	0.12	0.35	0.57	− 0.50	− 1.50	1.20	2.91	3.71
Big	− 0.04	− 0.07	− 0.07	0.20	0.21	− 0.49	− 0.95	− 0.70	1.89	1.41

				(iii) $R(t) - RF(t) = a + sSMB(t) + hHML(t) + e(t)$						
Small	0.24	0.46	0.49	0.53	0.55	0.97	1.92	2.24	2.52	2.49
2	0.52	0.58	0.64	0.58	0.64	2.00	2.40	2.76	2.61	2.56
3	0.52	0.61	0.52	0.60	0.66	2.00	2.58	2.25	2.66	2.61
4	0.69	0.39	0.50	0.62	0.79	2.78	1.55	2.07	2.51	2.85
Big	0.76	0.52	0.43	0.51	0.44	3.41	2.23	1.84	2.20	1.70
				(iv) $R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + e(t)$						
Small	−0.34	−0.12	−0.05	0.01	0.00	−3.16	−1.47	−0.73	0.22	0.14
2	−0.11	−0.01	0.08	0.03	0.02	−1.24	−0.20	1.04	0.51	0.34
3	−0.11	0.04	−0.04	0.05	0.05	−1.42	0.47	−0.47	0.71	0.56
4	0.09	−0.22	−0.08	0.03	0.13	1.07	−2.65	−0.99	0.33	1.24
Big	0.21	−0.05	−0.13	−0.05	−0.16	3.27	−0.67	−1.46	−0.69	−1.41
				(v) $R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + mTERM(t) + dDEF(t) + e(t)$						
Small	−0.35	−0.13	−0.05	0.01	0.00	−3.24	−1.58	−0.79	0.20	0.09
2	−0.11	−0.02	0.08	0.04	0.02	−1.29	−0.24	1.10	0.67	0.29
3	−0.12	0.04	−0.03	0.06	0.05	−1.45	0.48	−0.42	0.79	0.56
4	0.08	−0.22	−0.08	0.04	0.13	1.04	−2.67	−0.94	0.47	1.23
Big	0.21	−0.05	−0.13	−0.06	−0.17	3.29	−0.72	−1.46	−0.73	−1.51

[a]See footnote under table 9c.

Table 9b

Intercepts form excess bond return regressions for two government and five corporate bond portfolios: July 1963 to December 1991, 342 months.[a]

	Bond portfolio						
	1–5G	6–10G	Aaa	Aa	A	Baa	LG
	(i) $R(t) - RF(t) = a + mTERM(t) + dDEF(t) + e(t)$						
a	0.08	0.09	− 0.02	− 0.00	− 0.00	0.06	0.06
$t(a)$	2.70	2.16	− 1.10	− 0.55	− 0.29	1.42	0.67
	(ii) $R(t) - RF(t) = a + b[RM(t) - RF(t)] + e(t)$						
a	0.08	0.08	− 0.03	− 0.02	− 0.01	0.04	0.00
$t(a)$	1.27	0.76	− 0.24	− 0.15	− 0.11	0.37	0.03
	(iii) $R(t) - RF(t) = a + sSMB(t) + hHML(t) + e(t)$						
a	0.12	0.16	0.07	0.07	0.07	0.11	0.08
$t(a)$	1.70	1.47	0.52	0.58	0.55	0.82	0.58
	(iv) $R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + e(t)$						
a	0.06	0.07	− 0.07	− 0.07	− 0.08	− 0.05	− 0.11
$t(a)$	0.89	0.62	− 0.62	− 0.64	− 0.69	− 0.41	− 1.00
	(v) $R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + mTERM(t) + dDEF(t) + e(t)$						
a	0.09	0.11	− 0.00	− 0.00	− 0.00	0.02	− 0.07
$t(a)$	2.84	2.77	− 0.17	− 0.25	− 0.57	0.52	− 0.77

[a]See footnote under table 9c.

two-factor (SMB and HML) regressions to values close to 0. Only three of the 25 intercepts in the three-factor regressions differ from 0 by more than 0.2% per month; 16 are within 0.1% of 0. Intercepts close to 0 say that the regressions that use $RM-RF$, SMB, and HML to absorb common time-series variation in returns do a good job explaining the cross-section of average stock returns.

There is an interesting story for the smaller intercepts obtained when the excess market return is added to the two-factor (SMB and HML) regressions. In the three-factor regressions, the stock portfolios produce slopes on $RM-RF$ close to 1. The average market risk premium (0.43% per month) then absorbs the similar strong positive intercepts observed in the regressions of stock returns on SMB and HML. In short, the size and book-to-market factors can explain the differences in average returns across stocks, but the market factor is needed to explain why stock returns are on average above the one-month bill rate.

TERM and DEF – Table 9a shows that adding the term-structure factors, $TERM$ and DEF, to the time-series regressions for stocks has almost no effect on the intercepts produced by the three stock-market factors. Likewise, in spite of the strong slopes on $TERM$ and DEF when they are used alone to explain stock

Table 9c

F-statistics testing the intercepts in the excess-return regressions against 0 and matching probability levels of bootstrap and F-distributions.[a]

	Regression (from tables 9a and 9b)				
	(i)	(ii)	(iii)	(iv)	(v)
F-statistic	2.09	1.91	1.78	1.56	1.66
Probability level					
Bootstrap	0.998	0.996	0.985	0.951	0.971
F-distribution	0.999	0.996	0.990	0.961	0.975

[a] RM is the value-weighted monthly percent return on all stocks in the 25 size-BE/ME portfolios, plus the negative-BE stocks excluded from the 25 portfolios. RF is the one-month Treasury bill rate, observed at the beginning of the month. SMB (small minus big), the return on the mimicking portfolio for the common size factor in stock returns, is the difference each month between the simple average of the returns on the three small-stock portfolios (S/L, S/M, and S/H) and the simple average of the returns on the three big-stock portfolios (B/L, B/M, and B/H). HML (high minus low), the return on the mimicking portfolio for the common book-to-market equity factor in returns, is the difference each month between the simple average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low-BE/ME portfolios (S/L and B/L). $TERM$ is $LTG-RF$, where LTG is the long-term government bond return. DEF is $CB-LTG$, where CB is the return on a proxy for the market portfolio of corporate bonds.

The seven bond portfolios used as dependent variables in the excess-return regressions are 1- to 5-year and 6- to 10-year governments (1–5G and 6–10G) and corporate bonds rated Aaa, Aa, A, Baa, and below Baa (LG) by Moody's. The 25 size-BE/ME stock portfolios are formed as follows. Each year t from 1963 to 1991 NYSE quintile breakpoints for size (ME, stock price times shares outstanding), measured at the end of June, are used to allocate NYSE, Amex, and NASDAQ stocks to five size quintiles. NYSE quintile breakpoints for BE/ME are also used to allocate NYSE, Amex, and NASDAQ stocks to five book-to-market equity quintiles. In BE/ME, BE is book common equity for the fiscal year ending in calendar year $t-1$, and ME is for the end of December of $t-1$. The 25 size-BE/ME portfolios are the intersections of the five size and the five BE/ME groups. Value-weighted monthly percent returns on the portfolios are calculated from July of year t to June of $t+1$.

Regressions (i)–(v) in table 9c correspond to the regressions in tables 9a and 9b. The F-statistic is

$$F = (A'\Sigma^{-1}A)(N-K-L+1)/(L*(N-K)*\omega_{1,1}),$$

where $N = 342$ observations, $L = 32$ regressions, K is 1 plus the number of explanatory variables in the regression, A is the (column) vector of the 32 regression intercepts, Σ ($L \times L$) is the unbiased covariance matrix of the residuals from the 32 regressions, and $\omega_{1,1}$ is the diagonal element of $(X'X)^{-1}$ corresponding to the intercept. Gibbons, Ross, and Shanken (1989) show that this statistic has an F-distribution with L and $N-K-L+1$ degrees of freedom under the assumption that the returns and explanatory variables are normal and the true intercepts are 0.

In the bootstrap simulations, the slopes (with intercepts set to 0), explanatory variables, and residuals from the regressions for July 1963 to December 1991 in tables 3 to 7 are used to generate 342 monthly excess returns for the 25 stock and seven bond portfolios for each regression model. These model returns and the explanatory returns, $RM-RF$, SMB, HML, $TERM$, and DEF, for July 1963 to December 1991, are the population for the simulations. Each simulation takes a random sample, with replacement, of 342 paired observations (the same set of observations for each of the five regression models) on the model returns and the explanatory variables, and estimates the regressions. For each model, the table shows the proportion of 10,000 simulations in which the F-statistic is smaller than the empirical estimate. The table also shows the probability that a value drawn from an F-distribution is smaller than the empirical estimate.

returns (table 3), the two variables produce intercepts close to the average excess returns for the 25 stock portfolios in table 2.

The reason for these results is straightforward. The average *TERM* and *DEF* returns (the average risk premiums for the term-structure factors) are puny, 0.06% and 0.02% per month. The high volatility of *TERM* and *DEF* (table 2) allows them to capture substantial common variation in bond and stock returns in the two-factor regressions of table 3 and the five-factor regressions of table 8. But the low average *TERM* and *DEF* returns imply that the two term-structure factors can't explain much of the cross-sectional variation in average stock returns.

5.2. The cross-section of average bond returns

Tables 3, 7b and 8b say that the common variation in bond returns is dominated by the bond-market factors, *TERM* and *DEF*. Only the low-grade bond portfolio (LG) has nontrivial slopes on the stock-market factors when *TERM* and *DEF* are in the bond regressions. Like the average values of *TERM* and *DEF*, the average excess returns on the bond portfolios are close to 0 (table 2), so it is not surprising that the intercepts in the time-series regressions for bonds (table 9b) are close to 0.

Do low average *TERM* and *DEF* premiums imply that the term-structure factors are irrelevant in a well-specified asset-pricing model? Hardly. *TERM* and *DEF* are the dominant variables in the common variation in bond returns. Moreover, Fama and French (1989) and Chen (1991) find that the expected values of variables like *TERM* and *DEF* vary through time and are related to business conditions. The expected value of *TERM*, the term premium for discount-rate risks, is positive around business cycle troughs and negative near peaks. The expected value of the default premium in *DEF* is high when economic conditions are weak and default risks are high, and it is low when business conditions are strong. Thus, the common sensitivity of stocks and bonds to *TERM* and *DEF* implies interesting intertemporal variation in expected stock and bond returns.

5.3. Joint tests on the regression intercepts

We use the F-statistic of Gibbons, Ross, and Shanken (1989) to formally test the hypothesis that a set of explanatory variables produces regression intercepts for the 32 bond and stock portfolios that are all equal to 0. The F-statistics, and bootstrap probability levels, for the five sets of intercepts produced by the explanatory variables in tables 3 to 8 are in table 9c.

The F-tests support the analysis of the intercepts above. The tests reject the hypothesis that the term-structure returns, *TERM* and *DEF*, suffice to explain the average returns on bonds and stocks at the 0.99 level. This confirms the conclusion, obvious from the regression intercepts in table 9a, that the low average *TERM* and *DEF* returns cannot explain the cross-section of average stock returns. The *F*-test rejects the hypothesis that *RM–RF* suffices to explain average returns at the 0.99 level. This confirms that the excess market return cannot explain the size and book-to-market effects in average stock returns. The large positive intercepts for stocks observed when *SMB* and *HML* are the only explanatory variables produce an *F*-statistic that rejects the zero-intercepts hypothesis at the 0.98 level.

In terms of the *F*-test, the three stock-market factors, *RM–RF*, *SMB*, and *HML*, produce the best-behaved intercepts. Nevertheless, the joint test that all intercepts for the seven bond and 25 stock portfolios are 0 rejects at about the 0.95 level. The rejection comes largely from the lowest-*BE/ME* quintile of stocks. Among stocks with the lowest ratios of book-to-market equity (growth stocks), the smallest stocks have returns that are too low (− 0.34% per month, $t = -3.16$) relative to the predictions of the three-factor model, and the biggest stocks have returns that are too high (0.21% per month, $t = 3.27$). Put a bit differently, the rejection of a three-factor model in table 9c is due to the absence of a size effect in the lowest-*BE/ME* quintile. The five portfolios in the lowest-*BE/ME* quintile produce slopes on the size factor *SMB* that are strongly negatively related to size (table 6). But unlike the other *BE/ME* quintiles, average returns in the lowest-*BE/ME* quintile show no relation to size (table 2).

Despite its marginal rejection in the *F*-tests, our view is that the three-factor model does a good job on the cross-section of average stock returns. The rejection of the model simply says that because *RM–RF*, *SMB*, and *HML* absorb most of the variation in the returns on the 25 stock portfolios (the typical R^2 values in table 6 are above 0.93), even small abnormal average returns suffice to show that the three-factor model is just a model, that is, it is false. To answer the important question of whether the model can be useful in applications, the interesting result is that only one of the 25 three-factor regression intercepts for stocks (for the portfolio in both the smallest-size and the lowest-*BE/ME* quintiles) is much different from 0 in practical terms.

Indeed, our view is that the three-factor regressions that use *RM–RF*, *SMB*, and *HML* to explain average returns do surprisingly well, given the simple way the mimicking returns *SMB* and *HML* for the size and book-to-market factors are constructed. The regressions produce intercepts for stocks that are close to 0, even though *SMB* and *HML* surely contain some firm-specific noise as proxies for the risk factors in returns related to size and book-to-market equity.

Adding the term-structure returns, *TERM* and *DEF*, to regressions that also use *RM–RF*, *SMB*, and *HML* as explanatory variables increases *F*. The larger *F* comes from bonds. The five-factor regression intercepts and R^2 values for stocks are close to those produced by the three stock-market factors. But for bonds, adding *TERM* and *DEF* results in much lower residual standard errors, and the increased precision pushes the five-factor intercepts for the two government bond portfolios beyond two standard errors from 0. The two intercepts are, however, rather small, 0.09% and 0.11% per month.

The three stock-market factors produce a lower F, but we think the five-factor regressions provide the best model for returns and average returns on bonds and stocks. $TERM$ and DEF dominate the variation in bond returns. And the variation in the expected values of $TERM$ and DEF with business conditions is an interesting part of the variation through time in the expected returns on stocks and bonds that is missed by the F-test, which is concerned only with long-term average returns.

6. Diagnostics

In this section we check the robustness of our inference that five common risk factors explain the cross-section of expected stock and bond returns. We first use the residuals from the five-factor time-series regressions to check that the regressions capture the variation through time in the cross-section of expected returns. We then examine whether our five risk factors capture the January seasonals in stock and bond returns. Next come split-sample regressions that use one set of stocks in the explanatory returns and another, disjoint, set in the dependent returns. These tests address the concern that the evidence of size and book-to-market factors in the regressions above is spurious, arising only because we use size and book-to-market portfolios for both our dependent and explanatory returns. The last and most interesting tests examine whether the stock-market factors that capture the average returns on size–BE/ME portfolios work as well on portfolios formed on other variables known to be informative about average returns, in particular, earnings/price and dividend/price ratios.

6.1. The predictability of the regression residuals

There is evidence that stock and bond returns can be predicted using (a) dividend yields (D/P), (b) spreads of low-grade over high-grade bond yields (default spreads, DFS), (c) spreads of long-term over short-term bond yields (term spreads, TS), and (d) short-term interest rates. [See Fama (1991) and the references therein.] If our five risk factors capture the cross-section of expected returns, the predictability of stock and bond returns should be embodied in the explanatory returns (the month-by-month risk premiums) in the five-factor regressions. The regression residuals should be unpredictable. To test this hypothesis, we estimate the 32 time-series regressions,

$$e_p(t+1) = k_0 + k_1 D(t)/P(t) + k_2 DFS(t) + k_3 TS(t) + k_4 RF(t) + \eta_p(t+1). \tag{2}$$

The $e_p(t+1)$ in (2) are the time series of residuals for our 25 stock and seven bond portfolios from the five-factor regressions of table 7. The dividend yield, $D(t)/P(t)$, is dividends on the value-weighted portfolio of NYSE stocks for the year ending in month t divided by the value of the portfolio at the end of t. The default spread, $DFS(t)$, is the difference at the end of month t between the yield on a market portfolio of corporate bonds and the long-term government bond yield (from Ibbotson Associates). The term spread, $TS(t)$, is the difference between the long-term government bond yield at the end of month t and the one-month bill rate, $RF(t)$.

The estimates of (2) produce no evidence that the residuals from the five-factor time-series regressions are predictable. In the 32 regressions, 15 produce negative values of R^2 (adjusted for degrees of freedom). Only four of the 32 R^2 values exceed 0.01; the largest is 0.03. Out of 128 (32×4) slopes in the residual regressions, ten are more than two standard errors from 0; they are split evenly between positive and negative values, and they are scattered randomly across the 32 regressions and the four explanatory variables.

The fact that variables known to predict stock and bond returns do not predict the residuals from our five-factor regressions supports our inference that the five risk factors capture the cross-section of expected stock and bond returns. The residual tests are also interesting information on a key regression specification. Since we estimate regression slopes on returns for the entire 1963–1991 period, we implicitly assume that the sensitivities of the dependent returns to the risk factors are constant. If the true slopes vary through time, the regression residuals may be spuriously predictable. The absence of predictability suggests that the assumption of constant slopes is reasonable, at least for the portfolios used here.

6.2. *January seasonals*

Since the work Roll (1983) and Keim (1983), documenting that stock returns, especially returns on small stocks, tend to be higher in January, it is standard in tests of asset-pricing models to look for unexplained January effects. We are leery of judging models on their ability to explain January seasonals. If the seasonals are, in whole or in part, sampling error, the tests can contain a data-snooping bias toward rejection [Lo and MacKinlay (1990)]. Nevertheless, we test for January seasonals in the residuals from our five-factor regressions. Despite our fears, we find that, except for the smallest stocks, residual January seasonals are weak at best. The strong January seasonals in the returns on stocks and bonds are largely absorbed by strong seasonals in our risk factors.

Table 10 shows regressions of returns on a dummy variable that is 1 in January and 0 in other months. The regression intercepts are average returns for non-January months, and the slopes on the dummy measure differences between average January returns and average returns in other months.

Table 10

Tests for January seasonals in the dependent returns, explanatory returns, and residuals from the five-factor regressions: July 1963 to December 1991, 342 months.[a]

$$R(t) = a + bJAN(t) + e$$

	a	b	$t(a)$	$t(b)$	R^2	a	b	$t(a)$	$t(b)$	R^2
Factor	Five-factor explanatory returns									
RM–RF	0.31	1.49	1.22	1.67	0.00					
RMO	0.40	1.19	2.03	1.70	0.00					
SMB	0.05	2.74	0.30	4.96	0.06					
HML	0.21	2.29	1.53	4.70	0.06					
TERM	0.10	− 0.41	0.56	− 0.69	− 0.00					
DEF	− 0.07	1.10	− 0.81	3.56	0.03					
Stock portfolio	Excess stock returns					Five-factor regression residuals				
	Smallest-size quintile									
BE/ME Low	− 0.13	6.31	− 0.30	4.23	0.05	− 0.12	1.51	− 1.17	4.09	0.04
BE/ME 2	0.24	5.62	0.63	4.27	0.05	− 0.05	0.56	− 0.57	2.01	0.00
BE/ME 3	0.31	5.91	0.90	4.93	0.06	− 0.06	0.69	− 0.88	3.06	0.02
BE/ME 4	0.37	6.29	1.14	5.55	0.08	− 0.06	0.76	− 1.02	3.57	0.03
BE/ME High	0.40	7.39	1.20	6.31	0.10	− 0.09	1.13	− 1.41	4.94	0.06
	Size quintile 2									
BE/ME Low	0.20	2.92	0.48	2.04	0.00	0.02	− 0.23	0.21	− 0.74	− 0.00
BE/ME 2	0.37	4.17	1.04	3.34	0.03	0.00	− 0.04	0.04	− 0.15	− 0.00
BE/ME 3	0.53	3.95	1.63	3.48	0.03	0.04	− 0.55	0.62	− 2.16	0.01
BE/ME 4	0.48	4.32	1.65	4.22	0.05	0.02	− 0.22	0.28	− 0.97	− 0.00
BE/ME High	0.55	5.76	1.66	4.99	0.07	− 0.01	0.12	− 0.14	0.49	− 0.00
	Size quintile 3									
BE/ME Low	0.24	2.35	0.62	1.78	0.00	0.04	− 0.49	0.50	− 1.74	0.00
BE/ME 2	0.42	2.87	1.31	2.57	0.02	0.03	− 0.41	0.42	− 1.48	0.00
BE/ME 3	0.43	3.06	1.47	2.99	0.02	0.07	− 0.80	0.83	− 2.90	0.02
BE/ME 4	0.52	3.51	1.92	3.68	0.04	0.04	− 0.46	0.52	− 1.80	0.00
BE/ME High	0.60	4.53	1.91	4.12	0.04	0.03	− 0.34	0.33	− 1.15	0.00

				Size quintile 4						
BE/ME Low	0.39	1.12	1.16	0.95	− 0.00	0.04	− 0.46	0.46	− 1.60	0.00
BE/ME 2	0.21	1.77	0.68	1.65	0.00	0.06	− 0.73	0.73	− 2.54	0.02
BE/ME 3	0.40	2.08	1.40	2.11	0.01	0.08	− 0.93	0.93	− 3.27	0.03
BE/ME 4	0.52	3.12	1.88	3.24	0.03	0.03	− 0.37	0.34	− 1.17	0.00
BE/ME High	0.68	4.45	2.15	4.00	0.04	0.00	− 0.03	0.03	− 0.09	− 0.00
				Biggest-size quintile						
BE/ME Low	0.37	0.34	1.34	0.35	− 0.00	− 0.03	0.38	− 0.48	1.67	0.00
BE/ME 2	0.27	1.11	1.02	1.19	0.00	0.00	− 0.00	0.00	− 0.02	− 0.00
BE/ME 3	0.23	1.11	0.92	1.28	0.00	0.01	− 0.17	0.16	− 0.57	− 0.00
BE/ME 4	0.37	2.38	1.54	2.85	0.02	− 0.00	0.08	− 0.09	0.31	− 0.00
BE/ME High	0.32	3.38	1.17	3.59	0.03	− 0.02	0.25	− 0.18	0.63	− 0.00
Bond portfolio			Excess bond returns					Five-factor regression residuals		
1–5G	0.11	0.05	1.58	0.20	− 0.00	0.00	− 0.04	0.12	− 0.40	− 0.00
6–10G	0.16	− 0.22	1.35	− 0.56	− 0.00	0.00	− 0.11	0.23	− 0.79	− 0.00
Aaa	0.03	0.34	0.21	0.74	− 0.00	0.01	− 0.17	0.62	− 2.17	0.01
Aa	0.03	0.51	0.23	1.15	0.00	0.00	− 0.11	0.53	− 1.85	0.00
A	0.00	0.86	0.05	1.94	0.00	− 0.01	0.12	− 0.60	2.08	0.01
Baa	0.05	1.14	0.35	2.48	0.01	− 0.01	0.14	− 0.29	1.01	0.00
LG	0.00	1.56	0.05	3.17	0.03	− 0.02	0.19	− 0.17	0.58	− 0.00

[a]*JAN*(*t*) is a dummy variable that is 1 if month *t* is January and 0 otherwise. *RMO* is the sum of the intercept and residuals from the regression of *RM–RF* on *SMB*, *HML*, *TERM*, and *DEF*. *RM* is the value-weighted monthly stock-market return. *RF* is the one-month Treasury bill rate, observed at the beginning of the month. *SMB* and *HML* are the returns on the mimicking portfolios for the size and book-to-market equity factors in stock returns. *TERM* is *LTG–RF*, where *LTG* is the long-term government bond return. *DEF* is *CB–LTG*, where *CB* is the return on a proxy for the market portfolio of corporate bonds.

The seven bond portfolios are 1- to 5-year and 6- to 10-year governments (1–5G and 6–10G) and bonds rated Aaa, Aa, A, Baa, and below Baa (LG) by Moody's. The 25 size-*BE/ME* portfolios are formed as the intersections of independent sorts of stocks into size and book-to-market equity quintiles in June of each year from 1963–1991. The variables are described in more detail in table 8.

The table confirms that there are January seasonals in excess stock returns, and the seasonals are related to size. The slopes on the January dummy are all more than 2.92% per month and more than two standard errors from 0 for the portfolios in the two smallest size quintiles. Controlling for BE/ME, the extra January return declines monotonically with increasing size. More interesting, the January seasonal in stock returns is also related to book-to-market equity. In every size quintile, the slopes on the January dummy tend to increase with BE/ME. The extra January return for the two highest-BE/ME portfolios in a size quintile is always at least 2.38% per month and 2.85 standard errors from 0.

January seasonals are not limited to stock returns. The slopes on the January dummy for corporate bonds increase monotonically from the Aaa to the LG portfolio. The extra January returns are 0.86%, 1.14%, and 1.56% per month for the A, Baa, and LG portfolios, and these extra average returns are at least 1.94 standard errors from 0.

If our five-factor time-series regressions are to explain the January seasonals in stock and bond returns, there must be January seasonals in the risk factors. Table 10 shows that, except for $TERM$, the risk factors have extra January returns in excess of 1% per month and at least 1.67 standard errors from 0. The seasonals in the size and book-to-market factors are especially strong. The average SMB and HML returns in January are 2.74% and 2.29% per month greater than in other months, and the extra January returns are 4.96 and 4.70 standard errors from 0. Indeed, like the excess returns on the 25 stock portfolios and the five corporate bond portfolios that are the dependent variables in the five-factor regressions, the extra January returns on the risk factors are generally much larger and more reliably different from 0 than the average returns for non-January months.

Finally, table 10 shows that the January seasonals in our risk factors largely absorb the seasonals in stock and bond returns. In the regressions of the five-factor residuals on the January dummy, only the stock portfolios in the smallest-size quintile produce systematically positive slopes; even these slopes are only one-quarter to one-tenth the positive January seasonals in the raw excess returns on the portfolios. If anything, the five-factor residuals for the remaining size quintiles show negative January seasonals, but the slopes on the January dummy for these stock portfolios, and for the bond portfolios, are small and mostly within two standard errors of 0. In short, whether spurious or real, the January seasonals in the returns on stocks and corporate bonds seem to be largely explained by the corresponding seasonals in the risk factors of our five-factor model.

6.3. Split-sample tests

In the time-series regressions for stocks, the dependent returns and the two explanatory returns SMB and HML are portfolios formed on size and book-to-market equity. Many readers worry that the apparent explanatory power of SMB and HML is spurious, induced by the regression setup. We think this is unlikely, given that the dependent returns are based on much finer size and BE/ME sorts (25 portfolios) than the SMB and HML returns. It also seems unlikely that we have stumbled on two mimicking returns for size and BE/ME factors that (a) measure strong common variation in the returns on 25 portfolios when really there is none, and (b) produce exactly the patterns in the regression slopes on SMB and HML needed to explain the size and book-to-market effects in the average returns on the 25 portfolios. Still, an independent test is of interest.

We split the stocks in each of the 25 size-BE/ME portfolios into two equal groups. One group is used to form the 25 dependent value-weighted portfolio returns for the time-series regressions. The other is used to form half-sample versions of the explanatory returns, $RM-RF$, SMB, and HML. The roles of the two groups are then reversed, and another set of regressions is run. In this way we have two sets of regressions. In each set, the explanatory and dependent returns are from disjoint groups of stocks.

Without showing all the details, we can report that the results for the two sets of regressions of excess returns for 25 size-BE/ME portfolios on disjoint versions of $RM-RF$, SMB, and HML are similar to the full-sample results in tables 6 and 9. The slopes on $RM-RF$, SMB, and HML in the split-sample regressions are close to those in table 6, and the intercepts, like those for the full-sample three-factor regressions in table 9, are close to 0. In short, the split-sample regressions confirm that there are common risk factors in returns related to size and book-to-market equity. They also confirm that market, size, and book-to-market factors seem to capture the cross-section of average stock returns.

If anything, the split-sample regressions show less power to reject the hypothesis that $RM-RF$, SMB, and HML capture the cross-section of average stock returns than the full-sample regressions. Since the 25 dependent portfolio returns in the split-sample regressions use half the available stocks, the portfolios are less diversified than those in table 6. Although the three-factor split-sample regressions produce high values of R^2 (mostly greater than 0.88), they are a bit lower than those in table 6 (mostly greater than 0.9). As a result, the F-tests of the zero-intercepts hypothesis are weaker for the split-sample regressions than for the full-sample regressions.

6.4. Portfolios formed on E/P

The most interesting check on our inferences about the role of size and book-to-market risk factors in returns is to examine whether these variables explain the returns on portfolios formed on other variables known to be informative about average returns. Table 11 shows summary statistics, as well as one-factor ($RM-RF$) and three-factor ($RM-RF$, SMB, and HML) regressions for portfolios formed on earnings/price (E/P) and dividend/price (D/P) ratios.

The average returns on the E/P portfolios have the U-shape documented in Jaffe, Keim, and Westerfield (1989) and Fama and French (1992a). The portfolio of firms with negative earnings and the portfolio of firms in the highest-E/P quintile have the highest average returns. For the positive-E/P portfolios, average return increases from the lowest- to the highest-E/P quintile. This pattern is an interesting challenge for our risk factors.

Table 11

Summary statistics for value-weighted monthly excess returns (in percent) on portfolios formed on dividend/price (D/P) and earnings/price (E/P), and regressions of excess portfolio returns on (i) the excess market return ($RM-RF$) and (ii) the excess market return ($RM-RF$) and the mimicking returns for the size (SMB) and book-to-market equity (HML) factors: July 1963 to December 1991, 342 months.[a]

$$\text{(i)}\quad R(t) - RF(t) = a + b[RM(t) - RF(t)] + e(t)$$

$$\text{(ii)}\quad R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + e(t)$$

	Portfolios formed on E/P			Portfolios formed on D/P		
Portfolio	Mean	Std.	$t(mn)$	Mean	Std.	$t(mn)$
$\leqslant 0$	0.72	7.77	1.72	0.48	7.36	1.20
Low	0.27	5.23	0.96	0.39	5.48	1.30
2	0.47	4.76	1.82	0.44	4.83	1.68
3	0.46	4.68	1.83	0.47	4.65	1.87
4	0.55	4.48	2.27	0.57	4.32	2.42
High	0.86	4.84	3.30	0.56	3.86	2.67

Portfolios formed on E/P

	Regression (i)			Regression (ii)				
Portfolio	a	b	R^2	a	b	s	h	R^2
$E/P \leqslant 0$	0.13	1.37	0.64	−0.30	1.24	1.13	0.46	0.82
	(0.50)	(24.70)		(−1.68)	(27.82)	(17.42)	(6.10)	
Low	−0.20	1.10	0.91	0.04	0.99	−0.01	−0.50	0.96
	(−2.35)	(57.42)		(0.70)	(66.78)	(−0.55)	(−19.73)	
2	0.03	1.01	0.94	0.03	1.01	0.02	−0.00	0.94
	(0.46)	(70.24)		(0.40)	(61.17)	(1.01)	(−0.08)	
3	0.04	0.99	0.92	−0.00	1.00	0.01	0.09	0.92
	(0.50)	(61.62)		(−0.12)	(55.46)	(0.40)	(2.86)	
4	0.15	0.93	0.88	−0.02	0.98	0.05	0.33	0.91
	(1.76)	(49.78)		(−0.28)	(53.57)	(1.95)	(10.44)	
High	0.46	0.94	0.78	0.08	1.03	0.24	0.67	0.91
	(3.69)	(34.73)		(1.01)	(51.56)	(8.34)	(19.62)	

Table 11 (continued)

	Portfolios formed on D/P							
	Regression (i)			Regression (ii)				
Portfolio	a	b	R^2	a	b	s	h	R^2
$D/P = 0$	− 0.15	1.45	0.80	− 0.23	1.20	0.99	− 0.21	0.94
	(− 0.86)	(37.18)		(− 2.30)	(49.45)	(28.09)	(− 5.17)	
Low	− 0.11	1.15	0.91	0.11	1.03	0.09	− 0.48	0.95
	(− 1.29)	(59.15)		(1.64)	(65.09)	(3.92)	(− 17.92)	
2	− 0.01	1.04	0.96	0.06	1.01	− 0.01	− 0.14	0.96
	(− 0.19)	(85.34)		(1.17)	(77.07)	(− 0.66)	(− 6.49)	
3	0.04	0.99	0.93	− 0.03	1.02	0.02	0.14	0.94
	(0.64)	(69.14)		(− 0.44)	(64.43)	(0.72)	(5.09)	
4	0.17	0.91	0.91	0.04	0.98	− 0.06	0.30	0.94
	(2.45)	(58.42)		(0.59)	(66.51)	(− 2.80)	(12.00)	
High	0.24	0.72	0.73	− 0.01	0.85	− 0.05	0.54	0.84
	(2.22)	(30.16)		(0.16)	(40.08)	(− 1.77)	(15.04)	

[a]Portfolios are formed in June of year t, 1963–1991. The dividend yield (D/P) for year t is the dividends paid from July of $t-1$ to June of t [measured using the procedure described in Fama and French (1988)], divided by market equity in June of $t-1$. The earnings/price ratio (E/P) for year t is the equity income for the fiscal year ending in calendar year $t-1$, divided by market equity in December of $t-1$. Equity income is income before extraordinary items, plus income-statement deferred taxes, minus preferred dividends. The quintile breakpoints for D/P or E/P are determined using only NYSE firms with positive dividends or earnings. Regression t-statistics are in parentheses. See table 7 for definitions of $RM-RF$, SMB, and HML.

Table 11 confirms the evidence in Basu (1983) that the one-factor Sharpe–Lintner model leaves the relation between average return and E/P largely unexplained. For the positive-E/P portfolios, the intercepts in the one-factor regressions increase monotonically, from − 0.20% per month ($t = -2.35$) for the lowest-E/P quintile to 0.46% ($t = 3.69$) for the highest. The failure of the one-factor model has a simple explanation. The market βs for the positive-E/P portfolios are all close to 1.0, so the one-factor model cannot explain the positive relation between E/P and average return.

In contrast, the three-factor model that uses $RM-RF$, SMB, and HML to explain returns leaves no residual E/P effect in average returns. The three-factor intercepts for the five positive-E/P portfolios are within 0.1 of 0 (t's from − 0.12 to 1.01). Interestingly, the three-factor regressions say that the increasing pattern in the average returns on the positive-E/P portfolios is due to their loadings on the book-to-market factor HML. The lowest positive-E/P quintile has an HML slope, − 0.50, like those produced by portfolios in the lowest-BE/ME quintile in the three-factor regressions in table 6. The highest-E/P quintile has an HML slope, 0.67, like those for portfolios in the highest-BE/ME quintile in table 6. Table 1 confirms that there is also a positive relation between E/P and BE/ME for our 25 portfolios formed on size and BE/ME.

Fama and French (1992b) find that low BE/ME is characteristic of growth stocks, that is, stocks with persistently high earnings on book equity that result in high stock prices relative to book equity. High BE/ME, on the other hand, is associated with distress, that is, persistently low earnings on book equity that result in low stock prices. The loadings on HML in the three-factor regressions of table 11 then say that low-E/P stocks have the low average returns typical of (low-BE/ME) growth stocks, while high-E/P stocks have the high average returns associated with distress (high-BE/ME).

The negative-E/P portfolio produces the only hint of evidence against the three-factor model. In spite of the portfolio's high average excess return (0.72% per month), the three-factor model says that its average return is 0.3% per month too low, given its strong loadings on SMB (1.13, like the smallest-size portfolios in table 6) and HML (0.46, like the higher-BE/ME portfolios in table 6). In other words, according to the three-factor model, the average return on this portfolio should be higher because its return behaves like those of small, relatively depressed, stocks. The three-factor intercept for the negative-E/P portfolio is, however, only 1.68 standard errors from 0.

In short, E/P portfolios produce a strong spread in average returns, which seems to be absorbed by the three common risk factors in stock returns. The E/P portfolios are thus interesting corroboration of our inferences that (a) there are common risk factors in stock returns related to size and book-to-market equity, and (b) $RM-RF$, SMB, and HML, the mimicking returns for market, size, and BE/ME risk factors, capture the cross-section of average stock returns.

6.5. Portfolios formed on D/P

Table 11 shows that, as in Keim (1983), average returns on portfolios formed on D/P are also U-shaped; they drop from the zero-dividend portfolio to the lowest positive-D/P portfolio, and then increase across the positive-D/P portfolios. The U-shaped pattern, and the overall spread in average returns, are, however, much weaker for the D/P portfolios than for the E/P portfolios.

Table 11 also confirms Keim's (1983) finding that the one-factor Sharpe–Lintner model leaves a pattern in average returns that looks like a tax penalty on dividends. The one-factor intercepts increase monotonically from the lowest- to the highest-D/P portfolios. This suggests that pre-tax returns on higher-D/P stocks must be higher to equalize after-tax risk-adjusted returns.

But the apparent tax effect in average returns does not survive in the three-factor regressions that use $RM-RF$, SMB, and HML to explain returns. The three-factor intercepts for the five positive-D/P portfolios are close to 0 and show no relation to D/P. The three-factor regressions say that the increasing pattern in the average returns on the positive-D/P portfolios is due to the increasing pattern in their loadings on the book-to-market factor HML. The lowest-(positive)-D/P quintile has a strong negative HML slope, -0.48, and the highest-D/P portfolio has a strong positive slope, 0.54. Again, the three-factor model says that low-D/P stocks have the low average returns typical of growth stocks, whereas high-D/P stocks have the high average returns associated with relative distress. Table 1 confirms that there is also a positive relation between D/P and BE/ME for our 25 portfolios formed on size and BE/ME.

The zero-dividend portfolio produces the strongest evidence against the three-factor model. The three-factor model says that the high average excess return on this portfolio (0.48% per month) is 0.23% too low ($t = -2.30$), given its strong loading (0.99) on SMB, the mimicking return for the size factor. In other words, because the return on the zero-dividend portfolio varies like the return on a portfolio of small stocks, the three-factor model says that the high return on this portfolio is not high enough. But the three-factor intercept for the zero-dividend portfolio is small in practical terms. Moreover, the three-factor model produces intercepts for the five positive-D/P portfolios that are all close to 0, both statistically and practically. We conclude that, overall, the D/P portfolios are consistent with our inference that the three stock-market factors, $RM-RF$, SMB, and HML, capture the cross-section of average stock returns.

7. Interpretation and applications

This paper studies the common risk factors in stock and bond returns and tests whether these shared risks capture the cross-section of average returns. There are at least five common factors in returns. Three stock-market factors produce common variation in stock returns. Except for low-grade corporate bonds, the stock-market factors have little role in returns on government and corporate bonds. The stock and bond markets are linked, however, through two shared term-structure factors.

7.1. Interpretation

Table 2 shows that the three stock-market factors, RMO, SMB, and HML, are largely uncorrelated with one another and with the two term-structure factors, $TERM$ and DEF. The regressions in table 8 that use RMO, SMB, HML, $TERM$, and DEF to explain stock and bond returns thus provide a good summary of the separate roles of the five factors in the volatility of returns and in the cross-section of average returns.

The 25 stock portfolios produce slopes on the orthogonalized market return, RMO, that are all around 1. Thus RMO, which has a standard deviation of 3.55% per month, accounts for similar common variation in the returns on all the stock portfolios. The average RMO return, 0.50% per month ($t = 2.61$), is also a common part of the average excess returns on stocks. Since the RMO slopes for stocks are all around 1, we can interpret the average RMO return as the premium for being a stock (rather than a one-month bill) and sharing general stock-market risk.

For stocks, the slopes on the two term-structure returns in table 8 are all around 0.8. The standard deviations of *TERM* and *DEF*, 3.02% and 1.60% per month (table 2), then say that *TERM* accounts for similar variation in the returns on all the stock portfolios, on the order of that captured by *RMO*, while *DEF* captures less common variation in returns. The average *TERM* and *DEF* returns are only 0.06% and 0.02% per month, so they explain almost none of the average excess returns on stocks. But the expected *TERM* and *DEF* returns vary through time with business conditions [Fama and French (1989) and Chen (1991)]. Thus *TERM* and *DEF* produce interesting time-series variation in expected bond and stock returns.

Except for low-grade corporate bonds, *TERM* and *DEF* capture almost all the common variation in bond returns identified in the five-factor regressions of table 8. Thus the low average excess returns on bonds fit nicely with the low average *TERM* and *DEF* returns. R^2 values near 1 in tables 3 and 8 say that *TERM* and *DEF* explain almost all the variation in high-grade (Aaa, Aa, A) corporate returns. Since the *TERM* and *DEF* slopes for corporate bonds (around 1) are similar to the slopes for stocks (around 0.8), we can infer that stocks share almost all the variation in high-grade corporate bond returns. Stocks, however, have substantial additional common volatility due to stock-market factors.

In the five-factor regressions of table 8, the slopes on *RMO*, *TERM*, and *DEF* do not vary much across the 25 stock portfolios. As a result, the roles of *RMO*, *TERM*, and *DEF* in stock returns are captured well by the excess market return, *RM–RF*, in table 7. The slopes on *RM–RF* in table 7 are, however, the same as the slopes on *RMO* in table 8. Thus, like *RMO*, *TERM*, and *DEF*, the excess market return does not explain the strong cross-sectional differences in average stock returns and their volatilities (table 2). That job is left to *SMB* and *HML*, the mimicking returns for the risk factors related to size and book-to-market equity.

The slopes on *SMB* in table 8 exceed 1.5 for portfolios in the smallest-size quintile, and they drop to around 0.3 for portfolios in the biggest-size quintile. The standard deviation of *SMB* is large, 2.89% per month. The common size-related factor in returns is thus important in explaining why small-stock returns are much more variable than big-stock returns (table 2). The average *SMB* return is only 0.27% per month ($t = 1.73$). The *SMB* slopes in table 8 range from 1.92 to 0.20, however, so the predicted spread in average returns across the 25 stock portfolios due to the size-related risk factor is large, 0.46% per month.

The slopes on *HML* in table 8 range from about -1 for portfolios in the lowest-book-to-market quintile to values near 0 in the highest-*BE/ME* quintile. *HML* thus tends to increase the volatility of low-*BE/ME* stock returns. Table 2 confirms that, within the size quintiles, the returns on the lowest-*BE/ME* portfolios are more volatile than the highest-*BE/ME* returns, especially for the three smallest-size quintiles, where the five-factor regressions produce R^2 values near 1. The average *HML* return, 0.40% per month ($t = 2.91$, table 2), then says that portfolios in the lowest-*BE/ME* quintile, with *HML* slopes close to -1, have their average returns reduced by about 0.40% per month relative to portfolios in the highest-*BE/ME* quintile, which have *HML* slopes close to 0.

Fama and French (1992b) find that book-to-market equity is related to relative profitability. On average, low-*BE/ME* firms have persistently high earnings and high-*BE/ME* firms have persistently poor earnings. The evidence here then suggests that *HML*, the difference between the returns on high- and low-*BE/ME* stocks, captures variation through time in a risk factor that is related to relative earnings performance. *HML* lowers the average returns on low-*BE/ME* stocks because their negative slopes on *HML* indicate that they hedge against the common factor in returns related to relative profitability.

A caveat is in order, however, about detailed stories for the slopes and average premiums in the time-series regressions. Many transformations of the five explanatory returns yield the same intercepts and R^2 values. Thus they yield the same inferences about the total common variation in returns and the ability of five factors to capture the cross-section of average returns. But different transformations change the slopes and average premiums for the factors. For example, the average value of *RMO*, the orthogonalized market return, is 0.50% per month ($t = 2.61$) versus 0.43% ($t = 1.76$) for *RM–RF*. Using *RMO* rather than *RM–RF* in the five-factor regressions also changes the slopes on *SMB*, *HML*, *TERM*, and *DEF* (compare tables 7 and 8). But *RMO* and *RM–RF* produce the same intercepts and R^2 values for testing a five-factor asset-pricing model.

At a minimum, our results show that five factors do a good job explaining (a) common variation in bond and stock returns and (b) the cross-section of average returns. We think there is appeal in the simple way we define mimicking returns for the stock-market and bond-market factors. But the choice of factors, especially the size and book-to-market factors, is motivated by empirical experience. Without a theory that specifies the exact form of the state variables or common factors in returns, the choice of any particular version of the factors is somewhat arbitrary. Thus detailed stories for the slopes and average premiums associated with particular versions of the factors are suggestive, but never definitive.

7.2. *Applications*

In principle, our results can be used in any application that requires estimates of expected stock returns. The list includes (a) selecting portfolios, (b) evaluating portfolio performance, (c) measuring abnormal returns in event studies, and (d) estimating the cost of capital. The applications depend on the evidence that the five factors provide a good description of the cross-section of average returns, but they do not require that we have identified the true factors.

If the five factors capture the cross-section of average returns, they can be used to guide portfolio selection. The exposures of a candidate portfolio to the five risk factors can be estimated with a regression of the portfolio's past excess returns on the five explanatory returns. The regression slopes and the historical average premiums for the factors can then be used to estimate the (unconditional) expected return on the portfolio. A similar procedure can be used to estimate the expected return on a firm's securities, for the purpose of judging its cost of capital. (We predict, however, that sampling error will be a serious problem in the five-factor parameter estimates for individual securities.)

If our results are taken at face value, evaluating the performance of a managed portfolio is straightforward. The intercept in the time-series regression of the managed portfolio's excess return on our five explanatory returns is the average abnormal return needed to judge whether a manager can beat the market, that is, whether he can use special information to generate average returns greater than those on passive combinations of the mimicking returns for the five risk factors.

Using our results for portfolio formation and performance evaluation is even simpler for portfolios that hold only stocks. Tables 5 to 8 say that a model that uses only the three stock-market factors, $RM-RF$, SMB, and HML, does as well as the five-factor model in explaining the common time-series variation in stock returns and the cross-section of average stock returns.

Many continue to use the one-factor Sharpe–Lintner model to evaluate portfolio performance and to estimate the cost of capital, despite the lack of evidence that it is relevant. At a minimum, the results here and in Fama and French (1992a) should help to break this common habit.

Finally, in event studies of the stock-price response to firm-specific information, the residuals from a one-factor regression of the stock's return on a market return are often used to abstract from common variation in returns. Our results suggest that the residuals from three-factor regressions that also use SMB and HML will do a better job isolating the firm-specific components of returns.

Using a three-factor alternative is especially important if the tests impose a cross-section constraint on average stock returns. For example, Agrawal, Jaffe, and Mandelker (1991) use the residuals from the Sharpe–Lintner model to judge the post-merger stock returns of acquiring firms. Aware that post-merger returns may seem too low because acquiring firms tend to be large, they control for size as well as the excess market return when measuring abnormal returns. Still, they find that the average abnormal returns of acquiring firms are negative and similar in size in each of the five years after mergers.

We conjecture that the persistent negative abnormal returns of acquiring firms are a book-to-market effect. We guess that acquiring firms tend to be successful firms that have high stock prices relative to book value and low loadings on HML. In our three-factor model, low loadings on HML would reduce the average stock returns of acquiring firms, and produce persistent negative abnormal returns in tests that adjust only for market and size factors.

7.3. Open questions

Taken together, the results here and in Fama and French (1992b) suggest that there is an economic story behind the size and book-to-market effects in average stock returns. The tests here show that there are common return factors related to size and book-to-market equity that help capture the cross-section of average stock returns in a way that is consistent with multifactor asset-pricing models. Fama and French (1992b) show that size and BE/ME are related to systematic patterns in relative profitability and growth that could well be the source of common risk factors in returns.

But our work leaves many open questions. Most glaring, we have not shown how the size and book-to-market factors in returns are driven by the stochastic behavior of earnings. How does profitability, or any other fundamental, produce common variation in returns associated with size and BE/ME that is not picked up by the market return? Can specific fundamentals be identified as state variables that lead to common variation in returns that is independent of the market and carries a different premium than general market risk? These and other interesting questions are left to future work.

References

Agrawal, Anup, Jeffrey F. Jaffe, and Gershon N. Mandelker, 1991, The post-merger performance of acquired firms: A re-examination of an anomaly, Journal of Finance, forthcoming.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, Journal of Financial Economics 9, 3–18.

Banz, Rolf W. and William J. Breen, 1986, Sample dependent results using accounting and market data: Some evidence, Journal of Finance 41, 779–793.

Basu, Sanjoy, 1983, The relationship between earnings yield, market value, and return for NYSE common stocks: Further evidence, Journal of Financial Economics 12, 129–156.

Bhandari, Laxmi Chand, 1988, Debt/equity ratio and expected common stock returns: Empirical evidence, Journal of Finance 43, 507–528.

Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in: M. Jensen, ed., Studies in the theory of capital markets (Praeger, New York, NY).

Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, Journal of Financial Economics 7, 265–296.

Breeden, Douglas T., Michael R. Gibbons, and Robert H. Litzenberger, 1989, Empirical tests of the consumption-oriented CAPM, Journal of Finance 44, 231–262.

Chan, K. C., Nai-fu Chen, and David Hsieh, 1985, An exploratory investigation of the firm size effect, Journal of Financial Economics 14, 451–471.

Chen, Nai-fu, 1991, Financial investment opportunities and the macroeconomy, Journal of Finance 46, 529–554.

Chen, Nai-fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, Journal of Business 59, 383–403.

Fama, Eugene F., 1991, Efficient markets: II, Journal of Finance 46, 1575–1617.
Fama, Eugene F. and Kenneth R. French, 1988, Dividend yields and expected stock returns, Journal of Financial Economics 22, 3–25.
Fama, Eugene F. and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, Journal of Financial Economics 25, 23–49.
Fama, Eugene F. and Kenneth R. French, 1992a, The cross-section of expected stock returns, Journal of Finance 47, 427–465.
Fama, Eugene F. and Kenneth R. French, 1992b, The economic fundamentals of size and book-to-market equity, Working paper (Graduate School of Business, University of Chicago, Chicago, IL).
Fama, Eugene F. and James MacBeth, 1973, Risk, return and equilibrium: Empirical tests, Journal of Political Economy 81, 607–636.
Gibbons, Michael R., Stephen A. Ross and Jay Shanken, 1989, A test of the efficiency of a given portfolio, Econometrica 57, 1121–1152.
Jaffe, Jeffrey, Donald B. Keim, and Randolph Westerfield, 1989, Earnings yields, market values and stock returns, Journal of Finance 44, 135–148.
Keim, Donald B., 1983, Size-related anomalies and stock return seasonality, Journal of Financial Economics 12, 13–32.
Keim, Donald B., 1988, Stock market regularities: A synthesis of the evidence and explanations, in: E. Dimson, ed., Stock market anomalies (Cambridge University Press, Cambridge).
Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, Review of Economics and Statistics 47, 13–37.
Lo, Andrew W. and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, Review of Financial Studies 3, 431–467.
Merton, Robert C., 1973, An intertemporal capital asset pricing model, Econometrica 41, 867–887.
Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, Journal of Financial Economics 8, 323–361.
Reinganum, Marc R., 1981, A new empirical perspective on the CAPM, Journal of Financial and Quantitative Analysis 16, 439–462.
Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, Journal of Portfolio Management 11, 9–17.
Roll, Richard, 1983, Vas ist das? The turn-of-the-year effect and the return premia of small firms, Journal of Portfolio Management 9, 18–28.
Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, Journal of Economic Theory 13, 341–360.
Shanken, Jay, 1982, The arbitrage pricing theory: Is it testable?, Journal of Finance 37, 1129–1140.
Shanken, Jay and Mark I. Weistein, 1990, Macroeconomic variables and asset pricing: Estimation and tests, Working paper (Simon School of Business Administration, University of Rochester, Rochester, NY).
Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance 19, 425–442.

Journal of Economic Perspectives—Volume 18, Number 3—Summer 2004—Pages 25–46

The Capital Asset Pricing Model: Theory and Evidence

Eugene F. Fama and Kenneth R. French

The capital asset pricing model (CAPM) of William Sharpe (1964) and John Lintner (1965) marks the birth of asset pricing theory (resulting in a Nobel Prize for Sharpe in 1990). Four decades later, the CAPM is still widely used in applications, such as estimating the cost of capital for firms and evaluating the performance of managed portfolios. It is the centerpiece of MBA investment courses. Indeed, it is often the only asset pricing model taught in these courses.[1]

The attraction of the CAPM is that it offers powerful and intuitively pleasing predictions about how to measure risk and the relation between expected return and risk. Unfortunately, the empirical record of the model is poor—poor enough to invalidate the way it is used in applications. The CAPM's empirical problems may reflect theoretical failings, the result of many simplifying assumptions. But they may also be caused by difficulties in implementing valid tests of the model. For example, the CAPM says that the risk of a stock should be measured relative to a comprehensive "market portfolio" that in principle can include not just traded financial assets, but also consumer durables, real estate and human capital. Even if we take a narrow view of the model and limit its purview to traded financial assets, is it

[1] Although every asset pricing model is a capital asset pricing model, the finance profession reserves the acronym CAPM for the specific model of Sharpe (1964), Lintner (1965) and Black (1972) discussed here. Thus, throughout the paper we refer to the Sharpe-Lintner-Black model as the CAPM.

■ *Eugene F. Fama is Robert R. McCormick Distinguished Service Professor of Finance, Graduate School of Business, University of Chicago, Chicago, Illinois. Kenneth R. French is Carl E. and Catherine M. Heidt Professor of Finance, Tuck School of Business, Dartmouth College, Hanover, New Hampshire. Their e-mail addresses are ⟨eugene.fama@gsb.uchicago.edu⟩ and ⟨kfrench@dartmouth.edu⟩, respectively.*

legitimate to limit further the market portfolio to U.S. common stocks (a typical choice), or should the market be expanded to include bonds, and other financial assets, perhaps around the world? In the end, we argue that whether the model's problems reflect weaknesses in the theory or in its empirical implementation, the failure of the CAPM in empirical tests implies that most applications of the model are invalid.

We begin by outlining the logic of the CAPM, focusing on its predictions about risk and expected return. We then review the history of empirical work and what it says about shortcomings of the CAPM that pose challenges to be explained by alternative models.

The Logic of the CAPM

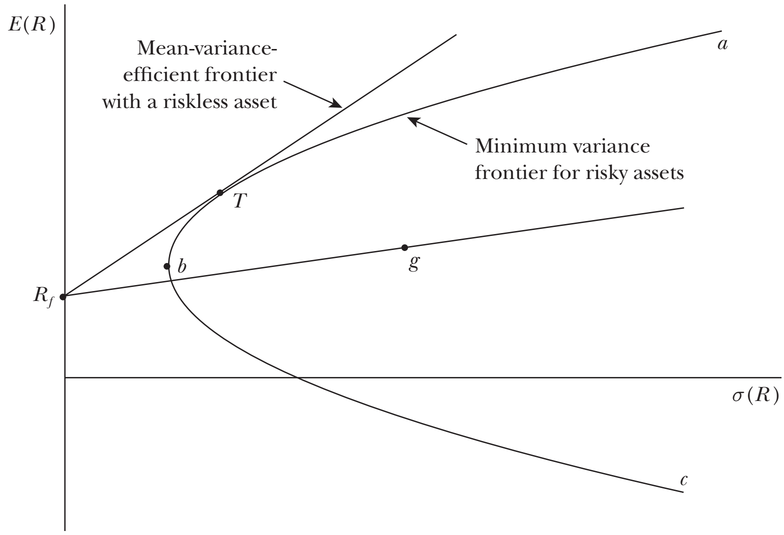
The CAPM builds on the model of portfolio choice developed by Harry Markowitz (1959). In Markowitz's model, an investor selects a portfolio at time $t - 1$ that produces a stochastic return at t. The model assumes investors are risk averse and, when choosing among portfolios, they care only about the mean and variance of their one-period investment return. As a result, investors choose "mean-variance-efficient" portfolios, in the sense that the portfolios 1) minimize the variance of portfolio return, given expected return, and 2) maximize expected return, given variance. Thus, the Markowitz approach is often called a "mean-variance model."

The portfolio model provides an algebraic condition on asset weights in mean-variance-efficient portfolios. The CAPM turns this algebraic statement into a testable prediction about the relation between risk and expected return by identifying a portfolio that must be efficient if asset prices are to clear the market of all assets.

Sharpe (1964) and Lintner (1965) add two key assumptions to the Markowitz model to identify a portfolio that must be mean-variance-efficient. The first assumption is *complete agreement*: given market clearing asset prices at $t - 1$, investors agree on the joint distribution of asset returns from $t - 1$ to t. And this distribution is the true one—that is, it is the distribution from which the returns we use to test the model are drawn. The second assumption is that there is *borrowing and lending at a risk-free rate*, which is the same for all investors and does not depend on the amount borrowed or lent.

Figure 1 describes portfolio opportunities and tells the CAPM story. The horizontal axis shows portfolio risk, measured by the standard deviation of portfolio return; the vertical axis shows expected return. The curve *abc*, which is called the minimum variance frontier, traces combinations of expected return and risk for portfolios of risky assets that minimize return variance at different levels of expected return. (These portfolios do not include risk-free borrowing and lending.) The tradeoff between risk and expected return for minimum variance portfolios is apparent. For example, an investor who wants a high expected return, perhaps at point *a*, must accept high volatility. At point *T*, the investor can have an interme-

Figure 1
Investment Opportunities



diate expected return with lower volatility. If there is no risk-free borrowing or lending, only portfolios above b along abc are mean-variance-efficient, since these portfolios also maximize expected return, given their return variances.

Adding risk-free borrowing and lending turns the efficient set into a straight line. Consider a portfolio that invests the proportion x of portfolio funds in a risk-free security and $1 - x$ in some portfolio g. If all funds are invested in the risk-free security—that is, they are loaned at the risk-free rate of interest—the result is the point R_f in Figure 1, a portfolio with zero variance and a risk-free rate of return. Combinations of risk-free lending and positive investment in g plot on the straight line between R_f and g. Points to the right of g on the line represent borrowing at the risk-free rate, with the proceeds from the borrowing used to increase investment in portfolio g. In short, portfolios that combine risk-free lending or borrowing with some risky portfolio g plot along a straight line from R_f through g in Figure 1.[2]

[2] Formally, the return, expected return and standard deviation of return on portfolios of the risk-free asset f and a risky portfolio g vary with x, the proportion of portfolio funds invested in f, as

$$R_p = xR_f + (1 - x)R_g,$$

$$E(R_p) = xR_f + (1 - x)E(R_g),$$

$$\sigma(R_p) = (1 - x)\sigma(R_g),\ x \leq 1.0,$$

which together imply that the portfolios plot along the line from R_f through g in Figure 1.

To obtain the mean-variance-efficient portfolios available with risk-free borrowing and lending, one swings a line from R_f in Figure 1 up and to the left as far as possible, to the tangency portfolio T. We can then see that all efficient portfolios are combinations of the risk-free asset (either risk-free borrowing or lending) and a single risky tangency portfolio, T. This key result is Tobin's (1958) "separation theorem."

The punch line of the CAPM is now straightforward. With complete agreement about distributions of returns, all investors see the same opportunity set (Figure 1), and they combine the same risky tangency portfolio T with risk-free lending or borrowing. Since all investors hold the same portfolio T of risky assets, it must be the value-weight market portfolio of risky assets. Specifically, each risky asset's weight in the tangency portfolio, which we now call M (for the "market"), must be the total market value of all outstanding units of the asset divided by the total market value of all risky assets. In addition, the risk-free rate must be set (along with the prices of risky assets) to clear the market for risk-free borrowing and lending.

In short, the CAPM assumptions imply that the market portfolio M must be on the minimum variance frontier if the asset market is to clear. This means that the algebraic relation that holds for any minimum variance portfolio must hold for the market portfolio. Specifically, if there are N risky assets,

$$\text{(Minimum Variance Condition for } M) \quad E(R_i) = E(R_{ZM}) + [E(R_M) - E(R_{ZM})]\beta_{iM}, \; i = 1, \ldots, N.$$

In this equation, $E(R_i)$ is the expected return on asset i, and β_{iM}, the market beta of asset i, is the covariance of its return with the market return divided by the variance of the market return,

$$\text{(Market Beta)} \quad \beta_{iM} = \frac{\text{cov}(R_i, R_M)}{\sigma^2(R_M)}.$$

The first term on the right-hand side of the minimum variance condition, $E(R_{ZM})$, is the expected return on assets that have market betas equal to zero, which means their returns are uncorrelated with the market return. The second term is a risk premium—the market beta of asset i, β_{iM}, times the premium per unit of beta, which is the expected market return, $E(R_M)$, minus $E(R_{ZM})$.

Since the market beta of asset i is also the slope in the regression of its return on the market return, a common (and correct) interpretation of beta is that it measures the sensitivity of the asset's return to variation in the market return. But there is another interpretation of beta more in line with the spirit of the portfolio model that underlies the CAPM. The risk of the market portfolio, as measured by the variance of its return (the denominator of β_{iM}), is a weighted average of the covariance risks of the assets in M (the numerators of β_{iM} for different assets).

Thus, β_{iM} is the covariance risk of asset i in M measured relative to the average covariance risk of assets, which is just the variance of the market return.[3] In economic terms, β_{iM} is proportional to the risk each dollar invested in asset i contributes to the market portfolio.

The last step in the development of the Sharpe-Lintner model is to use the assumption of risk-free borrowing and lending to nail down $E(R_{ZM})$, the expected return on zero-beta assets. A risky asset's return is uncorrelated with the market return—its beta is zero—when the average of the asset's covariances with the returns on other assets just offsets the variance of the asset's return. Such a risky asset is riskless in the market portfolio in the sense that it contributes nothing to the variance of the market return.

When there is risk-free borrowing and lending, the expected return on assets that are uncorrelated with the market return, $E(R_{ZM})$, must equal the risk-free rate, R_f. The relation between expected return and beta then becomes the familiar Sharpe-Lintner CAPM equation,

$$\text{(Sharpe-Lintner CAPM)} \quad E(R_i) = R_f + [E(R_M) - R_f)]\beta_{iM},\ i = 1, \ldots, N.$$

In words, the expected return on any asset i is the risk-free interest rate, R_f, plus a risk premium, which is the asset's market beta, β_{iM}, times the premium per unit of beta risk, $E(R_M) - R_f$.

Unrestricted risk-free borrowing and lending is an unrealistic assumption. Fischer Black (1972) develops a version of the CAPM without risk-free borrowing or lending. He shows that the CAPM's key result—that the market portfolio is mean-variance-efficient—can be obtained by instead allowing unrestricted short sales of risky assets. In brief, back in Figure 1, if there is no risk-free asset, investors select portfolios from along the mean-variance-efficient frontier from a to b. Market clearing prices imply that when one weights the efficient portfolios chosen by investors by their (positive) shares of aggregate invested wealth, the resulting portfolio is the market portfolio. The market portfolio is thus a portfolio of the efficient portfolios chosen by investors. With unrestricted short selling of risky assets, portfolios made up of efficient portfolios are themselves efficient. Thus, the market portfolio is efficient, which means that the minimum variance condition for M given above holds, and it is the expected return-risk relation of the Black CAPM.

The relations between expected return and market beta of the Black and Sharpe-Lintner versions of the CAPM differ only in terms of what each says about $E(R_{ZM})$, the expected return on assets uncorrelated with the market. The Black version says only that $E(R_{ZM})$ must be less than the expected market return, so the

[3] Formally, if x_{iM} is the weight of asset i in the market portfolio, then the variance of the portfolio's return is

$$\sigma^2(R_M) = Cov(R_M, R_M) = Cov\left(\sum_{i=1}^{N} x_{iM}R_i,\ R_M\right) = \sum_{i=1}^{N} x_{iM}Cov(R_i, R_M).$$

premium for beta is positive. In contrast, in the Sharpe-Lintner version of the model, $E(R_{ZM})$ must be the risk-free interest rate, R_f, and the premium per unit of beta risk is $E(R_M) - R_f$.

The assumption that short selling is unrestricted is as unrealistic as unrestricted risk-free borrowing and lending. If there is no risk-free asset and short sales of risky assets are not allowed, mean-variance investors still choose efficient portfolios—points above *b* on the *abc* curve in Figure 1. But when there is no short selling of risky assets and no risk-free asset, the algebra of portfolio efficiency says that portfolios made up of efficient portfolios are not typically efficient. This means that the market portfolio, which is a portfolio of the efficient portfolios chosen by investors, is not typically efficient. And the CAPM relation between expected return and market beta is lost. This does not rule out predictions about expected return and betas with respect to other efficient portfolios—if theory can specify portfolios that must be efficient if the market is to clear. But so far this has proven impossible.

In short, the familiar CAPM equation relating expected asset returns to their market betas is just an application to the market portfolio of the relation between expected return and portfolio beta that holds in any mean-variance-efficient portfolio. The efficiency of the market portfolio is based on many unrealistic assumptions, including complete agreement and either unrestricted risk-free borrowing and lending or unrestricted short selling of risky assets. But all interesting models involve unrealistic simplifications, which is why they must be tested against data.

Early Empirical Tests

Tests of the CAPM are based on three implications of the relation between expected return and market beta implied by the model. First, expected returns on all assets are linearly related to their betas, and no other variable has marginal explanatory power. Second, the beta premium is positive, meaning that the expected return on the market portfolio exceeds the expected return on assets whose returns are uncorrelated with the market return. Third, in the Sharpe-Lintner version of the model, assets uncorrelated with the market have expected returns equal to the risk-free interest rate, and the beta premium is the expected market return minus the risk-free rate. Most tests of these predictions use either cross-section or time-series regressions. Both approaches date to early tests of the model.

Tests on Risk Premiums

The early cross-section regression tests focus on the Sharpe-Lintner model's predictions about the intercept and slope in the relation between expected return and market beta. The approach is to regress a cross-section of average asset returns on estimates of asset betas. The model predicts that the intercept in these regressions is the risk-free interest rate, R_f, and the coefficient on beta is the expected return on the market in excess of the risk-free rate, $E(R_M) - R_f$.

Two problems in these tests quickly became apparent. First, estimates of beta for individual assets are imprecise, creating a measurement error problem when they are used to explain average returns. Second, the regression residuals have common sources of variation, such as industry effects in average returns. Positive correlation in the residuals produces downward bias in the usual ordinary least squares estimates of the standard errors of the cross-section regression slopes.

To improve the precision of estimated betas, researchers such as Blume (1970), Friend and Blume (1970) and Black, Jensen and Scholes (1972) work with portfolios, rather than individual securities. Since expected returns and market betas combine in the same way in portfolios, if the CAPM explains security returns it also explains portfolio returns.[4] Estimates of beta for diversified portfolios are more precise than estimates for individual securities. Thus, using portfolios in cross-section regressions of average returns on betas reduces the critical errors in variables problem. Grouping, however, shrinks the range of betas and reduces statistical power. To mitigate this problem, researchers sort securities on beta when forming portfolios; the first portfolio contains securities with the lowest betas, and so on, up to the last portfolio with the highest beta assets. This sorting procedure is now standard in empirical tests.

Fama and MacBeth (1973) propose a method for addressing the inference problem caused by correlation of the residuals in cross-section regressions. Instead of estimating a single cross-section regression of average monthly returns on betas, they estimate month-by-month cross-section regressions of monthly returns on betas. The times-series means of the monthly slopes and intercepts, along with the standard errors of the means, are then used to test whether the average premium for beta is positive and whether the average return on assets uncorrelated with the market is equal to the average risk-free interest rate. In this approach, the standard errors of the average intercept and slope are determined by the month-to-month variation in the regression coefficients, which fully captures the effects of residual correlation on variation in the regression coefficients, but sidesteps the problem of actually estimating the correlations. The residual correlations are, in effect, captured via repeated sampling of the regression coefficients. This approach also becomes standard in the literature.

Jensen (1968) was the first to note that the Sharpe-Lintner version of the

[4] Formally, if x_{ip}, $i = 1, \ldots, N$, are the weights for assets in some portfolio p, the expected return and market beta for the portfolio are related to the expected returns and betas of assets as

$$E(R_p) = \sum_{i=1}^{N} x_{ip}E(R_i), \text{ and } \beta_{pM} = \sum_{i=1}^{N} x_{ip}\beta_{pM}.$$

Thus, the CAPM relation between expected return and beta,

$$E(R_i) = E(R_f) + [E(R_M) - E(R_f)]\beta_{iM},$$

holds when asset i is a portfolio, as well as when i is an individual security.

relation between expected return and market beta also implies a time-series regression test. The Sharpe-Lintner CAPM says that the expected value of an asset's excess return (the asset's return minus the risk-free interest rate, $R_{it} - R_{ft}$) is completely explained by its expected CAPM risk premium (its beta times the expected value of $R_{Mt} - R_{ft}$). This implies that "Jensen's alpha," the intercept term in the time-series regression,

$$\text{(Time-Series Regression)} \quad R_{it} - R_{ft} = \alpha_i + \beta_{iM}(R_{Mt} - R_{ft}) + \varepsilon_{it},$$

is zero for each asset.

The early tests firmly reject the Sharpe-Lintner version of the CAPM. There is a positive relation between beta and average return, but it is too "flat." Recall that, in cross-section regressions, the Sharpe-Lintner model predicts that the intercept is the risk-free rate and the coefficient on beta is the expected market return in excess of the risk-free rate, $E(R_M) - R_f$. The regressions consistently find that the intercept is greater than the average risk-free rate (typically proxied as the return on a one-month Treasury bill), and the coefficient on beta is less than the average excess market return (proxied as the average return on a portfolio of U.S. common stocks minus the Treasury bill rate). This is true in the early tests, such as Douglas (1968), Black, Jensen and Scholes (1972), Miller and Scholes (1972), Blume and Friend (1973) and Fama and MacBeth (1973), as well as in more recent cross-section regression tests, like Fama and French (1992).

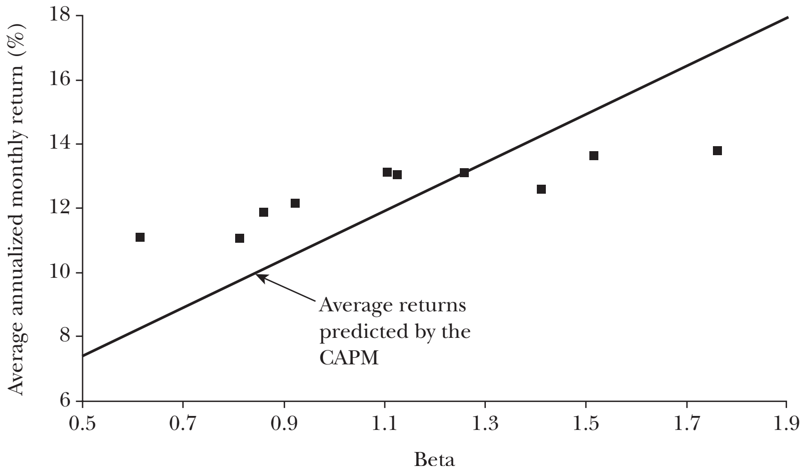
The evidence that the relation between beta and average return is too flat is confirmed in time-series tests, such as Friend and Blume (1970), Black, Jensen and Scholes (1972) and Stambaugh (1982). The intercepts in time-series regressions of excess asset returns on the excess market return are positive for assets with low betas and negative for assets with high betas.

Figure 2 provides an updated example of the evidence. In December of each year, we estimate a preranking beta for every NYSE (1928–2003), AMEX (1963–2003) and NASDAQ (1972–2003) stock in the CRSP (Center for Research in Security Prices of the University of Chicago) database, using two to five years (as available) of prior monthly returns.[5] We then form ten value-weight portfolios based on these preranking betas and compute their returns for the next twelve months. We repeat this process for each year from 1928 to 2003. The result is 912 monthly returns on ten beta-sorted portfolios. Figure 2 plots each portfolio's average return against its postranking beta, estimated by regressing its monthly returns for 1928–2003 on the return on the CRSP value-weight portfolio of U.S. common stocks.

The Sharpe-Lintner CAPM predicts that the portfolios plot along a straight

[5] To be included in the sample for year t, a security must have market equity data (price times shares outstanding) for December of $t - 1$, and CRSP must classify it as ordinary common equity. Thus, we exclude securities such as American Depository Receipts (ADRs) and Real Estate Investment Trusts (REITs).

Figure 2
Average Annualized Monthly Return versus Beta for Value Weight Portfolios Formed on Prior Beta, 1928–2003



line, with an intercept equal to the risk-free rate, R_f, and a slope equal to the expected excess return on the market, $E(R_M) - R_f$. We use the average one-month Treasury bill rate and the average excess CRSP market return for 1928–2003 to estimate the predicted line in Figure 2. Confirming earlier evidence, the relation between beta and average return for the ten portfolios is much flatter than the Sharpe-Lintner CAPM predicts. The returns on the low beta portfolios are too high, and the returns on the high beta portfolios are too low. For example, the predicted return on the portfolio with the lowest beta is 8.3 percent per year; the actual return is 11.1 percent. The predicted return on the portfolio with the highest beta is 16.8 percent per year; the actual is 13.7 percent.

Although the observed premium per unit of beta is lower than the Sharpe-Lintner model predicts, the relation between average return and beta in Figure 2 is roughly linear. This is consistent with the Black version of the CAPM, which predicts only that the beta premium is positive. Even this less restrictive model, however, eventually succumbs to the data.

Testing Whether Market Betas Explain Expected Returns

The Sharpe-Lintner and Black versions of the CAPM share the prediction that the market portfolio is mean-variance-efficient. This implies that differences in expected return across securities and portfolios are entirely explained by differences in market beta; other variables should add nothing to the explanation of expected return. This prediction plays a prominent role in tests of the CAPM. In the early work, the weapon of choice is cross-section regressions.

In the framework of Fama and MacBeth (1973), one simply adds predetermined explanatory variables to the month-by-month cross-section regressions of returns on beta. If all differences in expected return are explained by beta, the average slopes on the additional variables should not be reliably different from zero. Clearly, the trick in the cross-section regression approach is to choose specific additional variables likely to expose any problems of the CAPM prediction that, because the market portfolio is efficient, market betas suffice to explain expected asset returns.

For example, in Fama and MacBeth (1973) the additional variables are squared market betas (to test the prediction that the relation between expected return and beta is linear) and residual variances from regressions of returns on the market return (to test the prediction that market beta is the only measure of risk needed to explain expected returns). These variables do not add to the explanation of average returns provided by beta. Thus, the results of Fama and MacBeth (1973) are consistent with the hypothesis that their market proxy—an equal-weight portfolio of NYSE stocks—is on the minimum variance frontier.

The hypothesis that market betas completely explain expected returns can also be tested using time-series regressions. In the time-series regression described above (the excess return on asset i regressed on the excess market return), the intercept is the difference between the asset's average excess return and the excess return predicted by the Sharpe-Lintner model, that is, beta times the average excess market return. If the model holds, there is no way to group assets into portfolios whose intercepts are reliably different from zero. For example, the intercepts for a portfolio of stocks with high ratios of earnings to price and a portfolio of stocks with low earning-price ratios should both be zero. Thus, to test the hypothesis that market betas suffice to explain expected returns, one estimates the time-series regression for a set of assets (or portfolios) and then jointly tests the vector of regression intercepts against zero. The trick in this approach is to choose the left-hand-side assets (or portfolios) in a way likely to expose any shortcoming of the CAPM prediction that market betas suffice to explain expected asset returns.

In early applications, researchers use a variety of tests to determine whether the intercepts in a set of time-series regressions are all zero. The tests have the same asymptotic properties, but there is controversy about which has the best small sample properties. Gibbons, Ross and Shanken (1989) settle the debate by providing an F-test on the intercepts that has exact small-sample properties. They also show that the test has a simple economic interpretation. In effect, the test constructs a candidate for the tangency portfolio T in Figure 1 by optimally combining the market proxy and the left-hand-side assets of the time-series regressions. The estimator then tests whether the efficient set provided by the combination of this tangency portfolio and the risk-free asset is reliably superior to the one obtained by combining the risk-free asset with the market proxy alone. In other words, the Gibbons, Ross and Shanken statistic tests whether the market proxy is the tangency portfolio in the set of portfolios that can be constructed by combining the market portfolio with the specific assets used as dependent variables in the time-series regressions.

Enlightened by this insight of Gibbons, Ross and Shanken (1989), one can see a similar interpretation of the cross-section regression test of whether market betas suffice to explain expected returns. In this case, the test is whether the additional explanatory variables in a cross-section regression identify patterns in the returns on the left-hand-side assets that are not explained by the assets' market betas. This amounts to testing whether the market proxy is on the minimum variance frontier that can be constructed using the market proxy and the left-hand-side assets included in the tests.

An important lesson from this discussion is that time-series and cross-section regressions do not, strictly speaking, test the CAPM. What is literally tested is whether a specific proxy for the market portfolio (typically a portfolio of U.S. common stocks) is efficient in the set of portfolios that can be constructed from it and the left-hand-side assets used in the test. One might conclude from this that the CAPM has never been tested, and prospects for testing it are not good because 1) the set of left-hand-side assets does not include all marketable assets, and 2) data for the true market portfolio of all assets are likely beyond reach (Roll, 1977; more on this later). But this criticism can be leveled at tests of any economic model when the tests are less than exhaustive or when they use proxies for the variables called for by the model.

The bottom line from the early cross-section regression tests of the CAPM, such as Fama and MacBeth (1973), and the early time-series regression tests, like Gibbons (1982) and Stambaugh (1982), is that standard market proxies seem to be on the minimum variance frontier. That is, the central predictions of the Black version of the CAPM, that market betas suffice to explain expected returns and that the risk premium for beta is positive, seem to hold. But the more specific prediction of the Sharpe-Lintner CAPM that the premium per unit of beta is the expected market return minus the risk-free interest rate is consistently rejected.

The success of the Black version of the CAPM in early tests produced a consensus that the model is a good description of expected returns. These early results, coupled with the model's simplicity and intuitive appeal, pushed the CAPM to the forefront of finance.

Recent Tests

Starting in the late 1970s, empirical work appears that challenges even the Black version of the CAPM. Specifically, evidence mounts that much of the variation in expected return is unrelated to market beta.

The first blow is Basu's (1977) evidence that when common stocks are sorted on earnings-price ratios, future returns on high E/P stocks are higher than predicted by the CAPM. Banz (1981) documents a size effect: when stocks are sorted on market capitalization (price times shares outstanding), average returns on small stocks are higher than predicted by the CAPM. Bhandari (1988) finds that high debt-equity ratios (book value of debt over the market value of equity, a measure of leverage) are associated with returns that are too high relative to their market betas.

Finally, Statman (1980) and Rosenberg, Reid and Lanstein (1985) document that stocks with high book-to-market equity ratios (B/M, the ratio of the book value of a common stock to its market value) have high average returns that are not captured by their betas.

There is a theme in the contradictions of the CAPM summarized above. Ratios involving stock prices have information about expected returns missed by market betas. On reflection, this is not surprising. A stock's price depends not only on the expected cash flows it will provide, but also on the expected returns that discount expected cash flows back to the present. Thus, in principle, the cross-section of prices has information about the cross-section of expected returns. (A high expected return implies a high discount rate and a low price.) The cross-section of stock prices is, however, arbitrarily affected by differences in scale (or units). But with a judicious choice of scaling variable X, the ratio X/P can reveal differences in the cross-section of expected stock returns. Such ratios are thus prime candidates to expose shortcomings of asset pricing models—in the case of the CAPM, shortcomings of the prediction that market betas suffice to explain expected returns (Ball, 1978). The contradictions of the CAPM summarized above suggest that earnings-price, debt-equity and book-to-market ratios indeed play this role.

Fama and French (1992) update and synthesize the evidence on the empirical failures of the CAPM. Using the cross-section regression approach, they confirm that size, earnings-price, debt-equity and book-to-market ratios add to the explanation of expected stock returns provided by market beta. Fama and French (1996) reach the same conclusion using the time-series regression approach applied to portfolios of stocks sorted on price ratios. They also find that different price ratios have much the same information about expected returns. This is not surprising given that price is the common driving force in the price ratios, and the numerators are just scaling variables used to extract the information in price about expected returns.

Fama and French (1992) also confirm the evidence (Reinganum, 1981; Stambaugh, 1982; Lakonishok and Shapiro, 1986) that the relation between average return and beta for common stocks is even flatter after the sample periods used in the early empirical work on the CAPM. The estimate of the beta premium is, however, clouded by statistical uncertainty (a large standard error). Kothari, Shanken and Sloan (1995) try to resuscitate the Sharpe-Lintner CAPM by arguing that the weak relation between average return and beta is just a chance result. But the strong evidence that other variables capture variation in expected return missed by beta makes this argument irrelevant. If betas do not suffice to explain expected returns, the market portfolio is not efficient, and the CAPM is dead in its tracks. Evidence on the size of the market premium can neither save the model nor further doom it.

The synthesis of the evidence on the empirical problems of the CAPM provided by Fama and French (1992) serves as a catalyst, marking the point when it is generally acknowledged that the CAPM has potentially fatal problems. Research then turns to explanations.

One possibility is that the CAPM's problems are spurious, the result of data dredging—publication-hungry researchers scouring the data and unearthing contradictions that occur in specific samples as a result of chance. A standard response to this concern is to test for similar findings in other samples. Chan, Hamao and Lakonishok (1991) find a strong relation between book-to-market equity (B/M) and average return for Japanese stocks. Capaul, Rowley and Sharpe (1993) observe a similar B/M effect in four European stock markets and in Japan. Fama and French (1998) find that the price ratios that produce problems for the CAPM in U.S. data show up in the same way in the stock returns of twelve non-U.S. major markets, and they are present in emerging market returns. This evidence suggests that the contradictions of the CAPM associated with price ratios are not sample specific.

Explanations: Irrational Pricing or Risk

Among those who conclude that the empirical failures of the CAPM are fatal, two stories emerge. On one side are the behavioralists. Their view is based on evidence that stocks with high ratios of book value to market price are typically firms that have fallen on bad times, while low B/M is associated with growth firms (Lakonishok, Shleifer and Vishny, 1994; Fama and French, 1995). The behavioralists argue that sorting firms on book-to-market ratios exposes investor overreaction to good and bad times. Investors overextrapolate past performance, resulting in stock prices that are too high for growth (low B/M) firms and too low for distressed (high B/M, so-called value) firms. When the overreaction is eventually corrected, the result is high returns for value stocks and low returns for growth stocks. Proponents of this view include DeBondt and Thaler (1987), Lakonishok, Shleifer and Vishny (1994) and Haugen (1995).

The second story for explaining the empirical contradictions of the CAPM is that they point to the need for a more complicated asset pricing model. The CAPM is based on many unrealistic assumptions. For example, the assumption that investors care only about the mean and variance of one-period portfolio returns is extreme. It is reasonable that investors also care about how their portfolio return covaries with labor income and future investment opportunities, so a portfolio's return variance misses important dimensions of risk. If so, market beta is not a complete description of an asset's risk, and we should not be surprised to find that differences in expected return are not completely explained by differences in beta. In this view, the search should turn to asset pricing models that do a better job explaining average returns.

Merton's (1973) intertemporal capital asset pricing model (ICAPM) is a natural extension of the CAPM. The ICAPM begins with a different assumption about investor objectives. In the CAPM, investors care only about the wealth their portfolio produces at the end of the current period. In the ICAPM, investors are concerned not only with their end-of-period payoff, but also with the opportunities they will have to consume or invest the payoff. Thus, when choosing a portfolio at time $t - 1$, ICAPM investors consider how their wealth at t might vary with future *state variables*, including labor income, the prices of consumption goods and the nature of portfolio opportunities at t, and expectations about the labor income, consumption and investment opportunities to be available after t.

Like CAPM investors, ICAPM investors prefer high expected return and low return variance. But ICAPM investors are also concerned with the covariances of portfolio returns with state variables. As a result, optimal portfolios are "multifactor efficient," which means they have the largest possible expected returns, given their return variances and the covariances of their returns with the relevant state variables.

Fama (1996) shows that the ICAPM generalizes the logic of the CAPM. That is, if there is risk-free borrowing and lending or if short sales of risky assets are allowed, market clearing prices imply that the market portfolio is multifactor efficient. Moreover, multifactor efficiency implies a relation between expected return and beta risks, but it requires additional betas, along with a market beta, to explain expected returns.

An ideal implementation of the ICAPM would specify the state variables that affect expected returns. Fama and French (1993) take a more indirect approach, perhaps more in the spirit of Ross's (1976) arbitrage pricing theory. They argue that though size and book-to-market equity are not themselves state variables, the higher average returns on small stocks and high book-to-market stocks reflect unidentified state variables that produce undiversifiable risks (covariances) in returns that are not captured by the market return and are priced separately from market betas. In support of this claim, they show that the returns on the stocks of small firms covary more with one another than with returns on the stocks of large firms, and returns on high book-to-market (value) stocks covary more with one another than with returns on low book-to-market (growth) stocks. Fama and French (1995) show that there are similar size and book-to-market patterns in the covariation of fundamentals like earnings and sales.

Based on this evidence, Fama and French (1993, 1996) propose a three-factor model for expected returns,

$$\text{(Three-Factor Model)} \quad E(R_{it}) - R_{ft} = \beta_{iM}[E(R_{Mt}) - R_{ft}] + \beta_{is}E(SMB_t) + \beta_{ih}E(HML_t).$$

In this equation, SMB_t (small minus big) is the difference between the returns on diversified portfolios of small and big stocks, HML_t (high minus low) is the difference between the returns on diversified portfolios of high and low B/M stocks, and the betas are slopes in the multiple regression of $R_{it} - R_{ft}$ on $R_{Mt} - R_{ft}$, SMB_t and HML_t.

For perspective, the average value of the market premium $R_{Mt} - R_{ft}$ for 1927–2003 is 8.3 percent per year, which is 3.5 standard errors from zero. The average values of SMB_t, and HML_t are 3.6 percent and 5.0 percent per year, and they are 2.1 and 3.1 standard errors from zero. All three premiums are volatile, with annual standard deviations of 21.0 percent ($R_{Mt} - R_{ft}$), 14.6 percent (SMB_t) and 14.2 percent (HML_t) per year. Although the average values of the premiums are large, high volatility implies substantial uncertainty about the true expected premiums.

One implication of the expected return equation of the three-factor model is that the intercept α_i in the time-series regression,

$$R_{it} - R_{ft} = \alpha_i + \beta_{iM}(R_{Mt} - R_{ft}) + \beta_{is}SMB_t + \beta_{ih}HML_t + \varepsilon_{it},$$

is zero for all assets i. Using this criterion, Fama and French (1993, 1996) find that the model captures much of the variation in average return for portfolios formed on size, book-to-market equity and other price ratios that cause problems for the CAPM. Fama and French (1998) show that an international version of the model performs better than an international CAPM in describing average returns on portfolios formed on scaled price variables for stocks in 13 major markets.

The three-factor model is now widely used in empirical research that requires a model of expected returns. Estimates of α_i from the time-series regression above are used to calibrate how rapidly stock prices respond to new information (for example, Loughran and Ritter, 1995; Mitchell and Stafford, 2000). They are also used to measure the special information of portfolio managers, for example, in Carhart's (1997) study of mutual fund performance. Among practitioners like Ibbotson Associates, the model is offered as an alternative to the CAPM for estimating the cost of equity capital.

From a theoretical perspective, the main shortcoming of the three-factor model is its empirical motivation. The small-minus-big (SMB) and high-minus-low (HML) explanatory returns are not motivated by predictions about state variables of concern to investors. Instead they are brute force constructs meant to capture the patterns uncovered by previous work on how average stock returns vary with size and the book-to-market equity ratio.

But this concern is not fatal. The ICAPM does not require that the additional portfolios used along with the market portfolio to explain expected returns "mimic" the relevant state variables. In both the ICAPM and the arbitrage pricing theory, it suffices that the additional portfolios are well diversified (in the terminology of Fama, 1996, they are multifactor minimum variance) and that they are sufficiently different from the market portfolio to capture covariation in returns and variation in expected returns missed by the market portfolio. Thus, adding diversified portfolios that capture covariation in returns and variation in average returns left unexplained by the market is in the spirit of both the ICAPM and the Ross's arbitrage pricing theory.

The behavioralists are not impressed by the evidence for a risk-based explanation of the failures of the CAPM. They typically concede that the three-factor model captures covariation in returns missed by the market return and that it picks up much of the size and value effects in average returns left unexplained by the CAPM. But their view is that the average return premium associated with the model's book-to-market factor—which does the heavy lifting in the improvements to the CAPM—is itself the result of investor overreaction that happens to be correlated across firms in a way that just looks like a risk story. In short, in the behavioral view, the market tries to set CAPM prices, and violations of the CAPM are due to mispricing.

The conflict between the behavioral irrational pricing story and the rational risk story for the empirical failures of the CAPM leaves us at a timeworn impasse. Fama (1970) emphasizes that the hypothesis that prices properly reflect available information must be tested in the context of a model of expected returns, like the CAPM. Intuitively, to test whether prices are rational, one must take a stand on what the market is trying to do in setting prices—that is, what is risk and what is the relation between expected return and risk? When tests reject the CAPM, one cannot say whether the problem is its assumption that prices are rational (the behavioral view) or violations of other assumptions that are also necessary to produce the CAPM (our position).

Fortunately, for some applications, the way one uses the three-factor model does not depend on one's view about whether its average return premiums are the rational result of underlying state variable risks, the result of irrational investor behavior or sample specific results of chance. For example, when measuring the response of stock prices to new information or when evaluating the performance of managed portfolios, one wants to account for known patterns in returns and average returns for the period examined, whatever their source. Similarly, when estimating the cost of equity capital, one might be unconcerned with whether expected return premiums are rational or irrational since they are in either case part of the opportunity cost of equity capital (Stein, 1996). But the cost of capital is forward looking, so if the premiums are sample specific they are irrelevant.

The three-factor model is hardly a panacea. Its most serious problem is the momentum effect of Jegadeesh and Titman (1993). Stocks that do well relative to the market over the last three to twelve months tend to continue to do well for the next few months, and stocks that do poorly continue to do poorly. This momentum effect is distinct from the value effect captured by book-to-market equity and other price ratios. Moreover, the momentum effect is left unexplained by the three-factor model, as well as by the CAPM. Following Carhart (1997), one response is to add a momentum factor (the difference between the returns on diversified portfolios of short-term winners and losers) to the three-factor model. This step is again legitimate in applications where the goal is to abstract from known patterns in average returns to uncover information-specific or manager-specific effects. But since the momentum effect is short-lived, it is largely irrelevant for estimates of the cost of equity capital.

Another strand of research points to problems in both the three-factor model and the CAPM. Frankel and Lee (1998), Dechow, Hutton and Sloan (1999), Piotroski (2000) and others show that in portfolios formed on price ratios like book-to-market equity, stocks with higher expected cash flows have higher average returns that are not captured by the three-factor model or the CAPM. The authors interpret their results as evidence that stock prices are irrational, in the sense that they do not reflect available information about expected profitability.

In truth, however, one can't tell whether the problem is bad pricing or a bad asset pricing model. A stock's price can always be expressed as the present value of expected future cash flows discounted at the expected return on the stock (Campbell and Shiller, 1989; Vuolteenaho, 2002). It follows that if two stocks have the same price, the one with higher expected cash flows must have a higher expected return. This holds true whether pricing is rational or irrational. Thus, when one observes a positive relation between expected cash flows and expected returns that is left unexplained by the CAPM or the three-factor model, one can't tell whether it is the result of irrational pricing or a misspecified asset pricing model.

The Market Proxy Problem

Roll (1977) argues that the CAPM has never been tested and probably never will be. The problem is that the market portfolio at the heart of the model is theoretically and empirically elusive. It is not theoretically clear which assets (for example, human capital) can legitimately be excluded from the market portfolio, and data availability substantially limits the assets that are included. As a result, tests of the CAPM are forced to use proxies for the market portfolio, in effect testing whether the proxies are on the minimum variance frontier. Roll argues that because the tests use proxies, not the true market portfolio, we learn nothing about the CAPM.

We are more pragmatic. The relation between expected return and market beta of the CAPM is just the minimum variance condition that holds in any efficient portfolio, applied to the market portfolio. Thus, if we can find a market proxy that is on the minimum variance frontier, it can be used to describe differences in expected returns, and we would be happy to use it for this purpose. The strong rejections of the CAPM described above, however, say that researchers have not uncovered a reasonable market proxy that is close to the minimum variance frontier. If researchers are constrained to reasonable proxies, we doubt they ever will.

Our pessimism is fueled by several empirical results. Stambaugh (1982) tests the CAPM using a range of market portfolios that include, in addition to U.S. common stocks, corporate and government bonds, preferred stocks, real estate and other consumer durables. He finds that tests of the CAPM are not sensitive to expanding the market proxy beyond common stocks, basically because the volatility of expanded market returns is dominated by the volatility of stock returns.

One need not be convinced by Stambaugh's (1982) results since his market proxies are limited to U.S. assets. If international capital markets are open and asset prices conform to an international version of the CAPM, the market portfolio should include international assets. Fama and French (1998) find, however, that betas for a global stock market portfolio cannot explain the high average returns observed around the world on stocks with high book-to-market or high earnings-price ratios.

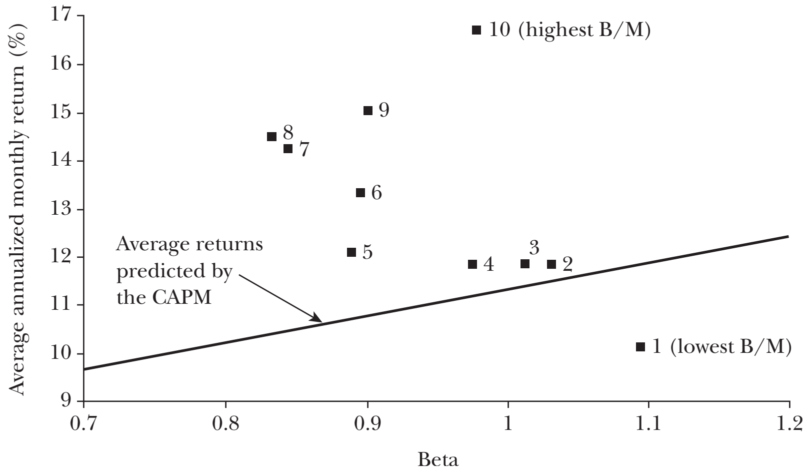
A major problem for the CAPM is that portfolios formed by sorting stocks on price ratios produce a wide range of average returns, but the average returns are not positively related to market betas (Lakonishok, Shleifer and Vishny, 1994; Fama and French, 1996, 1998). The problem is illustrated in Figure 3, which shows average returns and betas (calculated with respect to the CRSP value-weight portfolio of NYSE, AMEX and NASDAQ stocks) for July 1963 to December 2003 for ten portfolios of U.S. stocks formed annually on sorted values of the book-to-market equity ratio (B/M).[6]

Average returns on the B/M portfolios increase almost monotonically, from 10.1 percent per year for the lowest B/M group (portfolio 1) to an impressive 16.7 percent for the highest (portfolio 10). But the positive relation between beta and average return predicted by the CAPM is notably absent. For example, the portfolio with the lowest book-to-market ratio has the highest beta but the lowest average return. The estimated beta for the portfolio with the highest book-to-market ratio and the highest average return is only 0.98. With an average annualized value of the riskfree interest rate, R_f, of 5.8 percent and an average annualized market premium, $R_M - R_f$, of 11.3 percent, the Sharpe-Lintner CAPM predicts an average return of 11.8 percent for the lowest B/M portfolio and 11.2 percent for the highest, far from the observed values, 10.1 and 16.7 percent. For the Sharpe-Lintner model to "work" on these portfolios, their market betas must change dramatically, from 1.09 to 0.78 for the lowest B/M portfolio and from 0.98 to 1.98 for the highest. We judge it unlikely that alternative proxies for the market portfolio will produce betas and a market premium that can explain the average returns on these portfolios.

It is always possible that researchers will redeem the CAPM by finding a reasonable proxy for the market portfolio that is on the minimum variance frontier. We emphasize, however, that this possibility cannot be used to justify the way the CAPM is currently applied. The problem is that applications typically use the same

[6] Stock return data are from CRSP, and book equity data are from Compustat and the Moody's Industrials, Transportation, Utilities and Financials manuals. Stocks are allocated to ten portfolios at the end of June of each year t (1963 to 2003) using the ratio of book equity for the fiscal year ending in calendar year $t - 1$, divided by market equity at the end of December of $t - 1$. Book equity is the book value of stockholders' equity, plus balance sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock. Depending on availability, we use the redemption, liquidation or par value (in that order) to estimate the book value of preferred stock. Stockholders' equity is the value reported by Moody's or Compustat, if it is available. If not, we measure stockholders' equity as the book value of common equity plus the par value of preferred stock or the book value of assets minus total liabilities (in that order). The portfolios for year t include NYSE (1963–2003), AMEX (1963–2003) and NASDAQ (1972–2003) stocks with positive book equity in $t - 1$ and market equity (from CRSP) for December of $t - 1$ and June of t. The portfolios exclude securities CRSP does not classify as ordinary common equity. The breakpoints for year t use only securities that are on the NYSE in June of year t.

Figure 3
Average Annualized Monthly Return versus Beta for Value Weight Portfolios Formed on B/M, 1963–2003



market proxies, like the value-weight portfolio of U.S. stocks, that lead to rejections of the model in empirical tests. The contradictions of the CAPM observed when such proxies are used in tests of the model show up as bad estimates of expected returns in applications; for example, estimates of the cost of equity capital that are too low (relative to historical average returns) for small stocks and for stocks with high book-to-market equity ratios. In short, if a market proxy does not work in tests of the CAPM, it does not work in applications.

Conclusions

The version of the CAPM developed by Sharpe (1964) and Lintner (1965) has never been an empirical success. In the early empirical work, the Black (1972) version of the model, which can accommodate a flatter tradeoff of average return for market beta, has some success. But in the late 1970s, research begins to uncover variables like size, various price ratios and momentum that add to the explanation of average returns provided by beta. The problems are serious enough to invalidate most applications of the CAPM.

For example, finance textbooks often recommend using the Sharpe-Lintner CAPM risk-return relation to estimate the cost of equity capital. The prescription is to estimate a stock's market beta and combine it with the risk-free interest rate and the average market risk premium to produce an estimate of the cost of equity. The typical market portfolio in these exercises includes just U.S. common stocks. But empirical work, old and new, tells us that the relation between beta and average return is flatter than predicted by the Sharpe-Lintner version of the CAPM. As a result, CAPM estimates of the cost of equity for high beta stocks are too high (relative to historical average returns) and estimates for low beta stocks are too low (Friend and Blume, 1970). Similarly, if the high average returns on value stocks (with high book-to-market ratios) imply high expected returns, CAPM cost of equity estimates for such stocks are too low.[7]

The CAPM is also often used to measure the performance of mutual funds and other managed portfolios. The approach, dating to Jensen (1968), is to estimate the CAPM time-series regression for a portfolio and use the intercept (Jensen's alpha) to measure abnormal performance. The problem is that, because of the empirical failings of the CAPM, even passively managed stock portfolios produce abnormal returns if their investment strategies involve tilts toward CAPM problems (Elton, Gruber, Das and Hlavka, 1993). For example, funds that concentrate on low beta stocks, small stocks or value stocks will tend to produce positive abnormal returns relative to the predictions of the Sharpe-Lintner CAPM, even when the fund managers have no special talent for picking winners.

The CAPM, like Markowitz's (1952, 1959) portfolio model on which it is built, is nevertheless a theoretical tour de force. We continue to teach the CAPM as an introduction to the fundamental concepts of portfolio theory and asset pricing, to be built on by more complicated models like Merton's (1973) ICAPM. But we also warn students that despite its seductive simplicity, the CAPM's empirical problems probably invalidate its use in applications.

■ *We gratefully acknowledge the comments of John Cochrane, George Constantinides, Richard Leftwich, Andrei Shleifer, René Stulz and Timothy Taylor.*

[7] The problems are compounded by the large standard errors of estimates of the market premium and of betas for individual stocks, which probably suffice to make CAPM estimates of the cost of equity rather meaningless, even if the CAPM holds (Fama and French, 1997; Pastor and Stambaugh, 1999). For example, using the U.S. Treasury bill rate as the risk-free interest rate and the CRSP value-weight portfolio of publicly traded U.S. common stocks, the average value of the equity premium $R_{Mt} - R_{ft}$ for 1927–2003 is 8.3 percent per year, with a standard error of 2.4 percent. The two standard error range thus runs from 3.5 percent to 13.1 percent, which is sufficient to make most projects appear either profitable or unprofitable. This problem is, however, hardly special to the CAPM. For example, expected returns in all versions of Merton's (1973) ICAPM include a market beta and the expected market premium. Also, as noted earlier the expected values of the size and book-to-market premiums in the Fama-French three-factor model are also estimated with substantial error.

References

Ball, Ray. 1978. "Anomalies in Relationships Between Securities' Yields and Yield-Surrogates." *Journal of Financial Economics.* 6:2, pp. 103–26.

Banz, Rolf W. 1981. "The Relationship Between Return and Market Value of Common Stocks." *Journal of Financial Economics.* 9:1, pp. 3–18.

Basu, Sanjay. 1977. "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis." *Journal of Finance.* 12:3, pp. 129–56.

Bhandari, Laxmi Chand. 1988. "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence." *Journal of Finance.* 43:2, pp. 507–28.

Black, Fischer. 1972. "Capital Market Equilibrium with Restricted Borrowing." *Journal of Business.* 45:3, pp. 444–54.

Black, Fischer, Michael C. Jensen and Myron Scholes. 1972. "The Capital Asset Pricing Model: Some Empirical Tests," in *Studies in the Theory of Capital Markets.* Michael C. Jensen, ed. New York: Praeger, pp. 79–121.

Blume, Marshall. 1970. "Portfolio Theory: A Step Towards its Practical Application." *Journal of Business.* 43:2, pp. 152–74.

Blume, Marshall and Irwin Friend. 1973. "A New Look at the Capital Asset Pricing Model." *Journal of Finance.* 28:1, pp. 19–33.

Campbell, John Y. and Robert J. Shiller. 1989. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies.* 1:3, pp. 195–228.

Capaul, Carlo, Ian Rowley and William F. Sharpe. 1993. "International Value and Growth Stock Returns." *Financial Analysts Journal.* January/February, 49, pp. 27–36.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance.* 52:1, pp. 57–82.

Chan, Louis K.C., Yasushi Hamao and Josef Lakonishok. 1991. "Fundamentals and Stock Returns in Japan." *Journal of Finance.* 46:5, pp. 1739–789.

DeBondt, Werner F. M. and Richard H. Thaler. 1987. "Further Evidence on Investor Overreaction and Stock Market Seasonality." *Journal of Finance.* 42:3, pp. 557–81.

Dechow, Patricia M., Amy P. Hutton and Richard G. Sloan. 1999. "An Empirical Assessment of the Residual Income Valuation Model." *Journal of Accounting and Economics.* 26:1, pp. 1–34.

Douglas, George W. 1968. *Risk in the Equity Markets: An Empirical Appraisal of Market Efficiency.* Ann Arbor, Michigan: University Microfilms, Inc.

Elton, Edwin J., Martin J. Gruber, Sanjiv Das and Matt Hlavka. 1993. "Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios." *Review of Financial Studies.* 6:1, pp. 1–22.

Fama, Eugene F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance.* 25:2, pp. 383–417.

Fama, Eugene F. 1996. "Multifactor Portfolio Efficiency and Multifactor Asset Pricing." *Journal of Financial and Quantitative Analysis.* 31:4, pp. 441–65.

Fama, Eugene F. and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance.* 47:2, pp. 427–65.

Fama, Eugene F. and Kenneth R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics.* 33:1, pp. 3–56.

Fama, Eugene F. and Kenneth R. French. 1995. "Size and Book-to-Market Factors in Earnings and Returns." *Journal of Finance.* 50:1, pp. 131–55.

Fama, Eugene F. and Kenneth R. French. 1996. "Multifactor Explanations of Asset Pricing Anomalies." *Journal of Finance.* 51:1, pp. 55–84.

Fama, Eugene F. and Kenneth R. French. 1997. "Industry Costs of Equity." *Journal of Financial Economics.* 43:2 pp. 153–93.

Fama, Eugene F. and Kenneth R. French. 1998. "Value Versus Growth: The International Evidence." *Journal of Finance.* 53:6, pp. 1975–999.

Fama, Eugene F. and James D. MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy.* 81:3, pp. 607–36.

Frankel, Richard and Charles M.C. Lee. 1998. "Accounting Valuation, Market Expectation, and Cross-Sectional Stock Returns." *Journal of Accounting and Economics.* 25:3 pp. 283–319.

Friend, Irwin and Marshall Blume. 1970. "Measurement of Portfolio Performance under Uncertainty." *American Economic Review.* 60:4, pp. 607–36.

Gibbons, Michael R. 1982. "Multivariate Tests of Financial Models: A New Approach." *Journal of Financial Economics.* 10:1, pp. 3–27.

Gibbons, Michael R., Stephen A. Ross and Jay Shanken. 1989. "A Test of the Efficiency of a Given Portfolio." *Econometrica.* 57:5, pp. 1121–152.

Haugen, Robert. 1995. *The New Finance: The*

Case against Efficient Markets. Englewood Cliffs, N.J.: Prentice Hall.

Jegadeesh, Narasimhan and Sheridan Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance.* 48:1, pp. 65–91.

Jensen, Michael C. 1968. "The Performance of Mutual Funds in the Period 1945–1964." *Journal of Finance.* 23:2, pp. 389–416.

Kothari, S. P., Jay Shanken and Richard G. Sloan. 1995. "Another Look at the Cross-Section of Expected Stock Returns." *Journal of Finance.* 50:1, pp. 185–224.

Lakonishok, Josef and Alan C. Shapiro. 1986. Systemaitc Risk, Total Risk, and Size as Determinants of Stock Market Returns." *Journal of Banking and Finance.* 10:1, pp. 115–32.

Lakonishok, Josef, Andrei Shleifer and Robert W. Vishny. 1994. "Contrarian Investment, Extrapolation, and Risk." *Journal of Finance.* 49:5, pp. 1541–578.

Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics.* 47:1, pp. 13–37.

Loughran, Tim and Jay. R. Ritter. 1995. "The New Issues Puzzle." *Journal of Finance.* 50:1, pp. 23–51.

Markowitz, Harry. 1952. "Portfolio Selection." *Journal of Finance.* 7:1, pp. 77–99.

Markowitz, Harry. 1959. *Portfolio Selection: Efficient Diversification of Investments.* Cowles Foundation Monograph No. 16. New York: John Wiley & Sons, Inc.

Merton, Robert C. 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica.* 41:5, pp. 867–87.

Miller, Merton and Myron Scholes. 1972. "Rates of Return in Relation to Risk: A Reexamination of Some Recent Findings," in *Studies in the Theory of Capital Markets.* Michael C. Jensen, ed. New York: Praeger, pp. 47–78.

Mitchell, Mark L. and Erik Stafford. 2000. "Managerial Decisions and Long-Term Stock Price Performance." *Journal of Business.* 73:3, pp. 287–329.

Pastor, Lubos and Robert F. Stambaugh. 1999. "Costs of Equity Capital and Model Mispricing." *Journal of Finance.* 54:1, pp. 67–121.

Piotroski, Joseph D. 2000. "Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers." *Journal of Accounting Research.* 38:Supplement, pp. 1–51.

Reinganum, Marc R. 1981. "A New Empirical Perspective on the CAPM." *Journal of Financial and Quantitative Analysis.* 16:4, pp. 439–62.

Roll, Richard. 1977. "A Critique of the Asset Pricing Theory's Tests' Part I: On Past and Potential Testability of the Theory." *Journal of Financial Economics.* 4:2, pp. 129–76.

Rosenberg, Barr, Kenneth Reid and Ronald Lanstein. 1985. "Persuasive Evidence of Market Inefficiency." *Journal of Portfolio Management.* Spring, 11, pp. 9–17.

Ross, Stephen A. 1976. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory.* 13:3, pp. 341–60.

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance.* 19:3, pp. 425–42.

Stambaugh, Robert F. 1982. "On The Exclusion of Assets from Tests of the Two-Parameter Model: A Sensitivity Analysis." *Journal of Financial Economics.* 10:3, pp. 237–68.

Stattman, Dennis. 1980. "Book Values and Stock Returns." *The Chicago MBA: A Journal of Selected Papers.* 4, pp. 25–45.

Stein, Jeremy. 1996. "Rational Capital Budgeting in an Irrational World." *Journal of Business.* 69:4, pp. 429–55.

Tobin, James. 1958. "Liquidity Preference as Behavior Toward Risk." *Review of Economic Studies.* 25:2, pp. 65–86.

Vuolteenaho, Tuomo. 2002. "What Drives Firm Level Stock Returns?" *Journal of Finance.* 57:1, pp. 233–64.

This article has been cited by:

1. Kam Fong Chan, Terry Marsh. 2021. Asset prices, midterm elections, and political uncertainty. *Journal of Financial Economics* **141**:1, 276-296. [Crossref]
2. Christoph Merkle, Christoph J. Sextroh. 2021. Value and momentum from investors' perspective: Evidence from professionals' risk-ratings. *Journal of Empirical Finance* **62**, 159-178. [Crossref]
3. Shaoping Wang, Lu Yu, Qing Zhao. 2021. Do factor models explain stock returns when prices behave explosively? Evidence from China. *Pacific-Basin Finance Journal* **67**, 101535. [Crossref]
4. Joel M. Vanden. 2021. Equilibrium asset pricing and the cross section of expected returns. *Annals of Finance* **17**:2, 153-186. [Crossref]
5. Abdul Qoyum, Rizqi Umar Al Hashfi, Alyta Shabrina Zusryn, Hadri Kusuma, Ibnu Qizam. 2021. Does an Islamic-SRI portfolio really matter? Empirical application of valuation models in Indonesia. *Borsa Istanbul Review* **21**:2, 105-124. [Crossref]
6. Brandon Flores, Blessing Ofori-Atta, Andrey Sarantsev. 2021. A stock market model based on CAPM and market size. *Annals of Finance* **4**. . [Crossref]
7. Semen Yu. BOGATYREV. 2021. Simulation of emotional differences in the structured query language for databases of financial markets. *Financial Analytics: Science and Experience* **14**:2, 156-173. [Crossref]
8. Jian Zhang, Jie Li. 2021. Factorized estimation of high-dimensional nonparametric covariance models. *Scandinavian Journal of Statistics* **36**. . [Crossref]
9. Emanuel Kohlscheen, Előd Takáts. 2021. What can commercial property performance reveal about bank valuations?. *Journal of International Money and Finance* **113**, 102350. [Crossref]
10. Yiannis Karavias, Stella Spilioti, Elias Tzavalis. 2021. Investor sentiment effects on share price deviations from their intrinsic values based on accounting fundamentals. *Review of Quantitative Finance and Accounting* **56**:4, 1593-1621. [Crossref]
11. Sergei Yu. BOGATYREV. 2021. Looking into bubbles in financial markets and the emotional side of corporate forecast completion through modeling in the structured query language of financial databases. *Finance and Credit* **27**:4, 833-850. [Crossref]
12. Jessie Y. Zhu, Wally Smieliauskas. 2021. Evidence on the Economic Consequences of Marriage Equality and LGBT Human Rights. *Journal of Business Ethics* **5**. . [Crossref]
13. Muhammad Imran, Mengyun Wu, Linrong Zhang, Yun Zhao, Noor Jehan, Hee Cheol Moon. 2021. Market Premium and Macroeconomic Factors as Determinants of Industry Premium: Evidence from Emerging Economies. *Complexity* **2021**, 1-11. [Crossref]
14. Vu Tuan Chu, Trang Hanh Lam Pham. 2021. Zero leverage and product market competition. *SN Business & Economics* **1**:4. . [Crossref]
15. Levan Efremidze, Darrol J. Stanley, Clemens Kownatzki. 2021. Entropy trading strategies reveal inefficiencies in Japanese stock market. *International Review of Economics & Finance* **45**. . [Crossref]
16. Sergei Yu. BOGATYREV. 2021. The sentiment analysis method in finance: The psychological-financial index. *Finance and Credit* **27**:3, 561-584. [Crossref]
17. Ade Imam Muslim, Doddy Setiawan. 2021. Information Asymmetry, Ownership Structure and Cost of Equity Capital: The Formation for Open Innovation. *Journal of Open Innovation: Technology, Market, and Complexity* **7**:1, 48. [Crossref]
18. Catalin Dragomirescu-Gaina, Dionisis Philippas, Mike G. Tsionas. 2021. Trading off accuracy for speed: Hedge funds' decision-making under uncertainty. *International Review of Financial Analysis* **127**, 101728. [Crossref]

19. Karl-Heinz Tödter, Gerhard Ziebarth. 2021. Lifetime Cost of Living and Effective Prices: Theory and Evidence for Germany. *Jahrbücher für Nationalökonomie und Statistik* **241**:1, 29-69. [Crossref]

20. Nadia Anjum, Suresh Kumar Oad Rajput. 2021. Forecasting Islamic equity indices alpha. *International Journal of Islamic and Middle Eastern Finance and Management* **14**:1, 183-203. [Crossref]

21. Xiang Lin, Martin Thomas Falk. 2021. Nordic stock market performance of the travel and leisure industry during the first wave of Covid-19 pandemic. *Tourism Economics* **454**, 135481662199093. [Crossref]

22. Ze Zhang, Zilai Tang. 2021. Examination and Interpretation of the Quantitative Validity in China's Corporate-based Urban Network Analysis. *Chinese Geographical Science* **31**:1, 41-53. [Crossref]

23. Brittany Tarufelli, Brian Snyder, David Dismukes. 2021. The Potential Impact of the U.S. Carbon Capture and Storage Tax Credit Expansion on the Economic Feasibility of Industrial Carbon Capture and Storage. *Energy Policy* **149**, 112064. [Crossref]

24. Elena Shakina, Iuliia Naidenova, Angel Barajas. 2021. Shadow prices for intangible resources. *Journal of Intellectual Capital* **ahead-of-print**:ahead-of-print. . [Crossref]

25. Alfonso A. Rojo-Ramírez. 2021. Rendimiento mínimo del inversor-propietario. El caso de la empresa pyme familiar. *Small Business International Review* **5**:1, e287. [Crossref]

26. Marcel Aloy, Floris Laly, Sébastien Laurent, Christelle Lecourt. Modeling Time-Varying Conditional Betas. A Comparison of Methods with Application for REITs 229-264. [Crossref]

27. Mario Situm. 2021. Determination of expected cost of equity with the CAPM: Theoretical extension using the law of error propagation. *Managerial and Decision Economics* **42**:1, 77-84. [Crossref]

28. Olivier Dessaint, Jacques Olivier, Clemens A Otto, David Thesmar. 2021. CAPM-Based Company (Mis)valuations. *The Review of Financial Studies* **34**:1, 1-66. [Crossref]

29. Jhumur Sengupta. Application of Econometrics in Business Research 137-159. [Crossref]

30. Maurizio Montone, Remco C.J. Zwinkels. 2021. Risk, return, and sentiment in a virtual asset market. *SSRN Electronic Journal* **25**. . [Crossref]

31. James W. Kolari, Wei Liu, Jianhua Z. Huang. Capital Asset Pricing Models 25-52. [Crossref]

32. James W. Kolari, Wei Liu, Jianhua Z. Huang. Cross-Sectional Tests of the ZCAPM 159-195. [Crossref]

33. James W. Kolari, Wei Liu, Jianhua Z. Huang. Synopsis of Asset Pricing and the ZCAPM 261-285. [Crossref]

34. Konstantin Kostin, Philippe Runge, Ronald Adams. 2021. Investment strategies in pandemic situations: An analysis and comparison of prospective returns between developed and emerging markets. *Strategic Management* **26**:1, 34-52. [Crossref]

35. Jamshid Karimov, Faruk Balli, Hatice Ozer-Balli, Anne Bruin. 2020. Firm-level political risk and Shari'ah compliance: equity capital cost and payouts policy. *Accounting & Finance* **57**. . [Crossref]

36. Rai Imtiaz Hussain, Shahid Bashir, Shahbaz Hussain. 2020. Financial Sustainability and Corporate Social Responsibility Under Mediating Effect of Operational Self-Sustainability. *Frontiers in Psychology* **11**. . [Crossref]

37. MICHAEL UNGEHEUER, MARTIN WEBER. 2020. The Perception of Dependence, Investment Decisions, and Stock Prices. *The Journal of Finance* **85**. . [Crossref]

38. Fabio B. Gaertner, Asad Kausar, Logan B. Steele. 2020. Negative accounting earnings and gross domestic product. *Review of Accounting Studies* **25**:4, 1382-1409. [Crossref]

39. Augustin Landier, David Thesmar. 2020. Earnings Expectations during the COVID-19 Crisis*. *The Review of Asset Pricing Studies* **10**:4, 598-617. [Crossref]

40. Terrence Hendershott, Dmitry Livdan, Dominik Rösch. 2020. Asset pricing: A tale of night and day. *Journal of Financial Economics* **138**:3, 635-662. [Crossref]

41. Matthew Wang, Yi-Hong Lin, Ilya Mikhelson. 2020. Regime-Switching Factor Investing with Hidden Markov Models. *Journal of Risk and Financial Management* **13**:12, 311. [Crossref]

42. Yuan Hu, Abootaleb Shirvani, W. Brent Lindquist, Frank J. Fabozzi, Svetlozar T. Rachev. 2020. Option Pricing Incorporating Factor Dynamics in Complete Markets. *Journal of Risk and Financial Management* **13**:12, 321. [Crossref]

43. Steven D. Silver, Marko Raseta, Alina Bazarova. 2020. Dynamics of Phase Transitions in Expectations for Financial Markets: An Agent-Based, Multicomponent Model. *Journal of Behavioral Finance* **55**, 1-15. [Crossref]

44. Gregory Price, Warren Whatley. 2020. Did profitable slave trading enable the expansion of empire?: The Asiento de Negros, the South Sea Company and the financial revolution in Great Britain. *Cliometrica* **102**. . [Crossref]

45. Balázs Csillag, Gábor Neszveda. 2020. A gazdasági várakozások hatása a tőzsdei momentumstratégiára. *Közgazdasági Szemle* **67**:11, 1093-1111. [Crossref]

46. Ganggang Guo, Yulei Rao, Feida Zhu, Fang Xu. 2020. Innovative deep matching algorithm for stock portfolio selection using deep stock profiles. *PLOS ONE* **15**:11, e0241573. [Crossref]

47. Abhijeet Ghadge, Sarat Kumar Jena, Sachin Kamble, Dheeraj Misra, Manoj Kumar Tiwari. 2020. Impact of financial risk on supply chains: a manufacturer-supplier relational perspective. *International Journal of Production Research* **12**, 1-16. [Crossref]

48. Chariton Chalvatzis, Dimitrios Hristu-Varsakelis. 2020. High-performance stock index trading via neural networks and trees. *Applied Soft Computing* **96**, 106567. [Crossref]

49. Jonas Puck, Igor Filatotchev. 2020. Finance and the multinational company: Building bridges between finance and global strategy research. *Global Strategy Journal* **10**:4, 655-675. [Crossref]

50. Chinh Duc Pham, Le Tan Phuoc. 2020. An augmented capital asset pricing model using new macroeconomic determinants. *Heliyon* **6**:10, e05185. [Crossref]

51. Gerson de Souza Raimundo Júnior, Rafael Baptista Palazzi, Ricardo de Souza Tavares, Marcelo Cabus Klotzle. 2020. Market Stress and Herding: A New Approach to the Cryptocurrency Market. *Journal of Behavioral Finance* **33**, 1-15. [Crossref]

52. Lin Chen, Junbo Wang, Chunchi Wu, Hongquan Zhu. Divergent Opinion, Trading Information, and Stock Price Co-movements 1-21. [Crossref]

53. Ali Boloorforoosh, Peter Christoffersen, Mathieu Fournier, Christian Gouriéroux. 2020. Beta Risk in the Cross-Section of Equities. *The Review of Financial Studies* **33**:9, 4318-4366. [Crossref]

54. Michael Curran, Adnan Velic. 2020. The CAPM, National Stock Market Betas, and Macroeconomic Covariates: a Global Analysis. *Open Economies Review* **31**:4, 787-820. [Crossref]

55. Jamshid Karimov, Faruk Balli, Hatice Ozer Balli, Anne de Bruin. 2020. Shari'ah compliance requirements and the cost of equity capital. *Pacific-Basin Finance Journal* **62**, 101349. [Crossref]

56. Vanita Tripathi, Priti Aggarwal. 2020. Is value premium sector-specific? Evidence from India. *Managerial Finance* **46**:12, 1605-1628. [Crossref]

57. Tong Zhang. Stock Picking Strategy Based on Exploration of Chip Distribution Indicators 276-282. [Crossref]

58. Lerzan Aksoy, Sabine Benoit, Shreekant G. Joag, Jay Kandampully, Timothy Lee Keiningham, An L. Yan. 2020. Enterprise feedback management (EFM): what lies beyond the hype?. *Journal of Service Management* **32**:1, 53-69. [Crossref]

59. Patricia Sepúlveda Orejuela, María Dolores Guerrero-Baena, José A. Gómez-Limón. 2020. Desempeño económico-financiero de los distintos modelos empresariales en el sector del aceite de oliva en España. *Revista de Estudios Empresariales. Segunda Época* :1, 227-248. [Crossref]
60. Yingyi Hu, Tiao Zhao, Lin Zhang. 2020. Noise trading, institutional trading, and opinion divergence: Evidence on intraday data in the Chinese stock market. *International Review of Economics & Finance* **68**, 74-89. [Crossref]
61. Myrthe van Dieijen, Abhishek Borah, Gerard J. Tellis, Philip Hans Franses. 2020. Big Data Analysis of Volatility Spillovers of Brands across Social Media and Stock Markets. *Industrial Marketing Management* **88**, 465-484. [Crossref]
62. Chinh Duc Pham, Le Tan Phuoc. 2020. Is estimating the Capital Asset Pricing Model using monthly and short-horizon data a good choice?. *Heliyon* **6**:7, e04339. [Crossref]
63. A. S. Shaghikyan, H. N. Hayrapetyan. 2020. Equity Crowdfunding in the Eurasian Economic Union (EAEU). *Finance: Theory and Practice* **24**:3, 45-59. [Crossref]
64. Seong Mi Bae, Hyoung-Tae An, Jong Dae Kim. 2020. Mediators Linking Information Quality and the Cost of Equity Capital*. *Asia-Pacific Journal of Financial Studies* **49**:3, 410-437. [Crossref]
65. Naji Massad, Jørgen Vitting Andersen. 2020. Defining an intrinsic "stickiness" parameter of stock price returns. *Physica A: Statistical Mechanics and its Applications* **547**, 124464. [Crossref]
66. Faruk Balli, Md Iftekhar Hasan Chowdhury, Anne de Bruin. 2020. Transition to Islamic equities: Systematic risk and Shari'ah compliance. *Global Finance Journal* 100552. [Crossref]
67. Antonio Salvi, Filippo Vitolla, Nicola Raimo, Michele Rubino, Felice Petruzzella. 2020. Does intellectual capital disclosure affect the cost of equity capital? An empirical analysis in the integrated reporting context. *Journal of Intellectual Capital* **21**:6, 985-1007. [Crossref]
68. Nisa Vinodkumar, Hadeel Khalid AlJasser. 2020. Financial evaluation of Tadawul All Share Index (TASI) listed stocks using Capital Asset Pricing Model. *Investment Management and Financial Innovations* **17**:2, 69-75. [Crossref]
69. Thorsten Hens, Fatemeh Naebi. 2020. Behavioural heterogeneity in the capital asset pricing model with an application to the low-beta anomaly. *Applied Economics Letters* **1**, 1-7. [Crossref]
70. Jie Qin. 2020. Regret-based capital asset pricing model. *Journal of Banking & Finance* **114**, 105784. [Crossref]
71. 2020. The Quantified Analysis of Causes of Market Risk Fluctuations in the Group of Construction, Real Estate and Construction Material Companies in Vietnam During and After the Global Crisis 2007-2011. *WSEAS TRANSACTIONS ON ENVIRONMENT AND DEVELOPMENT* **16**. . [Crossref]
72. Saad Faysal, Mahdi Salehi, Mahdi Moradi. 2020. The impact of ownership structure on the cost of equity in emerging markets. *Management Research Review* **43**:10, 1221-1239. [Crossref]
73. Saejoon Kim, Soong Kim. 2020. Index tracking through deep latent representation learning. *Quantitative Finance* **20**:4, 639-652. [Crossref]
74. Benjamin Pfister, Manfred Schwaiger, Tobias Morath. 2020. Corporate reputation and the future cost of equity. *Business Research* **13**:1, 343-384. [Crossref]
75. Panagiotis Anastasiadis, Efthimios Katsaros, Anastasios-Taxiarchis Koutsioukis, Athanasios Pandazis. 2020. Performance-Risk Nexus of Global Low-Rated ETFs During the QE-Tapering Period. *Studies in Business and Economics* **15**:1, 194-211. [Crossref]
76. Yehui Wang, Jianxu Liu, Yuxuan Tang, Songsak Sriboonchitta. 2020. Housing Risk and Its Influence on House Price: An Expected Utility Approach. *Mathematical Problems in Engineering* **2020**, 1-16. [Crossref]

77. Anastasia Simmet, Winfried Pohlmeier. 2020. The CAPM with Measurement Error: 'There's life in the old dog yet!'. *Jahrbücher für Nationalökonomie und Statistik* **240**:4, 417-453. [Crossref]
78. O. E. Medvedeva, A. I. Artemenkov. 2020. THEORETICAL BASES OF ECONOMIC MEASUREMENT OF VALUE IN THE CONDITIONS OF GLOBAL TECHNOLOGICAL SHIFTS AND CRISES. NEWEST METHODOLOGICAL BASE (PART 2). *Vestnik Universiteta* :1, 114-120. [Crossref]
79. Semra Bank, Evrim Erdogan Yazar, Ugur Sivri. 2020. The portfolios with strong brand value: More returns? Lower risk?. *Borsa Istanbul Review* **20**:1, 64-79. [Crossref]
80. Damian Shaw-Williams, Connie Susilawati. 2020. A techno-economic evaluation of Virtual Net Metering for the Australian community housing sector. *Applied Energy* **261**, 114271. [Crossref]
81. Prince Worzie. 2020. FACTORS AFFECTING INVESTMENT DECISIONS AMONG LISTED FIRMS IN THE NAIROBI SECURITIES EXCHANGE. *International Journal of Engineering Technologies and Management Research* **7**:2, 124-142. [Crossref]
82. Mehmet Levent ERDAŞ. 2020. Belirli Kısıtlar Altında Doğrusal Programlamaya Dayalı Bir Portföy Optimizasyonu Modelinin Geliştirilmesi: Borsa İstanbul 30 Endeksi Üzerine Bir Uygulama. *TESAM Akademi Dergisi* 115-141. [Crossref]
83. Fenghua Wen, Nan Wu, Xu Gong. 2020. China's carbon emissions trading and stock returns. *Energy Economics* **86**, 104627. [Crossref]
84. Le Tan Phuoc, Chinh Duc Pham. 2020. The systematic risk estimation models: A different perspective. *Heliyon* **6**:2, e03371. [Crossref]
85. Filippo Vitolla, Antonio Salvi, Nicola Raimo, Felice Petruzzella, Michele Rubino. 2020. The impact on the cost of equity capital in the effects of integrated reporting quality. *Business Strategy and the Environment* **29**:2, 519-529. [Crossref]
86. Shea D. Chen, Andrew E. B. Lim. 2020. A Generalized Black–Litterman Model. *Operations Research* . [Crossref]
87. Oussama Tilfani, Paulo Ferreira, My Youssef El Boukfaoui. 2020. Multiscale optimal portfolios using CAPM fractal regression: estimation for emerging stock markets. *Post-Communist Economies* **32**:1, 77-112. [Crossref]
88. Andres F. Cantillo. 2020. Production commitments and the financial foundations of specialized economies. *Journal of Post Keynesian Economics* **43**:1, 90-111. [Crossref]
89. Tarana Azimova. Challenges in Estimation of Beta 79-98. [Crossref]
90. Luis Javier Sanchez-Barrios, Benedicto Kulwizira Lukanima, Natalia Hernandez-Vargas, Luis Ricardo Almanza Herazo. Estimating the CAPM Beta for Public and Private Firms 99-125. [Crossref]
91. Stefan Behringer. Die Entwicklung der angelsächsischen Unternehmensbewertung – kapitalmarktorientierter Ansatz 79-102. [Crossref]
92. Tolulope Latunde, Lukman Shina Akinola, Damilola Deborah Dare. 2020. Analysis of capital asset pricing model on Deutsche bank energy commodity. *Green Finance* **2**:1, 20-34. [Crossref]
93. Sandra Andraszewicz. Stock Markets, Market Crashes, and Market Bubbles 205-231. [Crossref]
94. Hasan A Fallahgoul, Vincentius Franstianto. 2020. Towards Explaining Deep Learning: Significance Tests for Multi-Layer Perceptrons. *SSRN Electronic Journal* . [Crossref]
95. Augustin Landier, David Thesmar. 2020. Earnings Expectations in the COVID Crisis. *SSRN Electronic Journal* . [Crossref]
96. 2020. Wavelet-based systematic risk estimation: application on GCC stock markets: the Saudi Arabia case. *Quantitative Finance and Economics* **4**:4, 542-595. [Crossref]

97. Duane W Rockerbie, Stephen T. Easton. Contract Options for Buyers and Sellers of Talent 53-67. [Crossref]
98. Petri Jylha, Michael Ungeheuer. 2020. Growth Expectations out of WACC. *SSRN Electronic Journal* . [Crossref]
99. Emanuel Kohlscheen, Előd Takáts. 2020. What Can Commercial Property Performance Reveal about Bank Valuations?. *SSRN Electronic Journal* . [Crossref]
100. Anne M. Tucker, Yusen Xia, Susan Smelcer. 2020. It Ain't Just What Funds Disclose (It's The Way That They Do It). *SSRN Electronic Journal* **35**. . [Crossref]
101. Stefan Anchev, Nicha Lapanan. 2020. Investor Base Size and Stock Return Anomalies. *SSRN Electronic Journal* . [Crossref]
102. Majeed Simaan. 2020. Working with CRSP/COMPUSTAT in R: Reproducible Empirical Asset Pricing. *SSRN Electronic Journal* **1**. . [Crossref]
103. Kam Fong Chan, Terry Marsh. 2020. Asset Pricing around Earnings Announcement Days. *SSRN Electronic Journal* . [Crossref]
104. Phuong Duong, Jinghui Liu, Ian Eddie. 2020. New financial regulatory philosophy: A paradigm shift in securities market supervision. *Corporate Ownership and Control* **17**:4, 8-17. [Crossref]
105. Nguyen Huu Anh, Nguyen La Soa, Ha Hong Hanh. 2020. Environmental accounting practices and cost of capital of enterprises in Vietnam. *Cogent Economics & Finance* **8**:1, 1790964. [Crossref]
106. Marília Cordeiro Pinheiro, André Luiz Marques Serrano. 2019. Analysis of the impact of Fies on the stock returns from the higher education sector. *Revista Contabilidade & Finanças* **30**:81, 368-380. [Crossref]
107. Tomislav Gelo, Željko Vrban, Dalibor Pudić. 2019. Allowed Revenue of Network System Operators in the Croatian Energy Sector and Interest Rate Changes on the Croatian Capital Market. *Zagreb International Review of Economics and Business* **22**:s2, 73-91. [Crossref]
108. Wei Kang Loo. 2019. Predictability of HK-REITs returns using artificial neural network. *Journal of Property Investment & Finance* **38**:4, 291-307. [Crossref]
109. Saurabh Gupta, Saumitra N. Bhaduri. 2019. Skin in the game – investor behavior in asset pricing, the Indian context. *Review of Behavioral Finance* **11**:4, 373-392. [Crossref]
110. Liping Liu. 2019. The reducibility of matrix sweeping operations: A computational issue in linear belief functions. *International Journal of Approximate Reasoning* **114**, 226-247. [Crossref]
111. Babak Jafarizadeh, Reidar B. Bratvold. 2019. Exploration economics: taking opportunities and the risk of double-counting risk. *Mineral Economics* **32**:3, 323-335. [Crossref]
112. Sabine Elmiger. 2019. CAPM-anomalies: quantitative puzzles. *Economic Theory* **68**:3, 643-667. [Crossref]
113. Levan Efremidze, Darrol J. Stanley, Abraham Park, Nikolai Wasilewski. 2019. Empirical implementation of entropy risk factor model: A test on Chilean peso. *Physica A: Statistical Mechanics and its Applications* **532**, 121836. [Crossref]
114. Joanne Hamet, Frantz Maurer. 2019. Valeur intrinsèque et valeur temps de la recherche en sciences de gestion. *Revue Française de Gestion* **45**:284, 103-123. [Crossref]
115. Marc Schaffer. 2019. The role of competition, innovation, and regulation on financial intermediary risk. *Managerial Finance* **ahead-of-print**:ahead-of-print. . [Crossref]
116. Tzu-Lun Huang. 2019. Is the Fama and French five-factor model robust in the Chinese stock market?. *Asia Pacific Management Review* **24**:3, 278-289. [Crossref]
117. Natalia Bailey, George Kapetanios, M. Hashem Pesaran. 2019. Exponent of Cross-sectional Dependence for Residuals. *Sankhya B* **81**:S1, 46-102. [Crossref]

118. Yang, Nguyen. 2019. Skewness Preference and Asset Pricing: Evidence from the Japanese Stock Market. *Journal of Risk and Financial Management* **12**:3, 149. [Crossref]
119. Nesrin Özkan. 2019. q-Faktör Modelinin Borsa İstanbul'da Geçerliliğinin Test Edilmesi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi* **14**:2, 441-456. [Crossref]
120. Anwar Hasan Abdullah Othman, Syed Musa Alhabshi, Razali Haron. 2019. The effect of symmetric and asymmetric information on volatility structure of crypto-currency markets. *Journal of Financial Economic Policy* **11**:3, 432-450. [Crossref]
121. Shaun Cox, James Britten. 2019. The Fama-French five-factor model: Evidence from the Johannesburg Stock Exchange. *Investment Analysts Journal* **48**:3, 240-261. [Crossref]
122. Richard A. Michelfelder, Pauline Ahern, Dylan D'Ascendis. 2019. Decoupling impact and public utility conservation investment. *Energy Policy* **130**, 311-319. [Crossref]
123. Shijun Wang, Baocheng Zhu, Lintao Ma, Yuan Qi. A Riemannian Primal-dual Algorithm Based on Proximal Operator and its Application in Metric Learning 1-8. [Crossref]
124. Johan Knif, James W. Kolari, Gregory Koutmos, Seppo Pynnönen. 2019. Measuring the relative return contribution of risk factors. *Journal of Asset Management* **20**:4, 263-272. [Crossref]
125. Jana Šimáková, Daniel Stavárek, Tomáš Pražák, Marie Ligocká. 2019. Macroeconomic factors and stock prices in the food and drink industry. *British Food Journal* **121**:7, 1627-1641. [Crossref]
126. Ricardo Méndez Romero, Hernán Rocha Pavés. 2019. ANÁLISIS DE LOS FONDOS DE PENSIONES EN CHILE: PERIODO 2011-2018. *Multidisciplinary Business Review* **12**:1, 1-9. [Crossref]
127. Thomas Gramespacher, Armin Bänziger. 2019. The Bias in Two-Pass Regression Tests of Asset-Pricing Models in Presence of Idiosyncratic Errors with Cross-Sectional Dependence. *Review of Pacific Basin Financial Markets and Policies* **22**:02, 1950012. [Crossref]
128. Muhammad Adnan Arshad, Saira Munir, Bashir Ahmad, Muhammad Waseem. 2019. Do factors matter for predicting high-risk stock returns? Comparison of single-, three- and five-factor CAPM. *International Journal of Financial Engineering* **06**:02, 1950015. [Crossref]
129. Qianwei Ying, Tahir Yousaf, Qurat ul Ain, Yasmeen Akhtar, Muhammad Shahid Rasheed. 2019. Stock Investment and Excess Returns: A Critical Review in the Light of the Efficient Market Hypothesis. *Journal of Risk and Financial Management* **12**:2, 97. [Crossref]
130. Glen Lehman, Chris Mortensen. 2019. Finance, Nature and Ontology. *Topoi* **28**. . [Crossref]
131. Semra Bank, Evrim Erdogan Yazar, Ugur Sivri. 2019. Can social media marketing lead to abnormal portfolio returns?. *European Research on Management and Business Economics* **25**:2, 54-62. [Crossref]
132. Defeng Yang, Zhanqing Wang, Fangmin Lu. 2019. The Influence of Corporate Governance and Operating Characteristics on Corporate Environmental Investment: Evidence from China. *Sustainability* **11**:10, 2737. [Crossref]
133. Roberto Savona, Cesare Orsini. 2019. Taking the right course navigating the ERC universe. *Journal of Asset Management* **20**:3, 157-174. [Crossref]
134. Salman Ahmed Shaikh, Mohd Adib Ismail, Abdul Ghafar Ismail, Shahida Shahimi, Muhammad Hakimi Mohd. Shafiai. 2019. Cross section of stock returns on Shari'ah -compliant stocks: evidence from Pakistan. *International Journal of Islamic and Middle Eastern Finance and Management* **12**:2, 282-302. [Crossref]
135. Nilanjana Chakraborty, Mohammed M Elgammal, David McMillan. 2019. Rational functions: an alternative approach to asset pricing. *Applied Economics* **51**:20, 2091-2119. [Crossref]
136. Khaled Elkhal. 2019. Business uncertainty and financial leverage: should the firm double up on risk?. *Managerial Finance* **45**:4, 536-544. [Crossref]

137. Malik Shahzad Shabbir, Iftikhar Muhammad. 2019. The dynamic impact of foreign portfolio investment on stock prices in Pakistan. *Transnational Corporations Review* **11**:2, 166-178. [Crossref]
138. Nordine Abidi, Burcu Hacibedel, Mwanza Nkusu. 2019. Frontier and Emerging Markets: A Perspective from Portfolio Flows and Financial Integration. *Journal of Banking and Financial Economics* **1/2019**:11, 19-45. [Crossref]
139. Benjamin R. Auer, Tobias Hiller. 2019. Can cooperative game theory solve the low-risk puzzle?. *International Journal of Finance & Economics* **24**:2, 884-889. [Crossref]
140. Sergio Bravo. 2019. The Corporate Life Cycle and the Cost of Equity. *Journal of Business Valuation and Economic Loss Analysis* **14**:1. . [Crossref]
141. Marisa Basten, Antonio Sánchez Serrano. 2019. European banks after the global financial crisis: a new landscape. *Journal of Banking Regulation* **20**:1, 51-73. [Crossref]
142. Moting Su, Zongyi Zhang, Ye Zhu, Donglan Zha. 2019. Data-Driven Natural Gas Spot Price Forecasting with Least Squares Regression Boosting Algorithm. *Energies* **12**:6, 1094. [Crossref]
143. Yang Ning, Liu Chun Wah, Luo Erdan. 2019. Stock price prediction based on error correction model and Granger causality test. *Cluster Computing* **22**:S2, 4849-4858. [Crossref]
144. Henry Leung, Jeffrey Tse, P. Joakim Westerholm. 2019. CEO traders and corporate acquisitions. *Journal of Corporate Finance* **54**, 107-127. [Crossref]
145. José Roberto Ferreira Savoia, José Roberto Securato, Daniel Reed Bergmann, Fabiana Lopes da Silva. 2019. Comparing results of the implied cost of capital and capital asset pricing models for infrastructure firms in Brazil. *Utilities Policy* **56**, 149-158. [Crossref]
146. Jaromír Antoch, Jan Hanousek, Marie Hušková, Jiří Trešl. 2019. Detection of Changes in Panel Data: Change in Fama-French Model Parameters for Selected European Stocks During the Financial Crisis. *Politická ekonomie* **67**:1, 3-19. [Crossref]
147. Octávio Valente Campos, Ana Carolina Vasconcelos Colares, Renata Turola Takamatsu, José Roberto de Souza Francisco. 2019. Precificação de ativos: análise do fator book-to-market após o deemed cost. *Revista Catarinense da Ciência Contábil* **18**, 1-16. [Crossref]
148. Mukail Aremu Akinde, Eriki Peter, Ochei Ailemen Ikpefan. 2019. Growth versus value investing: a case of Nigerian Stock Market. *Investment Management and Financial Innovations* **16**:1, 30-45. [Crossref]
149. Lanh Tran. How Annualized Wavelet Trading "Beats" the Market 124-137. [Crossref]
150. Dorota Witkowska. Is the Three-Factor Better Than Single-Factor Capital Asset Pricing Model? Case of Polish Capital Market 225-238. [Crossref]
151. Susann Ihlau, Hendrik Duscha. Grundlagen der Unternehmensbewertung 17-115. [Crossref]
152. Manfred Gilli, Dietmar Maringer, Enrico Schumann. Optimization problems in finance 219-228. [Crossref]
153. . Bibliography 599-608. [Crossref]
154. Babak Mahdavi-Damghani. 2019. Data-Driven Models & Mathematical Finance: Opposition or Apposition?. *SSRN Electronic Journal* . [Crossref]
155. Sassan Zaker. 2019. How Wealth Management Lost Clients in Translation. *SSRN Electronic Journal* . [Crossref]
156. Alessandro Vercelli. The Emergence of Modern Financial Economics 61-91. [Crossref]
157. Jakob Thomä, Michael Hayne, Nikolaus Hagedorn, Clare Murray, Rebecca Grattage. 2019. The alignment of global equity and corporate bonds markets with the Paris Agreement. *Journal of Applied Accounting Research* **20**:4, 439. [Crossref]

158. Michael Hasler, Charles Martineau. 2019. Does the CAPM Predict Returns?. *SSRN Electronic Journal* . [Crossref]

159. Michael Hasler, Charles Martineau. 2019. The CAPM Holds. *SSRN Electronic Journal* . [Crossref]

160. Anh Phong Nguyen, Hoang Anh Nguyen, Thi Hong Minh Ho, Phu Thanh Ngo, David McMillan. 2019. Risk and returns of different foreign ownership portfolios: Evidence from Vietnam stock market. *Cogent Economics & Finance* **7**:1, 1589412. [Crossref]

161. Michael Thicke. 2018. Market epistemology. *Synthese* **195**:12, 5571-5594. [Crossref]

162. Sabri Boubaker, Taher Hamza, Javier Vidal-García. 2018. Financial distress and equity returns: A leverage-augmented three-factor model. *Research in International Business and Finance* **46**, 1-15. [Crossref]

163. Madhavi Latha Challa, Venkataramanaiah Malepati, Siva Nageswara Rao Kolusu. 2018. Forecasting risk using auto regressive integrated moving average approach: an evidence from S&P BSE Sensex. *Financial Innovation* **4**:1. . [Crossref]

164. Patrick O'Sullivan. 2018. The Capital Asset Pricing Model and the Efficient Markets Hypothesis: The Compelling Fairy Tale of Contemporary Financial Economics. *International Journal of Political Economy* **47**:3-4, 225-252. [Crossref]

165. Keshav Sahadev, Michael Ward, Chris Muller. 2018. The impact of reference-day risk on beta estimation and a proposed solution. *Investment Analysts Journal* **47**:4, 327-342. [Crossref]

166. JONGSEOK LIM. 2018. A Study on the Estimation of the Cost of Equity Capital and Performance for Listed Logistics Service Providers based on CAPM. *Korean Journal of Logistics* **26**:3, 17-32. [Crossref]

167. Fang Zhang, Hong Fang, Xu Wang. 2018. Impact of Carbon Prices on Corporate Value: The Case of China's Thermal Listed Enterprises. *Sustainability* **10**:9, 3328. [Crossref]

168. Marta Olivia Rovedder de Oliveira, Aline Armanini Stefanan, Mauri Leodir Lobler. 2018. Brand equity, risk and return in Latin America. *Journal of Product & Brand Management* **27**:5, 557-572. [Crossref]

169. Stelios Bekiros, Nikolaos Loukeris, Iordanis Eleftheriadis, Gazi Uddin. 2018. Revisiting the three factor model in light of circular behavioural simultaneities. *Review of Behavioral Finance* **10**:3, 210-230. [Crossref]

170. John L. Glascock, Ran Lu-Andrews. 2018. The Asymmetric Conditional Beta-Return Relations of REITs. *The Journal of Real Estate Finance and Economics* **57**:2, 231-245. [Crossref]

171. I. Ferguson. 2018. Discount rates for corporate forest valuations. *Australian Forestry* **81**:3, 142-147. [Crossref]

172. Alfonso A. Rojo Ramírez, Maria J. Martínez Romero. 2018. Required and obtained equity returns in privately held businesses: the impact of family nature—evidence before and after the global economic crisis. *Review of Managerial Science* **12**:3, 771-801. [Crossref]

173. Matevž Skočir, Igor Lončarski. 2018. Multi-factor asset pricing models: Factor construction choices and the revisit of pricing factors. *Journal of International Financial Markets, Institutions and Money* **55**, 65-80. [Crossref]

174. Serge Darolles, Christian Francq, Sébastien Laurent. 2018. Asymptotics of Cholesky GARCH models and time-varying conditional betas. *Journal of Econometrics* **204**:2, 223-247. [Crossref]

175. Tomer Shushi. 2018. Stein's lemma for truncated elliptical random vectors. *Statistics & Probability Letters* **137**, 297-303. [Crossref]

176. Masato Sasaki, Anas Laamrani, Mitsuo Yamashiro. 2018. An interactive genetic algorithm for portfolio optimization considering the decision maker's preference. *Journal of Information and Optimization Sciences* **39**:4, 989-1008. [Crossref]

177. Lucas Bretschger, Filippo Lechthaler. 2018. Stock performance and economic growth: lessons from the Japanese case. *Macroeconomics and Finance in Emerging Market Economies* **11**:2, 195-217. [Crossref]

178. Geeta Lakshmi. 2018. Gekko and black swans: Finance theory in UK undergraduate curricula. *Critical Perspectives on Accounting* **52**, 35-47. [Crossref]

179. Christophe Schinckus. 2018. Pataphysics of finance: An essay of visual epistemology. *Critical Perspectives on Accounting* **52**, 57-68. [Crossref]

180. C. Janse van Rensburg, J.D. Krige. 2018. Paying the High Price of Active Management: A New Look at Unit Trust Fees. *Studies in Economics and Econometrics* **42**:1, 23-40. [Crossref]

181. Milica Radović, Snežana Radukić, Vladimir Njegomir. 2018. The Application of the Markowitz's Model in Efficient Portfolio Forming on the Capital Market in the Republic of Serbia. *Economic Themes* **56**:1, 17-34. [Crossref]

182. Huu T. Huynh, Yi Su, Gunnar Lucko, Richard C. Thompson. Beta Index and Complexity in Schedule Performance Measurement 471-480. [Crossref]

183. Pablo Koch-Medina, Jan Wenzelburger. 2018. Equilibria in the CAPM with non-tradeable endowments. *Journal of Mathematical Economics* **75**, 93-107. [Crossref]

184. Jae Wook Song, Bonggyun Ko, Woojin Chang. 2018. Analyzing systemic risk using non-linear marginal expected shortfall and its minimum spanning tree. *Physica A: Statistical Mechanics and its Applications* **491**, 289-304. [Crossref]

185. Harold L. Vogel. Rationality Rules 219-269. [Crossref]

186. Toan Luu Duc Huynh, Sang Phu Nguyen, Duy Duong. Pricing Assets with Higher Co-moments and Value-at-Risk by Quantile Regression Approach: Evidence from Vietnam Stock Market 953-986. [Crossref]

187. Stefan Behringer. Kennzahlen im Konzerncontrolling 89-133. [Crossref]

188. Fengmin Xu, Meihua Wang, Yu-Hong Dai, Dachuan Xu. 2018. A sparse enhanced indexation model with chance and cardinality constraints. *Journal of Global Optimization* **70**:1, 5-25. [Crossref]

189. Massimo Guidolin, Manuela Pedio. Multivariate GARCH and Conditional Correlation Models 229-266. [Crossref]

190. Bruce D. McNevin, Joan Nix. 2018. The beta heuristic from a time/frequency perspective: A wavelet analysis of the market risk of sectors. *Economic Modelling* **68**, 570-585. [Crossref]

191. Muhammad Hafidz Anuwar, Maheran Mohd Jaffar. The standardized credit rating grade for Malaysian listed companies in Bursa Malaysia 020060. [Crossref]

192. Ross Kingwell, Quenten Thomas, David Feldman, Imma Farré, Brad Plunkett. 2018. Traditional farm expansion versus joint venture remote partnerships. *Australian Journal of Agricultural and Resource Economics* **62**:1, 21-44. [Crossref]

193. Wei Liu, James W. Kolari, Seppo Pynnonen. 2018. CAPM Beta Lives for Expected Returns. *SSRN Electronic Journal* . [Crossref]

194. Deng-Ta Chen. 2018. : CAPM (Security Prices in Mean-Variance Equilibrium: A Further Study on CAPM). *SSRN Electronic Journal* . [Crossref]

195. Jessie Y. Zhu, Wally Smieliauskas. 2018. Evidence on the Economic Consequences of Marriage Equality and LGBT Human Rights. *SSRN Electronic Journal* . [Crossref]

196. Alex R. Horenstein. 2018. Leverage and Performance Metrics in Asset Pricing. *SSRN Electronic Journal* . [Crossref]

197. Kesh Sahadev, Mike Ward, Chris Muller. 2018. A Volume-Weighted-Average-Price (VWAP) Method for Estimating Beta in the Context of Reference-Day Risk. *SSRN Electronic Journal* . [Crossref]

198. Malcolm McLelland. 2018. CAPM Failure in Private Equity Valuation and the Alternative APT Method. *SSRN Electronic Journal* . [Crossref]
199. Malcolm McLelland. 2018. International Cost of Capital Estimation: The No-Arbitrage Method. *SSRN Electronic Journal* . [Crossref]
200. Wei Liu, James W. Kolari, Jianhua Huang. 2018. Return Dispersion and the Cross-Section of Stock Returns. *SSRN Electronic Journal* . [Crossref]
201. Zhichuan Frank Li, Dylan Minor, Jun Wang, Chong Yu. 2018. A Learning Curve of the Market: Chasing Alpha of Socially Responsible Firms. *SSRN Electronic Journal* . [Crossref]
202. Natalia Bailey, George Kapetanios, M. Hashem Pesaran. 2018. Exponent of Cross-Sectional Dependence for Residuals. *SSRN Electronic Journal* . [Crossref]
203. Weihao Han. 2018. Persistence and Performance of Chinese Mutual Funds. *SSRN Electronic Journal* . [Crossref]
204. Lisa R. Goldberg, Alex Papanicolaou, Alexander Shkolnik. 2018. Better Betas. *SSRN Electronic Journal* . [Crossref]
205. Terrence Hendershott, Dmitry Livdan, Dominik RRsch. 2018. Asset Pricing: A Tale of Night and Day. *SSRN Electronic Journal* . [Crossref]
206. Ben Matthies. 2018. Biases in the Perception of Covariance. *SSRN Electronic Journal* . [Crossref]
207. Fred Sporta. 2018. The Distressing Effect of Non-Performing Assets to Asset Quality for Commercial Banks in Kenya. *INTERNATIONAL JOURNAL OF INNOVATION AND ECONOMIC DEVELOPMENT* **3**:6, 71-83. [Crossref]
208. Gonul Colak, Dimitrios Gounopoulos, Panagiotis Loukopoulos, Georgios Loukopoulos. 2018. Local Policy Risk and IPO Performance. *SSRN Electronic Journal* . [Crossref]
209. Todd D. Gerarden, Richard G. Newell, Robert N. Stavins. 2017. Assessing the Energy-Efficiency Gap. *Journal of Economic Literature* **55**:4, 1486-1525. [Abstract] [View PDF article] [PDF with links]
210. Serdar Neslihanoglu, Vasilios Sogiakas, John H. McColl, Duncan Lee. 2017. Nonlinearities in the CAPM: Evidence from Developed and Emerging Markets. *Journal of Forecasting* **36**:8, 867-897. [Crossref]
211. Bogdan Negrea, Mihai Toma. 2017. Dynamic CAPM under ambiguity—An experimental approach. *Journal of Behavioral and Experimental Finance* **16**, 22-32. [Crossref]
212. José María Díez-Esteban, Conrado Diego García-Gómez, Félix Javier López-Iturriaga, Marcos Santamaría-Mariscal. 2017. Corporate risk-taking, returns and the nature of major shareholders: Evidence from prospect theory. *Research in International Business and Finance* **42**, 900-911. [Crossref]
213. Yinqiao Li, Renée Spigt, Laurens Swinkels. 2017. The impact of FinTech start-ups on incumbent retail banks' share prices. *Financial Innovation* **3**:1. . [Crossref]
214. Diego Broz, Gastón Milanesi, Daniel Alejandro Rossit, Diego Gabriel Rossit, Fernando Tohmé. 2017. Forest management decision making based on a real options approach: An application to a case in northeastern Argentina. *Forestry Studies* **67**:1, 97-108. [Crossref]
215. Hung-Chi Li, Syouching Lai, James A. Conover, Frederick Wu, Bin Li. Stock Returns and Financial Distress Risk: Evidence from the Asian-Pacific Markets 123-158. [Crossref]
216. Marc K. Chan, Simon Kwok. 2017. Risk-sharing, market imperfections, asset prices: Evidence from China's stock market liberalization. *Journal of Banking & Finance* **84**, 166-187. [Crossref]
217. Turan G. Bali, Robert F. Engle, Yi Tang. 2017. Dynamic Conditional Beta Is Alive and Well in the Cross Section of Daily Stock Returns. *Management Science* **63**:11, 3760-3779. [Crossref]

218. Márcio André Veras Machado, Robert Faff, Suelle Cariele de Souza e Silva. 2017. Applicability of Investment and Profitability Effects in Asset Pricing Models. *Revista de Administração Contemporânea* **21**:6, 851-874. [Crossref]
219. David G. Carmichael. 2017. Adjustments within discount rates to cater for uncertainty—Guidelines. *The Engineering Economist* **62**:4, 322-335. [Crossref]
220. Gastón Silverio Milanesi. 2017. Valuación de empresas: enfoque integral para mercados emergentes e inflacionarios. *Estudios Gerenciales* **33**:145, 377-390. [Crossref]
221. Júlio Lobão, Cristiano Pereira. 2017. Psychological barriers in stock market indices: Evidence from four southern European countries. *Cuadernos de Economía* **40**:114, 268-278. [Crossref]
222. Matheus Duarte Valente Vieira, Vinicius Mothé Maia, Marcelo Cabús Klotzle, Antonio Carlos Figueiredo. 2017. Modelo de Cinco Fatores de Risco: precificando carteiras setoriais no mercado acionário brasileiro. *Revista Catarinense da Ciência Contábil* **16**:48. . [Crossref]
223. Ikrame Ben Slimane, Makram Bellalah, Hatem Rjiba. 2017. Time-varying beta during the 2008 financial crisis – evidence from North America and Western Europe. *The Journal of Risk Finance* **18**:4, 398-431. [Crossref]
224. Karen Paul. 2017. The effect of business cycle, market return and momentum on financial performance of socially responsible investing mutual funds. *Social Responsibility Journal* **13**:3, 513-528. [Crossref]
225. Sara Aghakhani, Reda Alhajj, Jon Rokne, Philip Chang. An Effective LmRMR for Financial Variable Selection and its Applications 535-543. [Crossref]
226. Daniel Botero Guzman, Jhon Alexis Díaz Contreras. 2017. Análisis de la relación rentabilidad-riesgo en el mercado accionario internacional para un mundo parcialmente integrado. *Ensayos de Economía* **27**:51, 109-124. [Crossref]
227. Jill A. Brown, Anne Anderson, Jesus M. Salas, Andrew J. Ward. 2017. Do Investors Care About Director Tenure? Insights from Executive Cognition and Social Capital Theories. *Organization Science* **28**:3, 471-494. [Crossref]
228. Abel Rodríguez, Ziwei Wang, Athanasios Kottas. 2017. Assessing systematic risk in the S&P500 index between 2000 and 2011: A Bayesian nonparametric approach. *The Annals of Applied Statistics* **11**:2. . [Crossref]
229. Chandana Shahi, Sherrill Shaffer. 2017. CAPM and the changing distribution of historical returns. *Applied Economics Letters* **24**:9, 639-642. [Crossref]
230. Kwanglim Seo, Ellen Eun Kyoo Kim, Amit Sharma. 2017. Examining the determinants of long-term debt in the US restaurant industry. *International Journal of Contemporary Hospitality Management* **29**:5, 1501-1520. [Crossref]
231. Kevin E. Dow, Marcia Weidenmier Watson, Vincent J. Shea. 2017. Riding the waves of technology through the decades: The relation between industry-level information technology intensity and the cost of equity capital. *International Journal of Accounting Information Systems* **25**, 18-28. [Crossref]
232. Xiao-Ping Steven Zhang, Fang Wang. 2017. Signal Processing for Finance, Economics, and Marketing: Concepts, framework, and big data applications. *IEEE Signal Processing Magazine* **34**:3, 14-35. [Crossref]
233. Federico Etro, Elena Stepanova. 2017. Art Auctions and Art Investment in the Golden Age of British Painting. *Scottish Journal of Political Economy* **64**:2, 191-225. [Crossref]
234. Joseph Nyangon, John Byrne, Job Taminiau. 2017. An assessment of price convergence between natural gas and solar photovoltaic in the U.S. electricity market. *WIREs Energy and Environment* **6**:3. . [Crossref]

235. Devi Lusyana, Mohamed Sherif. 2017. Shariah -compliant investments and stock returns: evidence from the Indonesian stock market. *Journal of Islamic Accounting and Business Research* **8**:2, 143-160. [Crossref]
236. Vance Lesseig, Janet D. Payne. 2017. The precision of asset beta estimates. *International Journal of Managerial Finance* **13**:2, 213-224. [Crossref]
237. Guillaume Coqueret. 2017. Empirical properties of a heterogeneous agent model in large dimensions. *Journal of Economic Dynamics and Control* **77**, 180-201. [Crossref]
238. Zarah Puspitaningtyas. 2017. Estimating systematic risk for the best investment decisions on manufacturing company in Indonesia. *Investment Management and Financial Innovations* **14**:1, 46-54. [Crossref]
239. Vijay Gondhalekar, Kevin Lehnert. Financial Performance and the Competitive Effects of Corporate Social Responsibility 1-21. [Crossref]
240. Danielle Claire Sanderson, Steven Devaney. 2017. Occupier satisfaction and its impact on investment returns from UK commercial real estate. *Journal of Property Investment & Finance* **35**:2, 135-159. [Crossref]
241. Ching-Ping Wang, Hung-Hsi Huang, Jin-Sheng Hu. 2017. Reverse-Engineering and Real Options–Adjusted CAPM in the Taiwan Stock Market. *Emerging Markets Finance and Trade* **53**:3, 670-687. [Crossref]
242. Supriya Maheshwari, Raj Singh Dhankar. 2017. Momentum anomaly: evidence from the Indian stock market. *Journal of Advances in Management Research* **14**:1, 3-22. [Crossref]
243. Bin Mei. 2017. Investment returns of US commercial timberland: insights into index construction methods and results. *Canadian Journal of Forest Research* **47**:2, 226-233. [Crossref]
244. Shaista Arshad. Investigating the Integration 85-118. [Crossref]
245. Raquel J. Fonseca. Capital Asset Pricing Model—A Structured Robust Approach 385-392. [Crossref]
246. Efundem Agboraw, Aled Jones. Finance and Natural Resource Constraints 41-91. [Crossref]
247. Imad A. Moosa, Vikash Ramiah. The Rise and Fall of Neoclassical Finance 1-25. [Crossref]
248. Mª José Martínez Romero, Alfonso A. Rojo Ramírez. 2017. Socioemotional wealth's implications in the calculus of the minimum rate of return required by family businesses' owners. *Review of Managerial Science* **11**:1, 95-118. [Crossref]
249. . Bibliography 315-328. [Crossref]
250. Numan Ülkü. 2017. Monday effect in Fama–French's RMW factor. *Economics Letters* **150**, 44-47. [Crossref]
251. Richard C. Thompson, Yi Su, Gunnar Lucko. 2017. Measuring Project Performance Inspired by Stock Index. *Procedia Engineering* **196**, 706-713. [Crossref]
252. Ricardo G. Barcelona. Doing the Managerial Flexibility Maths 455-472. [Crossref]
253. Arun Muralidhar, Robert Savickas, Tzu-Jui Mao. 2017. Asset Pricing Anomalies: Two Hedge Factors with Negative Risk Premia Embedded in Portfolios!. *SSRN Electronic Journal* . [Crossref]
254. Pablo Fernandez. 2017. Finanzas y Economma Financiera (Finance and Financial Economics). *SSRN Electronic Journal* . [Crossref]
255. Pablo Fernandez. 2017. Finance and Financial Economics: A Debate About Common Sense and Illogical Models. *SSRN Electronic Journal* . [Crossref]
256. Kulab Jamar. 2017. Property Funds and REITs in Thailand: A CAPM Investigation. *SSRN Electronic Journal* . [Crossref]
257. Ming Li Chew, Sahil Puri. 2017. Using Natural Language Processing Techniques for Stock Return Predictions. *SSRN Electronic Journal* . [Crossref]

258. Gevorg Hunanyan. 2017. The Systematic Risk of Short-Sale Restrictions. *SSRN Electronic Journal* . [Crossref]
259. Mike Ward, Chris Muller, Pravin Semnarayan. 2017. Negative Investment Returns in a Developing Market Context. *SSRN Electronic Journal* . [Crossref]
260. Eben Otuteye. 2017. Re-Evaluating the Value of Modern Portfolio Theory and Asset Pricing Models Based on Behavioral Insights from Benjamin Graham's Value Investing Paradigm. *SSRN Electronic Journal* . [Crossref]
261. Min Deng. 2017. Death of the Capital Asset Pricing Model. *SSRN Electronic Journal* . [Crossref]
262. Eben Otuteye, Mohammad Siddiquee. 2017. A Critique of Modern Portfolio Theory and Asset Pricing Models Based on Behavioral Insights from Benjamin Graham's Value Investing Paradigm. *SSRN Electronic Journal* . [Crossref]
263. Jose Maria Diez Esteban, Conrado DiegoGarcca-GGmez, FFlix J. LLpez-Iturriaga, Marcos Santamarra-Mariscal. 2017. Corporate Risk-Taking, Returns and the Nature of Major Shareholders: Evidence from Prospect Theory. *SSRN Electronic Journal* . [Crossref]
264. Samuel Johnson, David Tuckett. 2017. Narrative Decision-Making in Investment Choices: How Investors Use News About Company Performance. *SSRN Electronic Journal* . [Crossref]
265. Johnny Kang, Tom B. Parker, Scott Radell, Ralph Smith. 2017. Reach for Safety. *SSRN Electronic Journal* . [Crossref]
266. Eben Otuteye, Mohammad Siddiquee. 2017. Buffett's Alpha: Further Explanations from Behavioral Value Investing Perspective. *SSRN Electronic Journal* . [Crossref]
267. Imran Riaz Malik, Attaullah Shah. 2017. Single Stock Futures and Their Impact on Risk Characteristics of the Underlying Stocks: A Dynamic CAPM Approach. *SSRN Electronic Journal* . [Crossref]
268. James Ming Chen. 2017. Baryonic Beta Dynamics: An Econophysical Model of Systematic Risk. *SSRN Electronic Journal* . [Crossref]
269. Yang Song. 2017. The Mismatch between Mutual Fund Scale and Skill. *SSRN Electronic Journal* . [Crossref]
270. Antonio Amendola, Dennis Marco Montagna, Mario Maggi. 2017. Analysis of Equity β Principal Components in Low Risk Framework: New Results and Prospectives. *SSRN Electronic Journal* . [Crossref]
271. Pieter Klaassen, Idzard van Eeghen. 2017. Bank Capital Allocation and Performance Management under Multiple Capital Constraints. *SSRN Electronic Journal* . [Crossref]
272. Jessie Y. Zhu, Wally Smieliauskas. 2017. Evidence on the Economic Consequences of Marriage Equality and LGBT Human Rights. *SSRN Electronic Journal* . [Crossref]
273. Gastón Silverio Milanesi. 2017. DESCUENTO DE FLUJO DE FONDOS E INFLACIÓN PARA LA VALORACIÓN DE EMPRESAS EN DOS MONEDAS. *Semestre Económico* **20**:44, 189-218. [Crossref]
274. Emon Kalyan Chowdhury. 2017. Functioning of Fama-French Three-Factor Model in Emerging Stock Markets: An Empirical Study on Chittagong Stock Exchange, Bangladesh. *Journal of Financial Risk Management* **06**:04, 352-363. [Crossref]
275. José Carlos de Souza Santos, Elias Cavalcante Filho. 2017. Investing on the CAPM Pricing Error. *Technology and Investment* **08**:01, 67-82. [Crossref]
276. Wei Liu, James W. Kolari. 2017. Do Multi-Factors Proxy More Efficient Market Indexes?. *SSRN Electronic Journal* . [Crossref]

277. Gaëlle Lenormand, Lionel Touchais. 2017. L'impact de l'IFRS 8 sur le coût des capitaux propres. *Management & Avenir* **92**:2, 133. [Crossref]
278. Prashant Das, Gabrielle Bodenmann. Minimizing the Cost of Capital in Hotel Investments 191-207. [Crossref]
279. Arun Muralidhar. 2017. A Very Simple Goals - and Risk-Based Asset Pricing Model (or Asset Pricing with Heterogeneous Investors). *SSRN Electronic Journal* . [Crossref]
280. Steven R. Ratner. 2017. Compensation for Expropriations in a World of Investment Treaties: Beyond the Lawful/Unlawful Distinction. *American Journal of International Law* **111**:1, 7-56. [Crossref]
281. Yehuda (Yud) Izhakian. 2017. Knight Meets Sharpe: Capital Asset Pricing under Ambiguity. *SSRN Electronic Journal* . [Crossref]
282. Daniel Andrei, Julien Cujean, Mungo Ivor Wilson. 2017. The Lost Capital Asset Pricing Model. *SSRN Electronic Journal* . [Crossref]
283. Seung C. Ahn, Alex R. Horenstein. 2017. Asset Pricing and Excess Returns Over the Market Return. *SSRN Electronic Journal* . [Crossref]
284. David J. Moore. 2016. A look at the actual cost of capital of US firms. *Cogent Economics & Finance* **4**:1, 1233628. [Crossref]
285. Igor Deplano, Giovanni Squillero, Alberto Tonda. 2016. Anatomy of a portfolio optimizer under a limited budget constraint. *Evolutionary Intelligence* **9**:4, 125-136. [Crossref]
286. Eric Le Fur, Hachmi Ben Ameur, Benoit Faye. 2016. Time-Varying Risk Premiums in the Framework of Wine Investment. *Journal of Wine Economics* **11**:3, 355-378. [Crossref]
287. Mareike Hornung, Robert Luther, Peter Schuster. 2016. Retrievability bias in explaining the hurdle rate premium puzzle. *Journal of Applied Accounting Research* **17**:4, 440-455. [Crossref]
288. Remmer Sassen, Anne-Kathrin Hinze, Inga Hardeck. 2016. Impact of ESG factors on firm risk in Europe. *Journal of Business Economics* **86**:8, 867-904. [Crossref]
289. Marie-Claude Beaulieu, Marie-Hélène Gagnon, Lynda Khalaf. 2016. Less is more: Testing financial integration using identification-robust asset pricing models. *Journal of International Financial Markets, Institutions and Money* **45**, 171-190. [Crossref]
290. Te-Feng Chen, San-Lin Chung, Wei-Che Tsai. 2016. Option-Implied Equity Risk and the Cross Section of Stock Returns. *Financial Analysts Journal* **72**:6, 42-55. [Crossref]
291. W. D. Chen, H. C. Li. 2016. Wavelet decomposition of heterogeneous investment horizon. *Journal of Economics and Finance* **40**:4, 714-734. [Crossref]
292. Brad M. Barber, Xing Huang, Terrance Odean. 2016. Which Factors Matter to Investors? Evidence from Mutual Fund Flows. *Review of Financial Studies* **29**:10, 2600-2642. [Crossref]
293. Nicholas G. Hall. Research and Teaching Opportunities in Project Management 329-388. [Crossref]
294. Thomas Lindner, Jakob Muellner, Jonas Puck. 2016. Cost of Capital in an International Context: Institutional Distance, Quality, and Dynamics. *Journal of International Management* **22**:3, 234-248. [Crossref]
295. Richard J. Agnello. 2016. Do U.S. paintings follow the CAPM? Findings disaggregated by subject, artist, and value of the work. *Research in Economics* **70**:3, 403-411. [Crossref]
296. Celine Gimet, Sandra Montchaud. 2016. What Drives European Football Clubs' Stock Returns and Volatility?. *International Journal of the Economics of Business* **23**:3, 351-390. [Crossref]
297. Robert F. Engle. 2016. Dynamic Conditional Beta. *Journal of Financial Econometrics* **14**:4, 643-667. [Crossref]
298. Rajesh H Acharya, Anver C Sadath. 2016. On the interaction between energy price and firm size in Indian economy. *OPEC Energy Review* **40**:3, 300-315. [Crossref]

299. Claes Fornell, Forrest V. Morgeson, G. Tomas M. Hult. 2016. Stock Returns on Customer Satisfaction Do Beat the Market: Gauging the Effect of a Marketing Intangible. *Journal of Marketing* **80**:5, 92-107. [Crossref]
300. Paul Windolf. 2016. Riding the Bubble: Financial Market Crises in Twenty-Two OECD Countries. *Journal of Economic Issues* **50**:3, 788-813. [Crossref]
301. Won Kang, Jungsoon Shin. 2016. Derivation of Corporate Debt Pricing Model and Its Empirical Implications. *Asia-Pacific Journal of Financial Studies* **45**:3, 439-462. [Crossref]
302. SWARN CHATTERJEE, AMY HUBBLE. 2016. DAY-OF-THE-WEEK EFFECT IN US BIOTECHNOLOGY STOCKS — DO POLICY CHANGES AND ECONOMIC CYCLES MATTER?. *Annals of Financial Economics* **11**:02, 1650008. [Crossref]
303. Syed Jawad Hussain Shahzad, Saniya Khalid, Saba Ameer. 2016. CAPM estimates: Can data frequency and time period lend a hand?. *International Journal of Financial Engineering* **03**:02, 1650018. [Crossref]
304. ##, ###. 2016. Why Do Some Asset Pricing Models Perform Poorly? Evidence from Irrationality, Transaction Costs, and Missing Factors. *Seoul Journal of Business* **22**:1, 1-64. [Crossref]
305. Mohammad Iftekhar Khan, Amit Banerji. 2016. Corporate Governance and Foreign Investment in India. *Indian Journal of Corporate Governance* **9**:1, 19-43. [Crossref]
306. Robert East. 2016. Bias in the evaluation of research methods. *Marketing Theory* **16**:2, 219-231. [Crossref]
307. Carmine De Franco, Bruno Monnier, Ksenya Rulik. 2016. Factor Exposure of Alternative Beta Strategiesacross Market Regimes. *The Journal of Index Investing* **7**:1, 78-91. [Crossref]
308. . Index 331-338. [Crossref]
309. . References 319-322. [Crossref]
310. Amabile Millani Rebeschini, Ricardo P. C. Leal. 2016. Stock Fund Returns and Macroeconomic Variables in Brazil. *Latin American Business Review* **17**:2, 139-161. [Crossref]
311. Christian Fieberg, Armin Varmaz, Thorsten Poddig. 2016. Covariances vs. characteristics: what does explain the cross section of the German stock market returns?. *Business Research* **9**:1, 27-50. [Crossref]
312. Mathijs Cosemans, Rik Frehen, Peter C. Schotman, Rob Bauer. 2016. Estimating Security Betas Using Prior Information Based on Firm Fundamentals. *Review of Financial Studies* **29**:4, 1072-1112. [Crossref]
313. Iman Mirzadeh, Bjorn Birgisson. 2016. Evaluation of Highway Projects under Government Support Mechanisms Based on an Option-Pricing Framework. *Journal of Construction Engineering and Management* **142**:4, 04015094. [Crossref]
314. . References 549-561. [Crossref]
315. Dawood Ashraf. 2016. Does Shari'ah Screening Cause Abnormal Returns? Empirical Evidence from Islamic Equity Indices. *Journal of Business Ethics* **134**:2, 209-228. [Crossref]
316. Marshall L. Stocker. 2016. The price of freedom: A Fama–French freedom factor. *Emerging Markets Review* **26**, 1-19. [Crossref]
317. Ton van den Bremer, Frederick van der Ploeg, Samuel Wills. 2016. The Elephant In The Ground: Managing Oil And Sovereign Wealth. *European Economic Review* **82**, 113-131. [Crossref]
318. Ran Leshem, Lisa R. Goldberg, Alan Cummings. 2016. Optimizing Value. *The Journal of Portfolio Management* **42**:2, 77-89. [Crossref]
319. Ran Leshem, Lisa R Goldberg, Alan Cummings. 2016. Optimizing Value. *The Journal of Portfolio Management* . [Crossref]
320. Pertti Lahdenperä. 2016. Preparing a framework for two-stage target-cost arrangement formulation. *International Journal of Managing Projects in Business* **9**:1, 123-146. [Crossref]

321. Nikodem Szumilo, Pascal Gantenbein, Werner Gleißner, Thomas Wiegelmann. 2016. Predicting uncertainty: the impact of risk measurement on value of real estate portfolios. *Journal of Property Research* **33**:1, 1-17. [Crossref]
322. Jean-Luc Besson, Michel M. Dacorogna, Paolo de Martin, Michael Kastenholz, Michael Moller. How Much Capital Does a Reinsurance Need? 235-253. [Crossref]
323. Arlie O. Petters, Xiaoying Dong. Capital Market Theory and Portfolio Risk Measures 151-208. [Crossref]
324. Julius Hemminki, Vesa Puttonen. Fundamental Indexation in Europe 323-330. [Crossref]
325. Les Coleman. Current Paradigm: Neoclassical Investment Theory 15-28. [Crossref]
326. Mihály Ormos, Dusán Timotity. 2016. Generalized asset pricing: Expected Downside Risk-based equilibrium modeling. *Economic Modelling* **52**, 967-980. [Crossref]
327. Peter C. Dawson. 2016. Is Opportunity Cost Synonymous with Cost of Capital and Required Rate of Return?: Untangling the Present Value Discount Rate. *SSRN Electronic Journal* . [Crossref]
328. Somayeh Moazeni. 2016. Risk-Averse Dynamic Arbitrage in Illiquid Markets. *SSRN Electronic Journal* . [Crossref]
329. John L. Glascock. 2016. The Asymmetric Conditional Beta-Return Relations of REITs. *SSRN Electronic Journal* . [Crossref]
330. Gregory Gadzinski, Markus Schuller, Andrea Vacchino. 2016. The Global Capital Stock. A Proxy for the Unobservable Global Market Portfolio. *SSRN Electronic Journal* . [Crossref]
331. James Ming Chen. 2016. Baryonic Beta Dynamics: Splitting the Atom of Systematic Risk. *SSRN Electronic Journal* . [Crossref]
332. Paul Schneider, Christian Wagner, Josef Zechner. 2016. Low Risk Anomalies?. *SSRN Electronic Journal* . [Crossref]
333. Stefano Dova. 2016. From the CAPM to Fama-French: A Road Paved with Leverage. *SSRN Electronic Journal* . [Crossref]
334. Nadiyah Amatul Haq, Alvina Syafira Fauzia, Anggun Puspita Khoirunnisa. 2016. Promoting Sustainable Financial System in Indonesia Towards SRI-KEHATI Index. *SSRN Electronic Journal* . [Crossref]
335. Sabin Bikram Panta, Niranjan Phuyal, Rajesh Sharma, Gautam Vora. 2016. The Cross-Section of Stock Returns: An Application of Fama-French Approach to Nepal. *Modern Economy* **07**:02, 223-231. [Crossref]
336. Nordine Abidi, Burcu Hacibedel, Mwanza Nkusu. 2016. Changing Times for Frontier Markets: A Perspective from Portfolio Investment Flows and Financial Integration. *IMF Working Papers* **16**:177, 1. [Crossref]
337. Elena N. Asparouhova, Peter Bossaerts. 2016. Competitive Off-Equilibrium: Theory and Experiment. *SSRN Electronic Journal* . [Crossref]
338. Iman MIRZADEH, Bjorn BIRGISSON. 2015. ACCOMMODATING ENERGY PRICE VOLATILITY IN LIFE CYCLE COST ANALYSIS OF ASPHALT PAVEMENTS. *JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT* **22**:8, 1001-1008. [Crossref]
339. David Geltner. 2015. Real Estate Price Indices and Price Dynamics: An Overview from an Investments Perspective. *Annual Review of Financial Economics* **7**:1, 615-633. [Crossref]
340. F. Echterling, B. Eierle, S. Ketterer. 2015. A review of the literature on methods of computing the implied cost of capital. *International Review of Financial Analysis* **42**, 235-252. [Crossref]
341. G. van Vuuren, R. Yacumakis. 2015. Hedge Fund Performance Evaluation Using the Kalman Filter. *Studies in Economics and Econometrics* **39**:3, 1-24. [Crossref]

342. Aris Spanos, Deborah G. Mayo. 2015. Error statistical modeling and inference: Where methodology meets ontology. *Synthese* **192**:11, 3533-3555. [Crossref]
343. Ozan Kocadağlı, Rıdvan Keskin. 2015. A novel portfolio selection model based on fuzzy goal programming with different importance and priorities. *Expert Systems with Applications* **42**:20, 6898-6912. [Crossref]
344. Gustavo A. Marrero, Luis A. Puch, Francisco J. Ramos-Real. 2015. Mean-variance portfolio methods for energy policy risk management. *International Review of Economics & Finance* **40**, 246-264. [Crossref]
345. Fabian T. Lutzenberger. 2015. Multifactor Models and their Consistency with the ICAPM: Evidence from the European Stock Market. *European Financial Management* **21**:5, 1014-1052. [Crossref]
346. C. A. Valle, N. Meade, J. E. Beasley. 2015. Factor neutral portfolios. *OR Spectrum* **37**:4, 843-867. [Crossref]
347. F. Echterling, B. Eierle. 2015. Mean reversion adjusted betas used in business valuation practice: a research note. *Journal of Business Economics* **85**:7, 759-792. [Crossref]
348. Marcelo Sandoval, Santiago Grijalva. Future grid business model innovation: A prosumer-based cost-benefit framework for valuation of Distributed Energy Resources 450-455. [Crossref]
349. . Bibliography 273-276. [Crossref]
350. . Bibliography 443-445. [Crossref]
351. Nazeeruddin Mohammad, Dawood Ashraf. 2015. The market timing ability and return performance of Islamic equities: An empirical study. *Pacific-Basin Finance Journal* **34**, 169-183. [Crossref]
352. Yasser Alhenawi. 2015. On the interaction between momentum effect and size effect. *Review of Financial Economics* **26**, 36-46. [Crossref]
353. Martin Širůček, Lukáš Křen. 2015. Application of Markowitz Portfolio Theory by Building Optimal Portfolio on the US Stock Market. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **63**:4, 1375-1386. [Crossref]
354. Harry S. Marmer. 2015. Fire! Fire! Is U.S. Low Volatility a Crowded Trade?. *The Journal of Investing* **24**:3, 17-37. [Crossref]
355. . References 321-329. [Crossref]
356. Yannick Malevergne, Didier Sornette. Multi-moment Method for Portfolio Management: Generalised Capital Asset Pricing Model in Homogeneous and Heterogeneous Markets 166-193. [Crossref]
357. K. J. Barnard, M. B. Bunting. 2015. Value and size investment strategies during the global financial crisis: evidence from the South African equity market. *South African Journal of Accounting Research* **29**:2, 177-196. [Crossref]
358. Richard A. Michelfelder. 2015. Empirical analysis of the generalized consumption asset pricing model: Estimating the cost of capital. *Journal of Economics and Business* **80**, 37-50. [Crossref]
359. Ruggiero Cavallo, R. Preston Mcafee, Sergei Vassilvitskii. 2015. Display Advertising Auctions with Arbitrage. *ACM Transactions on Economics and Computation* **3**:3, 1-23. [Crossref]
360. Seungho Baek, John F.O. Bilson. 2015. Size and value risk in financial firms. *Journal of Banking & Finance* **55**, 295-326. [Crossref]
361. Yuntaek Pae, Navid Sabbaghi. 2015. Equally weighted portfolios vs value weighted portfolios: Reasons for differing betas. *Journal of Financial Stability* **18**, 203-207. [Crossref]
362. Shweta Bajpai, Anil K. Sharma. 2015. An Empirical Testing of Capital Asset Pricing Model in India. *Procedia - Social and Behavioral Sciences* **189**, 259-265. [Crossref]

363. Roger Buckland, Julian Williams, Janice Beecher. 2015. Risk and regulation in water utilities: a cross-country comparison of evidence from the CAPM. *Journal of Regulatory Economics* **47**:2, 117-145. [Crossref]

364. Lin Chen, Lu Qin, Hongquan Zhu. 2015. Opinion divergence, unexpected trading volume and stock returns: Evidence from China. *International Review of Economics & Finance* **36**, 119-127. [Crossref]

365. Philippe Bertrand, Vincent Lapointe. 2015. How performance of risk-based strategies is modified by socially responsible investment universe?. *International Review of Financial Analysis* **38**, 175-190. [Crossref]

366. Hersh Shefrin. 2015. Investors' Judgments, Asset Pricing Factors and Sentiment. *European Financial Management* **21**:2, 205-227. [Crossref]

367. Jan Bastin. 2015. Volatility Effect: An Application on the German Stock Market. *Český finanční a účetní časopis* **2015**:1, 36-54. [Crossref]

368. Peter C. Dawson. 2015. The capital asset pricing model in economic perspective. *Applied Economics* **47**:6, 569-598. [Crossref]

369. Jan Doppegieter. The Capital Asset Pricing Model 1-4. [Crossref]

370. Virginie Coudert, Karine Hervé, Pierre Mabille. 2015. Internationalization Versus Regionalization in the Emerging Stock Markets. *International Journal of Finance & Economics* **20**:1, 16-27. [Crossref]

371. Dennis Schlegel. Background: Cost-of-Capital in the Finance Literature 9-70. [Crossref]

372. Marcus Schulmerich, Yves-Michel Leporcher, Ching-Hwa Eu. Modern Portfolio Theory and Its Problems 101-173. [Crossref]

373. Cherry Muijsson, Ed Fishwick, Steve Satchell. The Low Beta Anomaly and Interest Rates 305-328. [Crossref]

374. Maurício Vasconcellos Leão Lyrio, Wlademir Prates, Marcus Vinícius Andrade de Lima, Rogério João Lunkes. 2015. Análise da implementação de uma estratégia de investimento em ações baseada em um instrumento de apoio à decisão. *Contaduría y Administración* **60**:1, 113-143. [Crossref]

375. Usman Ayub, Syed Zulfiqar Ali Shah, Qaisar Abbas. 2015. Robust analysis for downside risk in portfolio management for a volatile stock market. *Economic Modelling* **44**, 86-96. [Crossref]

376. Josanco Floreani, Maurizio Polato, Andrea Paltrinieri, Flavio Pichler. Credit Quality, Bank Provisioning and Systematic Risk in Banking Business 1-34. [Crossref]

377. Federico Beltrame, Daniele Previtali, Luca Grassetti. The Estimation of Banks' Cost of Capital through the Capital at Risk Model: An Empirical Investigation across European Banks 35-65. [Crossref]

378. Nuraini Ismail, Mohd Tahir Ismail, Samsul Ariffin Abdul Karim, Firdaus Mohamad Hamzah. Modeling and forecasting the volatility of Islamic unit trust in Malaysia using GARCH model 050009. [Crossref]

379. Sven Carlin. 2015. A Real Value Risk Estimation Model for an Emerging Market. *SSRN Electronic Journal* . [Crossref]

380. Todd Gerarden, Richard G. Newell, Robert N. Stavins. 2015. Assessing the Energy-Efficiency Gap. *SSRN Electronic Journal* . [Crossref]

381. Christian-Oliver Ewald, Pariya Salehi. 2015. Salmon Futures and the Fish Pool Market in the Context of the CAPM and the Fama & French Three-Factor Model. *SSRN Electronic Journal* . [Crossref]

382. John Griffin. 2015. Risk Premia and Knightian Uncertainty in an Experimental Market Featuring a Long-Lived Asset. *SSRN Electronic Journal* . [Crossref]

383. Danilo Tilloca. 2015. A Risk-Neutral Approach for the Evaluation of Commercial Loans. *SSRN Electronic Journal* . [Crossref]

384. Ran Leshem, Lisa R. Goldberg, Alan Cummings. 2015. Optimizing Value. *SSRN Electronic Journal* . [Crossref]
385. Daniel Felix Ahelegbey. 2015. The Econometrics of Networks: A Review. *SSRN Electronic Journal* . [Crossref]
386. Piotr Wisniewski. 2015. The Valuation of Social Media Public Companies: There Is a Method to This Madness!. *SSRN Electronic Journal* . [Crossref]
387. MMrio Correia Fernandes. 2015. An Application of Risk-Return Metrics on Public-Private Partnerships: The Case of Portuguese Road Sector. *SSRN Electronic Journal* . [Crossref]
388. Marc K. Chan, Simon Kwok. 2015. The Effect of Risk Sharing on Asset Prices: Natural Experiment from the Chinese Stock Market Liberalization. *SSRN Electronic Journal* . [Crossref]
389. Christoph Becker, Wolfgang M. Schmidt. 2015. Value, Size, Momentum and the Average Correlation of Stock Returns. *SSRN Electronic Journal* . [Crossref]
390. Brandes Institute. 2015. Redefining Risk and Return in Common Stock Investment. *SSRN Electronic Journal* . [Crossref]
391. Pablo Fernandez. 2015. CAPM: An Absurd Model. *Business Valuation Review* **34**:1, 4-23. [Crossref]
392. Marie-Claude Beaulieu, Jean-Marie Dufour, Lynda Khalaf. 2015. Identification-Robust Factor Pricing: Canadian Evidence. *L'Actualité économique* **91**:1-2, 235-252. [Crossref]
393. Michael Olbrich, Tobias Quill, David J. Rapp. 2015. Business Valuation Inspired by the Austrian School. *Journal of Business Valuation and Economic Loss Analysis* **10**:1, 1-43. [Crossref]
394. Petros Messis, Achilleas Zapranis. 2014. Asset pricing with time-varying betas for stocks traded on S&P 500. *Applied Economics* **46**:36, 4508-4518. [Crossref]
395. Myuran Rajaratnam, Bala Rajaratnam, Kanshukan Rajaratnam. 2014. A novel equity valuation and capital allocation model for use by long-term value-investors. *Journal of Banking & Finance* **49**, 483-494. [Crossref]
396. Petros Messis, Achilleas Zapranis. 2014. Herding towards higher moment CAPM, contagion of herding and macroeconomic shocks: Evidence from five major developed markets. *Journal of Behavioral and Experimental Finance* **4**, 1-13. [Crossref]
397. Claudio Morana. 2014. Insights on the global macro-finance interface: Structural sources of risk factor fluctuations and the cross-section of expected stock returns. *Journal of Empirical Finance* **29**, 64-79. [Crossref]
398. Robin Zorzi, Bettina Friedl. 2014. The Optimal Hedge Ratio — An Analytical Decision Model Considering Periodical Accounting Constraints. *Review of Pacific Basin Financial Markets and Policies* **17**:04, 1450024. [Crossref]
399. Aidan Parkinson, Peter Guthrie. 2014. Evaluating the energy performance of buildings within a value at risk framework with demonstration on UK offices. *Applied Energy* **133**, 40-55. [Crossref]
400. Marek Kolodziej, Robert K. Kaufmann, Nalin Kulatilaka, David Bicchetti, Nicolas Maystre. 2014. Crude oil: Commodity or financial asset?. *Energy Economics* **46**, 216-223. [Crossref]
401. Janick Christian Mollet, Andreas Ziegler. 2014. Socially responsible investing and stock performance: New empirical evidence for the US and European stock markets. *Review of Financial Economics* **23**:4, 208-216. [Crossref]
402. Sangwon Suh, Wonho Song, Bong-Soo Lee. 2014. A new method for forming asset pricing factors from firm characteristics. *Applied Economics* **46**:28, 3463-3482. [Crossref]
403. Pei-Tha Gan. 2014. The precise form of financial integration: Empirical evidence for selected Asian countries. *Economic Modelling* **42**, 208-219. [Crossref]

404. Mario Filiasi, Giacomo Livan, Matteo Marsili, Maria Peressi, Erik Vesselli, Elia Zarinelli. 2014. On the concentration of large deviations for fat tailed distributions, with application to financial data. *Journal of Statistical Mechanics: Theory and Experiment* **2014**:9, P09030. [Crossref]
405. Terry Benzschawel,, Liang Fu,, Austin Murphy. 2014. An Empirical Analysis of Segmented Pricing of Bond Systematic Risk. *Credit and Capital Markets – Kredit und Kapital* **47**:3, 439-464. [Crossref]
406. Carolin Fritzsche,, Lars Vandrei. 2014. Keiner will sie haben – Theoretische Ursachen für Immobilienleerstand. *Credit and Capital Markets – Kredit und Kapital* **47**:3, 465-483. [Crossref]
407. Mary E. Barth, Eric C. So. 2014. Non-Diversifiable Volatility Risk and Risk Premiums at Earnings Announcements. *The Accounting Review* **89**:5, 1579-1607. [Crossref]
408. Richard A. Michelfelder. 2014. Asset characteristics of solar renewable energy certificates: market solution to encourage environmental sustainability. *Journal of Sustainable Finance & Investment* **4**:3, 280-296. [Crossref]
409. J. Augusto Felício, Irina Ivashkovskaya, Ricardo Rodrigues, Anastasia Stepanova. 2014. Corporate governance and performance in the largest European listed banks during the financial crisis. *Innovar* **24**:53, 83-98. [Crossref]
410. Duc Hung Tran. 2014. Multiple corporate governance attributes and the cost of capital – Evidence from Germany. *The British Accounting Review* **46**:2, 179-197. [Crossref]
411. Urs von Arx, Andreas Ziegler. 2014. The effect of corporate social responsibility on stock performance: new evidence for the USA and Europe. *Quantitative Finance* **14**:6, 977-991. [Crossref]
412. Austin Murphy, Liang Fu, Terry Benzschawel. 2014. An Empirical Examination of Ex AnteEstimates of the Market Risk Premium. *The Journal of Investing* **23**:2, 51-58. [Crossref]
413. Debasish Majumder. 2014. Asset pricing for inefficient markets: Evidence from China and India. *The Quarterly Review of Economics and Finance* **54**:2, 282-291. [Crossref]
414. . Bibliography 1213-1238. [Crossref]
415. Yi-Cheng Shih, Sheng-Syan Chen, Cheng-Few Lee, Po-Jung Chen. 2014. The evolution of capital asset pricing models. *Review of Quantitative Finance and Accounting* **42**:3, 415-448. [Crossref]
416. Kao-Yi Shen, Min-Ren Yan, Gwo-Hshiung Tzeng. 2014. Combining VIKOR-DANP model for glamor stock selection and stock performance improvement. *Knowledge-Based Systems* **58**, 86-97. [Crossref]
417. Madhu Acharyya, Chris Brady. 2014. Designing an Enterprise Risk Management Curriculum for Business Studies: Insights From a Pilot Program. *Risk Management and Insurance Review* **17**:1, 113-136. [Crossref]
418. Christoph Schwarzbach, Frederik Kunze, Norman Rudschuck, Torsten Windels. 2014. Stock investments for German life insurers in the current low interest environment: more homework to do. *Zeitschrift für die gesamte Versicherungswissenschaft* **103**:1, 45-63. [Crossref]
419. Ekaterina Svetlova. 2014. Modelling Beyond Application: Epistemic and Non-epistemic Values in Modern Science. *International Studies in the Philosophy of Science* **28**:1, 79-98. [Crossref]
420. Katsiaryna Salavei Bardos, Dev Mishra. 2014. Financial restatements, litigation and implied cost of equity. *Applied Financial Economics* **24**:1, 51-71. [Crossref]
421. René Aïd. A Review of Optimal Investment Rules in Electricity Generation 3-40. [Crossref]
422. Stefan Behringer. Kennzahlen im Konzerncontrolling 89-132. [Crossref]
423. Nicholas Apergis, James E. Payne. 2014. Resurrecting the size effect: Evidence from a panel nonlinear cointegration model for the G7 stock markets. *Review of Financial Economics* **23**:1, 46-53. [Crossref]
424. Hersh Shefrin. Distinguishing Rationality and Bias in Prices: Implications from Judgments of Risk and Expected Return 7-49. [Crossref]

425. Pablo Koch-Medina, Jan Wenzelburger. 2014. Equilibria in the CAPM with Nontradeable Endowments. *SSRN Electronic Journal* . [Crossref]
426. Arun Muralidhar, Kazuhiko Ohashi, Sung Hwan Shin. 2014. The Relative Asset Pricing Model: The Third Dimension. *SSRN Electronic Journal* . [Crossref]
427. Robert F. Engle. 2014. Dynamic Conditional Beta. *SSRN Electronic Journal* . [Crossref]
428. Ian Kaplan. 2014. Value Factors Do Not Forecast Returns for S&P 500 Stocks. *SSRN Electronic Journal* . [Crossref]
429. Marcio Andre Veras Machado, MMrcia Reis Machado. 2014. Liquidity and Asset Pricing: Evidence from the Brazilian Market. *SSRN Electronic Journal* . [Crossref]
430. Myuran Rajaratnam, Christo J. Auret, Robert William Vivian. 2014. Bashing Barmy Banks: Rethinking Bank Valuation Models Post the Global Financial Crisis (GFC). *SSRN Electronic Journal* . [Crossref]
431. Abel Rodriguez, Ziwei Wang, Athanasios Kottas. 2014. Assessing Systematic Risk in the S&P500 Index between 2000 and 2011: A Bayesian Nonparametric Approach. *SSRN Electronic Journal* . [Crossref]
432. Roberto Savona, Cesare Orsini. 2014. Taking the Right Course Navigating the ERC Universe. *SSRN Electronic Journal* . [Crossref]
433. William R. Pratt, Gokce Soydemir, Elena Bastida. 2014. Global Convergence of Health Care Financing in OECD Countries: An Equilibrium Based Asset Pricing Approach. *SSRN Electronic Journal* . [Crossref]
434. Arun Muralidhar. 2014. Modern Prospect Theory: The Missing Link Between Modern Portfolio Theory and Prospect Theory. *SSRN Electronic Journal* . [Crossref]
435. ZZlia Cazalet, Thierry Roncalli. 2014. Facts and Fantasies About Factor Investing. *SSRN Electronic Journal* . [Crossref]
436. Yasser Alhenawi. 2014. On the Interaction between Momentum Effect and Size Effect. *SSRN Electronic Journal* . [Crossref]
437. Paola De Santis, Carlo Drago. 2014. Systematic Risk Asymmetry of the American Real Estate Securities: Some New Econometric Evidences. *SSRN Electronic Journal* . [Crossref]
438. David Geltner. 2014. Real Estate Price Indices and Price Dynamics: An Overview from an Investments Perspective. *SSRN Electronic Journal* . [Crossref]
439. Siti Hajar Nadrah Mohamad Ghouse, Noryati Ahmad. 2014. Conceptual Paper of the Trading Strategy: Dogs of the Dow Theory (Dod). *SSRN Electronic Journal* . [Crossref]
440. Lynda Khalaf. 2014. L'économétrie et l'évidence fallacieuse : erreurs et avancées. *L'Actualité économique* **90**:1, 5. [Crossref]
441. Alfonso A. Rojo-Ramírez. 2014. Privately Held Company Valuation and Cost of Capital. *Journal of Business Valuation and Economic Loss Analysis* **9**:1, 1-21. [Crossref]
442. Stefan Lutz, Daniel Kleinfeldt. 2013. Risk as Determinant of Income and Cross-border Pricing of Multinational Enterprises. *Studies in Microeconomics* **1**:2, 185-212. [Crossref]
443. Andrew Walsh. 2013. Core–Satellite Strategies: Combining Stability and Opportunity in an ETF Portfolio. *The Journal of Index Investing* **4**:3, 50-53. [Crossref]
444. Andrés Mauricio Gómez Sánchez, José Gabriel Astaiza Gómez. 2013. Ciclo económico y prima por riesgo en el mercado accionario colombiano. *Ecos de Economía* **17**:37, 109-124. [Crossref]
445. Antonina Waszczuk. 2013. Do local or global risk factors explain the size, value and momentum trading pay-offs on the Warsaw Stock Exchange?. *Applied Financial Economics* **23**:19, 1497-1508. [Crossref]

446. Shima Lashgari, Jurgita Antuchevičienė, Alireza Delavari, Omid Kheirkhah. 2013. BEYOND CAPM: AN INNOVATIVE FACTOR MODEL TO OPTIMIZE THE RISK AND RETURN TRADE-OFF. *Journal of Business Economics and Management* **15**:4, 615-630. [Crossref]
447. Andreas Ehrenmann, Yves Smeers. 2013. Risk adjusted discounted cash flows in capacity expansion models. *Mathematical Programming* **140**:2, 267-293. [Crossref]
448. Bastiaan Pluijmers, Imke Hollander, Ramon Tol, Dimitris Melas. 2013. On the Commonality of Characteristics of Managed Volatility Portfolios. *The Journal of Investing* **22**:3, 86-98. [Crossref]
449. Stefan Lutz. 2013. Risk premia in multi-national enterprises. *The North American Journal of Economics and Finance* **25**, 293-305. [Crossref]
450. Markus Ampenberger, Thomas Schmid, Ann-Kristin Achleitner, Christoph Kaserer. 2013. Capital structure decisions in family firms: empirical evidence from a bank-based economy. *Review of Managerial Science* **7**:3, 247-275. [Crossref]
451. YongHee Kim, MinChung Kim, John W. O'Neill. 2013. Advertising and Firm Risk: A Study of the Restaurant Industry. *Journal of Travel & Tourism Marketing* **30**:5, 455-470. [Crossref]
452. M.-C. Beaulieu, J.-M. Dufour, L. Khalaf. 2013. Identification-Robust Estimation and Testing of the Zero-Beta CAPM. *The Review of Economic Studies* **80**:3, 892-924. [Crossref]
453. Paulo Augusto P. Britto, Carlos Henrique Rocha. 2013. Determinação do valor da tarifa-leilão do serviço de transporte de passageiros por ônibus: um modelo alternativo. *Journal of Transport Literature* **7**:3, 177-191. [Crossref]
454. Will Gans, Beat Hintermann. 2013. Market Effects of Voluntary Climate Action by Firms: Evidence from the Chicago Climate Exchange. *Environmental and Resource Economics* **55**:2, 291-308. [Crossref]
455. Jaspal Singh, Kiranpreet Kaur. 2013. Testing the Performance of Graham's Net Current Asset Value Strategy in Indian Stock Market. *Asia-Pacific Journal of Management Research and Innovation* **9**:2, 171-179. [Crossref]
456. John H. Hall, Wim Westerman. Basic Risk Adjustment Techniques in Capital Budgeting 215-239. [Crossref]
457. Robert B. Durand, Rick Newby, Leila Peggs, Michelle Siekierka. 2013. Personality. *Journal of Behavioral Finance* **14**:2, 116-133. [Crossref]
458. Pertti Lahdenperä. 2013. Determining 'the most economically advantageous tender' based on capability and fee-percentage criteria. *Journal of Public Procurement* **13**:4, 409-446. [Crossref]
459. Nicholas C. Barberis. 2013. Thirty Years of Prospect Theory in Economics: A Review and Assessment. *Journal of Economic Perspectives* **27**:1, 173-196. [Abstract] [View PDF article] [PDF with links]
460. L. Peter Jennergren. 2013. Firm Valuation with Bankruptcy Risk. *Journal of Business Valuation and Economic Loss Analysis* **8**:1. . [Crossref]
461. Jian Zhou. 2013. Conditional market beta for REITs: A comparison of modeling techniques. *Economic Modelling* **30**, 196-204. [Crossref]
462. Imad A. Moosa. 2013. The Capital Asset Pricing Model (CAPM): The History of a Failed Revolutionary Idea in Finance? Comments and Extensions. *Abacus* **49**, 62-68. [Crossref]
463. Arun Muralidhar, Kazuhiko Ohashi, Sung Hwan Shin. 2013. LDI: Does it Turn CAPM into RAPM?. *SSRN Electronic Journal* . [Crossref]
464. Marcio Andre Veras Machado, Otavio Ribeiro de Medeiros. 2013. Does the Liquidity Effect Exist in the Brazilian Stock Market?. *SSRN Electronic Journal* . [Crossref]
465. Argiro Svingou. 2013. Cross-Sectional Analysis of Stock Returns in Athens Stock Exchange for the Period 2004-2011. *SSRN Electronic Journal* . [Crossref]
466. Issam S. Strub. 2013. Tail Hedging Strategies. *SSRN Electronic Journal* . [Crossref]

467. Gueorgui I. Kolev. 2013. Two Gold Return Puzzles. *SSRN Electronic Journal* . [Crossref]
468. Philippe Bertrand, Vincent Lapointe. 2013. Socially Responsible Investment Performance: Impacts of Weighting by Capitalization. *SSRN Electronic Journal* . [Crossref]
469. Eben Otuteye, Mohammad Siddiquee. 2013. Redefining Risk: Propositions to Motivate a Re-Examination of the Standard Risk vs. Return Relationship in Common Stock and Bond Portfolio Management. *SSRN Electronic Journal* . [Crossref]
470. Sabine Elmiger. 2013. Can CRRA Preferences Explain CAPM-Anomalies in the Cross-Section of Stock Returns?. *SSRN Electronic Journal* . [Crossref]
471. Deling Chen. 2013. Empirical Analysis of Stock Returns of Banks in China's A-Share Markets. *SSRN Electronic Journal* . [Crossref]
472. Muthucattu Thomas Paul, Fosuhene Akua Asarebea. 2013. Validity of the Capital Assets Pricing Model: Evidence from the Indian Companies The NSE India. *SSRN Electronic Journal* . [Crossref]
473. Jialiu Lu. 2013. The CAPM: A Reformulation. *SSRN Electronic Journal* . [Crossref]
474. Kevin Kurt Robinson. 2013. Technical Analysis: Does Recent Market Data Substantiate the Efficient Market Hypothesis?. *SSRN Electronic Journal* . [Crossref]
475. Peter C. Dawson. 2013. An Economic Analysis of the Competitive Risk-Return Paradigm. *SSRN Electronic Journal* . [Crossref]
476. Dinh Tran Ngoc Huy. 2013. The risk level of Vietnam non-banking investment and financial services industry under financial leverage during and after the global crisis 2007-2011. *Risk Governance and Control: Financial Markets and Institutions* **3**:3, 48-55. [Crossref]
477. Nicolai Striewe, Nico Rottke, Joachim Zietz. 2013. The Impact of Institutional Ownership on REIT Performance. *Journal of Real Estate Portfolio Management* **19**:1, 17-30. [Crossref]
478. Joseph A. Cerniglia, Petter N. Kolm, Frank J. Fabozzi. Cross-Sectional Factor-Based Models and Trading Strategies . [Crossref]
479. Hans Marius Eikseth, Snorre Lindset. 2012. Are taxes sufficient for CAPM rejection?. *Applied Economics Letters* **19**:18, 1813-1816. [Crossref]
480. Michael McKenzie, Ólan T. Henry. 2012. The determinants of short selling: evidence from the Hong Kong equity market. *Accounting & Finance* **52**, 183-216. [Crossref]
481. Mohamed El Hedi Arouri, Duc Khuong Nguyen, Kuntara Pukthuanthong. 2012. An international CAPM for partially integrated markets: Theory and empirical evidence. *Journal of Banking & Finance* **36**:9, 2473-2493. [Crossref]
482. Moshe Levy. 2012. On the Spurious Correlation Between Sample Betas and Mean Returns. *Applied Mathematical Finance* **19**:4, 341-360. [Crossref]
483. Feng-jun Liu. An examination of CAPM based on data from Shenzhen Stock Exchange 1560-1566. [Crossref]
484. S. Maier, A. Street. Model for the economic feasibility of energy recovery from municipal solid waste in Brazil 1-10. [Crossref]
485. Saumya Ranjan Dash, Jitendra Mahakud. 2012. Investor sentiment, risk factors and stock return: evidence from Indian non-financial companies. *Journal of Indian Business Research* **4**:3, 194-218. [Crossref]
486. Tony Chieh-Tse Hou. 2012. Return persistence and investment timing decisions in Taiwanese domestic equity mutual funds. *Managerial Finance* **38**:9, 873-891. [Crossref]
487. Iftekhar Hasan, Heiko Schmiedel, Liang Song. 2012. Returns to Retail Banking and Payments. *Journal of Financial Services Research* **41**:3, 163-195. [Crossref]

488. Raul Leote de Carvalho, Xiao Lu, Pierre Moulin. 2012. Demystifying Equity Risk–Based Strategies: A Simple Alpha plus Beta Description. *The Journal of Portfolio Management* **38**:3, 56-70. [Crossref]
489. Xiao Lu, Pierre Moulin. 2012. Demystifying Equity Risk–Based Strategies: A Simple Alpha plus Beta Description. *The Journal of Portfolio Management* 120308233706009. [Crossref]
490. Manfred Gilli, Enrico Schumann. 2012. Heuristic optimisation in financial modelling. *Annals of Operations Research* **193**:1, 129-158. [Crossref]
491. Amit Goyal. 2012. Empirical cross-sectional asset pricing: a survey. *Financial Markets and Portfolio Management* **26**:1, 3-38. [Crossref]
492. Mohamed El Hedi Arouri, Philippe Foulquier. 2012. Financial market integration: Theory and empirical results. *Economic Modelling* **29**:2, 382-394. [Crossref]
493. Alexandros Kostakis, Kashif Muhammad, Antonios Siganos. 2012. Higher co-moments and asset pricing on London Stock Exchange. *Journal of Banking & Finance* **36**:3, 913-922. [Crossref]
494. Melih Madanoglu, Murat Kizildag, Ersem Karadag. 2012. Estimating Cost of Equity in the Restaurant Industry: What IS Your Required Rate of Return?. *The Journal of Hospitality Financial Management* **20**:1, 57-74. [Crossref]
495. Roger Dayala. 2012. The Capital Asset Pricing Model: A Fundamental Critique. *Business Valuation Review* **31**:1, 23-34. [Crossref]
496. Konstantinos A. Chrysafis. 2012. Corporate investment appraisal with possibilistic CAPM. *Mathematical and Computer Modelling* **55**:3-4, 1041-1050. [Crossref]
497. Zhi Da, Re-Jin Guo, Ravi Jagannathan. 2012. CAPM for estimating the cost of equity capital: Interpreting the empirical evidence. *Journal of Financial Economics* **103**:1, 204-220. [Crossref]
498. Woraphon Wattanatorn, Termkiat Kanchanapoom. 2012. Oil Prices and Profitability Performance: Sector Analysis. *Procedia - Social and Behavioral Sciences* **40**, 763-767. [Crossref]
499. Fernando Gómez-Bezares, Luis Ferruz, María Vargas. 2012. Can we beat the market with beta? An intuitive test of the CAPM. *Spanish Journal of Finance and Accounting / Revista Española de Financiación y Contabilidad* **41**:155, 333-352. [Crossref]
500. Mary E. Barth, Eric C. So. 2012. Non-Diversifiable Volatility Risk and Risk Premiums at Earnings Announcements. *SSRN Electronic Journal* . [Crossref]
501. Giuseppe Marzo. 2012. A Theory-of-The-Firm Perspective on Discount Rates Formulation in Investment Valuation. *SSRN Electronic Journal* . [Crossref]
502. Jun Yuan, Leonard MacLean, Kuan Xu, Yonggan Zhao. 2012. How Do Local Markets Respond to Global Risk Factor Differently in Various Market Regimes? A Study of Country Exchange Traded Funds. *SSRN Electronic Journal* . [Crossref]
503. Matthias Xaver Hanauer, Christoph Kaserer, Marc Steffen Rapp. 2012. Risikofaktoren und Multifaktormodelle für den Deutschen Aktienmarkt (Risk Factors and Multi-Factor Models for the German Stock Market). *SSRN Electronic Journal* . [Crossref]
504. Min Bai. 2012. Short-Selling Status and Asset-Pricing Models. *SSRN Electronic Journal* . [Crossref]
505. Mike Ward, Chris Muller. 2012. Empirical Testing of the CAPM on the Johannesburg Stock Exchange. *SSRN Electronic Journal* . [Crossref]
506. M. Hashem Pesaran, Takashi Yamagata. 2012. Testing CAPM with a Large Number of Assets. *SSRN Electronic Journal* . [Crossref]
507. Zvika Afik, Yaron Lahav. 2012. Risk Transfer Valuation in Advance Pricing Agreements between Multinational Enterprises and Tax Authorities. *SSRN Electronic Journal* . [Crossref]
508. Lord Mensah, R. K. Avuglah, Vincent Dedu. 2012. Asset Allocation on the Ghana Stock Exchange. *SSRN Electronic Journal* . [Crossref]

509. Gohar G. Stepanyan. 2012. Revisiting Firm Life Cycle Theory for New Directions in Finance. *SSRN Electronic Journal* . [Crossref]
510. Myuran Rajaratnam, Balakanapathy Rajaratnam, Kanshukan Rajaratnam. 2012. Mauling Mr. Market: Valuing Equity Capital of Businesses by Long-Term Value-Investors. *SSRN Electronic Journal* . [Crossref]
511. David Nanigian. 2012. Capitalizing on the Greatest Anomaly in Finance with Mutual Funds. *SSRN Electronic Journal* . [Crossref]
512. Saumya Ranjan Dash. 2012. Does Sentiment Risk Persist as Priced Risk Factor? A Multifactor Approach for Stock Return. *SSRN Electronic Journal* . [Crossref]
513. Arun Muralidhar, Sung Hwan Shin. 2012. The Relative Asset Pricing Model Incorporating Liabilities and Delegation to CIOs: Version 0.1. *SSRN Electronic Journal* . [Crossref]
514. Thierry Roncalli, Guillaume Weisang. 2012. Risk Parity Portfolios with Risk Factors. *SSRN Electronic Journal* . [Crossref]
515. Roman Brückner, Patrick Lehmann, Richard Stehle. 2012. In Germany the CAPM is Alive and Well. *SSRN Electronic Journal* . [Crossref]
516. Nicholas Barberis. 2012. Thirty Years of Prospect Theory in Economics: A Review and Assessment. *SSRN Electronic Journal* . [Crossref]
517. Fabian Echterling, Brigitte Eierle. 2012. Implied Cost of Capital Under Heterogeneous Expectations. *SSRN Electronic Journal* . [Crossref]
518. javed bin kamal. 2012. Optimal Portfolio Selection in Ex Ante Stock Price Bubble and Furthermore Bubble Burst Scenario from Dhaka Stock Exchange with Relevance to Sharpe's Single Index Model. *SSRN Electronic Journal* . [Crossref]
519. Juan Mascareñas. 2012. Gestión De Carteras Ii: Modelo De Valoración De Activos (Portfolio Management II: Capital Asset Pricing Model). *SSRN Electronic Journal* . [Crossref]
520. Wilson N. Sy. 2012. Scale and Competition in Australian Superannuation. *SSRN Electronic Journal* . [Crossref]
521. Yi-Jang Yu. 2012. The Asset Pricing System. *Modern Economy* **03**:05, 473-480. [Crossref]
522. Stefan Lutz. 2012. Risk Premia in Multi-National Enterprises. *SSRN Electronic Journal* . [Crossref]
523. Bartley J. Madden. Applying a Systems Mindset to Stock Valuation 43-66. [Crossref]
524. . References 159-165. [Crossref]
525. Pauline M. Ahern, Frank J. Hanley, Richard A. Michelfelder. 2011. New approach to estimating the cost of common equity capital for public utilities. *Journal of Regulatory Economics* **40**:3, 261-278. [Crossref]
526. Steve Keen. 2011. Debunking Macroeconomics. *Economic Analysis and Policy* **41**:3, 147-167. [Crossref]
527. Akhilesh Chandra, Alan Reinstein. 2011. Investment appeal of small growth stocks. *Advances in Accounting* **27**:2, 308-317. [Crossref]
528. Mathijs A. van Dijk. 2011. Is size dead? A review of the size effect in equity returns. *Journal of Banking & Finance* **35**:12, 3263-3274. [Crossref]
529. A Charteris, B Strydom. 2011. An Examination of the Volatility of South African Risk-Free Rate Proxies: A Component Garch Analysis. *Studies in Economics and Econometrics* **35**:3, 49-64. [Crossref]
530. Andreas Ziegler, Timo Busch, Volker H. Hoffmann. 2011. Disclosed corporate responses to climate change and stock performance: An international empirical analysis. *Energy Economics* **33**:6, 1283-1294. [Crossref]
531. Ajit Dayanandan, Han Donker. 2011. Oil prices and accounting profits of oil and gas companies. *International Review of Financial Analysis* **20**:5, 252-257. [Crossref]

532. Fengjun Liu. Can the Industry Index Become Another Market Variable? Empirical Evidence from the Listed Banks in Shanghai Stock Exchange 315-319. [Crossref]
533. S. Paulo. 2011. The South African Companies Act of 2008 (SACA2008), and the Sarbanes-Oxley Act of 2002. *International Journal of Law and Management* **53**:5, 340-354. [Crossref]
534. Gerald Spindler. 2011. Behavioural Finance and Investor Protection Regulations. *Journal of Consumer Policy* **34**:3, 315-336. [Crossref]
535. Liu Feng-jun, Li Fei. Empirical study of investment performance on listed banks from Shanghai Stock Exchange based on one-quarter holding period 932-937. [Crossref]
536. Norliza Muhamad Yusof, Maheran Mohd. Jaffar. Predicting the credit risk through Merton model 162-166. [Crossref]
537. TAEJUN DAVID LEE, ERIC HALEY, TAI WOONG YUN, WONJUN CHUNG. 2011. US Retirement Financial Services Advertising's Financial Information Provisions, Communication Strategies and Judgmental Heuristic Cues. *Journal of Consumer Affairs* **45**:3, 391-418. [Crossref]
538. Baibing Li, Xiangkang Yin. 2011. Information and capital asset pricing. *The European Journal of Finance* **17**:7, 505-523. [Crossref]
539. Cristian Diego Albuja, Fabio Gallo Garcia, Luiz Maurício Franco Moreiras, Elmo Tambosi Filho. 2011. Onde investir nos BRICS? Uma análise sob o prisma da organização industrial. *Revista de Administração de Empresas* **51**:4, 349-369. [Crossref]
540. Mohamed El Hedi Arouri. 2011. Does crude oil move stock markets in Europe? A sector investigation. *Economic Modelling* **28**:4, 1716-1725. [Crossref]
541. Konstantinos Kassimatis. 2011. Risk Aversion with Local Risk Seeking and Stock Returns: Evidence from the UK Market. *Journal of Business Finance & Accounting* **38**:5-6, 713-739. [Crossref]
542. Duc Hung Tran. 2011. Corporate Governance und Eigenkapitalkosten - Bestandsaufnahme des Schrifttums unter besonderer Berücksichtigung des Informationsaspektes und Forschungsperspektiven. *Zeitschrift für Betriebswirtschaft* **81**:5, 551-585. [Crossref]
543. Tanja Hribernik, Uroš Vek. 2011. Mutual Fund Performance in Slovenia: An Analysis of Mutual Funds with Investment Policies in Europe and the Energy Sector. *South East European Journal of Economics and Business* **6**:1, 61-69. [Crossref]
544. Simon Huston, Clive Warren, Peter Elliott. 2011. Elixir or delusion. *Journal of Property Investment & Finance* **29**:1, 49-58. [Crossref]
545. . Bibliography 563-576. [Crossref]
546. Robert B. Durand, Yihui Lan, Andrew Ng. 2011. Conditional beta: Evidence from Asian emerging markets. *Global Finance Journal* **22**:2, 130-153. [Crossref]
547. Amir Amel-Zadeh. 2011. The Return of the Size Anomaly: Evidence from the German Stock Market. *European Financial Management* **17**:1, 145-182. [Crossref]
548. Michael A. Crain. 2011. A Literature Review of the Size Effect. *SSRN Electronic Journal* . [Crossref]
549. Marie-Claude Beaulieu, Jean-Marie Dufour, Lynda Khalaf. 2011. Identification-Robust Estimation and Testing of the Zero-Beta CAPM. *SSRN Electronic Journal* . [Crossref]
550. Alfonso A. Rojo Rojo-Ramirez, Salvador Cruz-Rambaud Cruz-Rambaud, Juana Alonso Canadas. 2011. A Note on the Operating Return of a Company Under Modigliani-Miller Assumptions. *SSRN Electronic Journal* . [Crossref]
551. Lammertjan Dam, Pim Heijnen. 2011. Asset Pricing with Fixed Asset Supply. *SSRN Electronic Journal* . [Crossref]
552. Mohammad Ali Tareq, Saburo Horimoto. 2011. Can We Use Historical Mean Returns?. *SSRN Electronic Journal* . [Crossref]

553. Maria de Lourdes Trevino. 2011. Time Varying Arbitrage Pricing Factors in the Mexican Stock Market. *SSRN Electronic Journal* . [Crossref]
554. Baitshepi Tebogo. 2011. Valuing Securities and Managing Portfolios Under Uncertainty: A Reminder of the Underlying Assumptions. *SSRN Electronic Journal* . [Crossref]
555. Myuran Rajaratnam, Balakanapathy Rajaratnam, Kanshukan Rajaratnam. 2011. Murdering Mr. Market: An Equity Valuation and Capital Allocation Model for Long-Term Value-Investors. *SSRN Electronic Journal* . [Crossref]
556. Linda H. Chen, Lucia Silva Gao. 2011. The Pricing of Climate Risk. *SSRN Electronic Journal* . [Crossref]
557. Muhammad Hanif, Abubakar Javaid Dar. 2011. Comparative Testing of Capital Asset Pricing Model (CAPM) and Shari'a Compliant Asset Pricing Model (SCAPM): Evidence from Karachi Stock Exchange - Pakistan. *SSRN Electronic Journal* . [Crossref]
558. Tony Berrada, Reda Jürg Messikh, Gianluca Oderda, Olivier V. Pictet. 2011. Beta-Arbitrage Strategies: When Do They Work, and Why?. *SSRN Electronic Journal* . [Crossref]
559. Alfonso A. Rojo Rojo-Ramirez, Juana Alonso Canadas, Salvador Cruz-Rambaud Cruz-Rambaud. 2011. Discount Rate and Cost of Capital: Some More about the Puzzle. *SSRN Electronic Journal* . [Crossref]
560. Stefan Lutz. 2011. Simultaneous Determination of Market Value and Risk Premium in the Valuation of Firms. *SSRN Electronic Journal* . [Crossref]
561. Nagendra Marisetty. 2011. An Empirical Study on CAPM with Respect to NSE NIFTY Stocks. *SSRN Electronic Journal* . [Crossref]
562. Martin Grandes, Demian T. Panigo, Ricardo A. Pasquini. 2010. On the estimation of the cost of equity in Latin America. *Emerging Markets Review* **11**:4, 373-389. [Crossref]
563. Young-Soon Hwang, Hong-Ghi Min, Judith A. McDonald, Hwagyun Kim, Bong-Han Kim. 2010. Using the credit spread as an option-risk factor: Size and value effects in CAPM. *Journal of Banking & Finance* **34**:12, 2995-3009. [Crossref]
564. Ravi Jagannathan, Ernst Schaumburg, Guofu Zhou. 2010. Cross-Sectional Asset Pricing Tests. *Annual Review of Financial Economics* **2**:1, 49-74. [Crossref]
565. Maria Elena De Giuli, Mario Alessandro Maggi, Claudia Tarantola. 2010. Bayesian outlier detection in Capital Asset Pricing Model. *Statistical Modelling: An International Journal* **10**:4, 375-390. [Crossref]
566. Yash Pal Taneja. 2010. Revisiting Fama French Three-Factor Model in Indian Stock Market. *Vision: The Journal of Business Perspective* **14**:4, 267-274. [Crossref]
567. Marie-Claude Beaulieu, Jean-Marie Dufour, Lynda Khalaf. 2010. Asset-pricing anomalies and spanning: Multivariate and multifactor tests with heavy-tailed distributions. *Journal of Empirical Finance* **17**:4, 763-782. [Crossref]
568. Olivier Brandouy, Walter Briec, Kristiaan Kerstens, Ignace Van de Woestyne. 2010. Portfolio performance gauging in discrete time using a Luenberger productivity indicator. *Journal of Banking & Finance* **34**:8, 1899-1910. [Crossref]
569. S. Paulo. 2010. The United Kingdom's Companies Act of 2006 and the capital asset pricing model. *International Journal of Law and Management* **52**:4, 253-264. [Crossref]
570. Chen Jianbao, Xu Yanping, Cheng Tingting. Quantile Regression Analysis of Cross-Section Returns in Chinese Stock Market 169-172. [Crossref]
571. Christophe Schinckus. 2010. Semiotics of Financial Marketplace. *Journal of Interdisciplinary Economics* **22**:4, 317-333. [Crossref]

572. Erlend Kvaal. 2010. The Discount Rate of IAS 36 – A Comment. *Accounting in Europe* **7**:1, 87-95. [Crossref]

573. Shinn-Shyr Wang, Kyle W. Stiegert, Tirtha P. Dhar. 2010. Strategic Pricing Behavior under Asset Value Maximization. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* **58**:2, 151-170. [Crossref]

574. S. Paulo. 2010. The UK Companies Act of 2006 and the Sarbanes-Oxley Act of 2002. *International Journal of Law and Management* **52**:3, 173-181. [Crossref]

575. Kurt Niquidet. 2010. Equity pricing in the forest sector: evidence from North American stock markets. *Canadian Journal of Forest Research* **40**:5, 943-952. [Crossref]

576. Michail Koubouros, Dimitrios Malliaropulos, Ekaterini Panopoulou. 2010. Long-run cash flow and discount-rate risks in the cross-section of US returns. *The European Journal of Finance* **16**:3, 227-244. [Crossref]

577. JEFFREY A. BUSSE, AMIT GOYAL, SUNIL WAHAL. 2010. Performance and Persistence in Institutional Investment Management. *The Journal of Finance* **65**:2, 765-790. [Crossref]

578. S. Paulo. 2010. Hamada's equation, the Sarbanes-Oxley Act of 2002 and the UK Companies Act of 2006. *International Journal of Law and Management* **52**:1, 54-63. [Crossref]

579. Caroline Fohlin, Steffen Reinhold. 2010. Common stock returns in the pre-WWI Berlin Stock Exchange. *Cliometrica* **4**:1, 75-96. [Crossref]

580. Frank Sortino. The Big Picture 1-12. [Crossref]

581. Cherif Guermat, Mark C. Freeman. 2010. A net beta test of asset pricing models. *International Review of Financial Analysis* **19**:1, 1-9. [Crossref]

582. Knut Sandberg Eriksen, Svenn Jensen. 2010. The cost of second best pricing and the value of risk premium. *Research in Transportation Economics* **30**:1, 29-37. [Crossref]

583. Herbert Kalthoff, Uwe Vormbusch. Representing and Modelling: The Case of Portfolio Management 174-188. [Crossref]

584. Syou-Ching Lai, Hung-Chih Li, James A. Conover, Frederick Wu. O-score financial distress risk asset pricing 51-94. [Crossref]

585. Haim Levy. 2010. The CAPM is Alive and Well: A Review and Synthesis. *European Financial Management* **16**:1, 43-71. [Crossref]

586. Mathijs Cosemans, Rik G. P. Frehen, Peter C. Schotman, Rob Bauer. 2010. Estimating Security Betas Using Prior Information Based on Firm Fundamentals. *SSRN Electronic Journal* . [Crossref]

587. Reda Jürg Messikh, Gianluca Oderda. 2010. A Proof of the Outperformance of Beta Arbitrage Strategies. *SSRN Electronic Journal* . [Crossref]

588. Kenneth Monroe Norton, MBA, DBA, Pan G. Yatrakis. 2010. The Effects of Cultural Differences on Knowledge Assets and U.S. MNCS' Firms Value: A Three-Valuation Model Approach. *SSRN Electronic Journal* . [Crossref]

589. Matthias Thomas, Daniel Piazolo, Sebastian Michael Gläsner. 2010. Analyzing the Changing Risk and Return Structure of German Open-Ended Funds Using Semivariance Based Performance Measures. *SSRN Electronic Journal* . [Crossref]

590. Zhi Da, Re J. Guo, Ravi Jagannathan. 2010. CAPM for Estimating the Cost of Equity Capital: Interpreting the Empirical Evidence. *SSRN Electronic Journal* . [Crossref]

591. George Bragues. 2010. The Politics of Financial Markets: An Introductory Discussion. *SSRN Electronic Journal* . [Crossref]

592. Pierre Hereil, Philippe Mitaine, Nicolas Moussavi, Thierry Roncalli. 2010. Mutual Fund Ratings and Performance Persistence. *SSRN Electronic Journal* . [Crossref]

593. S. Trevis Certo, Tim R. Holcomb, R. Michael Holmes. 2009. IPO Research in Management and Entrepreneurship: Moving the Agenda Forward. *Journal of Management* **35**:6, 1340-1378. [Crossref]
594. Christophe Schinckus. 2009. Economic uncertainty and econophysics. *Physica A: Statistical Mechanics and its Applications* **388**:20, 4415-4423. [Crossref]
595. Claes Fornell, Sunil Mithas, Forrest V. Morgeson. 2009. Commentary —The Economic and Statistical Significance of Stock Returns on Customer Satisfaction. *Marketing Science* **28**:5, 820-825. [Crossref]
596. Stefan Veith, Jörg R. Werner, Jochen Zimmermann. 2009. Capital market response to emission rights returns: Evidence from the European power sector. *Energy Economics* **31**:4, 605-613. [Crossref]
597. Moawia Alghalith. 2009. Alternative theory of asset pricing. *Journal of Asset Management* **10**:2, 73-74. [Crossref]
598. Marie-Claude Beaulieu, Jean-Marie Dufour, Lynda Khalaf. 2009. Finite sample multivariate tests of asset pricing models with coskewness. *Computational Statistics & Data Analysis* **53**:6, 2008-2021. [Crossref]
599. Jean-Luc Besson, Michel M Dacorogna, Paolo de Martin, Michael Kastenholz, Michael Moller. 2009. How Much Capital Does a Reinsurance Need?. *The Geneva Papers on Risk and Insurance - Issues and Practice* **34**:2, 159-174. [Crossref]
600. Ric Thomas, Robert Shapiro. 2009. Managed Volatility: A New Approach to Equity Investing. *The Journal of Investing* **18**:1, 15-23. [Crossref]
601. C. Emre Alper, Oya Pinar Ardic, Salih Fendoglu. 2009. THE ECONOMICS OF THE UNCOVERED INTEREST PARITY CONDITION FOR EMERGING MARKETS. *Journal of Economic Surveys* **23**:1, 115-138. [Crossref]
602. Stuart Hyde, Mohamed Sherif. 2009. Tests of the conditional asset pricing model: further evidence from the cross-section of stock returns. *International Journal of Finance & Economics* **5**, n/a-n/a. [Crossref]
603. Ekaterina Svetlova. Theoretical Models as Creative Resources in Financial Markets 121-135. [Crossref]
604. Andreas Ziegler, Timo Busch, Volker H. Hoffmann. 2009. Corporate Responses to Climate Change and Financial Performance: The Impact of Climate Policy. *SSRN Electronic Journal* . [Crossref]
605. Haim Levy. 2009. Behavioral Economics and Asset Pricing. *SSRN Electronic Journal* . [Crossref]
606. Markus Ampenberger, Thomas Schmid, Ann-Kristin Achleitner, Christoph Kaserer. 2009. Capital Structure Decisions in Family Firms - Empirical Evidence from a Bank-Based Economy. *SSRN Electronic Journal* . [Crossref]
607. Pablo Fernandez. 2009. Betas Used by Professors: A Survey with 2,500 Answers. *SSRN Electronic Journal* . [Crossref]
608. Pablo Fernandez, Vicente J. Bermejo. 2009. Betas Used by Companies and Professors in Europe: A Survey. *SSRN Electronic Journal* . [Crossref]
609. Marius Bausys. 2009. The Performance of Minimum Variance Portfolios in the Baltic Equity Markets. *SSRN Electronic Journal* . [Crossref]
610. Rafał Wolski. 2009. The Influence of Negative Beta Assets on the Empirical SML in the Polish Capital Market. *Folia Oeconomica Stetinensia* **8**:1, 140-153. [Crossref]
611. Sudipta Basu. 2008. Panel on Big Unanswered Questions in Accounting—Synopsis. *Accounting Horizons* **22**:4, 449-451. [Crossref]
612. Brian J. Jacobsen, Xiaochun Liu. 2008. China's segmented stock market: An application of the conditional international capital asset pricing model. *Emerging Markets Review* **9**:3, 153-173. [Crossref]

613. Andy Lockett, Rory P. O'Shea, Mike Wright. 2008. The Development of the Resource-based View: Reflections from Birger Wernerfelt 1. *Organization Studies* **29**:8-9, 1125-1141. [Crossref]
614. Chyi Lin Lee, Jon Robinson, Richard Reed. 2008. Listed property trusts and downside systematic risk sensitivity. *Journal of Property Investment & Finance* **26**:4, 304-328. [Crossref]
615. Lerzan Aksoy, Bruce Cooil, Christopher Groening, Timothy L. Keiningham, Atakan Yalçın. 2008. The Long-Term Stock Market Valuation of Customer Satisfaction. *Journal of Marketing* **72**:4, 105-122. [Crossref]
616. D. Sornette, V.F. Pisarenko. 2008. Properties of a simple bilinear stochastic model: Estimation and predictability. *Physica D: Nonlinear Phenomena* **237**:4, 429-445. [Crossref]
617. Anusha Chari, Peter Blair Henry. 2008. Firm-specific information and the efficiency of investment. *Journal of Financial Economics* **87**:3, 636-655. [Crossref]
618. Fernando de Holanda Barbosa. 2008. Banco Nacional: jogo de Ponzi, PROER e FCVS. *Revista de Economia Política* **28**:1, 97-115. [Crossref]
619. Dimo Dimov, Gordon Murray. 2008. Determinants of the Incidence and Scale of Seed Capital Investments by Venture Capital Firms. *Small Business Economics* **30**:2, 127-152. [Crossref]
620. Julius Hemminki, Vesa Puttonen. 2008. Fundamental indexation in Europe. *Journal of Asset Management* **8**:6, 401-405. [Crossref]
621. Herbert Kimura, Leonardo Cruz Basso, Eduardo Kazuo Kayo. 2008. The Interplay between Strategy and Finance. *SSRN Electronic Journal* . [Crossref]
622. Hendri Adriaens, Bas Donkers, Bertrand Melenberg. 2008. The CAPM with Endogenous Beliefs. *SSRN Electronic Journal* . [Crossref]
623. Bartley J. Madden. 2008. Applying a Systems Mindset to Stock Valuation. *SSRN Electronic Journal* . [Crossref]
624. Vanita Tripathi. 2008. Company Fundamentals and Equity Returns in India. *SSRN Electronic Journal* . [Crossref]
625. Najah Attig, Omrane Guedhami, Dev R. Mishra. 2008. Multiple Large Shareholders, Control Contests, and Implied Cost of Equity. *SSRN Electronic Journal* . [Crossref]
626. Peter Christoffersen, Kris Jacobs, Gregory Vainberg. 2008. Forward-Looking Betas. *SSRN Electronic Journal* . [Crossref]
627. Chyi Lee, Jon Robinson, Richard Reed. 2008. Downside Beta and the Cross-sectional Determinants of Listed Property Trust Returns. *Journal of Real Estate Portfolio Management* **14**:1, 49-62. [Crossref]
628. Ron Bird, Lorenzo Casavecchia. 2007. Sentiment and Financial Health Indicators for Value and Growth Stocks: The European Experience. *The European Journal of Finance* **13**:8, 769-793. [Crossref]
629. Michael J. Mauboussin. 2007. The Wisdom and Whims of the Collective. *CFA Institute Conference Proceedings Quarterly* **24**:4, 1-8. [Crossref]
630. Peter Blair Henry. 2007. Capital Account Liberalization: Theory, Evidence, and Speculation. *Journal of Economic Literature* **45**:4, 887-935. [Abstract] [View PDF article] [PDF with links]
631. Moshe Levy. 2007. Conditions for a CAPM equilibrium with positive prices. *Journal of Economic Theory* **137**:1, 404-415. [Crossref]
632. Gerhard Kristandl, Nick Bontis. 2007. The impact of voluntary disclosure on cost of equity capital estimates in a temporal setting. *Journal of Intellectual Capital* **8**:4, 577-594. [Crossref]
633. Sven Husmann, Andreas Stephan. 2007. On estimating an asset's implicit beta. *Journal of Futures Markets* **27**:10, 961-979. [Crossref]

634. Marie-Claude Beaulieu, Jean-Marie Dufour, Lynda Khalaf. 2007. Multivariate Tests of Mean–Variance Efficiency With Possibly Non-Gaussian Errors. *Journal of Business & Economic Statistics* **25**:4, 398-410. [Crossref]

635. Arvind Pai, David Geltner. 2007. Stocks Are from Mars, Real Estate Is from Venus. *The Journal of Portfolio Management* **33**:5, 134-144. [Crossref]

636. Maik Eisenbeiss, Göran Kauermann, Willi Semmler. 2007. Estimating Beta-Coefficients of German Stock Data: A Non-Parametric Approach. *The European Journal of Finance* **13**:6, 503-522. [Crossref]

637. Michael E. Drew, Alastair Marsden, Madhu Veeraraghavan. 2007. Does Idiosyncratic Volatility Matter? New Zealand Evidence. *Review of Pacific Basin Financial Markets and Policies* **10**:03, 289-308. [Crossref]

638. Y. Malevergne, D. Sornette. 2007. Self-consistent asset pricing models. *Physica A: Statistical Mechanics and its Applications* **382**:1, 149-171. [Crossref]

639. Richard Saito, Rodrigo de Losso da Silveira Bueno. 2007. Fundamentos teóricos e empíricos de apreçamento de ativos. *Revista de Administração de Empresas* **47**:2, 81-85. [Crossref]

640. Robert R. Prechter, Wayne D. Parker. 2007. The Financial/Economic Dichotomy in Social Behavioral Dynamics: The Socionomic Perspective. *Journal of Behavioral Finance* **8**:2, 84-108. [Crossref]

641. Michail Koubouros, Ekaterini Panopoulou. 2007. Intertemporal Market Risks and the Cross–Section of Greek Average Returns. *Journal of Emerging Market Finance* **6**:2, 203-227. [Crossref]

642. Malcolm Baker, Jeffrey Wurgler. 2007. Investor Sentiment in the Stock Market. *Journal of Economic Perspectives* **21**:2, 129-151. [Abstract] [View PDF article] [PDF with links]

643. Beverly J. Hirtle, Kevin J. Stiroh. 2007. The return to retail and the performance of US banks. *Journal of Banking & Finance* **31**:4, 1101-1133. [Crossref]

644. Graham Bornholt. 2007. Extending the capital asset pricing model: the reward beta approach. *Accounting & Finance* **47**:1, 69-83. [Crossref]

645. Najet Rhaiem ., Saloua Ben Ammou ., Anouar Ben Mabrouk .. 2007. Estimation of Capital Asset Pricing Model at Different Time Scales Application to French Stock Market. *The International Journal of Applied Economics and Finance* **1**:2, 79-87. [Crossref]

646. Roland Burgman, Göran Roos. 2007. The importance of intellectual capital reporting: evidence and implications. *Journal of Intellectual Capital* **8**:1, 7-51. [Crossref]

647. Yu Xiao. 2007. Chinese Stock Market Systematic Risk#An Empirical Study. *SSRN Electronic Journal* . [Crossref]

648. Magdalena Morgese Borys. 2007. Testing Multi-Factor Asset Pricing Models in the Visegrad Countries. *SSRN Electronic Journal* . [Crossref]

649. Peter Christoffersen, Kris Jacobs, Gregory Vainberg. 2007. Forward-Looking Betas. *SSRN Electronic Journal* . [Crossref]

650. Peter Blair Henry. 2007. Capital Account Liberalization: Theory, Evidence, and Speculation. *SSRN Electronic Journal* . [Crossref]

651. Joon Chae, Cheol Won Yang. 2007. Why an Asset Pricing Model Fails to Explain the Cross Section of Stock Returns in the Korean Market'. *SSRN Electronic Journal* . [Crossref]

652. Malcolm P. Baker, Jeffrey A. Wurgler. 2007. Investor Sentiment in the Stock Market. *SSRN Electronic Journal* . [Crossref]

653. Martin Grandes. 2007. The Cost of Equity Beyond Capm: Evidence from Latin American Stocks (1986-2004). *SSRN Electronic Journal* . [Crossref]

654. Baibing Li, Xiangkang Yin. 2007. Information and Capital Asset Pricing. *SSRN Electronic Journal* . [Crossref]

655. Madhusudan Karmakar. 2007. Asymmetric Volatility and Risk-return Relationship in the Indian Stock Market. *South Asia Economic Journal* **8**:1, 99-116. [Crossref]

656. Kevin J. Stiroh. 2006. New Evidence on the Determinants of Bank Risk. *Journal of Financial Services Research* **30**:3, 237-263. [Crossref]

657. Christine A. Botosan. 2006. Disclosure and the cost of capital: what do we know?. *Accounting and Business Research* **36**:sup1, 31-40. [Crossref]

658. Holger Daske, Günther Gebhardt. 2006. Zukunftsorientierte Bestimmung von Risikoprämien und Eigenkapitalkosten für die Unternehmensbewertung. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* **58**:4, 530-551. [Crossref]

659. Michelle L. Barnes, Jose A. Lopez. 2006. Alternative measures of the Federal Reserve Banks' cost of equity capital. *Journal of Banking & Finance* **30**:6, 1687-1711. [Crossref]

660. Holger Daske. 2006. Economic Benefits of Adopting IFRS or US-GAAP - Have the Expected Cost of Equity Capital Really Decreased?. *Journal of Business Finance <html_ent glyph="@amp;" ascii="&"/ > Accounting*, ahead of print060428035417008. [Crossref]

661. Holger Daske. 2006. Economic Benefits of Adopting IFRS or US-GAAP - Have the Expected Cost of Equity Capital Really Decreased?. *Journal of Business Finance <html_ent glyph="@amp;" ascii="&"/ > Accounting* **33**:3-4, 329-373. [Crossref]

662. Holger Daske, Günther Gebhardt, Stefan Klein. 2006. Estimating the Expected Cost of Equity Capital Using Analysts' Consensus Forecasts. *Schmalenbach Business Review* **58**:1, 2-36. [Crossref]

663. Graham N. Bornholt. 2006. Extending the CAPM: The Reward Beta Approach. *SSRN Electronic Journal* . [Crossref]

664. Graham N. Bornholt. 2006. Expected Utility and Mean-Risk Asset Pricing Models. *SSRN Electronic Journal* . [Crossref]

665. Peter L. Bernstein. 2005. Capital Ideas: From the Past to the Future. *Financial Analysts Journal* **61**:6, 55-59. [Crossref]

666. Sudipta Basu. 2005. Discussion—The Effect of Risk on Price Responses to Unexpected Earnings. *Journal of Accounting, Auditing & Finance* **20**:4, 483-494. [Crossref]

667. Ho-Chuan (River) Huang *, Pei-Shan Wu. 2005. Tests of the CAPM with structural instability and asymmetry. *Applied Financial Economics Letters* **1**:5, 321-327. [Crossref]

668. Michelle L. Barnes, Jose A. Lopez. 2005. Alternative Measures of the Federal Reserve Banks' Cost of Equity Capital. *SSRN Electronic Journal* . [Crossref]

669. Vijay Gondhalekar, C.R. Narayanaswamy, Sridhar Sundaram. The Long-Term Risk Effects of the Gramm-Leach-Bliley Act (GLBA) on the Financial Services Industry 361-377. [Crossref]

The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXI, NO. 1 • FEBRUARY 2016

Does Academic Research Destroy Stock Return Predictability?

R. DAVID MCLEAN and JEFFREY PONTIFF*

ABSTRACT

We study the out-of-sample and post-publication return predictability of 97 variables shown to predict cross-sectional stock returns. Portfolio returns are 26% lower out-of-sample and 58% lower post-publication. The out-of-sample decline is an upper bound estimate of data mining effects. We estimate a 32% (58%–26%) lower return from publication-informed trading. Post-publication declines are greater for predictors with higher in-sample returns, and returns are higher for portfolios concentrated in stocks with high idiosyncratic risk and low liquidity. Predictor portfolios exhibit post-publication increases in correlations with other published-predictor portfolios. Our findings suggest that investors learn about mispricing from academic publications.

FINANCE RESEARCH HAS UNCOVERED many cross-sectional relations between predetermined variables and future stock returns. Beyond their historical insights, these relations are relevant to the extent that they provide insights into the future. Whether the typical relation continues outside a study's original sample is an open question, the answer to which can shed light on why cross-sectional return predictability is observed in the first place.[1] Although several papers

*R. David McLean is at DePaul University and Jeffrey Pontiff is at Boston College and is an unpaid director at a non-profit called the Financial Research Association. We are grateful to the Q Group, the Dauphine-Amundi Chair in Asset Management, and SSHRC for financial support. We thank participants at the Financial Research Association's 2011 early ideas session, Auburn University, Babson College, Bocconi University, Boston College, Brandeis University, CKGSB, Georgia State University, HBS, HEC Montreal, MIT, Northeastern University, Simon Fraser, University of Georgia, University of Maryland, University of South Carolina, University of Toronto, University of Wisconsin, Asian Bureau of Finance and Economic Research Conference, City University of Hong Kong International Conference, Finance Down Under Conference 2012, Wilfred Laurier, University of Washington Summer Conference, European Finance Association (Copenhagen), 1st Luxembourg Asset Management Conference, Ivey Business School, and Pontificia Universidad Catholica de Chile and Pierluigi Balduzzi, Turan Bali, Brad Barber, Mark Bradshaw, David Chapman, Shane Corwin, Alex Edmans, Lian Fen, Wayne Ferson, Francesco Franzoni, Xiaohui Gao, Thomas Gilbert, Robin Greenwood, Bruce Grundy, Cam Harvey, Clifford Holderness, Darren Kisgen, Owen Lamont, Borja Larrain, Juhani Linnainmaa, Jay Ritter, Ronnie Sadka, Paul Schultz, Andrei Shleifer, Ken Singleton, Bruno Skolnik, Jeremy Stein, Noah Stoffman, Matti Suominen, Allan Timmermann, Michela Verado, Artie Woodgate, Jianfeng Yu, William Ziemba, three anonymous referees, and an anonymous Associate Editor for helpful comments.

[1] Similar to Mittoo and Thompson's (1990) study of the size effect, we use a broad set of predictors to focus on out-of-sample cross-sectional predictability. For an analysis of the performance of

DOI: 10.1111/jofi.12365

note whether a specific cross-sectional relation continues out-of-sample, no study compares in-sample returns, post-sample returns, and post-publication returns for a large sample of predictors. Moreover, previous studies produce contradictory messages. As examples, Jegadeesh and Titman (2001) show that the relative returns to high momentum stocks increased after the publication of their 1993 paper, whereas Schwert (2003) argues that, since the publication of the value and size effects, index funds based on these variables fail to generate alpha.[2]

In this paper, we synthesize information for 97 characteristics shown to predict cross-sectional stock returns in peer-reviewed finance, accounting, and economics journals. Our goal is to better understand what happens to return predictability outside a study's sample period. We compare each predictor's returns over three distinct periods: (i) the original study's sample period, (ii) the period after the original sample but before publication, and (iii) the post-publication period. Previous studies attribute cross-sectional return predictability to statistical biases, rational pricing, and mispricing. By comparing the return predictability of the three periods, we can better differentiate between these explanations.

A. Statistical Bias

If return predictability in published studies results solely from statistical biases, then predictability should disappear out of sample. We use the term "statistical biases" to describe a broad array of biases inherent to research. Fama (1991, p. 1585) addresses this issue when he notes that "With many clever researchers on both sides of the efficiency fence, rummaging for forecasting variables, we are sure to find instances of 'reliable' return predictability that are in fact spurious." To the extent that the results of the studies in our sample are driven by such biases, we should observe a decline in return predictability out-of-sample.

B. Rational Expectations Versus Mispricing

Differences between in-sample and post-publication returns can be determined by both statistical biases and the extent to which investors learn from

out-of-sample time-series predictability, see LeBaron (2000) and Goyal and Welch (2008). For an analysis of cross-sectional predictability using international data, see Fama and French (1998), Rouwenhorst (1998), and McLean, Pontiff, and Watanabe (2009). For an analysis of calendar effects, see Sullivan, Timmermann, and White (2001).

[2] Lewellen (2014) uses 15 variables to produce a singular rolling cross-sectional return proxy and shows that it predicts, with decay, next period's cross section of returns. Haugen and Baker (1996) and Chordia, Subrahmanyan, and Tong (2013) compare characteristics in two separate subperiods. Haugen and Baker show that each of their characteristics produces statistically significant returns in their second subperiod, whereas Chordia, Subrahmanyan, and Tong show that none of their characteristics are statistically significant in their second subperiod. Green, Hand, and Zhang (2013) identify 300 published and unpublished characteristics but they do not estimate characteristic decay parameters as a function of publication or sample-end dates.

the publication. Cochrane (1999, p. 71) explains that, if predictability reflects risk, it is likely to persist "Even if the opportunity is widely publicized, investors will not change their portfolio decisions, and the relatively high average return will remain." Cochrane's logic follows Muth's (1961) rational expectations hypothesis, and thus can be broadened to nonrisk models such as Amihud and Mendelson's (1986) transaction-based model and Brennan's (1970) tax-based model. If return predictability reflects only rational expectations, then publication will not convey information that induces a rational agent to behave differently. Thus, once the impact of statistical bias is removed, pre- and post-publication return predictability should be equal.

If return predictability reflects mispricing and publication leads sophisticated investors to learn about and trade against the mispricing, then we expect the returns associated with a predictor should disappear or at least decay after the paper is published.[3] Decay, as opposed to disappearance, will occur if frictions prevent arbitrage from fully eliminating mispricing. Examples of such frictions include systematic noise trader risk (Delong et al. (1990)) and idiosyncratic risk and transaction costs (Pontiff (1996, 2006)). These effects can be magnified by principal-agent problems between investors and investment professionals (Shleifer and Vishny (1997)).[4]

C. Findings

We conduct our analysis using 97 different predictors from 79 different academic studies. We use long-short portfolio strategies that simultaneously buy and sell extreme quintiles that are based on each predictor. The average predictor's long-short return declines by 26% out-of-sample. This is an upper bound on the effect of statistical biases, since some traders are likely to learn about the predictor before publication, and their trading will cause the return decay to be greater than the pure decay from statistical bias.

The average predictor's long-short return shrinks 58% post-publication. Combining this finding with an estimated statistical bias of 26% implies a lower bound on the publication effect of about 32%. We can reject the hypothesis that return predictability disappears entirely, and we can also reject the hypothesis that post-publication return predictability does not change. This post-publication decline is robust to a general time trend, to time indicators used by other authors, and to time fixed effects.

The decay in portfolio returns is larger for predictor portfolios with higher in-sample returns and higher in-sample t-statistics. We also find that the decay is larger for predictors that can be constructed with only price and trading data and therefore are likely to represent violations of weak-form market efficiency.

[3] We do not distinguish between mispricing and "risk-reward deals" since both are inconsistent with rational expectations. Liu et al. (2014) develop a model of risk-reward deals and learning that is a framework for our findings.

[4] For evidence of limited arbitrage in short sellers and mutual funds, see Duan, Hu, and McLean (2009, 2010).

Post-publication returns are lower for predictors that are less costly to arbitrage, that is predictor portfolios more concentrated in liquid stocks and low idiosyncratic risk stocks. Our findings are consistent with mispricing accounting for some or all of the original return predictability, and investors learning about this mispricing.

We further investigate the effects of publication by studying traits that reflect trading activity. We find that stocks within the predictor portfolios observe post-publication increases in trading volume, and that the difference in short interest between stocks in the short and long sides of each portfolio increases after publication. These findings are consistent with the idea that academic research draws attention to predictors.[5]

Publication also affects the correlations between predictor portfolio returns. Yet-to-be-published predictor portfolios returns are correlated, and after a predictor is featured in a publication its portfolio return correlation with other yet-to-be-published predictor portfolios decreases while its correlation with already-published predictor portfolio returns increases. One interpretation of this finding is that some portion of predictor portfolio returns results from mispricing, and mispricing has a common source. This is why in-sample predictor portfolios returns are correlated. This interpretation is consistent with the irrational comovement models proposed in Lee, Shleifer, and Thaler (1991) and Barberis and Shleifer (2003). Publication could then cause more arbitrageurs to trade on the predictor, causing predictor portfolios to become more correlated with already published predictor portfolios that are also pursued by arbitrageurs, and less correlated with yet-to-be-published predictor portfolios.

Our findings are related to contemporaneous research that investigates how the magnitude of sophisticated capital affects anomaly returns (Hanson and Sundareram (2014), Kokkonen and Suominen (2014), Akbas et al. (2014)). Unlike these papers, we do not consider proxies for variation in sophisticated capital levels. Rather, our results suggest that academic publications transmit information to sophisticated investors.

The paper is organized as follows. In Section I we describe our research method. In Section II we describe our anomaly sample and discuss some summary statistics. Section III presents the main empirical findings. We conclude in Section IV.

I. Research Method

We begin by identifying studies that find cross-sectional relations between observable variables and future stock returns. We do not study time-series predictability. We limit ourselves to studies in peer-reviewed finance, accounting, and economics journals in which the null of no return predictability is rejected

[5] Drake, Rees, and Swanson (2011) demonstrate that short interest is more pronounced in the low-return segment of several characteristic-sorted portfolios. Their study does not estimate the difference between in- and out-of-sample short interest.

at the 5% level. We also require that the predicting variable be constructed with publicly available data. The studies were mostly identified with search engines such as Econlit by searching for articles in finance and accounting journals using words such as "cross-section." Some studies were identified in reference lists in books or other papers. We also contacted other finance professors and inquired about predictive variables we may have missed.

Most studies that we identify demonstrate cross-sectional predictability with either Fama-MacBeth (1973) slope coefficients or long-short portfolio returns. Some of the studies demonstrate a univariate relation between the given variable and subsequent returns, whereas other studies include additional control variables. Some studies that we identify are not truly cross-sectional, but instead present event study evidence that seems to imply a cross-sectional relation. Since we expect the results of these studies to be useful to investors, we include them in our analyses.

Our search process identifies 79 different studies. Based on these studies, we examine 97 cross-sectional relations. The various predictors and their associated studies are detailed in the paper's Internet Appendix.[6] We include all variables that relate to cross-sectional returns, including those with strong theoretical motivation such as Fama and MacBeth's (1973) market beta and Amihud's (2002) liquidity measure.

Our goal is not to perfectly replicate the findings in each paper. This is impossible since CRSP data change over time and papers often omit details about precise calculations. Moreover, in some cases we are unable to exactly reconstruct a given predictor. In such cases, we calculate a characteristic that captures the intent of the study. As an example, Franzoni and Marin (2006) show that a pension funding variable predicts future stock returns. This variable is no longer covered by Compustat, so with the help of the paper's authors we use available data from Compustat to construct a similar variable that we expect to contain much of the same information. As another example, Dichev and Piotroski (2001) show that firms that are downgraded by Moody's experience negative future abnormal returns. Compustat does not cover Moody's ratings but does cover S&P ratings, so we use S&P rating downgrades.

For some characteristics such as momentum, higher characteristic values are associated with higher returns, while for other characteristics such as size, higher characteristic values are associated with lower returns. We form long-short portfolios based on the extreme 20$^{\text{th}}$ percentiles of the characteristic. The long side is the side with the higher returns as documented by the original publication.

Sixteen of our predictors are indicator variables. For these cases, if the original paper demonstrates higher returns for firms assigned with the indicator, then these firms are included in the long-side portfolio, and an equal-weighted

[6] The Internet Appendix is available in the online version of the article on the *Journal of Finance* website.

portfolio of all other stocks is used to form the short-side portfolio. If the original paper demonstrates lower returns for indicated firms, then non-indicated firms form the long-side portfolio and the indicated firms form the short-side portfolio.

Three predictors are variables with three discrete values (long, short, and neutral). For example, Barth and Hutton (2004) develop a strategy of buying low accrual stocks with increases in analysts' earnings forecasts, and selling low accrual stocks with decreases in earnings forecasts. In cases like this, our long-short portfolio follows the original paper. We provide detailed descriptions of all 97 predictors in the paper's Internet Appendix.

The average correlation across the returns of the 97 predictor portfolios is 0.033. This finding is consistent with Green, Hand, and Zhang (2013), who report an average correlation of 0.09 among 60 quantitative portfolios. There are of course both higher and negative correlations among the predictors in our sample. As we explain below, we explicitly control for such cross-correlations when computing the standard errors of our test statistics.

In an earlier version of the paper we also calculate monthly Fama-MacBeth (1973) slope coefficient estimates using a continuous measure of the characteristic (e.g., firm size or past returns). As Fama (1976) shows, Fama-MacBeth (1973) slope coefficients are returns from long-short portfolios with unit net exposure to the characteristic. We obtain similar findings using both methods, so for the sake of brevity we only report quintile returns.

II. Important Dates and Summary Statistics

We segment periods based on both the end-of-sample date and the publication date. We do so because these are easily identifiable dates that may be associated with changes in predictability. The end of the original sample provides a clear demarcation for estimating statistical bias. The publication date, in contrast, provides only a proxy for when market participants learn about a predictor. As we mention above, we assume that more investors know about a predictor after the publication date as compared to before the publication date. However, some market participants may not read the paper until years after publication. Post-publication decay in return predictability may therefore be a slow process. We are unaware of theories on how long the decay should take or on the functional form of the decay. Despite the simplicity of our approach, the publication date generates robust estimates of return decay.

Table I provides summary statistics for the predictor portfolios that we study. For the 97 portfolios, the average monthly in-sample return is 0.582%. The average out-of-sample, pre-publication return is 0.402%, whereas the average post-publication return is 0.264%. Returns are equal-weighted unless the primary study presents value-weighted portfolio results as its primary finding, and the only study in our sample that does this is Ang et al. (2006).

The average length of time between the end-of-the sample and publication dates is 56 months. In comparison, the average original in-sample span is 323

Table I
Summary Statistics

This table reports summary statistics for the predictor portfolios studied in this paper. The returns are equal-weighted by predictor portfolio, that is, we first estimate the statistic for each predictor portfolio, and then take an equal-weighted average across the portfolios. The reported standard deviations are the standard deviations of the predictors' mean returns. Our sample period ends in 2013.

Number of predictor portfolios	97
Predictors portfolios with t-statistic > 1.5	85 (88%)
Mean publication year	2000
Median publication year	2001
Predictors from finance journals	68 (70%)
Predictors from accounting journals	27 (28%)
Predictors from economics journals	2 (2%)
Mean portfolio return in-sample	0.582
Standard deviation of mean in-sample portfolio return	0.395
Mean observations in-sample	323
Mean portfolio return out-of sample	0.402
Standard deviation of mean out-of-sample portfolio return	0.651
Mean observations out-of-sample	56
Mean portfolio return post-publication	0.264
Standard deviation of mean post-publication portfolio return	0.516
Mean observations post-publication	156

months, and the average post-publication span is 156 months. Our sample ends in 2013.

The publication date is determined by the year and month on the cover of the journal. We consider two variations. A previous version of this paper considers publication dates based on arrival time stamps at Boston metropolitan libraries. This variation produced nearly identical results. Another version considers the publication date to be the earlier of the actual publication date and the first time the paper appears on the SSRN. The average number of months between the end-of-sample and SSRN dates is 44 months, and we again obtain the same results.

Although we include all 97 predictors in our tests, 12 of our predictors produce portfolio returns with in-sample t-statistics that are less than 1.50. Thus, a total of 85 ($= 97 - 12$) or 88% of the predictors produce t-statistics that are greater than 1.50. With respect to the 12 predictors that do not reach this significance level, in some cases the original paper reports event study abnormal returns that do not survive in our portfolio sorts. In other cases, we do not have the same data used by the original authors. Portfolio formation also contributes to differences in statistical significance. We focus on long-short quintile returns, while some of the original papers that demonstrate predictability use Fama-MacBeth (1973) slope coefficients or buy-and-hold returns.

III. Empirical Analyses and Results

A. Portfolio Returns Relative to End-of-Sample and Publication Dates

In this section we formally study the returns of each predictor relative to its sample-end and publication dates. Our baseline regression model is described in equation (1):

$$R_{it} = \alpha_i + \beta_1\ Post\ Sample\ Dummy_{i,t} + \beta_2\ Post\ Publication\ Dummy_{i,t}\ +\ e_{it}. \tag{1}$$

In equation (1), the dependent variable is the monthly return for predictor i in month t. The post-sample dummy is equal to one if month t is after the end of the original sample but still pre-publication, and zero otherwise, whereas the post-publication dummy is equal to one if the month is post-publication and zero otherwise. The variable α_i captures predictor fixed effects.

As we mention previously, correlations across predictor portfolios are low, averaging only 0.033. However, there is variation in the correlations, with some portfolios being highly correlated and others being uncorrelated. We therefore compute our standard errors using feasible generalized least squares (FGLS) under the assumption of contemporaneous cross-correlation between returns. Clustering on time (as in previous drafts) produces similar results, with slightly smaller standard errors in most cases.

The post-sample coefficient, β_1, estimates the impact of statistical biases on predictor in-sample performance. This is an upper bound estimate, as it could be the case that sophisticated traders are aware of the working paper before publication. The post-publication coefficient, β_2, estimates both the impact of statistical biases and the impact of publication. If statistical biases are the source of in-sample predictability, then the coefficients on both the post-sample and post-publication dummies should be -0.582, which is the negative of the average in-sample mean return (reported in Table I). Such a finding would be consistent with Fama's (1991) conjecture that much of the return predictability in academic studies is the outcome of data mining. In contrast, if predictors' returns are entirely the result of mispricing and arbitrage resulting from publication corrects all mispricing, then the post-publication coefficient should be equal to -0.582 and the post-sample dummy should be close to zero. In the remaining extreme, if there are no statistical biases and academic papers have no influence on investors' actions, then both of the coefficients should equal zero.

B. Predictor Return Dynamics Relative to End-of-Sample and Publication Dates

Table II presents regression estimates of how predictability changes out-of-sample and post-publication. Column (1) reports the results for our main specification, which estimates equation (1) on our full sample of 97 predictors. The post-sample coefficient in this regression is -0.150%, and statistically

Table II

Regression of Predictor Portfolio Returns on Post-Sample and Post-Publication Indicators

The regressions test for changes in returns relative to the predictor's sample-end and publication dates. The dependent variable is the monthly return to a long-short portfolio that is based on the extreme quintiles of each predictor. *Post-Sample (S)* is equal to one if the month is after the sample period used in the original study and zero otherwise. *Post-Publication (P)* is equal to one if the month is after the official publication date and zero otherwise. Mean is the in-sample mean return of the predictor portfolio during the original sample period. t-statistics are the in-sample t-statistic of each predictor portfolio. Standard errors (in parentheses) are computed under the assumption of contemporaneous cross-sectional correlation between panel portfolio residuals. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The bottom three rows report p-values from tests of whether post-sample and post-publication changes in returns are statistically different from one another and whether any declines are 100% of the in-sample mean (the effects disappears entirely).

Variables	(1)	(2)	(3)	(4)
Post-Sample (S)	−0.150***	−0.180**	0.157	0.067
	(0.077)	(0.085)	(0.103)	(0.112)
Post-Publication (P)	−0.337***	−0.387***	−0.002	−0.120
	(0.090)	(0.097)	(0.078)	(0.114)
S × Mean			−0.532***	
			(0.221)	
P × Mean			−0.548***	
			(0.178)	
S × t-statistic				−0.061***
				(0.023)
P × t-statistic				−0.063***
				(0.018)
Predictor FE?	Yes	Yes	Yes	Yes
Observations	51,851	45,465	51,851	51,944
Predictors (N)	97	85	97	97
Null : S = P	0.024	0.021		
Null: P = −1 × (mean)	0.000	0.000		
Null: S = −1 × (mean)	0.000	0.000		

significant. Thus, our best estimate of the post-sample decline is 15.0 bps. The post-publication coefficient is −0.337, and it is also statistically significant. These results show that, on average, predictor portfolios are 33.7 bps lower post-publication compared to before publication. Table I shows that the average predictor has an in-sample mean return of 58.2 bps per month. Hence, post-sample and post-publication returns decline relative to the in-sample mean by 26% and 58%, respectively.

The regression in the second column includes only 85 predictors. It excludes the 12 predictors that generate t-statistics with values less than 1.5. Exclusion of these predictors does not change the basic inference reported in column (1). The post-sample and post-publication coefficients are −0.180 and −0.387, respectively, in column (2), similar to the results in column (1). The average in-sample return for the 85 predictors is 0.652, so the post-publication decay in percentage terms is similar if these other 12 predictors are included. The average in-sample return is larger when we exclude the 12 predictors because we are excluding the 12 predictors that lack significant in-sample predictability.

At the bottom of Table II, we report tests of whether the post-publication coefficient and out-of-sample but pre-publication coefficient are equal. In both of the regressions described above, the coefficients are significantly different at the 5% level. This difference tells us that there is an effect associated with publication that cannot be explained by statistical biases, that should be fully reflected in the out-of-sample but pre-publication coefficient.

The bottom of Table II also reports tests of whether predictor portfolio returns disappear entirely post-publication. This test is generated from a linear restriction that equates the post-publication coefficient to the average sum of the fixed effects and the intercept.[7] This test, along with the *t*-test on the post-publication coefficient, allows us to easily reject both nulls, that is, we reject the null that anomaly returns decay entirely post-publication, and we reject the null that they do not decay.

The regression in the third column includes the predictor fixed effects along with interactions between the in-sample mean return of each predictor and the out-of-sample and post-publication dummy variables. The interactions test whether predictor portfolio returns with higher in-sample means decline more post-publication. We do not include the in-sample mean in the regression by itself because it does not vary over time and we include predictor fixed effects.

In column (3), the coefficient on the post-sample dummy is 0.157, whereas the coefficient on the interaction between the post-sample dummy and the in-sample mean is −0.532. As we mention above, the average in-sample monthly return for the 97 portfolios is 0.582% (see Table I), so the overall post-sample effect is $0.157 + (-0.532 \times 0.582) = -0.153$, similar to the post-sample coefficient in column (1). The standard deviation of the in-sample mean return is 0.395 (see Table I). Hence, a portfolio with an in-sample mean return that is one standard deviation more than average has a $-0.532 \times 0.395 = -0.210$ bp decline in post-sample monthly return. This could reflect predictors with larger in-sample returns having a larger degree of statistical bias. Alternatively, it could reflect arbitrageurs being more likely to learn about and trade on predictors with higher returns before publication. Similarly, the the post-publication dummy is −0.002, and the interaction between the post-sample dummy and in-sample mean is −0.548. This relation is also displayed in Figure 1 (Panel A), which plots the average in-sample mean for each predictor against its post-publication decline, and shows that predictors with larger in-sample returns have greater post-publication declines.

The final regression in Table II interacts the post-sample and post-publication dummies with a predictor's in-sample *t*-statistic. The average

[7] The expected return of a predictor in-sample is the sum of the regression intercept and the predictor's fixed effect. We take the average of these sums, which is equal to the average predictor's in-sample return. We then test whether this value minus the coefficient on either publication or post-sample is equal to zero.

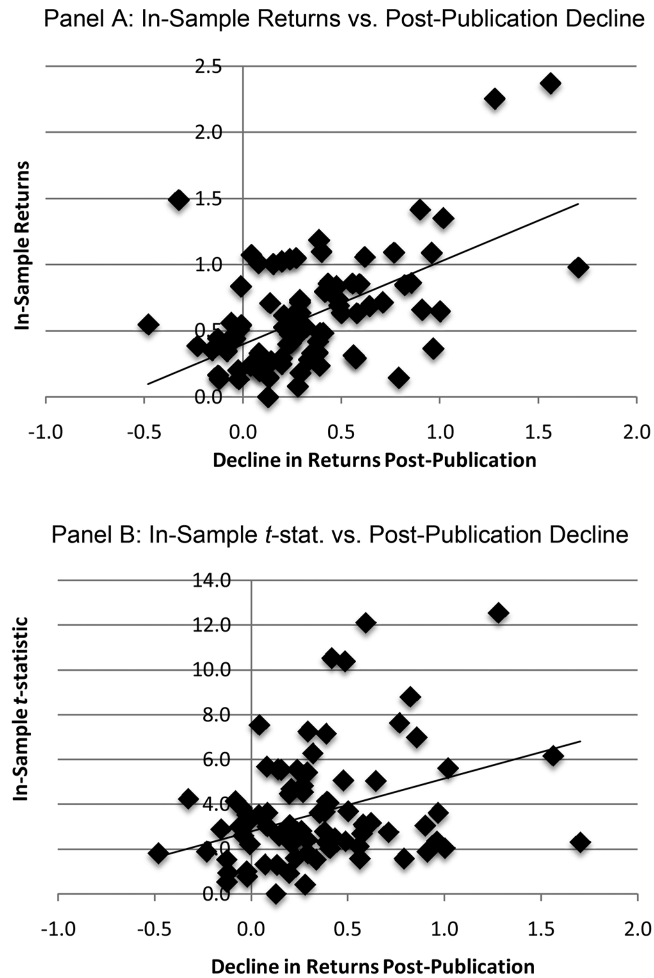


Figure 1. The relation between in-sample returns and post-publication decline in returns. Panel A plots the relation between in-sample returns and post-publication declines in returns. For each predictor, we estimate the mean return to a long-short portfolio that contemporaneously buys and sells the extreme quintiles during the sample period of the original study. We then estimate the mean return for the period after the paper is published through 2013. To be included in the figure, a predictor's in-sample return has to generate a t-statistic greater than 1.5; 85 of the 97 predictors that we examine meet this criterion. The predictor also has to have at least three years of post-publication return data. This excludes 10 of the 85 predictors, resulting in a sample of 75 predictors. Panel B repeats this exercise, but it plots in-sample t-statistic against post-publication declines. The returns are reported in percent, e.g., 1.5 is a monthly return of 1.5%.

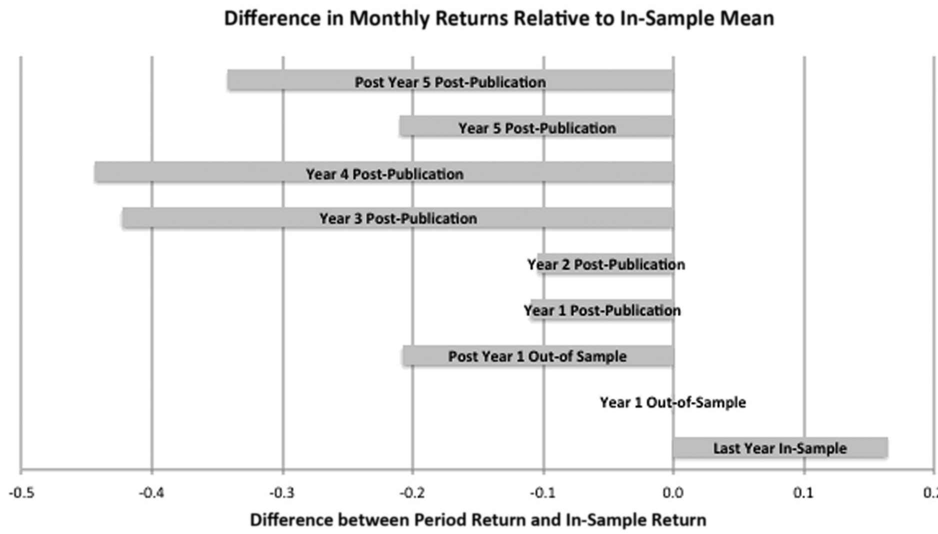


Figure 2. Predictor return dynamics around the sample-end and publication dates. This figure explores changes in predictability by examining finer post-sample and post-publication partitions. The figure plots the coefficients from a regression containing dummy variables that signify the last 12 months of the original sample, the first 12 months out-of sample, and the other out-of-sample months. In addition, the publication dummy is split into six different variables, namely, one dummy for each of the first five years post-publication and one dummy for all of the months that are at least five years after publication. The returns are reported in percent, e.g., 1.5 is a monthly return of 1.5%.

in-sample t-statistic is 3.55 and the standard deviation of the t-statistics is 2.39. Hence, the regression estimates an incremental decline for a characteristic portfolio with a t-statistic that is one standard deviation higher than the average of −0.146 post-sample and −0.151 post-publication. The post-publication effect is plotted in Figure 1 (Panel B). The results here are consistent with the idea that arbitrageurs devote more capital to characteristic portfolios that are associated with higher in-sample returns. In an untabulated specification we condition decay on in-sample Sharpe ratios and estimate very similar results.

Previous versions of the paper consider whether decay is related to the cumulative number of academic citations to the publication that introduces the portfolio returns associated with the predictor. After controlling for the publication date, this measure has little incremental value in explaining decay.

C. A Closer Look at Predictor Return Dynamics around the End-of-Sample and Publication Dates

Figure 2 further considers changes in predictability by examining finer post-sample and post-publication partitions. The figure plots the coefficients from a regression of predictor returns on dummy variables for the last 12 months of the original sample, the first 12 months out-of sample, and the remaining out-of-sample months. In addition, the publication dummy is split into six variables, namely, a dummy variable for each of the first five years post-publication and a dummy variable for all of the months that are at least five years after publication. Some caution is needed in interpreting this figure. Although the estimates in this figure are interesting, statistical power is lower from partitioning the results, and theory does not guide us regarding the appropriate partitions.

The publication process often takes years. This gives unscrupulous researchers an opportunity to choose their sample-ends so as to report stronger results. Figure 2 shows that the coefficient on the dummy for the last 12 months of the original sample is positive, which means that the last 12 months of the sample have higher returns than the other in-sample months, which could be consistent with researchers ending their samples opportunistically. However, the coefficient on the dummy for the first 12 months post-sample is virtually zero, which indicates that the first 12 months post-sample have on average the same returns as the in-sample months. If authors were selectively choosing their sample periods, then this coefficient should be negative.

Figure 2 shows that, after the first 12 months out-of-sample, returns are lower as compared to in-sample. After the first 12 months post-sample and during the remaining out-of-sample months but before publication, returns are more than 20 basis points lower. Returns remain at this level the first two years post-publication, and then decay further: In the third year we estimate a decay of 40.8 bps; in the fourth year decay is 43.3 bps; and in the fifth year decay is 20.5 bps. After the fifth year post-publication predictors returns are on average 33.9 bps lower as compared to in-sample.

One might suggest that we examine post-publication returns as a function of the predictor's persistence (i.e., how often the portfolio turns over). Initial decay may be muted if new capital flows into portfolios that are determined by a persistent predictor. For example, new flows into high book-to-market stocks might cause a temporary increase in the returns of book-to-market portfolios, whereas portfolios based on less persistent predictors, such as last month's stock return, would not generate such an effect. We consider this possibility in an earlier version of the paper. We find some evidence that portfolio returns to more persistent predictors decay less following publication, but the effect is not statistically significant.

D. Controlling for Time Trends and Persistence

It could be the case that the dissemination of academic research has no effect on return predictability, and that our end-of-sample and publication coefficients reflect a time trend or a trend that proxies for lower costs of corrective trading. For example, anomalies' returns may drop post-publication if anomalies reflect mispricing and declining trading costs have made arbitrage less costly (see Goldstein et al. (2009) and Anand et al. (2012)). Consistent with this idea, Chordia, Subrahmanyam, and Tong (2013) show that the returns of the 12 anomalies decline after 1993, which they attribute to an increase in hedge funds and lower trading costs.

To examine the possibility that our results reflect a time effect and not a publication effect, we construct a time variable that is equal to 1/100 in January 1926 and increases by 1/100 each month in our sample. Tale III presents the results. In column (1), we estimate a regression of monthly portfolio returns on the time variable and predictor fixed effects. The time variable produces a negative slope coefficient that is significant at the 1% level, which is consistent with the idea that portfolio returns have declined over time.

In column (2), we estimate the effect of a dummy variable that is equal to one if the year is after 1993 and zero otherwise. We use this specification because, as we mention above, Chordia, Subrahmanyam, and Tong (2013) show that 12 predictors have lower returns after 1993. The post-1993 coefficient is negative but insignificant in our sample.

In column (3), we relate decay to a time trend, the post-1993 indicator, and the post-sample and post-publication indicator variables. The time trend variable is still negative and significant, however the post-1993 dummy variable is now *positive* and statistically significant. The post-publication coefficient is −0.362 and statistically significant, similar to the estimate reported in our main specification in Table II. Thus, adding a time trend and a 1993 break has little impact on post-publication return decay.

An alternative way to control for time effects is to include time fixed effects. Time fixed effects demean each monthly anomaly return by the average anomaly return in the same month. Hence, including time fixed effects allows for parameter estimation that is free from all forms of time-series decay. Column (4) of Table III reports an estimation that includes time fixed effects. The estimated coefficients are very close to those in Table II. In particular, predictor returns are 17.9 bps lower out-of-sample and 31.0 bps lower post-publication, both significant at the 5% level. Based on the average in-sample return of 58.2 bps, this specification implies a sizeable 53% drop in post-publication predictability after all time effects have been removed.

In the final two regressions in Table III, we test whether predictor returns are persistent, and whether controlling for persistence changes the publication effect. Recent work by Moskowitz, Ooi, and Pedersen (2013) and Asness, Moskowitz, and Pedersen (2013) finds broad momentum across asset classes and correlation of momentum returns across classes, whereas Grundy and Martin (2001) fail to find significant momentum in the Fama-French factors. We include the predictor's prior-month's return and the sum of its returns over the last 12 months' returns in columns (5) and (6), respectively. Both of the lagged return coefficients are positive and significant, which is broadly consistent with Moskowitz, Ooi, and Pedersen (2013). The post-publication coefficient remains significant in each of these regressions, suggesting a post-publication decline of about 25 to 30 bps once persistence is controlled for.

E. Do Returns and Post-Publication Decay Vary across Predictor Types?

In this section, we group predictors into four broad categories and examine variations in pre-publication returns, post-publication returns, and the

Table III

Time Trends and Persistence in Predictor Returns

The regressions reported in this table test for time trends and persistence in predictor returns. *Post-Sample (S)* is equal to one if the month is after the sample period used in the original study and zero otherwise. *Post-Publication (P)* is equal to one if the month is after the official publication date and zero otherwise. *Time* is the number of months divided by 100 post-January 1926. *Post-1993* is equal to one if the year is greater than 1993 and zero otherwise. *1-Month Return* and *12-Month Return* are the predictor's return from the last month and the cumulative return over the last 12 months. Standard errors (in parentheses) are computed under the assumption of contemporaneous cross-sectional correlation between panel portfolio residuals. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Time	−0.069***		−0.069***			
	(0.011)		(0.026)			
Post-1993		−0.120	0.303***			
		(0.074)	(0.118)			
Post-Sample			−0.190**	−0.179**	−0.132*	−0.128
			(0.081)	(0.080)	(0.076)	(0.078)
Post-Publication			−0.362***	−0.310**	−0.295***	−0.258***
			(0.124)	(0.122)	(0.089)	(0.093)
1-Month Return					0.114***	
					(0.015)	
12-Month Return						0.020***
						(0.004)
Observations	51,851	51,851	51,851	51,851	51,754	50,687
Char. FE?	Yes	Yes	Yes	Yes	Yes	Yes
Time FE?	No	No	No	Yes	No	No

post-publication return decay. We designate the predictor categories as (i) Event, (ii) Market, (iii) Valuation, and (iv) Fundamentals.

Event predictors are based on events within the firm, external events that affect the firm, and changes in firm performance. Examples of event predictors include share issues, changes in financial analyst recommendations, and unexpected increases in R&D spending. Market predictors are predictors that can be constructed using only financial data, such as volume, prices, returns, and shares outstanding. Momentum, long-term reversal, and market value of equity are included in our sample of market predictors. Valuation predictors are ratios, where one of the numbers reflects a market value and the other reflects fundamentals. Examples of valuation predictors include sales-to-price and book-to-market. Finally, fundamental predictors are constructed from financial statement data and analysts' expectations of financial statement data. Debt, taxes, and accruals (all scaled by total assets) are examples of fundamental predictors.

As we mention previously, the average correlation among the predictor portfolio returns is 0.033, whereas the median is 0.018. The correlation is not higher within the groups. Valuation predictor portfolios' returns have the highest within-group correlation, averaging 0.058, whereas market predictor portfolios

Table IV

Predictor Returns across Different Predictor Types

This table tests whether predictor returns and changes in returns post-publication vary across types of predictors. To conduct this exercise, we split our predictors into four groups: (i) event, (ii) market, (iii) valuation, and (iv) fundamentals. We regress monthly predictor returns on dummy variables that signify each predictor group. Each column reports how each predictor type differs from the other three types. The bottom two rows test whether post-publication returns for each predictor type are different than those of the other three types. Standard errors (in parentheses) are computed under the assumption of contemporaneous cross-sectional correlation between panel portfolio residuals. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Variable	(1)	(2)	(3)	(4)
Post-Publication (P)	−0.208***	−0.316***	−0.310***	−0.301***
	(0.059)	(0.097)	(0.080)	(0.089)
Market	0.304***			
	(0.079)			
P × Market	−0.244			
	(0.169)			
Event		−0.098**		
		(0.046)		
P × Event		0.105		
		(0.091)		
Valuation			−0.056	
			(0.063)	
P × Valuation			0.186	
			(0.131)	
Fundamental				−0.201***
				(0.045)
P × Fundamental				0.025
				(0.089)
Constant	0.482***	0.606***	0.585***	0.630***
	(0.036)	(0.052)	(0.000)	(0.053)
Observations	51,851	51,851	51,851	51,851
Predictors	97	97	97	97
Type + (P × Type)	0.060	0.007	0.121	−0.176
p-value	0.210	0.922	0.256	0.012

have the lowest, averaging 0.021. The reason for this is that there can be both very high and very low return correlations within each group. As an example, the highest correlation in our sample is 0.933, which is between the returns of the price and size portfolios. The lowest correlation is −0.895, which is between the returns of the price and 52-week high portfolios. Similarly, the momentum and price portfolios' returns have a correlation of −0.715. All of these predictors are market predictors. As in the previous tables, we estimate our standard errors via FGLS, which accounts for contemporaneous cross-correlations.

We formally test for differences between the four predictor portfolio groups in Table IV. Using all data, monthly returns are regressed on a dummy variable representing one of the four predictor types, a post-publication dummy, and the interaction between the post-publication and the predictor type dummy:

$$R_{i,t} = \alpha_{i,} + \beta_1 \, Post\ Publication\ Dummy_i + \beta_2 \, Predictor\ Type\ Dummy_i$$
$$+ \beta_3 \, Post\ Publication\ Dummy_i \times Predictor\ Type\ Dummy_i + e_{it}. \quad (2)$$

The coefficient on the *Predictor Type Dummy*, β_2, estimates whether the in-sample average returns of a group are different from those of the other groups. The results show that, compared to the other categories of predictors, market-based predictors have the highest pre-publication returns, while fundamental predictors have the lowest pre-publication returns.

The coefficient on the interaction, β_3, tests whether post-publication declines vary across predictor groups. The decline for the market-based predictor portfolio returns is largest, although it is not significantly different from the declines for the other predictors. Valuation predictor returns have the smallest declines post-publication.

Differences in post-publication expected returns are given by the sum of the type coefficient and the interaction coefficients $\beta_2 + \beta_3$. The sums and their associated p-values are reported in the bottom two rows of Table IV. Despite the high pre-publication returns of market-based predictors, post-publication market-based predictor returns are not significantly higher than those of the non-market-based predictors. This result is consistent with the results in Table II, which shows that predictors with higher in-sample returns have larger declines in returns post-publication. The bottom two rows also show that post-publication returns are significantly lower for fundamental predictors, so this pre-publication difference in returns is persistent post-publication.

F. Costly Arbitrage

The results above are consistent with the idea that publication attracts arbitrageurs, which results in lower returns post-publication. As we explain in the introduction, Pontiff (1996, 2006) and Shleifer and Vishny (1997) suggest that costs associated with arbitrage can prevent arbitrageurs from fully eliminating mispricing. By this logic, predictor portfolios concentrated in stocks that are costlier to arbitrage (e.g., smaller stocks, less liquid stocks, stocks with more idiosyncratic risk) should decline less post-publication. If predictor returns are the outcome of rational asset pricing, then the post-publication decline should not be related to arbitrage costs.[8]

Previous papers in the costly arbitrage literature relate arbitrage costs to differences in returns across stocks within a predictor portfolio (see Pontiff (2006), Duan, Hu, and McLean (2010), and McLean (2010)). In contrast, we estimate

[8] Our exercise recognizes that, if returns reflect mispricing, then, in equilibrium, portfolios that incur higher costs will deliver higher returns. This approach deviates from an earlier literature, such as Lesmond, Schill, and Zhou (2004) and Korajczyk and Sadka (2004), who question whether costs eliminate the excess return of a particular portfolio.

the relation between arbitrage costs and expected returns across (instead of within) portfolios. Another difference between our tests and the previous literature is that previous studies assume that the informed trader had knowledge of the predictor before (and after) the publication date. Our tests consider the possibility that publication informs arbitrageurs, which affects the decay in return predictability post-publication.

Our costly arbitrage variables include three transaction cost variables—size, bid-ask spreads, and dollar volume—and two holding cost variables—idiosyncratic risk and a dividend-payer dummy. We also create a costly arbitrage index, which is the first principal component of the five costly arbitrage variables.

Large stocks, stocks with high dollar volume, and stocks with low spreads are more liquid and therefore less costly to arbitrage. Hence, we expect long-short returns to be lower in predictor portfolios concentrated in such stocks. Firm size is measured as the market value of equity. Average monthly spreads are estimated from daily high and low prices using the method of Corwin and Schultz (2012). Dollar volume is the number of shares traded during the past month multiplied by the month-end stock price.

Idiosyncratic risk limits the amount that an investor will invest in a mispriced stock (Treynor and Black (1973) and Pontiff (1996, 2006)), and thus returns should be higher in predictor portfolios concentrated in high idiosyncratic risk stocks. We compute monthly idiosyncratic risk by regressing daily returns on the 12 value-weighted industry portfolios from Ken French's website. We estimate a regression for each stock using the last 24 months of daily data. For each day, we square that day's residuals and, to correct for autocorrelation, add two times the product of that day's and the previous day's residuals. The monthly idiosyncratic risk measure is created by cumulating the daily sum of residual products for a given month. If the industry factor model regression contains less than 30 observations, the stock is not assigned an idiosyncratic risk measure in that month.

Pontiff (1996, 2006) explains that dividends mitigate holding costs since they decrease the effective duration of the position. The intuition is that dividends reduce the future amount of capital devoted to arbitrage, thus reducing the cumulative holding costs.[9] We use a dummy variable equal to one if a firm paid a dividend and zero otherwise. We expect returns to be lower for predictor portfolios concentrated in stocks that pay dividends.

The costly arbitrage index is based on the first principal component of the five costly arbitrage variables. A higher value of the index is associated with lower arbitrage costs and therefore lower expected portfolio returns. The index has positive correlations with the size, dividends, and dollar volume variables, and negative correlations with the spreads and idiosyncratic risk variables.

[9] This result assumes that the level of the mispricing is unaffected by the dividend payout. The result also holds for the case in which the level of the mispricing is influenced by mispricing but the relative mispricing is not. For proof, see the Appendix in Pontiff (2006).

We estimate the arbitrage cost of each predictor portfolio as follows. First, for each month, we compute the average cross-sectional ranking for a trait (e.g., size or idiosyncratic risk) among all of the stocks in CRSP. Each stock-month observation is therefore assigned a ranking between zero and one. Next, for each month, we estimate the average rank of the stocks that are in either the long or the short sides of each predictor portfolio. This creates a time series of monthly rank-averages for each trait. We then take the average of each time-series to estimate a single costly arbitrage variable for each predictor. We only use in-sample months to create the costly arbitrage variables, as it could be the case that trading caused by publication affects the costly arbitrage variables. We then estimate the following regression,

$$R_{i,t} = \alpha_{i,} + \beta_1\ \mathit{Post\ Publication\ Dummy}_{i,t} + \beta_2\ \mathit{Arbitrage\ Cost}_i + \beta_3\ \mathit{Post\ Publication\ Dummy}_{i,t} \times \mathit{Arbitrage\ Cost}_i + e_{it}, \qquad (3)$$

where the dependent variable is a predictor's monthly return. We report the results in Table V. The results largely support the notion that some sophisticated traders exert price pressure pre-publication, but the price pressure is tempered by arbitrage costs. If some sophisticated traders implement predictor strategies pre-publication, then portfolios with higher arbitrage costs should have higher pre-publication returns. This effect is given by the slopes on the non-interacted arbitrage cost variables, β_2. Five of the costly arbitrage variables (including the index) have slopes with the expected sign, and all five are statistically significant. The dollar volume variable produces a slope in the opposite direction—predictor portfolios concentrated in stocks with high dollar volume of trading tend to have higher in-sample returns, although this effect is not statistically significant.

Post-publication knowledge of a predictor should be widespread, and we thus expect portfolios that are easier to arbitrage to have lower post-publication returns. The sum of the costly arbitrage coefficient, β_2, plus the coefficient on the interaction between the post-publication dummy and the arbitrage cost variable, β_3, should therefore reflect higher expected returns for predictors that are more costly to arbitrage. The sum of these coefficients and their associated p-values are presented in the last two rows of Table V. All six of these sums have the correct expected sign, and five of the six are statistically significant.

For brevity, we do not report a specification that simultaneously includes all five of the primary costly arbitrage variables and all five of the interactions. Caution is needed in interpreting such results due to high correlations between the right-hand-side variables. Regarding in-sample returns, idiosyncratic risk is the only costly arbitrage variable that commands a statistically significant slope with the expected sign. Post-publication, returns are lower for predictor portfolios that contain stocks with more idiosyncratic risk. The post-publication effects for spreads and size have the expected signs but are insignificant. Idiosyncratic risk's post-publication p-value is 0.000. These findings are consistent with Pontiff's (2006, p. 35) review of the literature that

Table V

Costly Arbitrage and the Persistence of Predictor Returns

These regressions test whether arbitrage costs are associated with declines in predictability post-publication. The dependent variable is a predictor portfolio's monthly long-short return. The independent variables reflect various traits of the stocks in each predictor portfolio. To measure the strength of the traits of the stocks within a portfolio, we first rank all of the stocks in CRSP on the trait (e.g., size or spreads), assigning each stock a value between zero and one based on its rank. We then take the average rank of all of the stocks in the portfolio for that month. Finally, we take an average of the predictor's monthly trait averages, using all of the months that are in-sample. Hence, in the size regression reported in the first column, the independent variable is the average market value rank of the stocks in the predictor's portfolio during the in-sample period for the predictor. Average monthly *Spreads* are the average monthly bid-ast Spreads estimated from daily high and low prices using the method of Corwin and Schultz (2012). *Dollar Volume* is shares traded multiplied by stock price. *Idiosyncratic Risk* is daily stock return variance, which is orthogonal to the market and industry portfolios. *Dividends* is a dummy equal to one if the firm paid a dividend during the last year and zero otherwise. *Index* is the first principal component of the other five measures. Standard errors (in parentheses) are computed under the assumption of contemporaneous cross-sectional correlation between panel portfolio residuals. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The bottom two rows test whether the sum of the costly arbitrage variable (*CA*) plus the interaction between the publication dummy and the costly arbitrage variable ($P \times CA$) is statistically different from zero.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Post-Pub. (P)	−0.190	−0.139	0.215	−0.242	−0.321	−0.264**
	(0.274)	(0.235)	(0.230)	(0.273)	(0.211)	(0.078)
P × Size	−0.138					
	(0.459)					
Size	−1.064**					
	(0.236)					
P × Spreads		−0.301				
		(0.603)				
Spreads		1.228**				
		(0.252)				
P × Dol.Vol.			−1.059*			
			(0.500)			
Dol. Vol.			0.215			
			(0.308)			
P × Idio. Risk				−0.047		
				(0.554)		
Idio. Risk				2.064***		
				(0.330)		
P × Div.					−0.321	
					(0.211)	
Div.					−0.526***	
					(0.145)	
P × Index						−0.009
						(0.019)
Index						−0.056***
						(0.011)
Constant	1.145***	0.146*	0.476***	−0.469***	0.855***	0.565***
	(0.130)	(0.174)	(0.144)	(0.171)	(0.097)	(0.000)
Observations	51,851	51,851	51,851	51,851	51,851	51,851
CA + (P × CA)	−1.202	0.927	−0.844	2.017	−0.847	−0.065
p-value	0.003	0.096	0.000	0.000	0.144	0.000

leads him to conclude that "idiosyncratic risk is the single largest cost faced by arbitrageurs."

G. Post-Publication Trading Activity in Predictor Portfolios

If academic publication provides market participants with information, then informed trading activity should affect not only prices, but also other indicators of trading. We therefore test whether trading volume, dollar trading volume, variance, and short interest increase in predictor portfolios after publication. To do so, we re-estimate the regression described in equation (1), but replace monthly stock returns with a monthly measure of one of the traits.

Trading volume is measured as shares traded, whereas dollar volume is measured as shares traded multiplied by price. Variance is the monthly stock return squared. We compute the average value of each variable among the stocks that enter either the long or the short side of the predictor portfolio each month, and test whether the means change post-publication. We use the logs of these variables as the dependent variables in our regressions. Short interest is measured as shares shorted scaled by shares outstanding. We measure the difference in short interest between the short and long sides of each portfolio each month, and use the difference as the dependent variable in our regressions. If publication draws short sellers to predictors, then this relative shorting measure should increase post-publication. Previous studies show that all of these variables increase over time during our sample period, so we include time fixed effects in all but the short interest specification, which measures the difference between the long and short sides in each cross-section.

We report the results in Table VI. The results show that trading volume and dollar volume are significantly higher during the period that is post-sample but pre-publication, while variance is significantly lower. Hence, there appears to be an increase in trading among predictor portfolio stocks even before a paper is published, suggesting that the information content of papers may get to some investors before the paper is published.

The post-publication coefficients show that trading volume and dollar volume are significantly higher in predictor portfolios after publication. The dependent variables are logs, so the coefficients show that post-publication trading volume and dollar volume increase by 18.7% and 9.7%, respectively. Variance, in contrast, declines by 6.5% post-publication. Lower volatility could reflect less noise trading (Shiller (1981) and Pontiff (1997)).

The final column reports results from the short interest regression. Recall that the short interest variable is the short interest on the short side minus the short interest on the long side. The coefficients in this regression are reported in percent (the dependent variable is multiplied by 100). If investors recognize that predictor portfolio stocks are mispriced, then there should be more shorting on the short side than on the long side. The average difference in short interest between the short and long sides of the characteristic portfolios in-sample is 0.143% (not in tables). The mean and median levels of short interest in our sample (1976 to 2013) are 3.45% and 0.77%, respectively, so this

Table VI

Trading Activity Dynamics in Predictor Portfolio Stocks

This regression models the dynamics of the traits of stocks in predictor portfolios, relative to the predictor's original sample period and the publication date. We perform monthly ranks based on turnover, dollar value of trading volume, and stock return variance. *Trading Volume* is measured as shares traded, whereas *Dollar Volume* is measured as shares traded multiplied by price. Variance is the monthly stock return squared. For each predictor portfolio, we compute the average of each variable among the stocks that enter either the long or the short side of the characteristic portfolio each month, and test whether it increases out-of-sample and post-publication. For short interest (shares shorted scaled by shares outstanding), we take the average short interest in the long quintile for each characteristic, and subtract from it the average short interest in the short quintile. The short interest findings are reported in percent (the dependent variable is multiplied by 100). *Post-Sample* is equal to 1 if the month is after the end of the sample but pre-publication. *Post-Sample (S)* is equal to one if the month is after the sample period used in the original study and zero otherwise. *Post-Publication (P)* is equal to one if the month is after the official publication date and zero otherwise. Standard errors (in parentheses) are computed under the assumption of contemporaneous cross-sectional correlation between panel portfolio residuals. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Variables	Variance	Trading volume	Dollar volume	Short-long short interest
Post-Sample (S)	−0.054***	0.092***	0.066***	0.166***
	(0.007)	(0.001)	(0.007)	(0.014)
Post-Publication (P)	−0.065***	0.187***	0.097***	0.315***
	(0.008	(0.013)	(0.007)	(0.013)
Observations	52,632	52,632	52,632	41,026
Time FE?	Yes	Yes	Yes	No
Predictor FE?	Yes	Yes	Yes	Yes
Null: S = P	0.156	0.000	0.000	0.000

difference is economically meaningful. This result suggests that some practitioners know prior to publication that stocks in the predictor portfolios are mispriced and trade accordingly. This could be because practitioners are trading on the predictor, or it could reflect practitioners trading on other strategies that happen to be correlated with the predictor. As an example, if short sellers evaluate firms individually with fundamental analysis, their resulting positions may be stocks with low book-to-market, high accruals, high stock returns over the last few years, etc., even though short sellers are not choosing stocks based on these traits.

Post-sample, relative shorting increases by 0.166%, and, post-publication, relative shorting increases by 0.315%. Economically, the post-publication effect represents a three-fold increase in shorting post-publication relative to in-sample. So, although some practitioners may know about these strategies before publication, the results here suggest that publication makes the effects more widely known. These short interest results are consistent with Hanson and Sunderam (2014), who use short interest as a proxy for sophisticated investors and find that increases in short interest are associated with lower future returns in value and momentum stocks.

Table VII
Regressions of Predictor Returns on Return Indices of Other Predictors

This regression models the returns of each predictor relative to the returns of other predictors. The dependent variable is a predictor's monthly long-short return. *Post-Publication (P)* is equal to one if the month is after the official publication date and zero otherwise. In-Sample Index Return is the equal-weighted return of all other unpublished predictor portfolios. *Post-Publication Index Return* is an equal-weighted return of all other published predictor portfolios. Standard errors (in parentheses) are computed under the assumption of contemporaneous cross-sectional correlation between panel portfolio residuals. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Variables	Coefficients
In-Sample Index Returns	0.748***
	(0.027)
Post-Publication Index Return	−0.008
	(0.004)
P × In-Sample Index Returns	−0.674***
	(0.033)
P × Post-Publication Index Return	0.652***
	(0.045)
Publication (P)	−0.081
	(0.042)
Constant	0.144***
	(0.019)
Observations	42,975
Predictors	97

H. The Effects of Publication on Correlations among Characteristic Portfolios

In this section, we study the effects that publication has on correlations between characteristic portfolios. If predictor returns reflect mispricing and if mispricing has a common source (e.g., investor sentiment), then we might expect in-sample predictor portfolios to be correlated with other in-sample predictor portfolios. This effect is suggested in Lee, Shleifer, and Thaler (1991), Barberis and Shleifer (2003), and Barberis, Shleifer, and Wurgler (2005). If publication causes arbitrageurs to trade on a predictor, then publication could also cause a predictor portfolio to become more highly correlated with other published predictors and less correlated with unpublished predictors because of fund flows or other factors common to arbitrage portfolios.

In Table VII, we regress predictor portfolio returns on the returns of an equal-weighted index of all other predictor portfolios that are pre-publication, and a second equal-weighted index of all of the other predictor portfolios that are post-publication. We include a dummy variable that indicates whether the predictor is post-publication, and interactions between this dummy variable and the pre-publication and post-publication indices.

The results show that pre-publication predictor returns are significantly related to the returns of other pre-publication predictor portfolios. The coefficient (or beta) on the pre-publication predictor portfolio is 0.748 and statistically significant. In contrast, the β for a pre-publication portfolio on other post-publication portfolios is −0.008 and insignificant. Hence, the returns of unpublished predictors are correlated with the returns of other unpublished predictors, but not with the returns of published predictors.

The interactions show that, once a predictor is published, its returns are less correlated with the returns of other pre-publication predictor portfolios and more correlated with the returns of other post-publication predictor portfolios. The coefficient for an interaction between the post-publication dummy and the return of the portfolio consisting of in-sample predictors is −0.653 and highly significant. These results show that, once a predictor is published, the beta of its returns with the returns of other yet-to-be-published predictors' returns virtually disappears, as the overall coefficient decreases to 0.748 – 0.674 = 0.074. The coefficient on the interaction between the post-publication dummy and the returns of the other post-publication predictors is 0.652 and significant at the 1% level, suggesting that there is a significant relation between the portfolio returns of published predictors and other published predictors.

IV. Conclusion

This paper studies 97 characteristics shown to explain cross-sectional stock returns in peer-reviewed finance, accounting, and economics journals. Using portfolios based on the extreme quintiles for each predictor, we compare each predictor's return predictability over three distinct periods: (i) the original study's sample period, (ii) the period outside the original sample period but before publication, and (iii) the post-publication period.

We use the period during which a predictor is outside of its original sample but still pre-publication to estimate an upper bound on the effect of statistical biases. We estimate the effect of statistical bias to be about 26%. This is an upper bound because some investors could learn about a predictor while the study is still a working paper. The average predictor's return declines by 58% post-publication. We attribute this post-publication effect both to statistical biases and to the price impact of sophisticated traders. Combining this finding with an estimated statistical bias of 26% implies a publication effect of 32%. Our estimate of post-publication decay in predictor returns is statistically significant relative to both the null of no post-publication decay and to the null that post-publication returns decay entirely.

Several of our findings support the idea that some or all of the original cross-sectional predictability is the result of mispricing. First, the returns of predictor portfolios with larger in-sample means decline more post-publication, and strategies concentrated in stocks that are more costly to arbitrage have higher expected returns post-publication. Arbitrageurs should pursue trading strategies with the highest after-cost returns, so these results are consistent with the idea that publication attracts sophisticated investors. Second, we find that turnover, dollar volume, and especially short interest increase significantly in predictor portfolios post-publication. This result is also consistent with the idea that academic research draws trading attention to the predictors. Finally, we find that, before a predictor is featured in an academic publication, its returns are correlated with the returns of other yet-to-be-published predictors, but its returns are not correlated with those of published predictors. This finding is consistent with behavioral finance models of comovement. After publication, a predictor's correlation with yet-to-be-published predictors is close to zero, and its correlation with already published predictors becomes positive significant.

Initial submission: May 16, 2013; Final version received: May 21, 2015
Editor: Kenneth Singleton

REFERENCES

Akbas, Ferhat, Will J. Armstrong, Sorin Sorescu, and Aanidhar Subrahmanyam, 2014, Time varying market efficiency in the cross-section of expected stock returns, Working paper, UCLA.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223–249.

Anand, Amber, Paul Irvine, Andy Puckett, and Kumar Venkataraman, 2012, Performance of institutional trading desks: An analysis of persistence in trading costs," *Review of Financial Studies* 25, 557–698.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse H. Pedersen, 2013, Value and momentum everywhere, *Journal of Finance* 68, 929–985.

Bali, Turan G., and Nusret Cakici, 2008, Idiosyncratic volatility and the cross section of expected returns, *Journal of Financial and Quantitative Analysis* 43, 29–58.

Barberis, Nicholas, and Andrei Shleifer, 2003, Style investing, *Journal of Financial Economics* 68, 161–199.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283–317.

Brennan, Michael J., 1970, Taxes, market valuation, and corporate financial policy, *National Tax Journal* 23, 417–427.

Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong, 2013, Trends in the cross-section of expected stock returns, Working paper, Emory University.

Cochrane, John H., 1999, Portfolio advice for a multifactor world, *Economic Perspectives: Federal Reserve Bank of Chicago* 23, 59–78.

Corwin, Shane A., and Paul Schultz, 2012, A simple way to estimate bid-ask spreads from daily high and low prices, *Journal of Finance* 67, 719–759.

De Long, J Bradford., Andrei Shleifer, Lourence H. Summers, and Robert J. Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.

Dichev, Ilia D., 1998, Is the risk of bankruptcy a systematic risk?, *Journal of Finance* 53, 1131–1148.

Dichev, Ilia D., and Joseph D. Piotroski, 2001, The long-run stock returns following bond ratings changes, *Journal of Finance* 56, 173–203.

Drake, Michael S., Lynn Rees, and Edward P. Swanson, 2011, Should investors follow the prophets or the bears? Evidence on the use of public information by analysts and short sellers, *Accounting Review* 82, 101–130.

Duan, Ying, Gang Hu, and R. David McLean, 2009, When is stock-picking likely to be successful? Evidence from mutual funds, *Financial Analysts Journal* 65, 55–65.

Duan, Ying, Gang Hu, and R. David McLean, 2010, Costly arbitrage and idiosyncratic risk: Evidence from short sellers, *Journal of Financial Intermediation* 19, 564–579.

Fama, Eugene F., 1976, *Foundations of Finance* (Basic Books, New York).

Fama, Eugene F., 1991, Efficient capital markets: II, *Journal of Finance* 46, 1575–1617.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 1998, Value versus growth: The international evidence, *Journal of Finance* 53, 1975–1999.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Franzoni, Francesco, and Jose M. Marin, 2006, Pension plan funding and stock market efficiency, *Journal of Finance* 61, 921–956.

Goldstein, Michael, Paul Irvine, Eugene Kandel, and Zvi Weiner, 2009, Brokerage commissions and institutional trading patterns, *Review of Financial Studies* 22, 5175–5212.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang, 2013, The supraview of return predictive signals, *Review of Accounting Studies* 18, 692–730.

Greenwood, Robin, 2008, Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights, *Review of Financial Studies* 21, 1153–1186.

Grundy, Bruce D., and Spencer J. Martin, 2001, Understanding the nature of the risks and the source of the rewards to momentum investing, *Review of Financial Studies* 14, 29–78.

Hanson, Samuel G., and Adi Sunderam, 2014, The growth and limits of arbitrage: Evidence from short interest, *Review of Financial Studies* 27, 1238–1286.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2013, . . . and the cross-section of expected returns, Working paper, Duke University.

Haugen, Robert A., and Nardin L. Baker, 1996, Commonality in the determinants of expected stock returns, *Journal of Financial Economics* 41, 401–439.

Heckman, James, 1979, Sample selection bias as a specification error, *Econometrica* 47, 153–161.

Hedges, Larry V., 1992, Modeling publication selection effects in meta-analysis, *Statistical Science* 7, 246–255.

Hutton, Amy, and Mary Barth, 2004, Analyst earnings forecast revisions and the pricing of accruals, *Review of Accounting Studies* 9, 59–96.

Jegadeesh, Narasimhan, and Sheridan Titman, 2001, Profitability of momentum strategies: An evaluation of alternative explanations, *Journal of Finance* 56, 699–720.

Kokkonen, Joni, and Matti Suominen, 2014, Hedge funds and stock market efficiency, Working paper, Aalto University.

Korajczyk, Robert, and Ronnie Sadka, 2004, Are momentum profits robust to trading costs, *Journal of Finance* 59, 1039–1082.

Leamer, Edward E., 1978, *Specification Searches: Ad Hoc Inference with Nonexperimental Data* (John Wiley & Sons, New York).

LeBaron, Blake, 2000, The stability of moving average technical trading rules on the Dow Jones Index, *Derivatives Use, Trading and Regulation* 5, 324–338.

Lee, Charles, Andrei Shleifer, and Richard Thaler, 1991, Investor sentiment and the closed-end fund puzzle, *Journal of Finance* 46, 75–109.

Lesmond, David A., Michael J. Schill, and Chunsheng Zhou, 2004, The illusory nature of momentum profits, *Journal of Financial Economics* 71, 349–380.

Lewellen, Johnathan, 2014, The cross-section of expected returns, *Critical Finance Review* 4, 1–44.

Liu, Qi, Lei Lu, Bo Sun, and Hongjun Yan, 2014, A model of anomaly discovery, Working paper, Yale School of Management.

Lo, Andrew, and Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431–467.

McLean, R. David, 2010, Idiosyncratic risk, long-term reversal, and momentum, *Journal of Financial and Quantitative Analysis*, 45, 883–906.

McLean, R. David, Jeffrey Pontiff, and Akiko Watanabe, 2009, Share issuance and cross-sectional returns: International evidence, *Journal of Financial Economics* 94, 1–17.

Michaely, Roni, Richard Thaler, and Kent L. Womack, 1995, Price reactions to dividend initiations and omissions: Overreaction or drift?, *Journal of Finance* 50, 573–608.

Mittoo, Usha, and Rex Thompson, 1990, Do capital markets learn from financial economists?, Working paper, Southern Methodist University.

Moskowitz, Tobias, Yao Hua Ooi, and Lasse H. Pedersen, 2013, Time series momentum, *Journal of Financial Economics* 104, 228–250.

Muth, John F., 1961, Rational expectations and the theory of price movements, *Econometrica* 29, 315–335.

Pontiff, Jeffrey, 1996, Costly arbitrage: Evidence from closed-end funds, *Quarterly Journal of Economics* 111, 1135–1151.

Pontiff, Jeffrey, 1997, Excess volatility and closed-end funds, *American Economic Review* 87, 155–169.

Pontiff, Jeffrey, 2006, Costly arbitrage and the myth of idiosyncratic risk, *Journal of Accounting and Economics* 42, 35–52.

Rouwenhorst, K. Geert, 1998, International momentum strategies, *Journal of Finance* 53, 267–284.

Schwert, G. William, 2003, Anomalies and market efficiency, in George M. Constantinides, Milton Harris, and René Stulz eds.: *Handbook of the Economics of Finance* (Elsevier Science B.V., Amsterdam).

Shiller, Robert, 1981, Do stock prices move too much to be justified by subsequent changes in dividends, *American Economic Review* 71, 421–436.

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits to arbitrage, *Journal of Finance* 52, 35–55.

Sullivan, Ryan, Alan Tinmerman, and Halbert White, 2001, Dangers of data mining: The case of calendar effects in stock returns, *Journal of Econometrics* 105, 249–286.

Treynor, Jack, and Fischer Black, 1973, How to use security analysis to improve portfolio selection, *Journal of Business* 46, 66–86.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.

Journal of Financial Economics 101 (2011) 313–332

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Time-varying rare disaster risk and stock returns ☆

Henk Berkman [a], Ben Jacobsen [b,*], John B. Lee [a]

[a] University of Auckland, Auckland, New Zealand
[b] Massey University, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 20 September 2009
Received in revised form
25 August 2010
Accepted 27 September 2010
Available online 8 March 2011

JEL classification:
G12
G15

Keywords:
Equity premium
Volatility
Rare disasters
International political crises
Consumption

ABSTRACT

This study provides empirical support for theoretical models that allow for time-varying rare disaster risk. Using a database of 447 international political crises during the period 1918–2006, we create a crisis index that shows substantial variation over time. Changes in this crisis index, our proxy for changes in perceived disaster probability, have a large impact on both the mean and volatility of world stock market returns. Crisis risk is positively correlated with the earnings–price ratio and the dividend yield. Cross-sectional tests also show that crisis risk is priced: Industries that are more crisis risk sensitive yield higher returns.

1. Introduction

The study of rare disasters has recently attracted renewed interest, as the probability of rare disasters might explain longstanding asset-pricing puzzles, such as the equity premium puzzle and the volatility puzzle.[1] For example, extending the rare disaster model in Rietz (1988), Barro (2006) finds that his model explains the high observed equity premium with realistic degrees of risk aversion. Essential inputs in this and related rare disaster models are disaster probability and disaster size. Barro (2006) estimates these parameters from actual contractions in gross domestic product (GDP) associated with World War I, the Great Depression, and World War II across several different countries.

More recently, Gabaix (2009) extends the Barro-Rietz model by allowing cross-sectional and time-series variation in the expected loss from disasters. Also extending the Barro-Rietz model, Wachter (2009) introduces a time-varying probability of rare disasters and finds that her

☆ This paper (previously titled "War, Peace, and Stock Markets") has benefited from the very helpful comments of two anonymous referees, and the editor (G.W. Schwert). We also thank Utpal Bhattacharya, Michael Brecher, Charles Corrado, Paul Koch, Martin Lally, Micah Officer, Qinghai Wang, and participants at the European Finance Association meetings in Zurich, Switzerland, August 2006; the Asian FMA meetings in Auckland, New Zealand, July 2006; the New Zealand Finance Colloquium in Dunedin, New Zealand, January 2006; and presentations at several universities. This research is supported by the Marsden Fund administered by the Royal Society of New Zealand. The usual disclaimer applies.

* Corresponding author.
E-mail addresses: h.berkman@auckland.ac.nz (H. Berkman), b.jacobsen@massey.ac.nz (B. Jacobsen), byong.lee@auckland.ac.nz (J.B. Lee).

[1] See Mehra and Prescott (1985) and Shiller (1981). Gabaix (2009) uses the rare disaster model to suggest solutions to ten puzzles in macro-finance.

0304-405X/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jfineco.2011.02.019

model can account for the high equity premium and high volatility in the stock market, while generating low mean and volatility for the government bill rate. Similar results are obtained in the time-varying disaster risk model of Gourio (2008a).

An obstacle to empirical verification of the rare disaster models is that individual countries rarely face actual major disasters. For example, as we will show later, at the country level, an international political crisis occurs on average once every 15 years, a full-scale war once every 74 years, and a war on home territory once every 119 years. The difficulty of empirical verification is serious, because calibrations based on disaster models are highly sensitive to underlying assumptions, and there is only limited evidence for the underlying mechanisms these models rely on.

Our study avoids the small sample problem inherent in the use of *actual* rare disasters by focusing on a much larger sample of *potential* disasters: international political crises that are likely to cause changes in perceived rare disaster probabilities. Even though the large majority of these international crises did not escalate into a full-scale war or major conflict, the fact that financial market prices are forward-looking allows us to gauge the impact of changes in the probability of rare disasters on stock market prices. Consequently, whereas other studies infer the existence of a disaster risk premium from asset prices, we directly test whether a link exists between changes in disaster risk and changes in stock market prices.

Our source of events, that is, changes in disaster probability, is a detailed database of all international political crises that occurred during the period 1918–2006. This database, known as the International Crisis Behavior project (ICB) (see http://www.cidcm.umd.edu/icb/), has been developed and refined over a period of more than 30 years by the Center for International Development and Conflict Management. The ICB database contains detailed information on more than 400 international political crises. The most important events for each crisis (such as start and end dates) have been carefully documented, with crises classified based on numerous characteristics, such as superpower involvement, duration, and gravity. In addition to the large number of observations and the consistent way in which crises are measured, this ICB database is also attractive due to its definition of *crisis*. A crisis does not necessarily start with an attack or military action; rather, it is defined as a perceived change in the probability of a threat that results in the start or end of an international political crisis. This perceived change in the probability of a threat is likely to be closely aligned with the news events to which investors might react.

Our empirical results, which are based on 447 major international political crises and 938 crisis actors, are consistent with the implications of rare disaster models. First, we find that international political crises result in large, negative world stock market returns when they start, lower-than-average world market returns as they continue, and positive world market returns when they end. Our estimates indicate that disaster risk does weigh heavily on the minds of investors: The negative return due to international crises cumulates to almost 4% per annum.[2]

Second, consistent with time-varying disaster probability models, the start of an international crisis significantly increases the volatility of world market returns, and for each crisis that comes to an end, volatility decreases. This latter finding is also consistent with the suggestion in Schwert (1989) and the theoretical model in Veronesi (2004) that a higher level of political instability can result in increased stock market volatility.

Third, we show that stock price reactions are stronger for more severe crises, such as crises involving a territorial threat, a threat of grave damage (large casualties in a war), or a threat to existence. Markets also react more strongly when a great power or a superpower is involved on both sides of the conflict and when the crisis is part of a protracted conflict. Using a crisis severity index that reflects several aspects of the seriousness of a crisis, we again find that the more severe a crisis and, therefore, the greater the probability of a dramatic decline in consumption, the stronger the effect on stock returns.

Fourth, while predictive regressions do not show a significant relation between crisis risk and future market returns, the earnings–price ratio and dividend yield for the S&P 500 index are significantly positively correlated with crisis risk. This latter result is consistent with time-varying disaster models, and suggests that expected market returns are higher in periods with high crisis risk. Further support for the notion that disaster risk is priced comes from our cross-sectional asset pricing tests, where we find that U.S. industries with higher crisis risk sensitivity yield higher returns, on average.

Finally, consistent with a central assumption in time-varying disaster risk models, we show that changes in disaster risk negatively affect expected future consumption growth, proxied by consumer confidence and GDP-growth forecasts. This finding is important, because it is through this process that changes in disaster risk generate changes in expected returns and volatility over time. We also find that the start (end) of a crisis for an individual country significantly increases (decreases) the probability of that country being at the start of an actual consumption disaster as identified by Barro and Ursúa (2009).[3]

[2] This estimate is likely to be too optimistic due to survivorship bias (see also Jorion and Goetzmann, 1999). For example, investors in Russian stocks and bonds during the Revolution of 1917 lost all their money. However, as we lack long data series for Russian stock markets, this negative return is not included in our analysis. In addition, there are several months with missing observations for some European countries around the start of World War II, and some monthly returns are missing for the German stock market at the end of World War II. Furthermore, to the extent that crises are anticipated, our estimates are biased downward. Finally, our analysis does not include civil wars, crises not identified by ICB as international crises, and other rare disasters that may have consumption effects.

[3] While this evidence directly links our crisis probability measures with actual consumption disasters, increases in crisis probability could also be associated with small, persistent decreases in the expected consumption growth rate and increases in its volatility. To the extent that this is the case, our results can also be interpreted along the lines of long-run risk models (see Bansal and Yaron, 2004).

Our results still hold if we exclude the first 30 years of our sample period, and thus exclude World War I, the Great Depression, and World War II. This shows that the strong stock market reactions, the high correlation between the earnings–price ratio and crisis risk, and the cross-sectional asset pricing findings are not the result of the actual disasters used in the calibrations in Barro (2006). We also verify that our results are robust to alternative definitions of the world market index. In addition to the Global Financial Data world market index, we use an index based on equal country weights and an index that reflects the importance of countries based on annual GDP, obtaining similar results. Our final robustness test investigates stock market reactions for individual countries that are directly involved in crises as a crisis actor. The start of an international political crisis reduces stock returns in crisis-actor countries by 2%, and for every month that a crisis continues, the stock market drops by an average 1.7% per month. These results again show that investors care strongly about rare disaster risk.

In related studies, Barro and Ursúa (2008) estimate crisis probabilities using actual GDP and consumer expenditure declines for a large set of countries, and Barro and Ursúa (2009) consider stock market crashes and depressions worldwide. Their focus differs from ours, as they consider periods when actual drops in consumption occurred. There are also numerous case studies of financial market reactions to actual crises. For example, Waldenström and Frey (2002), and Frey and Kucher (2000) study the impact of events during World War II on the prices of several countries' government bonds traded in Zurich and Sweden, respectively. More recently, the war in Iraq has created interest in the consequences of war on financial markets. Rigobon and Sack (2005) study the impact of the Iraq War on several financial variables. Moreover, Amihud and Wohl (2004) find that the likelihood of Saddam Hussein's fall from power, as reflected in a traded futures contract that paid out if Saddam were ousted, is related to stock market returns. The same "Saddam Security" is used by Wolfers and Zitzewitz (2009) to estimate the expected cost of the war in Iraq. All of these studies show that war is costly. For example, Wolfers and Zitzewitz (2009) estimate that in the pre-war period, the Iraq War was expected to lower the value of U.S. equities by around 15% (or $1.1 trillion in market value of all stocks in the S&P 500 index). Similarly, there is some evidence of a relation between stock market volatility and political crises. Bittlingmayer (1998) focuses on Germany in the period 1880–1940 and presents evidence suggesting that political uncertainty in the early 1920s generated greater stock volatility. Voth (2002) studies the relation between volatility and political uncertainty in ten individual countries during the interwar period. His results reveal that political uncertainty, proxied by anti-government demonstrations, riots, and assassinations, can explain a substantial part of the increase in stock volatility during the Great Depression. Finally, Brown, Burdekin, and Weidenmier (2006) argue that the low volatility of consols during Pax Britannica (1816–1913) is probably the result of the political stability in that period.

Our study differs from this previous work in several ways. First, we focus on changes in disaster probability for a large sample of events over an extended period, including both disasters that did eventuate and also potential ones that did not. Second, we examine how international political crisis risk affects world stock market returns and show that disaster risk cannot simply be diversified away.[4] Finally, consistent with time-varying disaster risk models, we show that disaster risk affects future consumption growth and expected stock returns.

2. The ICB database, crisis variables, and stock returns

2.1. The ICB database

The ICB database, which was started in 1975, provides detailed information on 447 major international political crises that occurred over the period 1918–2006, and on the 983 crisis actors (countries) involved in these crises. An extensive discussion of the database, definitions of variables, and specific choices made by ICB members can be found in Brecher and Wilkenfeld (1997). ICB data have been used in hundreds of articles and numerous books in the political science literature,[5] and also in studies on war and economics (e.g., Hess and Orphanides, 1995; Blomberg, Hess, and Orphanides, 2004).

The intention of ICB is to create an exhaustive list of international political crises since World War I. To this end, ICB has assembled a list of all candidate crises and established whether they meet the following definition of a *crisis*. A *foreign policy crisis*, that is, a crisis for an individual state, is a situation with three necessary and sufficient conditions deriving from a change in the state's internal or external environment. All three conditions are perceptions held by the highest-level decision makers of the state actor concerned: (1) a threat to one or more basic values, along with (2) an awareness of finite time for response to the value threat, and (3) a heightened probability of involvement in military hostilities.

As pointed out earlier, a particularly attractive feature of the definition used by Brecher and Wilkenfeld (1997) in relation to studying the interaction between financial markets and rare disasters is their emphasis on perceived probabilities. A crisis does not necessarily begin with an attack or military action, but a perceived change in the probability of threat can result in the identification of the start or end of a crisis. Consequently, the start and end dates of a crisis are likely to be closely aligned with the news events to which investors react, thus lowering the chance that the market has already incorporated the news. Another advantage of using events related to international political crises is that such events are likely to be exogenous. Other events, for example, events related to financial crises, might contain information regarding disaster probability, but, at the same time, will most likely

[4] Copeland and Zhu (2007) argue that insofar as extreme events are less than perfectly correlated internationally, disaster risk is diversifiable, thereby mitigating their effect on the required return on equity.

[5] The ICB Web site (http://www.cidcm.umd.edu/icb/) contains an overview of studies that have employed the ICB data.

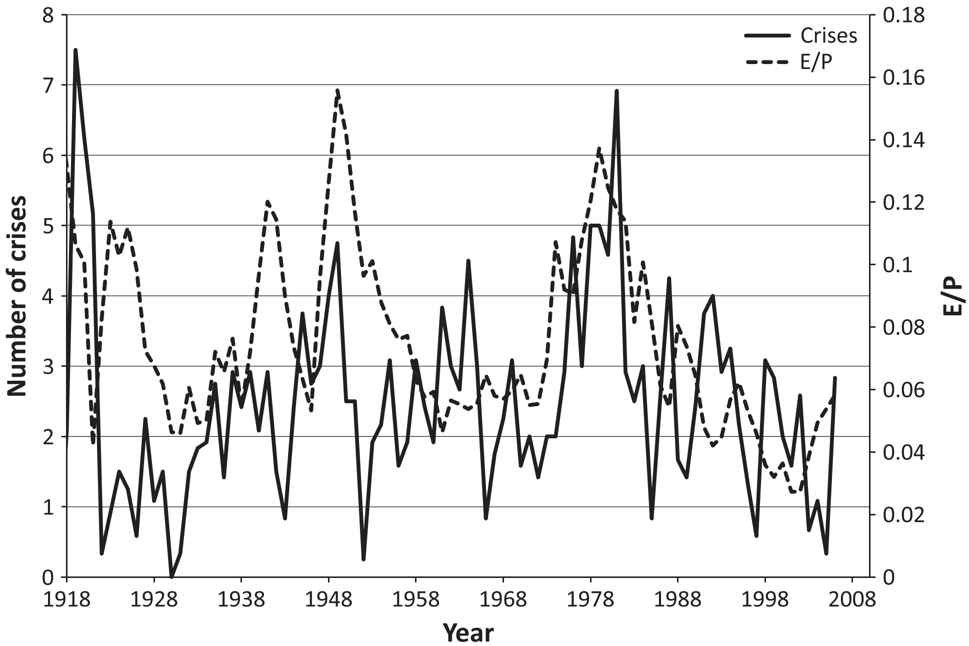


Fig. 1. For every year in the period 1918–2006, the figure plots the average number of international crises per month according to the ICB database. The dotted line plots the earnings/price ratio (E/P) for the S&P 500 index.
(*Source*: Global Financial Data).

contain information regarding expected consumption if no disaster occurs. Finally, the database also contains several measures of crisis severity, allowing us to identify more serious crises (i.e., crises with more severe threats or with a broader international impact) that arguably resulted in greater changes in disaster probability.

For every crisis, ICB distinguishes 66 crisis dimensions and control variables including specific dates, crisis triggers, gravity of the crisis, great power involvement, location, and crisis outcomes. In a related database, ICB provides details on all crisis actors in relation to the different crises (in total, 80 dimensions, control variables, and actor attributes). The ICB Web site, http://www.cidcm.umd.edu/icb/, contains detailed background information on all 447 international crises. ICB defines the trigger date for a crisis as the date when the earliest actor in an international crisis perceives a crisis. This date is based on the occurrence of an act or event or derived from, for example, diaries, memoirs, or speeches. In our analysis of the impact of crises on world market returns, we use the crisis trigger date as the most objective start date of a crisis.[6] We exclude six crises from the analysis, because the database did not specify an explicit end date (in terms of month and year); thus, the total number of observations we use is 441.

For every year in the period 1918–2006, Fig. 1 plots the average monthly number of international crises in that year, according to the ICB database.

Even though one could argue the total number of crises is a crude measure of disaster probability, this crisis measure seems to correlate well with global political instability. The first peak in the graph occurs circa 1920. This period was marked by substantial international tension due to crises in both Central and Eastern Europe in the wake of the First World War, and in and around Russia during and following its civil war (1918–1920). The number of crises peaks again around the beginning of World War II and again in 1945 when it ended. Immediately after the war, political and ideological disputes between the era's two superpowers, the Soviet Union and the United States, began to emerge. This was the origin of the Cold War that culminated in major international crises, such as the Berlin Blockade in July 1948 and the Korean War (1950). Stalin's death in 1953 perhaps helped to relax tensions, but the political situation in the second half of the 1950s and early 1960s remained very grave (e.g., the protracted conflict in Indochina that

[6] In our analysis of individual countries, we use crisis trigger dates of individual countries that might differ from the general crisis trigger date. An example should clarify this distinction: The trigger date for the crisis related to the Japanese attack on Pearl Harbor on December 7, 1941 is November 26, 1941. On that date, U.S. Secretary of State Hull presented two Japanese envoys with a Ten Point Plan, viewed by Japan as an ultimatum and triggering a crisis for Tokyo. However, in our analysis of individual countries, we use December 7, 1941 as the trigger date for the United States (as well as Australia, Canada, the Netherlands, New Zealand, and the United Kingdom), as the attack itself triggered a crisis for these countries. One could argue that we should also use

(*footnote continued*)
December 7, 1941 in our analysis of world market returns, because that was, for this crisis, the day stock markets dropped. However, to avoid arbitrary choices, we choose to use the general trigger dates and end dates of ICB in our analysis of world market returns. Moreover, the crisis trigger month for individual countries is the same as the general crisis trigger month for more than 75% of our sample.

escalated into the Vietnam War, the Berlin crisis in 1961, and the Cuban Missile Crisis in 1962, which brought the world close to a nuclear war). The second half of the 1960s and the 1970s was a period of relative calm (détente), but indirect conflict between the superpowers continued in the Middle East, Ethiopia, and Angola. Conflicts flared up again in 1979 and the early 1980s with the Soviet intervention in Afghanistan (described by President Carter as the "most serious threat to peace since the Second World War"). Other major international crises in this period are the labor strikes in Poland led by Solidarity and the start of the Iran-Iraq War. The end of the Cold War began with Gorbachev's ascension to power in 1985 and ended with the dissolution of the Soviet Union in December 1991. In the next two decades, several conflicts emerged that can be traced back to the Cold War and the break-up of the Soviet Union (e.g., crises involving North Korea, crises in and around the Balkans, and crises in the Caucasus). However, in this new uni-polar world with the United States as the only remaining superpower, the number of international crises has declined noticeably. Nevertheless, several major crises have broken out in this period. Prominent examples are the Taiwan Strait conflicts, the Gulf War and post-Gulf War crises, conflicts between Israel and neighboring countries, and crises in response to terrorist attacks by the Al Qaeda network.

In this study, we examine whether political events that take place in one part of the world, with crisis actors whose contribution to the world stock market portfolio might be small, have the potential to affect world stock market returns. Before turning to the empirical evidence, it is important to point out that more than 50% of the crises we consider had five or more "involved actors" and more than 10% of crises had more than ten involved actors.[7] These numbers, which are similar for the first and the second half of our sample period, indicate the potentially substantial impact of the crises in our sample. Also note that while the economy of a crisis actor is likely to be affected by a crisis, there is reason to believe that, throughout our sample period, crises could also affect the economies of non-crisis actors. For example, O'Rourke and Williamson (2002) claim that at the beginning of the twentieth century "the world economy was extremely well-integrated even by late-twentieth-century standards." In a related vein, Obstfeld and Taylor (2003) point out that global financial integration was well-entrenched at the beginning of the twentieth century, and that foreign capital stocks and correlations between stock markets in this period were as high as at the end of that century.

2.2. Crisis variables

The ICB data sets contain numerous variables relating to different aspects of crises. We use many of these variables and transform them in such a way that we can answer our main questions (See Appendix A). To study the impact of international crises on stock returns, the most general variable is the total number of crises in month t according to the ICB database: $Crisis_t$. We make a further distinction in this general crisis variable and define the following variables: $Start_t$—the number of crises that start in month t; End_t—the number of crises that end in month t; and $During_t$—the number of ongoing crises in month t (where month t is not the start or end month). Note that $During_t$ differs from $Crisis_t$ only in the sense that all crises starting or ending in month t are excluded from $During_t$ (and included in $Crisis_t$). It is likely that $Start_t$ and End_t represent the periods with the largest changes in disaster probabilities; however, we also include $During_t$ as a variable in our analysis. For example, investors might consider an ongoing crisis as a further increase in disaster probability.[8]

Some crises are more damaging than others, and we expect that more severe crises indicate higher disaster probability and will lead to a stronger market impact. We use six dummy variables to capture different aspects of the severity of a crisis: whether or not a crisis started with violence, violence used during the crisis, full-scale wars, gravity of value threat, whether the crisis is part of a protracted conflict, and great power or superpower involvement.

The ICB database includes a variable that indicates whether a crisis started with a violent act. Our subgroup *Violent break* contains the crises that begin with a violent act.[9] *Violent crises* are crises with either serious clashes or full-scale wars (as opposed to crises with minor clashes or crises with no violence), and crises in the subgroup *War* are all full-scale wars. ICB defines "gravity of value threat" as the most salient object of threat identified by any actor in a crisis. ICB identifies the following value threats: an economic threat, limited military damage, a political threat,[10] threat to influence,[11] territorial threat,[12] threat of grave damage,[13] and the most basic value threat: a threat to existence.[14] For our analysis, we consider the subset *Grave threat*: the number of crises in month t (*Start*, *During*, or *End*) in which the value threat involves a territorial threat, a threat of grave damage, or a threat to existence.[15] We also separate crises in protracted

[7] Involved actors are states that have substantial involvement in the crisis, where substantial involvement is one of the following types of activity: direct military; semi-military; covert; economic; and political other than mere statements of approval or disapproval by officials. In total, there are 2,534 involved actors in the 441 crises in our sample (including the 983 crisis actors in the sample).

[8] Excluding $During_t$ does not alter our conclusions regarding start and end effects.

[9] ICB defines a violent act as a border clash, border crossing by limited force, invasion of airspace, sinking of an airship, sea-air incident, bombing of large target, large-scale military attack, or war.

[10] Threat of overthrow of regime, change of institutions, intervention in domestic politics, subversion.

[11] Threat of declining power in the global system and/or regional subsystem.

[12] Threat of integration, annexation of part of state's territory.

[13] Threat of large casualties in war, mass bombings as a result of grave damage.

[14] Threat of survival of population, genocide, threat to existence of entity, of total annexation, colonial rule, occupation.

[15] After transformation to binary variables, several variables in the data set are highly correlated with $Grave\ threat_t$: crisis management technique, centrality of violence, intensity of violence, extent of violence, and timing of violence.

conflicts, *Protracted,* and crises outside protracted conflicts. According to Brecher and Wilkenfeld (1997), protracted conflicts tend to have more violent triggers, and crisis actors tend to perceive more basic threats. The authors also suggest that protracted conflicts often reflect political involvement by major powers and can be destabilizing, as they can affect the status of dominant powers in the international system. To measure the involvement of major powers more directly, we use several variables in the ICB data set to construct the subgroup *Major power* that contains crises in which at least one great power or superpower[16] is involved on both sides of the conflict.[17] Each of the six subgroups above has a matching dummy variable that assumes a value of one if a crisis is in that subgroup, and zero otherwise. For example, a crisis in the subgroup *Violent break* has a value of one for the dummy variable *Violent break*, and zero otherwise.

Finally, we construct a *crisis severity index* that summarizes different aspects of crisis severity into one measure by aggregating the six variables above and adding one.[18] For example, a war with great power involvement that is part of a protracted conflict, has a score of four on the crisis severity index (one for being a crisis, one for being a war, one for having great power involvement, and one for being part of a protracted conflict). Because the number of crises ($Crisis_t$) and the crisis severity index are highly correlated (correlation coefficient is 0.91), we do not include a separate figure for the evolution of the crisis severity index over time.

Table 1 presents descriptive statistics of the crisis variables. An average month has 2.47 crises. The greatest number of crises in a month is ten and occurred during April-July 1919. Apart from a crisis between Nicaragua and Costa Rica, due to a military coup in the latter country, the other crises in that period occurred in the proximity of Russia.

Months without any crises occur approximately 10% of the time. To be precise: 123 months out of the 1,068 months were free from any crisis. Crisis-free periods generally do not last long. The longest period without any international crisis lasted 20 months. It began on December 22, 1929, with the resolution of a crisis between the Soviet Union and China over the Chinese Eastern Railway, and lasted until the night of September 18, 1931, when the Japanese Kwantung Army clashed with Chinese troops at Mukden. This triggered a crisis for China and Japan, and was followed by the steady occupation by Japanese troops of the northeastern provinces of China.

The maximum number of international crises that start in a particular month is four. In January 1981, France and Libya experienced a crisis over Libya's plan to merge with Chad; a Peruvian helicopter was shot down, which led to a new crisis between Ecuador and Peru over a long-disputed border; a raid by the South African Defensive Forces destroyed an African National Congress headquarters in Maputo, Mozambique creating a crisis for Mozambique and South Africa; and a crisis for Israel was triggered by a French announcement that an Osirak nuclear reactor in Iraq would be fully operational by July 14. The maximum number of crises that ended in a given month is also four. In March 1939, an agreement between Germany and Lithuania regarding Germany's demand for the annexation of Memel ended a crisis for Lithuania; Germany's annexation of Czechoslovakia began and ended a crisis; a dispute between France and Italy ended with a speech by Mussolini; and the occupation of Madrid by the Nationalists ended the Spanish Civil War.

The column in Table 1, Panel A with the heading "Sum," also provides several interesting observations of recent history. Out of a total of 441 crises, 184 crises began with a violent break, 201 crises involved serious violence, and 95 crises were full-scale wars. There are 199 crises that involved threats of the most basic values at some time during the crisis. In 137 of the crises, at least one major power was involved on both sides of the conflict, and 258 crises were part of a protracted conflict. On average, 1.4 violent crises and 0.8 wars are taking place in any given month (add *Start*, *During*, and *End*, and divide by 1,068).

The column with the first-order autocorrelations, $\rho(1)$, shows that a crisis commencing in one month increases the likelihood of a crisis commencing the next. Similarly, if a crisis comes to an end in a certain month, it increases the likelihood of more crises ending in the near future. Because first-order autocorrelations for our main variables (particularly *During*) are sometimes high, we test whether these main variables might contain a unit root. Using the Elliot-Rothenberg-Stock DF-GLS test, we strongly reject the presence of a unit root for all series.

Table 1, Panel B shows significant correlation between the different crisis variables. For example, crises that begin with a violent act (*Violent break*) tend to result in crises exhibiting either serious clashes or full-scale wars (*Violent crises* and *War*). The correlations involving *Grave threat* and *Major power* are slightly lower but still always significant at the 1% level. Not surprisingly, the correlation between the crisis severity index and all other measures is always high.

2.3. Crises, wars, and consumption disasters

Table 2 provides information on crisis probabilities based on sample frequencies. While on a world scale crises occur frequently, for individual countries international political crises are rare. With 983 crisis actors and 167 countries at present, the likelihood of becoming involved in a crisis in any given year, based on our sample of 89 years, equals 6.6%, or once every 15 years. For

[16] ICB uses the following hierarchy of powers: superpowers (the United States and the Soviet Union 1945–2002; great powers (France, Germany, Italy, Japan, the United Kingdom, the United States, and the Soviet Union 1919–1939), and China (since 1949); France and the United Kingdom (since 1945); medium powers; and small powers.

[17] This variable subsumes several variables in the database: great power involvement in the crisis, content of great power activity, effectiveness of great power activity, the level and number of international systems that are affected, and several variables that describe the perception by crisis actors of great power involvement.

[18] We thank an anonymous referee for the suggestion to construct a summary measure of crisis severity. We follow La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1998) in adding a set of dummy variables to create an aggregate score.

Table 1
Descriptive statistics and correlation of crisis variables.

Panel A reports the mean, standard deviation, minimum, maximum, and sum for all crisis variables used in our analysis; $\rho(1)$ is the first-order autocorrelation of each series. Crisis denotes the number of crises that take place in any month. This variable is split into *Start*, *During*, and *End*, which denote the monthly number of starting, ongoing, and ending crises, respectively. *Violent break* gives the number of crises that start with a violent act. *Violent crises* are crises with either serious clashes or full-scale wars, and crises in the subgroup War include all full-scale wars. *Grave threat* denotes the number of crises that involve a threat to existence, a threat of great damage, or a territorial threat. *Major power* denotes the number of crises where a great power or a superpower is involved on both sides of the conflict. *Protracted* is the number of crises that are part of a protracted conflict. The crisis severity index summarizes different aspects of crisis severity into one measure by aggregating six variables (*Violent break*, Violent crisis, War, *Grave threat*, *Major power*, and *Protracted*) and adding 1. The critical value at the 1% level for the Elliott-Rothenberg-Stock DF-GLS test statistic for tests of the null hypothesis of a unit root equals -2.57. * Denotes significance at the 10% level; ** denotes significance at the 5% level; and *** denotes significance at the 1% level. Panel B reports the correlation coefficients between the different crisis variables based on *Start*.

Panel A: Basic characteristics of crisis variables

Variables	Mean	Std. dev.	Min	Max	Sum	$\rho(1)$	DF-GLS
Crisis	2.471	1.814	0	10	2,639	0.819***	−4.30***
All							
Start	0.413	0.671	0	4	441	0.116***	−5.61***
During	1.645	1.393	0	10	1,757	0.858***	−5.85***
End	0.413	0.676	0	4	441	0.098**	−3.54***
Violent break							
Start	0.172	0.435	0	3	184	0.056*	−8.50***
During	0.598	0.765	0	4	639	0.805***	−8.44***
End	0.172	0.420	0	3	184	0.081	−4.47***
Violent crises							
Start	0.188	0.436	0	3	201	0.060**	−8.05***
During	1.001	1.138	0	8	1,069	0.887***	−6.04***
End	0.188	0.447	0	3	201	0.104	−4.51***
War							
Start	0.089	0.295	0	2	95	0.017	−7.21***
During	0.589	0.748	0	3	629	0.866***	−6.90***
End	0.089	0.310	0	2	95	0.064***	−15.05***
Grave threat							
Start	0.186	0.429	0	2	199	0.050	−5.54***
During	0.709	0.867	0	4	757	0.840***	−7.45***
End	0.186	0.448	0	3	199	−0.024	−9.00***
Major power							
Start	0.128	0.367	0	3	137	0.073***	−7.93***
During	0.530	0.868	0	7	566	0.891***	−5.91***
End	0.128	0.361	0	3	137	0.118***	−8.87***
Protracted							
Start	0.242	0.503	0	3	258	0.114***	−9.58***
During	0.885	1.018	0	5	945	0.851***	−7.02***
End	0.242	0.506	0	4	258	0.046	−6.43***
Crisis severity index							
Start	1.419	2.459	0	14	1,515	0.099***	−3.14***
During	5.957	5.387	0	37	6,362	0.859***	−6.06***
End	1.419	2.173	0	19	1,515	0.089**	−3.70***

Panel B: Correlation among crisis severity variables

	All crises	*Violent break*	*Violent crises*	*War*	*Grave threat*	*Major power*	*Protracted*	*Crisis severity index*
All crises	1							
Violent break	0.65***	1						
Violent crises	0.66***	0.62***	1					
War	0.43***	0.43***	0.69***	1				
Grave threat	0.69***	0.49***	0.53***	0.31***	1			
Major power	0.51***	0.23***	0.41***	0.37***	0.28***	1		
Protracted	0.78***	0.59***	0.62***	0.48***	0.53***	0.48***	1	
Crisis severity index	0.91***	0.76***	0.83***	0.64***	0.73***	0.59***	0.86***	1

individual countries, these probabilities can differ greatly. For instance, Australia and Canada have only been involved in three crises, while the United States has been involved in 64 crises (see also the last column in Table 4).

While crises in our database are reasonably frequent (2.5 per month), the occurrence of what one might consider an actual disaster is rare. Table 2 shows that the sample probability for a country to be involved in a

Table 2
Crisis probabilities.
Annual average probabilities for countries – assuming 167 countries – to become involved in an international political crisis. The subdivision with respect to crisis gravity and violence is based on classification of the variables from the ICB database.

Annual crisis probabilities			
Gravity		*Violence*	
Economic threat	0.25%	No violence	3.10%
Limited military damage	0.59%	Minor clashes	0.74%
Political threat	1.10%	Serious clashes	1.41%
Territorial threat	1.68%	Full-scale war	1.35%
Threat to influence	1.44%		
Threat of great damage	1.10%		
Threat to existence	0.45%		
Total	6.61%	Total	6.61%

Table 3
Depressions and crisis actors.
Results from the following probit model of the relation between the start of country-specific depressions (multi-year declines of consumption or GDP by 10% or more) and the involvement of that country in international political crises:

$$Depression\_start_{c,y} = \alpha + \beta_1 Start_{c,y} + \beta_2 During_{c,y} + \beta_3 End_{c,y} + \varepsilon_{c,y}.$$

The dependent variable is a dummy variable that is equal to one for country/year pairs that are indicated as the start of a depression in Barro and Ursúa (2009) and is equal to zero otherwise. There are 37 depressions in the period 1918–2006, for the 19 countries in our sample. $Start_{c,y}$, $During_{c,y}$, and $End_{c,y}$ refer to the number of international crises in year y in which country c was a crisis actor. The Z-statistics are based on cluster-corrected standard errors.

	Coefficient	Marginal effect	Z-statistic
Start	0.537	0.028	3.86
During	0.009	0.001	0.39
End	−0.435	−0.023	−2.85

crisis that is a threat to existence or a threat of great damage equals 1.55% per annum. The probability to become involved in a full-scale war is 1.35% a year, or once every 74 years.[19] On average, a country gets involved in a new crisis with serious clashes, or a full-scale war, once every 36 years, or around 2.76% of the time during our sample period. The sample probability for a country to experience the launch of a war on home territory is 0.84% per year (not reported in the table).

Of all crises in our sample, crises that started as a war, or developed into a war, are most likely to be associated with consumption disasters.[20] We consider all wars in our database and calculate the percentage change in real GDP per capita for the war-actor countries in the years around actual wars. The change in GDP is measured from the year of the start of a war, or the year before the start of the war if GDP per capita is higher, to the first minimum. In almost all cases, these minima occurred within five years. There are 95 separate wars in the database, and we can match these wars with GDP per capita data for 88 different war-actor countries.[21] We find that GDP does not decrease below the level at the start of the war in 24 cases. For the remaining 64 observations, we find that the average change in GDP from the start of the war to the minimum is −24.9%, and the median decrease is −19.8%.

To provide further evidence on the relation between the number of international crises a country is involved in and actual consumption disasters, we use data from Barro and Ursúa (2009) on country-specific depressions (multi-year declines of consumption or GDP by 10% or more). For our sample period from 1918 to 2006, we match the 19 countries for which we have stock price data with the data in Barro and Ursúa, and find 37 country-specific depressions. For these depressions, we obtain the start year as indicated in Table 2 in Barro and Ursúa. Next, we test if the starts of these country-specific depressions are related to country-specific crisis indicators ($Start_{cy}$, $During_{cy}$, and End_{cy}), using the following probit model:

$$Depression\_start_{c,y} = \alpha + \beta_1 Start_{c,y} + \beta_2 During_{c,y} + \beta_3 End_{c,y} + \varepsilon_{c,y}. \tag{1}$$

In Table 3, we present the results from this probit model, where the dependent variable is a dummy variable that is equal to one for country/year pairs that are indicated as the start of a depression in Barro and Ursúa (2009), and is equal to zero otherwise. $Start_{c,y}$, $During_{c,y}$, and $End_{c,y}$ refer to the number of international crises in year y in which country c was a crisis actor. For each of these variables, Table 3 presents the parameter estimate, the marginal effect, and the Z-statistic based on cluster-corrected standard errors.

The results in Table 3 show that years in which a country becomes an actor in an international political crisis have a 2.8% higher probability of being the start of a consumption disaster. Years in which a country ends being an actor in an international political crisis, have a 2.3% lower probability of being at the start of a consumption disaster. The unconditional probability of a depression for our sample is 2.2% per annum, which implies that these marginal effects are economically significant. Both effects are also highly statistically significant.

2.4. Stock market data

Table 4 presents descriptive statistics of stock market index returns for the world market and for the 19 individual countries in this study. These are monthly stock returns based on general market price indices taken

[19] This number is calculated as the number of crisis actors involved in a full-scale war (200) divided by (167 times 89). The other probabilities in Table 2 are calculated in the same way.

[20] Some crises start relatively innocently but develop into a war. For example, out of the 96 crises that started with a political act like subversion, alliance formation by adversaries, diplomatic sanctions, severance of diplomatic relation, or violation of treaties, 17% escalated into a full-scale war. Similarly, 30% of the crises that started with a revolt in another country, or a violent act directed at an ally or a friendly state, developed into a war.

[21] The real GDP per capita data are from Angus Maddison (http://www.ggdc.net/Maddison/). For many war-actor countries, consumption data are missing, especially in the first half of our sample period.

Table 4
Basic characteristics of monthly stock returns.

Start date is the first month for which stock market price index data are available in Global Financial Data. *# Obs. missing* is the number of missing observations in the period between the start and end of the return series. *# Crises* is the total number of crises for each country in the period for which stock returns are available. The last row reports the averages across all 19 countries.

Country	Start date	Mean (%)	Standard deviation (%)	Minimum (%)	Maximum (%)	# Obs.	# Obs. missing	# Crises
World market index	1919:03	0.38	3.68	−20.14	14.06	1,055	0	441
Australia	1918:01	0.53	4.24	−55.24	20.10	1,068	0	3
Austria	1922:03	0.85	7.91	−39.78	114.37	963	37	4
Belgium	1926:01	0.39	5.08	−26.03	23.13	958	16	9
Canada	1918:12	0.45	4.65	−33.46	20.59	1,057	0	3
Columbia	1927:02	0.83	5.75	−22.58	64.08	959	0	2
Finland	1922:03	0.80	5.78	−32.56	35.16	1,014	5	6
France	1919:01	0.70	5.88	−27.61	63.04	1,044	12	31
Germany	1918:01	0.29	8.90	−146.07	69.02	1,025	59	24
India	1922:09	0.52	5.39	−25.72	35.06	1,000	2	18
Italy	1918:01	0.61	7.29	−30.74	46.79	1,038	2	12
Japan	1918:01	0.57	6.26	−31.83	50.87	1,058	11	15
Netherlands	1919:02	0.36	4.79	−26.59	22.50	1,031	26	11
Peru	1933:01	1.14	11.25	−30.38	115.41	888	1	6
Portugal	1934:01	0.79	7.10	−29.00	62.90	841	35	3
South Africa	1918:01	0.68	4.97	−35.14	21.67	1,068	0	4
Spain	1918:01	0.55	5.38	−33.48	80.70	1,025	44	8
Sweden	1918:01	0.51	4.97	−37.65	24.29	1,068	0	5
United Kingdom	1918:01	0.46	4.66	−30.92	42.32	1,068	0	43
United States	1918:01	0.50	5.42	−35.66	35.27	1,068	0	64
Average		0.61	6.08	−38.44	49.86	1015	11.8	14.3

from Global Financial Data.[22] The world stock price index series is based on the Morgan Stanley Capital International (MSCI) world index from 1970 onward. Before 1970, the world market return is the weighted average of the returns of individual countries.[23]

Our selection of countries is dictated by the availability of stock market data and the involvement of a country in international crises.[24] To be included in our sample, we require stock market data starting before 1935. We also require a country be involved in at least two crises in the period for which we have stock market data for that country.[25] For every country, we report the number of crises in which that country was an actor in the last column in Table 4.

As we use log returns, values lower than −100% are possible. This explains the high negative minimum return (−146%) for Germany. This occurred in July 1948, when Germany reformed its currency, and all financial assets were converted at a rate of one Deutschemark for ten Reichsmark, leading to an 80% decline in stocks. Not surprisingly, we find that the world market index has the lowest standard deviation.

3. Main results

3.1. World market mean returns

To measure the impact of international crises on stock returns, we test whether world market returns depend on the number of crises in any given month

$$r_t^{world} = \mu + \alpha_1 Crisis_t + \varepsilon_t, \tag{2}$$

where r_t^{world} is the return on the world market in month t, and ε_t denotes the error term. As there is no world market interest rate, we focus on total returns rather than excess returns. (We find similar results if we use excess returns calculated using the three-month U.S. T-bill rates as a proxy for the risk free rate.) We report the regression results of Eq. (2) in Table 5.

From Table 5, we see an economically and statistically significant negative relation between world stock market returns and the aggregate number of international crises in any month. On average, a crisis reduces monthly stock returns by 0.12% (White-adjusted t-value is −2.03). Since there are, on average, 2.47 crises per month, our point estimate implies a negative return due to international crises of 0.30% per month, or 3.6% annually. Using this

[22] We use price indices instead of total return indices, because price data are available over a longer period. The results are similar if we use the shorter total return series.

[23] The weights in January 1919 are as follows: North America 44% (the United States 41%, Canada 3%), Europe 44% (the United Kingdom 12%, Germany 8%, France 8%, Italy 4%, Switzerland 2.5%, the Netherlands 2.5%, Belgium 2%, Spain 2%, Denmark 1%, Norway 1%, and Sweden 1%), Asia and the Far East 12% (Japan 6%, India 2%, Australia 2%, South Africa Gold 1%, and South Africa Industrials 1%). Global Financial Data used these weights from January 1919 until the MSCI indices were introduced in 1970.

[24] For some individual countries, markets were sometimes closed in reaction to severe crises. For example, we have no stock market returns for several European countries for early 1940, when markets were closed in reaction to the German invasion. In these cases, our use of the world market return is likely to underestimate the total impact of the crisis on stock prices.

[25] As a consequence of the latter restriction, we exclude Denmark, Norway, and Switzerland from our analysis.

Table 5
The impact of crises on stock returns.

This table contains the estimation results for model (2): $r_t^{world} = \mu + \alpha_1 Crisis_t + \varepsilon_t$, where $Crisis_t$ is the total number of crises in month t, and r_t^{World} is the world stock market return in month t (in percent). t-Statistics are based on heteroskedasticity-consistent standard errors. t-Statistics in bold denote significance at least at the 10% level.

	Estimate (%)	t-Value
Constant, μ	0.674	**3.73**
Crisis, α_1	−0.120	−**2.03**

information, we can also calculate the estimated average cost per crisis, which is equal to 0.72% (the cost per month (0.3%) times the number of months (1,068) divided by the number of crises (441)).

To test whether investors respond negatively to the start of a crisis and positively to the end of a crisis, we distinguish between start effects, effects during the crisis, and end effects

$$r_t^{world} = \mu + \alpha_1 Start_t + \alpha_2 During_t + \alpha_3 End_t + \varepsilon_t. \quad (3)$$

In the first row in Table 6, we report the results for regression (3), where we use all crises, irrespective of the nature of the crisis. The other rows in Table 6 present the results of specifications designed to test the hypothesis that the reaction of investors is stronger if the probability of a drastic reduction in consumption is higher. In these tests, we restrict the crises used in the estimation of model (3) to crises in one of the following subcategories: *Violent crises*, *Wars*, *Violent break*, *Grave threat*, *Protracted conflicts*, and *Major power* involvement. The last row in Table 6 gives the results, based on the crisis severity index.

Consistent with the rare disaster models, the increase in crisis probability at the start of a crisis has a significant and negative impact on stock returns. World market returns are almost half a percent lower (0.43%), with a t-value of −2.63. During crises, we see a monthly loss of 0.15% that is less dramatic than the *Start* effect and significant at the 10% level (t-value of −1.88). Finally, in months when crises end, we see a positive effect. Average monthly stock returns are 0.29% higher in those months. Again, the t-value of 1.73 indicates that this effect is statistically significant at the 10% level.

Stock prices react more strongly to the start and end of more serious crises. The absolute value of the coefficient for *Start* and *End* is higher when we limit the sample to more severe crises.[26] For instance, the start of a full-scale war has an average impact of −1.25%, which is almost three times as large as in the average crisis. Similarly, and consistent with the observations in Brecher and Wilkenfeld (1997), investors seem to perceive crises related to protracted conflicts as more serious. In the last row of Table 6, we report the results of a regression that

Table 6
Start, during, and end effects of crises on world stock market returns.

Estimation results for model (3): $r_t^{world} = \mu + \alpha_1 Start_t + \alpha_2 During_t + \alpha_3 End_t + \varepsilon_t$, where $Start_t$ and End_t denote the number of crises that start and end in month t, respectively, and $During_t$ is the total number of crises that occurs in month t, excluding all crises that start or end in that month. r_t^{World} is the return on the world market in month t (in percent). We report results for all crises and subsets based on different measures of crisis severity, as defined in Table 1. t-Statistics are based on heteroskedasticity-consistent standard errors. t-Statistics in bold denote significance at least at the 10% level.

	Constant	*Start*	*During*	*End*
All crises				
Coefficient (%)	0.680	−0.428	−0.151	0.292
t-Statistic	**3.76**	−**2.63**	−**1.88**	**1.73**
Violent crises				
Coefficient (%)	0.566	−0.754	−0.134	0.449
t-Statistic	**3.89**	−**2.60**	−1.34	**1.71**
War				
Coefficient (%)	0.516	−1.254	−0.092	0.326
t-Statistic	**3.63**	−**3.34**	−0.59	0.92
Violent break				
Coefficient (%)	0.313	−0.659	0.044	0.874
t-Statistic	**2.19**	−**2.25**	0.28	**3.09**
Grave threat				
Coefficient (%)	0.718	−0.638	−0.396	0.337
t-Statistic	**4.96**	−**2.21**	−**2.82**	1.25
Major power				
Coefficient (%)	0.680	−0.702	−0.450	0.157
t-Statistic	**5.04**	−**2.14**	−**3.60**	0.54
Protracted				
Coefficient (%)	0.528	−0.578	−0.194	0.652
t-Statistic	**3.43**	−**2.49**	−**1.74**	**3.18**
Crisis severity index				
Coefficient (%)	0.712	−0.153	−0.045	0.104
t-Statistic	**4.24**	−**3.06**	−**2.14**	**2.29**

uses the crisis severity index. As expected, we find significantly negative coefficients for *Start* and *During* and a significantly positive coefficient for *End* (all significant at the 5% level). We conclude that more severe crises, with a higher probability of a dramatic negative impact on future consumption, have a stronger impact on stock returns.

3.2. World market volatility effects

According to Wachter (2009), an increase in disaster probability results in an increase in stock market volatility, and a decrease in disaster probability results in a decrease in stock market volatility. Veronesi (2004) develops a model with dynamic learning in which investors know that there is a very small probability that the economy may enter into a serious recession. Consistent with a suggestion in Schwert (1989), this model also shows that volatility is not only related to single events but also depends on global political stability.

To test whether the volatility of world market returns is related to the occurrence of international crises, we estimate a GARCH(1,1) model with exogenous

[26] In further tests, we find that for the subgroups *Violent crises*, *Wars*, and *Grave threats*, the stock price reaction to the start of a crisis is significantly stronger (at the 5% level) than the stock price reaction to the start of remaining crises.

Table 7
World stock market returns and international crises: GARCH(1,1)-X models.

Estimates of the GARCH-X models: $r_t^{world} = \mu + \alpha_1 Start_t + \alpha_2 During_t + \alpha_3 End_t + \varepsilon_t,\ \varepsilon_t \sim N(0,\sigma_t^2),\ \sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2 + \beta_3 Start_t + \beta_4 During_t + \beta_5 End_t$. $Start_t$ and End_t denote the number of crises that start and end in month t, respectively, and $During_t$ is the total number of crises that occurs in month t, excluding all crises that start or end in that month. r_t^{World} is the return on the world market in month t (in percent). We report results for all crises and subsets based on different measures of crisis severity, as defined in Table 1. t-Statistics in bold denote significance at least at the 10% level.

	Mean				Volatility					
	Const.	*Start*	*During*	*End*	Const.	β_1	β_2	*Start*	*During*	*End*
All crises										
Coeff. (%)	0.882	−0.426	−0.237	0.231	0.298	0.084	0.874	1.930	0.082	−1.500
t-Statistic	**5.81**	**−2.83**	**−3.27**	1.46	1.59	**3.47**	**27.61**	**2.87**	0.83	**−2.46**
Violent crises										
Coeff. (%)	0.776	−0.753	−0.262	0.376	0.406	0.085	0.869	2.980	0.014	−1.690
t-Statistic	**6.26**	**−2.72**	**−2.97**	**1.65**	**2.25**	**3.36**	**25.95**	**2.66**	0.10	**−1.66**
War										
Coeff. (%)	0.726	−1.100	−0.347	0.440	0.376	0.086	0.881	4.630	−0.218	−1.960
t-Statistic	**6.19**	**−2.77**	**−2.51**	1.37	**2.51**	**4.28**	**37.71**	**3.14**	−1.53	**−1.88**
Violent break										
Coeff. (%)	0.592	−0.790	−0.161	0.855	0.298	0.088	0.863	2.270	0.015	0.006
t-Statistic	**5.01**	**−2.69**	−1.24	**2.91**	1.50	**2.90**	**18.78**	1.58	0.05	0.00
Grave threat										
Coeff. (%)	0.931	−0.693	−0.626	0.252	0.271	0.089	0.854	1.810	0.447	−0.730
t-Statistic	**7.73**	**−2.71**	**−4.64**	1.14	**1.72**	**3.33**	**23.37**	1.25	1.46	−0.59
Major power										
Coeff. (%)	0.710	−0.897	−0.434	0.224	0.201	0.092	0.882	3.810	0.090	−2.510
t-Statistic	**5.94**	**−2.70**	**−3.43**	0.78	1.18	**4.16**	**34.49**	**3.30**	0.67	**−2.29**
Protracted										
Coeff. (%)	0.674	−0.578	−0.298	0.535	0.363	0.104	0.854	2.500	0.208	−2.190
t-Statistic	**5.20**	**−2.58**	**−2.28**	**2.89**	**1.72**	**4.12**	**24.68**	**2.66**	1.24	**−2.57**
Crisis severity index										
Coeff. (%)	0.899	−0.156	−0.068	0.096	0.225	0.086	0.871	0.562	0.017	−0.384
t-Statistic	**6.27**	**−3.39**	**−3.53**	**2.31**	1.41	**3.57**	**27.51**	**3.31**	0.71	**−2.85**

regressors.[27] We use GARCH models, as we are interested in volatility effects but at the same time want to control for the effect international crises may have on mean returns

$$r_t^{world} = \mu + \alpha_1 Start_t + \alpha_2 During_t + \alpha_3 End_t + \varepsilon_t,$$
$$\varepsilon_t \sim N(0,\sigma_t^2),$$
$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_t^2 + \beta_2 \sigma_{t-1}^2 + \beta_3 Start_t + \beta_4 During_t + \beta_5 End_t + \varepsilon_t. \tag{4}$$

In this specification, the variance of the error term is allowed to depend on past squared residuals, past variance, and $Start_t$, $During_{t,}$, and End_t. We expect that $Start_t$ is positively related to volatility and that End_t is negatively related to stock market volatility.

The results for this extended GARCH(1,1) model are found in Table 7.[28] Table 7 also presents the results of similar models, using different crisis definitions depending on the severity of the crisis, and the last row gives the results using the crisis severity index.

We make three observations based on the results in Table 7. First, focusing on the mean, we find that both start and end effects are comparable with the results in Table 6. However, after modeling volatility more explicitly, the mean effect of the *During* variable is also significant in most cases. Second, there is conditional heteroskedasticity in this long series of monthly stock returns. The parameter estimates for both the lagged squared error term and the lagged variance are statistically significant. Third, the results offer strong support for the hypothesis that international crises increase world market volatility: The start of every additional international crisis significantly increases monthly variance, and every reduction in international crises significantly decreases the variance. We conclude that international

[27] See, for example, Engle and Patton (2001) for a discussion of these models. There are numerous variations of GARCH models, or as Engle (2002) puts it: "The alphabet soup of volatility models continually amazes." We use the GARCH(1,1) model with exogenous regressors, as it is simple and robust (Engle, 2002).

[28] Our main conclusions are robust with respect to alternative variations of the GARCH model. For instance, if we use the E-GARCH model, we find similar results. We also conduct a simulation analysis to establish if our estimates are consistent and unbiased. In each month t, we generate a random error from $N(0,\sigma_t^2)$, where σ_t^2 is computed from the variance equation with the estimated parameters in Table 7. The error terms are then added to the mean equation to form a monthly return series on which the GARCH model is estimated. The process is repeated 1,000 times. The results indicate that the median and the mean of the estimated GARCH model based on the simulated return series are very close to the true parameters. (The simulation results are available on request.)

political uncertainty has a significant impact on world stock market volatility.

3.3. Crisis risk and expected returns

3.3.1. Time-series evidence: predictive regressions

According to time-varying disaster models, the expected stock market excess return is an increasing function of expected rare disaster risk. Following the methodology in French, Schwert, and Stambaugh (1987) and Amihud (2002), we proxy the level of expected disaster risk with an AR(1) model and use this estimate to test the following two hypotheses:

H1. The expected stock market excess return is an increasing function of expected disaster risk.

H2. Unexpected disaster risk has a negative effect on contemporaneous stock market return.

Our measure of disaster risk at the start of month t is the sum of the number of ongoing crises in month $t-1$ and the number of crises that started in month $t-1$.[29] We assume investors predict disaster risk for month t based on the following AR(1) process:

$$Start_t + During_t = \alpha + \beta[Start_{t-1} + During_{t-1}] + \varepsilon_t. \quad (5)$$

Using the number of crises as a crisis measure, the estimate for the intercept in model (5) is 0.361 (t-statistic=8.1), and the estimated slope coefficient is 0.827 (t-statistic=48.4). A Chow test shows that the coefficients of the model are stable over time. We therefore interpret the fitted value from model (5) estimated over the entire sample period as the level of expected disaster risk, and the residual from model (5) is our measure of unexpected disaster risk. To test the hypotheses at the start of this section, we estimate the following model:

$$r_t^{world} = \alpha + \beta_1 Expected\ Disaster\ Risk_t + \beta_2\ Unexpected\ Disaster\ Risk_t + \varepsilon_t. \quad (6)$$

Inclusion of unexpected disaster risk in model (6) is important because, as shown in the previous section, it helps explain variations in market returns over time. Moreover, the coefficient of the unexpected disaster risk variable provides indirect evidence of the relation between the expected level of disaster risk and the expected market risk premium. Following the reasoning in French, Schwert, and Stambaugh (1987) and Amihud (2002), a positive relation between the expected level of disaster risk and the expected market risk premium induces a negative relation between unexpected disaster risk and excess market returns. Time-varying disaster risk models therefore predict that $\beta_1 > 0$ and $\beta_2 < 0$.

The results for model (6) using the monthly world market return as the dependent variable are shown in Panel A of Table 8. Panel B of Table 8 presents the results with the excess return for the U.S. stock market as the dependent variable.[30] The U.S. market excess return is of interest because of the availability of reliable excess return data for most of our sample period, and also because of the prominence of this variable in the finance literature. Note that for each of the alternative crisis measures in Table 8, we re-estimate model (5) to obtain the expected and unexpected values for that particular crisis measure.

The results in Table 8 provide strong support for the second hypothesis. Consistent with the results in Table 6, innovations in disaster risk have a very significant negative contemporaneous impact on stock market returns. However, for both world market returns and U.S. stock market excess returns, the coefficient of the expected disaster risk measure is almost always insignificant. Using predictive regressions, we conclude that there is no evidence to directly support the hypothesis that the expected stock market excess return is an increasing function of expected disaster risk.

3.3.2. Time-series evidence: valuation ratios

Several authors point out that realized market returns are a noisy proxy for expected market returns and propose alternative measures for expected market returns.[31] Time-varying disaster risk models predict that when the probability of disasters is high, the earnings–price ratio and dividend yield are high because expected stock returns relative to bonds are high (see, e.g., Gabaix, 2009; Wachter, 2009). Therefore, as an alternative test of the hypothesis that the expected stock market excess return is an increasing function of expected rare disaster risk, we examine the correlations between the E/P ratio and the dividend yield and our measures of disaster risk.

Fig. 1 provides visual evidence of the positive correlation between the number of crises over our sample period and the earnings–price ratio of the S&P 500 index (*source*: Global Financial Data). Consistent with the visual evidence in Fig. 1, we find that the correlation between the number of crises and the E/P ratio is significantly positive,

[29] Recall that $During_{t-1}$ is the number of ongoing crises in month $t-1$, where month $t-1$ is not the start or end month of the crisis. The number of crises at the start of month t is therefore equal to $Start_{t-1} + During_{t-1}$.

[30] The U.S. market excess return is the value-weighted return of all NYSE, AMEX, and NASDAQ stocks from CRSP, minus the one-month Treasury bill rate from Ibbotson Associates. We obtain these data, and the portfolio returns used in the next section, from Kenneth French's Web site at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[31] For example, Fama and French (2002, p. 643) state,

> Indeed, an advantage of the expected return estimates from fundamentals is that they are likely to be less sensitive than the average return to long lived shocks to dividend and earnings growth rates or the expected stock return. In short, the estimates of the expected stock return from fundamentals are likely to be more precise than the average stock return.

In addition, Elton (1999, p. 1200) states,

> I believe that developing better measures of expected return and alternative ways of testing asset pricing theories that do not require using realized returns have a much higher payoff than any additional development of statistical tests that continue to rely on realized returns as a proxy for expected returns.

Table 8
Expected stock market returns and international crises.
Estimation results for model (6):

$$r_t^{world} = \alpha + \beta_1 \ Expected\ Disaster\ Risk_t + \beta_2 \ Unexpected\ Disaster\ Risk_t + \varepsilon_t. \tag{6}$$

Expected Disaster Risk is the fitted value and *Unexpected Disaster Risk* is the residual from the following model (5):

$$Start_t + During_t = \alpha + \beta[Start_{t-1} + During_{t-1}] + \varepsilon_t. \tag{5}$$

$Start_t$ denotes the number of crises that start in month t, and $During_t$ is the total number of crises that occurs in month t, excluding all crises that start or end in that month. The results for the world market return are in Panel A, and the results for the excess return on the U.S. stock market are in Panel B. Note that for each of the alternative crisis measures, we re-estimate model 5 to obtain expected and unexpected values for that particular crisis measure. We report results for all crises and subsets based on different measures of crisis severity, as defined in Table 1. *t*-Statistics are based on heteroskedasticity-consistent standard errors. *t*-Statistics in bold denote significance at least at the 10% level.

	Panel A: World stock market			Panel B: U.S. equity premium		
	Constant	Expected	Unexpected	Constant	Expected	Unexpected
All						
Coefficient (%)	0.559	−0.089	−0.413	0.419	0.113	−0.456
t-Statistic	**2.67**	−1.04	**−3.23**	1.20	0.75	**−2.28**
Violent						
Coefficient (%)	0.148	0.297	−0.631	0.376	0.237	−0.658
t-Statistic	0.87	**1.82**	**−3.22**	1.38	1.28	**−2.21**
War						
Coefficient (%)	0.416	−0.055	−0.816	0.343	0.454	−1.240
t-Statistic	**2.63**	−0.34	**−2.98**	1.41	**1.80**	**−2.90**
Violent break						
Coefficient (%)	0.480	−0.089	−0.622	0.408	0.313	−0.742
t-Statistic	**2.89**	−0.86	**−3.27**	1.54	1.24	**−2.50**
Grave threat						
Coefficient (%)	0.656	−0.307	−0.668	0.554	0.106	−0.670
t-Statistic	**3.83**	**−2.15**	**−3.47**	**2.02**	0.44	**−2.24**
Major power						
Coefficient (%)	0.397	−0.171	−0.763	0.617	0.164	−0.824
t-Statistic	**3.12**	−0.38	**−2.47**	**3.11**	0.23	**−1.67**
Protracted						
Coefficient (%)	0.434	−0.053	−0.625	0.497	0.136	−0.875
t-Statistic	**2.49**	−0.45	**−3.70**	**1.82**	0.73	**−3.40**
Crisis severity index						
Coefficient (%)	0.563	−0.025	−0.144	0.351	0.041	−0.172
t-Statistic	**2.83**	−1.14	**−4.24**	1.06	1.03	**−3.23**

at 0.21 (p-value is 0.001). Similarly, the correlation between the crisis severity index and the earnings–price ratio is highly significant at 0.25. When we consider the dividend yield, we again find highly significant correlations: The correlation between the number of crises and the dividend yield is 0.14, and the correlation between the crisis severity index and the dividend yield is 0.20.[32] We conclude that the evidence based on the correlations between these valuation ratios and our crisis measures is consistent with time-varying disaster models, suggesting that expected market returns are higher during periods of high crisis risk.

3.3.3. *Cross-sectional evidence*

A natural cross-sectional implication of disaster models is that assets that do better during a disaster should have lower expected returns. Since actual disasters are rare, we again use time variation in our crisis measures to test the hypothesis that assets that do relatively well when there is an increase in disaster probability should have lower expected returns (see Gabaix, 2009; Gourio, 2008b).

We run two-step Fama-MacBeth (1973) regressions to study the cross-sectional relation between crisis sensitivity and one-month-ahead returns across the 30 Fama-French industry portfolios. Our choice of value-weighted industry portfolios as test assets is motivated by the finding that standard asset pricing models do a poor job explaining the substantial heterogeneity in expected returns among these portfolios (Fama and French, 1997; Lewellen, Nagel, and Shanken, 2010).

To measure crisis sensitivity for each industry, we include unexpected changes in disaster probability, that is, innovations obtained from model (5), as an additional

[32] Using expected disaster estimates from model (5), instead of the actual number of crises, yields almost identical results. When we use the natural log of the price-earnings ratio or the price-dividend ratio, the correlations with the number of crises are −0.17 and −0.15, respectively, and the correlations with the crisis severity index are −0.21 and −0.19, respectively. All these correlations are significant at the 1% level.

factor in the Fama-French three-factor model. Thus, following other research in this area, we use the previous 60 monthly observations ($\tau = t-60$ to $t-1$) and run the following time-series regression to estimate factor loadings for the Fama-French factors and the crisis sensitivity for each industry portfolio i:

$$r_{i,\tau} = \alpha_{i,t-1} + \beta_{i,t-1}^{MKTRF} MKTRF_\tau + \beta_{i,t-1}^{SMB} SMB_\tau + \beta_{i,t-1}^{HML} HML_\tau + \beta_{i,t-1}^{Crisis} Crisis_\tau + \eta_{i,\tau}, \quad (7)$$

where $r_{i,\tau}$ is the excess return on industry portfolio i in month τ; $MKTRF_\tau$, SMB_τ and HML_τ are the Fama-French market, size, and B/M factors, respectively, in month τ; and $Crisis_\tau$ is the innovation in the crisis measure in month τ, where the innovation is the residual in model (5) for that crisis measure.

Before we run the second-stage cross-sectional regressions, we first transform the estimated crisis sensitivities, $\beta_{i,t-1}^{Crisis}$, into decile ranks each month and scale them back to range between zero and one. This transformation into decile ranks renders our results less sensitive to measurement errors and facilitates the interpretation of coefficient estimates (see, e.g., Nagel, 2005). Next, we estimate the following cross-sectional regression in each month t:

$$r_{i,t} = \gamma_t + \gamma_{MKTRF,t}\beta_{i,t-1}^{MKTRF} + \gamma_{SMB,t}\beta_{i,t-1}^{SMB} + \gamma_{HML,t}\beta_{i,t-1}^{HML} + \gamma_{Crisis,t}\beta_{i,t-1}^{Crisis} + \varepsilon_{i,t}. \quad (8)$$

Finally, we average the risk premium associated with each factor over the sample period and report the results in Table 9.

The coefficient estimates shown in Table 9 are the time-series averages of monthly estimates of risk premiums. Standard errors are adjusted for autocorrelation using the Newey-West method. Table 9 shows that estimated risk premiums for the Fama-French factors are positive but statistically insignificant. The risk premium for crisis sensitivity is always negative, and, with the exception of crises involving major powers and protracted crises, the crisis risk premiums are statistically significant. Thus, portfolios that do relatively well when there is an increase in disaster probability yield lower returns, on average.

The coefficients for the two broad crisis measures (all crises and the crisis severity index) are more than three standard errors away from zero and imply that the average difference in returns between the top decile and bottom decile portfolios in terms of crisis sensitivity is about 0.3% per month. Because we use U.S. industry portfolios, we also restrict the crisis definition to crises with U.S. involvement. The results are in the last two rows of Table 9, and again show a significantly negative coefficient for crisis risk sensitivity. We conclude the cross-sectional asset pricing tests are consistent with the hypothesis that crisis risk is priced: industries that are more crisis risk sensitive yield a higher return, on average.[33]

Table 9
Risk premiums for Fama-French factor loadings and crisis sensitivity.

This table reports average estimated risk premiums from second-stage cross-sectional regressions for each month t:

$$r_{i,t} = \gamma_t + \gamma_{MKTRF,t}\beta_{i,t-1}^{MKTRF} + \gamma_{SMB,t}\beta_{i,t-1}^{SMB} + \gamma_{HML,t}\beta_{i,t-1}^{HML} + \gamma_{Crisis,t}\beta_{i,t-1}^{Crisis} + \varepsilon_{i,t}. \quad (8)$$

Factor loadings for each of the 30 Fama-French industry portfolios in each month are estimated from first-stage time-series regressions over the previous 60 months ($\tau = t-60$ to $t-1$):

$$r_{i,\tau} = \alpha_{i,t-1} + \beta_{i,t-1}^{MKTRF} MKTRF_\tau + \beta_{i,t-1}^{SMB} SMB_\tau + \beta_{i,t-1}^{HML} HML_\tau + \beta_{i,t-1}^{Crisis} Crisis_\tau + \eta_{i,\tau}. \quad (7)$$

$Crisis_t$ is the unexpected change in disaster probability proxied by the innovation obtained from model (5) for the corresponding type of crisis. The results are based on industry crisis sensitivities, $\beta_{i,t-1}^{Crisis}$, that are transformed each month into decile ranks (scaled back to range between zero and one). We report results for all crises and subsets based on different measures of crisis severity, as defined in Table 1. t-Statistics presented are based on Newey-West standard errors. t-Statistics in bold denote significance at least at the 10% level.

	MKTRF	SMB	HML	Crisis
All crises				
Coefficient	0.1036	0.2137	0.1330	−0.3171
t-Statistic	0.45	1.37	0.83	**−3.79**
Violent crises				
Coefficient	0.1305	0.2289	0.1045	−0.2241
t-Statistic	0.57	1.43	0.65	**−2.49**
War				
Coefficient	0.1458	0.2364	0.1279	−0.2011
t-Statistic	0.63	1.49	0.83	**−2.45**
Violent break				
Coefficient	0.0507	0.2049	0.0927	−0.1685
t-Statistic	0.22	1.32	0.60	**−2.10**
Grave threat				
Coefficient	0.1516	0.2273	0.1337	−0.1491
t-Statistic	0.65	1.46	0.85	**−1.71**
Major power				
Coefficient	0.1120	0.2309	0.1741	−0.0092
t-Statistic	0.48	1.48	1.07	−0.10
Protracted				
Coefficient	0.0375	0.2192	0.1705	−0.1227
t-Statistic	0.16	1.36	1.05	−1.37
Crisis severity index				
Coefficient	0.1415	0.2202	0.1367	−0.2901
t-Statistic	0.61	1.40	0.85	**−3.27**
US crises				
Coefficient	0.1275	0.2272	0.1172	−0.1359
t-Statistic	0.55	1.47	0.75	**−2.18**

4. Additional tests

4.1. Consumption growth

The results in the previous section are consistent with recent theoretical models in which variation in expected

[33] Our conclusions are robust with respect to alternative variations of the two-step regressions. If estimated crisis sensitivities are not transformed into decile ranks, all estimates for the crisis risk premium are negative and significantly different from zero for seven out of nine types of crisis measures. Very similar results are obtained if we exclude actual consumption disasters in the United States and start the sample in 1948, or if we use $Start_t$ as a measure of crisis innovation in the first-stage regression (7) instead of the residual from model (5). Expanding the test asset pool with the 25 Fama-French portfolios based on size and book-to-market ratio also leads to qualitatively similar results. The results of these robustness tests are available on request.

Table 10
Consumer confidence and international crises.

We use the University of Michigan Consumer Sentiment Index (MCSI) as our measure of consumer confidence and estimate the following model:

$$MCSI_t = \mu + \alpha_1 Start_t + \alpha_2 During_t + \alpha_3 End_t + \varepsilon_t,$$

where $Start_t$ and End_t denote the number of crises that start and end in month t, respectively, and $During_t$ is the total number of crises that occurs in month t excluding all crises that start or end in that month. $MCSI_t$ is the value of the MCSI in month t and has been available since January 1978. We report results for all crises and subsets based on different measures of crisis severity, as defined in Table 1. t-Statistics in bold denote significance at least at the 10% level.

	Constant	*Start*	*During*	*End*
All				
Coefficient	95.53	−2.05	−3.19	−2.43
t-Statistic	**128.09**	**−2.37**	**−8.60**	**−3.10**
Grave threat				
Coefficient	89.48	−3.90	0.02	−2.72
t-Statistic	**104.86**	**−2.56**	0.02	**−1.88**
Major power				
Coefficient	90.94	−0.98	−11.52	0.33
t-Statistic	**138.77**	−0.44	**−8.74**	0.14
Violent				
Coefficient	91.36	−3.79	−1.68	−4.53
t-Statistic	**121.99**	**−2.65**	**−2.61**	**−3.17**
War				
Coefficient	88.64	−6.80	0.59	−7.10
t-Statistic	**106.89**	**−2.79**	0.60	**−3.30**
Violent break				
Coefficient	89.48	−1.78	−0.60	−2.48
t-Statistic	**109.05**	−1.35	−0.82	**−1.91**
Protracted				
Coefficient	92.29	−2.17	−3.77	−2.44
t-Statistic	**121.49**	**−1.66**	**−6.39**	**−1.97**
Crisis severity index				
Coefficient	93.64	−0.677	−0.609	−0.706
t-Statistic	**126.25**	**−2.49**	**−4.80**	**−2.99**
U.S. crises				
Coefficient	89.02	−1.19	−7.38	3.55
t-Statistic	**130.79**	−0.36	**−3.55**	1.42

returns over time is generated by a time-varying probability of disaster. This section extends the analysis by examining the underlying mechanism in these models: the relation between changes in disaster probability and expected future consumption growth.[34]

We present two sets of results. In the first set of tests, we use consumer confidence as proxy for expected consumption growth, and in the second set of tests, we use expected growth in real GDP as proxy for expected consumption growth. Unfortunately, international data for a sufficiently long period are unavailable, and we are restricted to U.S. data.

The University of Michigan's Consumer Sentiment Index (MCSI) is one of the most widely followed measures of U.S. consumer confidence and has been used extensively in academic research. Ludvigson (2004) provides details on measurement and reporting of this consumer confidence index. He also presents evidence that consumer confidence contains information about future consumption growth. Monthly consumer confidence data are available from January 1978 forward. We use this monthly index and estimate a model similar to regression (3), with the consumer confidence index as dependent variable.

Focusing on all crises, we find that in the period 1978–2006, U.S. consumer confidence is significantly negatively related to the start of crises and deteriorates further during crises. We also find that the negative impact of crises on consumer confidence tends to linger even after crises end. A possible explanation for this result is that, on average, survey respondents are skeptical about whether these crises have really been resolved and remain pessimistic. We observe similar results when the independent variables are based on the stand-alone crisis severity variables, or on the crisis severity index. Because we use survey data on U.S. consumer confidence, we also report the results for a smaller subset of crises in which the U.S. was involved. These results are in the last two rows in Table 10 and show that *During* is still significant and negative, *Start* is negative and insignificant, and *End* is positive and insignificantly different from zero.

Table 11 shows the impact of disaster probability on forecast GDP growth. GDP growth forecasts are obtained from the Survey of Professional Forecasters which has been available on a quarterly basis since the last quarter of 1968.[35] To obtain estimates of real GDP growth, we use identification codes for each forecaster to match expectations of price index and nominal GDP. We use the mean real GDP-growth forecast averaged across all forecasts in each quarter as our measure of forecast GDP growth. To align the GDP forecasts with our crisis variables, we use the crisis values in the first month of each quarter.[36] GDP growth forecasts are available for the next five quarters. Because we find similar results for all horizons, we only report results for forecast GDP growth over the next year.

The results in Table 11 indicate a significant negative impact of ongoing crises on forecast GDP growth for all crisis measures apart from *Violent break* and *Protracted*. When the independent variables are based on the crisis severity index, we again find a significant negative impact of ongoing crises. The last two rows of Table 11 present the results for the smaller subset of crises with U.S. involvement. These results show that the start of a crisis has a significant negative impact on next year's forecast GDP growth, as do continuing crises. Crises with U.S. involvement that come to an end tend to increase forecast GDP, but this effect is not statistically significant.

The criteria used by ICB to select the crises in their database and the results in Section 2.3, suggest a direct link between our crisis probability measures and actual

[34] We thank an anonymous referee for suggesting this analysis.

[35] These data are from http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/. The same data set is used by Buraschi and Jitsov (2006).

[36] The survey is sent after the release of the U.S. Bureau of Economic Analysis' advance report of the national income and product accounts, at the end of the first month of each quarter. After 1990, the deadline for responses is before the end of the middle month of each quarter.

Table 11
Annual GDP forecasts and international crises.
This table shows the results of the following model:

$$Forecast\ \ GDP\ \ Growth_t = \mu + \alpha_1 Start_t + \alpha_2 During_t + \alpha_3 End_t + \varepsilon_t.$$

GDP-growth forecasts for the next year are obtained from the Survey of Professional Forecasters, which has been available on a quarterly basis since the last quarter of 1968 (http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters). $Start_t$ and End_t denote the number of crises that start and end in month t, respectively, and $During_t$ is the total number of crises that occurs in month t, excluding all crises that start or end in that month. We report results for all crises and subsets based on different measures of crisis severity, as defined in Table 1. t-Statistics in bold denote significance at least at the 10% level.

	Constant	*Start*	*During*	*End*
All				
Coefficient	3.175	0.049	−0.290	−0.055
t-Statistic	**19.82**	0.33	**−2.97**	−0.36
Grave threat				
Coefficient	2.915	−0.143	−0.258	0.084
t-Statistic	**16.64**	−0.64	**−1.94**	0.39
Major power				
Coefficient	2.900	0.302	−0.820	0.212
t-Statistic	**25.73**	0.49	**−3.58**	0.63
Violent				
Coefficient	3.020	−0.091	−0.258	−0.147
t-Statistic	**21.03**	−0.29	**−2.14**	−0.49
War				
Coefficient	2.883	0.215	−0.280	−0.295
t-Statistic	**18.96**	0.41	**−1.98**	−0.67
Violent break				
Coefficient	2.765	−0.068	−0.052	−0.061
t-Statistic	**18.69**	−0.25	−0.41	−0.30
Protracted				
Coefficient	2.814	−0.132	−0.160	0.228
t-Statistic	21.17	−0.62	−1.13	0.93
Crisis severity index				
Coefficient	3.111	−0.009	−0.066	−0.004
t-Statistic	**20.12**	−0.19	**−2.72**	−0.08
U.S. crises				
Coefficient	2.814	−0.748	−0.739	0.452
t-Statistic	**23.46**	**−1.76**	**−1.89**	1.25

consumption disasters. Consequently, we attribute the negative relation between changes in crisis risk and expected future consumption growth to changes in the (unobservable) probability of a consumption disaster. However, this interpretation does not rule out the possibility that increased political instability also has a negative impact on long-term expected consumption growth and, therefore, on stock prices (see Bansal and Yaron, 2004). In addition, it is possible that international political instability creates uncertainty about expected consumption growth, which might also negatively impact stock prices and increase volatility (see Bansal and Shaliastovich, 2010).

4.2. Subsamples and event windows

To test whether our results are affected by the actual disasters used in the calibrations in Barro (2006), that is, crises during World War I, the Great Depression, and World War II, we re-estimate the GARCH(1,1) model (4) on a sample that starts in January 1948. The results are found in Table 12.

Consistent with our earlier results, we find that in the post-war period, world stock markets react negatively when crises start and continue, and that world stock markets react positively when crises end. Furthermore, for the post-war sample we also find that the start of international crises increases the volatility of world market returns, and for each crisis that comes to an end, volatility decreases.

Our tests thus far are based on the assumption that the start and end dates in the ICB database are correct. However, investors may anticipate crises earlier than indicated in the ICB data (in the next section, when we consider individual country results, we provide an example based on the *Anschluss* (annexation) between Austria and Germany). Moreover, investors may be skeptical that a crisis has really been resolved and delayed end effects may appear. To establish the sensitivity of our results to the trigger and end dates in the ICB database, we include two additional variables in our model: $Start_{t+1}$, which equals $Start_t$ with a one-month lead; and End_{t-1}, which equals End_t with a one-month lag. Table 13 reports the results.

We find that $Start_{t+1}$ and End_{t-1} are insignificant, indicating that current returns are not affected by future values of *Start* and lagged values of *End*.

4.3. Alternative definitions of the world market index

To establish whether our results are robust to the definition of the world stock market index, we construct two alternative indices. The first alternative world market index gives equal weight to all 19 countries in our sample. The second alternative world market index is meant to deal directly with the fact that for the period 1919–1970, the world market index reported in Global Financial Data is based on constant weights. To better reflect the shifting economic importance of the constituent countries in this period, we calculate the weighted world market return, in which we re-weigh each country annually based on its GDP.[37]

The results for the two alternative definitions of the world market index are found in Table 14.

Based on the equally weighted index, the coefficients for *Start* and *During* are negative in all regressions and generally significant. With the exception of *Major power*, the estimated coefficient for *End* is positive as expected, and significant in two of the eight models. When we use the GDP-weighted world stock market return, we confirm significantly negative start effects and find that end effects are always positive and generally significant. With the exception of *Violent break* and *War*, the coefficient for *During* is negative, but it is only significantly different from zero for *Violent crises*.

[37] We use the annual GDP per capita data from Barro and Ursúa (http://www.economics.harvard.edu/faculty/barro/data_sets_barro) and population and GDP (2006) data from Angus Maddison (http://www.ggdc.net/maddison) to construct a stock market index that is annually re-weighted based on relative GDPs of all countries in our study.

Table 12
World market returns and international crises: GARCH estimates 1948–2006.

This table reports the estimates of the GARCH-X models estimated over the period 1948–2006. Full sample results are reported in Table 7. The GARCH model is $r_t^{world} = \mu + \alpha_1 Start_t + \alpha_2 During_t + \alpha_3 End_t + \varepsilon_t,\ \varepsilon_t \sim N(0,\sigma_t^2),\ \sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2 + \beta_3 Start_t + \beta_4 During_t + \beta_5 End_t$. $Start_t$ and End_t denote the number of crises that start and end in month t, respectively, and $During_t$ is the total number of crises that occurs in month t, excluding all crises that start or end in that month. r_t^{World} is the return on the world market in month t (in percent). We report results for all crises and subsets based on different measures of crisis severity, as defined in Table 1. t-Statistics in bold denote significance at least at the 10% level.

	Mean				Volatility					
	Const.	*Start*	*During*	*End*	Const.	β_1	β_2	*Start*	*During*	*End*
All crises										
Coeff. (%)	0.856	−0.376	−0.116	0.235	0.149	0.076	0.892	1.312	−0.013	−0.603
t-Statistic	**4.28**	**−2.12**	−1.25	1.38	0.82	**3.74**	**31.34**	**1.82**	−0.12	−0.84
Violent crises										
Coeff. (%)	0.794	−0.867	−0.147	0.650	0.406	0.075	0.889	2.050	−0.020	−1.577
t-Statistic	**4.92**	**−2.63**	−1.19	**2.75**	**1.76**	**3.62**	**27.63**	**1.74**	−0.15	−1.40
War										
Coeff. (%)	0.797	−1.485	−0.374	1.135	0.248	0.062	0.913	4.999	−0.062	−3.037
t-Statistic	**5.75**	**−3.15**	**−2.14**	**4.08**	**1.72**	**3.31**	**41.48**	**3.61**	−0.47	**−2.97**
Violent break										
Coeff. (%)	0.743	−0.738	−0.140	0.658	0.120	0.080	0.881	0.793	0.095	0.906
t-Statistic	**4.97**	**−2.66**	−0.88	**2.24**	0.81	**3.64**	**28.39**	0.73	0.44	0.81
Grave threat										
Coeff. (%)	1.034	−0.503	−0.524	0.090	0.122	0.066	0.886	1.744	0.325	−0.266
t-Statistic	**6.62**	−1.62	**−3.15**	0.35	**0.76**	**2.93**	**27.62**	1.13	1.23	−0.19
Major power										
Coeff. (%)	0.658	−0.819	−0.252	0.410	0.199	0.078	0.902	4.376	0.040	−3.487
t-Statistic	**4.47**	**−1.78**	−1.15	1.01	1.42	**4.03**	**39.66**	**3.12**	0.16	**−2.48**
Protracted										
Coeff. (%)	0.720	−0.654	−0.088	0.526	0.069	0.076	0.898	1.880	−0.039	−0.554
t-Statistic	**4.41**	**−2.33**	−0.77	**2.47**	**0.58**	**3.97**	**36.36**	**2.55**	−0.33	−0.73
Crisis severity index										
Coeff. (%)	0.846	−0.156	−0.040	0.130	0.097	0.062	0.912	0.524	−0.001	−0.344
t-Statistic	**4.55**	**−2.83**	−1.61	**2.87**	0.72	**3.48**	**39.44**	**2.76**	−0.03	**−1.85**

4.4. Individual country returns

In this section, we examine the stock market reaction of individual countries to international crises in which these countries were actors. For this analysis, we use individual country trigger and end dates and estimate a model analogous to model (3), but now the dependent variable is the stock market return in country i in month t, and $Start_{i,t}$, $During_{i,t}$, and $End_{i,t}$ refer to the number of international crises in month t in which country i was a crisis actor.

A word of caution applies when interpreting the individual country crisis results below. First, as a result of market closures, there are several months with missing observations for some European countries around the beginning of World War II, and some monthly returns are missing for the German and the Japanese stock market at the end of World War II. Our estimates for *Start* and *During* might therefore be too high. Second, financial markets sometimes anticipate crises. Austria offers a clear example. According to the ICB database, the *Anschluss* crisis between Germany and Austria began on February 12, 1938, when Hitler presented Austrian chancellor Kurt Schuschnigg with an ultimatum. The Austrian stock market return in February, our *Start* month, was only marginally negative with −0.5%. Earlier however, in January 1938, the Austrian market had dropped 39%, its largest drop in history (see Table 4). This stock market decline could be caused by an unrelated event, but according to the *Wiener Zeitung*, investors had already anticipated the crisis: "Mid-January 1938, it was clear to insiders that Vienna was willing, or forced, to give in to threats and wishes from Berlin."[38]

The results for the individual countries are seen in Table 15. The second to last row in Table 15 reports the average coefficient across the 19 countries in our sample. The significance test at the bottom of the table is based on a heteroskedasticity-adjusted Wald test that the coefficients are jointly significantly different from zero.[39]

[38] "Wie sehr man in Wien bereit oder gezwungen war, den Drohungen und Wünschen aus Berlin nachzugeben, war Insidern schon Mitte Jänner 1938 klar." http://www.wienerzeitung.at/linkmap/personen/miklaspopup.htm. According to the same newspaper article, Schuschnigg was already under considerable pressure from Germany before the February meeting. This may be seen in the demand to remove the chief of staff of the Austrian army, Alfred Jansa, from his position in January 1938. Jansa and his staff had developed a scenario for Austria's defense against a German attack, a situation Hitler wanted to avoid at all costs. Schuschnigg complied with the demand.

[39] We could include the world market index as a control variable, or use a system of seemingly unrelated regressions to adjust for correlation between stock market returns. However, doing so might reduce the effect we are trying to measure if a crisis in one country has a substantial impact on stock prices in other countries. We therefore use a system of ordinary least squares regressions.

Table 13
Additional *Start* month and additional *End* month.
Estimation results for the following model:

$$r_t^{World} = \mu + \alpha_1 Start_{t+1} + \alpha_2 Start_t + \alpha_3 During_t + \alpha_4 End_t + \alpha_5 End_{t-1} + \varepsilon_t,$$

where $Start_t$ and End_t denote the number of crises that start and end in month *t*, respectively, and $During_t$ is total number of crises that occurs in month *t*, excluding all crises that start or end in that month. $Start_{t+1}$ equals $Start_t$ with a one-month lead, and End_{t-1} equals End_t with a one-month lag. r_t^{World} is the return on the world market in month *t* (in percent). We report results for all crises and subsets based on different measures of crisis severity, as defined in Table 1. *t*-Statistics are based on heteroskedasticity-consistent standard errors. *t*-Statistics in bold denote significance at least at the 10% level.

	Constant	*Start*(*t*+1)	*Start*	*During*	*End*	*End*(*t*−1)
All						
Coefficient	0.552	0.251	−0.465	−0.159	0.261	0.155
t-Statistic	**2.81**	1.44	**−2.89**	**−1.97**	1.53	0.91
Grave threat						
Coefficient	0.707	0.072	−0.638	−0.395	0.332	−0.018
t-Statistic	**4.54**	0.23	**−2.24**	**−2.80**	1.24	−0.07
Major power						
Coefficient	0.686	−0.046	−0.696	−0.447	0.166	−0.03
t-Statistic	**4.75**	−0.17	**−2.09**	**−3.59**	0.56	−0.12
Violent						
Coefficient	0.511	0.152	−0.779	−0.137	0.417	0.207
t-Statistic	**3.34**	0.52	**−2.71**	−1.37	1.57	0.87
War						
Coefficient	0.491	−0.214	−1.315	−0.084	0.324	0.487
t-Statistic	**3.13**	−0.56	**−3.42**	−0.56	0.89	1.33
Violent break						
Coefficient	0.513	0.071	−0.689	−0.158	0.847	0.469
t-Statistic	**2.89**	0.27	**−2.42**	**−1.99**	**2.98**	**1.77**
Protracted						
Coefficient	0.44	0.217	−0.624	−0.202	0.626	0.237
t-Statistic	**2.59**	1.02	**−2.71**	**−1.81**	**3.01**	1.24
Crisis severity index						
Coefficient	0.622	0.033	−0.16	−0.045	0.098	0.045
t-Statistic	**3.40**	0.70	**−3.25**	**−2.15**	**2.11**	1.01

In line with expectations, crises have a strong impact on stock returns if a country is a crisis actor. Similar to the world market, individual countries exhibit a strong negative stock market reaction in the month of the outbreak of an international crisis. The average coefficient of −2.0% is highly significant. The individual country coefficients range from a high of 0.9% for Japan, to a low of −6.0% for South Africa. Investors in individual countries clearly suffer when their country becomes involved in a crisis. Out of all 19 countries, only three have a positive coefficient. On average, an ongoing crisis in an actor country has a significant negative effect of −1.7% a month. The estimated coefficient is negative for 15 countries and positive for four countries. In the month a crisis ends, the stock market in countries that are involved in the crisis goes up, on average, 0.5% (insignificant). In unreported tests, we also include $Start_{t+1}$ and End_{t-1} similar to the model in Table 13. For this extended model, the average coefficients for $Start_{t+1}$ and $Start_t$ are −1.8% and −2.1%, the average coefficient for *During* is −1.8%, and the average coefficients for End_t and End_{t-1} are 0.5% and 0.3%. All these coefficients are significant at the 5% level, suggesting that anticipation of crises and skepticism about the resolution of crises are relevant factors to consider at the level of individual countries.

5. Conclusions

This study uses a large sample of major international political crises to directly test the link between changes in disaster risk and changes in stock market prices. Consistent with time-varying disaster probability models, we show that changes in the number of international crises have a large impact on world stock market returns—mean and volatility. In particular, crises starts – all of which had the potential to develop into a political disaster – have an economically large, negative impact on stock returns. We also document that changes in the number of international crises have a significant impact on world market volatility: The start of an international crisis increases monthly volatility of world market returns, and for each crisis that comes to an end, monthly volatility decreases. Moreover, we find that markets react more strongly (in terms of average return and volatility) when crises are more severe and when there is more at stake due to the involvement of major powers. Predictive regressions do not show a significant relation between crisis risk and future market returns. However, consistent with time-varying disaster models, we do find that crisis risk is positively correlated with the earnings–price ratio and the dividend yield. Cross-sectional tests provide further

Table 14
Alternative world stock market indices: equally weighted and GDP-weighted.

This table presents the results of model (3), using two alternative measures for the monthly world market return as dependent variable. The first alternative world market index gives equal weight to all 19 countries in our sample. The second alternative world market index is intended to reflect the shifting economic importance of the constituent countries in the sample period. It is based on a weighted world market return, in which country weights are determined annually based on each country's GDP. We use the annual GDP per capita data from Barro and Ursúa (http://www.economics.harvard.edu/faculty/barro/-data_sets_barro) and population and GDP (2006) data from Angus Maddison (http://www.ggdc.net/maddison/).

Crises	Equally weighted index				GDP-weighted index			
	Constant	*Start*	*During*	*End*	Constant	*Start*	*During*	*End*
All								
Coefficient (%)	0.914	−0.290	−0.126	0.117	0.608	−0.455	−0.036	0.261
t-Statistic	**6.34**	**−2.46**	**−2.11**	0.96	**3.49**	**−2.93**	−0.50	**1.64**
Violent break								
Coefficient (%)	0.689	−0.506	−0.094	0.513	0.413	−0.633	0.060	0.749
t-Statistic	**6.03**	**−2.22**	−0.83	**2.53**	**2.98**	**−2.25**	0.40	**2.95**
Violent crises								
Coefficient (%)	0.813	−0.509	−0.120	0.196	0.755	−0.568	−0.173	0.393
t-Statistic	**7.09**	**−2.38**	**−1.71**	1.12	**6.43**	**−2.44**	**−1.97**	**2.07**
War								
Coefficient (%)	0.774	−0.831	−0.138	0.174	0.501	−1.074	0.026	0.541
t-Statistic	**6.88**	**−3.20**	−1.16	0.70	**3.72**	**−3.03**	0.18	**1.68**
Grave threat								
Coefficient (%)	0.904	−0.372	−0.327	0.166	0.682	−0.617	−0.204	0.252
t-Statistic	**8.02**	**−1.74**	**−3.30**	0.87	**4.92**	**−2.20**	−1.63	1.02
Major power								
Coefficient (%)	0.795	−0.633	−0.145	−0.025	0.617	−0.660	−0.174	0.229
t-Statistic	**7.60**	**−2.64**	**−1.71**	−0.11	**4.75**	**−2.18**	−1.63	0.81
Protracted								
Coefficient (%)	0.716	−0.324	−0.087	0.302	0.483	−0.609	−0.012	0.598
t-Statistic	**5.77**	**−2.00**	−1.13	**2.08**	**3.23**	**−2.76**	−0.12	**3.07**
Crisis severity index								
Coefficient (%)	0.913	−0.104	−0.034	0.048	0.606	−0.149	−0.012	0.101
t-Statistic	**6.74**	**−2.96**	**−2.23**	1.55	**3.75**	**−3.15**	−0.61	**2.44**

Table 15
Stock market returns for crisis actors: Individual country effects.

Estimation results for the following model: $r_{i,t,}=\mu+\alpha_1 Start_{i,t}+\alpha_2 During_{i,t}+\alpha_3 End_{i,t}+\varepsilon_{i,t}$, where $r_{i,t}$ is the stock market return in country *i* in month *t*, and $Start_{i,t}$, $During_{i,t}$, and $End_{i,t}$ refer to the number of international crises in month *t* in which country *i* was a crisis actor. *t*-Statistics are based on heteroskedasticity-consistent standard errors. *t*-Statistics in bold denote significance at least at the 10% level.

Country	Constant		Country Start		Country During		Country End	
	Coefficient	*t*-Value	Coefficient	*t*-Value	Coefficient	*t*-Value	Coefficient	*t*-Value
Australia	0.558	**4.27**	−3.738	**−3.72**	−3.618	**−3.20**	−0.621	−0.87
Austria	0.841	**3.25**	−0.409	−0.51	−2.217	**−2.75**		
Belgium	0.528	**3.19**	−1.566	−0.81	−2.016	−0.95	1.386	0.71
Canada	0.449	**3.08**	−4.444	**−4.38**	−0.450	−0.37	2.815	**2.37**
Colombia	0.875	**4.71**	−2.627	**−6.09**	−0.996	−0.73	−7.272	−1.43
Finland	0.839	**4.57**	−5.377	**−2.67**	−10.931	**−2.60**	−1.753	−0.89
France	0.734	**3.70**	0.070	0.06	−0.227	−0.37	−0.612	−0.59
Germany	0.109	0.39	−0.976	−0.33	1.631	**1.74**	5.401	1.36
India	0.543	**3.07**	−1.783	−1.21	−0.972	−1.36	0.811	0.64
Italy	0.571	**2.43**	−0.670	−0.48	0.532	0.41	2.150	0.82
Japan	0.567	**2.80**	0.903	0.55	−0.102	−0.12	−0.399	−0.48
Netherlands	0.391	**2.52**	0.404	0.70	−1.476	**−1.76**	1.363	**1.84**
Peru	1.084	0.65	−2.781	−0.41	−4.105	**−1.84**	4.716	0.98
Portugal	0.798	**3.19**	−2.172	−1.57	−3.471	−1.64	−1.398	−0.82
South Africa	0.688	**4.48**	−5.965	−1.25	−0.976	−0.59	0.325	0.08
Spain	0.548	**3.20**	−3.981	**−1.78**	−0.665	−0.42	1.573	0.54
Sweden	0.551	**3.58**	−0.793	−0.35	−3.747	**−2.87**	1.693	0.55
United Kingdom	0.496	**3.23**	−1.048	−1.36	0.832	**1.93**	−1.691	**−1.81**
United States	0.501	**2.56**	−0.961	**−1.74**	0.275	0.73	0.268	0.49
Average	0.614	**0.00%**	−1.995	**0.00%**	−1.721	**0.00%**	0.488	21.71%
	$\chi^2(19)$	205.69	$\chi^2(19)$	93.19	$\chi^2(19)$	53.43	$\chi^2(18)$	22.34

support for the hypothesis that crisis risk is priced: U.S. industries with higher crisis risk sensitivity yield higher returns, on average. Finally, using U.S. survey data, we also provide support for an important assumption underlying time-varying disaster models, namely, that expected consumption growth is negatively related to disaster probability.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.jfineco.2011.02.019.

References

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.
Amihud, Y., Wohl, A., 2004. Political news and stock prices: the case of the Saddam Hussein contracts. Journal of Banking and Finance 28, 1185–1200.
Bansal, R., Shaliastovich, I., 2010. Confidence risk and asset prices. Unpublished Working Paper, Duke University.
Bansal, R., Yaron, A., 2004. Risks for the long run: a potential resolution of asset pricing puzzles. Journal of Finance 59, 1481–1509.
Barro, R., 2006. Rare disasters and asset markets in the twentieth century. Quarterly Journal of Economics 121, 823–865.
Barro, R., Ursúa, J., 2008. Consumption disasters in the twentieth century. American Economic Review 98, 58–63.
Barro, R., Ursúa, J., 2009. Stock market crashes and depressions. Unpublished Working Paper, NBER Working Paper No. 14760.
Bittlingmayer, G., 1998. Output, stock volatility, and political uncertainty in a natural experiment: Germany, 1880–1940. Journal of Finance 53, 2243–2257.
Blomberg, S., Hess, G., Orphanides, A., 2004. The macroeconomic consequences of eerrorism. Journal of Monetary Economics 51, 1007–1032.
Brecher, M., Wilkenfeld, J., 1997. A Study of Crisis. University of Michigan Press, Ann Arbor.
Brown, W., Burdekin, R., Weidenmier, M., 2006. Volatility in an era of reduced uncertainty: lessons from Pax Brittanica. Journal of Financial Economics 79, 693–707.
Buraschi, A., Jitsov, A., 2006. Model uncertainty and option markets with heterogeneous agents. Journal of Finance 61, 2841–2898.
Copeland, L., Zhu, Y., 2007. Rare disasters and the equity premium in a two-country world. Unpublished Working Paper, Cardiff University.
Elton, E., 1999. Expected return, realized return, and asset pricing tests. Journal of Finance 54, 1199–1220.
Engle, R., 2002. New frontiers for ARCH models. Journal of Applied Econometrics 17, 425–446.
Engle, R., Patton, A., 2001. What good is a volatility model? Quantitative Finance 1, 245–273.
Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 71, 607–636.
Fama, E., French, K., 1997. Industry costs of equity. Journal of Financial Economics 43, 153–193.
Fama, E., French, K., 2002. The equity premium. Journal of Finance 62, 637–659.
French, K., Schwert, G., Stambaugh, R., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.
Frey, B., Kucher, M., 2000. History as reflected in capital markets: the case of World War II. Journal of Economic History 60, 468–496.
Gabaix, X., 2009. Variable rare disasters: an exactly solved framework for ten puzzles in macro-finance. Unpublished Working Paper, New York University.
Gourio, F., 2008a. Time-series predictability in the disaster model. Finance Research Letters 5, 191–203.
Gourio, F., 2008b. Disasters, recoveries, and predictability. Unpublished Working Paper, Boston University.
Hess, G., Orphanides, A., 1995. War politics: an economic, rational-voter framework. The American Economic Review 85, 828–846.
Jorion, P., Goetzmann, W., 1999. Global stock markets in the twentieth century. The Journal of Finance 54, 953–980.
La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R., 1998. Law and finance. Journal of Political Economy 106, 1113–1155.
Ludvigson, S., 2004. Consumer confidence and consumer spending. Journal of Economic Perspectives 18, 29–50.
Mehra, R., Prescott, E., 1985. The equity premium: a puzzle? Journal of Monetary Economics 15, 145–161.
Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. Journal of Financial Economics 78, 277–309.
Lewellen, J., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset pricing tests. Journal of Financial Economics 96, 175–194.
Obstfeld, M., Taylor, A., 2003. Globalization and capital markets. In: Bordo, M., Taylor, A., Williamson, J. (Eds.), Globalization in Historical Perspective. University of Chicago Press, Chicago.
O'Rourke, K., Williamson, J., 2002. When did globalisation begin? European Review of Economic History 6, 23–50.
Rietz, T., 1988. The equity risk premium: a solution. Journal of Monetary Economics 22, 117–131.
Rigobon, R., Sack, B., 2005. The effects of war risk on U.S. financial markets. Journal of Banking and Finance 29, 1769–1789.
Schwert, W., 1989. Why does stock market volatility change over time? Journal of Finance 44, 1115–1153.
Shiller, R., 1981. Do stock prices move too much to be justified by subsequent changes in dividends? American Economic Review 71, 421–436.
Veronesi, P., 2004. The Peso problem hypothesis and stock market returns. Journal of Economic Dynamics & Control 28, 707–725.
Voth, H., 2002. Why was stock market volatility so high during the Great Depression? Evidence from 10 countries during the interwar period. Unpublished Working Paper, Massachusetts Institute of Technology.
Wachter, J., 2009. Can time varying risk of rare disasters explain aggregate stock market volatility? Unpublished Working Paper, NBER Working Paper No. 14386.
Waldenström, D., Frey, B., 2002. How government bond prices reflect wartime events: the case of the Stockholm market. Unpublished Working Paper, University of Zurich.
Wolfers, J., Zitzewitz, E., 2009. Using markets to inform policy: the case of the Iraq War. Economica 76, 225–250.

THE JOURNAL OF FINANCE • VOL. LXI, NO. 4 • AUGUST 2006

Investor Sentiment and the Cross-Section of Stock Returns

MALCOLM BAKER and JEFFREY WURGLER*

ABSTRACT

We study how investor sentiment affects the cross-section of stock returns. We predict that a wave of investor sentiment has larger effects on securities whose valuations are highly subjective and difficult to arbitrage. Consistent with this prediction, we find that when beginning-of-period proxies for sentiment are low, subsequent returns are relatively high for small stocks, young stocks, high volatility stocks, unprofitable stocks, non-dividend-paying stocks, extreme growth stocks, and distressed stocks. When sentiment is high, on the other hand, these categories of stock earn relatively low subsequent returns.

CLASSICAL FINANCE THEORY LEAVES NO ROLE FOR INVESTOR SENTIMENT. Rather, this theory argues that competition among rational investors, who diversify to optimize the statistical properties of their portfolios, will lead to an equilibrium in which prices equal the rationally discounted value of expected cash flows, and in which the cross-section of expected returns depends only on the cross-section of systematic risks.[1] Even if some investors are irrational, classical theory argues, their demands are offset by arbitrageurs and thus have no significant impact on prices.

In this paper, we present evidence that investor sentiment may have significant effects on the cross-section of stock prices. We start with simple theoretical predictions. Because a mispricing is the result of an uninformed demand shock in the presence of a binding arbitrage constraint, we predict that a broad-based wave of sentiment has cross-sectional effects (that is, does not simply raise or lower all prices equally) when sentiment-based demands *or* arbitrage

*Baker is at the Harvard Business School and National Bureau of Economic Research; Wurgler is at the NYU Stern School of Business and the National Bureau of Economic Research. We thank an anonymous referee, Rob Stambaugh (the editor), Ned Elton, Wayne Ferson, Xavier Gabaix, Marty Gruber, Lisa Kramer, Owen Lamont, Martin Lettau, Anthony Lynch, Jay Shanken, Meir Statman, Sheridan Titman, and Jeremy Stein for helpful comments, as well as participants of conferences or seminars at Baruch College, Boston College, the Chicago Quantitative Alliance, Emory University, the Federal Reserve Bank of New York, Harvard University, Indiana University, Michigan State University, NBER, the Norwegian School of Economics and Business, Norwegian School of Management, New York University, Stockholm School of Economics, Tulane University, the University of Amsterdam, the University of British Columbia, the University of Illinois, the University of Kentucky, the University of Michigan, the University of Notre Dame, the University of Texas, and the University of Wisconsin. We gratefully acknowledge financial support from the Q Group and the Division of Research of the Harvard Business School.

[1] See Gomes, Kogan, and Zhang (2003) for a recent model in this tradition.

constraints vary across stocks. In practice, these two distinct channels lead to quite similar predictions because stocks that are likely to be most sensitive to speculative demand, those with highly subjective valuations, also tend to be the riskiest and costliest to arbitrage. Concretely, then, theory suggests two distinct channels through which the shares of certain firms—newer, smaller, more volatile, unprofitable, non-dividend paying, distressed or with extreme growth potential, and firms with analogous characteristics—are likely to be more affected by shifts in investor sentiment.

To investigate this prediction empirically, and to get a more tangible sense of the intrinsically elusive concept of investor sentiment, we start with a summary of the rises and falls in U.S. market sentiment from 1961 through the Internet bubble. This summary is based on anecdotal accounts and thus by its nature can only be a suggestive, ex post characterization of fluctuations in sentiment. Nonetheless, its basic message appears broadly consistent with our theoretical predictions and suggests that more rigorous tests are warranted.

Our main empirical approach is as follows. Because cross-sectional patterns of sentiment-driven mispricing would be difficult to identify directly, we examine whether cross-sectional predictability patterns in stock returns depend upon proxies for beginning-of-period sentiment. For example, low future returns on young firms relative to old firms, conditional on high values for proxies for beginning-of-period sentiment, would be consistent with the ex ante relative overvaluation of young firms. As usual, we are mindful of the joint hypothesis problem that any predictability patterns we find actually reflect compensation for systematic risks.

The first step is to gather proxies for investor sentiment that we can use as time-series conditioning variables. Since there are no perfect and/or uncontroversial proxies for investor sentiment, our approach is necessarily practical. Specifically, we consider a number of proxies suggested in recent work and form a composite sentiment index based on their first principal component. To reduce the likelihood that these proxies are connected to systematic risk, we also form an index based on sentiment proxies that have been orthogonalized to several macroeconomic conditions. The sentiment indexes visibly line up with historical accounts of bubbles and crashes.

We then test how the cross-section of subsequent stock returns varies with beginning-of-period sentiment. Using monthly stock returns between 1963 and 2001, we start by forming equal-weighted decile portfolios based on several firm characteristics. (Our theory predicts, and the empirical results confirm, that large firms will be less affected by sentiment, and hence value weighting will tend to obscure the relevant patterns.) We then look for patterns in the average returns across deciles conditional upon the beginning-of-period level of sentiment. We find that when sentiment is low (below sample average), small stocks earn particularly high subsequent returns, but when sentiment is high (above average), there is no size effect at all. Conditional patterns are even sharper when we sort on other firm characteristics. When sentiment is low, subsequent returns are higher on very young (newly listed) stocks than older stocks, high-return volatility than low-return volatility stocks, unprofitable stocks than profitable ones, and nonpayers than dividend payers. When sentiment is high, these patterns completely reverse. In other words, several characteristics that do not have any unconditional predictive power actually display sign-flipping predictive ability, in the hypothesized directions, once one conditions on sentiment. These are our most striking findings. Although earlier data are not as rich, some of these patterns are also apparent in a sample that covers 1935 through 1961.

The sorts also suggest that sentiment affects extreme growth and distressed firms in similar ways. Note that when stocks are sorted into deciles by sales growth, book-to-market, or external financing activity, growth and distress firms tend to lie at opposing extremes, with more "stable" firms in the middle deciles. We find that when sentiment is low, the subsequent returns on stocks at both extremes are especially high relative to their unconditional average, while stocks in the middle deciles are less affected by sentiment. (The result is not statistically significant for book-to-market, however.) This U-shaped pattern in the conditional difference is also broadly consistent with theoretical predictions: both extreme growth and distressed firms have relatively subjective valuations and are relatively hard to arbitrage, and so they should be expected to be most affected by sentiment. Again, note that this intriguing conditional pattern would be averaged away in an unconditional study.

We then consider a regression approach, which allows us to control for comovement in size and book-to-market-sorted stocks using the Fama-French (1993) factors. We use the sentiment indexes to forecast the returns of various high-minus-low portfolios (in terms of sensitivity to sentiment). Not surprisingly, given that our decile portfolios are equal-weighted and several of the characteristics we examine are correlated with size, the inclusion of *SMB* as a control tends to reduce the magnitude of the predictability, although some predictive power generally remains.

We then turn to the classical alternative explanation, namely, that they simply reflect a complex pattern of compensation for systematic risk. This explanation would account for the predictability evidence by either time variation in rational, market-wide risk premia or time variation in the cross-sectional pattern of risk, that is, beta loadings. Further tests cast doubt on these hypotheses. We test the second possibility directly and find no link between the patterns in predictability and patterns in betas with market returns or consumption growth. If risk is not changing over time, then the first possibility requires not just time variation in risk premia, but also changes in sign. Put simply, it would require that in half of our sample period (when sentiment is relatively low), older, less volatile, profitable, and/or dividend-paying firms actually require a risk premium over very young, highly volatile, unprofitable, and/or nonpayers. This is counterintuitive. Other aspects of the results also suggest that systematic risk is not a complete explanation.

The results challenge the classical view of the cross-section of stock prices and, in doing so, build on several recent themes. First, the results complement earlier work that shows sentiment helps to explain the time series of returns (Kothari and Shanken (1997), Neal and Wheatley (1998), Shiller (1981, 2000), Baker and Wurgler (2000)). Campbell and Cochrane (2000), Wachter (2000), Lettau and Ludvigson (2001), and Menzly, Santos, and Veronesi (2004) examine the effects of conditional systematic risks; here we condition on investor sentiment. Daniel and Titman (1997) test a characteristics-based model for the cross-section of expected returns; we extend their specification into a *conditional* characteristics-based model. Shleifer (2000) surveys early work on sentiment and limited arbitrage, two key ingredients here. Barberis and Shleifer (2003), Barberis, Shleifer, and Wurgler (2005), and Peng and Xiong (2004) discuss category-level trading, and Fama and French (1993) document comovement of stocks of similar sizes and book-to-market ratios; uninformed demand shocks for categories of stocks with similar characteristics are central to our results. Finally, we extend and unify known relationships among sentiment, IPOs, and small stock returns (Lee, Shleifer, and Thaler (1991), Swaminathan (1996), Neal and Wheatley (1998)).

Section I discusses theoretical predictions. Section II provides a qualitative history of recent speculative episodes. Section III describes our empirical hypotheses and data, and Section IV presents the main empirical tests. Section V concludes.

I. Theoretical Effects of Sentiment on the Cross-Section

A mispricing is the result of both an uninformed demand shock and a limit on arbitrage. One can therefore think of two distinct channels through which investor sentiment, as defined more precisely below, might affect the cross-section of stock prices. In the first channel, sentimental demand shocks vary in the cross-section, while arbitrage limits are constant. In the second, the difficulty of arbitrage varies across stocks but sentiment is generic. We discuss these in turn.

A. Cross-Sectional Variation in Sentiment

One possible definition of investor sentiment is the propensity to speculate.[2] Under this definition, sentiment drives the relative demand for speculative investments, and therefore causes cross-sectional effects even if arbitrage forces are the same across stocks.

What makes some stocks more vulnerable to broad shifts in the propensity to speculate? We suggest that the main factor is the subjectivity of their valuations. For instance, consider a canonical young, unprofitable, extreme growth stock. The lack of an earnings history combined with the presence of apparently unlimited growth opportunities allows unsophisticated investors to defend, with equal plausibility, a wide spectrum of valuations, from much too low to much too high, as suits their sentiment. During a bubble period, when the propensity to speculate is high, this profile of characteristics also allows investment bankers (or swindlers) to further argue for the high end of valuations. By contrast, the value of a firm with a long earnings history, tangible assets, and

[2] Aghion and Stein (2004) develop a model with both rational expectations and bounded rationality in which investors periodically emphasize growth over profitability. While the emphasis is on the corporate and macroeconomic effects, the bounded-rationality version of the model offers some similar predictions for the cross-section of returns.

stable dividends is much less subjective, and thus its stock is likely to be less affected by fluctuations in the propensity to speculate.[3]

While the above channel suggests how variation in the propensity to speculate may generally affect the cross-section, it does not take a stand on how sentimental investors actually choose stocks. We suggest that they simply demand stocks that have the bundle of salient characteristics that is compatible with their sentiment.[4] That is, investors with a low propensity to speculate may demand profitable, dividend-paying stocks not because profitability and dividends are correlated with some unobservable firm property that defines safety to the investor, but precisely because the salient characteristics "profitability" and "dividends" are essentially taken to define safety.[5] Likewise, the salient characteristics "no earnings," "young age," and "no dividends" mark the stock as speculative. Casual observation suggests that such an investment process may be a more accurate description of how typical investors pick stocks than the process outlined by Markowitz (1959), in which investors view individual securities purely in terms of their statistical properties.

B. Cross-Sectional Variation in Arbitrage

One might also define investor sentiment as optimism or pessimism about stocks in general. Indiscriminate waves of sentiment still affect the cross-section, however, if arbitrage forces are relatively weaker in a subset of stocks.

This channel is better understood than the cross-sectional variation in sentiment channel. A body of theoretical and empirical research shows that arbitrage tends to be particularly risky and costly for young, small, unprofitable, extreme growth, or distressed stocks. First, their high idiosyncratic risk makes relative-value arbitrage especially risky (Wurgler and Zhuravskaya (2002)). Moreover, such stocks tend to be more costly to trade (Amihud and Mendelsohn (1986)) and particularly expensive, sometimes impossible, to sell short (D'Avolio (2002), Geczy, Musto, and Reed (2002), Jones and Lamont (2002), Duffie, Garleanu, and

[3] The favorite-longshot bias in racetrack betting is a static illustration of the notion that investors with a high propensity to speculate (racetrack bettors) have a relatively high demand for the most speculative bets (longshots have the most negative expected returns; see Hausch and Ziemba (1995)).

[4] The idea that investors view securities as a vector of salient characteristics borrows from Lancaster (1966, 1971), who views consumer demand theory from the perspective that the utility of a consumer good (e.g, oranges) derives from more primitive characteristics (fiber and vitamin C).

[5] The implications of categorization for finance are explored by Baker and Wurgler (2003), Barberis and Shleifer (2003), Barberis, Shleifer, and Wurgler (2005), Greenwood and Sosner (2003), and Peng and Xiong (2004). Note that if investors infer category membership from salient characteristics (some psychologists propose that category membership is determined by the presence of defining or characteristic features, see, for example, Smith, Shoben, and Rips (1974)), then sentiment-driven demand will be directly connected to characteristics even if sentimental investors undertake an intervening process of categorization and trade entirely at the category level. It is also empirically convenient to boil key investment categories down into vectors of stable and measurable characteristics: One can use the same empirical framework to study episodes such as the late 1960s growth stocks bubble and the Internet bubble. In other words, the term "Internet bubble" is interesting, but it does not make for a useful or testable theory. The key is to examine the recurring underlying characteristics.

Pedersen (2002), Lamont and Thaler (2003), Mitchell, Pulvino, and Stafford (2002)). Further, their lower liquidity also exposes would-be arbitrageurs to predatory attacks (Brunnermeier and Pedersen (2005)).

The key point of this discussion is that, in practice, *the same stocks that are the hardest to arbitrage also tend to be the most difficult to value*. While for expositional purposes we have outlined the two channels separately, they are likely to have overlapping effects. This may make them difficult to distinguish empirically; however, it only strengthens our predictions about what region of the cross-section is most affected by sentiment. Indeed, the two channels can reinforce each other. For example, the fact that investors can convince themselves of a wide range of valuations in some regions of the cross-section generates a noise-trader risk that further deters short-horizon arbitrageurs (De Long et al. (1990), Shleifer and Vishny (1997)).[6]

II. An Anecdotal History of Investor Sentiment, 1961–2002

Here we briefly summarize the most prominent U.S. stock market bubbles between 1961 and 2002 (matching the period of our main data). The reader eager to see results may skip this section, but it is useful for three reasons. First, despite great interest in the effects of investor sentiment, the academic literature does not contain even the most basic ex post characterization of most of the recent speculative episodes. Second, a knowledge of the rough timing of these episodes allows us to make a preliminary judgment about the accuracy of the quantitative proxies for sentiment that we develop later. Third, the discussion sheds some initial, albeit anecdotal, light on the plausibility of our theoretical predictions.

We distill our brief history of sentiment from several sources. Kindleberger (2001) draws general lessons from bubbles and crashes over the past few hundred years, while Brown (1991), Dreman (1979), Graham (1973), Malkiel (1990, 1999), Shiller (2000), and Siegel (1998) focus more specifically on recent U.S. stock market episodes. We take each of these accounts with a grain of salt, and emphasize only those themes that appear repeatedly.

We start in 1961, a year that Graham (1973), Malkiel (1990) and Brown (1991) note as characterized by a high demand for small, young, growth stocks; Dreman (1979, p. 70) confirms their accounts. For instance, Malkiel writes of a "new-issue mania" that was concentrated on new "tronics" firms. "... The tronics boom came back to earth in 1962. The tailspin started early in the year and exploded in a horrendous selling wave ... Growth stocks took the brunt of the decline, falling much further than the general market" (p. 54–57).

The next major bubble developed in 1967 and 1968. Brown writes that "scores of franchisers, computer firms, and mobile home manufactures seemed

[6] We do not incorporate the equilibrium prediction of DeLong et al. (1990), namely that securities with more exposure to sentiment have higher unconditional expected returns. Elton, Gruber, and Busse (1998) argue that expected returns are not higher on stocks that have higher sensitivities to the closed-end fund discount. However, Brown et al. (2003) argue that exposure to a sentiment factor constructed from daily mutual fund flows is a priced factor in the United States and Japan.

to promise overnight wealth . . . . [while] quality was pretty much forgotten" (p. 90). Malkiel and Dreman also note this pattern of a focus on firms with strong earnings growth or potential and an avoidance of "the major industrial giants, 'buggywhip companies,' as they were sometimes contemptuously called" (Dreman 1979, p. 74–75). Another characteristic apparently out of favor was dividends. According to the *New York Times*, "during the speculative market of the late 1960s many brokers told customers that it didn't matter whether a company paid a dividend—just so long as its stock kept going up" (9/13/1976). But "after 1968, as it became clear that capital losses were possible, investors came to value dividends" (10/7/1999). In summarizing the performance of stocks from the end of 1968 through August 1971, Graham (1973) writes: "[our] comparative results undoubtedly reflect the tendency of smaller issues of inferior quality to be relatively overvalued in bull markets, and not only to suffer more serious declines than the stronger issues in the ensuing price collapse, but also to delay their full recovery—in many cases indefinitely" (p. 212).

Anecdotal accounts invariably describe the early 1970s as a bear market, with sentiment at a low level. However, a set of established, large, stable, consistently profitable stocks known as the "nifty fifty" enjoyed notably high valuations. Brown (1991), Malkiel (1990), and Siegel (1998) each highlight this episode. Siegel writes, "All of these stocks had proven growth records, continual increases in dividends . . . and high market capitalization" (p. 106). Note that this speculative episode is a mirror image of those described above (and below). That is, the bubbles associated with high sentiment periods centered on small, young, unprofitable growth stocks, whereas the nifty fifty episode appears to be a bubble in a set of firms with an opposite set of characteristics (old, large, and continuous earnings and dividend growth) that happened in a period of *low* sentiment.

The late 1970s through mid 1980s are described as a period of generally high sentiment, perhaps associated with Reagan-era optimism. This period witnessed a series of speculative episodes. Dreman describes a bubble in gambling issues in 1977 and 1978. Ritter (1984) studies the hot-issue market of 1980, and finds greater initial returns on IPOs of natural resource start-ups than on large, mature, profitable offerings. Of 1983, Malkiel (p. 74–75) writes that "the high-technology new-issue boom of the first half of 1983 was an almost perfect replica of the 1960's episodes . . . The bubble appears to have burst early in the second half of 1983 . . . the carnage in the small company and new-issue markets was truly catastrophic." Brown confirms this account. Of the mid 1980s, Malkiel writes that "What electronics was to the 1960s, biotechnology became to the 1980s . . . . new issues of biotech companies were eagerly gobbled up . . . . having positive sales and earnings was actually considered a drawback" (p. 77–79). But by 1987 and 1988, "market sentiment had changed from an acceptance of an exciting story . . . to a desire to stay closer to earth with low-multiple stocks that actually pay dividends" (p. 79).

The late 1990s bubble in technology stocks is familiar. By all accounts, investor sentiment was broadly high before the bubble started to burst in 2000. Cochrane (2003) and Ofek and Richardson (2002) offer ex post perspectives on the bubble, while Asness et al. (2000) and Chan, Karceski, and Lakonishok (2000) were arguing even before the crash that late 1990s growth stock valuations were difficult to ascribe to rationally expected earnings growth. Malkiel draws parallels to episodes in the 1960s, 1970s, and 1980s, and Shiller (2000) draws parallels to the late 1920s. As in earlier speculative episodes that occurred in high sentiment periods, demand for dividend payers seems to have been low (*New York Times*, 1/6/1998). Ljungqvist and Wilhelm (2003) find that 80% of the 1999 and 2000 IPO cohorts had negative earnings per share and that the median age of 1999 IPOs was 4 years. This contrasts with an average age of over 9 years just prior to the emergence of the bubble, and of over 12 years by 2001 and 2002 (Ritter (2003)).

These anecdotes suggest some regular patterns in the effect of investor sentiment on the cross-section. For instance, canonical extreme growth stocks seem to be especially prone to bubbles (and subsequent crashes), consistent with the observation that they are more appealing to speculators and optimists and at the same time hard to arbitrage. The "nifty fifty" bubble is a notable exception, but anecdotal accounts suggest that this bubble occurred during a period of broadly low sentiment, so it may still be consistent with the cross-sectional prediction that an increase in sentiment increases the *relative* price of those stocks that are the most subjective to value and the hardest to arbitrage. We now turn to formal tests of this prediction.

III. Empirical Approach and Data

A. Empirical Approach

Theory and historical anecdote both suggest that sentiment may cause systematic patterns of mispricing. Because mispricing is hard to identify directly, however, our approach is to look for systematic patterns of mispricing *correction*. For example, a pattern in which returns on young and unprofitable growth firms are (on average) especially low when beginning-of-period sentiment is estimated to be high may represent the correction of a bubble in growth stocks.

Specifically, to identify sentiment-driven changes in cross-sectional predictability patterns, we need to control for two more basic effects, namely, the generic impact of investor sentiment on all stocks and the generic impact of characteristics across all time periods. Thus, we organize our analysis loosely around the following predictive specification:

$$E_{t-1}[R_{it}] = a + a_1 T_{t-1} + \mathbf{b}_1' \mathbf{x}_{it-1} + \mathbf{b}_2' T_{t-1} \mathbf{x}_{it-1}, \tag{1}$$

where i indexes firms, t denotes time, $\mathbf{x}$ is a vector of characteristics, and T is a proxy for sentiment. The coefficient a_1 picks up the generic effect of sentiment, and the vector $\mathbf{b}_1$ the generic effect of characteristics. Our interest centers on $\mathbf{b}_2$. The null is that $\mathbf{b}_2$ equals zero or, more precisely, that any nonzero effect is rational compensation for systematic risk. The alternative is that $\mathbf{b}_2$ is nonzero and reveals cross-sectional patterns in sentiment-driven mispricing. We call Equation (1) a "conditional characteristics model" because it adds conditional terms to the characteristics model of Daniel and Titman (1997).

B. Characteristics and Returns

The firm-level data are from the merged CRSP-Compustat database. The sample includes all common stock (share codes 10 and 11) between 1962 through 2001. Following Fama and French (1992), we match accounting data for fiscal year-ends in calendar year $t-1$ to (monthly) returns from July t through June $t+1$, and we use their variable definitions when possible.

Table I shows summary statistics. Panel A summarizes returns variables. Following common practice, we define momentum, *MOM*, as the cumulative raw return for the 11-month period from 12 through 2 months prior to the observation return. Because momentum is not mentioned as a salient characteristic in historical anecdote, and theory does not suggest a direct connection between momentum and the difficulty of valuation or arbitrage, we use momentum merely as a control variable to understand the independence of our results from known mispricing patterns.

The remaining panels summarize the firm and security characteristics that we consider. The previous sections' discussions point us directly to several variables. To that list, we add a few more characteristics that, by introspection, seem likely to be salient to investors. Overall, we roughly group characteristics as pertaining to firm size and age, profitability, dividends, asset tangibility, and growth opportunities and/or distress.

Size and age characteristics include market equity, *ME*, from June of year t, measured as price times shares outstanding from CRSP. We match *ME* to monthly returns from July of year t through June of year $t+1$. Firm age, *Age*, is the number of years since the firm's first appearance on CRSP, measured to the nearest month,[7] and *Sigma* is the standard deviation of monthly returns over the 12 months ending in June of year t. If there are at least nine returns available to estimate it, *Sigma* is then matched to monthly returns from July of year t through June of year $t+1$. While historical anecdote does not identify stock volatility itself as a salient characteristic, prior work argues that it is likely to be a good proxy for the difficulty of both valuation and arbitrage.

Profitability characteristics include the return on equity, $E+/BE$, which is positive for profitable firms and zero for unprofitable firms. Earnings (E) is income before extraordinary items (Item 18) plus income statement deferred taxes (Item 50) minus preferred dividends (Item 19), if earnings are positive; book equity (BE) is shareholders equity (Item 60) plus balance sheet deferred taxes (Item 35). The profitability dummy variable $E>0$ takes the value one for profitable firms and zero for unprofitable firms.

Dividend characteristics include dividends to equity, D/BE, which is dividends per share at the ex date (Item 26) times Compustat shares outstanding (Item 25) divided by book equity. The dividend payer dummy $D>0$ takes the value one for firms with positive dividends per share by the ex date. The decline noted by Fama and French (2001) in the percentage of firms that pay dividends is apparent.

[7] Barry and Brown (1984) use the more accurate term "period of listing." A large number of firms appear on CRSP for the first time in December 1972, when Nasdaq coverage begins. Excluding these firms from our analyses of age does not change any of our inferences.

Table I

Summary Statistics, 1963–2001

Panel A summarizes the returns variables. Returns are measured monthly. Momentum (*MOM*) is defined as the cumulative return for the 11-month period between 12 and 2 months prior to t. Panel B summarizes the size, age, and risk characteristics. Size is the log of market equity. Market equity (*ME*) is price times shares outstanding from CRSP in the June prior to t. Age is the number of years between the firm's first appearance on CRSP and t. Total risk (σ) is the annual standard deviation in monthly returns from CRSP for the 12 months ending in the June prior to t. Panel C summarizes profitability variables. The earnings-book equity ratio is defined for firms with positive earnings. Earnings (E) is defined as income before extraordinary items (Item 18) plus income statement deferred taxes (Item 50) minus preferred dividends (Item 19). Book equity (*BE*) is defined as shareholders equity (Item 60) plus balance sheet deferred taxes (Item 35). We also report an indicator variable equal to one for firms with positive earnings. Panel D reports dividend variables. Dividends (D) are equal to dividends per share at the ex date (Item 26) times shares outstanding (Item 25). We scale dividends by assets and report an indicator variable equal to one for firms with positive dividends. Panel E shows tangibility measures. Plant, property, and equipment (Item 7) and research and development (Item 46) are scaled by assets. We only record research and development when it is widely available after 1971; for that period, a missing value is set to zero. Panel F reports variables used as proxies for growth opportunities and distress. The book-to-market ratio is the log of the ratio of book equity to market equity. External finance (*EF*) is equal to the change in assets (Item 6) less the change in retained earnings (Item 36). When the change in retained earnings is not available we use net income (Item 172) less common dividends (Item 21) instead. Sales growth decile is formed using NYSE breakpoints for sales growth. Sales growth is the percentage change in net sales (Item 12). In Panels C through F, accounting data from the fiscal year ending in $t-1$ are matched to monthly returns from July of year t through June of year $t+1$. All variables are Winsorized at 99.5 and 0.5%.

	Full Sample					Subsample Means				
	N	Mean	SD	Min	Max	1960s	1970s	1980s	1990s	$2000-1$
					Panel A: Returns					
R_t (%)	1,600,383	1.39	18.11	−98.13	2,400.00	1.08	1.56	1.25	1.46	1.28
MOM_{t-1} (%)	1,600,383	13.67	58.13	−85.56	343.90	21.62	12.24	15.02	13.06	11.02
					Panel B: Size, Age, and Risk					
ME_{t-1} ($M)	1,600,383	621	2,319	1	23,302	388	238	395	862	1,438
Age_t (Years)	1,600,383	13.36	13.41	0.03	68.42	15.90	12.62	13.61	13.26	13.47
σ_{t-1} (%)	1,574,981	13.70	8.73	0.00	60.77	9.44	12.51	13.32	13.89	19.55
					Panel C: Profitability					
$E+/BE_{t-1}$ (%)	1,600,383	10.70	10.03	0.00	65.14	12.10	12.05	11.37	9.54	9.49
$E > 0_{t-1}$	1,600,383	0.78	0.41	0.00	1.00	0.95	0.91	0.78	0.71	0.68
					Panel D: Dividend Policy					
D/BE_{t-1} (%)	1,600,383	2.08	2.98	0.00	17.94	4.42	2.75	2.11	1.58	1.43
$D > 0_{t-1}$	1,600,383	0.48	0.50	0.00	1.00	0.77	0.66	0.50	0.37	0.33
					Panel E: Tangibility					
PPE/A_{t-1} (%)	1,476,109	54.66	37.15	0.00	187.69	70.21	59.14	55.49	51.28	45.49
RD/A_{t-1} (%)	1,452,840	2.97	7.27	0.00	54.75		1.22	2.29	3.86	4.68
					Panel F: Growth Opportunities and Distress					
BE/ME_{t-1}	1,600,383	0.94	0.86	0.02	5.90	0.70	1.37	0.95	0.76	0.82
EF/A_{t-1} (%)	1,549,817	11.44	24.24	−71.23	127.30	7.17	6.45	10.59	13.97	17.71
GS_{t-1} (Decile)	1,529,508	5.94	3.16	1.00	10.00	5.67	5.66	6.01	6.08	5.91

The referee suggests that asset tangibility may proxy for the difficulty of valuation. Asset tangibility characteristics are measured by property, plant and equipment (Item 7) over assets, *PPE*/*A*, and research and development expense over assets (Item 46), *RD*/*A*. One concern is the coverage of the R&D variable. We do not consider this variable prior to 1972, because the Financial Accounting Standards Board did not require R&D to be expensed until 1974 and Compustat coverage prior to 1972 is very poor. Also, even in recent years less than half of the sample reports positive R&D.

Characteristics indicating growth opportunities, distress, or both include book-to-market equity, *BE*/*ME*, whose elements are defined above. External finance, *EF*/*A*, is the change in assets (Item 6) minus the change in retained earnings (Item 36) divided by assets. Sales growth (*GS*) is the change in net sales (Item 12) divided by prior-year net sales. Sales growth *GS*/10 is the decile of the firm's sales growth in the prior year relative to NYSE firms' decile breakpoints.

As will become clear below, one must grasp the multidimensional nature of the growth and distress variables in order to understand how they interact with sentiment. In particular, book-to-market wears at least three hats: High values may indicate distress; low values may indicate high growth opportunities; and, as a scaled-price variable, book-to-market is also a generic valuation indicator that varies with any source of mispricing or rational expected returns. Similarly, sales growth and external finance wear at least two hats: Low values (which are negative) may indicate distress, and high values may reflect growth opportunities. Further, to the extent that market timing motives drive external finance, *EF*/*A* also wears a third hat as a generic misvaluation indicator.

All explanatory variables are Winsorized each year at their 0.5 and 99.5 percentiles. Finally, in Panels C through F, the accounting data for fiscal years ending in calendar year $t-1$ are matched to monthly returns from July of year t through June of year $t+1$.

C. Investor Sentiment

Prior work suggests a number of proxies for sentiment to use as time-series conditioning variables. There are no definitive or uncontroversial measures, however. We therefore form a composite index of sentiment that is based on the common variation in six underlying proxies for sentiment: the closed-end fund discount, NYSE share turnover, the number and average first-day returns on IPOs, the equity share in new issues, and the dividend premium. The sentiment proxies are measured annually from 1962 to 2001. We first introduce each proxy separately, and then discuss how they are formed into overall sentiment indexes.

The closed-end fund discount, *CEFD*, is the average difference between the net asset values (NAV) of closed-end stock fund shares and their market prices. Prior work suggests that *CEFD* is inversely related to sentiment. Zweig (1973) uses it to forecast reversion in Dow Jones stocks, and Lee et al. (1991) argue that sentiment is behind various features of closed-end fund discounts. We take the value-weighted average discount on closed-end stock funds for 1962 through 1993 from Neal and Wheatley (1998), for 1994 through 1998 from CDA/Wiesenberger, and for 1999 through 2001 from turn-of-the-year issues of the *Wall Street Journal*.

NYSE share turnover is based on the ratio of reported share volume to average shares listed from the *NYSE Fact Book*. Baker and Stein (2004) suggest that turnover, or more generally liquidity, can serve as a sentiment index: In a market with short-sales constraints, irrational investors participate, and thus add liquidity, only when they are optimistic; hence, high liquidity is a symptom of overvaluation. Supporting this, Jones (2001) finds that high turnover forecasts low market returns. Turnover displays an exponential, positive trend over our period and the May 1975 elimination of fixed commissions also has a visible effect. As a partial solution, we define *TURN* as the natural log of the raw turnover ratio, detrended by the 5-year moving average.

The IPO market is often viewed as sensitive to sentiment, with high first-day returns on IPOs cited as a measure of investor enthusiasm, and the low idiosyncratic returns on IPOs often interpreted as a symptom of market timing (Stigler (1964), Ritter (1991)). We take the number of IPOs, *NIPO*, and the average first-day returns, *RIPO*, from Jay Ritter's website, which updates the sample in Ibbotson, Sindelar, and Ritter (1994).

The share of equity issues in total equity and debt issues is another measure of financing activity that may capture sentiment. Baker and Wurgler (2000) find that high values of the equity share predict low market returns. The equity share is defined as gross equity issuance divided by gross equity plus gross long-term debt issuance using data from the *Federal Reserve Bulletin*.[8]

Our sixth and last sentiment proxy is the dividend premium, P^{D-ND}, the log difference of the average market-to-book ratios of payers and nonpayers. Baker and Wurgler (2004) use this variable to proxy for relative investor demand for dividend-paying stocks. Given that payers are generally larger, more profitable firms with weaker growth opportunities (Fama and French (2001)), the dividend premium may proxy for the relative demand for this correlated bundle of characteristics.

Each sentiment proxy is likely to include a sentiment component as well as idiosyncratic, non-sentiment-related components. We use principal components analysis to isolate the common component. Another issue in forming an index is determining the relative timing of the variables—that is, if they exhibit lead-lag relationships, some variables may reflect a given shift in sentiment earlier than others. For instance, Ibbotson and Jaffe (1975), Lowry and Schwert (2002), and Benveniste et al. (2003) find that IPO volume lags the first-day returns on IPOs. Perhaps sentiment is partly behind the high first-day returns, and this attracts additional IPO volume with a lag. More generally, proxies that involve firm supply responses (*S* and *NIPO*) can be expected to lag behind proxies

[8] While they both reflect equity issues, the number of IPOs and the equity share have important differences. The equity share includes seasoned offerings, predicts market returns, and scales by total external finance to isolate the composition of finance from the level. On the other hand, the IPO variables may better reflect demand for certain IPO-like regions of the cross-section that theory and historical anecdote suggest are most sensitive to sentiment.

that are based directly on investor demand or investor behavior (*RIPO*, P^{D-ND}, *TURN*, and *CEFD*).

We form a composite index that captures the common component in the six proxies and incorporates the fact that some variables take longer to reveal the same sentiment.[9] We start by estimating the first principal component of the six proxies and their lags. This gives us a first-stage index with 12 loadings, one for each of the current and lagged proxies. We then compute the correlation between the first-stage index and the current and lagged values of each of the proxies. Finally, we define *SENTIMENT* as the first principal component of the correlation matrix of six variables—each respective proxy's lead or lag, whichever has higher correlation with the first-stage index—rescaling the coefficients so that the index has unit variance.

This procedure leads to a parsimonious index

$$SENTIMENT_t = -0.241 CEFD_t + 0.242 TURN_{t-1} + 0.253 NIPO_t + 0.257 RIPO_{t-1} + 0.112 S_t - 0.283 P^{D-ND}_{t-1}, \tag{2}$$

where each of the index components has first been standardized. The first principal component explains 49% of the sample variance, so we conclude that one factor captures much of the common variation. The correlation between the 12-term first-stage index and the *SENTIMENT* index is 0.95, suggesting that little information is lost in dropping the six terms with other time subscripts.

The *SENTIMENT* index has several appealing properties. First, each individual proxy enters with the expected sign. Second, all but one enters with the expected timing; with the exception of *CEFD*, price and investor behavior variables lead firm supply variables. Third, the index irons out some extreme observations. (The dividend premium and the first-day IPO returns reached unprecedented levels in 1999, so for these proxies to work as individual predictors in the full sample, these levels must be matched exactly to extreme future returns.)

One might object to equation (2) as a measure of sentiment on the grounds that the principal components analysis cannot distinguish between a common sentiment component and a common business cycle component. For instance, the number of IPOs varies with the business cycle in part for entirely rational reasons. We want to identify when the number of IPOs is high for *no* good reason. We therefore construct a second index that explicitly removes business cycle variation from each of the proxies prior to the principal components analysis.

Specifically, we regress each of the six raw proxies on growth in the industrial production index (Federal Reserve Statistical Release G.17), growth in consumer durables, nondurables, and services (all from BEA National Income Accounts Table 2.10), and a dummy variable for NBER recessions. The residuals from these regressions, labeled with a superscript $\perp$, may be cleaner proxies for investor sentiment. We form an index of the orthogonalized proxies following the same procedure as before. The resulting index is

[9] See Brown and Cliff (2004) for a similar approach to extracting a sentiment factor from a set of noisy proxies.

$$SENTIMENT_t^{\perp} = -0.198CEFD_t^{\perp} + 0.225TURN_{t-1}^{\perp} + 0.234NIPO_t^{\perp} + 0.263RIPO_{t-1}^{\perp} + 0.211S_t^{\perp} - 0.243P_{t-1}^{D-ND,\perp}. \quad (3)$$

Here, the first principal component explains 53% of the sample variance of the orthogonalized variables. Moreover, only the first eigenvalue is above 1.00. In terms of the signs and the timing of the components, $SENTIMENT^{\perp}$ retains all of the appealing properties of *SENTIMENT*.

Table II summarizes and correlates the sentiment measures, and Figure 1 plots them. The figure shows immediately that orthogonalizing to macro variables is a second-order issue. It does not qualitatively affect any component of the index or the overall index (see Panel E). Indeed, Table II suggests that on balance the orthogonalized proxies are slightly *more* correlated with each other than are the raw proxies. If the raw variables were driven by common macroeconomic conditions (that we failed to remove through orthogonalization) instead of common investor sentiment, one would expect the opposite. In any case, to demonstrate robustness we present results for both indexes in our main analysis.

More importantly, Figure 1 shows that the sentiment measures roughly line up with anecdotal accounts of fluctuations in sentiment. Most proxies point to low sentiment in the first few years of the sample, after the 1961 crash in growth stocks. Specifically, the closed-end fund discount and dividend premium are high, while turnover and equity issuance-related variables are low. Each variable identifies a spike in sentiment in 1968 and 1969, again matching anecdotal accounts. Sentiment then tails off until, by the mid 1970s, it is low by most measures (recall that for turnover this is confounded by deregulation). The late 1970s through mid 1980s sees generally rising sentiment, and, according to the composite index, sentiment has not dropped far below a medium level since 1980. At the end of 1999, near the peak of the Internet bubble, sentiment is high by most proxies. Overall, $SENTIMENT^{\perp}$ is positive for the years 1968–1970, 1972, 1979–1987, 1994, 1996–1997, and 1999–2001. This correspondence with anecdotal accounts seems to confirm that the measures capture the intended variation.

There are other variables that one might reasonably wish to include in a sentiment index. The main constraint is availability and consistent measurement over the 1962–2001 period. We have considered insider trading as a sentiment measure. Unfortunately, a consistent series does not appear to be available for the whole sample period. However, Nejat Seyhun shared with us his monthly series, which spans 1975 to 1994, on the fraction of public firms with net insider buying (as plotted in Seyhun (1998, p. 117)). Lakonishok and Lee (2001) study a similar series. We average Seyhun's series across months to obtain an annual series. Over the overlapping 20-year period, insider buying has a significant negative correlation with both the raw and orthogonalized sentiment indexes, and also correlates with the six underlying components as expected.

Table II

Investor Sentiment Data, 1962–2000

Means, standard deviations, and correlations for measures of investor sentiment. In the first panel, we present raw sentiment proxies. The first (*CEFD*) is the year-end, value-weighted average discount on closed-end mutual funds. The data on prices and net asset values (NAVs) come from Neal and Wheatley (1998) for 1962 through 1993, CDA/Wiesenberger for 1994 through 1998, and turn-of-the-year issues of the *Wall Street Journal* for 1999 and 2000. The second measure (*TURN*) is detrended natural log turnover. Turnover is the ratio of reported share volume to average shares listed from the NYSE Fact Book. We detrend using the past 5-year average. The third measure (*NIPO*) is the annual number of initial public offerings. The fourth measure (*RIPO*) is the average annual first-day returns of initial public offerings. Both IPO series come from Jay Ritter, updating data analyzed in Ibbotson, Sindelar, and Ritter (1994). The fifth measure (S) is gross annual equity issuance divided by gross annual equity plus debt issuance from Baker and Wurgler (2000). The sixth measure (P^{D-ND}) is the year-end log ratio of the value-weighted average market-to-book ratios of payers and nonpayers from Baker and Wurgler (2004). Turnover, the average annual first-day return, and the dividend premium are lagged 1 year relative to the other three measures. *SENTIMENT* is the first principal component of the six sentiment proxies. In the second panel, we regress each of the six proxies on the growth in industrial production, the growth in durable, nondurable, and services consumption, the growth in employment, and a flag for NBER recessions. The orthogonalized proxies, labeled with a "$\perp$," are the residuals from these regressions. $SENTIMENT^{\perp}$ is the first principal component of the six orthogonalized proxies. Superscripts a, b, and c denote statistical significance at the 1%, 5%, and 10% level, respectively.

					Correlations with Sentiment		Correlations with Sentiment Components					
	Mean	SD	Min	Max	*SENTIMENT*	$SENTIMENT^{\perp}$	*CEFD*	*TURN*	*NIPO*	*RIPO*	S	P^{D-ND}
Panel A: Raw Data												
$CEFD_t$	9.03	8.12	−10.41	23.70	−0.71[a]	−0.60[a]	1.00					
$TURN_{t-1}$	11.99	18.27	−26.70	42.96	0.71[a]	0.68[a]	−0.29[c]	1.00				
$NIPO_t$	358.41	262.76	9.00	953.00	0.74[a]	0.66[a]	−0.55[a]	0.38[b]	1.00			
$RIPO_{t-1}$	16.94	14.93	−1.67	69.53	0.76[a]	0.80[a]	−0.42[a]	0.50[a]	0.35[b]	1.00		
S_t	19.53	8.34	7.83	43.00	0.33[b]	0.44[a]	−0.01	0.30[c]	0.16	0.26	1.00	
P^{D-ND}_{t-1}	0.20	18.67	−33.17	36.06	−0.83[a]	−0.76[a]	0.52[a]	−0.50[a]	−0.56[a]	−0.58[a]	−0.12	1.00
Panel B: Controlling for Macroeconomic Conditions												
$CEFD^{\perp}_t$	0.00	6.25	−18.32	9.60	−0.62[a]	−0.63[a]	1.00					
$TURN^{\perp}_{t-1}$	0.00	15.49	−26.03	26.37	0.69[a]	0.71[a]	−0.26	1.00				
$NIPO^{\perp}_t$	0.00	226.30	−435.98	484.15	0.73[a]	0.74[a]	−0.45[a]	0.39[b]	1.00			
$RIPO^{\perp}_{t-1}$	0.00	14.31	−23.55	46.54	0.77[a]	0.83[a]	−0.46[a]	0.53[a]	0.44[a]	1.00		
$S^{\perp}_t$	0.00	6.15	−12.17	14.29	0.55[a]	0.67[a]	−0.41[a]	0.32[b]	0.50[a]	0.47[a]	1.00	
$P^{D-ND\perp}_{t-1}$	0.00	16.89	−43.20	35.96	−0.78[a]	−0.77[a]	0.26	−0.60[a]	−0.46[a]	−0.68[a]	−0.28[c]	1.00

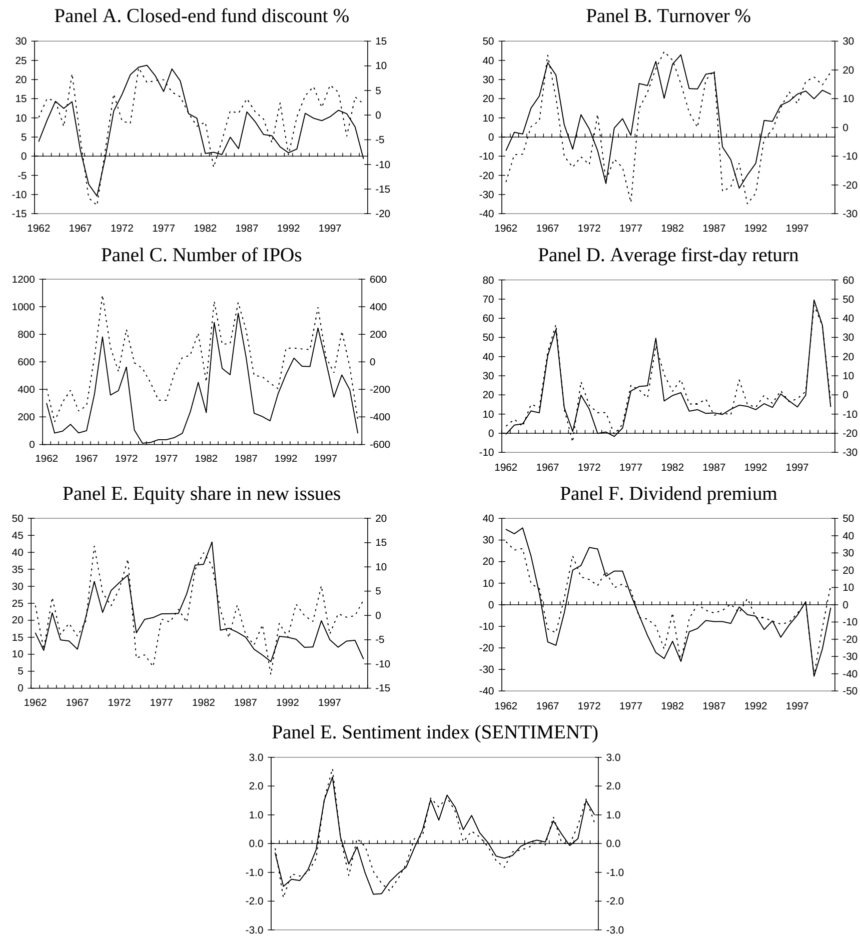


Figure 1. Investor sentiment, 1962–2001. The first panel shows the year-end, value-weighted average discount on closed-end mutual funds. The data on prices and net asset values (NAVs) come from Neal and Wheatley (1998) for 1962 through 1993, CDA/Wiesenberger for 1994 through 1998, and turn-of-the-year issues of the *Wall Street Journal* for 1999 through 2001. The second panel shows detrended log turnover. Turnover is the ratio of reported share volume to average shares listed from the NYSE Fact Book. We detrend using the past 5-year average. The third panel shows the annual number of initial public offerings. The fourth panel shows the average annual first-day returns of initial public offerings. Both series come from Jay Ritter, updating data analyzed in Ibbotson, Sindelar, and Ritter (1994). The fifth panel shows gross annual equity issuance divided by gross annual equity plus debt issuance from Baker and Wurgler (2000). The sixth panel shows the year-end log ratio of the value-weighted average market-to-book ratios of payers and nonpayers from Baker and Wurgler (2004). The solid line (left axis) is raw data. We regress each measure on the growth in industrial production, the growth in durable, nondurable, and services consumption, the growth in employment, and a flag for NBER recessions. The dashed line (right axis) is the residuals from this regression. The solid (dashed) line in the final panel is a first principal component index of the six raw (orthogonalized) measures. Both are standardized to have zero mean and unit variance. In the index, turnover, the average annual first-day return, and the dividend premium are lagged 1 year relative to the other three measures, as discussed in the text.

IV. Empirical Tests

A. Sorts

Table III looks for conditional characteristics effects in a simple, nonparametric way. We place each monthly return observation into a bin according to the decile rank that a characteristic takes at the beginning of that month, and then according to the level of $SENTIMENT^{\perp}$ at the end of the previous calendar year. To keep the meaning of the deciles similar over time, we define them based on NYSE firms. The trade-off is that there is not a uniform distribution of firms across bins in any given month. We compute the equal-weighted average monthly return for each bin and look for patterns. In particular, we identify time-series changes in cross-sectional effects from the conditional *difference* of average returns across deciles.

The first rows of Table III show the effect of size, as measured by *ME*, conditional on sentiment. These rows reveal that the size effect of Banz (1981) appears in low sentiment periods only. Specifically, Table III shows that when $SENTIMENT^{\perp}$ is negative, returns average 2.37% per month for the bottom *ME* decile and 0.92 for the top decile. A similar pattern is apparent when conditioning on *CEFD* (not reported). A link between the size effect and closed-end fund discounts is also noted by Swaminathan (1996). This pattern is consistent with some long-known results. Namely, the size effect is essentially a January effect (Keim (1983), Blume and Stambaugh (1983)), and the January effect, in turn, is stronger after a period of low returns (Reinganum (1983)), which is also when sentiment is likely to be low.

As an aside, note that the average returns across the first two rows of Table III illustrate that subsequent returns tend to be higher, across most of the cross-section, when sentiment is low. This is consistent with prior results that the equity share and turnover, for example, forecast market returns. More generally, it supports our premise that sentiment has broad effects, and so the existence of richer patterns within the cross-section is not surprising.

The conditional cross-sectional effect of *Age* is striking. In general, investors appear to demand young stocks when $SENTIMENT^{\perp}$ is positive and prefer older stocks when sentiment is negative. For example, when sentiment is pessimistic, top-decile *Age* firms return 0.54% per month *less* than bottom-decile *Age* firms. However, they return 0.85% *more* when sentiment is optimistic. When sentiment is positive, the effect is concentrated in the very youngest stocks, which are recent IPOs; when it is negative, the contrast is between the bottom and top several deciles of age. Overall, there is a nearly monotonic effect in the conditional difference of returns. This result is intriguing because *Age* has no unconditional effect.[10] *The strong conditional effects, of opposite sign, average out across high and low sentiment periods*.

[10] This conclusion is in seeming contrast to Barry and Brown's (1984) evidence of an unconditional negative period-of-listing effect; however, their sample excludes stocks listed for fewer than 61 months.

Table III

Future Returns by Sentiment Index and Firm Characteristics, 1963–2001

For each month, we form 10 equal-weighted portfolios according to the NYSE breakpoints of firm size (*ME*), *age*, total risk, earnings-book ratio for profitable firms (*E*/*BE*), dividend-book ratio for dividend payers (*D*/*BE*), fixed assets (*PPE*/*A*), research and development (*RD*/*A*), book-to-market ratio (*BE*/*ME*), external finance over assets (*EF*/*A*), and sales growth (*GS*). We also calculate portfolio returns for unprofitable firms, nonpayers, zero-PP&E firms, and zero-R&D firms. We then report average portfolio returns over months in which $SENTIMENT^{\perp}$ from the previous year-end is positive, months in which it is negative, and the difference between these two averages. $SENTIMENT^{\perp}$ is positive for 1968–1970, 1972, 1979–1987, 1994, 1996–1997, and 1999–2001 (the return series ends in 2001, so the last value used is 2000).

		Decile											Comparisons			
	$SENTIMENT^{\perp}_{t-1}$	≤0	1	2	3	4	5	6	7	8	9	10	10 − 1	10 − 5	5 − 1	>0–≤0
ME	Positive		0.73	0.74	0.85	0.83	0.92	0.84	1.06	0.99	1.02	0.98	0.26	0.06	0.20	
	Negative		2.37	1.68	1.66	1.51	1.67	1.35	1.26	1.25	1.05	0.92	−1.45	−0.75	−0.70	
	Difference		−1.65	−0.93	−0.81	−0.68	−0.75	−0.51	−0.20	−0.26	−0.03	0.06	1.71	0.81	0.90	
Age	Positive		0.25	0.83	0.94	0.95	1.18	1.19	0.96	1.18	1.09	1.11	0.85	−0.07	0.93	
	Negative		1.77	1.88	1.97	1.68	1.70	1.68	1.38	1.34	1.36	1.24	−0.54	−0.46	−0.08	
	Difference		−1.52	−1.05	−1.03	−0.74	−0.51	−0.49	−0.42	−0.16	−0.27	−0.13	1.39	0.39	1.00	
σ	Positive		1.44	1.41	1.25	1.20	1.24	1.08	1.01	0.88	0.75	0.30	−1.14	−0.94	−0.20	
	Negative		1.01	1.17	1.26	1.37	1.52	1.61	1.65	1.83	2.08	2.41	1.40	0.89	0.51	
	Difference		0.43	0.24	−0.01	−0.16	−0.28	−0.53	−0.65	−0.95	−1.33	−2.11	−2.54	−1.84	−0.71	
E/*BE*	Positive	0.35	0.68	0.85	0.86	0.89	0.92	0.88	0.92	1.05	1.10	0.93	0.24	0.01	0.24	0.61
	Negative	2.59	2.24	2.10	2.26	1.82	1.65	1.79	1.62	1.59	1.43	1.57	−0.67	−0.08	−0.59	−0.95
	Difference	−2.25	−1.56	−1.25	−1.40	−0.93	−0.73	−0.91	−0.70	−0.54	−0.34	−0.65	0.91	0.09	0.82	1.56
D/*BE*	Positive	0.44	1.08	1.09	1.29	1.11	1.24	1.17	1.31	1.24	1.19	1.15	0.07	−0.09	0.16	0.75
	Negative	2.32	1.87	1.63	1.59	1.51	1.38	1.30	1.20	1.12	1.16	1.18	−0.69	−0.19	−0.49	−0.89
	Difference	−1.88	−0.79	−0.54	−0.30	−0.40	−0.14	−0.14	0.11	0.12	0.03	−0.03	0.76	0.11	0.65	1.64
PPE/*A*	Positive	1.31	0.48	0.66	0.74	0.81	1.04	0.90	0.79	0.87	1.04	1.05	0.57	0.02	0.56	−0.53
	Negative	1.26	1.93	1.96	1.90	1.87	1.82	1.89	1.66	1.56	1.29	1.62	−0.31	−0.20	−0.11	0.53
	Difference	0.05	−1.45	−1.31	−1.17	−1.07	−0.78	−0.99	−0.87	−0.69	−0.25	−0.56	0.88	0.22	0.67	−1.05
RD/*A*	Positive	0.80	1.21	1.04	1.37	1.37	1.34	1.22	1.24	1.29	1.39	1.38	0.17	0.04	0.13	0.55
	Negative	1.63	1.57	1.47	1.58	1.73	1.66	1.81	1.97	2.04	2.13	2.44	0.87	0.78	0.09	0.43
	Difference	−0.83	−0.36	−0.43	−0.22	−0.36	−0.32	−0.60	−0.73	−0.75	−0.74	−1.05	−0.69	−0.74	0.04	0.12
BE/*ME*	Positive		0.03	0.61	0.82	0.87	0.96	1.09	1.17	1.18	1.29	1.27	1.24	0.31	0.93	
	Negative		1.41	1.43	1.46	1.54	1.61	1.69	1.87	1.94	2.18	2.45	1.04	0.84	0.20	
	Difference		−1.38	−0.81	−0.64	−0.67	−0.65	−0.60	−0.70	−0.76	−0.88	−1.18	0.20	−0.53	0.73	
EF/*A*	Positive		1.08	1.04	1.25	1.18	1.19	1.17	1.02	0.92	0.75	−0.01	−1.09	−1.20	0.11	
	Negative		2.43	2.09	1.85	1.75	1.59	1.53	1.51	1.51	1.71	1.53	−0.90	−0.06	−0.84	
	Difference		−1.35	−1.05	−0.59	−0.57	−0.40	−0.35	−0.49	−0.60	−0.96	−1.54	−0.18	−1.14	0.96	
GS	Positive		0.70	1.07	1.19	1.15	1.21	1.18	1.22	1.10	0.81	0.05	−0.65	−1.16	0.51	
	Negative		2.49	1.78	1.61	1.54	1.47	1.57	1.68	1.78	1.68	1.69	−0.80	0.22	−1.02	
	Difference		−1.79	−0.71	−0.42	−0.40	−0.26	−0.39	−0.46	−0.68	−0.87	−1.64	0.15	−1.38	1.53	

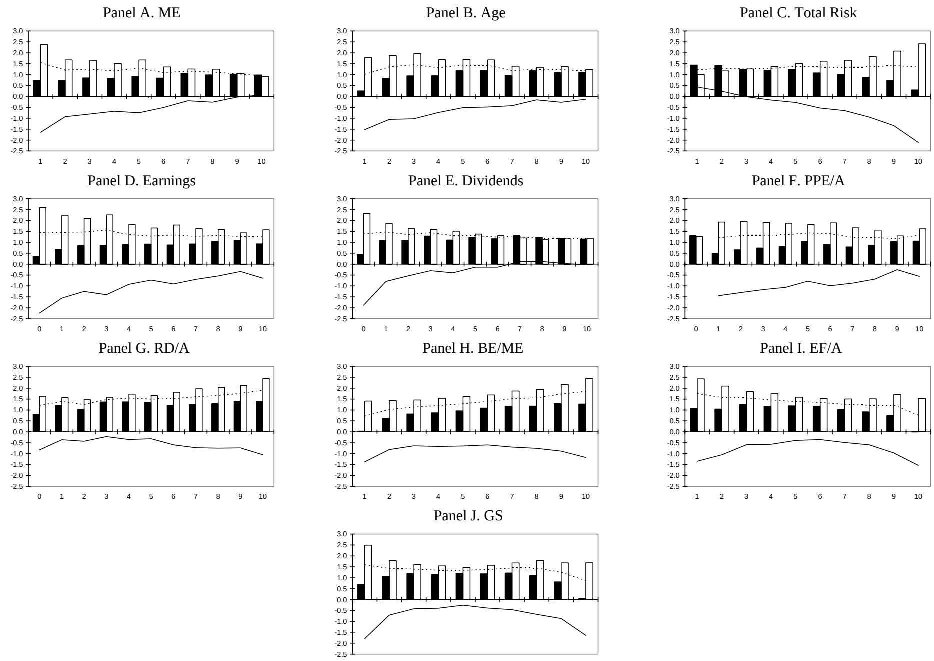


Figure 2. Two-way sorts: Future returns by sentiment index and firm characteristics, 1963–2001. For each month, we form 10 portfolios according to the NYSE breakpoints of firm size (*ME*), age, total risk, earnings-book ratio for profitable firms (*E/BE*), dividend-book ratio for payers (*D/BE*), fixed assets (*PPE/A*), research and development (*RD/A*), book-to-market ratio (*BE/ME*), external finance over assets (*EF/A*), and sales growth (*GS*). We also calculate portfolio returns for unprofitable, nonpaying, zero-PP&E, and zero-R&D firms. The solid bars are returns following positive $SENTIMENT^{\perp}$ periods, and the clear bars are returns following negative sentiment periods. The dashed line is the average across both periods and the solid line is the difference. $SENTIMENT^{\perp}$ is positive for 1968–1970, 1972, 1979–1987, 1994, 1996–1997, and 1999–2001 (returns end in 2001, so the last value used is 2000).

The next rows of Table III indicate that the cross-sectional effect of return volatility is conditional on sentiment in the hypothesized manner. In particular, high *Sigma* stocks appear to be out of favor when sentiment is low, as they earn returns of 2.41% per month over the next year. However, just as with *Age*, the cross-sectional effect of *Sigma* fully reverses in low sentiment conditions. Loosely speaking, when sentiment is high, "riskier" stocks earn *lower* returns. When sentiment is low, they earn *higher* returns. A natural interpretation is that highly volatile stocks are, like young stocks, relatively hard to value and relatively hard to arbitrage, making them especially prone to fluctuations in sentiment.

Figure 2 shows the results of Table III graphically. Panel C, for example, shows the unconditional average monthly returns across *Sigma* deciles (dashed line), which is essentially flat; the average monthly return in high sentiment periods (solid bar), which is decreasing with risk decile; the average monthly return in low sentiment periods (clear bars), which is increasing with risk deciles; and the difference in conditional returns (solid line). The solid line summarizes the difference in the relationship between *Sigma* and future returns across the two regimes and clearly illustrates that the future returns on high *Sigma* stocks are more sensitive to sentiment.

The next rows examine profitability and dividends. For average investors, perhaps the most salient comparisons are simply those between profitable and unprofitable ($E < 0$) firms and payers and nonpayers ($D = 0$). These contrasts are in the extreme right column, where we average returns across profitable (paying) firms and compare them to unprofitable (nonpaying) firms. These characteristics again display intriguing conditional sign-flip patterns. When sentiment is positive, monthly returns over the next year are 0.61% higher on profitable than unprofitable firms and 0.75% higher on payers than nonpayers. When it is negative, however, returns are 0.95% per month *lower* on profitable firms and 0.89% lower on payers. The left column shows that these patterns are driven mostly by conditional variation in the returns of unprofitable and nonpaying firms, although there are also some differences across levels of dividend payments and profitability. Again, this is consistent with unprofitable, nonpaying firms being generally harder both to value and to arbitrage, thus exposing them more to sentiment fluctuations.

The next two rows look at asset tangibility characteristics under the notion that firms with less tangible assets may be more difficult to value. The patterns here are not so strong, but there is a suggestion that firms with more intangible assets, as measured by less *PPE/A*, are more sensitive to fluctuations in sentiment. (This pattern is only apparent within firms that report positive *PPE/A*.) The clearest pattern in *RD/A* is a modest unconditional effect in which higher *RD/A* firms earn higher returns.

The remaining variables—book-to-market, external finance, and sales growth—also display intriguing patterns. Most simply, running across rows, one can see that each of them has some unconditional explanatory power. Future returns are generally higher for high *BE/ME* stocks, low *EF/A* stocks, and low *GS* decile stocks. The *EF/A* result is reminiscent of Loughran and Ritter (1995) and Spiess and Affleck-Graves (1995, 1999), while the *GS* result is suggested in Lakonishok, Shleifer, and Vishny (1994).

A closer look reveals that after controlling for these unconditional effects, a conditional pattern emerges. Specifically, there is a U-shaped pattern in the conditional difference. Consider the *GS* variable. The difference in returns on bottom-decile *GS* firms is −1.79% per month. For fifth-decile firms, the difference is only −0.26% per month. But for tenth-decile firms, the difference is again large, −1.64% per month. U-shaped patterns also appear in the conditional difference row for *BE/ME* and *EF/A*. The solid lines in Panels H–J of Figure 2 show these "frowns" graphically. The figure illustrates why one must control for the strong unconditional effects in these variables in order to see the conditional effects.

Thus, in all three of these growth and distress variables, firms with extreme values react more to sentiment than firms with middle values. What does the U reflect? It reflects the multidimensional nature of the growth and distress variables. Consider *GS*. High-*GS* firms include high-flying growth firms, low-*GS* firms are often distressed firms with shrinking sales, and middle-*GS* firms are steady, slow-growth firms. Thus, relative to firms in the middle deciles, firms with extreme values of *GS* are harder to value, and perhaps to arbitrage, and thus may be more sensitive to sentiment. Put differently, firms with extreme values of *GS* are likely to seem riskier, in a salient sense, than firms in the middle. The same explanation may help to explain the U-shaped patterns in the conditional difference row of *EF/A* and *BE/ME*. There again, low *EF/A* firms and high *BE/ME* firms include distressed firms, high *EF/A* and low *BE/ME* firms include high-flyers, and the middle deciles tend to be populated by the most "stable" firms.

In unreported results, we sort returns not just on positive and negative values of $SENTIMENT^{\perp}$ but also on >1 and <-1 standard deviation values. Not surprisingly, conditioning on more extreme values of sentiment leads to stronger results. We take more formal account of the continuous nature of the sentiment indexes in the next subsection. Also, for brevity, we omit sorts on *SENTIMENT* (the nonorthogonalized version), which give similar results. We present results for both indexes in the next section. Finally, we have also sorted returns on positive and negative $SENTIMENT^{\perp}$, where positive and negative are defined relative to a 10-year average. By requiring a 10-year history of sentiment, one loses a little more than one-quarter of the sample. The results are qualitatively identical to those in Table III, although slightly weaker except for *Age*, which is slightly stronger.

B. Predictive Regressions for Long–Short Portfolios

Another way to look for conditional characteristics effects is to use sentiment to forecast equal-weighted portfolios that are long on stocks with high values of a characteristic and short on stocks with low values. Above we see that the average payer, for example, earns higher returns than the average nonpayer when sentiment is high, so sentiment seems likely to forecast a long–short portfolio formed on dividend payment. But a regression approach allows us to conduct formal significance tests, incorporate the continuous nature of the sentiment indexes, and determine which characteristics have conditional effects that are distinct from well-known unconditional effects.

Table IV starts by plotting the average monthly returns on various long–short portfolios over time. The first several rows show that, not surprisingly, long–short portfolios formed on size (*SMB*), age, volatility, profitability, dividend payment, and (to a lesser extent) tangibility are typically highly correlated. Thus, a good question, which we address in subsequent tables, is whether the results from the sorts are all part of the same pattern or are somewhat distinct. This question is also relevant given that our portfolios are equal-weighted. By controlling for *SMB* in portfolio forecasting regressions, we can examine the extent to which the conditional predictability patterns are independent of size.

Table IV

Correlations of Portfolio Returns, 1963–2001

Correlations among characteristics-based portfolios. The sample period includes monthly returns from 1963 to 2001. The long–short portfolios are formed based on firm characteristics: firm size (*ME*), *age*, total risk (σ), profitability (*E*), dividends (*D*), fixed assets (*PPE*), research and development (*RD*), book-to-market ratio (*BE/ME*), external finance over assets (*EF/A*), and sales growth decile (*GS*). High is defined as a firm in the top three NYSE deciles, low is defined as a firm in the bottom three NYSE deciles, and medium is defined as a firm in the middle four NYSE deciles. Superscripts a, b, and c denote statistical significance at the 1%, 5%, and 10%, level, respectively.

		Size, Age, and Risk			Profitability, Dividends		Tangibility		Growth Opportunities and Distress			Growth Opportunities			Distress		
		ME	*Age*	σ	*E*	*D*	*PPE/A*	*RD/A*	*BE/ME*	*EF/A*	*GS*	*BE/ME*	*EF/A*	*GS*	*BE/ME*	*EF/A*	*GS*
ME	*SMB*	1.00															
Age	High-Low	-0.78^{a}	1.00														
σ	High-Low	0.77^{a}	-0.92^{a}	1.00													
E	$>0 - <0$	-0.64^{a}	0.85^{a}	-0.85^{a}	1.00												
D	$>0 - =0$	-0.76^{a}	0.94^{a}	-0.96^{a}	0.91^{a}	1.00											
PPE/A	High-Low	-0.74^{a}	0.85^{a}	-0.87^{a}	0.71^{a}	0.84^{a}	1.00										
RD/A	High-Low	0.61^{a}	-0.78^{a}	0.82^{a}	-0.69^{a}	-0.80^{a}	-0.79^{a}	1.00									
BE/ME	*HML*	0.09^{a}	0.47^{a}	-0.46^{a}	0.23^{a}	0.43^{a}	0.51^{a}	-0.73^{a}	1.00								
EF/A	High-Low	0.33^{a}	-0.55^{a}	0.56^{a}	-0.34^{a}	-0.51^{a}	-0.59^{a}	0.66^{a}	-0.65^{a}	1.00							
GS	High-Low	0.14^{a}	-0.17^{a}	0.23^{a}	0.10^{b}	-0.15^{a}	-0.34^{a}	0.29^{a}	-0.57^{a}	0.73^{a}	1.00						
BE/ME	Medium-Low	-0.37^{a}	0.66^{a}	-0.68^{a}	0.48^{a}	0.64^{a}	0.61^{a}	-0.80^{a}	0.81^{a}	-0.74^{a}	-0.49^{a}	1.00					
EF/A	High-Medium	0.65^{a}	-0.87^{a}	0.90^{a}	-0.76^{a}	-0.88^{a}	-0.83^{a}	0.76^{a}	-0.51^{a}	0.77^{a}	0.42^{a}	-0.73^{a}	1.00				
GS	High-Medium	0.65^{a}	-0.84^{a}	0.88^{a}	-0.69^{a}	-0.86^{a}	-0.81^{a}	0.80^{a}	-0.62^{a}	0.78^{a}	0.55^{a}	-0.77^{a}	0.92^{a}	1.00			
BE/ME	High-Medium	0.31^{a}	-0.30^{a}	0.29^{a}	-0.50^{a}	-0.35^{a}	-0.22^{a}	-0.01	0.40^{a}	-0.31^{a}	-0.53^{a}	0.16^{a}	0.14^{a}	0.03	1.00		
EF/A	Medium-Low	-0.61^{a}	0.69^{a}	-0.70^{a}	0.77^{a}	0.74^{a}	0.56^{a}	-0.39^{a}	-0.01	0.04	0.26^{a}	0.23^{a}	-0.60^{a}	-0.47^{a}	-0.61^{a}	1.00	
GS	Medium-Low	-0.61^{a}	0.80^{a}	-0.79^{a}	0.88^{a}	0.83^{a}	0.60^{a}	-0.62^{a}	0.17^{a}	-0.20^{a}	0.31^{a}	0.42^{a}	-0.65^{a}	-0.63^{a}	-0.53^{a}	0.77^{a}	1.00

In the last several rows of Table IV, we break the growth and distress variables into "high minus medium" and "medium minus low" portfolios. In the case of the *GS* variable, for example, these portfolios are highly *negatively* correlated with each other, at −0.63, indicating that high and low *GS* firms actually move together relative to middle *GS* firms. Likewise, the correlation between "high minus medium" and "medium minus low" *EF/A* is −0.60. Thus, simple "high minus low" analyses of these variables would omit crucial aspects of the cross-section.

The question is whether sentiment can predict the various long–short portfolios analyzed in Table IV. We run regressions of the type[11]

$$R_{X_{it}=High,t} - R_{X_{it}=Low,t} = c + dSENTIMENT_{t-1} + u_{it}. \quad (4)$$

The dependent variable is the monthly return on a long–short portfolio, such as *SMB*, and the monthly returns from January through December of t are regressed on the sentiment index that prevailed at the end of the prior year. We also distinguish novel predictability effects from well-known comovement using the multivariate regression

$$\begin{aligned} R_{X_{it}=High,t} - R_{X_{it}=Low,t} = c + dSENTIMENT_{t-1} + \beta RMKT_t + sSMB_t \\ + hHML_t + mUMD_t + u_{it}. \end{aligned} \quad (5)$$

The variable *RMRF* is the excess return of the value-weighted market over the risk-free rate. The variable *UMD* is the return on high-momentum stocks minus the return on low-momentum stocks, where momentum is measured over months [−12, −2]. As described in Fama and French (1993), *SMB* is the return on portfolios of small and big *ME* stocks that is separate from returns on *HML*, where *HML* is constructed to isolate the difference between high and low *BE/ME* portfolios.[12] We exclude *SMB* and *HML* from the right side when they are the portfolios being forecast. Standard errors are bootstrapped to correct for the bias induced if the autocorrelated sentiment index has innovations that are correlated with innovations in portfolio returns, as in Stambaugh (1999).

Table V shows the results. The results provide formal support to our preliminary impressions from the sorts. In particular, the first panel shows that when sentiment is high, returns on small, young, and high volatility firms are relatively low over the coming year. The coefficient on sentiment diminishes once we control for *RMRF, SMB*, *HML*, and *UMD*, but in most cases the significance of the predictive effect does not depend on including or excluding these controls. In terms of magnitudes, the coefficient for predicting *SMB*, for example, indicates that a one-unit increase in sentiment (which equals a one-SD increase, because the indexes are standardized) is associated with a −0.40% lower monthly return on the small minus large portfolio.

[11] Intuitively, in terms of equation (1), this amounts to a regression of $(b_1\Delta X + b_2T_{t-1}\Delta X)$ on sentiment proxies T_{t-1}, where ΔX is the difference between "high" and "low" levels of a characteristic.

[12] These portfolios are taken from Ken French's website and are described there.

Table V also shows that the coefficients on *SENTIMENT* and $SENTIMENT^{\perp}$ are very similar. Keep in mind that the coefficients on $SENTIMENT^{\perp}$ are essentially the same as one would find from regressing long–short portfolio returns directly on a raw sentiment index and controls for contemporaneous macroeconomic conditions—that is, regressing X on Z and using the residuals to predict Y is equivalent to regressing Y on X and Z. The similarity of the results on *SENTIMENT* and $SENTIMENT^{\perp}$ thus suggests that macroeconomic conditions play a minor role.

For profitability and dividend payment, we run regressions to predict the difference between the profitable and paying portfolios and the unprofitable and nonpaying portfolios, respectively, because the sorts suggest that these are likely to capture the main contrasts. The results show that sentiment indeed has significant predictive power for these portfolios, with higher sentiment forecasting relatively higher returns on payers and profitable firms. The patterns are little affected by controlling for *RMRF*, *SMB*, *HML*, and *UMD*.

As we find with the sorts, the tangibility characteristics do not exhibit strong conditional effects. Sentiment does have marginal predictive power for the *PPE/A* portfolio, with high sentiment associated with relatively low future returns on low *PPE/A* stocks, but this disappears after controlling for *RMRF*, *SMB*, *HML*, and *UMD*. The coefficients on the *RD/A* portfolio forecasts are not consistent in sign or magnitude.

Also as we find with the sorts, the "growth and distress" variables do not have simple monotonic relationships with sentiment. Panel D shows that sentiment does not predict simple high minus low portfolios formed on any of *BE/ME*, *EF/A*, or *GS*. However, Panels E and F show that when the multidimensional nature of these variables is incorporated, there is much stronger evidence of predictive power. We separate extreme growth opportunities effects from distress effects by constructing High, Medium, and Low portfolios based on the top three, middle four, and bottom three NYSE decile breakpoints, respectively.

The results show that when sentiment is high, subsequent returns on both low *and* high sales growth firms are low relative to returns on medium growth firms. This illustrates the U-shaped pattern in Table III in a different way, and shows that it is statistically significant. An equally significant U-shaped pattern is apparent with external finance; when sentiment is high, subsequent returns on both low *and* high external finance firms are low relative to more typical firms. In the case of *BE/ME*, however, although sentiment predicts the high minus medium and medium minus low portfolios with opposite signs, neither coefficient is reliably significant. This matches our inferences from the sorts, where we see that the U-shaped pattern in the conditional difference for *BE/ME* is somewhat weaker than for *EF/A* and *GS*.

Equations (4) and (5) offer a simple framework in which to address some robustness issues. To test whether the results are driven by an overall trend, we include a post-1982 dummy in the regressions, with no change in inferences from those in the last column of Table V. Also, the results are slightly stronger when returns from January and December are removed from the sample. This

Table V

Time Series Regressions of Portfolio Returns, 1963 to 2001

Regressions of long–short portfolio returns on lagged *SENTIMENT*, the market risk premium (*RMRF*), the Fama–French factors (*HML* and *SMB*), and a momentum factor (*UMD*).

$$R_{X_{it}=High,t} - R_{X_{it}=Low,t} = c + dSENTIMENT_{t-1} + \beta RMRF_t + sSMB_t + hHML_t + mUMD_t + u_t.$$

The sample period includes monthly returns from 1963 to 2001. The long–short portfolios are formed based on firm characteristics (X): firm size (*ME*), *age*, total risk (σ), profitability (*E*), dividends (*D*), fixed assets (*PPE*), research and development (*RD*), book-to-market ratio (*BE/ME*), external finance over assets (*EF/A*), and sales growth decile (*GS*). High is defined as a firm in the top three NYSE deciles, low is defined as a firm in the bottom three NYSE deciles, and medium is defined as a firm in the middle four NYSE deciles. Average monthly returns are matched to *SENTIMENT* from the previous year-end. $SENTIMENT^{\perp}$ index is based on six sentiment proxies that have been orthogonalized to growth in industrial production, the growth in durable, nondurable, and services consumption, the growth in employment, and a flag for NBER recessions; the components of *SENTIMENT* are not orthogonalized. The first and third sets of columns show univariate regression results, while the second and the fourth columns include *RMRF*, *SMB*, *HML*, and *UMD* as control variables. *SMB* (*HML*) is not included as a control variable when *SMB* (*HML*) is the dependent variable. Bootstrapped p-values are in brackets.

		$SENTIMENT_{t-1}$		$SENTIMENT_{t-1}$ Controlling for $RMRF$, SMB, HML, UMD		$SENTIMENT^{\perp}_{t-1}$		$SENTIMENT^{\perp}_{t-1}$ Controlling for $RMRF$, SMB, HML, UMD	
		d	**p(*d*)**	***d***	**p(*d*)**	***d***	**p(*d*)**	***d***	**p(*d*)**
Panel A: Size, Age, and Risk									
ME	***SMB***	−0.4	[0.04]	−0.3	[0.08]	−0.4	[0.06]	−0.3	[0.15]
Age	**High-Low**	0.5	[0.02]	0.2	[0.08]	0.5	[0.01]	0.2	[0.04]
σ	**High-Low**	−1.0	[0.01]	−0.5	[0.01]	−0.9	[0.00]	−0.4	[0.01]
Panel B: Profitability and Dividend Policy									
E	**>0 – <0**	0.8	[0.00]	0.5	[0.02]	0.8	[0.00]	0.4	[0.02]
D	**>0 – =0**	0.8	[0.00]	0.4	[0.00]	0.8	[0.00]	0.4	[0.00]

(continued)

Table V—*Continued*

		$SENTIMENT_{t-1}$		$SENTIMENT_{t-1}$ Controlling for *RMRF*, *SMB*, *HML*, *UMD*		$SENTIMENT^{\perp}_{t-1}$		$SENTIMENT^{\perp}_{t-1}$ Controlling for *RMRF*, *SMB*, *HML*, *UMD*	
		d	**p(*d*)**	***d***	**p(*d*)**	***d***	**p(*d*)**	***d***	**p(*d*)**
Panel C: Tangibility									
PPE/A	**High-Low**	0.4	[0.12]	0.1	[0.65]	0.4	[0.07]	0.1	[0.53]
RD/A	**High-Low**	−0.3	[0.25]	0.0	[0.77]	−0.3	[0.24]	0.1	[0.62]
Panel D: Growth Opportunities and Distress									
BE/ME	***HML***	0.1	[0.47]	0.0	[1.00]	0.2	[0.25]	0.1	[0.68]
EF/A	**High-Low**	−0.1	[0.24]	−0.1	[0.46]	−0.2	[0.13]	−0.1	[0.40]
GS	**High-Low**	−0.1	[0.53]	−0.0	[0.67]	−0.1	[0.23]	−0.1	[0.51]
Panel E: Growth Opportunities									
BE/ME	**Medium-Low**	0.2	[0.11]	0.1	[0.39]	0.3	[0.06]	0.1	[0.35]
EF/A	**High-Medium**	−0.4	[0.00]	−0.2	[0.01]	−0.4	[0.00]	−0.2	[0.00]
GS	**High-Medium**	−0.4	[0.00]	−0.2	[0.00]	−0.4	[0.00]	−0.2	[0.00]
Panel F: Distress									
BE/ME	**High-Medium**	−0.1	[0.21]	−0.1	[0.49]	−0.1	[0.30]	−0.1	[0.53]
EF/A	**Medium-Low**	0.3	[0.00]	0.2	[0.01]	0.2	[0.00]	0.2	[0.00]
GS	**Medium-Low**	0.3	[0.01]	0.2	[0.04]	0.3	[0.01]	0.2	[0.07]

indicates that tax-motivated trading and associated fluctuations in liquidity around the turn of the year do not drive the main results. Further, our portfolios are equal-weighted. As mentioned previously, the purpose of this is that theory predicts that small firms will be most affected by sentiment, and hence value weighting will obscure the relevant patterns. Yet by sorting on characteristics that are correlated with size, as several of our characteristics are, and then equal-weighting these characteristics portfolios, one worries that we are just picking up the size effect once again. By controlling for *SMB* in portfolio forecasting regressions, we can see that several of the conditional predictability patterns are distinguishable from size, though as theory predicts the predictive coefficient is attenuated. Finally, while we omit the results for brevity, the six individual sentiment components generally predict the portfolio returns with the expected sign. The number of IPOs and the closed-end fund discount offer the best individual performance, followed by the equity share, turnover, the average first-day return, and the dividend premium. (Those results are reported in the NBER working paper version of this paper.)

In summary, the regressions essentially confirm the significance of the patterns suggested in the sorts. When sentiment is high, future returns are relatively low for small firms, the youngest firms, firms with volatile stock returns, unprofitable firms, non-dividend-paying firms, high growth firms, and distressed firms. And vice-versa. In general, the results support predictions that sentiment has stronger effects on stocks that are hard to value and hard to arbitrage.

C. A Brief Look at Earlier Data

Reliable accounting information, especially on the sorts of firms most affected by sentiment, is not easy to obtain for the pre-Compustat era. Some of our sentiment proxies also are not available. However, using CRSP data, we can perform a reduced set of tests over a longer period. Specifically, we form a sentiment index from 1935 to 2001 using the first principal component of *CEFD*, *S*, and *TURN*, where *TURN* is lagged relative to the others, in the spirit of equation (2).[13] We also orthogonalize these sentiment proxies with respect to consumption growth variables and NBER recessions (industrial production is not available over the full period) to form an index in the spirit of equation (3). We use these indexes to forecast the return on *SMB* and long–short portfolios formed on *Age*, *Sigma*, and dividend payer status.

The results are in Table VI. With the exception of the *Age* portfolio, for which the results are not significant, the results from the full 1935–2001 period and the "out-of-sample" 1935–1961 period are similar to those in more

[13] The closed-end fund discount is first available in 1933 from Neal and Wheatley (1998): "Wiesenberger's survey has published end-of-year fund prices and net asset values since 1943. Moreover, the first edition of the survey contains end-of-year data from 1933 to 1942." Turnover and the equity share in new issues are available in earlier years. None of our inferences in Panel A of Table VI change when we use a longer sample period and a sentiment index based on these two variables alone.

Table VI

Time Series Regressions of Portfolio Returns, 1935 to 2001

Regressions of long–short portfolio returns on lagged *SENTIMENT*, the market risk premium (*RMRF*), the Fama–French factors (*HML* and *SMB*), and a momentum factor (*UMD*).

$$R_{X_{it}=High,t} - R_{X_{it}=Low,t} = c + dSENTIMENT_{t-1} + \beta RMRF_t + sSMB_t + hHML_t + mUMD_t + u_t.$$

The long–short portfolios are formed based on firm characteristics (X): firm size (*ME*), *age*, total risk (σ), and dividends (D). High is defined as a firm in the top three NYSE deciles, and low is defined as a firm in the bottom three NYSE deciles. The sentiment index is the first principal component of the closed-end fund discount (*CEFD*), the equity share (S), and the lag of detrended log turnover (*TURN*). Average monthly returns are matched to *SENTIMENT* from the previous year-end. $SENTIMENT^{\perp}$ index is based on six sentiment proxies that have been orthogonalized to growth in industrial production, the growth in durable, nondurable, and services consumption, the growth in employment, and a flag for NBER recessions; the components of *SENTIMENT* are not orthogonalized. The first and third sets of columns show univariate regression results, while the second and the fourth columns include *RMRF*, *SMB*, *HML*, and *UMD* as control variables. *SMB* (*HML*) is not included as a control variable when *SMB* (*HML*) is the dependent variable. Bootstrapped p-values are in brackets.

		$SENTIMENT_{t-1}$		$SENTIMENT_{t-1}$ Controlling for *RMRF*, *SMB*, *HML*, *UMD*		$SENTIMENT^{\perp}_{t-1}$		$SENTIMENT^{\perp}_{t-1}$ Controlling for *RMRF*, *SMB*, *HML*, *UMD*	
		d	**p(*d*)**	***d***	**p(*d*)**	***d***	**p(*d*)**	***d***	**p(*d*)**
		Panel A: 1935–2001							
ME	***SMB***	−0.3	[0.03]	−0.2	[0.07]	−0.3	[0.04]	−0.2	[0.07]
Age	**High-Low**	0.2	[0.18]	0.1	[0.38]	0.2	[0.10]	0.1	[0.26]
σ	**High-Low**	−1.0	[0.00]	−0.4	[0.00]	−0.8	[0.00]	−0.4	[0.00]
D	$> \mathbf{0} - = \mathbf{0}$	0.9	[0.00]	0.5	[0.01]	0.7	[0.00]	0.4	[0.01]
		Panel B: 1935–1961							
ME	**SMB**	−0.3	[0.05]	−0.3	[0.05]	−0.1	[0.33]	−0.1	[0.33]
Age	**High-Low**	−0.1	[0.41]	−0.1	[0.41]	−0.1	[0.04]	−0.1	[0.05]
σ	**High-Low**	−0.8	[0.01]	−0.8	[0.01]	−0.4	[0.14]	−0.4	[0.16]
D	$>\mathbf{0} - = \mathbf{0}$	0.9	[0.01]	0.9	[0.01]	0.5	[0.10]	0.5	[0.11]

recent data.[14] One possibility for the insignificant results on the *Age* portfolio is that we measure age as the number of months for which CRSP data are available. Anecdotal evidence suggests that in these early data, there are fewer truly "young" firms listing on the NYSE. In contrast, in recent years, many genuinely young IPOs start trading on Nasdaq, so our way of measuring age may be more meaningful.

The longer time series make it possible to conduct an out-of-sample test. In unreported results, we compare the in-sample reduction in root mean squared

[14] For a more detailed look at earlier data, see Gruber (1966). He documents changes in the cross-sectional determinants of stock prices between 1951 and 1963, and argues that they are connected to changes in the average investor's time horizon, that is, shifts in the term structure of discount rates. This is similar to our notion of sentiment as a shift in the propensity to speculate.

error (RMSE) in Table VI to the reduction in RMSE that an investor might see using only past data. The results suggest that a substantial fraction of the portfolio predictability would have been "knowable" in advance. The exceptions are the *Age* portfolio and the *SMB* portfolio in the post-1980 period (for which the in-sample predictive power for *SMB* is also modest). A table is available on request.

Together, the longer-sample results and the out-of-sample exercise rule out the possibility that a spurious correlation is behind the main results. The fact that there are at least several fluctuations in sentiment, and the fact that the cross-sectional patterns tend to work in the predicted directions, cast further doubt on that notion.

D. Systematic Risk

At face value, the conditional characteristics effects seem unlikely to be compensation for systematic risk. Among other considerations, the index $SENTIMENT^{\perp}$ is orthogonalized to macroeconomic conditions; the patterns match predictions about where sentiment should matter most; and the patterns line up with anecdotal accounts of bubbles and crashes. Intuitively, the systematic risk explanation requires that older, profitable, less volatile, dividend-paying firms often require *higher* returns than younger, unprofitable, more volatile, nonpaying firms, and are recognized as *riskier* in the relevant sense by the marginal investor. While this proposition already seems counterintuitive, we attempt to rule it out more rigorously.

Systematic risk explanations come in two basic flavors. One is that the systematic risks (beta loadings) of stocks with certain characteristics vary with the sentiment proxies, despite our effort to isolate them from macroeconomic conditions. We investigate this directly in Table VII, where we ask whether sentiment coincides with time-variation in market betas in a way that could at least qualitatively reconcile the earlier results with a conditional CAPM. Specifically, we predict returns on the characteristics portfolios

$$R_{X_{it}=High,t} - R_{X_{it}=Low,t} = c + dSENTIMENT_{t-1} + \beta(e + fSENTIMENT_{t-1})RMRF_t + u_{it}. \qquad (6)$$

The time-varying betas story predicts that the composite coefficient βf, reported in Table VII, has the same sign as the estimates of d in Table V. However, it turns out that when the coefficient βf is significant, it is typically of the wrong sign. We obtain similar results when we replace *RMRF* by aggregate consumption growth. A table is available upon request.

The second systematic risk story keeps stocks' betas fixed, but allows the risk premium to vary with sentiment, which means that the difference in required returns between the high and low beta stocks varies in proportion. However, this story runs into trouble with the simple fact that the predicted effect of several characteristics varies not just in magnitude over time, but also in sign. It would seem then that the bulk of the results do not reflect compensation for classical systematic risks.

Table VII

Conditional Market Betas, 1963–2001

Regressions of long–short portfolio returns on the market risk premium (*RMRF*) and the market risk premium interacted with *SENTIMENT*.

$$R_{X_{it}=High,t} - R_{X_{it}=Low,t} = c + dSENTIMENT_{t-1} + \beta(e + fSENTIMENT_{t-1})RMRF_t + u_t.$$

The long–short portfolios are formed based on firm characteristics (*X*): firm size (*ME*), *age*, total risk (σ), profitability (*E*), dividends (*D*), fixed assets (*PPE*), research and development (*RD*), book-to-market ratio (*BE/ME*), external finance over assets (*EF/A*), and sales growth decile (*GS*). High is defined as a firm in the top three NYSE deciles, low is defined as a firm in the bottom three NYSE deciles, and medium is defined as a firm in the middle four NYSE deciles. Monthly returns are matched to *SENTIMENT* from the previous year-end. $SENTIMENT^{\perp}$ index is based on six sentiment proxies that have been orthogonalized to growth in industrial production, the growth in durable, nondurable and services consumption, the growth in employment and a flag for NBER recessions; the components of *SENTIMENT* are not orthogonalized. Heteroskedasticity-robust *p*-values are in brackets. A superscript "a" indicates a statistically significant βf that matches the sign of the return predictability from Table V; "b" indicates a statistically significant βf that does not match.

		$SENTIMENT_{t-1}$		$SENTIMENT^{\perp}_{t-1}$	
		βf	$t(\beta f)$	βf	$t(\beta f)$
Panel A: Size, Age, and Risk					
ME	***SMB***	−0.03	[0.48]	−0.02	[0.62]
Age	**High-Low**	−0.05	[0.19]	−0.07	[0.09]
σ	**High-Low**	0.00	[0.98]	0.03	[0.58]
Panel B: Profitability and Dividend Policy					
E	**>0 – <0**	−0.04	[0.47]	−0.00	[0.98]
D	**>0 – = 0**	−0.01	[0.75]	−0.04	[0.35]
Panel C: Tangibility					
PPE/A	**High-Low**	−0.00	[0.94]	−0.01	[0.76]
RD/A	**High-Low**	0.12[b]	[0.01]	0.17[b]	[0.00]
Panel D: Growth Opportunities and Distress					
BE/ME	***HML***	−0.10[b]	[0.02]	−0.12[b]	[0.00]
EF/A	**High-Low**	0.06[b]	[0.00]	0.07[b]	[0.00]
GS	**High-Low**	0.42	[0.06]	0.37	[0.08]
Panel E: Growth Opportunities					
BE/ME	**Medium-Low**	−0.06	[0.05]	−0.09[b]	[0.01]
EF/A	**High-Medium**	0.03	[0.13]	0.04	[0.05]
GS	**High-Medium**	0.05[b]	[0.02]	0.06[b]	[0.01]
Panel F: Distress					
BE/ME	**High-Medium**	−0.07[a]	[0.00]	−0.07[a]	[0.00]
EF/A	**Medium-Low**	0.03	[0.08]	0.02	[0.17]
GS	**Medium-Low**	−0.01	[0.62]	−0.02	[0.27]

E. Predictive Regressions for Earnings Announcement Returns

Our last test is whether there are conditional characteristics effects in the returns around earnings announcements. La Porta et al. (1997) find that low book-to-market stocks have lower average returns at earnings announcements than high book-to-market stocks, suggesting systematic errors in earnings expectations. Likewise, if errors in earnings expectations account for some of our results, we might expect that the average earnings announcement return on small, young, volatile, unprofitable, nonpaying, extreme growth, and/or distress stocks would tend to be inversely related to sentiment.

This methodology, while appealing at first glance, has only limited power to detect how expectational errors affect our results. That is, our results are driven by the correlated correction of mispricing, but a firm's announcement event return picks up the expectational corrections that occur only to it alone, within its own announcement window. An anecdote from Malkiel (1999) illustrates the problem: "The music slowed drastically for the conglomerates on January 19, 1968. On that day, the granddaddy of the conglomerates, Litton Industries, announced that earnings for the second quarter of that year would be substantially less than forecast.... the announcement was greeted with disbelief and shock. In the selling wave that followed, conglomerate stocks declined by roughly 40 percent..." (p. 67). So although a study of announcement event returns captures the corrective effect of Litton Industries' announcement on its own stock, it picks up none of its broader effects, which appear to be important to our main results. Nevertheless, an analysis of earnings announcements may provide a lower bound on the effect that sentiment-driven expectational errors have on our results.

We gather quarterly earnings announcement dates from the merged CRSP-Compustat file. These dates are available beginning in January 1971. The quarterly earnings announcement sample represents approximately 75% of the firm-quarters (firm-months) analyzed in the main tables, so coverage is fairly complete. For each firm-quarter observation, we compute the cumulative abnormal return over the value-weighted market index over trading days $[-1, +1]$ around the report date. We then construct a quarterly series of average announcement effects for each characteristic decile, and attempt to predict it with the composite sentiment index, that is,

$$CAR_{X_{it}=Decile,t} = c + dSENTIMENT^{\perp}_{t-1} + u_t. \tag{7}$$

Table VIII reports the coefficient estimates for each characteristic decile using the orthogonalized sentiment index. The results for the raw index are very similar.

Perhaps the most striking feature of Table VIII is that most coefficients are negative, thus earnings announcement effects are in general lower following high sentiment periods. A very crude comparison can be made between the cross-sectional patterns in Table VIII and those in Table III. In Table VIII, 12 of the 104 coefficients are significant at the 5% level. In Table III, 9 of the 104 estimated conditional differences are larger than 1.5% per month in absolute

Table VIII

Earnings Announcement Effects, 1973–2001

Regressions of average quarterly earnings announcement effects on lagged *SENTIMENT*.

$$CAR_{X_{it}=Decile,t} = c + dSENTIMENT^{\perp}_{t-1} + u_t.$$

We report the coefficient b below. For each calendar quarter, we form 10 portfolios according to the NYSE breakpoints of firm size (*ME*), *age*, total risk, earnings-book ratio for profitable firms (*E/BE*), dividend-book ratio for dividend payers (*D/BE*), fixed assets (*PPE/A*), research and development (*RD/A*), book-to-market ratio (*BE/ME*), external finance over assets (*EF/A*), and sales growth (*GS*). We also calculate average announcement effects for unprofitable firms, nonpayers, zero-PP&E firms, and zero-R&D firms. Quarterly average announcement effects are matched to $SENTIMENT^{\perp}$ from the previous year-end. $SENTIMENT^{\perp}$ index is based on six sentiment proxies that have been orthogonalized to growth in industrial production, the growth in durable, nondurable, and services consumption, the growth in employment, and a flag for NBER recessions. Heteroskedasticity-robust p-values are in brackets.

	Decile										
	≤0	1	2	3	4	5	6	7	8	9	10
ME		−0.18	−0.05	−0.02	−0.07	−0.02	−0.05	0.00	−0.03	0.01	−0.03
		[0.03]	[0.35]	[0.76]	[0.27]	[0.64]	[0.36]	[0.99]	[0.50]	[0.90]	[0.48]
Age		−0.11	0.02	−0.06	−0.06	−0.12	0.00	0.06	−0.03	0.00	−0.12
		[0.20]	[0.79]	[0.27]	[0.35]	[0.08]	[1.00]	[0.21]	[0.64]	[1.00]	[0.01]
σ		0.10	0.05	0.05	0.05	−0.04	−0.03	0.03	−0.02	−0.08	−0.30
		[0.05]	[0.23]	[0.26]	[0.32]	[0.45]	[0.59]	[0.69]	[0.79]	[0.27]	[0.00]
E/BE	−0.31	−0.24	0.01	0.11	−0.08	−0.09	−0.02	−0.01	0.04	0.04	0.10
	[0.00]	[0.01]	[0.92]	[0.24]	[0.31]	[0.23]	[0.73]	[0.85]	[0.54]	[0.47]	[0.08]
D/BE	−0.18	0.00	−0.05	−0.02	0.01	0.00	0.03	−0.03	−0.08	0.04	−0.10
	[0.02]	[0.94]	[0.43]	[0.72]	[0.80]	[1.00]	[0.56]	[0.60]	[0.07]	[0.43]	[0.16]
PPE/A	−0.12	−0.12	−0.12	−0.14	−0.02	0.01	0.00	−0.12	−0.10	−0.06	−0.05
	[0.14]	[0.12]	[0.14]	[0.05]	[0.75]	[0.87]	[0.97]	[0.07]	[0.09]	[0.28]	[0.41]
RD/A	−0.05	−0.10	−0.07	−0.04	−0.10	−0.15	−0.28	−0.29	−0.14	−0.11	−0.08
	[0.32]	[0.45]	[0.46]	[0.66]	[0.22]	[0.12]	[0.00]	[0.01]	[0.10]	[0.21]	[0.42]
BE/ME		−0.11	−0.03	0.03	−0.04	−0.04	0.04	−0.05	0.01	−0.14	−0.12
		[0.06]	[0.67]	[0.62]	[0.38]	[0.43]	[0.54]	[0.37]	[0.87]	[0.05]	[0.24]
EF/A		−0.07	−0.01	0.05	−0.06	0.03	−0.04	−0.10	0.01	−0.11	−0.09
		[0.38]	[0.95]	[0.40]	[0.31]	[0.57]	[0.51]	[0.07]	[0.90]	[0.06]	[0.20]
GS		−0.20	0.00	−0.08	−0.02	−0.03	0.04	−0.03	0.03	0.00	−0.11
		[0.02]	[0.99]	[0.23]	[0.68]	[0.60]	[0.51]	[0.56]	[0.60]	[0.94]	[0.15]

value. The intersection of the two tables' "strong results" is six cells, and the signs of the effects are congruent in all cases.

Overall, this suggests that some portion of the conditional characteristics effects may reflect the correction of errors in earnings expectations. However, as noted above, this test is not powerful and provides only a lower bound on the contribution of expectational errors.

V. Conclusion

In classical finance theory, investor sentiment does not play any role in the cross-section of stock prices, realized returns, or expected returns. This paper challenges that view. We use simple theoretical arguments, historical accounts of speculative episodes, and most importantly a set of novel empirical results to demonstrate that investor sentiment, broadly defined, has significant cross-sectional effects.

Our main empirical finding is that the cross-section of future stock returns is conditional on beginning-of-period proxies for sentiment. The patterns are rich but intuitive. When sentiment is estimated to be high, stocks that are attractive to optimists and speculators and at the same time unattractive to arbitrageurs—younger stocks, small stocks, unprofitable stocks, non-dividend-paying stocks, high volatility stocks, extreme growth stocks, and distressed stocks—tend to earn relatively low subsequent returns. Conditional on low sentiment, however, these cross-sectional patterns attenuate or completely reverse. The most striking finding is that several firm characteristics that display no unconditional predictive power actually hide strong conditional patterns that become visible only after conditioning on sentiment. We consider the classical explanation that the results reflect compensation for systematic risks, but several aspects of the results are inconsistent with this explanation.

The results suggest several avenues for future work. In corporate finance, a better understanding of sentiment may shed light on patterns in security issuance and the supply of firm characteristics that seem to be conditionally relevant to share price. In asset pricing, the results suggest that descriptively accurate models of prices and expected returns need to incorporate a prominent role for investor sentiment.

REFERENCES

Aghion, Philippe, and Jeremy Stein, 2004, Growth versus margins: The destabilizing consequences of giving the stock market what it wants, Working paper, Harvard University.

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223–249.

Asness, Clifford S., Jacques A. Friedman, Robert J. Krail, and John M. Liew, 2000, Style timing: Value versus growth, *Journal of Portfolio Management* 26, 50–60.

Baker, Malcolm, and Jeremy Stein, 2004, Market liquidity as a sentiment indicator, *Journal of Financial Markets* 7, 271–299.

Baker, Malcolm, and Jeffrey Wurgler, 2000, The equity share in new issues and aggregate stock returns, *Journal of Finance* 55, 2219–2257.

Baker, Malcolm, and Jeffrey Wurgler, 2004, A catering theory of dividends, *Journal of Finance* 59, 1125–1165.

Banz, Rolf, 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Barberis, Nicholas, and Andrei Shleifer, 2003, Style investing, *Journal of Financial Economics* 68, 161–199.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283–318.

Barry, Christopher B., and Stephen J. Brown, 1984, Differential information and the small firm effect, *Journal of Financial Economics* 13, 283–294.

Benveniste, Lawrence M., Alexander P. Ljungqvist, William J. Wilhelm, Jr., and Xiaoyun Yu, 2003, Evidence of information spillovers in the production of investment banking services, *Journal of Finance* 58, 577–608.

Blume, Marshall, and Robert Stambaugh, 1983, Biases in computed returns: An application to the size effect, *Journal of Financial Economics* 12, 387–404.

Brown, Gregory W., and Michael T. Cliff, 2004, Investor sentiment and the near-term stock market, *Journal of Empirical Finance* 11, 1–27.

Brown, John Dennis, 1991, *101 Years on Wall Street* (Prentice Hall, Englewood Cliffs, NJ).

Brown, Stephen J., William N. Goetzmann, Takato Hiraki, Noriyoshi Shiraishi, and Masahiro Watanabe, 2003, Investor sentiment in Japanese and U.S. daily mutual fund flows, Working paper, Yale University.

Brunnermeier, Markus, and Lasse Pedersen, 2005, Predatory trading, *Journal of Finance* 60, 1825–1863.

Campbell, John Y., and John H. Cochrane, 2000, Explaining the poor performance of consumption-based asset pricing models, *Journal of Finance* 55, 2863–2878.

Chan, Louis K.C., Jason Karceski, and Josef Lakonishok, 2000, New paradigm or same old hype in equity investing? *Financial Analysts Journal* 56, 23–36.

Cochrane, John H., 2003, Stocks as money: Convenience yield and the tech-stock bubble, in William C. Hunter, George G. Kaufman, and Michael Pomerleano, eds.: *Asset Price Bubbles* (MIT Press, Cambridge).

D'Avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross-sectional variation in stock returns, *Journal of Finance* 46, 1739–1764.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.

Dreman, David, 1979, *Contrarian Investment Strategy*, Dreman Contrarian Group.

Duffie, Darrell, Nicolae Garleanu, and Lasse H. Pedersen, 2002, Securities lending, shorting, and pricing, *Journal of Financial Economics* 66, 307–339.

Elton, Edwin J., Martin J. Gruber, and Jeffrey A. Busse, 1998, Do investors care about sentiment?, *Journal of Business* 71, 477–500.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2001, Disappearing dividends: Changing firm characteristics or lower propensity to pay? *Journal of Financial Economics* 60, 3–44.

Geczy, Christopher C., David K. Musto, and Adam V. Reed, 2002, Stocks are special too: An analysis of the equity lending market, *Journal of Financial Economics* 66, 241–269.

Gomes, Joao, Leonid Kogan, and Lu Zhang, 2003, Equilibrium cross section of returns, *Journal of Political Economy* 111, 693–732.

Graham, Benjamin, 1973, *The Intelligent Investor*, *4th ed.* (Harper & Row, Amsterdam).

Greenwood, Robin, and Nathan Sosner, 2003, Trade and the comovement of stock returns: Evidence from Japan, Working paper, Harvard University.

Gruber, Martin J., 1966, The determinants of common stock prices, Ph.D. dissertation, Columbia University.

Hausch, Donald, and William Ziemba, 1995, Efficiency of sports and lottery betting markets, In: *Handbooks in Operations Research and Management Science*, Volume 9, Elsevier, New York, NY.

Ibbotson, Roger, and Jeffrey F. Jaffe, 1975, 'Hot issue' markets, *Journal of Finance* 30, 1027–1042.

Ibbotson, Roger, Jody Sindelar, and Jay Ritter, 1994, The market's problems with the pricing of initial public offerings, *Journal of Applied Corporate Finance* 7, 66–74.

Jones, Charles, 2001, A century of stock market liquidity and trading costs, Working paper, Columbia University.

Jones, Charles, and Owen Lamont, 2002, Short sale constraints and stock returns, *Journal of Financial Economics* 66, 207–239.

Keim, Donald, 1983, Size-related anomalies and stock return seasonality: Further empirical evidence, *Journal of Financial Economics* 12, 13–32.

Kothari, S. P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169–203.

Kindleberger, Charles, 2001, *Manias, Panics, and Crashes* (Wiley, New York).

La Porta, Rafael, Josef Lakonishok, Andrei Shleifer, and Robert W. Vishny, 1997, Good news for value stocks: Further evidence on market efficiency, *Journal of Finance* 52, 859–874.

Lakonishok, Josef, and Inmoo Lee, 2001, Are insider trades informative?, *Review of Financial Studies* 14, 79–111.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

Lamont, Owen A., and Richard H. Thaler, 2003, Can the market add and subtract? Mispricing and tech stock carve-outs, *Journal of Political Economy* 111, 227–268.

Lancaster, Kelvin J., 1966, A new approach to consumer theory, *Journal of Political Economy* 74, 132–157.

Lancaster, Kelvin J., 1971, *Consumer Demand: A New Approach* (Columbia UP, New York).

Lee, Charles, Andrei Shleifer, and Richard H. Thaler, 1991, Investor sentiment and the closed-end fund puzzle, *Journal of Finance* 46, 75–109.

Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Ljungqvist, Alexander, and William Wilhelm, 2003, IPO pricing in the dot-com bubble, *Journal of Finance* 58, 723–752.

Loughran, Tim, and Jay Ritter, 1995, The new issues puzzle, *Journal of Finance* 50, 23–51.

Lowry, Michelle, and G. William Schwert, 2002, IPO market cycles: Bubbles or sequential learning?, *Journal of Finance* 57, 1171–1200.

Malkiel, Burton, 1990, *A Random Walk Down Wall Street* (W.W. Norton, New York).

Malkiel, Burton, 1999, *A Random Walk Down Wall Street* (W.W. Norton, New York).

Markowitz, Harry, 1959, *Portfolio Selection: Efficient Diversification of Investments* (Wiley, New York).

Menzly, Lior, Tano Santos, and Pietro Veronesi, 2004, Understanding predictability, *Journal of Political Economy* 112, 1–47.

Mitchell, Mark L., Todd Pulvino, and Erik Stafford, 2002, Limited arbitrage and equity markets, *Journal of Finance* 57, 551–584.

Neal, Robert, and Simon Wheatley, 1998, Do measures of investor sentiment predict stock returns, *Journal of Financial and Quantitative Analysis* 34, 523–547.

Ofek, Eli, and Matthew Richardson, 2002, DotCom mania: The rise and fall of internet stocks, *Journal of Finance* 58, 1113–1138.

Peng, Lin, and Wei Xiong, 2004, Limited attention and asset prices, Working paper, Princeton University.

Reinganum, Marc R., 1983, The anomalous stock market behavior of small firms in January: Empirical tests for tax-loss selling effects, *Journal of Financial Economics* 12, 89–104.

Ritter, Jay, 1984, The 'hot issue' market of 1980, *The Journal of Business* 57, 215–240.

Ritter, Jay, 1991, The long-run performance of initial public offerings, *Journal of Finance* 46, 3–27.

Ritter, Jay, 2003, Some factoids about the 2002 IPO market, Working paper, University of Florida.

Seyhun, H. Nejat, 1998, *Investment Intelligence from Insider Trading* (MIT Press, Cambridge).

Shiller, Robert J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71, 421–436.

Shiller, Robert J., 2000, *Irrational Exuberance* (Princeton UP, Princeton).

Shleifer, Andrei, 2000, *Inefficient Markets* (Oxford UP, Oxford).

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.

Siegel, Jeremy, 1998, *Stocks for the Long Run* (McGraw-Hill, New York).

Smith, E. E., E. J. Shoben, and L. J. Rips, 1974, Structure and process in semantic memory: A featural model for semantic decisions, *Psychological Review* 81, 214–241.

Speiss, D. Katherine, and John Affleck-Graves, 1995, Underperformance in long-run stock returns following seasoned equity offerings, *Journal of Financial Economics* 38, 243–267.

Spiess, Katherine, and John Affleck-Graves, 1999, The long-run performance of stock returns following debt offerings, *Journal of Financial Economics* 54, 45–73.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Stigler, George J., 1964, Public regulation of the securities markets, *Journal of Business* 37, 117–142.

Swaminathan, Bhaskaran, 1996, Time-varying expected small firm returns and closed-end fund discounts, *Review of Financial Studies* 9, 845–887.

Wachter, Jessica, 2000, Habit formation and the cross-section of asset returns, Chapter 4 of Ph.D. dissertation, Harvard University.

Wurgler, Jeffrey, and Katia Zhuravskaya, 2002, Does arbitrage flatten demand curves for stocks?, *Journal of Business* 75, 583–608.

Zweig, Martin E., 1973, An investor expectations stock price predictive model using closed-end fund premiums, *Journal of Finance* 28, 67–87.

The Sum of All FEARS Investor Sentiment and Asset Prices

Zhi Da
Mendoza College of Business, University of Notre Dame

Joseph Engelberg
Rady School of Management, University of California at San Diego

Pengjie Gao
Mendoza College of Business, University of Notre Dame

We use daily Internet search volume from millions of households to reveal market-level sentiment. By aggregating the volume of queries related to household concerns (e.g., "recession," "unemployment," and "bankruptcy"), we construct a Financial and Economic Attitudes Revealed by Search (FEARS) index as a new measure of investor sentiment. Between 2004 and 2011, we find FEARS (i) predict short-term return reversals, (ii) predict temporary increases in volatility, and (iii) predict mutual fund flows out of equity funds and into bond funds. Taken together, the results are broadly consistent with theories of investor sentiment. (*JEL* G10)

John Maynard Keynes (1936) argued that markets can fluctuate wildly under the influence of investors' "animal spirits," which move prices in a way unrelated to fundamentals. Fifty years later, De Long, Shleifer, Summers, and Waldmann (1990; DSSW hereafter) formalized the role of investor sentiment in financial markets. DSSW demonstrate that if uninformed noise traders base their trading decisions on sentiment and risk-averse arbitrageurs encounter limits to arbitrage, sentiment changes will lead to more noise trading, greater mispricing, and excess volatility. Although the survival of noise traders in the long run remains open for debate (e.g., Kogan, Ross, Wang and Westerfield 2006, 2009), there is a growing consensus that noise traders can induce large

---

We thank Nick Barberis, Shane Corwin, Jennifer Conrad, David Hirshleifer (the editor), Bob Jennings, Shimon Kogan, Ralph Koijen, Owen Lamont, Paul Schultz, Andrei Shleifer, Paul Tetlock, Dick Thaler, Jeff Wurgler, Yu Yuan, two anonymous referees and seminar participants at University of Notre Dame, State of Indiana Annual Conference, the 2010 NBER Behavioral Economics meeting and the WFA 2010 meeting for helpful comments and discussions. We are grateful to Conrad Gann at TrimTabs, Inc. for his assistance with the daily mutual fund data used in this study, Shane Corwin for providing us with the intraday data for realized volatility calculation during the recent years, and Pavel Savor for sharing with us macro announcement dates. Jianfeng Zhu provided valuable research assistance. We are responsible for remaining errors. Send correspondence to Joseph Engelberg, Finance Department, UCSD Rady School of Management, 9500 Gilman Drive #0553, La Jolla, CA 92093-0553; telephone: 858-822-7912. E-mail: jengelberg@ucsd.edu.

doi:10.1093/rfs/hhu072 Advance Access publication October 17, 2014

price movements and excess volatility in the short run.[1] As Baker and Wurgler (2007) put it in their survey article: “Now, the question is no longer, as it was a few decades ago, whether investor sentiment affects stock prices, but rather how to measure investor sentiment and quantify its effects.”

In this paper we propose a possible answer: investor sentiment can be directly measured through the Internet search behavior of households. We aggregate the volume of Internet search queries such as “recession,” “bankruptcy,” and “unemployment” from millions of U.S. households to construct a Financial and Economic Attitudes Revealed by Search (FEARS) index. We then quantify the effects of FEARS on asset prices, volatility, and fund flows. We find that FEARS predict return reversals: although increases in FEARS correspond with low market-level returns today, they predict high returns (reversal) over the next few days. Moreover, increases in FEARS coincide with only temporary increases in market volatility and predict mutual fund flow out of equity funds and into bond funds. Such trading patterns and price reversals can also come from liquidity shocks, as modeled in Campbell, Grossman and Wang (1993; CGW hereafter). In this case, high-frequency investor sentiment, as measured by FEARS, turns out to be a powerful trigger of liquidity shocks that affect prices.

The appeal of our search-based sentiment measure is more transparent when compared with alternatives. Traditionally, empiricists have taken two approaches to measuring investor sentiment. Under the first approach, empiricists proxy for investor sentiment with market-based measures such as trading volume, closed-end fund discount, initial public offering (IPO) first-day returns, IPO volume, option implied volatilities (VIX), or mutual fund flows (see Baker and Wurgler (2007) for a comprehensive survey of the literature). Although market-based measures have the advantage of being readily available at a relatively high frequency, they have the disadvantage of being the equilibrium outcome of many economic forces other than investor sentiment. Qiu and Welch (2006) put it succinctly: “How does one test a theory that is about inputs → outputs with an output measure?”

Under the second approach, empiricists use survey-based indices such as the University of Michigan Consumer Sentiment Index, the UBS/GALLUP Index for Investor Optimism, or investment newsletters (Brown and Cliff (2005), Lemmon and Portniaguina (2006), and Qiu and Welch (2006)). Compared with survey-based measures of investor sentiment, the search-based sentiment measure we propose has several advantages. First, search-based sentiment measures are available at a high frequency.[2] Survey measures are often available

---

[1] A particularly interesting thread of this literature examines sentiment following non-economic events such as sports (Edmans, Garcia, and Norli 2007), aviation disasters (Kaplanski and Levy 2010), weather conditions (Hirshleifer and Shumway 2003), and seasonal affective disorder (SAD; Kamstra, Kramer, and Levi 2003), and shows these sentiment-changing events cause changes in asset prices.

[2] To date, high-frequency analysis of investor sentiment is found only in laboratory settings. For example, Bloomfield, O’Hara, and Saar (2009) use laboratory experiments to investigate the impact of uninformed traders on underlying asset prices.

monthly or quarterly. In fact, we find that our daily FEARS index can predict monthly survey results of consumer confidence and investor sentiment. Second, search-based measures *reveal* attitudes rather than inquire about them. Although many people answer survey questions for altruistic reasons, there is often little incentive to answer survey questions carefully or truthfully, especially when questions are sensitive (Singer 2002). Search volume has the potential to reveal more personal information where non-response rates in surveys are particularly high or the incentive for truth-telling is low. For example, eliciting the likelihood of job loss via survey may be a sensitive topic for a respondent. On the other hand, aggregate search volume for terms like "find a job," "job search," or "unemployment" reveals concern about job loss. Finally, some economists have been skeptical about answers in survey data that are not "cross-verif(ied) with data on actual (not self-reported) behavior observed by objective external measurement" (Lamont, quoted in Vissing-Jorgensen 2003). Search behavior is an example of such objective, external verification.

Google, the largest search engine in the world, makes public the Search Volume Index (SVI) of search terms via its product Google Trends (http://www.google.com/trends/).[3] When a user inputs a search term into Google Trends, the application returns the search volume history for that term scaled by the time-series maximum (a scalar). As an example, Figure 1 plots the SVI for the terms "recession" and "bankruptcy," respectively. The plots conform with intuition. For example, the SVI for "recession" began rising in the middle of 2007 and then increased dramatically beginning in 2008. All of this was well before the National Bureau of Economic Research (NBER) announced in December 2008 that the United States had been in a recession since December of 2007. The SVI for "bankruptcy" peaks once during 2005 and once again during 2009 before falling off. According to the American Bankruptcy Institute, actual bankruptcies in the United States follow a similar pattern with peaks in 2005 and 2009/2010.[4]

At the monthly frequency, SVI correlates well with alternative measures of market sentiment. For example, Figure 2 plots the monthly log SVI for "recession" (with a minus sign because higher SVI on "recession" signals pessimism) against the monthly University of Michigan Consumer Sentiment Index (MCSI), which asks households about their economic outlook. During our sample period from January 2004 to December 2011, the two time series are highly correlated with a correlation coefficient of 0.858. When we use the log change in "recession" SVI this month to predict next month's log change in the MCSI, we find that an increase in SVI predicts a decrease in the MCSI (t-value = 2.56). This predictive result suggests that SVI, revealing household

[3] By February 2009, Google accounted for 72.11% of all search queries performed in the United States, according to Hitwise, which specializes in tracking Internet traffic.

[4] See http://www.abiworld.org/AM/AMTemplate.cfm?Section=Home&CONTENTID=65139&TEMPLATE=/CM/ContentDisplay.cfm.

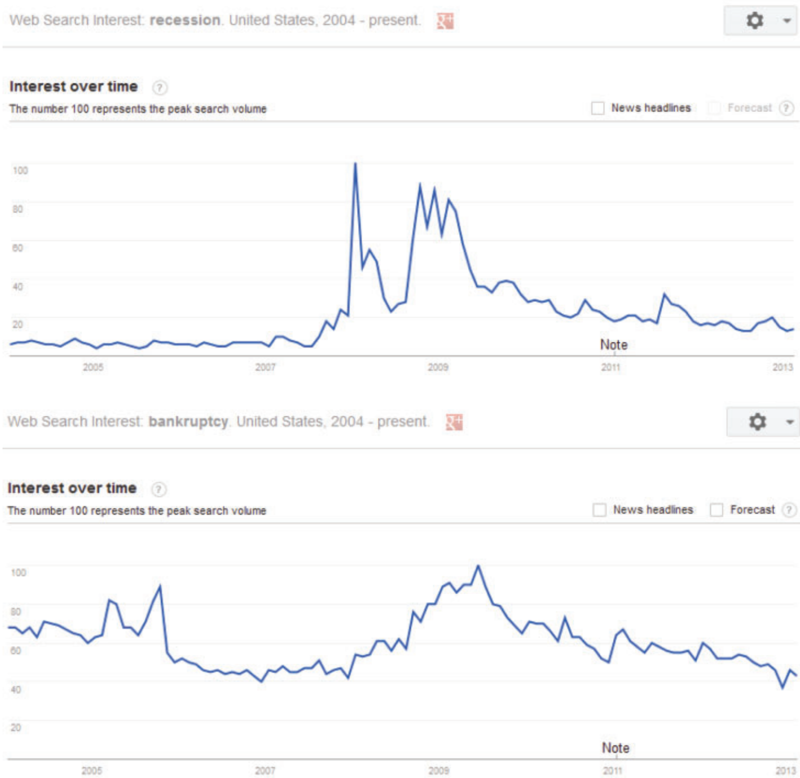


Figure 1
Illustrations of Google search volume
The figures represent the graphical output of weekly aggregate search frequency (SVI) from Google Trends (http://www.google.com/trends/). The top (bottom) panel plots weekly SVI for "recession" ("bankruptcy") in the United States. Plotted SVI is weekly search volume scaled by the maximum over the time period.

sentiment at a high frequency, leads survey-based sentiment measures by at least a month.

The key to the construction of our FEARS index is the identification of relevant sentiment-revealing search terms. To identify search terms in a way that is as objective as possible, we begin with well-known dictionaries in the finance and textual analytics literature (Tetlock 2007) and select the set of words classified as "economic" words with either "positive" or "negative" sentiment. This provides us with a list of 149 words such as "crisis," "gold," "inflation," "recession," and "security." Second, we download the associated top ten related search terms (provided by Google) in order to see how these economic words are used by search engine users in practice. Finally, we eliminate non-economic search terms and search terms with too few valid SVIs. This procedure results in a list of 118 search terms, for which we calculate daily log differences. To

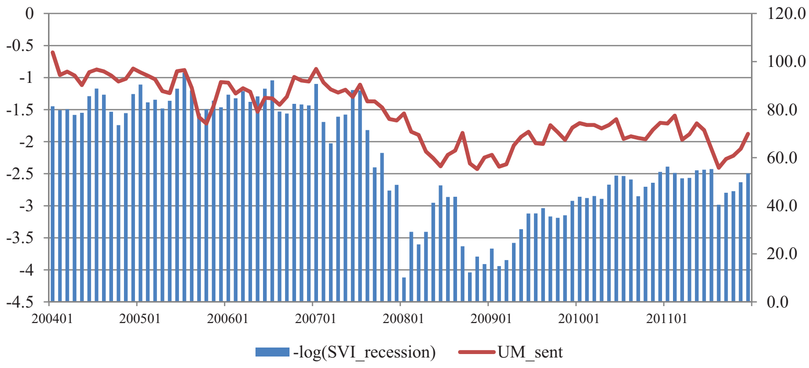


Figure 2
Search for "recession" and consumer confidence
We plot the monthly log SVI for "recession" (with a minus sign) against the monthly University of Michigan Consumer Sentiment Index. The data are from January 2004 to December 2011. The correlation between the two series is 0.858.

make these 118 terms comparable, we winsorize, remove intra-week and intra-year seasonality, and standardize each time series (as in Baker and Wurgler 2006). Finally, we run backward-looking rolling regressions to let the data tell us which of the 118 terms are most important. For example, when thinking about our FEARS list in January 2011, we run a regression to determine the historical relationship between search and contemporaneous market return for all of our search terms during the period between January 1, 2004, and December 31, 2010. Only the search terms that have historically been related to returns (through December 31, 2010) are used for our FEARS list beginning in January 2011. This procedure produces a dynamic list of thirty search terms whose search volume changes are then averaged to produce our FEARS index.

We then relate our FEARS index to asset prices. In Section 2, we find a negative contemporaneous correlation between FEARS and stock market returns. Increases (decreases) in FEARS correspond with low (high) returns. However, in the days following, this relationship reverses. Increases in FEARS today predict increases in stock market returns in the following two days, which is consistent with sentiment-induced temporary mispricing. Moreover, this reversal is strongest among stocks with higher beta, higher volatility, and greater downside risk, consistent with the predictions in Baker and Wurgler (2006, 2007). We find similar spike-reversal patterns among other asset classes. For example, among Treasury bonds, we find a positive contemporaneous correlation between FEARS (i.e., increases in FEARS correspond with high Treasury bond returns) consistent with the notion of flight-to-safety. Again, this relationship reverses in the following days.

In Section 3 we consider the prediction that high-frequency sentiment swings will generate excess volatility in the short term. We find a significant positive contemporaneous correlation between our FEARS index and daily market volatility measured as either realized volatility on the S&P 500 exchange traded fund (ETF) return or the Chicago Board of Exchange (CBOE) market volatility index (VIX). As volatility displays seasonal patterns and is well known to be persistent and long-lived (Engel and Patton 2001; Andersen, Bollerslev, Diebold and Ebens 2001; Andersen, Bollerslev, Diebold, and Labys 2003), we account for this long-range dependence through the fractional integrated autoregressive moving average (ARFIMA) model, $ARFIMA(1,d,1)$. In addition, parallel to our earlier analysis, we also examine the daily returns on a tradable volatility-based asset, the CBOE VIX futures contract. When we relate our FEARS index to these daily VIX futures returns, we first confirm the strong contemporaneous correlation between our FEARS index and VIX futures returns. As before, we find that our FEARS index predicts a reversal in VIX futures returns during the next two trading days.

As a more direct test of the "noise trading" hypothesis, we examine daily mutual fund flows in Section 4. Because individual investors hold about 90% of total mutual fund assets and they are more likely to be "noise" traders, daily flows to mutual fund groups likely aggregate "noise" trading at the asset class level.[5] We examine two groups of mutual funds that specialize in equity and intermediate Treasury bonds. We document strong persistence in fund flows and again use the $ARFIMA$ model to extract daily innovations to these fund flows. Our results suggest significant outflow from the equity market one day after an increase in FEARS. We also observe a significant inflow to bond funds one day after a significant withdrawal from equity funds. Taken together, the evidence indicates a "flight to safety," with investors shifting their investments from equities to bonds after a spike in FEARS.

1. Data and Methodology

Although the data for this study come from a variety of sources, we begin by discussing the construction of our FEARS index, which is the main variable in our analysis.

1.1 Construction of FEARS index

Our objective is to build a list of search terms that reveal sentiment toward economic conditions. We follow the recent text analytics literature in finance, which uses the Harvard IV-4 Dictionary and the Lasswell Value Dictionary (Tetlock 2007; Tetlock, Saar-Tsechansky, and Macskassy 2008). These dictionaries place words into various categories such as "positive," "negative," "weak," "strong," and so on. Because we are interested in household sentiment toward the economy, we select the set of words that are "economic"

---

[5] Source: *2007 Investment Company Fact Book* by the Investment Company Institute.

words that also have either “positive” or “negative” sentiment.[6] This results in 149 words such as “bankruptcy,” “crisis,” “gold,” “inflation,” “recession,” “valuable,” and “security.”

We call this list the “primitive” word list. Our next task is to understand how these words might be searched in Google by households. To do this, we input each primitive word into Google Trends, which, among other things, returns ten “top searches” related to each primitive word.[7] For example, a search for “deficit” results in the related searches “budget deficit,” “attention deficit,” “attention deficit disorder,” “trade deficit,” and “federal deficit” because this is how the term “deficit” is commonly searched in Google. Our 149 primitive words generate 1,490 related terms, which become 1,245 terms after removing duplicates.

Next we remove terms with insufficient data. Of our 1,245 terms, only 622 have at least 1,000 observations of daily data.[8] Finally, we remove terms that are not clearly related to economics or finance. For example, a search for “depression” results in the related searches “the depression,” “great depression,” “the great depression,” “depression symptoms,” “postpartum depression,” “depression signs,” etc. We keep the first three terms (which relate to an economic depression) and remove the last three terms (which relate to the mental disorder of depression). This leaves us with 118 search terms.

We download the SVI for each of these 118 terms over our sample period of January 2004 to December 2011 from Google Trends. Google Trends allows users to restrict SVI results to specific countries (e.g., search volume for “recession” from British households). Because most of the dependent variables of interest in this paper are related to U.S. indices, we restrict the SVI results to the United States. Thus, the measures we construct represent the sentiment of American households. W define the daily change in search term j as:

$$\Delta SVI_{j,t} = \ln\left(SVI_{j,t}\right) - \ln(SVI_{j,t-1}). \tag{1}$$

Figure 3 plots the daily log changes for two terms, “Inflation” and “Price of Gold,” during two different quarters in 2006. The figures demonstrate several important features of the search data. The first is seasonality: SVI change rises during the beginning of the week (e.g., Monday and Tuesday)

[6] Specifically, from http://www.wjh.harvard.edu/˜inquirer/spreadsheet_guide.htm we take all economic words (those with tags “Econ@” or “ECON”) which also have a positive or negative sentiment tag (those with tags “Ngtv,” “Negativ,” “Positiv,” or “Pstv”).

[7] According to Google, “Top searches refers to search terms with the most significant level of interest. These terms are related to the term you’ve entered. . . . Our system determines relativity by examining searches that have been conducted by a large group of users preceding the search term you’ve entered, as well as after.”

[8] To increase the response speed, Google currently calculates SVI from a random subset of the actual historical search data. This is why SVIs on the same search term might be slightly different when they are downloaded at different points in time. We believe that the impact of such sampling error is small for our study and should bias against finding significant results. As in Da, Gao, and Engelberg (2011), when we download the SVIs several times and compute their correlation, we find that the correlations are usually above 97%.

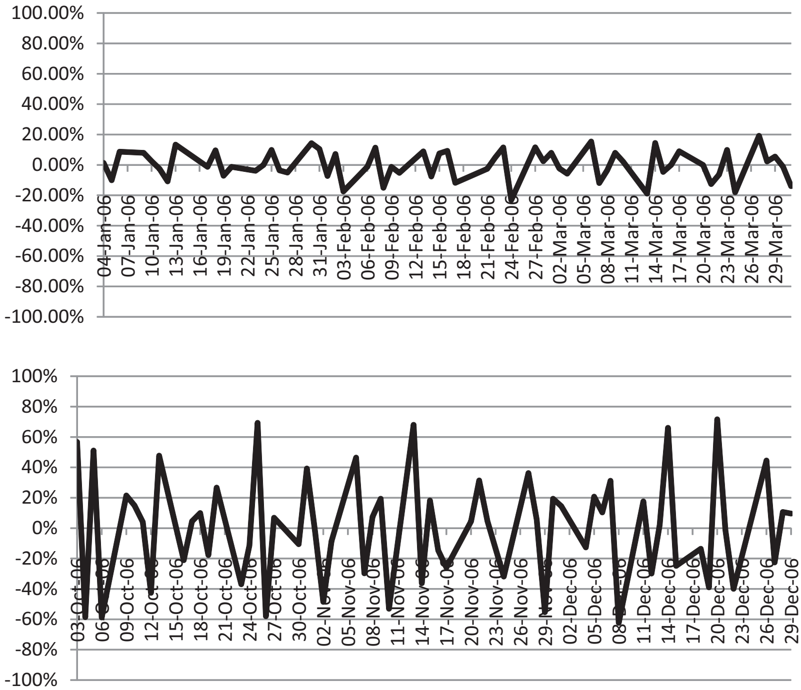


Figure 3
SVI change examples for "inflation" and "price of gold"
We plot two examples of daily changes in SVI. The first is for the term "Inflation" over the period January 2006–March 2006 plotted in the top panel. The second is for the term "Price of Gold" over the period October 2006–December 2006 plotted in the bottom panel.

and falls throughout the week, which generates the repeated five-day hump-shaped pattern depicted in Figure 3. Moreover, there is considerable variance differences across terms. SVI change for "Inflation" and "Price of Gold" are plotted on the same scale so that the heteroscedasticity is apparent. In fact, the standard deviation of SVI change for "Price of Gold" is nearly three times greater than that of "Inflation." Finally, the SVI change for "Price of Gold" indicates the presence of some extreme values. To mitigate any concerns about outliers and to address the issues of seasonality and heteroscedasticity in the data, we adjust the raw data in the following way. First, we winsorize each series at the 5% level (2.5% in each tail). Then, to eliminate seasonality from $\Delta SVI_{j,t}$, we regress $\Delta SVI_{j,t}$ on weekday dummies and month dummies and keep the residual. Finally, to address heteroscedasticity and make each times series comparable, we standardize each of the time series by scaling each by the time-series standard deviation as in Baker and Wurgler (2006). This leaves us with an adjusted (winsorized, deseasonalized, and standardized) daily change in search volume, $\Delta ASVI_t$, for each of our 118 terms.

Our final step is to let the data identify search terms that are most important for returns. To do this we run expanding backward rolling regressions of $\Delta ASVI$ on market returns every six months (every June and December) to determine the historical relationship between search and contemporaneous market return for all 118 of our search terms. When we do this it becomes clear that, given a search term that has a strong relationship with the market, the relationship is almost always negative. This is despite the fact that we began with economic words of both positive and negative sentiment when selecting words from the Harvard and Lasswell dictionaries. For example, when we use all 118 terms in the full sample (January 2004–December 2011) we find zero terms with a t-statistic on $\Delta ASVI$ above 2.5 but fourteen terms with a t-statistic below -2.5. These terms include "recession," "great depression," "gold price," and "crisis." As in Tetlock (2007) it appears that negative terms in English language are most useful for identifying sentiment. For this reason, we use only the terms that have the largest negative t-statistic on $\Delta ASVI$ to form our FEARS index. Formally, we define FEARS on day t as:

$$FEARS_t = \sum_{i=1}^{30} R^i(\Delta ASVI_t) \tag{2}$$

where $R^i(\Delta ASVI_t)$ is the $\Delta ASVI_t$ for the search term that had a t-statistic rank of i from the period January 2004 through the most recent six-month period, where ranks run from smallest ($i=1$) to largest ($i=118$). For example, at the end of June 2009, we run a regression of $\Delta ASVI$ on contemporaneous market return during the period January 1, 2004–June 30, 2009, for each of our 118 search terms. Then we rank the t-statistic on $\Delta ASVI$ from this regression from most negative ($i=1$) to most positive ($i=118$). We select the thirty most negative terms and use these terms to form our FEARS index for the period from July 1, 2009, to December 31, 2009. $FEARS$ on day t during this period is simply the average $\Delta ASVI$ of these thirty terms on day t. Given our relatively short sample period, we choose an expanding rolling window to maximize the statistical power of the selection. We choose a cutoff of thrity as it is often considered to be the minimum number of observations needed to diversify away idiosyncratic noise. Robustness to alternative cutoff choices (e.g., top twenty-five or top thirty-five) is shown in Table 5. Finally, due to the need for an initial window of at least six months, our FEARS index starts in July 2004.

There are several advantages to this historical, regression-based approach for selecting terms. First, using historical regressions to identify the most relevant terms is an objective way to "let the data speak for itself." Kogan et al. (2009) also take a similar regression approach to identify relevant words in firm 10-Ks and argue this approach not only helps the researcher identify terms that were not ex ante obvious but also is an objective way to select terms. This is also true in our case. For example, the word "gold" is considered an economic word of positive sentiment by the Harvard dictionary, and yet we find a strong negative

Table 1
FEARS terms from the full sample

	Search Term	T-Statistic
1	GOLD PRICES	−6.04
2	RECESSION	−5.60
3	GOLD PRICE	−4.81
4	DEPRESSION	−4.56
5	GREAT DEPRESSION	−4.15
6	GOLD	−3.98
7	ECONOMY	−3.52
8	PRICE OF GOLD	−3.23
9	THE DEPRESSION	−3.20
10	CRISIS	−2.93
11	FRUGAL	−2.87
12	GDP	−2.85
13	CHARITY	−2.63
14	BANKRUPTCY	−2.50
15	UNEMPLOYMENT	−2.46
16	INFLATION RATE	−2.32
17	BANKRUPT	−2.28
18	THE GREAT DEPRESSION	−2.17
19	CAR DONATE	−2.11
20	CAPITALIZATION	−2.10
21	EXPENSE	−1.97
22	DONATION	−1.89
23	SAVINGS	−1.82
24	SOCIAL SECURITY CARD	−1.71
25	THE CRISIS	−1.65
26	DEFAULT	−1.63
27	BENEFITS	−1.56
28	UNEMPLOYED	−1.55
29	POVERTY	−1.52
30	SOCIAL SECURITY OFFICE	−1.51

This table reports the 30 search terms derived from words of economic sentiment in the Harvard and Lasswell dictionaries (see the description in Section 1.1) that have had the largest negative correlation with the market. The terms are ordered from most negative (GOLD PRICES) to least negative (SOCIAL SECURITY OFFICE).

relationship between searches for "gold" and market returns, consistent with the evidence in Baur and Lucey (2010), who argue that gold represents a "safe haven" in times of distress, at least in view of retail investors who are most likely to be affected by sentiment. This only came to light given our data-driven approach for constructing the FEARS index.

Table 1 displays the top thirty terms over our entire sample (January 2004–December 2011). The terms that historically have the largest daily correlation with the market include "gold prices" (t-statistic = −6.04), "recession" (t-statistic = −5.60), "gold price" (t-statistic = −4.81), "depression" (t-statistic = −4.56), and "great depression" (t-statistic = −4.15).

1.2 Other data

Most of our empirical tests are carried out at the aggregate market or index level. Daily indices are either taken directly from CRSP or calculated from the individual stock prices and returns in the CRSP daily stock file. To ensure that illiquid index component stocks are not driving our results, we also examine four highly liquid index exchange-traded funds (ETFs): the SPDR S&P 500 (NYSEARCA: SPY), the PowerShares QQQ Trust (NASDAQ: QQQQ), the

Russell 1000 Index ETF (NYSEARCA: IWB), and the Russell 2000 Index ETF (NYSEARCA: IWM). We also obtain intraday data on SPY from TAQ in order to estimate realized market volatility. Finally, we obtain Treasury portfolio returns from the CRSP ten-year constant maturity Treasury file.

The Chicago Board Options Exchange (CBOE) daily market volatility index (VIX), which measures the implied volatility of options on the S&P 100 stock index, is well known as an "investor fear gauge" by practitioners. For example, Whaley (2001) discusses the spikes in the VIX series since its 1986 inception, which captures the crash of October 1987 and the 1998 Long Term Capital Management crisis. Baker and Wurgler (2007) consider it to be an alternative market sentiment measure. We include the VIX index as a control variable in most specifications. Later we use our FEARS index to predict VIX, as well as returns from VIX futures traded on the CBOE.

We obtain a high-frequency measure of concurrent macroeconomic conditions from the Federal Reserve Bank of Philadelphia.[9] Using a dynamic factor model to extract the latent state of macroeconomic activity from a large number of macroeconomic variables, Aruoba, Diebold, and Scotti (2009) construct a daily measure of macroeconomic activities (the "ADS" index). According to the Federal Reserve Bank of Philadelphia, construction of the ADS index includes a battery of seasonally adjusted macroeconomic variables of mixed frequencies: weekly initial jobless claims; monthly payroll employment, industrial production, personal income less transfer payments, manufacturing and trade sales; and quarterly real gross domestic product (GDP). The change in the ADS index reflects innovations driven by macroeconomic conditions. An increase in the ADS index indicates progressively better-than-average conditions, while a decrease in the ADS index indicates progressively worse-than-average conditions. We also obtain the dates of important macroeconomic announcements about consumer price index (CPI), producer price index (PPI), unemployment rates, or interest rates, as in Savor and Wilson (2013), for our sample period.

To capture uncertainty related to economic policies, we adopt a news-based measure of economic policy uncertainty (EPU) recently developed by Baker, Bloom, and Davis (2013).[10] This measure is constructed by counting the number of U.S. newspaper articles achieved by the NewsBank Access World News database with at least one term from each of the following three categories of terms: (i) "economic" or "economy"; (ii) "uncertain" or "uncertainty"; and (iii) "legislation," "deficit," "regulation," "congress," "Federal Reserve," or "White House." Baker, Bloom, and Davis (2013) provide evidence that the news-based measure of EPU seems to capture perceived economic policy uncertainty.

---

[9] The data are available at http://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index.

[10] The data are available at http://www.policyuncertainty.com/us_daily.html.

In a robustness table we also use a measure of news-based sentiment. Our news-based sentiment measure is the fraction of negative words in the *Wall Street Journal* "Abreast of the Market" column as in Tetlock (2007). To identify negative words, we follow Tetlock (2007) and use the dictionaries from the General Inquirer program. Loughran and McDonald (2011) argue that some negative words in these dictionaries do not have a truly negative meaning in the context of financial markets. For example, words like "tax," "cost," "vice," and "liability" simply describe company operations. Instead, they develop an alternative negative word list that better reflects the tone of financial text. We obtain qualitatively similar results when using either word list.

Our daily mutual fund flow data are obtained from TrimTabs, Inc. A description of TrimTabs data can be found in Edelen and Warner (2002) and Greene and Hodges (2002). TrimTabs collects daily flow information for about 1,000 distinct mutual funds that represent approximately 20% of the universe of U.S.-based mutual funds according to Greene and Hodges (2002). TrimTabs aggregates the daily flows for groups of mutual funds categorized using fund objectives from Morningstar. For our study, we focus on the daily flow of two groups of mutual funds. The first group (Equity) specializes in equity. The second group (MTB) specializes in "intermediate Treasury bonds." For each group, we compute the daily flow as the ratio between dollar flow (inflow minus outflow) and fund total net assets (TNA). The data we received from TrimTabs covers the five-year period from July 2004 to October 2009.

2. FEARS and Asset Returns

We first examine the relationship between FEARS and returns across various asset classes. We then examine how this relationship varies among the cross-section of stocks when we consider limits to arbitrage.

2.1 FEARS and average returns

One salient feature of sentiment theories is the heterogeneity of investors. In sentiment models, there is typically one class of investors who suffer from a bias, such as extrapolative expectations about future cash flows. These biases lead investors to make demands for assets that are not reflected by fundamentals and, in the presence of limits to arbitrage, push prices away from fundamental values. Thus, a central prediction of theories of investor sentiment is reversal. For example, when sentiment is high, prices are temporarily high but later become low.

We look for evidence of return reversals by running the following regressions:

$$return_{i,t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m \gamma_m Control^m_{i,t} + u_{i,t+k}. \quad (3)$$

In regression (3), $return_{i,t+k}$ denotes asset i's return on day $t+k$. We also consider two-day cumulative returns, $return_{i,[t+1,t+2]}$, to gain a perspective on the cumulative effects of return reversals. Control variables ($Control^m_{i,t}$) include lagged asset-class returns (up to five lags), changes in a news-based measure of economic policy uncertainty (EPU), the CBOE volatility index (VIX), and changes in the Aruoba-Diebold-Scotti (ADS) business conditions index.[11] We calculate bootstrapped standard errors, and our statistical inference is conservative.[12]

In Table 2, we examine the Standard and Poor's 500 index. When $k=0$, the negative and significant coefficient on $FEARS_t$ suggests a negative contemporaneous relationship between FEARS and a broad equity index. Days in which there were sharp declines in the equity indices there were also sharp increases in search for terms like "recession," "gold price," "depression," and so on. For example, the first column of Table 2 shows that a standard-deviation increase in FEARS corresponds with a contemporaneous decline of 19 basis points for the daily S&P 500 index, after controlling for lagged returns, contemporaneous VIX, EPU, and ADS.[13] This result is perhaps unsurprising. Recall that the search terms that compose the FEARS index were selected based on their historical correlation with the market. Table 2 suggests that they continue to be correlated out of sample.

Much of the day 0 effect, however, is temporary. In the ensuing days, the positive and significant coefficient on $FEARS$ suggests that increases in FEARS predict higher returns. As evident in columns 2 to 4, these reversals are significant on both the first and the second days ($k=1$ and 2).[14] Specifically, a standard-deviation increase in FEARS predicts an increase of 7.1 basis points in the S&P 500 at $k=1$ (significant at the 5% level), and an *increase* of 7.3 basis points at $k=2$ (significant at the 10% level). The cumulative impact of a standard-deviation increase in FEARS predicts a cumulative increase of 14.4 basis points in the S&P 500 over days 1 and 2 (significant at the 1% level). In other words, the initial impact of FEARS on the S&P 500 index on day 0 is almost completely reversed after two days. In Table 2, we also consider longer horizons, ranging from $k=3$ to $k=5$, but none of the coefficients on FEARS are statistically significant and point estimates are economically negligible,

11 We also find that replacing the VIX index with an increasingly popular alternative sentiment index, the Credit Suisse Fear Barometer (CSFB), has little effect on the results.

12 For all the empirical results reported in the paper, we have also computed standard errors that are robust to heteroscedasticity and autocorrelations. These unreported standard errors imply even higher t-values in general, and thus only strengthen our conclusions.

13 A one-standard-deviation change in the FEARS index corresponds to 0.3549. Recall that while each individual search term has been standardized so that its standard deviation is one by construction, the average across search terms will not have a standard deviation of one given correlation among search terms.

14 Note that search volume and returns are measured over different intervals. Daily search volume is measured over the interval 00:00–24:00 PST, while returns are measured over the interval 13:00 PST–13:00 PST. Therefore, the return on day t+1 overlaps with some search volume on day t. If FEARS measured after hours on day t spilled into day t+1 return, we would expect a negative coefficient in column 2. We do not find one, which suggests that the effect from this mismatch in measurement of intervals is small. Moreover, FEARS on day t predict returns on day t+2 where there is no overlap of measurement intervals.

Table 2
FEARS and S&P 500 index returns

	(1) Ret(t)	(2) Ret(t+1)	(3) Ret(t+2)	(4) Ret[t+1, t+2]	(5) Ret(t+3)	(6) Ret(t+4)	(7) Ret(t+5)
FEARS	−0.00532***	0.00200**	0.00207*	0.00409***	−0.000620	−0.000800	0.00104
	(0.00130)	(0.000966)	(0.00113)	(0.00137)	(0.000937)	(0.000943)	(0.00100)
VIX	−0.000187***	1.80e–05	1.50e–07	1.43e–05	–6.07e–06	–7.02e–06	–8.61e–06
	(6.31e–05)	(6.38e–05)	(6.36e–05)	(8.29e–05)	(6.28e–05)	(6.41e–05)	(6.04e–05)
EPU	4.73e–06	–1.33e–05*	1.20e–05	–1.42e–06	8.69e–06	–8.94e–06	2.68e–06
	(7.11e–06)	(7.45e–06)	(7.65e–06)	(9.27e–06)	(7.70e–06)	(7.92e–06)	(6.78e–06)
ADS	−0.0253	−0.0208	−0.0194	−0.0394	−0.0168	−0.0174	−0.0164
	(0.0298)	(0.0310)	(0.0315)	(0.0439)	(0.0319)	(0.0341)	(0.0361)
Ret(t)		−0.121***	−0.0600	−0.179***	0.0365	−0.0252	−0.0484
		(0.0376)	(0.0521)	(0.0595)	(0.0418)	(0.0495)	(0.0488)
Ret(t–1)	−0.155***	−0.0780	0.0370	−0.0424	−0.0165	−0.0597	0.0110
	(0.0378)	(0.0546)	(0.0403)	(0.0600)	(0.0488)	(0.0521)	(0.0484)
Ret(t–2)	−0.0896*	0.0167	−0.0210	−0.00436	−0.0496	0.00443	−0.0365
	(0.0541)	(0.0408)	(0.0465)	(0.0565)	(0.0485)	(0.0458)	(0.0500)
Ret(t–3)	0.00358	−0.0163	−0.0537	−0.0679	0.00338	−0.0323	0.0134
	(0.0394)	(0.0480)	(0.0481)	(0.0638)	(0.0497)	(0.0473)	(0.0425)
Ret(t–4)	−0.0318	−0.0507	0.00377	−0.0482	−0.0308	0.0102	−0.0109
	(0.0473)	(0.0474)	(0.0452)	(0.0564)	(0.0496)	(0.0426)	(0.0496)
Ret(t–5)	−0.0532	−0.00361	−0.0369	−0.0368	0.0184	−0.0182	0.0394
	(0.0462)	(0.0446)	(0.0503)	(0.0712)	(0.0426)	(0.0496)	(0.0458)
Constant	0.00424***	−0.000170	0.000167	5.48e–05	0.000299	0.000351	0.000361
	(0.00116)	(0.00117)	(0.00116)	(0.00153)	(0.00116)	(0.00119)	(0.00112)
Observations	1,891	1,890	1,889	1,889	1,888	1,887	1,886
Adjusted R^2	0.060	0.027	0.011	0.027	0.003	0.002	0.002

This table relates S&P 500 index daily returns to FEARS. The dependent variables are contemporaneous returns (column (1)), future S&P 500 index daily returns in the next five days (columns (2), (4), (5), (6), and (7), respectively), and future S&P 500 index return over the first two days (column (5)). The independent variable is the FEARS index. The set of control variables include lagged returns up to five lags, changes in a news-based measure of economic policy uncertainty (EPU), the CBOE volatility index (VIX), and changes in the Aruoba-Diebold-Scotti (ADS) business conditions index. The standard errors are bootstrapped standard errors. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

confirming that the effect of FEARS on asset prices operates mainly during the first three days. We also verify that this is true for other asset classes, and for this reason, we do not report the results for $k > 2$ in other tables.

Table 3 reports results using different test assets. Panels A and B focus on different equity portfolios, while Panel C focuses on Treasury securities. The test assets are the CRSP value weighted and equally weighted portfolios (Panel A), equity exchange-traded funds (Panel B), and the CRSP ten-year constant maturity Treasury portfolio (Panel C). The equity ETFs include the S&P 500 index ETF (SPY), the NASDAQ 100 ETF (QQQQ), the Russell 1000 Index ETF (IWB), and the Russell 2000 Index ETF (IWM). Across all assets, a contemporaneous increase in FEARS is always associated with a contemporaneous decrease of equity returns, and a contemporaneous increase of Treasury security returns. Moreover, an increase in FEARS today (i.e., $k=0$) always predicts a return reversal in the coming two days (i.e., $k=1$ and $k=2$). The effect of FEARS on equities is typically larger in both initial and future returns compared with Treasury securities. A standard-deviation increase in

FEARS corresponds with a contemporaneous decrease of 18 to 19 basis points among equities at $k=0$ (significant at the 1% level), and a reversal of 14 to 15 basis points during the next two days ($k=1$ and 2, significant at the 1%–5% level). In contrast, a standard-deviation increase in FEARS corresponds with a contemporaneous increase of 4 basis points for Treasury securities at $k=0$ (significant at the 1% level), and a complete reversal over the next two days (significant at the 1% level). Also, because the portfolios include more small stocks in Panel B (from the S&P 500 index, to the Russell 1000 Index, and then to the Russell 2000 Index) we observe a stronger reversal effect associated with our FEARS index.

Of course, such a short-term reversal can also be caused by a liquidity shock as in Campbell, Grossman and Wang (1993; GSW hereafter). As Baker and Stein (2004) point out, as sentiment and liquidity are intertwined, the difference between a sentiment-based story as in DSSW and a liquidity-based story as in GSW boils down how we view liquidity shocks and noise traders. Tetlock (2007) even goes so far as to say that "the difference between DSSW and CGW is philosophical rather than economic." Our results remain interesting even under the liquidity interpretation, as they suggest high-frequency investor sentiment, as measured by our FEARS, can be a powerful trigger of a liquidity shock.

Overall, Tables 2 and 3 illustrate that our FEARS index is strongly associated with contemporaneous returns and predicts future short-term return reversals.

2.2 FEARS and limits to arbitrage

As highlighted in Baker and Wurgler (2006, 2007), there are several additional channels that can exacerbate the effect of sentiment investors on asset prices. Perhaps the most important channel is limits to arbitrage (Pontiff 1996, Shleifer and Vishny 1997). Arbitrage capital moves slowly to take advantage of the irrational beliefs of sentiment investors. Motivated by limits to arbitrage, we consider several additional testing assets in order to explore the effect of sentiment on asset prices.

The first set of testing assets is the return spread from beta-sorted portfolios obtained from CRSP. CRSP computes a Scholes-Williams (1977) beta for common stocks traded on NYSE and AMEX using daily returns within a year and then forms decile portfolios. We take these beta-sorted decile portfolios, and compute the return spread between high beta stocks and low beta stocks.

According to Baker, Bradley, and Wurgler (2011), high-beta portfolios are prone to the speculative trading of sentiment investors. Moreover, high-beta stocks may be unattractive to arbitrageurs who face institutional constraints such as benchmarking. Because these two forces work in the same direction for high-beta stocks, it is natural to conjecture that investor sentiment may have a larger impact among high-beta stocks than among low-beta stocks. Thus, the return spreads between high-beta and low-beta stock portfolios should be negatively correlated with a contemporaneous increase in FEARS, while

Table 3
FEARS and returns to other asset classes

Panel A: FEARS and CRSP equally weighted and value-weighted index returns

	CRSP EW Index Returns				CRSP VW Index Returns			
	(1) Ret(t)	(2) Ret(t+1)	(3) Ret(t+2)	(4) Ret[t+1,t+2]	(5) Ret(t)	(6) Ret(t+1)	(7) Ret(t+2)	(8) Ret[t+1,t+2]
FEARS	–0.00519***	0.00195**	0.00185*	0.00381***	–0.00526***	0.00211**	0.00201*	0.00413***
	(0.00128)	(0.000949)	(0.000999)	(0.00133)	(0.00130)	(0.000981)	(0.00113)	(0.00139)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1,891	1,890	1,889	1,889	1,891	1,890	1,889	1,889
Adjusted R^2	0.035	0.005	0.005	0.006	0.052	0.018	0.009	0.019

Panel B: FEARS and selected exchange traded funds (ETFs) returns

	SPY ETF Returns				QQQQ ETF Returns			
	(1) Ret(t)	(2) Ret(t+1)	(3) Ret(t+2)	(4) Ret[t+1,t+2]	(5) Ret(t)	(6) Ret(t+1)	(7) Ret(t+2)	(8) Ret[t+1,t+2]
FEARS	–0.00527***	0.00213**	0.00164	0.00378***	–0.00457***	0.00214**	0.00173	0.00385***
	(0.00136)	(0.000971)	(0.00111)	(0.00136)	(0.00133)	(0.00106)	(0.00116)	(0.00149)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1,891	1,890	1,889	1,889	1,885	1,883	1,882	1,881
Adjusted R^2	0.058	0.026	0.012	0.026	0.030	0.009	0.002	0.008

	IWB ETF Returns				IWM ETF Returns			
	(1) Ret(t)	(2) Ret(t+1)	(3) Ret(t+2)	(4) Ret[t+1,t+2]	(5) Ret(t)	(6) Ret(t+1)	(7) Ret(t+2)	(8) Ret[t+1,t+2]
FEARS	–0.00506***	0.00214**	0.00174	0.00390***	–0.00529***	0.00272**	0.00196	0.00472***
	(0.00128)	(0.000960)	(0.00111)	(0.00137)	(0.00162)	(0.00124)	(0.00145)	(0.00178)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1,891	1,890	1,889	1,889	1,891	1,890	1,889	1,889
Adjusted R^2	0.052	0.019	0.009	0.019	0.041	0.014	0.005	0.015

Panel C: FEARS and Treasury returns

	(1) Ret(t)	(2) Ret(t+1)	(3) Ret(t+2)	(4) Ret[t+1,t+2]
FEARS	0.00112***	–0.000540*	–0.000624**	–0.00117***
	(0.000378)	(0.000324)	(0.000307)	(0.000426)
Constant	Yes	Yes	Yes	Yes
Observations	1,876	1,875	1,874	1,874
Adjusted R^2	0.020	0.007	0.008	0.013

This table relates several alternative index daily returns to FEARS. The dependent variables are contemporaneous returns (column (1) and column (5)) and future returns (columns (2) to (4), and columns (6) to (8)), while the independent variables are the FEARS index and a set of control variables (unreported), which include lagged returns up to five lags, changes in a news-based measure of economic policy uncertainty (EPU), the CBOE volatility index (VIX), and changes in the Aruoba-Diebold-Scotti (ADS) business conditions index. The test assets in Panel A include CRSP equally weighted and value-weighted portfolio daily returns. Panel B includes S&P Exchange Traded Fund (SPY) daily returns, NASDAQ Exchange Traded Fund (QQQQ) daily returns, Russell 1000 Exchange Traded Fund (IWB) daily returns, and Russell 2000 Exchange Traded Fund (IWM) daily returns. Panel C includes CRSP 10-year constant maturity Treasury portfolio daily returns. Bootstrapped standard errors are in parentheses. *, **, and *** denote the coefficient estimates are significant at 10%, 5%, and 1% significance levels, respectively.

future return spreads should be positively correlated with current increases in FEARS. Motivated by Wurgler and Zhuravskaya (2002), we also use total return volatility as a proxy for limits to arbitrage and examine the aforementioned reversal pattern for a portfolio of stocks with high volatility versus a portfolio of stocks with low volatility. The volatility-sorted portfolios are also obtained from CRSP. Using daily stock returns within a calendar year, CRSP computes the total return volatility of common stocks traded on NYSE and AMEX, and creates decile portfolios based on total return volatility.

Panel A from Table 4 confirms the hypothesis. As shown in Panel A, columns 1 and 2, sentiment has a more negative contemporaneous relationship with high-beta stocks. For example, a one-standard-deviation increase in FEARS is associated with a 22-basis-points decrease in the return spread between the high-beta and low-beta stock portfolio (statistically significant at the 1% level). Again, FEARS also predicts future return reversal effects. By $k=2$, the effect is almost completely reversed. Likewise in columns 3 and 4, we find FEARS to have stronger impact on high-volatility stocks than low-volatility stocks on day t, while the impact is almost completely reversed by the end of the second day ($k=2$).

Certain assets are also prone to "downside" risk. As Ang, Chen, and Xing (2006) observe, "downside" risk is not well captured by conventional beta from the capital asset pricing model (CAPM). If downside risk is particularly large when investor sentiment is high, we anticipate that a portfolio of stocks with high downside risk should underperform a portfolio of stocks with relatively low downside risk because downside risk limits arbitrageurs from correcting mispricing. Following Ang, Chen, and Xing (2006), we consider two measures of "downside risk." The first measure is "downside beta," which was first introduced by Bawa and Lindenberg (1977). Specifically, at the end of each month, we estimate the "downside beta" (i.e., β_i^-) for individual stocks as follows,

$$\beta_i^- = \frac{cov(r_i, r_m | r_m < \mu_m)}{var(r_m | r_m < \mu_m)}, \tag{4}$$

using the past year of daily returns.

The second measure of downside risk is "downside sigma" (i.e., σ_i^-), which is defined as follows:

$$\sigma_i^- = \sqrt{var(r_i | r_m < \mu_m)}, \tag{5}$$

and it is also estimated using the past year of daily returns on a monthly basis.

Analogous to the beta-sorted or the total return volatility-sorted portfolios constructed by CRSP, we create decile portfolios on the basis of the stock-level estimates of "downside beta" or "downside sigma" for individual stocks. We track daily portfolio returns over the next month, and rebalance the portfolio at the end of the next month. The return spreads between the returns of the high "downside beta" and low "downside beta" stock portfolios are the test assets in columns 5 and 6 of Panel A. Similarly, columns 7 and 8 relate FEARS and return

Table 4
FEARS and limits to arbitrage

Panel A: Single-sorted portfolio return spreads

	Beta		Total Volatility		Downside Beta		Downside Volatility	
	(1) Ret(t)	(2) Ret[t+1, t+2]	(3) Ret(t)	(4) Ret[t+1, t+2]	(5) Ret(t)	(6) Ret[t+1, t+2]	(7) Ret(t)	(8) Ret[t+1, t+2]
FEARS	–0.00620***	0.00641***	–0.0036***	0.00397***	–0.0106***	0.00716***	–0.00904***	0.00623**
	(0.00199)	(0.00226)	(0.00131)	(0.00146)	(0.00233)	(0.00243)	(0.00228)	(0.00251)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,891	1,889	1,891	1,889	1,891	1,889	1,891	1,889
Adjusted R^2	0.023	0.010	0.046	0.059	0.039	0.012	0.038	0.017

Panel B: Beta-neutral double-sorted portfolio return spreads

	Total Volatility		Downside Beta		Downside Volatility	
	(1) Ret(t)	(2) Ret[t+1, t+2]	(3) Ret(t)	(4) Ret[t+1, t+2]	(5) Ret(t)	(6) Ret[t+1, t+2]
FEARS	–0.00414***	0.00320***	–0.00614***	0.00474***	–0.00374***	0.00234**
	(0.00103)	(0.00119)	(0.00151)	(0.00148)	(0.000837)	(0.00102)
Controls	Yes	Yes	Yes	Yes	YES	YES
Observations	1,891	1,891	1,891	1,891	1,891	1,891
Adjusted R^2	0.088	0.055	0.052	0.013	0.092	0.065

This table links FEARS to daily high-minus-low return spreads on portfolios constructed by sorting on stock characteristics related to limits to arbitrage. In panel, these portfolios are constructed by single sorts on either the CAPM beta, the total volatility, the downside beta, or the downside volatility. In Panel B, we remove the effect from the beta using double sorts. For example, we conduct independent double sorts on beta and total volatility. We then compute high-volatility-minus-low-volatility return spreads using only stocks with similar betas. In each regression, the dependent variables are contemporaneous returns (column with odd numbers) and next-two-day returns (columns with even numbers) while the main independent variable is the FEARS index. The set of control variables (unreported) include lagged returns up to five lags, changes in a new-based measure of economic policy uncertainty (EPU), the CBOE volatility index (VIX), and changes in the Aruoba-Diebold-Scotti (ADS) business conditions index. The CAPM beta is computed following Scholes and Williams (1977) to account for the non-synchronicity in daily returns. Downside beta and downside volatility are computed following Ang, Chen, and Xing (2006). Bootstrapped standard errors are in parentheses. *, ** and *** denote the coefficient estimates are significant at 10%, 5% and 1% significance levels, respectively.

spreads between the high "downside sigma" and low "downside sigma" stock portfolios. The effect of sentiment on these return spreads is large. For instance, a one-standard-deviation increase in FEARS is associated with a decrease of 37 basis points in the return spreads between the high downside beta and low downside beta stock portfolio (statistically significant at the 1% level). Again, FEARS also predicts future return reversals. By $k=2$, the reversal of the return spreads associated with FEARS is about 25 basis points. Thus, sentiment has a stronger effect on high downside beta stocks than low downside beta stocks on day t, while the impact almost completely reverses back by the end of the second day ($k=2$) after event day t, or $k=0$. Similar results are obtained using the high-minus-low-downside-volatility portfolio return spreads.

We have shown earlier that FEARS predict a reversal in market return. Because stocks that are difficult to arbitrage tend to have higher betas, it is perhaps not surprising that FEARS predicts a stronger reversal among these stocks. In other words, the cross-sectional results in Panel A of Table 4 could be driven by a mechanical "beta effect." To examine whether "beta effect" is driving the results shown in Panel A, we construct a series of double-sorted portfolios to account for potential differences in betas across testing assets. Specifically, at the end of each month, we first compute the Scholes-Williams (1977) beta for each stock, using the past twelve-month daily returns. To ensure that our sample is comparable to various decile-sorted portfolios constructed by CRSP and further alleviate liquidity concerns, we restrict our sample to stocks from NYSE and AMEX. We sort these stocks into quintile portfolios. Within each quintile portfolio, we further sort stocks into another set of quintile portfolios based on total volatility, downside beta, or downside volatility (as estimated before). From each beta-sorted quintile portfolio, we compute the return spreads between the high and low total volatility, downside beta, or downside volatility portfolios, and take the average across the beta-sorted quintiles. These double-sorted portfolios generate return spreads with varying degrees of limits to arbitrage, but are beta-neutral.

Panel B of Table 4 reports our results. After removing the "beta effect," FEARS still significantly predicts reversals on the three beta-neutral return spreads due to differences in total volatility (columns 1 and 2), downside beta (columns 3 and 4), or downside volatility (columns 5 and 6), although the magnitudes of the reversals are in general smaller than those reported in Panel A. For example, a one-standard-deviation increase in FEARS is associated with a 22-basis-points decrease in the return spreads between the high and low downside-beta stock portfolio (statistically significant at the 1% level). By $k=2$, the reversal of the return spreads associated with FEARS is about 17 basis points (statistically significant at the 1% level)—or about 77.3% ($=17/22$) of reversal of initial return spreads.

Overall, this evidence provides additional support for the sentiment model of Baker and Wurgler (2006, 2007), which highlights the interaction between speculative trading and limits to arbitrage. It also provides cross-sectional evidence for sentiment-induced mispricing. Among the set of stocks for which sentiment is most likely to operate, we find the strongest evidence of temporary deviation from fundamentals.

2.3 Robustness checks

Construction of our FEARS index required several choices, and in this section we examine the robustness of our results to those choices and the inclusion of additional control variables.

For example, we use the thirty search terms whose $\Delta ASVI$s are most negatively correlated with the market return in our backward rolling window. Averaging FEARS across many search terms allows us to capture their common variation and, at the same time, alleviate idiosyncratic noise. In Panel A of Table 5, we construct alternative FEARS indices by averaging the top twenty-five search terms and top thirty-five search terms. Comparing the results in Table 5, Panel A, with those in Table 2, we find that the alternative FEARS indices produce very similar results. Moreover, to alleviate the effect from extreme outliers in the construction of the FEARS index, we also winsorized the series for each search term at the 5% level (2.5% in each tail). A potential concern about applying winsorization in the context of predictive regressions is that it could introduce a forward-looking bias. To address this concern, the final columns of Panel A report the results of using FEARS indices constructed without winsorization. The results are again very similar to those in Panel A of Table 2, if not slightly stronger.

In the main test specifications, we have been using a news-based measure of economic policy uncertainty (EPU), the CBOE volatility index (VIX), and Aruoba-Diebold-Scotti (ADS) business conditions index as our controls for economic uncertainty, investor sentiment, and macroeconomic conditions. There are also news-based investor sentiment measures. For example, Tetlock (2007) proposes a news-based sentiment measure using the fraction of negative words in the *Wall Street Journal* "Abreast of the Market" column. The news-based investor sentiment measure is available to us through 2010, and this is why we do not include it in our benchmark regressions. Nevertheless, the first two columns of Table 5, Panel B, shows that in this shorter sample our results are robustness to the inclusion of it as an additional control.

Another potential concern regarding our results is that FEARS could simply proxy for extreme market returns, which are more likely to revert in the future. Although we have included the market return and additional lags as control variables in our regressions, one may still be concerned that the FEARS index simply captures a nonlinear effect from large market returns. To address this concern, we include decile dummies for the market return in our regressions in columns 3 and 4 of Panel B. Little changes after the inclusion of these decile dummies.

The next two columns consider the effect of holidays on search and returns. Because search patterns may systematically change around holidays and there

Table 5
Robustness checks

Panel A: Construction

	Top 25		Top 35		No Winsorization	
	(1) Ret(t)	(2) Ret[t+1, t+2]	(3) Ret(t)	(4) Ret[t+1, t+2]	(5) Ret(t)	(6) Ret[t+1, t+2]
FEARS	–0.00513***	0.00374***	–0.00523***	0.00428***	–0.00578***	0.00437***
	(0.00124)	(0.00132)	(0.00133)	(0.00141)	(0.00146)	(0.00149)
Controls	Yes	Yes	Yes	Yes	YES	YES
Observations	1,891	1,889	1,891	1,889	1,861	1,859
Adjusted R^2	0.061	0.026	0.059	0.027	0.067	0.028

Panel B: Additional controls

	Media		Decile Return Fixed Effects		Holidays		Turnover	
	(1) Ret(t)	(2) Ret[t+1, t+2]	(3) Ret(t)	(4) Ret[t+1, t+2]	(5) Ret(t)	(6) Ret[t+1, t+2]	(7) Ret(t)	(8) Ret[t+1, t+2]
FEARS	–0.00564***	0.00448***	–0.00190**	0.00400***	–0.00533***	0.00409***	–0.00525***	0.00419***
	(0.00133)	(0.00148)	(0.000797)	(0.00134)	(0.00130)	(0.00137)	(0.00129)	(0.00137)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,639	1,639	1,891	1,889	1,891	1,889	1,891	1,889
Adjusted R^2	0.068	0.034	0.794	0.029	0.060	0.026	0.061	0.029

Panel C: Subsamples

	No Macro Announcements		After June 2006	
	(1) Ret(t)	(2) Ret[t+1, t+2]	(3) Ret(t)	(4) Ret[t+1, t+2]
FEARS	–0.00539***	0.00325**	–0.00759***	0.00563***
	(0.00144)	(0.00139)	(0.00202)	(0.00192)
Controls	Yes	Yes	Yes	Yes
Observations	1,586	1,584	1,408	1,406
Adjusted R^2	0.058	0.032	0.073	0.029

Panel D: Tradability

	(1) SPY	(2) QQQQ	(3) IWB	(4) IWM
FEARS	0.00262***	0.00265**	0.00268***	0.00273**
	(0.00101)	(0.00115)	(0.00101)	(0.00133)
Controls	Yes	Yes	Yes	Yes
Observations	1,508	1,502	1,508	1,508
Adjusted R^2	0.016	0.011	0.013	0.007

This table reports results from various robustness checks. The dependent variables are contemporaneous and future S&P 500 index daily returns All specification include a set of controls, including lagged returns up to five lags, changes in a new-based measure of economic policy uncertainty (EPU), the CBOE volatility index (VIX), and changes in the Aruoba-Diebold-Scotti (ADS) business conditions index. Panel A considers robustness with respect to the construction of the FEARS index, including estimates when the top 25 terms are used (columns 1 and 2), the top 35 terms are used (columns 3 and 4) and without winsorization (columns 5 and 6). Panel B considers additional controls, including a media-based sentiment measure as in Tetlock (2007), return decile fixed effects, turnover, and holiday controls. Holiday controls constitute dummy variables before and after each NYSE holiday. Panel C considers subsets of the data. Columns 1 and 2 consider the remaining sample when all macro announcements (Savor and Wilson (2012)) have been thrown out, and columns 3 and 4 consider the sample period after June 1, 2006, when Google Trends data became publicly available. Panel D considers tradability and reports results based on various exchange traded funds' open-to-close adjusted-returns on day (t+2). The set of exchange traded funds include the S&P Exchange Traded Fund (SPY), the NASDAQ Exchange Traded Fund (QQQQ), the Russell 1000 Exchange Traded Fund (IWB), and the Russell 2000 Exchange Traded Fund (IWM). *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

is some evidence of holiday-related return phenomenon (see Ariel 1990, for example), columns 5 and 6 of Panel B remove holiday effects by including additional dummy variables for the trading day before and the trading day after public holidays in our sample. Little changes after the inclusion of holiday controls.

Although we interpret the spike-reversal pattern herein as evidence of sentiment, such a pattern is potentially consistent with a liquidity shock following an economic event. The economic event could trigger spikes in both search volume and liquidity trades, pushing prices temporarily away from fundamentals (1993). This would also generate the predictable spike-reversal pattern we find. We address this alternative in a few ways. First, we include the turnover of the S&P 500 index as a control variable in columns 7 and 8 of Panel B.[15] Controlling for liquidity in this way does little to change the results. Second, we obtain macro announcement dates as in Savor and Wilson (2013) and remove all observations with macro announcements. The idea is that although periodical macro announcements may affect investor sentiment, they may also induce portfolio rebalancing and generate liquidity shocks. In Table 5, Panel C, we find the same spike-reversal pattern among observations without macro announcements. Third, recall that we find larger effects among hard-to-arbitrage stocks in Section 2.2. A liquidity hypothesis is unlikely to generate the same cross-sectional pattern in stocks.

Finally, we have been using the FEARS index on day t to "predict" asset returns on days $t+1$ and $t+2$ as we try to understand the economic impact of investor sentiment on contemporaneous and future prices. Because Google releases its SVI data with a one-day delay, these predictive regressions cannot be run in real time. For example, the SVI for a search term on Wednesday, January 23, will typically be released sometime during the evening of Thursday, January 24. Moreover, Google only made this data publicly available in June 2006.

Panels C and D demonstrate the robustness of our reversal results when the predictive regressions are implemented when data are available. The final columns of Panel C consider the subset of observations beginning in June 2006 when search data were available and finds little change in the main result. Panel D considers the predictability of day $t+2$ open-to-close returns with day t search volume. Continuing with the earlier example, it means that we use our FEARS index on Wednesday, January 23 (observable by the evening of Thursday, January 24) to predict the open-to-close returns on Friday, January 25. In other words, the predictive variables are strictly observable before the asset return can be computed. Because open prices are needed for this analysis, we focus our attention on ETFs. The results in Panel D confirm the strong and significant reversals across all four ETFs. The regression coefficients are only slightly smaller compared with those in Table 3, Panel B, reflecting the fact

[15] In untabulated results we also find that controlling for signed turnover — that is, the interaction of turnover on day t with the return on day t, does not change the reversal results on days t+1 and t+2.

that we are using open-to-close returns rather than the standard close-to-close returns.

3. FEARS and Volatility

A long strand of literature starting from Black (1986) suggests that investor sentiment and the resulting noise trading can affect both the level and the volatility of asset prices. If uninformed noise traders base their trading decisions on sentiment, then extreme sentiment changes will temporarily lead to more noise trading, greater mispricing, and excessive volatility. To our knowledge, no prior work has examined the relation between sentiment measures and market-level volatility at a high frequency.[16] In this section we examine the relationship between FEARS and various stock market return volatility measures. The results are reported in Table 6.

We start by examining two direct measures of stock market volatility. The first measure is realized volatility (RV), developed by Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2003). We implement the realized volatility estimation procedure by closely following Andersen, Bollerslev, Diebold and Ebens (2001). Since intraday transaction data are needed to calculate daily realized volatilities, we focus our attention on the SPDR S&P 500 ETF (NYSEARCA: SPY) as a close proxy for the stock market index. The SPY ETF is extremely liquid. For instance, the bid-ask spread is almost always one cent, the minimum tick size. Similar to Antweiler and Frank (2004), we choose fifteen-minute periods when we sample the intraday returns. $r_{t,d}$ denotes the intraday return for SPY during the d-th period on day t. SPY's (annualized) realized volatility on day t is given by

$$rv_t = 250 \sum_{d=1}^{N} r_{t,d}^2. \tag{6}$$

We then compute daily log realized volatility, rv, and remove potential seasonal effects by regressing it on day-of-the-week and month-of-the-year dummies. We focus on the residuals, or the seasonal-adjusted log RV time series (adj_rv). Because volatility is persistent and long-lived (Engel and Patton (2001); Andersen, Bollerslev, Diebold and Ebens 2001; Andersen, Bollerslev, Diebold, and Labys 2003), we also model the long-range dependence through the fractional integrated autoregressive moving average model, $ARFIMA(1,d,1)$:

$$(1-L)^d \left(adj\_rv_t - \beta_1 FEARS_t - \sum_m \beta_m Control_{i,t}^m \right) = (1-L)\varepsilon_t \tag{7}$$

[16] Using Yahoo! message board activities as a proxy for noise trading, Antweiler and Frank (2004) and Koski, Rice, and Tarhouni (2008) confirm the positive relation between noise trading and future volatility at the daily frequency for a small set of individual stocks.

Table 6
FEARS and volatility

Panel A: ARFIMA(1,d,1) on seasonal-adjusted log realized volatility and log VIX

	Realized Volatility on SPY			VIX		
	(1)	(2)	(3)	(4)	(5)	(6)
p	−0.060	−0.019	−0.104	0.844***	0.838***	0.838***
	(0.133)	(0.117)	(0.124)	(0.028)	(0.030)	(0.030)
q	−0.136	−0.204	−0.194	−0.520***	−0.513***	−0.513***
	(0.140)	(0.121)	(0.129)	(0.049)	(0.051)	(0.051)
d	0.484***	0.486***	0.486***	0.493***	0.494***	0.494***
	(0.019)	(0.017)	(0.017)	(0.010)	(0.009)	(0.009)
FEARS	0.233***			0.168***		
	(0.044)			(0.003)		
FEARS, 1st lag		−0.047			−0.000	
		(0.044)			(0.003)	
FEARS, 2nd lag			−0.013			−0.002
			(0.044)			(0.003)
Controls	YES	YES	YES	YES	YES	YES
Observations	1891	1890	1889	1891	1890	1889
Log Likelihood	−2146.1	−2158.5	−2158.0	2384.0	2368.5	2367.4

Panel B: Returns on VIX futures contract

	(1) Ret(t)	(2) Ret(t+1)	(3) Ret(t+2)	(4) Ret[t+1, t+2]
FEARS	0.0119***	−0.00303	−0.00493*	−0.00798**
	(0.00326)	(0.00266)	(0.00271)	(0.00368)
Controls	Yes	Yes	Yes	Yes
Observations	1,886	1,885	1,884	1,884
Adjusted R^2	0.011	0.001	−0.001	0.001

This table relates FEARS to stock market volatility. In Panel A, we model the seasonal–adjusted log realized volatility and log VIX as ARFIMA(1,d,1) processes that include FEARS (or its first or second lags) and other control variables. Realized volatility is computed using SPY intraday data. Both Panel A and B are estimated using the maximum likelihood method. Panel B relates Chicago Board of Exchange (CBOE) VIX futures daily returns to FEARS. For a contract with given settlement date, its daily return is computed as the change of logarithm of daily prices, or $R_t = \log(\frac{P_t}{P_{t-1}})$. Daily returns are obtained from the nearest to maturity contract until five trading day before the nearest–to–maturity contract's settlement date. Afterward, daily returns are obtained from the second–nearest–to–maturity contract. The control variables include changes in a news–based measure of economic policy uncertainty (EPU), and changes in the Aruoba–Diebold–Scotti (ADS) business conditions index.*, **, and *** denote the coefficient estimates are significant at 10%, 5%, and 1% significance levels, respectively.

where the fractional integration parameter is $d \in (0, 0.5)$. The control variables are changes in a news-based measure of economic policy uncertainty (EPU), and changes in the Aruoba-Diebold-Scotti (ADS) business conditions index. We estimate (7) using the maximum likelihood method. The key coefficient β_1 identifies the impact of the FEARS index on the realized volatility of the stock market after controlling for the persistent component in volatility, changes in EPU and ADS.

The second measure of the stock market volatility is the CBOE daily market volatility index (VIX). As in the case of the realized volatility, we also first compute the seasonal-adjusted log VIX time series (adj_vix) and then estimate a similar $ARFIMA(1,d,1)$ as in (7), except that we replace adj_rv_t with adj_vix_t. The results are reported in Panel B (columns 1 to 3 for the realized volatility and columns 4 to 6 for the VIX). We find that our

FEARS index is positively and significantly related to the market volatility measures only contemporaneously (see columns 1 and 4). Controlling for the persistent component in volatility, neither realized volatility nor VIX loads significantly on lagged FEARS index. These results again suggest that our FEARS index has only a transitory impact on the level of stock market volatility.

Second, parallel to our analysis in the previous sections, we also examine daily returns to a tradable asset based on volatility, the CBOE VIX futures contract. Working with the return series has the benefit of circumventing potential econometric issues associated with the VIX and RV time series and providing a clear interpretation of asset returns. For a contract with a given settlement date, its daily return is computed as the change of log daily prices.[17] We then use our FEARS index to predict these daily VIX futures returns using the same regression specifications in Equation (3). The results are reported in Panel C.

Panel C confirms the strong contemporaneous correlation between FEARS and volatility. For example, a one standard deviation increase in FEARS corresponds with a contemporaneous 42-basis-points increase in VIX futures return. In the next two trading days, we again find a reversal pattern. By the end of the second trading day, we observe a total reversal of 28 basis points.

Thus far analyzing both the levels and changes in stock market volatility, our results paint a consistent picture: an increase in our FEARS index coincides with an increase in market volatility that is temporary. To the extent that a spike in our FEARS index coincides with more noise trading, our evidence provides further support for the DSSW model, where noise trading leads to excessive volatility temporarily. The DSSW model also predicts a positive relation between the volatility of sentiment and the volatility of the asset price (see equation 11 in DSSW). The intuition is simple: if sentiment contributes to a temporary price deviation from fundamental value, then the more volatile sentiment is, the higher the excessive price volatility should be. While our focus is on the level of investor sentiment as measured by our FEARS index, we also try to analyze the joint volatility dynamics between our FEARS index and market returns. Specifically, we model daily stock market excess return and our FEARS index jointly using a multivariate GARCH with dynamic conditional correlation (see Engle (2000)). Unreported results confirm a significant positive correlation of 7.17% (p-value of 0.002) between the conditional variance of the stock market return and that of the FEARS index.

[17] Daily returns are calculated using the contract closest to maturity except when this contract is less than five days away. When the closest to maturity contract is less than five days away, daily returns are calculated from the second closest to maturity contract.

4. FEARS and Fund Flows

Noise traders affect asset prices via trading. To directly examine the sentiment effects of noise traders, we examine daily mutual fund flows in our last set of tests. Because individual investors hold about 90% of total mutual fund assets, and they are more likely to be sentiment traders, daily flows to mutual fund groups likely aggregate noise trading at the asset-class level (Brown, Goetzmann, Hiraki, Shiraishi, and Watanabe 2002 2002). Daily mutual fund flow data are obtained from TrimTabs for two groups of mutual funds that specialize in equity (Equity) and intermediate treasury bonds (MTB).

Bollerslev and Jubinski (1999) and Fleming and Kirby (2006) provide evidence that an individual stock's daily trading volume series exhibits long-run temporary dependencies, which can be modeled using a fractionally integrated process. Similar to observations made on the volume of individual stocks, we also find very strong persistence and long-memory components in daily fund flows. For this reason, we first demean each of the daily fund flow series, and apply the $ARFIMA(p,d,q)$ models to extract daily fund flow innovations. Our diagnostics indicate that the $ARFMA(1,d,1)$ model fits the underlying daily fund flows well. The integration parameter values are in the neighborhood of 0.40 and p-value less than 0.1%. In addition, the moving average (MA) as well as the autoregressive (AR) terms are all statistically significant at the 1% level or better.

There is one data issue worth pointing out. TrimTabs mutual fund flow is calculated using both publicly observable net asset value (NAV) and privately reported total asset value (NTA). Despite the obvious accuracy of NAV, the NTA information might be reported with a delay of one day for some funds. Both Edelen and Warner (2002) as well as Greene and Hodges (2002) document this issue, and analyze it in detail. Because of this potential one-day reporting delay, we note that TrimTabs flow in day $t+1$ may actually contain flow in day t (see also Yuan (2008)). We run regressions of contemporaneous fund flows and fund flows one to four days ahead. In particular, we run the following regression:

$$flow_{i,t+k} = \beta_0 + \beta_1 FEARS_t + \sum_m \gamma_m Control_{i,t}^m + u_{i,t+k} \tag{8}$$

where fund class i includes bond and equity funds. Control variables($Control_{i,t}^m$) include as usual VIX, ΔEPU, ΔADS, and five lags of market returns. The results of these regressions are reported in Table 7.

We find that our FEARS index has significant incremental predictive power for future daily fund flow innovations of both equity and bond funds. In the equity flow regressions, the coefficient on FEARS is negative on each day we consider ($t=0,1,2,3,$ and 4) and is statistically significant for days $t+1$ (p-value $< 10\%$), $t+2$ (p-value $< 5\%$) and $t+3$ (p-value $< 10\%$). The results suggest that investors start to withdraw from equity mutual funds the day during the spike in FEARS (recall that the outflow on $t+1$ may actually contain outflow in day t

Table 7
Sentiment and fund flows

Panel A: Equity fund flow

	Flow (t) (1)	Flow (t+1) (2)	Flow (t+2) (3)	Flow (t+3) (4)	Flow (t+4) (5)
FEARS	1.04e–05	–8.62e–05*	–8.95e–05**	–7.60e–05*	–6.49e–05
	(5.18e–05)	(4.82e–05)	(3.89e–05)	(4.29e–05)	(4.25e–05)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	1,339	1,338	1,337	1,336	1,335
Adjusted R^2	0.081	0.096	0.11	0.081	0.046

Panel B: Bond fund flow

	Flow (t) (1)	Flow (t+1) (2)	Flow (t+2) (3)	Flow (t+3) (4)	Flow (t+4) (5)
FEARS	0.000174	8.31e–05	0.000165	0.000231**	6.57e–05
	(0.000224)	(1.58e–05)	(0.000108)	(0.000108)	(9.47e–05)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	1,339	1,338	1,337	1,336	1,335
Adjusted R^2	0.014	0.013	0.017	0.013	0.016

This table reports the results of contemporaneous and predictive regressions. We consider two mutual fund groups specializing in equity (Panel A) and medium–term Treasury bonds (Panel B). For each mutual fund group, we obtain its daily fund flow (as a percentage of TNA) from TrimTabs. To remove the persistence in fund flow, we use a ARFIMA(1,d,1) model to extract daily flow innovations. The set of control variables include lagged returns up to five lags, changes in a news–based measure of economic policy uncertainty (EPU), the CBOE volatility index (VIX), and changes in the Aruoba–Diebold–Scotti (ADS) business conditions index. Bootstrapped standard errors are in parentheses. *, ** and *** denote the coefficient estimates are significant at 10%, 5%, and 1% significance levels, respectively.

due to a reporting delay in the TrimTabs data), and such an outflow persists for the next two days. Interestingly, in the bond flow regressions, the coefficient on FEARS is positive on all days but only significant on $t=3$, suggesting a significant inflow to bond funds one day after a significant withdrawl from equity funds. Taken together, the evidence highlights a flight to safety where investors are shifting their investments from equity funds to bond funds after a spike in FEARS.

Considering equity flows, the coefficients on FEARS are economically large. For example, an one standard-deviation increase in FEARS is associated with significant equity outflows of $-3.13\ \times 10^{-5}(=0.35\times -8.95\times 10^{-5})$ on $t=2$. Given the average equity fund flow of -5.06×10^{-5}, this is about 62% of the typical daily flows. Similarly, a one-standard-deviation increase in FEARS is associated with significant bond inflows of 8.1×10^{-5} on $t=3$, which is slightly larger than the average daily bond flow (7.49×10^{-5}).

In short, the evidence herein suggests that individual investors switch from equity funds to bond funds when negative sentiment is high.

5. Discussion of Alternative Interpretations

Just as many authors have understood the solicitation of household attitudes by survey as a measure of sentiment (e.g., Brown and Cliff (2005), Lemmon and Portniaguina (2006), and Qiu and Welch (2006)), we understand the revelation of household attitudes via search as a measure of sentiment. We then test many of the predictions of sentiment models such as DSSW. So far we have found evidence that the attitudes of households as revealed by their search behavior have predictability for short-term returns, short-term market volatility and both equity and bond mutual fund flows.

5.1 Endogenous search

Some readers may be concerned that search is endogenous to macroeconomic events. For example, there must be some macro events that coordinate the large spikes in search we observe in Figure 1. This does not disqualify search as a measure of sentiment. In fact, we should expect investor sentiment to be endogenous to macroeconomic events.[18] News arrives daily - some of it will affect investor sentiment and some of it will not. To the extent that daily returns, the policy uncertainty index and the business conditions index measure news arrival, we have explicitly controlled for news events in each of our specifications. Therefore, we can think of our FEARS index as describing the amount of sentiment generated by an event.

Other readers will be concerned about reverse causality in some of our prediction models if events are anticipated. We cannot conclude that sentiment today causes return tomorrow in the same way we cannot conclude that someone who buys an umbrella today in preparation for rain tomorrow causes the rain tomorrow. However, the predictability for returns (Section 2) likely mitigates such concerns. The fact that we find high FEARS today are correlated with low returns today but predict high returns tomorrow makes reverse causality unlikely. It is implausible that investors, anticipating a high return tomorrow, would search for terms like "recession" and "inflation" today. Return reversal following a spike in the FEARS index is more consistent with sentiment models, which predict temporary deviation from fundamentals.

5.2 Search as a measure of sentiment

Beyond endogeneity concerns, there are also other interpretations of our measure and its subsequent predictability for asset volatility. For instance, it is possible that search for terms like "recession" or "great depression" proxy for time-varying risk aversion. In Campbell and Cochrane (1999), a low surplus consumption ratio will jointly cause risk aversion and volatility to increase. In Kyle and Xiong (2001), when convergence traders have reduced capital as a result of losses, their risk aversion will increase (due to wealth effects) while asset volatility increases as they liquidate their positions. Both models generate a correlation between risk aversion and volatility in the time series.

---

[18] Qui and Welch (2006), p. 32, discuss this issue as well. They argue: "The theories are about sentiment, not about sentiment orthogonal to macroeconomic conditions. In what theory would we expect sentiment not to be related to unemployment, GDP, portfolio returns, wealth changes, etc.? (Answer: None!) Sentiment does not drop like manna from heaven."

While this is a possible interpretation of our evidence, there are two important caveats. First, neither model generates a predictable reversal in prices, which is what we find in Section 2. Second, there is little evidence that risk aversion changes rapidly (see Brunnermeier and Nagel 2008). Therefore, it seems unlikely that the large daily variation we observe in search volume represents time-varying risk aversion.

Alternatively, FEARS may be proxying for time-varying parameter uncertainty. Uncertainty about the parameters of models governing the dynamics of asset returns can be positively related to future asset volatility (see Veronesi 1999, among others). While the VIX index is commonly viewed as an indicator of aggregate uncertainty, we do not find any evidence that VIX is related to return reversal: FEARS remains a strong predictor of future VIX even after controlling for current VIX. Moreover, our policy uncertainty control variable (EPU) further alleviates this concern.

Finally, some readers may worry that search for FEARS is a neutral activity that does not reflect underlying pessimism or optimism. The argument is that households may search for terms like "inflation" or "recession" not because they are concerned about inflation or a recession but rather because they wish to gather information about inflation or recession. This claim is not supported by the evidence. First, even a cursory look at many of the FEARS components (such as "recession" or "bankruptcy") suggests they increase in bad times (Table 2). For example, (negative) search volume for the term "recession" has an 85.8% correlation with the University of Michigan's Consumer Sentiment Index, suggesting most of the time households search for "recession" when they are worried about a recession. Second, recall from Section 2 that we find a contemporaneous, negative relationship between FEARS and equity returns. The days in which equity returns are low are the same days in which households search for terms in our FEARS index.

6. Conclusion

By aggregating queries like "recession," "bankruptcy," and "depression," we construct a Financial and Economic Attitudes Revealed by Search (FEARS) index. We show that the FEARS index predicts aggregate market returns. In particular, the FEARS index is correlated with low returns today but predicts high returns tomorrow, a reversal pattern that is consistent with sentiment-induced temporary mispricing. Moreover, this effect is strongest among stocks that are favored by sentiment investors and are difficult to arbitrage. In addition, our FEARS index is strongly related to the transitory component of daily volatility, and it is also correlated with VIX futures returns. Finally, using daily aggregate mutual fund flows, we also provide direct evidence for "noise" trading. Increases in the FEARS index trigger daily mutual fund flows out of equity funds and into bond funds. The evidence is broadly consistent with the "noise trading" hypothesis of De Long et al. (1990).

More generally, this paper follows a new strand of the sentiment literature that proposes novel, high-frequency measures that do not rely on market outcomes like return and volume. Tetlock (2007) suggests that a journalist's tone as measured by the frequency of negative words in a *Wall Street Journal* column captures sentiment and also shows that this tone has predictability for returns. Tetlock (2007) argues that the results have two reasonable interpretations: the media reports investor sentiment before it is fully incorporated into market prices, or the media directly influences investors' attitudes toward stocks. Although we also find predictability for returns, our results have only one reasonable interpretation because aggregate search volume does not require a journalist or other intermediary. As such this paper underscores the usefulness of search data in financial applications. Search data has the potential to objectively and directly reveal to empiricists the underlying beliefs of an entire population of households. Given that many financial models link beliefs to equilibrium outcomes (such as returns or volume), search behavior has the potential to provide sharper tests of economic models. The tests herein constitute one possible application of search data. We leave the many other applications for future research.

References

Andersen, T., T. Bollerslev, F. X. Diebold and H. Ebens. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* 61:43–76.

Andersen, T., T. Bollerslev, F. X. Diebold, and P. Labys. (2003). Modeling and forecasting realized volatility. *Econometrica* 71:579–625.

Ang, A., J. Chen, and Y. Xing. (2006). Downside risk. *Review of Financial Studies* 19:1191–239.

Antweiler, W., and M. Z. Frank. (2004). Is all that talk just noise? The information content of Internet stock message boards. *Journal of Finance* 59:1259–94.

Ariel, R. A. (1990). High stock returns before holidays: Existence and evidence on possible causes. *Journal of Finance* 45:1611–26.

Aruoba, S. B., F. X. Diebold, and C. Scotti. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics* 27:417–27.

Baker, S. R., N. Bloom, and S. J. Davis. (2013). Measuring economic policy uncertainty. Working Paper, Stanford University.

Baker, M., B. Bradley, and J. Wurgler. (2011). Benchmarks as limits to arbitrage: Understanding the low volatility anomaly. *Financial Analysts Journal* 67: 40–54.

Baker, M., and J. Stein. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets* 7:271–99.

Baker, M., and J. Wurgler. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance* 61:1645–80.

———. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives* 21:129–51.

Baur, D. G., and B. M. Lucey. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds, and gold. *Financial Review* 45:217–29.

Bawa, V. S. and E. B. Lindenberg. (1977). Capital market equilibrium in a mean-lower partial moment framework. *Journal of Financial Economics* 5:189–200.

Black, F. (1986). Noise. *Journal of Finance* 41:529–43.

Bloomfield, R., M. O'Hara, and G. Saar. (2009). How noise trading affects markets: an experimental analysis. *Review of Financial Studies* 22:2275–302.

Bollerslev, T., and D. Jubinski. (1999). Equity trading volume and volatility: Latent information arrivals and common long-run dependencies. *Journal of Business and Economic Statistics* 17:9–21.

Brown, G., and M. Cliff. (2005). Investor sentiment and asset valuation. *Journal of Business* 78(2):405–40.

Brown, S. J., W. N. Goetzmann, William N., T. Hiraki, N. Shiraishi, and M. Watanabe. (2002). Investor sentiment in Japanese and U.S. daily mutual fund flows. Yale ICF Working Paper No. 02-09.

Brunnermeier, M., and S. Nagel. (2008). Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals' asset allocation. *American Economic Review* 98(3):713-36.

Campbell, J., and J. Cochrane, (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107 (2):205–51.

Campbell, J., S. Grossman and J. Wang. (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics* 108(4):905–39.

Da, Z., P. Gao, and J. Engelberg. (2011). In search of attention. *Journal of Finance* 66:1461–99.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann. (1990). Noise trader risk in financial markets. *Journal of Political Economy* 98:703–38.

Edelen, R., and J. Warner. (2002). Aggregate price effects of institutional trading: a study of mutual fund flows and market returns. *Journal of Financial Economics* 59:195–221.

Edmans, A., D. Garcia, and O. Norli. (2007). Sports sentiment and stock returns. *Journal of Finance* 62(4):1967–98.

Engel, R. F., and A. J. Patton. (2001). What good is a volatility model? *Quantitative Finance* 1:237–45.

Fleming, J., and C. Kirby. (2006). Long memory in volatility and trading volume. Working Paper. Rice University.

Greene, J. T., and C. W. Hodges. (2002). The dilution impact of daily fund flows on open-end mutual funds. *Journal of Financial Economics* 65:131–58.

Hellwig, Martin F. (1980). On the Aggregation of Information in Competitive Markets, *Journal of Economic Theory* 22, 477–98.

Hirsheifer, D., and T. Shumway. (2003). Good day sunshine: Stock returns and the weather. *Journal of Finance* 58:1009–1032.

Kamstra, M. J., L. A. Kramer, and M. D. Levi. (2003). Winter blues: A SAD stock market cycle. *American Economic Review* 93:324–43.

Kaplanski, G., and H. Levy. (2010). Sentiment and stock prices: The case of aviation disaster. *Journal of Financial Economics* 95:174–201.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. London: Macmillan.

Kogan, L., S. A. Ross, J. Wang, and M. M. Westerfield. (2006). The price impact and survival of irrational traders. *Journal of Finance* 61(1):195–229.

Kogan, S., D. Levin, B. Routledge, J. Sagi and N. Smith. (2009). Predicting risk from financial reports with regression. Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference.

Kogan, L., S. A. Ross, J. Wang, and M. M. Westerfield. (2009). Market selection. Working Paper, MIT.

Koski, J. L., E. M. Rice and A. Tarhouni. (2008). Noise trading and volatility: Evidence from day trading and message boards. Working Paper, University of Washington.

Kyle, A., and W. Xiong. (2001). Contagion as a wealth effect. *Journal of Finance* 56:1401–40.

Lemmon, M. L., and E. V. Portniaguina. (2006). Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies* 19(4):1499–529.

Loughran, T., and B. McDonald. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66:35–65.

Pontiff, J. (1996). Costly arbitrage: Evidence from closed-end funds. *Quarterly Journal of Economics* 111(4):1135–51.

Qiu, L., and I. Welch. (2006). Investor sentiment measures. Working Paper, Brown University.

Savor, P., and M. Wilson. (2013). How much do investors care about macroeconomic risk? Evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis* 48:343–75.

Scholes, M., and J. Williams. (1977). Estimating betas from nonsynchronous data. *Journal of Financial Economics* 5:309–27.

Shleifer, A., and R. W. Vishny. (1997). The limits of arbitrage. *Journal of Finance* 52:35–55.

Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. Working Paper, University of Michigan.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62:1139–68.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy. (2008). More than words: Quantifying language to firms' fundamentals. *Journal of Finance* 63(3):1437–67.

Veronesi, P. (1999). Stock market overreaction to bad news in good times: A rational expectations equilibrium model. *Review of Financial Studies* 12:975–1007.

Vissing-Jorgensen, A. (2003). Perspectives on behavioral finance: Does "irrationality" disappear with dealth? Evidence from expectations and actions. *NBER Macroeconomics Annual*.

Whaley, R. E. (2001). The investor fear gauge. *Journal of Portfolio Management* 26:12–7.

Wurgler, J., and E. Zhuravskaya. (2002). Does arbitrage flatten demand curves for stocks? *Journal of Business* 75:583–608.

Yuan, Yu. (2008). Attention and trading. Working Paper, University of Iowa.

It Depends on Where You Search:
A Comparison of Institutional and Retail Attention

Azi Ben-Rephael†, Zhi Da§, and Ryan D. Israelsen†

First Version: July 2015
This Version: January 2016

Abstract

We propose a direct measure of abnormal institutional investor attention (AIA) at the daily frequency using the news-searching and news-reading activity for specific stocks on Bloomberg terminals. We find AIA to be highly correlated with measures of contemporaneous institutional trading, and related to but different from other investor attention proxies, including user requests at EDGAR which are limited to specific regulatory filings. Importantly, AIA enables us to directly contrast institutional attention with retail attention measured using Google search frequency. We find that institutional attention responds more quickly to major news events, triggers more trading, and is less constrained, compared to retail attention. In sharp contrast to retail attention which results in positive and temporary price pressure, AIA facilitates permanent price adjustment and alleviates price under-reaction to news. The well-documented price drifts following both earnings announcements and analyst recommendation changes come only from announcements where institutional investors fail to pay attention, according to our measure.

---

* We thank Christa Bouwman, Joey Engelberg, Huseyin Gulen, Mark Grinblatt, Umit Gurun, Harrison Hong, Byoung-Hyoun Hwang, Chris Parsons, Anjan Thakor, seminar participants at Indiana University, IDC Herzliya, and Tel Aviv University, and participants at the 2016 Jackson Hole Finance Group Conference and 2015 Wabash River Finance Conference for helpful comments and suggestions.
† Azi Ben-Rephael and Ryan Israelsen are at the Kelley School of Business at Indiana University. Ben-Rephael: abenreph@indiana.edu, +1 (812) 856-4068. Israelsen: risraels@indiana.edu, +1 (812) 855-8435.
§ Zhi Da is at the Mendoza College of Business at Notre Dame University. zda@nd.edu, +1 (574) 631-0354.

1. Introduction

Information needs to attract investor attention before it can be processed and incorporated into asset prices via trading. Attention, however, is a scarce cognitive resource (Kahneman, 1973).[1] A voluminous literature has demonstrated that limited investor attention is often associated with slow information diffusion and under-reaction to news.[2]

When empirically examining the impact of limited investor attention on price reaction to information, it is important to differentiate retail investor attention from institutional investor attention. Retail investors are less likely to immediately act on information when it arrives. Additionally, since retail investors rarely short, retail attention on average results in positive and temporary price pressure (Barber and Odean, 2008). Such positive price pressure is confirmed by Da, Engelberg, and Gao (2011) using Google search as a direct measure of retail attention. In sharp contrast, institutional investors have greater resources and stronger incentives to quickly react to news and are more likely to be the marginal investors in moving prices. As such, institutional investor attention is likely far more important in facilitating the incorporation of new information into asset prices. In this paper, we focus on institutional investor attention and confirm that its impact on asset prices is very different from that of retail attention.

We propose a novel measure of institutional investor attention using the news-searching and news-reading activity for specific stocks on Bloomberg terminals. Because Bloomberg terminals are expensive – annual subscriptions cost $20,000 to $24,000, per machine – and are leased on a 2-year basis, they are much more likely to be used by institutional investors than retail investors.[3] In fact, the number of subscriptions is limited to about 320,000 worldwide.[4] Bloomberg records the number of times each article is read by its users as well as the number of times users search for news for a specific stock. They then rank these numbers against user behavior over the same stock during the previous 30 days and provide us with the transformed

[1] By attention, we mean the cognitive resources required in order to reduce the information entropy, following Sims (2003) and Peng (2005).

[2] Examples include Hirshleifer and Teoh (2003), Peng and Xiong (2006), Cohen and Frazzini (2008), DellaVigna and Pollet (2009), Hirshleifer, Lim, and Teoh (2009, 2011), Da, Gurun, and Warachka (2014), Hendershott, Li, Menkveld, and Seasholes (2013), among many others.

[3] Strasburg, J. (2013, May 15). This is How Much a Bloomberg Terminal Costs. *Quartz.* http://qz.com/84961/this-is-how-much-a-bloomberg-terminal-costs/

[4] Bloomberg. Retrieved from http://www.bloomberg.com/professional/tools-analytics/collaboration/ on August 27, 2014.

data. We define *abnormal institutional attention* (hereafter, "*AIA*") as a dummy variable equal to one when attention during the day exceeds levels of attention during at least 94% of the hours over the previous month, and zero otherwise.[5] In other words, an AIA equal to one indicates a spike in institutional investor attention on that stock during that day. Compared to other measures that are indirect or based on equilibrium outcomes such as return and institutional trading volume, AIA directly reveals institutional investor attention.[6]

Figure 1 contains an example of AIA for Overstock.com (NASDAQ: OSTK) during 2013. Vertical bars mark the days associated with abnormal institutional investor attention (AIA=1). The four quarterly earnings announcement days are indicated with an "E" above the figure. The figure shows that the company experienced institutional attention shocks on 15 days during the year. While three of these shocks are driven by earnings announcements, not all earnings announcements result in abnormal institutional attention.

Figure 1 also plots the daily number of relevant news articles (the right axis) with major events described below the figure. Almost all abnormal institutional attention can be traced back to some salient news on the firm (CEO turnover, outcome of lawsuit, analyst recommendation change, etc.). In other words, news coverage and institutional attention are clearly correlated. Nevertheless, news coverage does not guarantee attention and AIA directly identifies the news that attracts the institutional attention. Finally, abnormal institutional attention on Overstock.com is also correlated with extreme price movement.

We find similar determinants of AIA when we examine a broad sample of Russell 3000 stocks during the period from February 2010 to June 2013. Firm-specific news is the most important driver of AIA. Equilibrium outcomes during the day such as absolute return, trading volume, intra-day volatility, and closeness to 52 week high-low are also significantly related to AIA. In addition, AIA displays strong seasonality within the week. The likelihood of an institutional attention shock decreases monotonically from Monday to Friday. For example, a stock is 25% less likely to have an attention shock on a Friday compared to a Monday, consistent

[5] While 94% may seem like an arbitrary cutoff, this number based on the way Bloomberg constructs its measure. See Section 2.2 for more details on the data provided by Bloomberg and on our measure.

[6] Examples include extreme returns (Barber and Odean (2008)), trading volume (Barber and Odean (2008), Gervais, Kaniel, and Mingelgrin (2001), and Hou, Peng, and Xiong (2009)), news and headlines (Barber and Odean (2008) and Yuan (2015)), advertising expense (Chemmanur and Yan (2009), Grullon, Kanatas, and Weston (2004), Lou (2014), Madsen and Niessner (2014)), and price limits (Seasholes and Wu (2007)).

with the results in DellaVigna and Pollet (2009) and the pattern displayed by retail attention documented in Liu and Peng (2015). Finally, in the cross-section, larger and more volatile stocks with more analyst coverage are more likely to experience institutional attention shocks, while advertisement expenditure and institutional holdings are not significant related.

User request at the Securities and Exchange Commission's (SEC) EDGAR (Electronic Data Gathering, Analysis, and Retrieval) online system have also been used to track investor attention.[7] Compared to investors who search for information on Google, investors requesting information on EDGAR are more likely to be institutional investors. While the EDGAR measure is positively and significantly related to AIA, its explanatory power is small compared to the occurrence of news.[8] One important distinction between the two measures is that AIA is based on all news searches while hits on EDGAR are limited to specific regulatory filings. Not surprisingly, controlling for the EDGAR measure in our subsequent analysis hardly changes our results.

Most interestingly, we find our institutional investor attention measured using AIA to be distinct from retail investor attention measured using abnormal daily Google search. While AIA and the Google-search-based measure are positively and significantly correlated at the daily frequency, they explain less than 2% of each other's variation. When we correlate both measures to contemporaneous measures of abnormal trading volume, we find that only AIA has a significantly higher correlation with abnormal institutional trading volume than with abnormal total trading volume. This finding confirms that AIA, not the Google-search-based measure, directly measures institutional investors' attention. Importantly, conditional on major news events, AIA leads retail attentions, but not vice versa, confirming that institutional investors have greater resources and stronger incentives to quickly pay attention to news. Finally, attention constraints are more likely to be binding for retail investors. For example, we find retail attention to be significantly lower if there are more events during the same day, consistent with the evidence in Liu and Peng (2015). No such relation is observed with AIA.

[7] For example, see Bauguess, Cooney, and Hanley (2013), Drake, Roulstone, and Thornock (2015), Lee, Ma, and Wang (2015), deHaan, Shevlin, and Thornock (2015) and Loughran and McDonald (2015) for recent applications of the EDGAR data.

[8] The EDGAR measure is limited to a subset of mandatory filings while AIA captures abnormal institutional attention to a broader set of news events. Indeed, Drake, Roulstone, and Thornock (2015) find that 86% of the users accessing EDGAR do so infrequently and only around 2% of the users access EDGAR actively during a given quarter.

We then examine how institutional investor attention affects the incorporation of information into asset prices, we focus on two types of firm-level announcements — quarterly earnings announcements and analyst recommendation changes (that are not immediately driven by earnings announcements) — for four reasons. First, both announcements contain important value-relevant information to which institutional investors are likely to pay attention and react.[9,10] Second, information released in both announcements is quantifiable, which allows us to control for both the magnitude and implications of the information and tease out the incremental impact of the attention. Third, both announcements have been documented in the literature to generate post-announcement drift (see for example, Ball and Brown, 1968 and Livnat and Mendenhall, 2006 for earnings announcements; and Stickel, 1995 and Womack, 1996 for analyst recommendation changes). In other words, investors underreact to both announcements on average. It is only a natural question to ask if institutional attention on the announcement day facilitates information incorporation and alleviates price under-reaction to news. Finally, by examining two distinct types events of we can determine whether institutional attention plays a broad role, or is more limited.

We find strong and consistent evidence that institutional attention facilitates information incorporation for both types of announcements. The evidence is summarized in Figure 2 (earnings announcements) and Figure 3 (analyst recommendation changes). After controlling for the information content of the announcement and a comprehensive set of relevant stock characteristics, announcements accompanied with abnormal institutional attention experience larger return (in absolute term) during the announcement day and very little subsequent price drift. In other words, the well-documented post-announcement drifts come almost exclusively from announcements with limited institutional investor attention. When institutional investors fail to pay sufficient attention, price initially underreacts to information, resulting in a drift.

[9] For example, Schmidt (2015) finds that professional asset managers with a large fraction of portfolio stocks exhibiting an earnings announcement are significantly less likely to trade in other stocks, suggesting that many earnings announcements indeed grab institutional investor attention. In fact, since earnings announcements are usually pre-scheduled, investors may be prepared to allocate more attention on the earnings announcement days. We confirm that AIA is on average higher on earnings announcement days than on the days of recommendation changes which are usually not pre-scheduled.

[10] Along these lines, Boudoukh, Feldman, Kokan and Richardson (2013) use textual analysis to identify relevant news (from the set of all news). They show that focusing on relevant news, there is considerably more evidence of a strong relationship between stock price changes and information.

We confirm the incremental value of AIA in a two-step exercise. In the first step, we orthogonalize AIA by regressing it on a broad set of variables we show to be related to AIA. These variables include equilibrium outcomes such as abnormal volume and intraday volatility, firm characteristics, news, analyst coverage, institutional holdings, seasonality and other direct attention measures. The residual is the unexpected AIA which, by construction, captures abnormal institutional attention unrelated to the independent variables. In the second step, we replace AIA with the residual AIA in the return regressions and find very similar results. Thus, the relation between AIA and price reaction to news announcements is not driven by AIA's correlations with other variables that have been documented to be related to post-announcement-drifts.

We fully acknowledge the possibility that other features of the announcements that we have not controlled for may be driving the high AIA on the announcement day itself. After all, the allocation of attention is an endogenous decision. While it is easy for such "unidentified" features to explain the higher announcement-day return (in absolute terms), it is more challenging for them to also explain a lower post-announcement drift. For example, while important news may drive both higher AIA and a higher absolute announcement-day return, it tends to be associated with stronger, not weaker, drift going forward. In fact, Chan, Jegadeesh, and Lakonishok (1996) document that higher (absolute) earnings-announcement-window-returns predict stronger, not weaker, post-earnings-announcement drift on average. One may also argue that higher AIA reflects higher uncertainty or disagreement that prompts institutional investors to acquire more information on Bloomberg, but it is harder to explain why prices fully adjust to news associated with more uncertainty and disagreement but only partially adjust when there is little uncertainty. In the end, we believe that institutional investor attention seems the most natural economic force that simultaneously explains both a higher price response immediately upon the announcement and a lower subsequent price drift.

To rule out a reverse causality story where a higher announcement-day (absolute) return actually leads to high AIA, we focus on earnings announcements taking place after the market closes – from 4pm to 12am. For these announcements, which make up half of our sample, high

AIA on the same day cannot be driven by the earnings-announcement return.[11,12] Yet we find very similar results in this reduced sample: high AIA is associated with stronger price adjustment on the first trading day following the announcement and lower price drift afterwards even after controlling for the size of the earnings surprise and other stock characteristics.

Not surprisingly, in sharp contrast to institutional attention, we find that retail attention does not facilitate the incorporation of information during earnings and recommendation change announcements. In fact, regardless of the content of the news, retail attention tends to result in positive price pressure, consistent with evidence in the prior literature (see Barber and Odean, 2008, and others).[13]

The impact of investor attention on the price reaction to news announcements has been examined before. A few papers use indirect proxies for attention. For example, Hirshleifer, Lim and Teoh (2009) find that when there are more firms reporting earnings on the same day, stocks have smaller reactions on the announcement date and greater drift going forward. DellaVigna and Pollet (2009) find similar results when announcements are made on Fridays. Several papers use trading volume as a measure of attention. Hou, Peng, and Xiong (2009) document that stocks with higher trading volume experience smaller post-earnings-announcement-drift. Similarly, Loh (2010) finds that stocks with higher trading volume react more to stock recommendations during the announcement and experience smaller subsequent price drift. Boehmer and Wu (2013) use short selling volume as a proxy for investor attention and show that there is little drift when there are negative earnings surprises and short selling volume is high. The advantage of our AIA measure is twofold. First, it allows us to focus on institutional investor attention which is more important for driving permanent price change. Second, while trading volume and short interest are equilibrium outcomes that reflect many economic forces other than investor attention, AIA directly reveals institutional investor attention.

[11] In contrast to earnings announcements, only 15% of the recommendation changes in our sample take place after the market has closed.

[12] We acknowledge that trading does occur in OTC markets after market close. However, trading volume is by far smaller and less concentrated relative to the trading volume at the opening on day-*t*. Thus, it is fair to assume that institutional investors are more likely to notice news than prices in the OTC market, especially news of an earnings announcement which tends to come right after market close.

[13] Consistent with the positive price effect, Lee (1992) shows that small traders are net buyers following both positive and negative earnings surprises.

Our paper makes several contributions to the literature on investor attention. First, we introduce a new, direct measure of institutional investor attention based on institutional investors' news searching and news reading activity. Importantly, because this measure is not limited to events associated with a firm's regulatory filings, it can capture a more broad set of events that may draw the attention of institutional investors, which allows us to examine its role across multiple types of news events. Second, because AIA is broadly analogous to the direct measure of retail attention from Google searches, an additional contribution lies in documenting the relation between the two types of attention. For example, we show that institutional attention leads retail attention. Moreover, we find that institutional investor attention is less constrained that that of retail investors. Finally, using two distinct types of information events, and controlling for a comprehensive set of other proxies for attention, we confirm that institutional attention plays an important role in the quick incorporation of information into asset prices.

Our paper also contributes to the broader literature that links news media to asset prices, including Tetlock (2007), Fang and Peress (2009), Loughran and McDonald (2011), Engelberg and Parsons (2011) and Gurun and Butler (2012), Peress (2014), Peress and Schmidt (2014) among others. Our results suggest that institutional attention is necessary for new information to be incorporated into prices on a timely basis. Our findings confirm the important role played by institutional investors in various other financial contexts.[14]

The remainder of the paper is organized as follows. Section 2 describes our samples and the construction of our AIA measure in detail. Section 3 examines factors that are related to AIA and compares AIA to retail attention measure. Section 4 studies the impact of AIA on asset prices following the announcements of firm earnings and analyst recommendation changes. Section 5 concludes.

[14] Among many examples, see Alti and Sulaeman (2012) for the case of seasoned equity offerings (SEOs), Bauguess, Cooney, and Hanley (2013) for the case of initial public offerings (IPOs), and Sulaeman and Wei (2014) for the case of cost of equity capital.

2. Data and Summary Statistics

2.1 Sample Construction

Bloomberg provide us with data which include transformed measures of news reading and news searching activity on Bloomberg's terminals. Based on data availability, our sample period ranges from February 2010 to June 2013. Following Da, Engelberg and Gao (2011), we start with the sample of Russell 3000 stocks. We then require the stocks in our sample to satisfy the following conditions: (1) have search and news reading information on Bloomberg terminals; (2) have a share code of 10 or 11 in the Center for Research in Securities Prices (CRSP) database; (3) have book-to-market information for the DGTW risk adjustment (Daniel, Grinblatt, Titman and Wermers, 1997). These conditions reduce our sample from 3,000 to 2,298 stocks. This is the main sample of our analysis ("Full Sample").

To arrive at the sample for analyzing earnings announcements, we start with the Full sample and require at least two analysts in IBES making earnings forecasts prior to the announcements. According to Battalio and Mendenhall (2005), measures of institutional trading following earnings announcements respond more to analyst-consensus-based earnings surprises rather than time-series-based earnings surprises. As a result, we will compute quarterly standardized unexpected earnings (*SUE*) relative to the analyst forecast consensus. The requirement for analyst forecasts reduces the sample of stocks from 2,298 to 1,952 ("*EarnAnn* sample") and yields a final sample of 18,543 earnings announcements.

To arrive at the sample for analyzing analyst recommendation change, we start with the Full sample and follow the filters in Jegadeesh and Kim (2010), Loh and Stulz (2011) and Kadan, Michaely, and Moulton (2013). In particular, we: (1) remove recommendation changes that occur on the same day as, or the day following, earnings announcements; (2) remove recommendation changes on days when multiple analysts issue recommendations for the same firm; (3) require at least one analyst to have issued a recommendation for the stock and revised the recommendation within 180 calendar days; (4) require at least two analysts, other than the revising analyst, to have active recommendations for the stock as of the day before the revision; (5) consider a recommendation to be active for up to 180 days after it is issued or until I/B/E/S indicates that the analyst has stopped issuing recommendations for that stock. After applying all these filters, we end up with 7,041 recommendation changes covering 1,376 stocks. This forms the subsample of our recommendation change analysis ("*RecChng* Sample").

Finally, institutional trading activity data is obtained from Ancerno Ltd. Ancerno is a widely-recognized transaction-cost consulting firm to institutional investors, and our database contains all trades made by Ancerno's base of clients. Ancerno data mainly includes trades by mutual funds and pension plans. A detailed explanation about Ancerno variables can be found in the Appendix of Puckett and Yan (2011). Our sample of transactions from Ancerno ends on March 31, 2013. As a result, the sample used in our trading analysis ends on that date.

2.2 Abnormal Institutional Attention (AIA) Measure

In order to construct their own measure of attention, Bloomberg records the number of times news articles on a particular stock are read by its terminal users and the number of times users actively search for news for a specific stock. Bloomberg then assigns the value of 1 for each article read and 10 for each news search. These numbers are then aggregated into hourly counts. Using the hourly counts, Bloomberg then creates a numerical attention score each hour by comparing the average hourly count during the previous 8 hours to all hourly counts over the previous month for the same stock. They assign a score of 0 if the rolling average is less than 80% of the hourly counts over the previous 30 days. Similarly, Bloomberg assigns a score of 1, 2, 3 or 4 if the average is between 80% and 90%, 90% and 94%, 94% and 96%, or greater than 96% of the previous 30 days' hourly counts, respectively. Finally, Bloomberg aggregates up to the daily frequency by taking a maximum of all hourly scores throughout the calendar day. Bloomberg provides us with these latter transformed scores, but does not provide us with the raw hourly counts or scores. Since we are interested in abnormal attention, and not just the level of attention, our abnormal institutional attention measure (AIA) measure is a dummy variable which receives the value of 1 if Bloomberg's daily max is 3 or 4, and 0 otherwise. This captures the right tail of the measure's distribution.[15] In other words, an AIA equal to one indicates the existence of institutional investor attention shock on that stock during that day.

[15] Our empirical results are similar if we exclude 3 from the definition of abnormal or include 2.

2.3 Other Variables

We compare institutional attention to retail attention. Following Da, Engelberg and Gao (2011), retail attention is measured using Google's Search Volume Index (*SVI*). Abnormal daily *SVI* (hereafter, "*LnAbnDSVI*") is calculated as the natural log of the ratio of *SVI* to the average of *SVI* over the previous month.

We obtain news coverage of our sample stock from RavenPack. "*LnNews*" is the log of 1 plus the number of news articles published on the Dow Jones newswire during the day.

We obtain the EDGAR server logs data from the SEC. Each day, for each stock, we calculate the total number of hits. To filter the data in order to exclude mass automated hits and mistakes, we follow the procedure used in Loughran and McDonald (2015).[16] Specifically, we exclude hits flagged as webcrawlers and exclude ip addresses that access more than 50 unique firms' filings in a given day. We also exclude retrievals of index files and hits resulting in errors (defined as log file status codes 300 or above). After filtering out these observations, we define "*EDGAR*" as the total number of hits on a given day. "*LnAbnEDGAR*" is then calculated as the natural log of the ratio of *EDGAR* to the average of *EDGAR* over the previous month. We use the WRDS CIK-CUSIP table to link the EDGAR data with CRSP.

Other variables used in our analysis are constructed from the standard databases: COMPUSTAT / CRSP / IBES. "*SizeInM*" is the firm market capitalization, rebalanced every June, in millions of dollars. "*LnBM*" is the firm natural logarithm of the firm book-to-market ratio, rebalanced every June following Fama-French (1992). "*SDRET*" is the firm daily standard deviation of return, calculated based on the previous 21 trading days. "*Ret*" is the CRSP daily return, in %. "*Dgtw*" is *Ret* minus the stock's daily benchmark portfolio daily return following Daniel, Grinblatt, Titman and Wermers (1997). "*Turnover*", is the daily turnover. "*Dvol in M*" is the daily dollar trading volume, in millions of dollars. "*AbnVol*" is the stock's abnormal trading volume calculated following Barber and Odean (2008) as the stock's daily volume divided by the previous 252-day average trading volume. "*HLtoH*" is the ratio between the stock's daily high-and-low price difference and the daily high price. "*LnNumEst*" is the log of 1 plus the number of analysts covering the stock. "*52HighDum*" ("*52LowDum*") is a dummy variable which receives

[16] All of our results are robust to using the filters described in deHaan, Shevlin, and Thornock (2015).

the value of 1 if the stock's price beat its 52-week high (low) price and 0 otherwise. "*AdvExpToSales*" is the firm advertising expenses to sales.

2.4 Summary Statistics

Table 1 provides summary statistics of our full sample and the two subsamples used for the earnings announcement and recommendation change analysis. Panel A shows that AIA frequency is 0.098 in the full sample, suggesting that the average stock in our sample experiences institutional attention shocks on 9.8% of all trading days.[17]

AIA frequency increases to 0.567 for the *EarnAnn* sample, suggesting that 56.7% of the announcement days coincide with an institutional attention shock. This is not surprising as earnings announcements are likely to attract institutional investor attention. At the same time, we note that not all earnings announcements trigger institutional attention shocks. This heterogeneity is important and allows us to study the impact of institutional attention on asset prices after controlling for the magnitude of earnings surprise.

AIA frequency is slightly lower at 0.447 for the *RecChng* Sample, suggesting that 44.7% of the recommendation change days are associated with institutional attention shocks. One difference between earnings announcements and recommendation changes is that the former are usually pre-scheduled so institutional investors can optimally allocate more attention to the announcement day.

Exploring other stock characteristics across the three samples indicates that these are not small firms. The average (median) size is around 5.5 (1) Billion. Naturally, the firms in the *RecChng* sample are larger due to our recommendation filters which require at least 3 active analysts covering the firm. Not surprisingly, trading volume and intraday volatility are higher during the *EarnAnn* and *RecChng* announcement days. On average, institutional holdings make up around 60 to 70 percent of shares outstanding, consistent with the well-documented increase in institutional holdings over time. The number of analysts covering a stock is 9, on average, and

[17] Because AIA is calculated using the maximum hourly attention throughout the day, this number need not be the 6% that the 94% cutoff may suggest.

is naturally higher in the *RecChng* sample given the additional filters used in creating that sample. The average absolute value of the earnings surprise (change in analyst recommendation) is 2.68 (1.34).

Finally, Panel A also reports the sample statistics of the EDGAR and DSVI attention measures. On average, there are about 42 hits (excluding robots) on EDGAR on a given day for stocks in the full sample. This number increases to about 80 in the *EarnAnn* sample and 68 in the *RecChng* sample. These patterns are also evident in the abnormal measure. On average, *LnAbnEDGAR* increases by around 66% in *EarnAnn* sample and around 10% in *RecChng* sample.[18] The differential in abnormal attention across these two events may not be surprising since earnings announcements are scheduled events, while changes in analyst recommendations are typically less predictable. Strikingly, our measure of abnormal retail attention, *LnAbnDSVI*, only presents an increase of around 2% on earnings announcement days and is essentially 0 on days with recommendation changes. Our subsequent analysis confirms that AIA responds to these news events much faster than *LnAbnDSVI*.

In Panel B, we present sample averages conditioning on AIA for the three samples. The panel shows that across all three samples, absolute returns, turnover, dollar trading volume and intraday price movements are higher during attention shocks. The average number of analysts is also higher which is consistent with greater information processing. Interestingly, both the magnitude of the earnings surprise and magnitude of the changes in analyst recommendation are quite similar across the AIA subsamples. This suggests that the magnitude of the surprise is not the primary driver behind abnormal institutional investor attention. Finally, activity is higher on both EDGAR and Google, suggesting that AIA is contemporaneously positively correlated with these attention measures. In Table 2, we examine these relations in a multivariate regression framework.

[18] Since *LnAbnEDGAR* and *LnAbnDSVI* are calculated as a log of a ratio they have an interpretation of a percentage change.

3. What Drives Institutional Attention?

What factors are related to institutional attention shocks? How is institutional attention related to institutional trading? How are institutional attention shocks related to retail attention shocks? We examine these questions in this section.

3.1 Determinants of Institutional and Retail Abnormal Attention

We explore a wide set of variables which are associated with our abnormal institutional attention shocks. For comparison, we also explore how these variables are associated with abnormal retail attention shocks.[19] To examine these determinants, we conduct Probit panel regressions in Panel A of Table 2, using daily AIA as the dependent variable, and OLS panel regressions in Panel B or Table 2 using daily *LnAbnDSVI* as the dependent variable.

Motivated by the example of Overstock.com in Figure 1, we focus on five categories of variables. In column (1), we examine variables that are related to news. They include log number of news articles on that stock that day (*LnNews)*, and dummy variables to indicate earning announcements and recommendation changes. These news-related variables have the highest explanatory power of institutional attention shocks with a pseudo *R*-squared of 7.19%. All three news variables are highly significant.[20]

In column (2), we examine variables that are related to equilibrium outcomes of trading on that day. They include absolute DGTW-adjust return (*AbsDgtw*), abnormal trading volume (*AbnVol*), measure of intraday volatility (*HLtoH*), and dummy variables indicating if the current price beats 52-week high or low (*52 High Dum* and *52 Low Dum*). Many of these equilibrium outcomes have been used as proxies for investor attention (see Gervais, Kaniel, and Mingelgrin (2001), Barber and Odean (2008), and Hou, Peng, and Xiong (2009) among others). The regression coefficients reported to column (2) confirm that these equilibrium outcomes are

[19] See also Drake, Roulstone and Thornock (2012) who explore retail attention in the sample of S&P 500 stocks during 2005-2008.

[20] The pseudo *R*-squared from a univariate regression of AIA on *EarnAnnDum* (*RecChngDum*) is 1.74% (0.30%), consistent with the fact that earnings announcements are pre-scheduled.

related to institutional attention shocks as well. Nevertheless, equilibrium outcomes have lower explanatory power compared to news (pseudo *R*-squared is 2.49%).

In column (3), we examine various firm characteristics. We find that larger firms with more volatile returns, more analyst coverage are associated with significantly more institutional attention shocks on average. The results are similar to those documented in the prior literature using other measures of investor attention (see Grullon, Kanatas, and Weston (2004), Da, Engelberg and Gao (2011) and Liu and Peng (2015)). On the other hand, controlling for the other variables, we do not find a significant relation between advertisement expenses and institutional attention. Strikingly, the percentage of institutional holdings does not seem to have a significant relation with AIA. Altogether, firm characteristics have a combined pseudo *R*-squared of 4.89%.

In column (4), we include the other direct measures of attention, *LnAbnEDGAR* and abnormal retail attention, *LnAbnDSVI*. Both measures are positively related to AIA. Strikingly, the pseudo *R*-squared is only 1.63%. One possible reason is that the EDGAR measure is limited to a subset of mandatory filings while AIA captures abnormal institutional attention to a broader set of news events. Indeed, Drake, Roulstone, and Thornock (2015) find that 86% of the users accessing EDGAR do so infrequently and only around 2% of the users access EDGAR actively during a given quarter. Similarly, retail attention is more likely to be reactive (to occurrence of news) rather than proactive as a result of optimal attention allocation decision.

In column (5), we document strong within-week seasonality associated with institutional attention. The likelihood of an institutional attention shocks decreases monotonically from Monday to Friday. For example, a stock is 25% less likely to have an attention shock on a Friday compared to a Monday, consistent with the results in DellaVigna and Pollet (2009). The total explanatory power of the seasonality effect is low with a pseudo *R*-squared of only 0.26%.

Finally, in column (6), we include all five categories of explanatory variables and obtain a pseudo *R*-squared of 12.05%. The result suggests that existing proxies of investor attention explain a small fraction of institutional attention shocks. Of course, the low pseudo *R*-squared could be partially driven by measurement errors in AIA. Despite these errors, our subsequent analysis confirms that component of AIA orthogonal to other investor attention proxies continue to exert significant impact on asset prices.

Next, in order to better understand differences in what drives institutional attention and retail attention, we estimate OLS panel regressions of abnormal daily SVI on the same sets of variables as in the previous analysis and present results in Panel B.

Column (1) shows that the relation between *LnAbnDSVI* and news-related measures is qualitatively similar to what we find with institutional attention. However, with an adjusted *R*-squared of only 0.08%, these variables explain very little of the variation in retail attention. In fact, this is true of all 6 specifications in Panel B.

Results using the equilibrium outcome measures in Column (2) look relatively similar to those for AIA, with two exceptions. Instead of a positive relation between the intraday price range and attention, there is a negative relation. Additionally, while retail attention is likely to be higher when a stock hits its 52-week high, the same is not true of the 52-week low.

Column (3) of Panel B shows that abnormal institutional and retail attentions behave differently with respect to firm characteristics. While larger firms are more likely to draw both types of attention, the only additional variable with a statistically significant relation to abnormal retail attention is *SDRET*. In contrast to the case of institutional attention, less volatile stocks draw retail attention.

Column (4) shows that both AIA and *LnAbnEDGAR* are positively related to retail attention, though the adjusted *R*-squared is only 0.10%.

As was the case with AIA, there is within-week seasonality in *LnAbnDSVI.* Column (5) shows that retail attention is significantly lower on Friday than on Monday. Abnormal retail attention on Tuesday is actually greater than on Monday, though this result is only statistically significant at the 10% level.

Finally, in Column (6), we regress abnormal retail attention on all 5 categories of variables. Results are generally similar to those in the first five columns. However, *EarnAnnDum*, *RecChngDum*, and *HLtoH* are no longer statistically significant at 5%. Jointly, these variables explain less than 0.4% of the variation in the direct measure of abnormal retail attention. In similar analysis in Da, Engelberg and Gao (2011), a set of attention related variables explains about 3% of the variation in abnormal SVI at weekly frequency. Variations in daily abnormal SVI seem harder to explain.

3.2 Institutional Attention, Retail Attention, and Abnormal Trading Volume

Investor attention often triggers trading. If AIA truly measures abnormal institutional attention, we would expect there to be a strong contemporaneous correlation between AIA and investor trading. Moreover, we would expect the impact of AIA on trading to be the most pronounced for institutional investors. By contrast, we wouldn't expect to find similar patterns using abnormal retail attention. We examine this point in Table 3. In particular, we calculate two measures of abnormal attention using Ancerno and CRSP. Abnormal institutional trading volume (*Ancerno-AbnVol*) is calculated as the stock's Ancerno daily volume divided by the previous 8-week average Ancerno trading volume. As a benchmark, abnormal total trading volume (*CRSP-AbnVol*) is calculated as the stock's CRSP daily volume divided by the previous 8-week average CRSP trading volume.

We regress these abnormal trading volume measures on AIA (Panel A) and *LnAbnDSVI* (Panel B). The specifications are similar across the both panels. To make the analysis comparable across the two measures, we transform *LnAbnDSVI* into a dummy variable ("DSVIDUM") which receives the value of 1 if *LnAbnDSVI* is in the top decile on a given day and 0 otherwise. To top DSVIDUM decile captures the 10% of the cases with highest attention which are comparable in frequency to our AIA measure (see Table 1).

The panels include six regression specifications, where we sequentially add the five sets of control variables associated with institutional attention from Table 2. For each measure, we report the first difference (i.e., the difference in coefficients between AIA=0 and AIA=1 in Panel A and DSVIDUM=0 and DSVIDUM=1 in Panel B), together with the difference in difference ("Diff-in-Diff") and its statistical significance. For example, "*CRSP-AbnVol-Diff*" coefficient estimate in Panel A captures the additional response of CRSP's abnormal volume to a shock in AIA.

Focusing on the final column of Panel A, where we include all five sets of control variables, we find a strongly significant coefficient of 0.240 on *CRSP-AbnVol*. The result suggests that an institutional attention spike is accompanied with a 24.0% increase in abnormal total trading volume, relative to the case of AIA=0. The coefficient on *Ancerno-AbnVol* is larger with a value of 0.318, confirming that the same institutional attention spike correlates much more with abnormal institutional trading volume. The difference between the two coefficients (i.e., *Diff-in-Diff*) of 0.079 is highly significant with a t-statistic of 3.14.

Recall that Ancerno data primarily consist of trades by mutual funds and pension plans who are not the most active institutional investors. We would expect an even stronger link between AIA and institutional trading were we to focus on trading by hedge funds, for example. Moreover, the aggregate CRSP volume measure includes both individual and institutional investors, which biases the test against finding a significant difference. Thus, choosing a benchmark made up of only retail investors' trades should magnify our findings.

Overall the evidence in Panel A supports the notion that AIA measures attention shocks among institutional investors. In Panel B, we examine whether there is a similar relation between retail attention and institutional trading using the same analysis in Panel A.

Focusing on abnormal retail attention, Panel B of Table 3 clearly presents a different pattern. First, although the coefficients on the CRSP abnormal volume measure are positive in all six specifications, the magnitudes are only around 1/10 of the magnitude presented in Panel A. Moreover, the coefficients on the ANCERNO sample are not significant after controlling for firm characteristics. Finally, there is no statistically significant difference in differences in the impact on the two types of abnormal trading volume based on retail attention for any of the six specifications.

To summarize, we find that institutional attention measured using AIA is unique. While it is related to existing proxies of investor attention in an intuitive way, a large fraction of AIA remains unexplained even with existing proxies combined. Equipped with our AIA measures, we can then directly examine how institutional investor attention affects asset prices in response to information. This is the focus of our analysis in the next section.

4. Institutional Attention and Price Response to Information

The announcements of firm earnings and analyst recommendation changes are both important value-relevant information events. A voluminous literature has documented post-announcement price drift following both events. In other words, investors underreact to both announcements on average. In this section, we examine whether institutional attention on the announcement day facilitates information incorporation and alleviates price under-reaction to news.

4.1 Earnings Announcements

We examine the impact of institutional attention on earnings-announcement-window returns and post-earnings-announcement drifts using panel regressions. The results are reported in Table 4. If institutional investors facilitate information incorporation through attention and information processing, we would expect such information to be incorporated on day-0. More importantly, that would result in a less (if any) drift over subsequent days. Since many factors (observable and unobservable) can affect day-0 return, it is virtually impossible to provide direct evidence of a causal relation on day-0. However, less drift going forward would be clear evidence of information incorporation on day-0. Accordingly, in this section we provide clear evidence of less (if any) drift in stocks with high abnormal attention. Regarding the impact of AIA on day-0 returns, we discuss three potential explanations and argue that a causal effect of AIA on day-0 is the most likely explanation given the full set of our results. Finally, we show that the impact of retail attention is completely different and consistent with previous findings.

We match the timing of returns, AIA, and announcements by shifting earnings announcements that occur after trading hours to the following day. The dependent variables are day-0 *DGTW* risk adjusted return and *t*+1 to *t*+40 risk adjusted cumulative returns where day *0* represents the earnings announcement day. In panel A, the main dependent variables are: *AIA*, the quarterly standardized unexpected earnings (*SUE*), and their interaction term (*SUE_AIA*). To the extent that *SUE* controls for the fundamental information content at the announcement, the coefficient on *SUE_AIA* identifies the incremental impact of having institutional attention. We also include a comprehensive set of control variables that might affect returns.

The positive and significant coefficients on *SUE* confirm both the day-0 impact of the announcement and the existence of post-earnings-announcement drifts (PEAD). Stock prices react strongly to earnings surprises on the announcement day and continue to drift in the direction of *SUE* over the next 40 trading days. The coefficients on the interaction term *SUE_AIA* suggest that institutional attention facilitates information incorporation at the announcement and alleviates future drift. The coefficient of 0.0019 on the announcement day suggests that when institutional investors pay attention, stock price reaction is 19bps larger for a one unit change in *SUE*. Note that this additional price response is consistent with our prior.

Next, focusing on the drift, we find that the coefficients on the interaction term *SUE_AIA* are negative and significant starting from day *t*+1 up to day *t*+40. Strikingly, the magnitude of the coefficient is about -0.0015 by the end of *t*+40, which is close to the coefficient on SUE in absolute term by *t*+40 (0.0020). In other words, when institutional investors pay more attention at the earnings announcement, there is almost no PEAD at all.

Having established that abnormal institutional attention is associated with weaker drift, we return to the announcement return findings. In particular, there are three possible explanations for the day-0 result. First, institutional investor attention facilitates information incorporation, and as a result, price underreaction to information is smaller. Alternatively, the relation between attention and the large price response on the announcement day could be driven by endogeneity (both attention and the market respond without being directly related. In other words, holding *SUE* constant, institutional investors endogenously allocate more attention to the types of announcements that have greater price impact) or reverse causality (larger price reaction on the announcement day actually triggers institutional attention).[21] We address the potential reverse causality explanation in Panel C of Table 4. Importantly, although we cannot address the potential endogeneity explanation in our setting, the distinction between the three explanations is that only the first would predict a smaller price drift going forward, which we clearly find.[22]

[21] We use the term endogenous to refer to the relation between attention and return, not to investors' choices of attention allocation.

[22] One can rightly argue that while the magnitude of the earnings surprise is the same, unobservable factors which are associated with the content of the announcement may differ and could attract abnormal institutional search. One potential explanation could be that abnormal AIA captures disagreement. That is, there is ambiguity about the news which non-randomly triggers attention. However, although disagreement may explain abnormal searches, it cannot

Our main results thus far in this subsection are nicely summarized in Figure 2. To construct this figure, we use the estimated regression coefficients from Panel A of Table 4 and the conditional means of each group of interest (the four groups are based on the intersection between Positive SUE, negative SUE, AIA=0 and AIA=1). Figure 2 shows that the well-documented PEAD comes almost exclusively from announcements with limited institutional investor attention. When institutional investors fail to pay sufficient attention, price initially underreacts to information (as evident in panel A), resulting in a drift (as evident in Panel B). Thus, our results offer direct support that limited investor attention, especially those from institutional investors, is the driving force behind PEAD.

Recall that Table 2 documents a significant link between AIA and measures of equilibrium outcomes such as abnormal trading volume, return volatility, average spread, and size. The prior literature has used some of these equilibrium outcomes as investor attention proxy to study PEAD. For example, Hou, Peng, and Xiong (2009) documents that stocks with higher trading volume experience smaller post-earnings-announcement-drift. The advantage of our AIA measure is twofold. First, it allows us to focus on institutional investor attention which is more important for driving permanent price change. Second, while trading volume is an equilibrium outcome that reflects many economic forces other than investor attention, AIA directly reveals institutional investor attention. Table 2 also shows that AIA is also related to more direct measures of attention. In particular, Drake, Roulstone, and Thornock (2015) find that more hits on EDGAR on the day of, and the day after an earnings announcement are related to a smaller PEAD. Although *LnAbnEDGAR* explains only a small part of the variation in AIA, we directly control for *LnAbnEDGAR* in Panel A. Additionally, in an untabulated set of results, we include both *LnAbnEDGAR* and its interaction with *SUE* directly in the regression with AIA. We find that the coefficients on AIA and its interaction with *SUE* are qualitatively unchanged. More importantly, the coefficients on the interaction of the EDGAR measure with *SUE* are no longer statistically significant. Including *LnAbnDSVI* produces similar results.

To alleviate any remaining concerns, we confirm the incremental value of AIA in a two-step exercise. In the first step, we orthogonalize AIA using all AIA determinants explored in

---

explain the fact the prices fully adjust and there is virtually zero drift going forward. On the other hand, agreement might explain a full price adjustment but is not consistent with abnormal searching activity.

Table 2. In particular, these variables include equilibrium outcomes, firm characteristics, news and analyst coverage, institutional holdings, seasonality and other direct attention measures. The residual is the unexpected AIA which, by construction, captures abnormal institutional attention unrelated to equilibrium outcomes or other proxies for attention. The pseudo R-squared of the first stage regression captures 27.75% of the variation in AIA. In the second step, we replace AIA with the residual AIA and find very similar results as reported in panel B.[23] Importantly, residual AIA continues to predict a significant reduction in PEAD and stocks with high unexpected AIA have almost zero PEAD.

In our final set of tests in this subsection, we address the potential reverse causality explanation. In particular, because AIA is measured daily from 12am to 12am while return is measured from 4pm to 4pm (close-to-close), it's possible that announcement-day returns lead to attention, and not vice-versa. For example, consider a large earnings surprise announced in the morning before the market opens on day *t*. The earnings surprise is almost fully incorporated into price on day *t*, resulting in a large announcement-day return and little price drift going forward. The large earnings-announcement day return then causes institutional investor to pay abnormal attention after market-close on day *t*.

To rule out a reverse causality explanation, we focus on the subset of earnings announcements occurring between 4pm and 12am after the market has closed on day *t*-1.[24] Roughly 50% of our earnings announcements sample events (9,308 firm-quarter observations, 50.4%) take place between 4pm and 12am (consistent with Michaely, Rubin, and Vedrashko, 2014). If we observe AIA = 1 on day *t*-1 for these earnings announcements, the institutional attention cannot be caused by the earnings-announcement day return. Panel C reports the results where we repeat our regression analysis in Panel A for this reduced sample using AIA on day *t*-1. Our results are robust in this sample.

Finally, one may argue that higher return on day t – which is associated with AIA=1 – mechanically causes less drift going forward, which would make the attention-drift relation

[23] Note that this is equivalent to including variable interactions terms in a full regression.

[24] We acknowledge that trading does occur in OTC markets after market close. However, trading volume is by far smaller and less concentrated relative to the trading volume at the opening on day-*t*. Thus, it is fair to assume that institutional investors are more likely to notice news than prices in the OTC market, especially news of an earnings announcement which tends to come right after market close.

mechanical. This is unlikely for a few reasons. First, Chan, Jegadeesh, and Lakonishok (1996) document that higher (absolute) earnings-announcement-window-return predicts stronger, not weaker, post-earnings-announcement drift on average. Second, in untabulated results, we directly control for announcement-day return in the regressions when examining post-announcement drifts. We confirm that controlling for the return on announcement day t barely changes the impact of AIA on post-earnings announcement returns from day t+1 up to day t+40.

4.2 Analyst Recommendation Changes

In this subsection, we study price reaction during and after analyst recommendation changes using similar panel regressions. Similar to earnings announcements, we shift recommendations that occur after trading hours by one trading day. We focus on day-0 and subsequent ten trading days. The results are reported in Table 5. As detailed in Section 2.1, in constructing the *RecChng* sample, we only keep recommendation changes with unambiguous information content that is different from that in earnings announcements. In other words, our *RecChng* sample contains additional information events that are relatively independent from those in the *EarnAnn* sample. This additional set of tests provides strong evidence that our results are not specific to earnings announcements.

The regressions in Table 5A (5B) are similar to those in Table 4A (4B) except that we replace *SUE* with *RecChng* which measures the change in analyst recommendations. Specifically, *RecChng* ranges from -4 to 4, where a positive (negative) number refers to an upgrade (a downgrade).

The positive and significant coefficients on *RecChng* confirm that stock prices react to recommendation changes strongly on the announcement day and continue to drift in the direction of *RecChng* for the next 10 trading days.

The negative coefficients on the interaction term *RecChng_AIA* again suggest that institutional attention facilitates information incorporation at the announcement and alleviates future drift. In particular, the positive coefficient of 0.0081 on the announcement day suggests that when institutional investors pay attention, stock price reacts by 81 bps more for a one-notch change in the recommendation.

Focusing on the drift, the coefficients on the interaction term are negative and significant from t+1 to t+10. By the end of t+10, the coefficient is about -0.0027 by the end of t+10, equal to the corresponding coefficient on *RecChng* in absolute term (0.0027). In other words, when institutional investors pay more attention to analyst recommendation change, there is no post-announcement drift.

Similar to Figure 2, our results are nicely summarized in Figure 3. To construct this figure, we use the estimated regression coefficients from Panel A of Table 6 and the conditional means of each group of interest (the four groups are based on the intersection between Positive REC, negative REC, AIA=0 and AIA=1). Figure 3 confirms that price drift following a recommendation change comes almost exclusively from announcements with limited institutional investor attention. When institutional investors fail to pay sufficient attention, price initially underreacts to information (as evident in panel A), resulting in a drift (as evident in Panel B). The patterns in Figure 3 are very similar to those in Figure 2. The patterns remain the same when residual AIA is used in Panel B.

Similar to Panel 4B, residual AIA produces consistent results.[25] As for an analysis using data after the market close, in contrast to earnings announcements, the vast majority of recommendation changes in our sample take place before the market has closed.[26] While this prevents us from focusing directly on after-hour recommendation changes, in untabulated results, we find that including the announcement day return as an independent variable has no impact on the relation between AIA and future drift.

Finally, exploring the predictive power of EDGAR around analyst recommendation changes shows that EDGAR cannot explain the drift (even without controlling for AIA). This reveals the importance of AIA as a direct measure if institutional investor attention. In contrast to EDGAR which is limited to a set of firms' regulatory filings, AIA (which is based on direct news reading and searching) allows exploration of a broader set of information events for which there

[25] The Pseudo R-Squared from the first stage regression is 12.2% compared to the pseudo R-Squared of 27.75% found in the earnings announcements sample.
[26] Less than 15% of our 7,041 changes occur between 4pm and 12am.

may be no associated SEC filing. Consequently, using AIA in the setting of analyst recommendation changes delivers strikingly similar conclusions to those found using earnings surprises.

4.3 The Impact of Retail Attention

So far, our evidence suggests that institutional attention facilitates price discovery and alleviates under-reaction to news. Table 2 shows that institutional attention and retail attention are often correlated contemporaneously. It is natural to ask whether retail attention plays a similar role. We examine this important question in Table 6. We repeat the panel regressions analysis from Tables 4A and 5A after replacing AIA with *StndLnAbnDSVI* (the standardized abnormal search volume index).[27]

In contrast to AIA, the coefficients on *StndLnAbnDSVI* are almost always positive and increasing (though not significant) following earnings announcements and recommendation changes. In other words, regardless of the content of the news, retail attention almost always results in positive price pressure, consistent with the evidence in the prior literature (see Barber and Odean, 2008, and others).

More importantly, the coefficients on the interaction terms between *StndLnAbnDSVI* and *SUE* or *RecChng* are always positive and in some cases statistically significant. Thus, it is clear that having more retail attention does not alleviate post-announcement price drifts.

To summarize, while institutional attention facilitates permanent price discovery, retail attention only results in a positive price effect. In this regard, the impact of institutional attention and retail attention on asset prices is fundamentally different.

[27] For ease of interpretation, *StndLnAbnDSVI* is standardized to reflect the impact of 1 standard deviation of *LnAbnDSVI* on returns. This was not necessary for AIA which is an indicator variable.

4.4 The Relation between Institutional and Retail Abnormal Attention

Having documented how the impact of attention on the incorporation of information differs across institutional and retail investors, we next investigate more directly how institutional and retail attention shocks are related to each other using our pooled sample of earnings announcements and analyst recommendation changes events. We examine this question with two sets of lead-lag regressions reported in Table 7. Note that the two measures are analogous since both measures focus on active investor searches.

In our first set of tests (Specifications 1-3), we regress AIA on its lags and on lags of *LnAbnDSVI*. Since AIA takes only the value of 0 or 1, we estimate the relation between AIA (as dependent) and *LnAbnDSVI* using Probit models.[28] To distinguish differences in attention across the two types of events, we include the variable *EarnDum*, which capturers the incremental level of attention during earnings announcements. Additionally, following the *investor distraction* hypothesis of Hirshleifer, Lim and Teoh (2009), we include the variable "*LnNumEvents*", which captures the (log of) the total number of earnings announcements and analyst recommendation changes occurring on day *t*.

We immediately observe that AIA is persistent. We find positive and significant autocorrelations on all five lags of AIA. Moreover, in specification (1), the coefficient on the first lag of *LnAbnDSVI* is positive and statistically significant. However, once more lags of both measures are included, the latter coefficient is smaller and no longer significant. The coefficients on EarnDum are positive and statistically significant. This shouldn't come as a surprise since these events are pre-scheduled. Interestingly, the positive but statistically insignificant coefficient on *LnNumEvents* suggests that the frequency of simultaneous news events does not represent a constraint to institutional investor attention.

In the second set of tests (Specifications 4-6), we regress *LnAbnDSVI* on its own lags and on lags of AIA. We use OLS panel regressions, as *LnAbnDSVI* is a continuous variable. Similar to AIA, *StndASVI* is also positively autocorrelated – but only for the first lag. Interestingly, the coefficients on lagged AIAs are positive and many are statistically significant. Together with the results from specification (3), this provides strong evidence that institutional attention shocks

[28] Using linear probability regressions instead of Probit models yields qualitatively similar results.

lead retail attention shocks. These collective findings are not surprising as institutional investors have greater resources and stronger financial incentive to monitor the market and are more likely to pay attention to news and react immediately.

Along these lines, the regressions also provide some evidence which suggests that retail attention is reduced when there are many simultaneous news events. Thus, retail investor attention is more constrained than institutional investor attention, suggesting that the *investor distraction* hypothesis is more relevant for retail investors than for institutional investors.

5. Conclusion

To our best knowledge, we propose the first broad, direct measure of abnormal institutional investor attention. Our abnormal institutional investor attention measure (AIA) is based on the news-searching and news-reading frequency for specific stocks on Bloomberg terminals which are used almost exclusively by institutional investors. We find AIA to be related to but different from other investor attention proxies. In addition, AIA is highly correlated with measures of abnormal institutional trading contemporaneously.

More importantly, AIA enables us to directly contrast institutional attention with retail attention measured using Google search frequency. We find that institutional attention responds to major news events faster, triggers more trading, and is less constrained compared to retail attention.

Since institutional investors are more likely to react to news immediately and become the marginal investors who act, institutional investor attention is crucial in facilitating the incorporation of new information into asset prices. Indeed, we find that the well-documented price drifts following both earnings announcements and analyst recommendation changes come only from announcements where institutional investors fail to pay attention according to our measure. In sharp contrast, retail attention almost always results in a positive price pressure that is eventually reverted.

Earnings announcements and analyst recommendation changes are just two examples of important information events. It will be interesting to use AIA to examine the differential impact of institutional and retail attention on market reaction to other corporate events such as IPOs, M&As, product launches, and dividend cuts, we leave these and other exciting applications of AIA for future research.

References:

Alti A. and J. Sulaeman, 2012, When do high stock returns trigger equity issues? *Journal of Financial Economics* 103, 61-87.

Ball, R., & Brown, P.,1968, An empirical evaluation of accounting income numbers, *Journal of accounting research*, 159-178.

Barber, B. M., and T. Odean, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies* 21, 785–818.

Battalio, H. R., and R. R. Mendenhall, 2005, Earnings Expectations, investor trade size, and anomalous returns around earnings announcements, *Journal of Financial Economics* 77, 289-319.

Bauguess, S., J. Cooney, K. W. Hanley, 2013, Investor demand for information in newly issued securities, working paper.

Boehmer, E., and J. J. Wu, 2013, Short selling and the price discovery process, *Review of Financial Studies* 26, 287–322.

Boudoukh, J., R. Feldman, S. Kokan and M. Richardson, 2013, which news moves stock prices? A texual analysis, working Paper.

Chan, L., N. Jegadeesh, and Lakonishok, 1996, "Momentum Strategies," *Journal of Finance* 51, 1681-1713.

Cohen, L., and Frazzini, A., 2008, Economic links and predictable returns, *Journal of Finance*, *63*(4), 1977-2011.

Chemmanur, T., and A. Yan, 2009, Advertising, attention, and stock returns, Working paper, Boston College and Fordham University.

Da, Z., J. Engelberg, and P. Gao, 2011, In search of attention, *Journal of Finance* 66, 1461-1499.

Da, Z., U. G. Gurun, and M. Warachka, 2014, Frog in the pan: continuous information and momentum, *Journal of Finance*, forthcoming.

DeHaan, E., T. Shevlin, and J. Thornock, 2015, Market (in)attention and the strategic scheduling and timing of earnings announcements, *Journal of Accounting and Economics* forthcoming.

DellaVigna, S., and Pollet, J. M., 2009, Investor inattention and Friday earnings announcements, *Journal of Finance*, *64*(2), 709-749.

Daniel, K. D., M. Grinblatt, S. Titman, and R. Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* 52, 1035–1058.

Drake M., D. Roulstone, and J. Thornock, 2012, Investor Information Demand: Evidence from Google Searches Around Earnings Announcements, *journal of Acounting Research* 50(4), 1001-1040.

Drake M., D. Roulstone, and J. Thornock, 2015, The determinants and consequences of information acquisition via EDGAR, *Contemporary Accounting Research* forthcoming.

Engelberg, J. E., and Parsons, C. A. 2011, The causal impact of media in financial markets, *Journal of Finance*, 66(1), 67-97.

Fang, L., and J. Peress, 2009, Media coverage and the cross-section of stock returns, *Journal of Finance* 64(5), 2023-2052.

Fama, E. F., and K. R. French , 1992. The cross-section of expected stock returns, *Journal of Finance*, *47*(2), 427-465.

Gervais, S., R. Kaniel, and D. H. Mingelgrin, 2001, The high-volume return premium, *Journal of Finance* 56, 877–919.

Grullon, G., G. Kanatas, and J. P. Weston, 2004, Advertising, breath of ownership, and liquidity, *Review of Financial Studies* 17, 439–461.

Gurun, U. G., and A. W. Butler, 2012, Don't believe the hype: Local media slant, local advertising, and firm value, *Journal of Finance* 67(2), 561-598.

Hendershott, T., S. X. Li, A. J. Menkveld, and M. S. Seasholes, 2013, Asset price dynamics with limited attention, working paper.

Hirshleifer, D., and S. H. Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337–386.

Hirshleifer, D., Lim, S. S., and Teoh, S. H, 2009, Driven to distraction: Extraneous events and underreaction to earnings news, *Journal of Finance*, *64*(5), 2289-2325.

Hirshleifer, D., Lim, S. S., and Teoh, S. H., 2011, Limited investor attention and stock market misreactions to accounting information, *Review of Asset Pricing Studies* 1 (1): 35-73.

Hou, K., L. Peng, and W. Xiong, 2009, A tale of two anomalies: The implications of investor attention for price and earnings momentum, Working paper, Ohio State University and Princeton University.

Jegadeesh, N., and W. Kim, 2010, Do analysts herd? An analysis of recommendations and market reactions. *Review of Financial Studies* 23 (2): 901-937.

Kadan, O., R. Michaely, and P. Moulton, 2013, Who Trades on and Who Profits from Analyst Recommendations? working paper.

Kahneman, D., 1973. Attention and Effort. Prentice-Hall, Englewood Cliffs, NJ.

Lee, C. MC., 1992, Earnings news and small traders: An intraday analysis. *Journal of Accounting and Economics*, *15*(2), 265-302.

Lee, C. MC., P. Ma, and C. Y. Wang, 2015, Search-based peer firms: Aggregating investor perceptions through internet co-searches, *Journal of Financial Economics* 116, 410-431.

Liu, H., and L. Peng, 2015, Investor Attention: Seasonal Patterns and Endogenous Allocations, Working Paper.

Livnat, J., and Mendenhall, R. R., 2006, Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research*, *44*(1), 177-205.

Loh, R. K. 2010. Investor Inattention and the Underreaction to Stock Recommendations. Financial Management 39:1223–51.

Loh, R. K., and R. M. Stulz, 2011, When are analyst recommendation changes influential? *Review of Financial Studies* 24, 593-627.

Lou, D., 2014, Attracting investor attention through advertising, *Review of Financial Studies*, 27 (6): 1797-1829.

Loughran, T., and B. McDonald, 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance* 66(1), 35-65.

Loughran, T., and B. McDonald, 2015, Information decay and financial disclosures, working paper.

Madsen, J., and M. Niessner, 2014, Is investor attention for sale? the role of advertising in financial markets, Working paper.

Michaely, R., A. Rubin, and A. Vedrashko, 2014, Corporate governance and the timing of earnings announcements, *Review of Finance* 18(6), 2003-2044.

Peng, L., 2005, Learning with information capacity constraints, *Journal of Financial and Quantitative Analysis* 40, 307-329.

Peng, L., and W. Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563–602.

Peress, J., 2014, The media and the diffusion of information in financial markets: Evidence from newspaper strikes, *Journal of Finance* 69(5), 2007-2043.

Peress, J., and D. Schmidt, 2014, Glued to the TV: the trading activity of distracted investors, working paper.

Puckett, A., and X. S. Yan, 2011, The interim trading skills of institutional investors, *Journal of Finance* 66(2), 601-633.

Sims, C. A., 2003, Implications of rational inattention, *Journal of Monetary Economics* 50, 665-690.

Seasholes, Mark S., and Guojun Wu, 2007, Predictable behavior, profits, and attention, *Journal of Empirical Finance* 14.5: 590-610.

Stickel, Scott E., 1995, The anatomy of the performance of buy and sell recommendations, Financial Analysts Journal 51, 25-39.

Sulaeman, J. and C. Wei, 2014, Institutional Presence, working Paper.

Tetlock, P. C., 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62(3), 1139-1168.

Womack, K. L.,1996, Do brokerage analysts' recommendations have investment value?. *Journal of finance*, 137-167.

Yuan, Y, 2015, Market-wide attention, trading, and stock returns, *Journal of Financial Economics* 116 (3), 548-564.

Table 1 – Summary Statistics of Abnormal Institutional Attention (*AIA*) and Other Selected Variables

The table reports summary statistics of Abnormal Institutional Attention measure ("*AIA*") from Bloomberg (hereafter, "*AIA*") and other selected variables from February 2010 to June 2013. Our initial sample includes all Russell 3000 stocks with CRSP share codes 10 and 11, *AIA* information and book-to-market information for the DGTW risk adjustment (Daniel, Grinblatt, Titman and Wermers, 1997). We report results for the full sample ("Full Sample"), earnings announcements sample ("EarnAnn Sample") and analyst recommendation changes sample ("*RecChng* Sample"). The Full Sample includes 1,714,610 day-stock observations; the EarnAnn Sample includes 18,453 EarnAnn-stock observations; and the *RecChng* Sample includes 7,041 *RecChng*-stock observations. Panel A reports for each sample the mean, median and standard deviation of the firms' time series averages. Panel B reports the conditional means conditioning on AIA=0 and AIA=1.

In order to construct the measure, Bloomberg records the number of times news articles on a particular stock are read by its terminal users and the number of times users actively search for news for a specific stock. Bloomberg then assigns a value of 1 for each article read and 10 for each news search. These numbers are then aggregated into an hourly count. Using the hourly count, Bloomberg then creates a numerical attention score each hour by comparing past 8-hour average count to all hourly counts over the previous month for the same stock. They assign the value of 0 if the rolling average is less than 80% of the hourly counts over the previous 30 days. Similarly, Bloomberg assigns a score of 1, 2, 3 or 4 if the average is between 80% and 90%, 90% and 94%, 94% and 96%, or greater than 96% of the previous 30 days' hourly counts, respectively. Finally, Bloomberg aggregates up to the daily frequency by taking a maximum of all hourly scores throughout the day. These are the data provided to us by Bloomberg. Since we are interested in abnormal attention, our *AIA* measure is a dummy variable which receives the value of 1 if Bloomberg's score is 3 or 4, and 0 otherwise. This captures the right tail of the measure's distribution.

In the table, "*Num Firms*" reports the number of unique firms. "*AIA Frequency*" reports *AIA* frequency for all three samples. In the case of the *Full Sample*, we divide each firm's total number of days where *AIA* is equal to 1 by the firm's total trading days during its sample period. Then, we calculate the cross-sectional average. For the *EarnAnn* and *RecChng* samples, we divide the number of firm-event cases where *AIA* is equal to 1 by the total number of firm-event observations. For all other variables, Mean, Median and SD refer to the cross sectional average, median and standard deviation of the firms' time series averages. "*SizeInM*" is the stock's market capitalization, rebalanced every June, in millions of dollars. "*LnBM*" is the natural logarithm of the stock's book-to-market ratio, rebalanced every June following Fama-French (1992). "*SDRET*" is the daily standard deviation of stock returns, calculated based on the previous 21 trading days. "*Ret*" is the daily stock return, in %. "*AbsRet*" is the absolute value of *Ret*. "*DGTW*" is *Ret* minus the stock's daily benchmark portfolio daily return following Daniel, Grinblatt, Titman and Wermers (1997). "*AbsDGTW*" is the absolute value of *DGTW*. "*Turnover*", is the daily stock turnover. "*Dvol in M*" is the daily dollar trading volume, in millions of dollars. "*HLtoH*" is the ratio between the stock's daily high-and-low price difference and the daily high price. "*InstHold*" is the percentage of shares held by institutional investors obtained from the Thomson Reuters CDA/Spectrum institutional holdings' (S34) database. "*NumEst*" is the the number of analysts covering the stock. "Abs SUE/REC" is the absolute value of the surprise in analyst forecast and analyst recommendation change. SUE and REC are defined in Tables 4 and 5, respectively. "*Relative Spread*" is calculated as the [(Ask-Bid)/Midpoint]/2 using *CRSP* end of day quotes. Panel A also reposts two additional variables which are associated with attention. "*EDGAR*" is the daily number of unique requests for firm filings on the SEC EDGAR server (Loughran and McDonald, 2015). "*LnAbnEDGAR*" is the natural log of the ratio of *EDGAR* on day *t* to the average of *EDGAR* over the previous month. "*LnAbnDSVI*" is Da, Engelberg and Gao (2011)'s abnormal retail attention measure based on Google's Search Volume Index (*SVI*), calculated at the daily frequency (*DSVI*). Similar to *EDGAR*, "*LnAbnDSVI*" is calculated as the natural log of the ratio of *DSVI* on day *t* to the average of *DSVI* over the previous month. Due to daily *SVI* data availability, the *Full*, *EarnAnn* and *RecChng* samples include 1,012,855, 11,315, and 4,954 *LnAbnDSVI* observations, respectively.

Panel 1.A – Cross-Sectional Statistics

	Full Sample			*EarnAnn* Sample			*RecChng* Sample		
Variables	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Num Firms	2,298			1,952			1,376		
AIA frequency	0.098			0.567			0.447		
SizeInM	5,575	945	19,810	5,718	1,000	20,275	8,585	2,075	24,582
LnBM	-0.63	-0.51	0.83	-0.67	-0.54	0.82	-0.76	-0.68	0.80
SDRET	2.23	2.10	0.84	2.08	1.94	0.91	2.22	2.00	1.21
Ret %	0.09	0.08	0.30	0.33	0.15	2.84	0.15	0.13	4.10
DGTW %	0.01	0.01	0.30	0.31	0.11	2.66	0.09	0.09	3.91
Turnover	0.01	0.01	0.01	0.02	0.02	0.03	0.02	0.01	0.04
Dvol in M	52.82	8.31	207.27	130.70	25.15	489.25	121.91	41.16	307.79
HLtoH	0.03	0.03	0.01	0.07	0.06	0.03	0.04	0.03	0.03
InstHold	0.62	0.67	0.22	0.65	0.68	0.20	0.70	0.74	0.18
NumEst	9.06	7.16	7.05	9.46	7.52	6.74	13.02	11.75	6.96
Abs SUE/REC	N/A	N/A	N/A	2.68	2.22	1.92	1.34	1.25	0.35
EDGAR	42.17	25.38	67.97	79.64	46.33	115.36	68.08	38.66	105.94
LnAbnEDGAR	-0.03	-0.02	0.06	0.66	0.66	0.42	0.10	0.11	0.56
LnAbnDSVI	-0.01	0.00	0.09	0.02	0.02	0.37	0.00	0.02	0.40

Panel 1.B – Sample Averages Conditioning on AIA

	Full Sample		*EarnAnn* Sample		*RecChng* Sample	
Variables	AIA=0	AIA=1	AIA=0	AIA=1	AIA=0	AIA=1
AbsRet %	1.55	3.09	4.11	5.70	2.08	3.69
AbsDGTW %	1.22	2.75	3.79	5.29	1.75	3.37
Turnover	0.0080	0.0180	0.0172	0.0296	0.0172	0.0299
Dvol in M	48.01	73.36	65.18	148.52	106.91	174.63
HLtoH	0.0290	0.0450	0.0620	0.0709	0.0352	0.0427
Relative Spread	0.0010	0.0008	0.0011	0.0006	0.0004	0.0004
NumEst	9.05	9.45	7.63	10.26	13.36	14.71
InstHold	0.62	0.64	0.61	0.68	0.71	0.71
Abs SUE/REC	N/A	N/A	2.68	2.85	1.33	1.36
EDGAR	40.40	55.27	57.11	88.31	63.94	85.68
LnAbnEDGAR	-0.108	0.212	0.574	0.709	0.030	0.182
LnAbnDSVI	-0.0014	0.0166	-0.012	0.040	-0.030	0.038

Table 2 – The Contemporaneous Relation between Abnormal Institutional Attention, Abnormal Retail Attention, Attention Proxies and Other Explanatory Variables

The table reports results of the contemporaneous relation between Abnormal Institutional Attention measure ("*AIA"*) from Bloomberg (Panel A) and abnormal retail attention ("*LnAbnDSVI*") based on Google's daily Search Volume Index (Panel B) on selected explanatory variables. "*AIA*" and "*LnAbnDSVI*" are defined in Table 1. In particular, Panel A (B) analyzes Probit panel models (OLS panel regressions) where *AIA* (*LnAbnDSVI*) is the dependent variable. Each panel includes 6 identical specifications. Focusing on Panel A, Specification 1 explores the relation between *AIA* and "News" variables; Specification 2 explores the relation between *AIA* and price related variables; Specification 3 explores the relation between *AIA* and other firm characteristics; Specification 4 explores the relation between *AIA* and other attention measures; Specification 5 explores the effect of the day of the week effect on *AIA*; and Specification 6 explores all 5 categories together. Due to daily SVI data availability, Panel A includes 1,714,610 day-stock observations and Panel B includes 1,012,855 day-stock observations. We handle *LnAbnDSVI*'s missing observations when analyzing *AIA* in Panel A using Pontiff and Woodgate's (2008) approach. First, we define a dummy variable which takes the value of 1 whenever the *LnAbnDSVI* exists and 0 otherwise. Second, we replace *LnAbnDSVI* missing values with zeros.

In both panels, "*LnNews*" is the log of 1 plus the number of news articles published on the Dow Jones newswire during the day, provided by RavenPack. "*EarnDum*" is a dummy variable which receives the value of 1 on earnings announcements days and 0 otherwise. "*RecChngDum*" is a dummy variable which receives the value of 1 on days with change in analyst recommendations and 0 otherwise. "*Ret* is the CRSP daily stock return. "*Dgtw*" is *Ret* minus the stock's benchmark portfolio daily return following Daniel, Grinblatt, Titman and Wermers (1997). "*AbsDgtw*" is the absolute value of *Dgtw*. "*AbnVol*" is the stock's abnormal trading volume calculated following Barber and Odean (2008) as the stock's daily volume divided by the previous 252-day average trading volume. "*HLtoH*" is the ratio between the stock's daily high-and-low price difference and the daily high price. *52HighDum* (*52LowDum*) is a dummy variable which receives the value of 1 if the stock's price beat its 52-week high (low) price and 0 otherwise. "*LnSize*" is the log of the stock's average size in millions of dollars from day *t-27* to *t-6*. "*LnBM*" is the natural logarithm of the firm's book-to-market ratio, rebalanced every June following Fama-French (1992). *SDRET* is the standard deviation of daily stock returns from day *t*-27 to day *t*-6. "*InstHold*" is the percentage of shares held by institutional investors obtained from the Thomson Reuters CDA/Spectrum institutional holdings' (S34) database. "*LnNumEst*" is the log of 1 plus the number of analysts covering the stock, using the most recent information. "*AdvExpToSales*" is the firm advertising expenses to sales as in Da, Engelberg and Gao (2011), using the most recent information. "*LnAbnEDGAR*" is defined in Table 1. "Tuesday" – "Friday" are dummy variables which receive the value of 1 if the stock's day of the week is Tuesday-Friday, respectively, and 0 otherwise. "*P-RSQ*" ("*Adj-RSQ*") is the Probit model's (OLS panel regression's) pseudo (adjusted) *R*-squared. Standard errors are clustered by stock and day and *t*-statistics are reported below the coefficient estimates.

Panel 2.A – *AIA as a Dependent Variable*

Variable	(1)	(2)	(3)	(4)	(5)	(6)
LnNews t	0.382					0.205
	38.16					20.33
EarnAnnDum t	0.943					0.793
	32.24					27.47
RecChngDum t	0.866					0.655
	36.45					29.30
AbsDgtw t		0.074				0.072
		17.99				9.74
AbnVol t		0.059				0.049
		3.42				4.59
HLtoH t		2.243				9.371
		5.51				8.14
52 High Dum t		0.356				0.048
		23.48				2.80
52 Low Dum t		0.105				-0.235
		3.74				-5.02
LnSize			0.154			0.134
			23.56			12.47
LnBM			-0.014			-0.027
			-1.63			-2.97
SDRET			0.024			-0.078
			5.01			-9.03
InstHold			-0.010			0.042
			0.27			1.40
LnNumEst			0.235			0.231
			16.52			13.14
AdvExpToSale			-0.005			-0.403
			-0.02			-1.56
LnAbnDSVI				0.114		0.059
				7.72		5.86
LnAbnEDGAR				0.207		0.114
				29.82		18.61
Tuesday					-0.008	-0.081
					-0.41	-3.19
Wednesday					-0.043	-0.127
					-2.04	-4.99
Thursday					-0.046	-0.164
					-2.11	-6.25
Friday					-0.248	-0.329
					-11.27	-11.93
P-RSQ	7.19%	2.49%	4.89%	1.63%	0.26%	12.05%

Panel 2.B – *LnAbnDSVI as a Dependent Variable*

Variable	(1)	(2)	(3)	(4)	(5)	(6)
LnNews t	0.011					0.005
	6.16					2.69
EarnAnnDum t	0.017					0.000
	3.18					-0.08
RecChngDum t	0.017					0.011
	2.86					1.94
AbsDgtw t		0.004				0.004
		6.23				5.90
AbnVol t		0.004				0.003
		4.19				3.72
HLtoH t		-0.339				-0.039
		-3.94				-0.59
52 High Dum t		0.011				0.000
		3.36				-0.12
52 Low Dum t		0.000				-0.008
		0.05				-0.90
LnSize			0.007			0.005
			3.17			2.15
LnBM			0.003			0.002
			1.01			0.86
SDRET			-0.006			-0.007
			-3.57			-4.29
InstHold			0.000			0.004
			0.02			0.38
LnNumEst			-0.008			-0.009
			-1.71			-1.88
AdvExpToSale			-0.064			-0.076
			-0.70			-0.82
AIA				0.032		0.019
				9.77		7.29
LnAbnEDGAR				0.007		0.005
				5.10		3.24
Tuesday					0.008	0.007
					1.79	1.65
Wednesday					-0.001	-0.002
					-0.13	-0.32
Thursday					-0.007	-0.008
					-1.41	-1.61
Friday					-0.031	-0.030
					-6.13	-5.88
Adj-RSQ	0.08%	0.05%	0.12%	0.10%	0.11%	0.34%

Table 3 – Abnormal Institutional Attention, Abnormal Retail Attention and Abnormal Trading Volume

The table reports results of panel regressions of abnormal trading volume on abnormal institutional attention ("*AIA*") (Panel A) and abnormal retail attention ("*LnAbnDSVI*") (Panel B) controlling for Table 2's attention determinants. "*AIA*" and "*LnAbnDSVI*" are defined in Table 1. We explore two samples of trading volume. The first is based on CRSP, where the CRSP's daily abnormal trading volume (hereafter, "*CRSP-AbnVol*") is calculated as the stock's CRSP daily volume divided by the previous 8-week average trading volume. The second sample is obtained from Ancerno Ltd., and captures institutional investors' trading volume. We calculate the abnormal institutional trading volume the same way as *CRSP-AbnVol* (hereafter, "*Ancerno-AbnVol*"). The Ancerno data are available until March 2013. After matching the CRSP and Ancerno samples and accounting for *LnAbnDSVI*'s data availability, Panel A (B) includes 1,314,755 (821,098) day-stock observations.

Panel A includes six specifications, where we sequentially add the five sets of control variables associated with institutional attention explored in Table 2. For example, "*Control Set 1*" includes *LnNews*, *EarnAnnDum* and *RecChngDum* control variables. Note that we exclude abnormal volume from *ControlSet* 2 since abnormal volume is our dependent variable. Recall that *AIA* is a dummy variable, thus, its coefficient captures the additional effect abnormal institutional attention (i.e., AIA=1). For brevity, we only report AIA's coefficient. "*CRSP-AbnVol-Diff*" ("*Ancerno-AbnVol-Diff*") is the difference in average abnormal volume of AIA=1 and AIA=0, where *CRSP-AbnVol* (*Ancerno-AbnVol*) is the dependent variable. "*Diff-in-Diff*" is the difference between the samples' average differences, using the difference-in-difference regression approach. Panel B includes the same specifications. To employ the same methodology used in Panel A, we transform *LnAbnDSVI* into a dummy variable (*DSVIDUM*) that mimics AIA's sample frequency. In particular, each day we rank *LnAbnDSVI* into ten deciles based on *LnAbnDSVI* values. Then, *DSVIDUM* received the value of 1 if *LnAbnDSVI* is in the top decile and 0 otherwise. Similar to Panel A, "*CRSP-AbnVol-Diff*" ("*Ancerno-AbnVol-Diff*") is the difference between *DSVIDUM*=1 and *DSVIDUM*=0 where *CRSP-AbnVol* (*Ancerno-AbnVol*) is the dependent variable, and "*Diff-in-Diff*" is the difference using the difference-in-difference regression approach. Standard errors are clustered by stock and day. *t*-statistics are reported below the regression coefficients.

Panel 3.A – *AIA* and Abnormal Trading Volume

Variable	(1)	(2)	(3)	(4)	(5)	(6)
CRSP-AbnVol- Diff	0.650	0.461	0.263	0.235	0.230	0.240
	17.78	11.47	9.31	8.98	8.53	8.6
Ancerno-AbnVol- Diff	0.763	0.547	0.333	0.315	0.308	0.318
	20.51	14.57	12.19	11.96	11.09	11.12
Diff-In-Diff	0.114	0.086	0.070	0.081	0.078	0.079
	3.98	3.77	2.61	3.39	3.23	3.14
Table 2 Controls						
Control Set 1		Yes	Yes	Yes	Yes	Yes
Control Set 2			Yes	Yes	Yes	Yes
Control Set 3				Yes	Yes	Yes
Control Set 4					Yes	Yes
Control Set5						Yes
Adj-RSQ	0.73%	1.33%	4.37%	4.73%	4.77%	4.80%

Panel 3.B – *LnAbnDSVI* and Abnormal Trading Volume

Variable	(1)	(2)	(3)	(4)	(5)	(6)
CRSP-AbnVol- Diff	0.070	0.054	0.027	0.026	0.022	0.022
	5.73	5.08	2.89	3.28	2.63	2.65
Ancerno-AbnVol- Diff	0.075	0.057	0.023	0.026	0.020	0.019
	3.25	3.01	1.71	1.76	1.30	1.29
Diff-In-Diff	0.005	0.003	-0.004	-0.001	-0.002	-0.002
	0.21	0.04	0.052	0.01	0.03	0.03
Table 2 Controls						
Control Set 1		Yes	Yes	Yes	Yes	Yes
Control Set 2			Yes	Yes	Yes	Yes
Control Set 3				Yes	Yes	Yes
Control Set 4					Yes	Yes
Control Set5						Yes
Adj-RSQ	0.05%	1.11%	4.23%	4.64%	4.75%	4.77%

Table 4 – *Institutional Attention* and Earnings Announcements Returns

The table reports results of panel regressions of earnings announcements' day-0 and cumulative day *t*+1 to *t*+40 *DGTW* risk adjusted returns on abnormal institutional attention and other explanatory variables. The sample includes 18,543 firm-quarter observations (see Table 1). In Panels A and B, we presents results using *AIA* and unexpected *AIA*, respectively. In Panel C, we focus on a reduced sample of earnings announcements that occur between 4:00pm-12:00am of day *t*-1. The reduced sample includes 9,308 stock-quarter observations.

In Panel A, "*AIA*" is our Abnormal Institutional Attention measure. "*SUE*" is the quarterly standardized unexpected earnings, calculated from I/B/E/S as the quarter's actual earnings minus the average of the most recent analyst forecast, divided by the standard deviation of that forecast. "*SUE_AIA*" is the interaction between *SUE* and *AIA*. Since *AIA* is a dummy variable, that interaction between *SUE* and *AIA* measures the additional sensitivity of the *AIA* equals 1 group. "*LnNews*" is the log of 1 plus the number of news articles published on the Dow Jones newswire during the day, provided by RavenPack. "*EDGAR*" is the daily number of unique requests for firm filings on the SEC EDGAR server (Loughran and McDonald, 2015). "*LnAbnEDGAR*" is the natural log of the ratio of *EDGAR* on day *t* to the average of *EDGAR* over the previous month. "*LnAbnDSVI*" is the natural log of the ratio of *DSVI* on day *t* to the average of *DSVI* over the previous month (see Table 2 for Pontiff and Woodgate's (2008) missing values approach). "*AbnVol*" is the stock's abnormal trading volume calculated following Barber and Odean (2008) as the stock's daily volume divided by the previous 252-day average trading volume. "*HLtoH*" is the ratio between the stock's daily high-and-low price difference and the daily high price. "*Ret t*-5_*t*-1" is the cumulative return from day *t*-5 to *t*-1. "*Turnover t*-5_*t*-1" is the stock's average turnover from day *t*-5 to *t*-1. "*Spread t-5_t-1*" is the average relative half bid-ask spread from day *t*-5 to *t*-1, calculated as [(Ask-Bid)/Midpoint]/2 using *CRSP* end of day quotes. "*SDRET*" is the standard deviation of daily return from day *t*-27 to day *t*-6. "*LnSize*" is the log of the stock's average size in millions of dollars from day *t-27* to *t-6*. "*LnBM*" is the firm natural logarithm of the firm book-to-market ratio, rebalanced every June following Fama-French (1992). "*InstHold*" is the percentage of shares held by institutional investors obtained from the Thomson Reuters CDA/Spectrum institutional holdings' (S34) database. "*LnNumEst*" is the log of 1 plus the number of analysts covering the stock, using the most recent information.

In Panel B, we estimate a Probit model using the earnings announcement sample to predict the probability of being a firm with AIA=1 on earnings announcement days. The model includes the event day *AbnVol*, *LnAbnEDGAR*, *HLtH* and *Spread* and also includes *LnSize* and *SDRET* firm characteristics. Using the model estimates, we spilt *AIA* into predicted and residual *AIA*. We then replace *AIA* with the residual part (hereafter, "*ResidAIA*"). For example, *SUE_ResidAIA*, is the interaction between *SUE* and *ResidAIA*. The other controls are identical to Panel A's control variables.

In Panel C, AIA is estimated on day *t*-1 to match SUE timing (i.e., 4:00pm-12:00am). In a similar manner, *SUE t*-1_ *AIA t*-1 is the interaction between SUE and AIA on day *t*-1.

Standard errors are clustered by stock and day and each model includes quarter and day-of-week fixed effects. *t*-statistics are reported below the coefficient estimates.

Panel 4.A – *AIA*

Variables	*t*	*t+1_t+1*	*t+1_t+2*	*t+1_t+3*	*t+1_t+5*	*t+1_t+10*	*t+1_t+20*	*t+1_t+30*	*t+1_t+40*
AIA t	0.000	-0.001	0.000	-0.001	-0.001	-0.001	0.000	0.001	0.002
	-0.02	-1.11	-0.47	-1.11	-1.00	-0.72	0.12	0.24	0.76
SUE t	0.0046	0.0007	0.0009	0.0011	0.0011	0.0013	0.0017	0.0017	0.0020
	17.85	6.21	7.32	7.45	6.85	6.33	6.16	5.19	5.06
SUE_AIA t	0.0019	-0.0003	-0.0005	-0.0007	-0.0007	-0.0008	-0.0011	-0.0013	-0.0015
	5.59	-2.63	-2.96	-3.58	-3.09	-2.75	-2.91	-3.14	-3.02
LnNews t	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.001
	2.49	-0.41	-0.72	0.14	0.43	0.83	0.18	0.78	0.99
LnAbnEDGAR t	-0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000
	-1.01	0.77	0.73	-0.35	-0.43	1.10	0.04	0.85	0.15
LnAbnDSVI t	0.002	0.000	-0.001	0.000	0.001	0.001	0.003	0.004	0.003
	0.79	-0.05	-0.99	-0.22	0.64	0.61	1.17	1.31	0.93
AbnVol t	-0.001	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001
	-0.42	1.76	0.85	0.47	1.31	2.62	2.63	1.85	2.19
HLtoH t	-0.029	-0.030	-0.034	-0.020	-0.029	-0.043	-0.018	-0.006	-0.055
	-0.68	-3.36	-2.51	-1.31	-1.37	-1.76	-0.57	-0.18	-1.22
Ret t-5_t-1	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	-5.40	0.17	0.54	0.81	0.98	0.74	-0.08	0.54	1.23
Turnover t-5_t-1	-0.325	0.010	0.016	0.042	0.058	0.002	-0.128	-0.223	-0.424
	-3.20	0.48	0.47	1.01	1.03	0.02	-1.12	-1.67	-2.75
Spread t-5_t-1	0.249	-0.007	0.178	0.246	0.302	0.668	0.806	0.759	3.150
	0.44	-0.01	0.37	0.42	0.43	0.71	0.60	0.49	1.32
SDRET	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
	1.86	-0.64	0.25	-0.26	-0.34	-0.13	0.09	-0.18	0.63
LnSize	-0.003	0.000	0.000	0.000	-0.001	-0.001	-0.001	-0.002	-0.001
	-5.37	-1.84	-0.41	-0.84	-1.48	-1.98	-1.53	-1.68	-0.91
LnBM	-0.001	-0.001	-0.001	0.000	0.000	-0.001	-0.001	-0.002	-0.004
	-1.69	-1.60	-1.37	-0.30	-0.10	-0.69	-0.67	-1.51	-2.23
InstHold	0.004	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.002
	1.80	0.11	0.28	0.50	0.72	0.18	0.19	0.27	0.43
LnNumEst	0.002	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	-0.002
	1.81	0.40	-0.22	-0.19	-0.43	0.01	-0.04	-0.57	-0.77

Panel 4.B – Residual *AIA*

Variables	*t*	*t+1_t+1*	*t+1_t+2*	*t+1_t+3*	*t+1_t+5*	*t+1_t+10*	*t+1_t+20*	*t+1_t+30*	*t+1_t+40*
ResidAIA t	0.000	-0.001	0.000	-0.001	-0.001	-0.001	0.000	0.001	0.002
	0.12	-1.08	-0.50	-1.12	-1.05	-0.77	0.24	0.41	0.97
SUE t	0.0056	0.0005	0.0007	0.0007	0.0007	0.0009	0.0011	0.0010	0.0011
	29.87	8.12	7.81	7.46	6.30	5.95	5.78	4.81	4.30
SUE_ResidAIA t	0.0001	-0.0003	-0.0004	-0.0005	-0.0004	-0.0003	-0.0006	-0.0008	-0.0011
	0.29	-1.96	-1.97	-2.24	-1.42	-1.23	-1.54	-1.63	-2.04
LnNews t	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.001
	2.66	-0.47	-0.77	0.08	0.40	0.87	0.23	0.82	1.01
LnAbnEDGAR t	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000
	-0.77	0.69	0.65	-0.46	-0.51	0.91	0.02	0.88	0.17
LnAbnDSVI t	0.002	0.000	-0.001	0.000	0.001	0.001	0.003	0.004	0.003
	0.78	-0.05	-1.02	-0.24	0.70	0.58	1.15	1.33	0.93
AbnVol t	-0.001	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001
	-0.43	1.83	0.89	0.49	1.35	2.62	2.62	1.87	2.21
HLtoH t	-0.028	-0.032	-0.035	-0.023	-0.033	-0.046	-0.020	-0.009	-0.055
	-0.80	-3.56	-2.86	-1.52	-1.92	-2.18	-0.74	-0.26	-1.16
Ret t-5_t-1	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	-6.04	0.20	0.35	0.81	1.01	0.74	-0.12	0.53	1.31
Ave Turnover t-5_t-1	-0.315	0.010	0.016	0.042	0.058	0.001	-0.129	-0.224	-0.423
	-3.06	0.48	0.49	1.09	1.12	0.01	-1.18	-1.74	-2.95
Ave Spread t-5_t-1	0.257	0.009	0.191	0.271	0.334	0.696	0.817	0.790	3.142
	0.45	0.02	0.40	0.46	0.47	0.73	0.62	0.51	1.30
Daily SDRET t-1	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
	2.10	-0.66	0.24	-0.27	-0.34	-0.13	0.10	-0.18	0.63
LnSize	-0.003	-0.001	0.000	0.000	-0.001	-0.002	-0.002	-0.002	-0.001
	-5.54	-1.99	-0.59	-1.20	-1.92	-2.39	-1.83	-1.96	-0.91
LnBM	-0.001	-0.001	-0.001	0.000	0.000	-0.001	-0.001	-0.002	-0.004
	-1.87	-1.84	-1.59	-0.31	-0.10	-0.78	-0.75	-1.69	-2.41
InstHold	0.004	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.002
	1.71	-0.10	0.14	0.28	0.52	0.02	0.13	0.20	0.39
LnNumEst	0.002	0.000	0.000	0.000	-0.001	0.000	0.000	-0.001	-0.002
	1.96	0.07	-0.52	-0.58	-0.80	-0.20	-0.11	-0.67	-0.80

Panel 4.C – Earnings Announcements - After Market Hours

Variables	*t*	*t+1_t+1*	*t+1_t+2*	*t+1_t+3*	*t+1_t+5*	*t+1_t+10*	*t+1_t+20*	*t+1_t+30*	*t+1_t+40*
AIA t- 1	0.000	-0.001	-0.001	-0.001	0.000	0.000	0.003	0.002	0.003
	0.13	-1.70	-0.96	-1.24	-0.15	0.11	1.09	0.81	0.67
SUE t-1	0.0054	0.0006	0.0007	0.0007	0.0008	0.0012	0.0012	0.0012	0.0013
	13.69	4.60	4.36	3.85	3.48	3.85	3.09	2.50	2.29
SUE t -1_ AIA t- 1	0.0013	-0.0005	-0.0005	-0.0004	-0.0005	-0.0011	-0.0011	-0.0013	-0.0011
	2.47	-2.62	-2.00	-1.49	-1.55	-2.58	-2.08	-2.17	-1.85
LnNews t	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.002
	2.49	-0.09	-0.08	0.44	0.64	1.26	0.74	1.20	1.87
LnAbnEDGAR t	0.000	0.000	0.001	0.001	0.001	0.002	0.002	0.002	0.002
	-0.03	1.21	1.63	1.30	1.03	2.07	1.44	1.73	1.30
LnAbnDSVI t	0.002	0.001	0.001	0.002	0.003	0.003	0.005	0.006	0.004
	0.67	1.44	0.47	1.21	1.88	1.65	1.84	1.57	0.97
AbnVol t	-0.001	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001
	-0.35	2.15	1.18	0.22	1.13	2.45	1.82	0.98	1.69
HLtoH t	0.019	-0.036	-0.028	-0.016	-0.034	-0.063	-0.019	0.011	-0.050
	0.35	-2.78	-1.47	-0.74	-1.13	-1.74	-0.41	0.20	-0.76
Ret t-5_t-1	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	-2.91	0.41	-0.33	0.02	0.37	0.32	0.44	0.89	0.64
Turnover t-5_t-1	-0.342	0.012	0.021	0.067	0.095	0.061	0.014	-0.029	-0.300
	-2.37	0.52	0.57	1.32	1.28	0.56	0.10	-0.18	-1.54
Spread t-5_t-1	0.928	-0.079	0.156	0.750	0.749	1.790	2.983	3.485	3.128
	0.82	-0.25	0.34	1.43	1.49	2.34	2.17	2.23	1.87
SDRET	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.000	0.006
	0.66	1.32	1.55	0.58	0.49	0.66	0.51	-0.08	0.86
LnSize	-0.002	-0.001	0.000	-0.001	-0.001	-0.002	-0.002	-0.002	-0.002
	-2.93	-1.94	-0.50	-1.22	-1.93	-2.38	-1.56	-1.44	-0.88
LnBM	-0.002	-0.001	-0.001	0.000	0.000	-0.001	0.000	-0.002	-0.004
	-2.18	-1.25	-1.48	-0.57	-0.41	-0.79	-0.28	-1.08	-1.88
InstHold	0.009	0.001	-0.001	0.000	0.001	0.003	0.004	0.002	0.000
	2.78	0.95	-0.34	-0.27	0.58	0.92	0.82	0.45	-0.03
LnNumEst	0.000	0.000	0.000	0.000	-0.001	0.001	0.000	-0.001	-0.003
	0.00	0.27	-0.37	-0.01	-0.41	0.50	-0.18	-0.36	-0.92

Table 5 – *Institutional Attention* and Change-in-Analyst-Recommendations Returns

The table reports results of panel regressions of change in analyst recommendations' day-0 and cumulative day *t*+1 to *t*+10 *DGTW* risk adjusted returns on institutional attention and other explanatory variables. In constructing the sample, we basically follow Jegadeesh and Kim (2010), Loh and Stulz (2011) and Kadan, Michaely, and Moulton (2013). In particular, we: (1) remove recommendation changes that occur on the same day as, or the day following, earnings announcements; (2) remove recommendation changes on days when multiple analysts issue recommendations for the same firm; (3) require at least one analyst who to have issued a recommendation for the stock and revised the recommendation within 180 calendar days; (4) require at least two analysts, other than the revising analyst, to have active recommendations for the stock as of the day before the revision; (5) consider a recommendation to be active for up to 180 days after it is issued or until I/B/E/S indicates that the analyst has stopped issuing recommendations for that stock. After applying all these filters, we end up with 7,041 changes in recommendations.

Panel A (B) presents our main (robustness) model using *AIA* (unexpected *AIA*). In Panel A, "*AIA*" is our Abnormal Institutional Attention measure. "*RecChng*" is the change in analyst recommendations. The variable ranges from -4 to 4, where a positive (negative) number refers to an upgrade (a downgrade) (that is, a reverse scale). "*RecChng_AIA*" is the interaction between the *RecChng* and *AIA*. Similar to Table 4, since *AIA* is a dummy variable, that interaction measures the additional sensitivity of the *AIA* equals 1 group. "*LnNews*" is the log of 1 plus the number of news articles published on the Dow Jones newswire during the day, provided by RavenPack. "*EDGAR*" is the daily number of unique requests for firm filings on the SEC EDGAR server (Loughran and McDonald, 2015). "*LnAbnEDGAR*" is the natural log of the ratio of *EDGAR* on day *t* to the average of *EDGAR* over the previous month. "*LnAbnDSVI*" is the natural log of the ratio of *DSVI* on day *t* to the average of *DSVI* over the previous month (see Table 2 for Pontiff and Woodgate's (2008) missing values approach). "*AbnVol*" is the stock's abnormal trading volume calculated following Barber and Odean (2008) as the stock's daily volume divided by the previous 252-day average trading volume. "*HLtoH*" is the ratio between the stock's daily high-and-low price difference and the daily high price. "*Ret t-5_t-1*" is the cumulative return from day t-5 to t-1. *Turnover t-5_t-1* is the stock's average turnover from day t-5 to t-1. "*Spread t-5_t-1*" is the average relative half bid-ask spread from day t-5 to t-1, calculated as [(Ask-Bid)/Midpoint]/2 using CRSP end of day quotes. "*SDRET*" is the standard deviation of daily return from day t-27 to day t-6. "*LnSize*" is the log of the stock's average size in millions of dollars from day t-27 to t-6. "*LnBM*" is the natural logarithm of the firm book-to-market ratio, rebalanced every June following Fama-French (1992). "*InstHold*" is the percentage of shares held by institutional investors obtained from the Thomson Reuters CDA/Spectrum institutional holdings' (S34) database. "*LnNumEst*" is the log of 1 plus the number of analysts covering the stock, using the most recent information.

In Panel B, we estimate a Probit model using the earnings announcement sample to predict the probability of being a firm with AIA equals 1 on change in analyst recommendation days. The model includes the event day *AbnVol*, *LnAbnEDGAR*, *HLtH* and *Spread* and also includes *LnSize* and *SDRET* firm characteristics. Using the model estimates, we spilt *AIA* into predicted and residual *AIA*. We then replace *AIA* with the unexpected part (hereafter, "*UnExpAIA*"). For example, *RecChng_UnExpAIA*, is the interaction between *RecChng* and *UnExpAIA*. The other controls are identical to Panel A's control variables.

Standard errors are clustered by stock and day and each model includes quarter and day-of-week fixed effects. *t*-statistics are reported below the coefficient estimates.

Panel 5.A – *AIA*

Variables	t	t+1_t+1	t+1_t+2	t+1_t+3	t+1_t+4	t+1_t+5	t+1_t+6	t+1_t+7	t+1_t+8	t+1_t+9	t+1_t+10
AIA t	0.002	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	-0.001	-0.002
	1.93	-0.17	-0.19	0.02	-0.04	0.09	1.13	0.69	-0.07	-0.59	-0.96
RecChng t	0.0059	0.0016	0.0020	0.0018	0.0019	0.0024	0.0026	0.0029	0.0025	0.0027	0.0028
	18.27	5.99	5.33	3.93	3.25	4.45	4.66	4.65	3.84	3.95	3.87
RecChng_AIA t	0.0080	-0.0010	-0.0010	-0.0009	-0.0006	-0.0013	-0.0021	-0.0023	-0.0021	-0.0025	-0.0026
	11.33	-2.53	-1.85	-1.52	-0.82	-1.54	-2.36	-2.37	-2.01	-2.19	-2.18
LnNews t	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.41	1.93	1.10	0.99	0.43	0.31	-0.05	0.31	0.23	0.21	0.26
LnAbnEDGAR t	0.001	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000	0.000	0.00	0.00
	0.87	0.47	-1.01	-0.14	-0.55	-0.79	-0.94	-0.16	0.08	0.20	0.12
LnAbnDSVI t	0.001	0.001	0.000	0.001	0.002	0.000	0.000	-0.001	-0.001	0.00	0.00
	0.66	1.12	0.34	0.63	1.23	0.16	0.09	-0.47	-0.66	-0.73	-0.54
AbnVol t	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.001	0.001	0.001
	0.09	0.51	0.88	0.78	0.65	1.21	0.90	0.73	1.00	1.08	1.09
HLtoH t	-0.151	-0.008	-0.050	-0.035	0.054	0.024	0.012	0.000	0.012	0.036	0.024
	-1.83	-0.29	-1.73	-0.76	0.52	0.29	0.14	0.00	0.14	0.40	0.28
Ret t-5_t-1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.69	1.21	1.88	0.50	-0.11	-0.12	-0.31	0.02	-0.12	-0.07	-0.11
Turnover t-5_t-1	-0.122	-0.055	-0.069	-0.104	-0.133	-0.138	-0.136	-0.117	-0.151	-0.158	-0.144
	-1.64	-2.16	-2.26	-2.98	-2.84	-2.48	-2.39	-1.98	-2.44	-2.42	-2.09
Spread t-5_t-1	-0.822	0.743	1.663	0.885	2.790	3.163	1.523	1.055	0.605	2.401	2.702
	-0.26	0.54	0.82	0.41	0.80	1.12	0.53	0.35	0.21	0.72	0.70
SDRET	-0.002	-0.001	-0.001	-0.001	0.000	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	-2.20	-1.87	-1.43	-1.08	-0.14	-0.70	-1.19	-1.22	-1.38	-0.85	-0.83
LnSize	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000
	1.00	0.35	0.81	0.56	0.03	-0.12	-0.10	-0.16	-0.33	-0.46	-0.36
LnBM	0.001	0.001	0.001	0.000	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.92	1.82	1.87	-0.30	-0.76	-0.84	-1.01	-1.11	-0.78	-0.73	-0.51
InstHold	0.000	0.001	0.000	-0.001	0.002	0.004	0.005	0.004	0.003	0.006	0.005
	-0.07	0.38	-0.19	-0.29	0.83	1.36	1.74	1.10	1.01	1.60	1.36
LnNumEst	0.001	0.001	0.001	0.001	0.001	0.002	0.003	0.003	0.004	0.003	0.002
	0.69	0.75	1.11	0.72	0.69	1.10	1.69	1.37	1.85	1.31	0.89

Panel 5.B – Residual *AIA*

Variables	*t*	*t+1_t+1*	*t+1_t+2*	*t+1_t+3*	*t+1_t+4*	*t+1_t+5*	*t+1_t+6*	*t+1_t+7*	*t+1_t+8*	*t+1_t+9*	*t+1_t+10*
ResidAIA t	0.001	0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.000	-0.001	-0.002
	0.88	-0.09	-0.23	0.01	-0.03	0.09	1.21	0.72	-0.04	-0.59	-1.03
RecChng t	0.0094	0.0012	0.0016	0.0014	0.0016	0.0019	0.0017	0.0019	0.0016	0.0017	0.0016
	29.15	6.18	6.29	4.50	4.02	4.60	3.97	4.02	3.18	3.13	2.88
RecChng_ResidAIA t	0.0056	-0.0010	-0.0009	-0.0008	-0.0001	-0.0007	-0.0016	-0.0017	-0.0016	-0.0017	-0.0021
	9.49	-2.31	-1.77	-1.22	-0.15	-0.84	-1.78	-1.68	-1.58	-1.67	-1.85
LnNews t	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.65	1.88	1.05	0.94	0.38	0.27	-0.06	0.29	0.21	0.18	0.23
LnAbnEDGAR t	0.001	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000	0.000	0.00	0.00
	0.95	0.49	-1.10	-0.13	-0.56	-0.77	-0.87	-0.12	0.08	0.17	0.08
LnAbnDSVI t	0.001	0.001	0.000	0.001	0.002	0.000	0.000	-0.001	-0.001	0.00	0.00
	0.61	1.10	0.34	0.64	1.27	0.18	0.14	-0.48	-0.70	-0.76	-0.57
AbnVol t	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.001
	0.14	0.52	0.89	0.83	0.67	1.25	0.97	0.77	0.98	1.02	1.04
HLtoH t	-0.148	-0.008	-0.050	-0.035	0.054	0.025	0.017	0.004	0.012	0.033	0.019
	-1.78	-0.30	-1.73	-0.79	0.54	0.31	0.20	0.05	0.14	0.38	0.22
Ret t-5_t-1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.75	0.81	1.83	0.49	-0.11	-0.12	-0.31	0.02	-0.13	-0.08	-0.13
Turnover t-5_t-1	-0.124	-0.055	-0.069	-0.104	-0.133	-0.138	-0.136	-0.117	-0.151	-0.158	-0.144
	-1.69	-2.28	-1.74	-2.48	-2.65	-2.54	-2.39	-1.79	-2.31	-2.30	-2.00
Spread t-5_t-1	-1.000	0.765	1.683	0.900	2.794	3.174	1.522	1.065	0.638	2.456	2.779
	-0.32	0.56	0.83	0.41	0.81	1.13	0.53	0.35	0.22	0.73	0.71
SDRET	-0.002	-0.001	-0.001	-0.001	0.000	0.000	-0.001	-0.001	-0.001	-0.001	-0.001
	-2.14	-1.90	-1.46	-1.04	-0.14	-0.66	-1.08	-1.14	-1.43	-0.93	-0.94
LnSize	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000
	0.84	0.38	0.85	0.57	0.04	-0.12	-0.13	-0.17	-0.30	-0.40	-0.28
LnBM	0.001	0.001	0.001	0.000	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.84	1.67	1.67	-0.29	-0.79	-0.88	-1.06	-1.19	-0.88	-0.82	-0.58
InstHold	0.000	0.001	0.000	-0.001	0.002	0.004	0.005	0.004	0.004	0.006	0.005
	-0.17	0.39	-0.18	-0.29	0.79	1.37	1.73	1.11	1.06	1.67	1.46
LnNumEst	0.001	0.001	0.001	0.001	0.001	0.002	0.003	0.003	0.004	0.003	0.002
	0.86	0.78	1.06	0.70	0.68	1.12	1.81	1.43	1.99	1.38	0.89

Table 6 – Retail Attention and Earnings Announcement and Change-in-Analyst-Recommendation Returns

The table repeats Tables 4 and 5 analyses, where we replace *AIA* with the Da, Engelberg and Gao (2011) daily abnormal retail attention measure (*LnAbnDSVI*, defined in Table 1). In Panel A (B) we analyze the earnings announcements (change in analyst recommendations) sample. We standardized *LnAbnDSVI* to reflect the effect of 1 standard deviation on returns ("*StndLnAbnDSVI*"). In each panel, Specification 1 only includes *StndLnAbnDSVI* and Specification 2 also includes the interaction term with the relevant event variable (i.e., *SUE* or *RecChng*). For brevity, the panels do not report Table 4 and 5's other control variables. Standard errors are clustered by stock and day and each model includes quarter and day-of-week fixed effects. *t*-statistics are reported below the coefficient estimates. Finally, due to daily SVI data availability, Panel A (B) includes 11,315, (4,954) event-stock observations.

Panel 6.A – Earnings Announcements

SPC	Variables	*t*	*t+1_t+1*	*t+1_t+2*	*t+1_t+3*	*t+1_t+5*	*t+1_t+10*	*t+1_t+20*	*t+1_t+30*	*t+1_t+40*
(1)	*StndLnAbnDSVI t*	0.0005	0.0000	-0.0003	-0.0001	0.0002	0.0003	0.0009	0.0014	0.0012
		0.71	0.04	-0.99	-0.33	0.60	0.52	1.06	1.27	0.96
	SUE t	0.0050	0.0005	0.0007	0.0007	0.0007	0.0007	0.0009	0.0007	0.0009
		15.88	5.21	5.81	5.80	4.90	4.28	3.54	2.44	2.58
(2)	*StndLnAbnDSVI t*	0.0001	0.0000	-0.0005	-0.0003	0.0001	0.0002	0.0007	0.0011	0.0010
		0.11	-0.03	-1.44	-0.79	0.29	0.26	0.80	1.03	0.81
	SUE t	0.0050	0.0005	0.0006	0.0007	0.0007	0.0007	0.0009	0.0007	0.0009
		15.82	5.26	5.78	5.74	4.85	4.22	3.49	2.38	2.54
	SUE_StndLnAbnDSVI t	0.0005	0.0000	0.0002	0.0002	0.0001	0.0002	0.0003	0.0003	0.0002
		2.13	0.29	2.34	2.46	1.84	1.56	1.60	1.52	1.24

Panel 6.B – Analyst Recommendation Changes

SPC	Variables	*t*	*t+1_t+1*	*t+1_t+2*	*t+1_t+3*	*t+1_t+4*	*t+1_t+5*	*t+1_t+6*	*t+1_t+7*	*t+1_t+8*	*t+1_t+9*	*t+1_t+10*
(1)	*StndLnAbnDSVI t*	0.0001	0.0002	0.0003	0.0004	0.0003	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002
		0.08	0.56	0.63	0.66	0.40	-0.03	-0.06	0.12	0.07	0.07	0.21
	RecChng t	0.0085	0.0012	0.0013	0.0011	0.0011	0.0015	0.0015	0.0015	0.0012	0.0012	0.0010
		19.67	5.30	4.22	2.68	2.30	2.96	2.87	2.52	1.90	1.83	1.51
(2)	*StndLnAbnDSVI t*	0.0000	0.0002	0.0003	0.0004	0.0003	0.0000	-0.0001	0.0001	0.0000	0.0001	0.0002
		-0.03	0.54	0.62	0.62	0.37	-0.06	-0.09	0.10	0.05	0.06	0.20
	RecChng t	0.0082	0.0011	0.0013	0.0010	0.0011	0.0015	0.0014	0.0014	0.0011	0.0012	0.0010
		20.21	5.18	4.12	2.52	2.19	2.83	2.73	2.44	1.80	1.82	1.51
	RecChng_StndLnAbnDSVI t	0.0021	0.0002	0.0000	0.0005	0.0005	0.0005	0.0007	0.0004	0.0006	0.0002	0.0001
		2.76	0.62	0.12	1.41	0.97	0.95	1.14	0.67	0.88	0.25	0.14

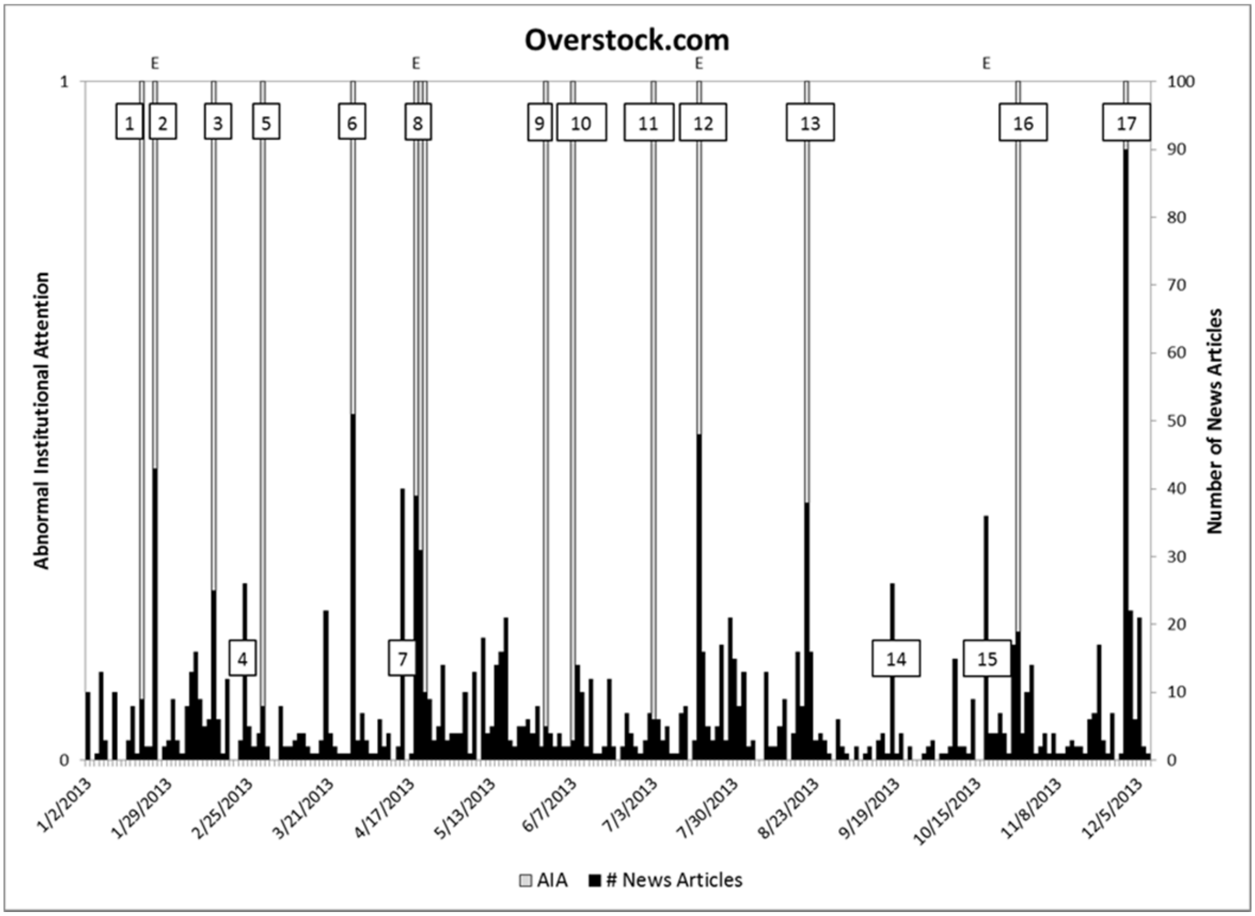
Table 7 – The Relation between Institutional and Abnormal Retail Attention around Information Events

The table explores the daily relation between Abnormal Institutional Attention measure ("*AIA*") and the abnormal retail attention measure ("*LnAbnDSVI*," defined in Table 1) around earnings announcements ("*EarnAnn"*) and change in analyst recommendations ("*RecChng"*) events. In particular, we pool together these events and estimate the lead-lag relation between *AIA* and *LnAbnDSVI*. We use Probit panel models when *AIA* is the dependent variable (Specifications 1-3) and OLS panel regressions when *LnAbnDSVI* is the dependent variable (Specifications 4-6). For each explanatory variable, the suffix *t*-j refers to the j_{th} lag of the corresponding variable, where *j* is from 1 to 5. For example "*AIA t*-1", is the first lag of *AIA*. "*EarnDum t*" is a dummy variable which receive the value of 1 for earnings announcements events and 0 otherwise. This captures the difference in attention between the two types of events. "*LnNumEvents t*" is the natural log of 1 plus the number of events on a given day. This captures the effect of multiple events on attention (Hirshleifer, Lim and Teoh, 2009). "*PSD/ADJ RSQ*" is the *Pseudo-RSQ* (*Adjusted-RSQ*) of the Probit panel models (OLS panel regressions). Standard errors are clustered by stock and day and each model includes quarter and day-of-week fixed effects. *t*-statistics are reported below the coefficient estimates.

	AIA *t*			LnAbnDSVI *t*		
	Probit			OLS		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
AIA t-1	0.884	0.783	0.754	0.015	0.009	0.008
	24.06	21.66	20.97	3.08	1.92	1.77
AIA t-2		0.328	0.269		0.009	0.007
		9.35	7.69		1.46	1.11
AIA t-3		0.431	0.318		0.017	0.013
		12.31	8.92		2.61	2.05
AIA t-5			0.309			0.009
			8.18			1.50
AIA t-5			0.364			0.013
			9.93			1.95
LnAbnDSVI t-1	0.098	0.063	0.056	0.632	0.664	0.660
	2.30	1.65	1.43	20.95	18.93	18.29
LnAbnDSVI t-2		0.004	0.001		-0.022	-0.019
		0.08	0.02		0.52	0.43
LnAbnDSVI t-3		0.028	0.004		-0.040	-0.031
		0.71	0.09		1.20	0.90
LnAbnDSVI t-4			0.019			-0.019
			0.51			0.46
LnAbnDSVI t-5			0.007			-0.001
			0.17			0.03
EarnDum t	0.607	0.704	0.748	0.029	0.031	0.032
	9.84	11.65	12.43	3.05	3.19	3.20
LnNumEvents t	0.0246	0.0280	0.0299	-0.0073	-0.0076	-0.0072
	0.91	1.06	1.14	1.75	1.83	1.71
PSD/ADJ RSQ	15.04%	16.94%	18.13%	42.18%	42.60%	42.08%

Figure 1 – Institutional Abnormal Attention, Earnings Announcements and News

The figure plots the daily *AIA* values for Overstock.com during 2013. As in Table 1, "*AIA*" is our measure of Abnormal Institutional Attention from Bloomberg. In addition, the figure plots earnings announcements days (indicated with an "E" above the plot) and the total number of news articles published on the firm in the RavenPack database. Additionally, sample headlines (from Factiva) for seventeen indicated events are listed below the figure.

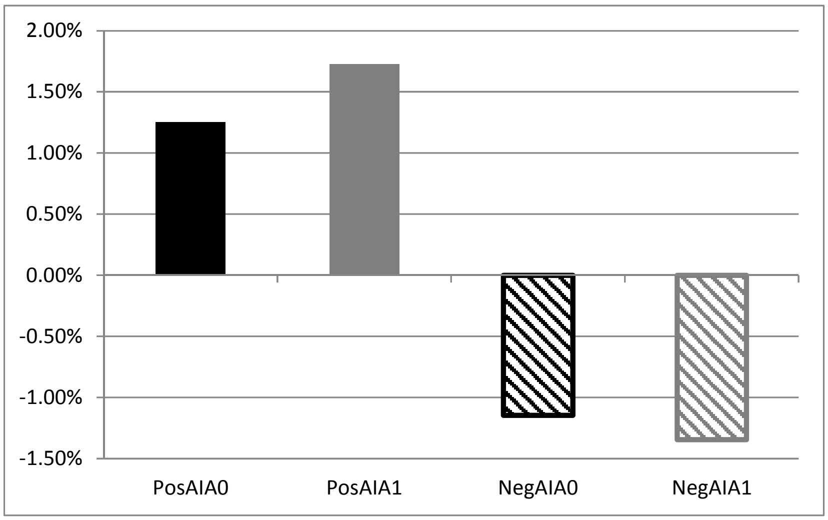


Event	Date	AIA	Sample Headline
1	1/18/2013	1	*Airport police: Overstock CEO arrested for having gun in luggage*
2	1/24/2013	1	*Overstock.com 4Q EPS 37c*
3	2/12/2013	1	*Overstock.com CEO to Take Medical Leave of Absence*
4	2/22/2013	0	*Overstock.com Inc Announces President Change-Form 8-K*
5	2/28/2013	1	*Groupon Shares Plunge on Profit, International Concerns*
6	3/28/2013	1	*Overstock.com Eyeing Supreme Court Appeal of Adverse NY Internet Tax Decision*
7	4/15/2013	0	*DJ Overstock CEO Byrne Resumes Duties After Medical Leave of Absence*
8	4/18/2013	1	*Overstock.com Reports Q1 2013 Results*
8	4/19/2013	1	*Overstock.com Raised to Buy From Underperform by BofA-Merrill Lynch*
9	5/30/2013	1	*Top 10 Nasdaq-traded stocks posting largest percentage decreases*
10	6/7/2013	1	*Officer JOHNSON III Sells 2,000 Of OVERSTOCK.COM INC*
11	7/3/2013	1	*Overstock.com downgraded at BofA/Merrill*
12	7/18/2013	1	*Overstock 2nd-quarter profit grows more than sevenfold, shares surge to multiyear high*
13	8/21/2013	1	*Overstock.com Victorious In Federal Lawsuit*
14	9/18/2013	0	*Overstock.com's 'The Good Good' sweepstakes awards entrants and their favorite charity*
15	10/17/2013	0	*Overstock.com's 3rd-quarter net income rises 31 percent as sales jump*
16	10/28/2013	1	*Overstock.com Presents Hot 99.5's Second Annual Jingle Ball in Washington, D.C.*
17	12/2/2013	1	*U.S. Supreme Court Won't Review New York Sales-Tax Law For Online Retailers*

Figure 2 – Abnormal Institutional Attention and Earnings Announcements Returns

The figure plots the effect of *SUE* on earnings announcements' day-0 *DGTW* risk adjusted returns (Graph 2.A) and day *t*+1 to *t*+40 cumulative risk adjusted returns (Graph 2.B) for the following four cases: positive *SUE* with *AIA* equals 0 ("*PosAIA0*"); positive *SUE* with *AIA* equals 1 ("*PosAIA1*"); negative *SUE* with *AIA* equals 0 ("*NegAIA0*"); and negative *SUE* with *AIA* equals 1 ("*NegAIA1*"). In order to estimate the conditional returns, for each group, we multiply the group's relevant *SUE* regression coefficient - estimated in Table 4.A - with the group's *SUE* average (i.e., the group's conditional mean). Since *AIA* is a dummy variable, we use the *SUE* regressions coefficient for *AIA* equals 0, and use the sum of *SUE* and *SUE_AIA* regression coefficients for *AIA* equals 1.

Graph 2.A – Day-0 Returns



Graph 2.B – *t*+1 – *t*+40 Cumulative Returns

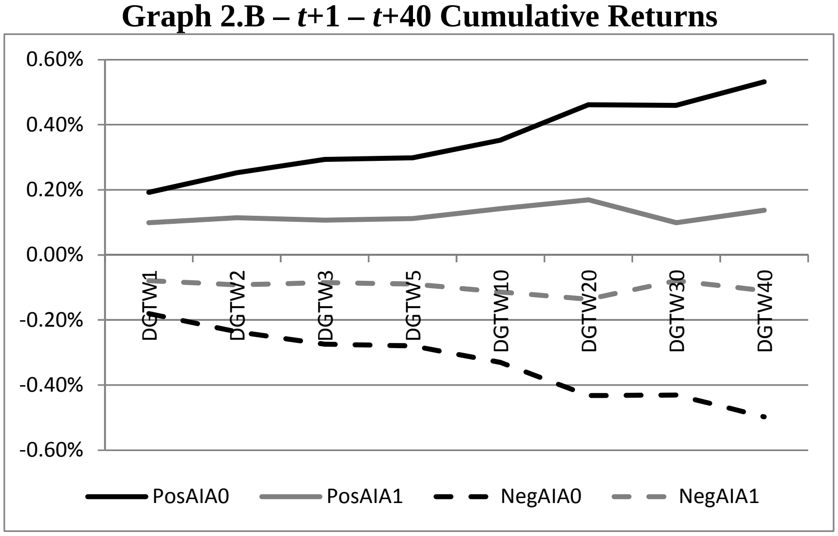
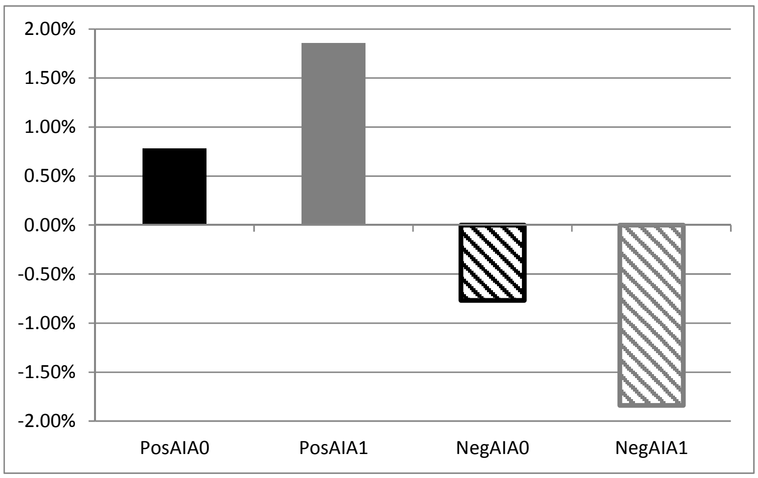


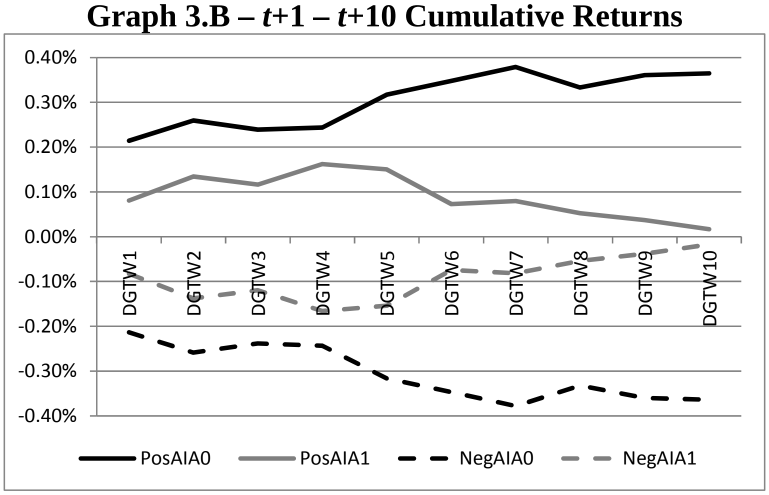
Figure 3 – Abnormal Institutional Attention and Change-in-Analyst-Recommendations Returns

The figure plots the effect of change in analyst recommendations *(RecChng)* on day-0 *DGTW* risk adjusted returns (Graph 3.A) and day *t*+1 to *t*+10 cumulative risk adjusted returns (Graph 3.B) for the following four cases: positive *RecChng* with *AIA* equals 0 ("*PosAIA0*"); positive *RecChng* with *AIA* equals 1 ("*PosAIA1*"); negative *RecChng* with *AIA* equals 0 ("*NegAIA0*"); and negative *RecChng* with *AIA* equals 1 ("*NegAIA1*"). In order to estimate the conditional returns, for each group, we multiply the group's relevant *RecChng* regression coefficient - estimated in Table 5.A - with the group's *RecChng* average (i.e., the group's conditional mean). Since *AIA* is a dummy variable, we use the *RecChng* regressions coefficient for *AIA* equals 0, and use the sum of *RecChng* and *RecChng _AIA* regression coefficients for *AIA* equals 1.

Graph 3.A – Day-0 Returns



Graph 3.B – *t*+1 – *t*+10 Cumulative Returns



Journal of Financial Economics 106 (2012) 660–684

Contents lists available at SciVerse ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Currency momentum strategies ☆

Lukas Menkhoff [a,1], Lucio Sarno [b,c,d,*], Maik Schmeling [a,3], Andreas Schrimpf [e,2]

[a] Department of Economics, Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, Germany
[b] Cass Business School, London, UK
[c] Singapore Management University, Singapore
[d] Centre for Economic Policy Research (CEPR), London, UK
[e] Bank for International Settlements, Centralbahnplatz 2, 4002 Basel, Switzerland

A R T I C L E I N F O

Article history:
Received 12 May 2011
Received in revised form
8 November 2011
Accepted 14 December 2011
Available online 4 July 2012

Keywords:
Momentum returns
Limits to arbitrage
Idiosyncratic volatility
Carry trades

JEL classification:
F31
G12
G15

A B S T R A C T

We provide a broad empirical investigation of momentum strategies in the foreign exchange market. We find a significant cross-sectional spread in excess returns of up to 10% per annum (p.a.) between past winner and loser currencies. This spread in excess returns is not explained by traditional risk factors, it is partially explained by transaction costs and shows behavior consistent with investor under- and overreaction. Moreover, cross-sectional currency momentum has very different properties from the widely studied carry trade and is not highly correlated with returns of benchmark technical trading rules. However, there seem to be very effective limits to arbitrage that prevent momentum returns from being easily exploitable in currency markets.

1. Introduction

Momentum returns in stock markets provide a strong challenge to standard finance theory. Simply buying assets with high recent returns and selling assets with low recent returns results in a very profitable investment strategy whose returns are difficult to understand by means of standard risk factors (Jegadeesh and Titman, 1993, 2001). Consequently, researchers have proposed various explanations that focus not only on conventional

☆ We would like to thank Bill Schwert (the Editor), Richard Levich (the Referee), Klaus Adam, Alessandro Beber, Craig Burnside, Darrell Duffie, Jacob Gyntelberg, Marcel Fratzscher, Jens Jackwerth, Gergana Jostova, Aneel Keswani, Michael Melvin, Christopher Neely, Jesper Rangvid, Stephan Siegel, Sheridan Titman, Christian Upper, and seminar participants at several institutions for helpful comments. We are also thankful for their input to several foreign exchange practitioners: Thomas Stolper (Goldman Sachs), Bilal Hafeez (Deutsche Bank), Gareth Berry (UBS), and Geoff Kendrick (Nomura). Sarno acknowledges financial support from the Economic and Social Research Council (No. RES-062-23-2340), Schmeling gratefully acknowledges financial support by the German Research Foundation (DFG), and Schrimpf is grateful for support from CREATES, funded by the Danish National Research Foundation. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank for International Settlements.
* Corresponding author. Faculty of Finance, Cass Business School, City University London, 106 Bunhill Row, London EC1Y 8TZ, UK.
Tel.: +44 20 7040 8772; fax: +44 20 7040 8881.
E-mail addresses: menkhoff@gif.uni-hannover.de (L. Menkhoff), lucio.sarno@city.ac.uk (L. Sarno), schmeling@gif.uni-hannover.de (M. Schmeling), andreas.schrimpf@bis.org (A. Schrimpf).
[1] Tel.: +49 511 7624552.
[2] Tel.: +41 61 280 8942.
[3] Tel.: +49 511 762 8213.

0304-405X/$ - see front matter
http://dx.doi.org/10.1016/j.jfineco.2012.06.009

risk-based models (see, e.g., Harvey and Siddique, 2000; Chordia and Shivakumar, 2002; Johnson, 2002; Pastor and Stambaugh, 2003; Liu and Zhang, 2011), but also on characteristics such as credit risk (Avramov, Chordia, Jostova, and Phlipov, 2007) or bankruptcy risk (Eisdorfer, 2008), limits to arbitrage (e.g., Chabot, Ghysels, and Jagannathan, 2009), behavioral explanations such as investor underreaction (e.g., Chui, Titman, and Wei, 2010), or high transaction costs (Korajczyk and Sadka, 2004). Despite this progress, the literature does not seem to have settled on a generally accepted explanation for momentum returns yet.

In this paper, we study foreign exchange (FX) markets as a natural laboratory for the analysis of momentum returns. Compared to stock markets, FX markets are more liquid and feature huge transaction volumes and low transaction costs, they are populated largely by sophisticated professional investors, and there are no natural short-selling constraints that prevent the shorting of past loser assets to fully implement momentum strategies. Hence, considering FX markets raises the hurdle for generating significant excess returns from momentum strategies considerably.

Surprisingly, there is little evidence on momentum in the *cross-section* of currencies. Large cross-country data sets were rare in the past so that the earlier literature has generally focused on momentum strategies in the time series of currencies, i.e., momentum strategies where individual currencies are bought and sold over time depending on various sorts of signals such as moving average cross-overs, filter rules, channel breakouts, etc. This literature has shown that certain technical trading rules were temporarily profitable but that their profits often tend to deteriorate over time as more traders learn about these strategies and start to exploit them (e.g., Levich and Thomas, 1993; Pukthuanthong-Le, Levich, and Thomas, 2007; Neely, Weller, and Ulrich, 2009, among others). A survey of this literature is provided by Menkhoff and Taylor (2007). However, some evidence on the existence of cross-sectional momentum profits in the FX market is provided by Okunev and White (2003), Asness, Moskowitz, and Pedersen (2009), and Burnside, Eichenbaum, and Rebelo (2011) in the context of small cross-sections of major currencies. Relative to our paper, these studies have a different focus, however, and do not provide a unifying analysis for understanding returns to cross-sectional currency momentum returns.

The main contribution of this paper is to study the economic anatomy of momentum profits in FX markets. We start by forming currency portfolios where an investor is long in currencies with high past excess returns (so-called "winners") and short in currencies with low past excess returns (so-called "losers"). We take the viewpoint of a U.S. investor and consider exchange rates against the U.S. dollar (USD). Our data cover the period from January 1976 to January 2010, and we study a cross-section of up to 48 currencies. We go beyond earlier research on currency momentum by (a) providing an in-depth analysis of the relative importance of systematic versus unsystematic risk for understanding momentum returns, (b) carefully comparing momentum strategies to carry trades and technical trading rules, (c) quantifying the importance of transaction costs, and investigating nonstandard sources of momentum returns, such as (d) under- and overreaction or (e) limits to arbitrage.

We find large and significant excess returns to currency momentum strategies of up to 10% per annum (p.a). As in Jegadeesh and Titman (2001), we find some evidence of return continuation and subsequent reversals over longer horizons of up to 36 months, which is consistent with behavioral biases, such as investor under- and overreaction, and suggests that momentum effects in different asset classes could share a common source. Importantly, currency momentum is very different from the popular carry trade in FX markets, providing high returns that are largely unrelated to carry trade returns.[4] Currency momentum returns are also different from returns generated by technical trading rules, which have been studied in a large empirical literature (e.g., Dooley and Shafer, 1976; Sweeney, 1986; Levich and Thomas, 1993; Neely, Weller, and Ulrich, 2009).

To rationalize these high excess returns of currency momentum strategies, we investigate whether currency momentum is significantly affected by (i) transaction costs, (ii) business cycle risk and other traditional risk factors, and (iii) different forms of limits to arbitrage. We find that momentum returns are indeed fairly sensitive to transaction costs. Adjusting returns for bid–ask spreads lowers the profitability of momentum strategies significantly since momentum portfolios are skewed towards currencies with high transaction costs. However, transaction costs are unable to completely account for currency momentum returns.

Also, momentum returns in FX markets are not systematically related to standard proxies for business cycle risk, liquidity risk (Brunnermeier, Nagel, and Pedersen, 2009), the carry trade risk factor proposed by Lustig, Roussanov, and Verdelhan (2011), volatility risk (Menkhoff, Sarno, Schmeling, and Schrimpf, 2012), the three Fama and French factors (Fama and French, 1992), or a four-factor model including a U.S. stock return momentum factor (Carhart, 1997). In short, there does not seem to be a systematic risk factor that would explain (net) momentum returns, a result that is akin to the corresponding findings based on U.S. equity momentum.

However, the profitability of currency momentum strategies varies significantly over time, which can induce limits to arbitrage for the major market participants in FX markets (e.g., proprietary traders and hedge funds), who usually have rather short investment horizons and could thus act myopically (e.g., Shleifer and Vishny, 1997).[5]

[4] The carry trade is a popular trading strategy that borrows in currencies with low interest rates and invests in currencies with high interest rates. According to uncovered interest parity, if investors are risk neutral and form expectations rationally, exchange rate changes will eliminate any gain arising from the differential in interest rates across countries. However, a number of empirical studies show that high interest rate currencies tend to appreciate, while low interest rate currencies tend to depreciate. As a consequence, carry trades form a profitable investment strategy, giving rise to the "forward premium puzzle" (Fama, 1984). See Burnside, Eichenbaum, Kleshchelski, and Rebelo (2011), Lustig, Roussanov, and Verdelhan (2011), and Menkhoff, Sarno, Schmeling, and Schrimpf (2012).

[5] We use the term "limits to arbitrage" here to mean that trading momentum strategies expose the investor to risks not captured by traditional covariance risk measures so that an anomaly like momentum returns is not easily exploitable. This definition is in line with much of

Furthermore, momentum returns are clearly related to currency characteristics. Returns are much higher in currencies with high (lagged) idiosyncratic volatility (about 8% p.a.) compared to currencies with low idiosyncratic volatility (about 4% p.a.). Returns are also related to measures of country risk, i.e., momentum strategies in countries with a high risk rating tend to yield significantly positive excess returns, whereas momentum strategies in countries with low risk ratings do not. Finally, a similar effect is found for a measure of exchange rate stability risk (i.e., the expected risk of observing large currency movements in the future).

In summary, we provide evidence that, despite FX markets' differences relative to stock markets, the properties of momentum strategies are fairly similar, which suggests that momentum profits in different asset classes could share a common root. Similar to stock markets, the high excess returns of currency momentum strategies can be (only) partially explained by their sensitivity to high transaction costs. Another piece of explanation of why momentum in currency markets persists is that there might be effective obstacles constraining the deployment of arbitrage capital to exploit the phenomenon. We find that currency momentum strategies are risky in that their returns are rather unstable over short time periods and that their exposure is subject to fundamental investment risk, captured by idiosyncratic characteristics of the currencies involved.

The remainder of this paper proceeds as follows. We selectively discuss earlier literature in Section 2. Section 3 details our data and portfolio formation procedure. Section 4 describes momentum returns in FX markets and compares momentum strategies with benchmark technical trading rules and the popular carry trade, while Section 5 discusses the results of our tests seeking to explain the high returns to currency momentum strategies. Section 6 provides robustness checks and Section 7 concludes. Additional results can be found in an Internet Appendix to this paper.

2. Related literature

Academic studies about momentum strategies are mostly focused on stock markets but momentum effects have been also detected in bond and commodity markets. To set the stage, we briefly survey this literature before we turn to FX markets and highlight the contributions of this paper.

Stock market momentum: Momentum effects are well documented in equity markets for almost two decades. The empirical literature is highly influenced by the work of Jegadeesh and Titman (1993), who show in a thorough analysis of the U.S. stock market that simple momentum strategies generate high returns, in the order of about 12% p.a., and are difficult to rationalize by standard asset pricing models. Subsequent studies extend the original research into new domains, including many countries worldwide beyond the U.S. (e.g., Rouwenhorst, 1998, 1999; Chan, Hameed, and Tong, 2000, Chui, Titman, and Wei, 2010) and higher frequencies (Gutierrez and Kelley, 2008).

While equity momentum is an established empirical fact, explanations have been heavily disputed. The major approaches to explain momentum can be classified as (i) risk-based and characteristics-based explanations, (ii) explanations invoking cognitive biases or informational issues, and (iii) explanations based on transaction costs or other forms of limits to arbitrage.

Starting with risk-based and characteristics-based explanations (i), early studies show that momentum returns are difficult to rationalize by covariance risk with standard factors (e.g., Fama and French, 1996; Jegadeesh and Titman, 2001). In the same vein, linking momentum to macroeconomic risk has proven rather challenging.[6] By contrast, firm-specific characteristics have been shown to be linked to momentum, e.g., momentum appears to be stronger among smaller firms (Hong, Lim, and Stein, 2000), among firms with lower credit rating (Avramov, Chordia, Jostova, and Philipov, 2007), and among firms with high revenue growth volatility (Sagi and Seascholes, 2007). Also, momentum returns appear to a large extent concentrated in firms with a high likelihood to go bankrupt (Eisdorfer, 2008).

Empirical work invoking behavioral biases (ii) in explaining momentum – focusing, for example, on investors' underreaction to news – also featured prominently since the beginning of the debate (Jegadeesh and Titman, 1993) and in subsequent work (e.g., Jegadeesh and Titman, 2001; Grinblatt and Han, 2005; Hvidkjaer, 2006).[7] Stressing how information is incorporated into prices, Chan, Jegadeesh, and Lakonishok (1996) provide early evidence that analysts' earnings forecasts respond gradually to news which can generate underreaction. Hong, Lim, and Stein (2000) demonstrate in detail the relation between weak analyst coverage and stronger momentum.[8] A final strand explores the role of transaction costs or limits to arbitrage (iii) in explaining momentum. Lesmond, Schill, and Zhou (2004) state that reasonably high transaction costs could wipe out momentum profits. Korajczyk and Sadka (2004) qualify this finding as they argue that momentum strategies can be designed in a way to limit transaction costs; this will lead to a more moderate cost level so that even very large momentum portfolios (with assets worth more than one billion U.S. dollars) are still highly profitable.

Momentum in bonds and commodities: Momentum has also been shown to exist in other asset classes. Regarding

(*footnote continued*)
the recent literature but it should be noted that the term (originally due to Keynes) initially referred to the market's inability to exploit risk-free arbitrage opportunities. Relative to this more precise definition, our tests are more closely related to "limits to speculation."

[6] For instance, Chordia and Shivakumar (2002) find support for time-varying risk factors explaining momentum returns, whereas Griffin and Martin (2003) and Cooper, Gutierrez, and Hameed (2004) do not.

[7] Behavioral models, e.g., by Daniel, Hirshleifer, and Subrahmanyam (1998), Barberis, Shleifer, and Vishny (1998), and Hong and Stein (1999) account for momentum effects by allowing for deviations from fully rational behavior such as overconfidence, slow updating of investor beliefs, and information imperfections.

[8] In addition, analyst behavior will lead, during the period of information incorporation, to information heterogeneity among investors, which is shown by Verardo (2009) to be related to momentum.

bond markets, momentum strategies do not work for investment-grade bonds (Gebhardt, Hvidkjaer, and Swaminathan, 2005) or bonds at the country level (Asness, Moskowitz, and Pedersen, 2009), but yield positive returns for non-investment grade *corporate* bonds (Jostova, Nikola, Philipov, and Stahel, 2010). Further analysis shows that momentum returns are not related to liquidity but seem to reflect default risk in the winner and loser portfolios. Regarding commodity markets, the high returns to momentum strategies are shown to be related to market states with low levels of inventories that indicate higher risk (Gorton, Hayashi, and Rouwenhorst, 2008). These findings tentatively suggest common sources of momentum profits that seem to be based on the risk characteristics of the underlying assets.

Currency momentum: In contrast to the extensive literature on momentum strategies in stock markets, the literature on currency momentum has mostly developed a somewhat different line of research. The most striking difference is the fact that currency momentum studies generally do not analyze momentum in a cross-section of currencies but in the time series of single exchange rates, often framed as "technical trading rules."[9] This literature is surveyed in Menkhoff and Taylor (2007) and we will discuss it in more depth below. This time-series literature has extensively examined which kinds of trading rules work best.

One exception from the time-series focus is Okunev and White (2003) who analyze a universe of eight currencies over 20 years, from January 1980 to June 2000. At the end of each month, the investor goes long in the currency with the best last-month performance and goes short in the currency with the worst last-month performance. This yields a return of about 6% p.a., which is largely independent of the base currency chosen and of the specific trading rule chosen, i.e., how exactly the best and worst currencies are identified. Thus, there is clear indication that currency momentum strategies can be profitable and thus worthy of a thorough examination.[10] Burnside, Eichenbaum, and Rebelo (2011) investigate returns to an equally weighted momentum portfolio that aggregates over momentum positions in individual currencies. They find (as we do in this paper) that standard risk factors cannot account for currency momentum returns.

Technical trading in FX markets: Technical trading in FX is in most cases the same as trend following, that is, exploiting the momentum of a market. These time-series momentum strategies include filter rules and moving average rules. A filter rule gives the signal to invest (to take a short position) in a currency if a defined upwards (downwards) exchange rate change has occurred, such as a 1% or 2% change. A moving average rule gives signals if short-term exchange rate averages become larger or smaller than longer-term averages.[11] Simple trend following trading strategies of this kind provide attractive returns, even considering interest rate differentials and transaction costs, as, for example, the early studies of Dooley and Shafer (1983) or Sweeney (1986) have demonstrated.[12]

These early studies have been challenged by subsequent work examining whether trend following trading strategies are also profitable in later periods. Whereas Dooley and Shafer (1983) and Levich and Thomas (1993) confirm profitability out-of-sample, studies also covering the 1990s and 2000s find that the above-mentioned simple trend following strategies applied to the same set of exchange rates no longer yield attractive returns (see, e.g., Olson, 2004; Pukthuanthong-Le, Levich, and Thomas, 2007; Neely, Weller, and Ulrich, 2009). However, profits are still found if either new forms of trend following strategies or new exchange rates are considered.[13]

Contributions of this paper: In contrast to the abundance of time-series studies, there is little evidence on cross-sectional aspects of currency momentum, whose importance has clearly risen in face of the realities of today's FX markets. Whereas there were about 10 convertible and liquid currencies in the 1970s, there are more than 30 currencies available to investors today. And while transaction volumes used to be dominated by banks' FX traders, asset managers of various kinds (including hedge funds) have emerged as some of the key players in today's FX markets. Overall, volumes, tradable assets, and participants have changed, which culminates in the perception of FX as a separate asset class, in parallel to, e.g., equities and bonds (King, Osler, and Rime, 2012). Even retail investors nowadays have access to various FX investment strategies via structured products. This naturally leads to studying cross-sectional currency momentum taking into account these new features and industry practices.[14]

In this paper, we go beyond earlier research in a number of directions. First, we analyze a much longer time span and, more importantly, a much larger cross-section of currencies that includes currencies of developed and emerging countries. This extended sample across time and currencies is crucial for our analysis of returns to currency momentum strategies since it allows

[9] See, e.g., Harris and Yilmaz (2009), Neely, Weller, and Ulrich (2009), and Serban (2010) in this respect.

[10] More recently, Asness, Moskowitz, and Pedersen (2009) have also investigated returns to a currency momentum strategy based on 10 currencies. The focus of their paper is very different from ours, however, with its primary objective being to explore the commonality of momentum across asset classes.

[11] For example, a 1,5 (or 5,20) rule suggests to buy Euro against US-dollar, if the 1- (5-) day US-dollar/Euro rate is higher than its 5-day (20-day) average.

[12] These strategies are also implemented in practice and the widespread use has led, e.g., Lequeux and Acar (1998), to build an index based on moving average rules to serve as a benchmark for Commodity Trading Advisors.

[13] Less well known and less studied forms include channel rules, genetic programming-based rules, Markov model-based rules, and others (e.g., Neely, Weller, and Dittmar, 1997). Neely and Weller (2012) provide a recent overview of different trading rules in currency trading. Neely, Weller, and Ulrich (2009) show that these rules are still profitable until the end of their sample period in 2005. Pukthuanthong-Le and Thomas (2008) confirm that standard trading rules in the main exchange rates do not generate profits when recent data are considered, whereas the same rules yield high returns in emerging markets' exchange rates.

[14] We thank the referee for pointing this out. An investment product such as the Currency Momentum ETF of Deutsche Bank, which is accessible even for retail investors, serves as an example of these new trends.

us to better identify return variation over time (and, hence, states of the business cycle) as well as across currencies that are structurally different and should have different exposures to global risk factors. Second, we can take explicit account of transaction costs, which is crucial since momentum returns are only relevant as long as they survive realistic transaction costs. Third, we take a close look at possible limits to arbitrage (which are a key theme in the recent literature on equity momentum) and investigate the role of idiosyncratic return volatility, country risk, and the risk of exchange rate stability. In sum, we provide a detailed account of the economic anatomy and drivers of currency momentum strategies that has been missing in the literature until now.

3. Data and currency portfolios

This section describes our data, the computation of currency excess returns, and the construction of momentum portfolios.

Data source and sample currencies: The data for spot exchange rates and one-month forward exchange rates cover the sample period from January 1976 to January 2010, and are obtained from Barclays Bank International (BBI) and Reuters (via Datastream). We denote the spot and forward rates in logs as s and f, respectively. Spot and forward rates are end-of-month data (last trading day in a given month) and are therefore not averaged over a month.

Our total sample consists of the following 48 countries: Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Euro area, Finland, France, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Ireland, Israel, Italy, Iceland, Japan, Kuwait, Malaysia, Mexico, Netherlands, New Zealand, Norway, Philippines, Poland, Portugal, Russia, Saudi Arabia, Singapore, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Ukraine, United Kingdom.

It is worth noting that, compared to, e.g., Lustig, Roussanov, and Verdelhan (2011) or Menkhoff, Sarno, Schmeling, and Schrimpf (2012), whose samples start in 1983 and have seven currency pairs in the beginning of the sample (mainly) based on BBI data quoted against the U.S. dollar, we employ a longer time series that extends back to 1976. We do so by complementing BBI data (which only start in 1983) with Reuters data quoted against the British Pound as in Burnside, Eichenbaum, Kleshchelski, and Rebelo (2011). We have a total of 16 currencies for this longer time span and convert these data to quotations against the U.S. dollar. These 16 countries are: Austria, Belgium, Canada, Denmark, France, Germany, Ireland, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. In addition to the larger cross-section and longer time series, we also have bid and ask quotes for spot and forward rates available so that we can adjust for transaction costs for the whole period from 1976 to 2010.

Finally, we note that our effective sample size varies over time as data for emerging countries become available or when currencies cease to exist, e.g., due to the adoption of the Euro. To illustrate this point, we plot the number of currencies with available data for each month of our sample in Fig. 1 (solid line). As can be inferred from this graph, our sample does not cover all 48 currencies at the same time since data availability varies naturally due to inclusion and exclusion of currencies. The total sum of actual observations (currency-month combinations) is

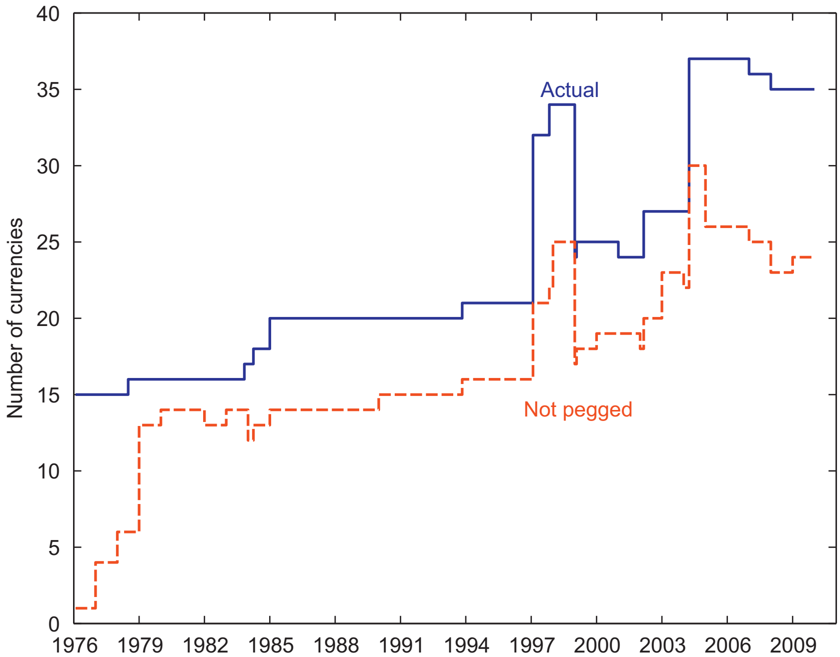


Fig. 1. Number of available currencies. The solid line shows the number of available currencies (i.e., currencies with available data for forward and spot exchange rates) and the dashed line shows the number of currencies with available data when excluding pegged currencies for each month of our sample period runs from January 1976 to January 2010.

9,403 as opposed to the theoretical maximum of 19,584 (408 months $\times$ 48 currencies). Individual start and end dates for each currency are shown in Table A1 in the Internet Appendix.

The dashed line in Fig. 1 also shows the number of available currencies that are not tightly pegged to other currencies. As can be expected, there are fewer currencies of this sort, especially in the very early part of the sample. While it is not problematic per se to perform momentum trading strategies in tightly linked currencies, one would expect that momentum profits should be relatively lower in the very early years of our sample. This is what we find in our empirical analysis below.[15]

Currency excess returns: Monthly excess returns to a U.S. investor for holding foreign currency k are given by

$$rx_{t+1}^k \equiv i_t^k - i_t - \Delta s_{t+1}^k \approx f_t^k - s_{t+1}^k, \tag{1}$$

where i^k denotes the one-month interest rate in country k, i without a superscript denotes the interest rate at home (the U.S. in our case), s and f denote the (log) spot and one-month forward rate (foreign currency unit per USD), respectively. Δs denotes the log spot rate change or return. If covered interest rate parity (CIP) holds, interest rate differentials $i_t^k - i_t$ equal forward discounts $f_t^k - s_t^k$. Akram, Rime, and Sarno (2008) show empirically that CIP holds even at very short horizons. Descriptive statistics for excess returns, forward discounts, and bid–ask spreads are reported in the Internet Appendix (Table A1).

For future reference, we also define net currency excess returns, i.e., currency excess returns after bid–ask spreads. These returns only apply when investigating dynamic investment strategies (momentum strategies in our case), where investors form portfolios of currencies. We detail the construction of portfolios below and simply define how we adjust for transaction costs here.

The net return for a currency that enters a portfolio at time t and exits the portfolio at the end of the month is computed as $rx_{t+1}^l = f_t^b - s_{t+1}^a$ for a long position and $rx_{t+1}^s = -f_t^a + s_{t+1}^b$ for a short position. An a (b) superscript indicates the ask (bid) quote. A currency that enters a portfolio but stays in the portfolio at the end of the month has a net excess return $rx_{t+1}^l = f_t^b - s_{t+1}$ for a long position and $rx_{t+1}^s = -f_t^a + s_{t+1}$ for a short position, whereas a currency that exits a portfolio at the end of month t but already was in the current portfolio the month before $(t-1)$ has an excess return of $rx_{t+1}^l = f_t^b - s_{t+1}^a$ for a long position and $rx_{t+1}^s = -f_t^a + s_{t+1}^b$ for a short position. Hence, since forward contracts in our sample have a maturity of one month, the investor always incurs transaction costs in the forward leg of his position but does not always have to trade the spot market leg of his position if he stays invested in a foreign currency. In addition, we assume that the investor has to establish a new position in each single currency in the first month (January 1976) and that he has to sell all positions in the last month (at the end of January 2010). Note that bid and ask rates are daily (not averaged over the month) so that they correspond exactly to the end-of-month data for spot and forward rates.

[15] We thank the referee for pointing this out.

However, one has to bear in mind that bid–ask spreads from BBI/Reuters are based on indicative quotes that are "too high" (see, e.g., Lyons, 2001) relative to actual effective spreads in FX markets so that our results with net returns (after deducting the bid–ask spread) should be understood as undercutting the lower bound on the profitability of momentum strategies and not as the "exact" return. For this reason, we frequently provide results with and without transaction costs below in our empirical analysis. We denote returns or spot rate changes after deducting bid–ask spreads as "net returns" and "net spot rate changes," respectively.

Portfolio construction: At the end of each month, we form six portfolios based on lagged returns over the previous $f=1,3,6,9,12$ months (f denotes the formation period) and these portfolios are held for $h=1,3,6,9,12$ months (h denotes the holding period). The one-sixth of all available currencies in a given month that have the lowest lagged returns are allocated to the first portfolio (denoted "Low"), the next sixth is allocated to portfolio 2, and so on, and the one-sixth of all currencies with the highest lagged returns are allocated to the sixth portfolio (denoted "High"). Hence, this procedure yields a time series of six currency momentum portfolios' excess returns and is analogous to the construction of momentum portfolios in the equity market literature.[16]

However, since interest rate differentials (forward discounts) contribute a significant share of the excess return of currency investments, we also track the pure spot rate changes of momentum portfolios themselves and report them separately in many tables. This way, we can check whether currency momentum is mainly driven by interest rate differentials or whether it occurs in spot rates, too.

Finally, in most analyses we work with the portfolio that is long in the winner currencies (portfolio "High") and short in the loser currencies (portfolio "Low"). These portfolios are denoted $MOM_{f,h}$ where f and h represent the formation and holding period, respectively, as defined above. We also refer to these portfolios simply as "long–short" momentum portfolios or "high-minus-low" portfolios. An important feature of these long–short portfolios is that they are dollar neutral, since the dollar component cancels out when taking the difference between (any) two portfolios.

4. Characterizing currency momentum returns

In this section, we present our main empirical results regarding the profitability and characteristics of currency

[16] Lustig and Verdelhan (2007) were the first to form portfolios of currency excess returns to be able to explain returns to the carry trade. This approach of forming currency portfolios has proved very useful in uncovering the economic drivers of carry trade risk premia and has been followed by several other papers afterwards. This way of constructing momentum returns differs from much of the earlier literature on technical trading in currency markets that mostly works in the time series of individual currency pairs (and then potentially aggregates across all currencies in the sample). Our approach is closer to how momentum is studied in the equity market literature and it is also closely related to how the financial industry sets up tradable momentum portfolios. For example, Deutsche Bank offers a currency momentum ETF based on G10 currencies and the underlying index is long (short) in the three best (worst) performing currencies over the last 12 months (Deutsche Bank, 2010).

momentum strategies (Section 4.1), the stability of the strategies out of sample (Section 4.2), the difference between currency momentum and technical trading rules (Section 4.3), the difference between currency momentum and carry trades (Section 4.4), and the long-run return behavior of momentum strategies (Section 4.5).

4.1. Returns to momentum strategies in currency markets

Table 1, Panel A, shows average annualized excess returns (left panel) and spot rate changes (right panel) for a number of high-minus-low momentum portfolios with formation and holding periods each varying between one and 12 months: $f,h=1,3,6,9,12$. Average excess returns in the left panel are based on sorting on lagged excess returns, and average spot rate changes in the right panel are based on sorting on lagged spot rate changes. To provide a perspective on profitability of FX momentum relative to risk, Panel B of Table 1 reports Sharpe Ratios for the same strategies.

Turning to excess returns in the left panel first, we find that momentum strategies yield substantial (and statistically highly significant) excess returns of about 6–10% for short holding periods of one month and their profits slowly fade out when increasing the holding period. The latter finding is quite pronounced since there is a monotone decline in average excess returns when moving from short holding periods to longer holding periods h for a given formation period f. However, we find many instances of significant momentum returns for strategies with longer holding periods as well, so that momentum is not confined to very short holding periods.

In the right panel of Table 1, Panel A, we also report the average difference between spot rate changes for the high and low portfolio. For ease of exposition, we actually report the negative of the log spot rate change (in the notation of Section 3) so that higher values indicate a positive contribution of spot rate movements to a momentum strategy's total excess return. Interestingly, the profitability of currency momentum strategies is also clearly visible in spot rate changes themselves and is thus not mostly driven by the interest rate differential as is the case for carry trades (see, e.g., Lustig, Roussanov, and Verdelhan, 2011). In fact, the strategy with a 12-month formation period is completely driven by favorable spot rate changes and the interest rate differential reduces the excess return somewhat.

Table 1
Momentum returns and Sharpe Ratios.

This table shows annualized average returns for different momentum strategies ($\overline{r^{f,h}}$) in Panel A. The rows show formation periods (f) whereas the columns indicate holding periods (h) in months. Numbers in brackets are t-statistics based on Newey and West (1987) heteroscedasticity and autocorrelation consistent (HAC) standard errors. The left part of the table shows currency excess returns (spot rate changes adjusted for interest rate differentials) whereas the right part shows pure spot rate returns. Panel B shows annualized Sharpe Ratios. t-Statistics based on a moving block-bootstrap are in squared brackets. The right panel shows average annualized spot rate changes (in percent) divided by the annualized standard deviation of mean exchange rate changes. The sample period is January 1976–January 2010 and we employ monthly returns.

Panel A: Excess returns and spot rate changes

	Excess returns						Spot rate changes				
	Holding period h						Holding period h				
f	1	3	6	9	12	f	1	3	6	9	12
1	9.46	7.00	6.17	5.15	5.75	1	7.91	4.42	3.38	4.75	3.13
	[5.31]	[4.11]	[3.13]	[2.73]	[3.6]		[4.55]	[3.07]	[1.93]	[2.94]	[2.02]
3	9.40	6.32	4.96	4.67	4.43	3	8.54	5.73	5.28	4.63	5.10
	[5.30]	[3.80]	[3.03]	[2.92]	[2.74]		[5.10]	[3.59]	[3.66]	[2.88]	[3.51]
6	8.54	6.31	3.66	3.25	3.14	6	6.50	5.75	3.47	3.64	3.17
	[4.78]	[3.63]	[2.06]	[1.79]	[1.69]		[3.88]	[4.00]	[2.15]	[2.32]	[1.80]
9	7.18	6.80	5.36	3.86	3.24	9	8.33	7.06	6.50	4.91	4.09
	[3.80]	[3.65]	[2.86]	[2.05]	[1.67]		[4.82]	[4.23]	[3.91]	[2.87]	[2.35]
12	6.16	5.48	3.02	2.05	1.89	12	7.59	6.04	3.94	3.19	3.03
	[3.40]	[3.24]	[1.75]	[1.17]	[1.04]		[4.63]	[4.02]	[2.59]	[1.97]	[1.92]

Panel B: Sharpe Ratios and normalized spot rate changes

	Excess returns						Spot rate changes				
	Holding period h						Holding period h				
f	1	3	6	9	12	f	1	3	6	9	12
1	0.95	0.76	0.59	0.56	0.61	1	0.84	0.53	0.37	0.57	0.37
	[5.48]	[4.10]	[3.15]	[2.47]	[2.95]		[5.52]	[4.23]	[3.25]	[2.81]	[3.21]
3	0.88	0.60	0.50	0.53	0.51	3	0.86	0.57	0.58	0.50	0.63
	[5.37]	[3.70]	[3.04]	[2.74]	[2.42]		[5.17]	[3.73]	[3.45]	[2.99]	[2.61]
6	0.79	0.60	0.37	0.34	0.33	6	0.64	0.60	0.38	0.41	0.35
	[4.55]	[3.53]	[1.94]	[1.76]	[1.48]		[4.76]	[3.70]	[2.06]	[2.05]	[1.43]
9	0.67	0.63	0.50	0.36	0.30	9	0.85	0.71	0.66	0.51	0.41
	[3.76]	[3.61]	[2.95]	[1.95]	[1.57]		[3.99]	[3.66]	[3.07]	[2.12]	[1.84]
12	0.61	0.56	0.32	0.21	0.19	12	0.77	0.64	0.44	0.35	0.33
	[3.18]	[3.05]	[1.64]	[1.17]	[1.05]		[3.48]	[3.32]	[1.89]	[1.27]	[1.14]

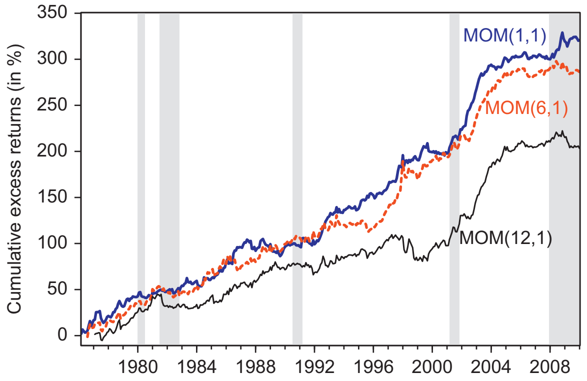


Fig. 2. Cumulative excess returns of momentum strategies. This figure shows cumulative log excess returns (not adjusted for transaction costs) accruing to three different momentum returns. The momentum strategies are for a formation period of 1, 6, and 12 months, respectively, and the holding period is one month. The bold line shows returns to the momentum strategy with a one-month formation period (MOM(1,1) in the figure), the dashed line shows returns to a strategy with a six-month formation period (MOM(6,1)), whereas the thin, black line shows returns to a momentum strategy with a 12-month formation period (MOM(12,1)). Shaded areas correspond to NBER recessions.

As noted above, results tend to be strongest for a holding period of $h=1$ month. We therefore focus on these strategies in most of the following analysis as they seem to present the hardest challenge when trying to understand momentum returns in currency markets. Since the level of average excess returns is also clearly dependent on the formation period f, we provide results for the three strategies with $f=1,6$, and 12 months in our empirical analyses below. In sum, most of our analysis in the remainder of the paper focuses on the three benchmark strategies $MOM_{1,1}$, $MOM_{6,1}$, and $MOM_{12,1}$.[17]

As a first and simple means of investigating a possible link between momentum returns and the state of the business cycle, and to provide a graphical exposition of momentum returns accruing to investors, Fig. 2 shows cumulative excess returns for the three benchmark momentum strategies $MOM_{1,1}$,$MOM_{6,1}$, and $MOM_{12,1}$ over the full sample period. Shaded areas correspond to NBER recessions. As illustrated by the figure, there is no obvious correlation of momentum returns with the state of the business cycle (as examined later in Section 5.2). However, the three benchmark momentum strategies show some comovement but are not perfectly correlated.

Sharpe Ratios: To get a first measure of risk-adjusted returns, Panel B of Table 1 presents Sharpe Ratios for the momentum strategies shown in Table 1 above in Panel A, and "normalized spot rate changes" (average spot rate changes divided by their standard deviation) in Panel B. Corroborating the evidence above, currency momentum strategies seem highly profitable, at least for a subset of strategies. For example, the annualized Sharpe Ratio of the MOM(1,1) strategy is 0.95, which seems very high, even in comparison to carry trades. See, e.g., Menkhoff, Sarno, Schmeling, and Schrimpf (2012) who report an annualized Sharpe Ratio of 0.82 for a carry trade strategy. Hence, even when taking risk into account on the basis of Sharpe Ratios, momentum strategies seem highly attractive. In addition, we see from Panel B of Table 1 that this performance is largely driven by spot rate changes and that it is not dominated by the interest rate component of excess returns.

Momentum returns and size of the cross-section: As noted above in the previous section, our effective sample size never exceeds 40 currencies and is therefore relatively small compared to sample sizes used in, e.g., the equity momentum literature. However, it is well known from earlier work that even small portfolios of currencies can yield large gains from diversification since currencies tend to be less correlated than stocks (e.g., Burnside, Eichenbaum, and Rebelo, 2008). To explore the link between the size of the cross-section and the magnitude of momentum returns, we conduct a stylized simulation experiment as follows. In each run i, we randomly draw (without replacement) a set of N currencies from the set of all 48 currencies while imposing the restriction that we have data for at least six currencies in each month of the sample period from January 1976 to January 2010. We then calculate average annualized momentum excess returns for a MOM(1,1) strategy and save this result. We do this 5,000 times for each cross-section size N and average over momentum profits to obtain an estimate of the "typical" momentum profit conditional on observing a cross-section of size N. For $N=48$, we simply report the momentum profit from Table 1.

Fig. 3 shows results from this exercise and it can be seen that expanding the size of the cross-section is very useful for small cross-sections but much less important for larger cross-sections. In other words, there are decreasing gains from expanding the size of the tradable currency universe. The maximum level of returns is roughly obtained for a

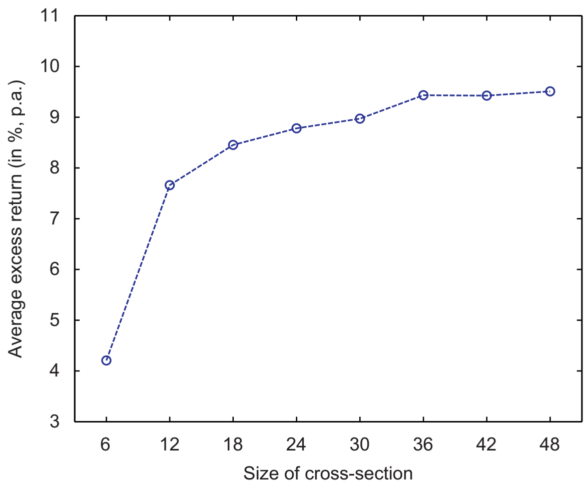


Fig. 3. Size of the cross section and momentum returns. This figure shows average annualized excess returns for a MOM(1,1) strategy implemented on a cross section of 6, 12, 18,...,48 currencies. We draw 5,000 random combinations of currencies for each size of the cross section (imposing the restriction that we have at least six currencies at each point in time) and plot the average (across simulations) annualized mean excess return of a MOM(1,1) strategy.

[17] One might worry that some currencies were not always tradable during our sample period due to, e.g., capital account restrictions. We provide robustness checks on this issue in Section 6.1.

Table 2
Momentum^2.

This table shows average momentum excess returns and Sharpe Ratios (SR) for momentum strategies based on other momentum portfolios. We first calculate monthly excess returns for all 144 possible momentum portfolios based on formation and holding periods of $f=1,2,\ldots,12$ and $h=1,2,\ldots,12$. Next we run a momentum strategy on these 144 momentum portfolios and sort momentum strategies from the first step into nine portfolios based on their lagged returns over an evaluation period. Lagged returns over the evaluation period (shown in the first column) vary from one to 120 months. We report results for the nine portfolios from the second stage (from "Worst" lagged strategy to "Best" lagged strategy) and a high-minus-low portfolio (best strategy minus worst strategy, "B − W") that is long in the best 16 (144/9) strategies and short in the worst 16 (144/9) strategies from the first step. Numbers in brackets are HAC t-statistics based on Newey and West (1987). The sample period is January 1976–January 2010 and we employ monthly returns.

		Momentum^2-Portfolios									
Lag		Worst	2	3	4	5	6	7	8	Best	B − W
1	Mean	1.03	2.45	3.44	3.78	4.34	5.46	5.82	6.85	8.70	7.67
	t	[0.79]	[1.74]	[2.40]	[2.54]	[2.90]	[3.50]	[3.76]	[4.46]	[5.68]	[4.31]
	SR	0.14	0.31	0.42	0.45	0.49	0.62	0.63	0.75	0.94	0.73
3	Mean	1.21	3.02	3.85	3.53	4.51	5.14	5.72	6.67	7.62	6.41
	t	[1.04]	[2.28]	[2.71]	[2.46]	[3.03]	[3.37]	[3.72]	[4.24]	[5.05]	[4.12]
	SR	0.20	0.43	0.50	0.43	0.53	0.58	0.62	0.72	0.83	0.71
6	Mean	3.40	2.99	3.47	3.92	3.93	4.69	6.09	6.54	7.88	4.48
	t	[2.87]	[2.39]	[2.63]	[2.87]	[2.62]	[3.17]	[3.94]	[4.27]	[5.04]	[3.08]
	SR	0.54	0.45	0.48	0.52	0.47	0.54	0.67	0.72	0.85	0.54
9	Mean	3.59	3.82	4.04	4.61	4.81	5.28	5.55	6.07	7.10	3.51
	t	[3.02]	[3.11]	[3.13]	[3.43]	[3.44]	[3.43]	[3.69]	[3.89]	[4.52]	[2.53]
	SR	0.57	0.57	0.56	0.60	0.59	0.60	0.63	0.68	0.79	0.44
12	Mean	3.29	4.57	4.57	4.42	4.86	4.98	5.54	5.87	7.25	3.96
	t	[2.76]	[3.69]	[3.39]	[3.28]	[3.46]	[3.51]	[3.44]	[3.75]	[4.83]	[3.14]
	SR	0.53	0.66	0.62	0.58	0.59	0.59	0.61	0.66	0.81	0.51
60	Mean	3.83	3.75	4.24	4.82	4.15	5.28	5.18	5.42	5.92	2.09
	t	[2.70]	[2.60]	[2.90]	[3.27]	[2.75]	[3.62]	[3.25]	[3.49]	[3.86]	[1.88]
	SR	0.55	0.53	0.58	0.63	0.52	0.66	0.60	0.64	0.68	0.33
120	Mean	3.69	4.21	4.82	4.60	4.54	5.01	5.75	5.76	6.40	2.70
	t	[2.16]	[2.48]	[2.80]	[2.61]	[2.62]	[2.88]	[3.20]	[3.18]	[3.59]	[2.03]
	SR	0.49	0.56	0.61	0.55	0.55	0.61	0.68	0.65	0.74	0.41

cross-section of size $N=36$. Hence, although our cross-section is far from what is used in the equity market literature, one can be confident that the results are quite representative of the currency market as a whole.

4.2. Out-of-sample perspective

Our setup to illustrate FX momentum profits, which is akin to the equity literature, has a clear out-of-sample component, since we form portfolios based on lagged information only. Hence, the momentum strategies discussed above are implementable in real time. However, average returns can vary markedly across different strategies (that is, different combinations of formation and holding period), and can also be fairly low. For example, the strategy with a 12-month formation and holding period only yields 1.89% p.a. over the full sample whereas the strategy with a one-month formation and holding period experienced an annualized average return close to 10%. This particular information is only available ex post and an investor could not have conditioned on this information in 1976. Hence, it is interesting to examine whether investors could have actually exploited these momentum profits taking into account that there is ex ante uncertainty about which specific momentum strategy to follow.[18] Put differently, do specific momentum strategies identified to be attractive in-sample continue to do well?

We tackle this question by investigating returns to what we term "Momentum^2" strategies. To do so, we imagine an investor who can invest in 144 different strategies (all combinations of $f=1,2,\ldots,12$ and $h=1,2,\ldots,12$) and has to rely on some mechanism to select between these different strategies. A natural mechanism in our context is to let the investor rely on momentum in lagged momentum returns (as measured over an evaluation period). More specifically, we form nine portfolios out of the universe of 144 possible momentum strategies. These nine portfolios are based on a ranking *of the momentum strategies* themselves by their lagged returns during an evaluation period, hence the term Momentum^2. Results for this exercise are shown in Table 2 which reports returns for all nine Momentum^2 portfolios (from "worst" lagged returns to "best" lagged returns) and a "best minus worst" portfolio. For robustness we show

[18] Silber (1994) also investigates whether trading strategies identified as profitable over an in-sample period continue to perform well in an out-of-sample period.

results for lags of 1,3,6,9,12,60, and 120 months over which individual momentum strategies are evaluated. As can be seen, using lagged momentum returns to identify future momentum returns seems feasible. For example, conditioning just on last month's return across all possible strategies leads to an annualized average excess return of 7.67% p.a. As for the simple momentum strategies above, we see a declining pattern in returns when moving to longer selection windows. For example, using a window of 120 months leads to much lower returns of only 2.70% p.a., which, however, are still significantly different from zero. Most importantly, these results indicate, however, that specific FX momentum strategies that performed well in the past tend to continue to do well and are thus quite stable.

While the above analysis confirms that momentum returns are exploitable in an out-of-sample setting, we further examine this issue from a somewhat different angle by a simple investigation of the subsample stability of momentum profits. To do so, Table A.2 in the Internet Appendix shows average annualized excess returns and Sharpe Ratios for four subperiods of equal length. We report results for formation periods of $f=1,3,6,9,12$ and a holding period of one month. As can be seen, the ranking of these five different strategies is fairly stable over the four subperiods. In other words, it is never the case that one strategy does extremely well in one subperiod but then produces large losses in the next subperiod. Overall, we conclude that it should have been possible for an investor to exploit momentum strategies in real time.

4.3. *Comparing momentum and technical trading rules*

The results presented above suggest that momentum effects in the cross-section of currencies are quite strong and that momentum strategies consequently yield high excess returns and Sharpe Ratios. However, an important question is whether the currency momentum returns documented above can be regarded as a novel phenomenon per se or whether they merely reflect returns to technical trading strategies that have been documented extensively in the earlier literature.

To investigate this issue, we compute returns to three benchmark moving average cross-over rules that have been employed frequently in earlier work on technical trading in FX markets. These strategies are based on moving averages of 1 and 20 days (1,20), 1 and 50 days (1,50), and 1 and 200 days (1,200) (see, among others, Dooley and Shafer, 1983; Levich and Thomas, 1993, Neely, Weller, and Ulrich, 2009).[19] While it is clearly not the case that these three strategies are perfect proxies for *all* possible technical trading strategies, their prominence in the earlier literature makes them interesting for comparison with our cross-sectional currency momentum strategies.

To set the stage, we first compute returns to these moving average rules for all currencies in our sample individually and then aggregate these strategies into an equally weighted portfolio. Panel A of Table 3 reports descriptive statistics for the three rules, which show that these strategies are profitable, with annual mean excess returns around 5% and high annual Sharpe Ratios between 0.77 and 0.88. Hence, these strategies form an interesting benchmark for our momentum returns.

To assess whether returns to the moving average rules described above capture returns to the currency momentum strategies, we run regressions of momentum returns for the MOM(1,1), MOM(6,1), and MOM(12,1) strategies on returns of the three moving average rules. Results are shown in Panel B of Table 3.

It can be seen that, even though moving average rule returns and currency momentum are to some extent correlated, the largest R^2 only amounts to 26%. More importantly, all intercept estimates (α's) are large in economic terms and strongly significant in statistical terms. Hence, it seems fair to conclude that currency momentum is not closely related to benchmark technical trading strategies as studied in the earlier literature, and that controlling for returns of these trading rules does not wipe out returns to our cross-sectional currency momentum.

In addition, we also examine returns to individual currencies' momentum strategies, i.e., where an investor is long or short in each currency depending on lagged returns in the same currency. This strategy is also studied in Moskowitz, Ooi, and Pedersen (2012). We report descriptive statistics for returns of each currency in Panel A of Table A.3 in the Internet Appendix, along with the average across countries, an equally weighted portfolio of all individual currencies' strategies, and, for comparison, the cross-sectional momentum strategy employed in this paper in Panel B. It can be seen that most of these time-series momentum strategies are profitable on average (Panel A of Table A.3) but that an aggregate strategy (the equally weighted portfolio, EW, in Panel B) is less profitable than a cross-sectional momentum strategy (MOM(1,1) in Panel B), as the latter strategy has a much higher average excess return (almost twice as high) and Sharpe Ratio.

4.4. *Comparing currency momentum and the carry trade*

An important question is to what extent momentum strategies simply capture the same information as the popular carry trade strategy in FX markets, where investors go long in high interest rate currencies and short in low interest rate currencies. After all, interest rate differentials are strongly autocorrelated and spot rate changes do not seem to adjust to compensate for this interest rate differential, which is well-known in the literature as the "forward premium puzzle" (Fama, 1984). Hence, it might be the case that lagged high returns simply proxy for lagged high interest rate differentials and that, therefore, currency momentum returns are very similar to carry trade returns. To address this concern, we perform a comprehensive comparison between momentum returns and carry trade returns in this section. The results clearly show that carry trade and momentum strategies, as well as their associated returns, are in fact very different.

Return correlations: Table 4, Panel A, shows correlation coefficients between returns to momentum portfolios and

[19] These trading strategies generate a buy (sell) signal, when the shorter moving average crosses the longer moving average from below (above).

Table 3

Moving average rules and cross-sectional momentum.

This table shows means, Sharpe Ratios (SR), standard deviations, skewness, and kurtosis for excess returns to three benchmark moving average (MA) rules in Panel A. Panel B shows results from regressions of cross-sectional momentum excess returns (i.e., high-minus-low portfolios) on a constant and excess returns to each of the three MA rules. Note that the adjusted R^2's in Panel B are in percent. The sample period is January 1976–January 2010 and we employ monthly returns.

Panel A: Descriptive statistics for MA rules

	(1,20)	(5,20)	(1,200)
Mean	5.27	5.14	5.23
	[5.56]	[5.73]	[4.64]
SR	0.88	0.83	0.77
St. dev.	5.98	6.22	6.81
Skewness	0.67	0.40	0.09
Kurtosis	4.63	4.71	4.97

Panel B: Regressions of cross-sectional momentum returns on MA rule returns

	MOM(1,1)			MOM(6,1)			MOM(12,1)		
$\alpha_{(1,20)}$	7.74			7.63			6.21		
	[4.54]			[4.60]			[3.62]		
$\beta_{(1,20)}$	0.33			0.17			−0.01		
	[3.57]			[1.41]			[−0.12]		
$\alpha_{(5,20)}$		7.80			7.49			5.84	
		[4.52]			[4.45]			[3.35]	
$\beta_{(5,20)}$		0.32			0.21			0.06	
		[3.95]			[1.72]			[0.60]	
$\alpha_{(1,200)}$			6.90			4.16			7.39
			[3.97]			[2.46]			[2.82]
$\beta_{(1,200)}$			0.47			0.81			0.03
			[5.67]			[7.56]			[0.16]
$\overline{R}^2$ (in %)	3.62	3.88	10.34	0.68	1.17	26.00	−0.25	−0.10	−0.23

carry trade portfolios. We show results for the long–short momentum strategies $MOM_{1,1}$, $MOM_{6,1}$, and $MOM_{12,1}$, and always report the correlation between corresponding portfolios; e.g., the correlation of momentum portfolio 2 and carry trade portfolio 2, or the correlation between the high-minus-low (H−L) carry trade and momentum portfolios. It can be seen that the correlations of excess returns for the six portfolios are rather high but that there is basically no correlation between the high-minus-low portfolios, and the latter represent the way carry trade and momentum strategies are typically implemented by market participants. Thus, the return to following a currency momentum strategy is basically uncorrelated with carry trade returns and this finding holds true regardless of the respective formation period underlying a momentum strategy.

In contrast, we show in Panel B that the high-minus-low portfolios of the three momentum strategies are much more highly correlated and reach correlations of more than 70% for $MOM_{6,1}$ and $MOM_{12,1}$. Hence, it seems fair to conclude that returns to different momentum strategies are likely to share a strong common component.

That excess returns to carry trades and momentum strategies are basically uncorrelated in FX markets appears in line with real-world strategies of many currency investors who combine momentum and carry trade positions in their portfolios to take advantage of an alleged diversification benefit from following the two strategies simultaneously.[20] For example, during the recent financial crisis from July 2007 to June 2009, the benchmark momentum strategy with $h=f=1$ experienced an average monthly return of 0.80% whereas the carry trade yielded a negative average monthly return of −0.05%. The return correlation of these two strategies was as low as −31% over these two years. Hence, the two strategies showed a clearly different behavior during this period.

Comparing portfolio properties: We additionally investigate characteristics of momentum and carry trade portfolios, which are reported in Table A.4 in the Internet Appendix. The table shows descriptive statistics for the six momentum portfolios with a formation and holding period of one month and six carry trade portfolios where currencies are sorted into portfolios depending on their lagged interest rate, as in, e.g., Lustig, Roussanov, and

[20] Patton and Ramadorai (in press), for example, show in a general universe of hedge funds (not necessarily currency funds) that there is significant exposure to carry trade and momentum-type returns and that this exposure is time-varying. Pojarliev and Levich (2010) show via style regressions that currency fund managers engage in both carry trade and momentum-type strategies. Melvin and Shand (2011) show that currency managers follow momentum strategies but that their exposure to momentum and the way momentum strategies are implemented change over time.

Table 4
Correlation of momentum and carry trade returns.

This table shows correlation coefficients between portfolio returns. Panel A shows correlation coefficients between momentum returns based on strategies with formation horizons of f equal to one, six, and 12 months and holding periods of $h=1$ month (denoted $MOM_{1,1}$, $MOM_{6,1}$, $MOM_{12,1}$, respectively) and forward discount-sorted portfolio returns (denoted C since they form the basis of the carry trade). Returns are based on six portfolios and a long-short portfolio for both momentum and the carry trade. We only report correlations for corresponding pairs of portfolios. For example, in row $\rho(M_{1,1},C)$, we report the correlation of the "Low" momentum portfolio with the "Low" carry trade portfolio in column "Low," the correlation of the third momentum portfolio with the third carry trade portfolio, and so on for all six portfolios and the long-short portfolios. Row $\rho(M_6,C)$ shows the correlations between portfolio pairs of the momentum strategy with a six-month formation period with the carry trade and row $\rho(M_{12},C)$ shows the correlations between portfolio pairs of the 12-month formation period momentum strategy and the carry trade. Panel B shows correlations for momentum portfolios with different formation horizons. The sample period is January 1976–January 2010 and we employ monthly returns.

Panel A: Momentum and carry trade portfolios							
	Low	2	3	4	5	High	H−L
$\rho(MOM_{1,1},C)$	0.68	0.84	0.83	0.85	0.81	0.73	0.04
$\rho(MOM_{6,1},C)$	0.63	0.84	0.82	0.83	0.81	0.74	0.01
$\rho(MOM_{12,1},C)$	0.67	0.85	0.81	0.87	0.82	0.74	0.07
Panel B: Momentum portfolios							
	Low	2	3	4	5	High	H−L
$\rho(MOM_{1,1},MOM_{6,1})$	0.77	0.83	0.88	0.85	0.83	0.79	0.45
$\rho(MOM_{1,1},MOM_{12,1})$	0.66	0.81	0.86	0.87	0.80	0.78	0.28
$\rho(MOM_{6,1},MOM_{12,1})$	0.82	0.89	0.89	0.89	0.91	0.89	0.73

Verdelhan (2011) or Menkhoff, Sarno, Schmeling, and Schrimpf (2012).[21]

As can be inferred from this table, there is a monotonically increasing pattern in average returns for both cross-sections but no clear pattern in higher moments of the return distribution. While the level of average returns and standard deviations of the high-minus-low momentum and carry trade portfolios is roughly similar, we find that the two long-short portfolios are clearly different in terms of their skewness. While the carry trade produces negatively skewed excess returns (Brunnermeier, Nagel, and Pedersen, 2009), we find a slightly positive skewness for the momentum strategy.

More interestingly, the last two rows of each panel show lagged average returns and lagged average forward discounts for each portfolio at the time of portfolio formation. Momentum portfolios do have a positive spread in forward discounts and carry trade portfolios have a positive spread in lagged returns, but these spreads are much lower in absolute value than the spread in the characteristic used for sorting currencies into portfolios. More specifically, the average cross-sectional spread in forward discounts (in annualized terms) at the time of portfolio formation is about 4.6% (5.13% versus 0.44%) for the momentum cross-section but averages more than 15% for the carry trade cross-section. Similarly, the average spread in lagged returns is almost 6% for the momentum portfolios (2.94% versus −2.93%) but only 0.84% for the carry trade cross-section. Hence, momentum and carry trade strategies seem far from being identical.

Double sorts: Next, we provide results based on double sorts. To this end, we first double sort currencies into two portfolios depending on whether a currency has a lagged forward discount above or below the median (of all available currencies), and then into three portfolios depending on their lagged excess return. Portfolios are rebalanced each month (i.e., $h=1$). Table 5 shows results for these double sorts for formation periods of $f=1,3,12$ months. There is no material difference between momentum returns among high versus low interest rate currencies. For example, the high-minus-low momentum return for a strategy with a one-month formation period based on low interest rate currencies is 5.06% p.a., on average, whereas it is 5.36% p.a. for high interest rate currencies. Hence, the difference between these two high-minus-low momentum portfolios is less than 0.30% p.a. and not statistically significant (with a t-statistic of only 0.17). Findings for the other two formation periods are very similar.

As above, we do not find a strong relation between momentum and carry trade strategies and the double sorts suggest that the two strategies are largely independent. In fact, going long in currencies with high lagged returns and high interest rates whilst shorting currencies with low returns and low interest rates generates an excess return of 10.52% p.a. that is even larger than the spread in both momentum or carry trade portfolios taken individually.

Cross-sectional regressions: Finally, we want to separate the effects of lagged excess returns and lagged interest rate differentials on future excess returns. To this end, we run Fama-MacBeth type cross-sectional regressions of currency excess returns (or spot rate changes) on (i) lagged excess returns over the last l months, (ii) lagged forward discounts, and/or (iii) lagged spot rate changes for each month of our sample, i.e.,

$$rx_t^k = \alpha_t + \beta_{rx,t} rx_{t-\ell;t-1}^k + \beta_{FD,t}(f_{t-1} - s_{t-1}) + \beta_{\Delta s,t} \Delta s_{t-\ell;t-1}^k + \varepsilon_t, \tag{2}$$

where the subscript $t-\ell;t-1$ refers to a variable defined over the last ℓ months using information available at time $t-1$. This procedure yields a time-series of coefficient estimates (α_t,β_t) and we report the mean of these time series and t-statistics based on Newey and West (1987) standard errors in Table 6 in the spirit of the approach by Fama and MacBeth (1973).[22]

These cross-sectional regressions serve to disentangle the information contained in lagged returns (or spot rate changes) and forward discounts for future excess returns (or spot rate changes) in a regression framework and on the level of individual currencies. Momentum strategies

[21] To conserve space in this table, we focus on the momentum strategy with $f=1$ and $h=1$. Results are similar for the other strategies.

[22] See, for example, Gutierrez and Kelley (2008), who employ a similar methodology.

Table 5
Double sorts.

This table shows annualized mean excess returns for double-sorted portfolios. All currencies in the sample are first sorted on lagged forward discounts (*FD*) into two portfolios along the median. Next, currencies within each of the two subgroups are allocated into three momentum portfolios depending on their lagged excess returns over $f=1,6$, or 12 months. Hence, row FD_L denotes the 50% of all currencies with the lowest (lagged) forward discount whereas FD_H denotes the 50% of all currencies with the highest (lagged) forward discounts. Columns M_L, M_M, and M_H denote the 33% of all currencies with the lowest, intermediate, and the highest (lagged) returns, respectively. Columns Δ_M show the return difference between high and low momentum portfolios (M_H-M_L) for each subgroup of currencies whereas, e.g., Δ_{FD} shows the return difference between the forward discount-sorted portfolios for each momentum subgroup. The lower-right cell in each subpanel shows the return difference between the two momentum "high-minus-low" portfolios of each forward discount category. We report annualized excess returns in percent for each portfolio and all high-minus-low portfolios. Numbers in brackets are Newey and West (1987) HAC *t*-statistics and the sample runs from January 1976 to January 2010.

Carry trade and momentum												
	$f=1$, $h=1$				$f=6$, $h=1$				$f=12$, $h=1$			
	M_L	M_M	M_H	Δ_M	M_L	M_M	M_H	Δ_M	M_L	M_M	M_H	Δ_M
FD_L	−4.52	−0.90	0.54	5.06	−4.40	−0.35	0.06	4.46	−3.94	−0.40	0.09	4.04
	[−2.90]	[−0.55]	[0.34]	[3.81]	[−2.81]	[−0.21]	[0.04]	[3.63]	[−2.34]	[−0.24]	[0.06]	[2.86]
FD_H	0.64	3.20	6.00	5.36	2.38	2.43	6.34	3.96	2.86	3.21	5.98	3.12
	[0.34]	[1.68]	[3.18]	[3.30]	[1.14]	[1.45]	[3.29]	[2.43]	[1.49]	[1.80]	[3.10]	[2.02]
Δ_{FD}	5.16	4.10	5.45	0.30	6.77	2.78	6.27	−0.50	6.80	3.61	5.89	−0.91
	[4.00]	[3.43]	[3.89]	[0.17]	[4.33]	[2.57]	[4.58]	[−0.26]	[4.71]	[3.22]	[4.56]	[−0.49]

Table 6
Cross-sectional regressions.

This table shows results for cross-sectional regressions of individual currencies' excess returns (left part) or spot rate changes (right part) on lagged excess returns, lagged forward discounts, and/or lagged spot rate changes. Numbers in parentheses are standard errors of the cross-sectional R^2's. For ease of interpretation, we have multiplied spot rate changes by minus one so that higher values indicate an appreciation of the foreign currency against the USD. The sample runs from January 1976 to January 2010.

Dependent: Excess returns					Dependent: Spot rate changes				
Const.	rx	$f-s$	Δs	R^2	Const.	rx	$f-s$	Δs	R^2
Panel A: One month									
−0.02	0.16			0.15	−0.16	0.08			0.13
[−0.17]	[5.65]			(0.01)	[−1.52]	[2.95]			(0.01)
0.00		0.63		0.14	0.00		−0.37		0.09
[0.01]		[4.87]		(0.01)	[0.01]		[−2.89]		(0.01)
0.02			0.13	0.13	−0.16			0.13	0.14
[0.22]			[4.46]	(0.01)	[−1.59]			[4.55]	(0.01)
−0.07	0.12	0.57		0.26	−0.07	0.12	−0.43		0.20
[−0.76]	[4.42]	[4.68]		(0.01)	[−0.76]	[4.42]	[−3.52]		(0.01)
−0.07		0.68	0.14	0.26	−0.07		−0.32	0.14	0.21
[−0.72]		[5.89]	[4.82]	(0.01)	[−0.72]		[−2.83]	[4.82]	(0.01)
Panel B: Six months									
0.06	0.30			0.17	−0.05	0.15			0.15
[0.57]	[5.65]			(0.01)	[−0.46]	[3.07]			(0.01)
0.04		0.46		0.13	0.04		−0.52		0.09
[0.33]		[2.98]		(0.01)	[0.31]		[−3.33]		(0.01)
0.12			0.19	0.14	−0.03			0.25	0.15
[1.20]			[3.24]	(0.01)	[−0.30]			[4.87]	(0.01)
0.08	0.21	0.36		0.27	0.07	0.23	−0.64		0.24
[0.82]	[3.89]	[2.36]		(0.02)	[0.82]	[4.39]	[−4.20]		(0.01)
0.06		0.57	0.23	0.27	0.07		−0.41	0.23	0.24
[0.71]		[4.01]	[4.27]	(0.02)	[0.77]		[−2.90]	[4.33]	(0.01)
Panel C: 12 Months									
−0.05	0.28			0.16	−0.17	0.12			0.15
[−0.52]	[3.97]			(0.01)	[−1.66]	[1.79]			(0.01)
0.04		0.42		0.12	0.03		−0.51		0.09
[0.36]		[2.66]		(0.01)	[0.29]		[−3.22]		(0.01)
0.03			0.20	0.14	−0.05			0.32	0.14
[0.24]			[2.45]	(0.01)	[−0.47]			[4.52]	(0.01)
−0.06	0.20	0.28		0.25	−0.06	0.25	−0.66		0.24
[−0.66]	[2.58]	[1.74]		(0.01)	[−0.62]	[3.21]	[−4.06]		(0.01)
−0.04		0.48	0.24	0.25	−0.04		−0.42	0.27	0.24
[−0.47]		[3.21]	[3.14]	(0.01)	[−0.42]		[−2.70]	[3.41]	(0.01)

require individual currencies' excess returns to vary cross-sectionally in a way that is predictable by lagged returns. Cross-sectional regressions allow us to test for this effect while simultaneously controlling for interest rate differentials and, hence, complement the double sorts above, which work on a portfolio level, and do not necessarily control for both factors at the same time due to sequential sorting.

Panel A shows results for regressions where we use lagged excess returns, forward discounts, and/or spot rate changes over the last month as explanatory variables, whereas Panels B and C show results for values of l equal to six and 12 months, respectively.[23]

Turning to results for excess returns first (left part of Table 6), we find that lagged returns, lagged forward discounts, as well as lagged spot rate changes are cross-sectionally positively related to subsequent currency returns even when including them in joint specifications. Hence, momentum effects are robust to controlling for forward discounts (interest rate differentials). Furthermore, it is noteworthy that lagged spot rate changes do about as well as lagged excess returns in the cross-sectional regressions so that momentum seems to originate from spot rate changes and not from lagged interest rate differentials, which corroborates our finding that carry trades and momentum are different.

The right part of Table 6 shows the same calculations but with spot rate changes as dependent variables. While the effect of lagged returns or spot rate changes is very similar to our results described above, we find that the forward discount has a *negative* impact on future spot rate changes. However, the coefficients based on univariate regressions are always smaller than one in absolute value. Hence, a one percent higher interest rate in a foreign country is only followed by a depreciation smaller than one percent relative to other currencies' excess returns against the USD, consistent with the existence of a forward bias (Fama, 1984). Note that these are cross-sectional regressions so that results do not necessarily translate into a time-series setting in which the forward premium puzzle has typically been studied.

4.5. Post-formation momentum returns

Jegadeesh and Titman (2001) suggest that momentum returns are driven by slow information diffusion that leads to underreaction and persistence in returns (see also Chui, Titman, and Wei, 2010). This initial underreaction can furthermore be accompanied by subsequent overreaction that magnifies the drift in returns but has to be corrected over the long run. To investigate these issues, Jegadeesh and Titman (2001) study the post-formation holding period returns of momentum strategies over longer time spans (i.e., the returns over long horizons after portfolio formation where the portfolio composition is held constant). They find a (roughly) "inverted U-shaped pattern", i.e., returns tend to increase for several months up to one year after portfolio formation but then peak and start to decrease significantly. Jegadeesh and Titman interpret this pattern as evidence of initial underreaction that drives prices and subsequent overreaction to the series of high returns, pushing prices up above the fundamental value of the asset. This overreaction is then corrected over longer periods, leading to the observed predictable pattern of increasing and decreasing returns after portfolio formation.[24]

As a first check of this hypothesis for currency markets, we plot cumulative post-formation excess returns over periods of 1,2,…,60 months for the zero-cost long-short momentum portfolios with a one, six, and 12 months formation periods (i.e., MOM_1, MOM_6, and MOM_{12}) in Fig. 4. Returns in the post-formation period are overlapping since we form new portfolios each month but track these portfolios for 60 months. There is a clear pattern of increasing returns that peaks after 8–12 months across strategies and a subsequent period of declining excess returns. The decline is more pronounced for momentum strategies with longer formation periods. Thus, on the face of it, this evidence looks very similar to the pattern identified in equity markets as in Jegadeesh and Titman (2001). This result is interesting since it suggests that currency and equity market momentum could have similar origins.[25]

In sum, these results on currency momentum are consistent with those on stock market momentum, where momentum returns could be (at least partly) driven by slow information processing and investor overreaction. However, given the highly liquid FX market, which is dominated by professional traders and investors, it is hard to believe that investor irrationalities of this kind are not quickly arbitraged away. Thus, it is worthwhile to examine possible limits to arbitrage activity that could explain the persistence of momentum profits in FX markets. This is addressed in the next section.

5. Understanding the results

5.1. Transaction costs

What role do transaction costs play for momentum returns? To address this question, we first report momentum returns after transaction costs in Table 7, Panel A,

[23] For ease of interpretation, we multiply spot rate changes by minus one, so that higher values mean that the foreign currency is appreciating against the USD.

[24] There is relatively little work on behavioral effects in currency markets (compared to equity markets). Burnside, Han, Hirshleifer, and Wang (2011) recently show, however, that concepts from behavioral finance can be useful to understand FX phenomena as well. In addition, Bacchetta and van Wincoop (2010) argue that many FX portfolios are still not actively managed but that portfolio decisions are often taken infrequently, which can be fully rational due to the costs of portfolio adjustments. This mechanism could also account for slow diffusion of information into prices in FX markets. Investors' infrequent portfolio adjustment decisions, slow-moving capital deployed to exploit arbitrage opportunities, and the implications of these aspects for the dynamics of asset price movements are also demonstrated recently in Duffie (2010).

[25] We also provide the same results for post-formation drift in cumulative spot rate changes in Fig. A1 in the Internet Appendix and find a very similar pattern (although with a somewhat lower magnitude with respect to the initial price increases) so that the result discussed above does not seem to be driven by interest rate differentials but also stems from price changes.

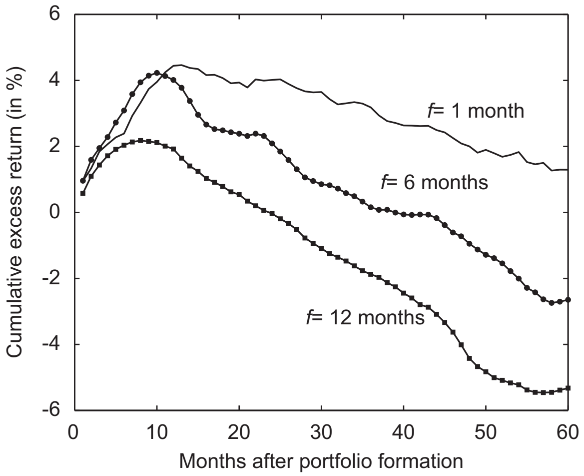


Fig. 4. Long-horizon momentum excess returns. This figure shows cumulative average excess returns to three different long-short currency momentum portfolios after portfolio formation. Momentum portfolios differ in their formation period ($f=1,6,12$ months) and post-formation returns are shown for 1,2,…,60 months following the formation period (i.e., we build new portfolios each months but track these portfolios for the first 60 months after their formation so that we are effectively using overlapping horizons). Excess returns are monthly and the sample period runs from January 1976 to January 2010.

where we impose the full quoted bid–ask spread. This spread is known to be too large relative to actual effective spreads Lyons (2001). Hence, these results are likely to underestimate momentum returns (or equivalently to provide a lower bound on profitability), whereas neglecting spreads clearly overstates momentum returns.

The results show that transaction costs could be an important factor for understanding momentum returns in currency markets (Burnside, Eichenbaum, Kleshchelski, and Rebelo, 2006; Burnside, Eichenbaum, and Rebelo, 2007). When applying the full spread, returns for the best strategy (with f, $h=1$) drop from nearly 10% to about 4% p.a. and they wipe out most of the profit of many other strategies. Interestingly, the effects of transaction costs on the average spot rate changes of portfolios, which are adjusted for bid–ask spreads in an analogous fashion to excess returns, are relatively less affected. To make the full effect of transaction costs more transparent, we also plot cumulative net excess returns (after transaction costs) for the three baseline strategies $MOM_{1,1}$, $MOM_{6,1}$, and $MOM_{12,1}$ in Fig. A2 in the Internet Appendix. Again, shaded areas correspond to NBER recessions. It can be seen that FX momentum strategies are much more profitable (after transaction costs) in the later part of the sample, but momentum strategies do not always deliver

Table 7
Momentum returns after transaction costs.

This table shows annualized average returns for different momentum strategies ($\overline{r^{f,h}}$) after adjusting for bid–ask spreads. Panel A shows results for net excess returns (left part) and net spot rate changes (right part) when deducting the full quoted spread. Numbers in brackets are t-statistics based on Newey and West (1987) standard errors. Panel B shows results only for net excess returns and for the case that effective spreads equal 75% (left part) or 50% (right part) of the quoted spread. The sample period is January 1976–January 2010 and we employ monthly returns.

Panel A: Quoted spreads

	Net excess returns						Net spot rate changes				
	Holding period h						Holding period h				
f	1	3	6	9	12	f	1	3	6	9	12
1	3.92	2.02	1.26	0.38	0.39	1	4.84	3.36	2.69	4.43	2.53
	[2.20]	[1.16]	[0.61]	[0.18]	[0.20]		[2.81]	[2.37]	[1.57]	[2.76]	[1.65]
3	4.41	2.12	0.88	0.97	−0.07	3	6.80	4.58	4.72	4.33	4.86
	[2.39]	[1.20]	[0.53]	[0.58]	[-0.04]		[3.99]	[2.81]	[3.18]	[2.58]	[3.32]
6	3.86	2.12	−0.27	−0.92	−1.28	6	5.06	4.83	3.06	3.27	3.29
	[2.09]	[1.19]	[−0.15]	[−0.49]	[−0.67]		[3.03]	[3.37]	[1.94]	[2.08]	[1.88]
9	2.48	2.43	0.99	−0.40	−1.06	9	7.53	6.73	6.19	4.81	3.84
	[1.26]	[1.27]	[0.51]	[−0.21]	[−0.54]		[4.34]	[4.00]	[3.69]	[2.88]	[2.20]
12	1.40	0.80	−1.46	−1.98	−2.44	12	6.65	5.53	3.75	2.92	2.77
	[0.74]	[0.45]	[−0.84]	[−1.11]	[−1.31]		[4.01]	[3.66]	[2.47]	[1.79]	[1.73]

Panel B: Effective spreads and net excess returns

	Effective spread of 75%						Effective spread of 50%				
	Holding period h						Holding period h				
f	1	3	6	9	12	f	1	3	6	9	12
1	5.28	3.24	2.51	1.53	1.69	1	6.64	4.47	3.77	2.69	3.00
	[2.98]	[1.89]	[1.25]	[0.76]	[0.88]		[3.76]	[2.62]	[1.89]	[1.36]	[1.61]
3	5.61	3.16	1.86	1.85	0.97	3	6.81	4.20	2.83	2.74	2.00
	[3.07]	[1.82]	[1.12]	[1.12]	[0.59]		[3.76]	[2.45]	[1.72]	[1.68]	[1.23]
6	5.03	3.17	0.70	0.15	−0.18	6	6.20	4.23	1.68	1.21	0.92
	[2.76]	[1.80]	[0.39]	[0.08]	[−0.10]		[3.43]	[2.41]	[0.94]	[0.66]	[0.49]
9	3.66	3.56	2.16	0.68	0.08	9	4.85	4.69	3.33	1.75	1.24
	[1.89]	[1.89]	[1.13]	[0.35]	[0.04]		[2.53]	[2.52]	[1.76]	[0.93]	[0.64]
12	2.60	1.97	−0.35	−0.94	−1.36	12	3.80	3.13	0.78	0.09	−0.28
	[1.39]	[1.12]	[−0.20]	[−0.53]	[−0.74]		[2.07]	[1.81]	[0.45]	[0.05]	[−0.15]

high returns to investors. Instead, there is much variation in profitability.

Next, given that the quoted spread is known to be too high relative to effective spreads, we follow Goyal and Saretto (2009) and report results for momentum excess returns after transaction costs of 75% (Panel B, left part) and 50% (Panel B, right part) of quoted spreads in Table 7. Results for these more realistic bid–ask spread adjustments indicate that transaction costs clearly matter but that they are not the sole driver of FX momentum returns as we find that many strategies still yield economically high and statistically significant returns on average.

Further scrutinizing this issue, we can break up the importance of transaction costs into turnover across portfolios and bid–ask spreads across portfolios. We provide results on both issues in the Internet Appendix (Table A.5). Two main conclusions emerge from this exercise. First, turnover can be extremely high, reaching values of more than 70% per month for the strategy with a one-month formation and holding period. Second, the winner and loser currencies do have higher transaction costs than the average exchange rate and the markup ranges from about 2.5 to 7 basis points per month. Accordingly, trading in the winner and loser currencies (as is necessary to set up a momentum strategy) is more costly than trading in the average currency pair. Hence, transaction costs clearly matter to a considerable extent.

However, given that transaction costs should be expected to decline over time due to more efficient trading technologies (such as electronic trading networks operated by, e.g., Electronic Broking Services (EBS) and Reuters), it seems unclear whether transaction costs are able to fully explain momentum returns. Fig. A3 in the Internet Appendix shows average bid–ask spreads across currencies for each month in our sample and separately for all countries and for the subsample of 15 developed countries as defined above. While there is a lot of time-series variation in average spreads, it is the case that spreads have trended downwards over our sample period. This downward trend is most clearly seen for the sample of developed countries for which we have almost complete data histories and for which average spreads are not driven by the frequent inclusion of emerging market currencies that induce some large spikes in average spreads when looking at the sample of all countries. Overall, the downward trend in bid–ask spreads seems to suggest that new technology has swamped the positive effect of volatility on bid–ask spreads. Thus, it is interesting to also investigate momentum strategies over a later part of our sample where bid-ask spreads tend to be lower on average since lower transaction costs could either imply (i) higher momentum returns due to lower trading costs or (ii) lower momentum returns since lower trading costs facilitate more capital being deployed for arbitrage activity.

Internet Appendix Table A12 shows results for the same calculations underlying Table 1 but we only include the period January 1992 to January 2010 to learn about whether the profitability of momentum strategies increases or declines over this recent period of low transaction costs. We find that unadjusted momentum returns reach levels similar to those for the full sample (Panel A) but that transaction cost-adjusted net excess returns (Panel B) are clearly higher and, for example, reach average annualized values of more than 7% for the one-month strategy $MOM_{1,1}$. Thus, lower bid–ask spreads do not necessarily lead to lower (unadjusted) excess returns, which further indicates that transaction costs are not the sole driving force behind momentum effects. This evidence also indicates that momentum returns are a phenomenon that is still exploitable nowadays.

5.2. Momentum returns and business cycle risk

Table 8, Panel A, shows results from univariate time-series regressions of momentum returns on various risk factors or business cycle state variables. See, e.g., Burnside, Eichenbaum, Kleshchelski, and Rebelo (2011) or Sarno, Schneider, and Wagner (2012) for similar regressions in the context of currency returns. These factors include macrovariables or other risk factors from the earlier literature: "Consumption" stands for real growth in non-durables and services consumption expenditures, "Employment" denotes U.S. total nonfarm employment growth, "ISM" denotes the ISM manufacturing index, "IP" denotes growth in real industrial production, "CPI" denotes the inflation rate, "M2" is the growth in real money balances, "Disp inc" is growth in real disposable personal income, "TED" denotes the TED spread (the difference between 3-month interbank rate, Libor and 3-month T-bill rate), "Term" denotes the term spread (20-year maturity minus 3-month T-bill rate), HML_{FX} is the return to the carry trade long-short portfolio (Lustig, Roussanov, and Verdelhan, 2011), and VOL_{FX} is a proxy for global FX volatility (Menkhoff, Sarno, Schmeling, and Schrimpf, 2012). We note that the alphas in these regressions cannot be interpreted as a measure of risk-adjusted returns for most specifications since we are mainly employing macrovariables or other nonreturn-based factors here. Statistical significance at the 5% level or below is indicated by bold numbers. However, looking across momentum strategies and macro-finance risk factors, there is little evidence that exposure to these factors is able to account for momentum returns. The adjusted R^2's are generally tiny and most slope coefficients are insignificantly different from zero.[26]

Panel B of Table 8 shows a multivariate regression of momentum returns on the three Fama-French factors augmented by the U.S. stock momentum factor (UMD), and it can again be seen that there is basically no explanatory power. Moreover, the alphas in these regressions, which are annualized and in percentages, can be

[26] As mentioned earlier, one exception is the momentum strategy with a 12-month formation period and global FX volatility. We find a highly significant slope coefficient here and a positive R^2. Menkhoff, Sarno, Schmeling, and Schrimpf (2012) show for this momentum strategy that innovations to global FX volatility do indeed capture a large amount of the cross-sectional spread in returns and that volatility risk is significantly priced. However, we do not find that FX volatility risk helps much for understanding momentum returns of the strategies with short formation periods of one month or six months.

Table 8
Macro risk.
This table shows time-series regression estimates of currency momentum returns (long–short portfolios $MOM_{1,1}$,$MOM_{6,1}$, and $MOM_{12,1}$) on various macrofactors and other risk factors. Consumption is real consumption growth, Employment denotes U.S. total nonfarm employment growth, ISM denotes the ISM manufacturing index, IP denotes growth in real industrial production, CPI denotes the inflation rate, M2 is the growth in real money balances, Disp inc is growth in real disposable personal income, TED denotes the TED spread, Term denotes the term spread (20 years minus 3 months), HML_{FX} is the return to the carry trade long–short portfolio (Lustig, Roussanov, Verdelhan, 2011), and VOL_{FX} is a proxy for global FX volatility (Menkhoff, Sarno, Schmeling, and Schrimpf, 2012). MKTRF, HML, and SMB are the Fama-French factors and UMD denotes the return to a long–short U.S. momentum portfolio. Panel A shows results for univariate regressions (intercepts α, slope coefficients β, and the adjusted R^2) whereas Panel B shows results from a multivariate regression of momentum returns on the three Fama-French factors and UMD. Bold numbers indicate significance at the 5%-level or below.

Panel A: Univariate regressions

	$MOM_{1,1}$			$MOM_{6,1}$			$MOM_{12,1}$		
	α	β	R^2	α	β	R^2	α	β	R^2
Consumption	**9.65**	−0.05	0.00	**8.95**	−0.12	0.00	**6.03**	0.07	0.00
Employment	**10.57**	−0.72	0.00	**7.74**	0.62	0.00	**5.86**	0.23	0.00
ISM	**9.46**	0.04	0.00	**8.60**	0.03	0.00	**6.14**	0.04	0.00
IP	**9.72**	0.11	0.00	**8.72**	0.04	0.00	**6.26**	0.03	0.00
CPI	**11.73**	−0.55	0.00	**9.11**	−0.12	0.00	**6.60**	−0.10	0.00
M2	**9.97**	0.34	0.00	**8.68**	0.02	0.00	**6.18**	−0.01	0.00
Disp inc	**9.33**	0.07	0.00	**8.42**	0.10	0.00	**5.95**	0.10	0.00
TED	**13.64**	−**0.38**	0.01	**11.95**	−0.30	0.01	**9.73**	−0.32	0.01
Term	4.48	**0.22**	0.01	**7.54**	0.05	0.00	5.05	0.05	0.00
HML_{FX}	**9.50**	0.04	0.00	**8.65**	0.02	0.00	**6.21**	0.08	0.00
VOL_{FX}	**11.70**	−0.44	0.00	**18.75**	−2.04	0.01	**27.59**	−**4.29**	0.04

Panel B: Multivariate regressions

	$MOM_{1,1}$			$MOM_{6,1}$			$MOM_{12,1}$		
	α	β	R^2	α	β	R^2	α	β	R^2
MKTRF	**8.73**	0.00	0.00	**8.02**	0.04	0.00	**5.16**	0.02	0.00
SMB		0.97			−0.54			0.71	
HML		0.06			0.01			0.06	
UMD		0.02			0.03			0.04	

interpreted as the risk-adjusted performance of momentum returns since the factors are excess returns in this case. Across strategies, the alphas are fairly high, as judged by this particular model for returns. Based on earlier research for the U.S. stock market, this result does not come as a surprise regarding the three Fama-French factors but it seems noteworthy that currency momentum is also unrelated to the UMD factor.[27]

In sum, there is little evidence that standard business cycle variables or portfolio-based risk factors help to understand momentum returns, i.e., it seems that the latter are largely disconnected from U.S. business cycle risk. This finding squares well with earlier results for U.S. equity momentum, which is hard to explain by relying on its covariance with macrorisk factors (e.g., Griffin and Martin, 2003; Cooper, Gutierrez, and Hameed, 2004).

5.3. Limits to arbitrage: time variation in momentum profitability

Next, we are interested in the stability of momentum returns over time. Since FX market participants (e.g., proprietary trading desks, asset managers, and hedge funds) generally have short investment horizons, time variation in momentum profits could also represent an important obstacle for taking arbitrage positions in FX markets.

Fig. 5 plots average excess returns to the three long-short momentum portfolios $MOM_{1,1}$, $MOM_{6,1}$, and $MOM_{12,1}$ over rolling windows of 36 months. The left part shows unadjusted returns while the right part of the figure shows net excess returns after transaction costs. It can be seen that the profitability of momentum strategies is time-varying and that both adjusted and unadjusted returns appear to be higher over the second part of the sample. In fact, momentum returns for all three strategies have been rather high between 2000 and 2005 reaching levels of monthly net excess returns of about 2% per month.

Most importantly, this figure also illustrates that momentum returns are far from being constant even over intermediate time intervals of several years. Hence, an investor seeking to profit from momentum returns has to have a long enough investment horizon. This result seems important, since the bulk of currency speculation is accounted for by professional market participants and proprietary traders who have a rather short horizon over which their performance is evaluated (Lyons, 2001). Hence, momentum strategies are potentially risky for myopic market participants, so that large time variation in the performance of

[27] We have also experimented with more elaborate cross-sectional asset pricing tests for both macrofactors and return-based factors but, as could be expected on the basis of the time-series results reported in Table 8, did not find any improvement in results.

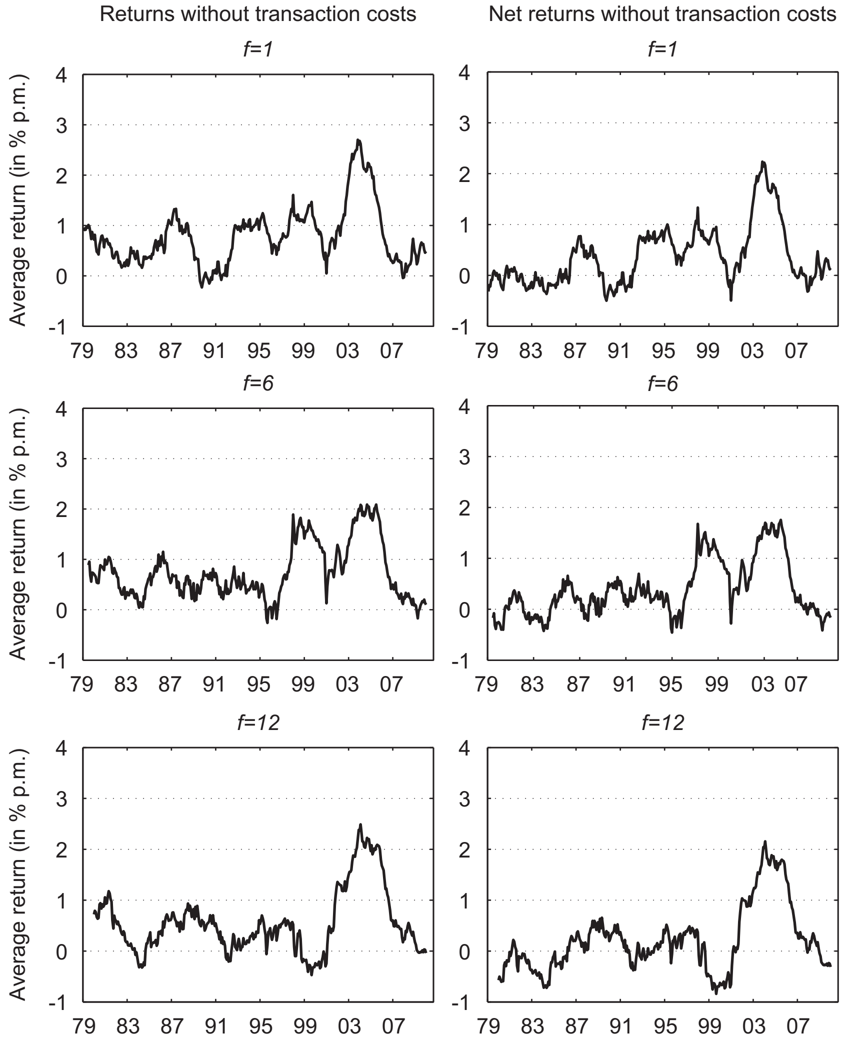


Fig. 5. Rolling average returns for three momentum strategies. This figure shows average excess returns per month (p.m.) over rolling windows of 36 months for three long-short momentum strategies: $MOM_{1,1}$, $MOM_{6,1}$, and $MOM_{12,1}$, where $MOM_{j,h}$ denotes the return difference between a portfolio long in currencies with the highest lagged excess returns (measured over the last f months) and a portfolio short in currencies with the lowest excess return over the last f months. Portfolios are held for $h=1$ month and we use excess returns without transaction costs (left part of the table) and net excess returns adjusted for transaction costs (right part). The sample runs from January 1976 to January 2010.

momentum returns could impede arbitrage activity by some of the key FX market players.[28]

5.4. Limits to arbitrage: idiosyncratic volatility

Unlike in stock markets, there are no natural short-selling constraints in FX markets. However, to conduct arbitrage in currency markets, an investor obviously has to set up positions that he may wish to hedge such that the position becomes a pure bet on return continuation but not on any sort of systematic risk. Hence, we investigate whether momentum returns are different between currencies with high or low idiosyncratic volatility (relative to an FX asset pricing model). Finding that currency momentum is stronger among high idiosyncratic volatility currencies would imply that attempts to arbitrage these momentum returns away could be risky since it will be hard to find a second pair of currencies that can be used as a hedge factor unrelated to simple return continuation.

To this end, Panel A of Table 9 shows results from double-sorting currencies first into two portfolios depending on whether a currency has a lagged idiosyncratic volatility above or below the median (of all available currencies), and then into three portfolios depending on their lagged

[28] The role of frictions (e.g., margin and capital constraints) on the deployment of arbitrage capital to investment opportunities by institutional investors is stressed, for instance, in recent work by Mitchell, Pedersen, and Pulvino (2007). Excellent recent surveys on limits to arbitrage and slow-moving capital that provide an obstacle to the corrective actions of rational arbitrageurs are provided by Duffie (2010) and Gromb and Vayanos (2010).

Table 9
Double sorts on idiosyncratic volatility or risk ratings and momentum.

The setup of this table is identical to Table 5 but here we sort on idiosyncratic volatility and momentum (Panel A), country risk and momentum (Panel B), and exchange rate stability risk and momentum (Panel C). We report annualized excess returns in percent for each portfolio and all high-minus-low portfolios. Numbers in brackets are Newey and West (1987) HAC *t*-statistics and the sample runs from January 1976 to January 2010.

Panel A: Idiosyncratic volatility and momentum

	$f=1,h=1$				$f=6,h=1$				$f=12,h=1$			
	M_L	M_M	M_H	Δ_M	M_L	M_M	M_H	Δ_M	M_L	M_M	M_H	Δ_M
$IVOL_L$	−1.04	0.92	2.93	3.97	−0.85	1.08	2.82	3.67	0.15	1.13	2.27	2.12
	[−0.65]	[0.55]	[1.75]	[2.81]	[−0.50]	[0.66]	[1.79]	[3.04]	[0.10]	[0.67]	[1.31]	[1.58]
$IVOL_H$	−3.52	1.00	4.57	8.09	−2.22	0.24	4.77	6.99	−0.78	0.20	4.38	5.16
	[−1.83]	[0.57]	[2.48]	[4.72]	[−1.16]	[0.14]	[2.44]	[4.28]	[−0.41]	[0.11]	[2.30]	[3.01]
Δ_{IVOL}	−2.48	0.07	1.64	4.11	−1.38	−0.84	1.95	3.33	−0.93	−0.94	2.11	3.04
	[−1.86]	[0.07]	[1.28]	[2.18]	[−1.15]	[−0.86]	[1.52]	[2.05]	[−0.80]	[−0.89]	[1.63]	[1.79]

Panel B: Country risk and momentum

	M_L	M_M	M_H	Δ_M	M_L	M_M	M_H	Δ_M	M_L	M_M	M_H	Δ_M
$CRISK_L$	0.01	3.41	4.51	4.50	0.95	3.14	4.26	3.31	1.65	3.24	3.86	2.21
	[0.01]	[1.78]	[2.52]	[3.12]	[0.49]	[1.67]	[2.33]	[2.67]	[0.80]	[1.67]	[2.10]	[1.51]
$CRISK_H$	−0.67	3.82	8.04	8.72	0.89	3.39	7.24	6.35	2.58	2.70	8.65	6.07
	[−0.34]	[1.90]	[3.72]	[4.19]	[0.40]	[1.94]	[3.24]	[2.92]	[1.29]	[1.43]	[3.56]	[2.34]
Δ_{CRISK}	−0.68	0.41	3.53	4.22	−0.06	0.25	2.97	3.04	0.93	−0.54	3.79	3.87
	[−0.46]	[0.35]	[2.21]	[2.12]	[−0.04]	[0.20]	[2.02]	[1.72]	[0.54]	[−0.45]	[2.61]	[1.93]

Panel C: Exchange rate stability risk and momentum

	M_L	M_M	M_H	Δ_M	M_L	M_M	M_H	Δ_M	M_L	M_M	M_H	Δ_M
$XSTAB_L$	1.27	0.15	3.25	1.98	1.56	0.30	3.60	2.04	0.80	1.40	3.22	2.42
	[0.83]	[0.10]	[2.17]	[1.39]	[0.96]	[0.23]	[2.32]	[1.31]	[0.50]	[1.04]	[2.15]	[1.70]
$XSTAB_H$	−0.48	4.04	6.09	6.56	0.51	3.35	6.06	5.55	1.58	3.38	6.36	4.78
	[−0.24]	[2.02]	[3.09]	[4.06]	[0.24]	[1.77]	[2.93]	[3.31]	[0.77]	[1.80]	[3.12]	[2.50]
Δ_{XSTAB}	−1.75	3.89	2.84	4.59	−1.05	3.05	2.47	3.51	0.78	1.98	3.14	2.35
	[−1.06]	[2.47]	[1.58]	[2.44]	[−0.59]	[2.01]	[1.31]	[1.70]	[0.43]	[1.21]	[1.82]	[1.11]

excess return.[29] For all three formation periods we study (i.e., f is either 1,6, or 12), we find that momentum returns are higher among currencies with high idiosyncratic volatility than among currencies with low idiosyncratic volatility (*IVOL*). The returns differences are quite large in economic terms. For example, sorting on lagged idiosyncratic volatility and lagged one-month returns leads to an annualized momentum excess return of 3.97% among currencies with low *IVOL*, whereas a momentum strategy among currencies with high *IVOL* yields an average excess return of 8.09% p.a. Thus, momentum strategies are much more profitable among currencies with high idiosyncratic risk.

5.5. Limits to arbitrage: country risk

A natural limit to arbitrage in foreign exchange markets is country risk. Institutional constraints such as country limits, for instance, can prevent position-taking in currencies of high risk countries. Arbitrage activity involving these countries' currencies also exposes investors to the risk of potential sudden capital account restrictions and sharp exchange rate moves. This implies that arbitrage strategies involving these countries' currencies are much more risky compared to those involving currencies of well-developed and highly stable countries with low risk ratings.

We now perform the same analysis as above but sort instead on a measure of country risk (*CRISK*) and a measure of exchange rate stability risk (*XSTAB*). These data are based on the International Country Risk Guide (ICRG) database from the Political Risk Services (PRS) group.[30] We employ the composite country risk rating, which comprises economic, political, and financial risk of a country, as a general proxy for the riskiness of a given country and exchange rate stability risk as a specific proxy for the risk of sharp currency movements.[31] Data for these risk proxies start in January 1985 and we employ the log

[29] Idiosyncratic volatility for each currency j in month t is computed from a regression of currency returns on a constant, the Dollar risk factor, and the HML_{FX} factor of Lustig, Roussanov, and Verdelhan (2011). Idiosyncratic volatility is then computed as the absolute value of the regression residual in month t. We find similar results to those reported below when we employ the volatility risk factor proposed by Menkhoff, Sarno, Schmeling, and Schrimpf (2012).

[30] These data are quite common as proxies for country risk; see, e.g., Bekaert, Harvey, Lundblad, and Siegel (2007), who also use risk indicators from this database.

[31] The exchange rate stability risk proxy measures the perceived risk of large exchange rate movements in the near future.

deviation of the risk rating of a country from the rating of the U.S. as a proxy of relative risk for a U.S. investor.

The setup here is somewhat akin to Avramov, Chordia, Jostovy, and Phlipov (2007, in press), who show that U.S. stock momentum is mainly concentrated in high credit risk firms that are illiquid and hard to sell short.[32] Hence, credit risk proxies for hurdles to arbitrage activity. In our context, we focus on country risk as a natural proxy for limits to arbitrage in FX markets. High risk countries are more politically unstable, economically less developed and more volatile so that establishing positions in the associated currencies poses nontrivial threats of sudden capital account restrictions and nonconvertibility of currency. In short, arbitrage activity involving these countries' currencies should be clearly more risky compared to well-developed and highly stable countries with low risk ratings similar to the U.S.

Panels B and C of Table 9 show results for double sorts on either country risk or exchange rate stability risk and momentum. Corroborating our earlier findings for idiosyncratic volatility, we find that momentum returns are significantly positive and always larger in high-risk countries than in low-risk countries, where momentum strategies do not yield significant excess returns. Hence, for an investor to profit from currency momentum strategies, it is necessary to operate in markets for currencies of risky countries. This is especially important since, unlike momentum strategies in domestic U.S. stocks, investments in foreign currency are always subject to risks of capital controls and nonconvertibility. Therefore, country risk should be an important limit to arbitrage activity in FX markets.

Finally, we examine whether our findings above are driven by country risk being related linearly to the cross-sectional spread in momentum returns and whether momentum is differently affected than carry trades. Table A16, which, as an example, is based on the strategy with a one month formation and holding period, in the Internet Appendix shows a clear pattern. Country risk and exchange rate stability risk are high for both winner *and* loser currencies (Panel A) in the momentum strategy. Hence, it is not the case that these risk ratings are simple proxies for interest rate differentials that drive our results. Instead, currency momentum strategies require that an investor has to go long and short in the most risky countries. This is especially true since momentum profits stem from both the long and short side of the position (see Table A.4, Panel A) so that it is necessary to set up both positions. Contrary to this, the cross-section of forward discount-sorted portfolios that form the basis of the carry trade (Table A16, Panel B) shows a very different pattern. Country risk is highest for carry trade target currencies (high interest rate currencies) and lowest for carry trade funding currencies (low interest rate currencies). This squares well with the finding that most of the carry trade return comes from the long position of the strategy (A.4, Panel B). In any case, these results indicate that country risk has a nonlinear impact on the cross-sectional spread in momentum portfolios' returns and, again, that the anatomy of carry trade strategies is very different from currency momentum.

Developed countries: Finally, a shortcut to looking at country risk could also be to define a sample of clearly developed countries that have stable exchange rate regimes and are most liquid. Table A13 in the Internet Appendix shows results before and after transaction costs similar to those in Table 1 but we limit the cross-section to 15 developed countries.[33] It is clear from this table that momentum returns are much smaller and basically nonexistent after transaction costs when looking at currencies of developed countries. This finding is interesting since it suggests that the profitability of momentum strategies depends on whether smaller and presumably less liquid currencies are included in the investment universe or not. Again, this shows that limits to arbitrage are an important factor in explaining the persistence of momentum returns in FX markets.

6. Robustness and additional tests

6.1. Capital account restrictions and tradability

We have shown above that momentum returns are large in FX markets when examining a broad cross-section that also includes smaller currencies from emerging markets. A potential concern regarding these results is whether all currencies have actually been tradable throughout the sample period as there can be capital controls for some countries or other issues rendering trading in these currencies infeasible. Many of these smaller currencies do indeed show up in the loser and winner portfolios quite frequently which is shown in Table A.6 in the Internet Appendix. This table reports the frequency with which each currency is included in the winner and loser portfolio of the MOM(1,1) strategy. The table shows, quite expectedly, that several larger currencies (e.g., Australia, Canada, Japan, New Zealand, Switzerland, United Kingdom) are often included in the momentum strategy but this dominance rests at least partly on the longer sample periods available for these currencies. However, the table also shows large inclusion frequencies for emerging markets such as Brazil, Indonesia, Poland, or Singapore. Hence, it seems worthwhile to investigate whether issues of tradability (or convertibility) affect our results.

As a first exercise, we limit the sample to currencies that have a positive score on the capital account openness index of Chinn and Ito (2006), both in the formation and holding period, to control for the possibility that some currencies are not tradable or that they are only traded in more opaque offshore markets which would not be adequately reflected in the data. We report results for

[32] In a similar vein, Jostova, Nikolova, Philipov, and Stahel (2010) show that momentum profits in U.S. corporate bond returns derive solely from long and short positions in non-investment grade bonds.

[33] These countries are Australia, Belgium, Canada, Denmark, Euro area, France, Germany, Italy, Japan, Netherlands, New Zealand, Norway, Sweden, Switzerland, and the United Kingdom.

Table 10
Momentum returns and capital controls.

The setup of this table is identical to Table 1 but here we exclude countries with capital controls. More specifically, at each point in time, we only include currencies of countries that have a score in excess of zero (Panel A) or a score higher than one (Panel B) in the updated capital account openness index of Chinn and Ito (2006). The sample periods runs from January 1976 to January 2010.

	Excess returns						Spot rate changes				
	Holding period h						Holding period h				
f	1	3	6	9	12	f	1	3	6	9	12
Panel A. Chinn-Ito index >0											
1	9.22	6.46	6.02	4.54	4.63	1	5.14	3.07	3.75	2.84	2.90
	[4.61]	[3.47]	[2.87]	[2.42]	[2.39]		[2.79]	[1.86]	[1.86]	[1.58]	[1.57]
3	9.90	7.07	6.18	5.39	5.33	3	7.04	5.33	4.91	3.62	4.00
	[5.30]	[3.96]	[3.42]	[3.44]	[3.07]		[3.81]	[2.81]	[2.70]	[2.06]	[2.30]
6	9.65	8.11	4.71	3.42	4.22	6	6.48	7.29	3.32	2.14	1.33
	[5.29]	[4.52]	[2.52]	[1.74]	[2.26]		[3.30]	[4.11]	[1.72]	[1.14]	[0.68]
9	8.95	8.46	5.60	4.53	2.64	9	6.83	5.76	5.30	3.64	1.82
	[4.53]	[4.48]	[2.85]	[2.21]	[1.37]		[3.42]	[2.81]	[2.66]	[1.79]	[0.92]
12	6.90	6.51	3.77	2.54	1.40	12	4.61	4.20	1.73	1.42	−0.53
	[3.61]	[3.58]	[2.06]	[1.38]	[0.77]		[2.33]	[2.20]	[0.96]	[0.78]	[−0.28]
Panel B. Chinn-Ito index >1											
1	8.97	5.80	6.51	4.65	4.25	1	5.06	2.96	4.10	2.27	2.75
	[4.61]	[3.01]	[3.07]	[2.48]	[2.02]		[2.79]	[1.77]	[2.02]	[1.25]	[1.45]
3	9.85	7.12	5.98	5.53	5.34	3	7.51	5.41	4.74	4.43	3.52
	[5.10]	[3.95]	[3.28]	[3.54]	[2.87]		[3.91]	[2.85]	[2.70]	[2.51]	[2.08]
6	10.50	8.86	5.40	3.43	2.34	6	6.19	6.61	3.17	2.43	−0.39
	[5.45]	[4.76]	[2.83]	[1.55]	[1.32]		[3.02]	[3.46]	[1.58]	[1.24]	[−0.20]
9	9.22	8.16	5.51	4.87	2.46	9	6.69	5.46	5.19	2.83	1.80
	[4.47]	[3.97]	[2.66]	[2.11]	[1.18]		[3.18]	[2.55]	[2.48]	[1.35]	[0.91]
12	6.88	6.05	2.63	2.18	1.54	12	4.73	3.72	1.46	1.09	−0.20
	[3.42]	[3.09]	[1.16]	[1.21]	[0.83]		[2.34]	[1.86]	[0.77]	[0.61]	[−0.10]

this restricted subset in Table 10, Panel A. As can be seen, the results are not affected by excluding these currencies. Moreover, countries with negative capital account openness index values do not account for a large share of the relevant corner portfolios (less than 20%, on average). While a positive score in the Chinn-Ito index already excludes a number of countries (even developed countries, e.g., the U.K. from 1976 to 1978), we additionally run the same exercise under the constraint that a country has to have an index score of at least one. This requirement eliminates several currencies almost completely from the sample (e.g., Brazil, Philippines, Poland, and South Korea) and significantly reduces the investable sample period for other countries (e.g., Belgium only becomes investable in the 1990s). Results for this filter are shown in Panel B of Table 10 but also do not indicate that momentum is primarily driven by currencies that exhibit limitations to investability.

While the above analysis suggests that tradability issues do not wipe out momentum profits in FX markets, we additionally ran a small survey among four large brokers in FX markets (Goldman Sachs, Deutsche Bank, UBS, Nomura) and asked which currencies would have been impossible (or nearly impossible) to trade in a dynamic portfolio strategy that requires frequent rebalancing. Based on their answers, we restricted our set of tradable currencies and computed momentum returns on the resulting sample. The following restrictions were imposed: Czech Republic (not tradable before 1999), Hungary (2000), Indonesia (1999), Malaysia (1999), Philippines (1999), Singapore (1999), South Africa (2001), Taiwan (1999), Hong Kong (1986), and Thailand (1999).[34] Results for this limited sample are shown in Table 11, Panel A. Corroborating the evidence based on the Chinn/Ito index above, we find that momentum profits are still significant after taking into account likely restrictions on tradability of countries.

As a final check, we augmented the market practitioner's list by eliminating all currencies with large trading in non-deliverable forwards in offshore markets that might not be adequately covered by our price and interest rate data. These currencies include: Brazil, Egypt, India, Indonesia, South Korea, Malaysia, the Philippines, and Taiwan. Results for this even more restricted set of currencies are shown in Panel B of Table 11 but only strengthen our findings above.

In sum, we find that accounting for capital account restrictions (or other trading restrictions) does not significantly weaken average momentum returns despite excluding many smaller emerging markets from our sample. This finding seems to be driven by the fact that most minor currencies, which are more likely to be subject to capital controls, only enter our sample very

[34] Most of the survey respondents' other restrictions, for example, regarding Egypt or Saudi Arabia, were actually already reflected in our data where data histories of several currencies start very late at the end of the 1990s or early 2000s (see Table A1 in the Internet Appendix).

Table 11
Momentum and tradability.

This table shows average annualized excess returns for six momentum portfolios sorted on lagged one-, six-, and 12-month returns and the corresponding high-minus-low momentum portfolios (H − L). Panel A shows results for a set of investable currencies as identified in a survey of FX professionals in major investment banks. Panel B additionally excludes all currencies with non-deliverable forward (NDF) trading in offshore markets. We refer to Section 6 in the main text for details. Numbers in brackets are *t*-statistics based on Newey and West (1987) and the sample period runs from January 1976 to January 2010.

Panel A: "Investable" currency universe

f	L	2	3	4	5	H	H − L
1	−3.28	−0.17	0.67	2.58	2.47	5.48	8.76
	[−1.89]	[−0.09]	[0.38]	[1.45]	[1.50]	[2.99]	[4.90]
6	−1.85	−0.79	1.45	1.16	2.31	5.81	7.65
	[−1.05]	[−0.44]	[0.88]	[0.68]	[1.36]	[2.98]	[4.80]
12	−2.14	0.38	0.94	1.43	2.97	4.75	6.89
	[−1.15]	[0.22]	[0.55]	[0.83]	[1.69]	[2.55]	[4.05]

Panel B: "Investable" currency universe ex NDF

f	L	2	3	4	5	H	H − L
1	−2.80	0.31	0.98	1.99	2.58	5.30	8.10
	[−1.64]	[0.17]	[0.52]	[1.07]	[1.56]	[2.88]	[4.69]
6	−1.68	−0.20	1.53	1.15	2.38	5.66	7.34
	[−0.97]	[−0.11]	[0.87]	[0.65]	[1.38]	[2.79]	[4.58]
12	−1.74	0.23	1.04	1.85	2.82	4.63	6.36
	[−0.97]	[0.13]	[0.58]	[1.04]	[1.58]	[2.37]	[3.58]

recently and, thus, do not drive the lion's share of our result.

6.2. Additional tests

Different base currencies: So far, we have investigated momentum profits from the viewpoint of a U.S. investor. For robustness, we also present results for a British (GBP), Swiss (CHF), Canadian (CAD), and Swedish (SEK) investor, i.e., we convert all data such that they are quoted against one of these four alternative numeraires. The effective size of the cross-section is, of course, unchanged since we lose one currency (the numeraire) but include the USD as a "new" currency.

Results are shown in Tables A.7 and A.8 in the Internet Appendix for excess returns and spot rate changes, respectively. It can be seen that results are basically unchanged relative to the benchmark case so that momentum is not a U.S. dollar phenomenon. This result is reasonable since our momentum portfolios are dollar neutral by construction (the USD component cancels out in the long minus short portfolio). Hence, changing the numeraire has little to no effect on the profitability of momentum strategies.

Furthermore, we also run regressions of momentum excess returns for the four different base currencies on a set of risk factors to rule out the possibility that momentum returns are more closely linked to traditional risk factors for non-U.S. investors. Due to data limitations, we cannot obtain data for all risk factors considered in Table 8 so that we focus on the following set of risk factors that should suffice to capture broad economic conditions in these four countries: growth in real industrial production (IP), CPI inflation, growth in real money balances, changes in the term spread, and (local) stock market returns. Results are reported in Table A.9 in the Internet Appendix and we find (similar to the U.S. case in Table 8) that momentum returns are not closely linked to any of these standard macro-finance risk factors.

Currency regimes: Another question of relevance is whether momentum strategies can be enhanced by considering information about currency regimes. Intuitively, currencies that are pegged or are only allowed to move in very small bands (or target zones) should be less useful in setting up a momentum strategy than freely floating currencies or currencies that are allowed to move in larger bands. To address this issue, we limit our sample of currencies to (i) free floats, managed floats, pre-announced crawling bands (wider than or equal to ± 2%), de facto crawling bands (narrower than or equal to ± 5%), moving bands (narrower than or equal to ± 2%) or (ii) free floats only. Sample (i) corresponds to category 3 whereas sample (ii) corresponds to category 4 of the International Monetary Fund (IMF) coarse classification of exchange rate regimes available on Carmen Reinhart's Web page.[35]

Results for these two samples of less heavily managed currencies are shown in Table A10. The sample period starts in 1986 here to have a large enough cross-section for free floats (also see Fig. 1). Panel A reports descriptive statistics for six momentum portfolios and the long minus short portfolio for sample (i). There is a monotonically increasing spread in average excess returns and a significantly positive average excess return for the momentum strategy long in winners and short in losers regardless of the formation period. Panel B shows results for sample (ii) that only comprises free floats. Average excess returns tend to be somewhat lower for formation periods of one and six months but somewhat higher for the 12-month formation period.

In sum, there does not seem to be a clear benefit from concentrating on only freely floating currencies. While freely floating currencies have more room for large price swings, excluding less flexible exchange rates results in a smaller cross-section and excludes a number of slowly trending rates that are managed in crawling bands.

Central bank interventions: Central bank interventions have been considered as one potential source of momentum profits early in the literature. For example, Silber (1994) shows that technical trading rules are more valuable when government agencies intervene in the market. However, later papers reach different conclusions so that the relation between official intervention and momentum trading is less

[35] http://www.carmenreinhart.com/data/browse-by-topic/topics/12. IMF categories 1 and 2 correspond to more restrictive regimes. It is important to note that for the last several years, the IMF classification of each country (published in the Annual Report on Exchange Arrangements and Exchange Restrictions) is based on the country's actual (de facto) policy, as determined by the IMF. For some countries, this classification could differ from the country's official (de jure) stated policy. For most of our sample, only the official stated policy is reported by the IMF.

clear-cut. In this vein, Neely (2002) finds that interventions do not influence technical trading profits and that momentum profits are more likely to precede intervention rather than being caused by them.[36]

Given the prominence of this topic in the earlier literature, we briefly examine the relationship between intervention and momentum returns in Table A11 in the Internet Appendix. We report results for regressions of momentum excess returns for our three benchmark strategies on contemporaneous and lagged central bank intervention activity. Intervention activity is proxied for by the sum of absolute intervention amounts of all central banks in the USD (against any foreign currency). Data for this exercise are obtained from the Federal Reserve Bank of St. Louis. Our results show that interventions are not very powerful in capturing momentum returns, broadly consistent with the findings in Neely (2002). However, it should be noted that our analysis is intentionally simple and that there are serious data issues with central bank interventions which are usually not made public.

European Monetary System (*EMS*): As an additional robustness check, we calculate momentum profits where we exclude all countries participating in the EMS (except for the Deutschmark) and focus on the 1990s where currencies of these countries moved in lock-step.[37] Since momentum in any of these countries should be very short-lived, it seems likely that excluding these currencies will yield larger momentum profits. We plot cumulative momentum excess returns (for the MOM(1,1) strategy) from 1990 to 1998 in Fig. A4 in the Internet Appendix and do indeed find that excluding EMS member countries leads to a somewhat better performance. Hence, the results reported in the main text seem conservative and it should be possible to increase the profitability of momentum strategies by carefully accounting for the correlation structure of currencies.

7. Conclusion

We have empirically investigated momentum strategies in FX markets, which rely on return continuation among winner and loser currencies. We find that these strategies yield surprisingly high unconditional average excess returns of up to 10% per year and that these returns are hard to understand in a framework that relies on covariance risk with standard risk factors. In contrast to an explanation based on systematic risk, we find evidence for under- and subsequent overreaction in long-horizon momentum returns. In this sense, the evidence for currency momentum seems similar to what has been found for equity markets in the earlier literature.

We also find that momentum returns are different from more conventional technical trading rules. As technical trading mostly aims at exploiting trends or momentum in currency movements, it could be expected that returns to these strategies are positively related to our cross-sectional momentum returns. We find, however, that returns to benchmark technical trading rules are somewhat lower and that the correlation with our momentum strategies is rather small. Moreover, currency momentum strategies are very different from the popular carry trade in FX markets. Hence, it comes as no surprise that momentum is not well captured by the global factors that have been shown to be related to carry trade returns in the earlier literature. Rather, momentum and the carry trade are different phenomena that require a different explanation.

However, currency momentum returns do not come as a free lunch for investors trying to exploit these strategies. We find that momentum portfolios in the FX market are significantly skewed towards minor currencies that have relatively high transaction costs, accounting for roughly 50% of momentum returns. Also, the concentration of minor currencies in momentum portfolios raises the need to set up trading positions in currencies with higher idiosyncratic volatility, higher country risk, and higher expected risk of exchange rate instabilities, which clearly imposes risks to investors that are not captured by standard risk factors in a covariance risk framework. Hence, there seem to be effective limits to arbitrage that prevent a straightforward exploitation of momentum returns. Furthermore, momentum profits are highly time-varying, which can also pose an obstacle to arbitrage activity for some of the key FX market participants (e.g., proprietary traders and hedge funds) who typically have fairly short-term investment horizons.

Seen from a broader perspective, there is mounting evidence that momentum can be seen as an ubiquitous phenomenon in financial markets (e.g., Asness, Moskowitz, and Pedersen, 2009). A key contribution of this paper is to show that momentum strategies deliver high excess returns in FX markets, comparable in magnitude to the excess returns documented in stock markets. This occurs despite the special characteristics of currency markets, such as huge trading volume, mostly professional traders, no short-selling constraints, and a considerable degree of central bank interference. However, we show that FX momentum returns are not driven by policy measures including monetary regimes, currency intervention, or the implementation of capital account controls. Momentum returns stem primarily from currencies that are hard to hedge and have high country risk, which is similar to recent findings that equity momentum is concentrated in stocks with high credit risk (Avramov, Chordia, Jostova, and Philipov, 2007), and momentum in corporate bonds is concentrated in non-investment grade bonds (Jostova, Nikolova, Philipov, and Stahel, 2010). In sum, these findings suggest that there could be a common source of momentum profits across asset classes.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jfineco.2012.06.009.

[36] Also, see Neely (1998) for an overview of several findings in the literature on interventions and returns to technical trading. See Sarno and Taylor (2001) for a comprehensive survey on the impact of official intervention on exchange rates.

[37] Neely and Weller (1999) investigate returns to technical trading rules in EMS currencies over the period from 1986 to 1996 and find that they generate significant excess returns.

References

Akram, Q.F., Rime, D., Sarno, L., 2008. Arbitrage in the foreign exchange market: turning on the microscope. Journal of International Economics 76, 237–253.

Asness, C., Moskowitz, T., Pedersen, L., 2009. Value and momentum everywhere. Unpublished working paper. AQR Capital Management, University of Chicago, Copenhagen Business School.

Avramov, D., Chordia, T., Jostova, G., Philipov, A., 2007. Momentum and credit rating. Journal of Finance 62, 2503–2520.

Avramov, D., Chordia, T., Jostova, G., Philipov, A. Anomalies and financial distress. Journal of Financial Economics, in press.

Bacchetta, P., van Wincoop, E., 2010. Infrequent portfolio decisions: a solution to the forward discount puzzle. American Economic Review 100, 870–904.

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343.

Bekaert, G., Harvey, C., Lundblad, C., Siegel, S., 2007. Global growth opportunities and market integration. Journal of Finance 62, 1081–1137.

Brunnermeier, M., Nagel, S., Pedersen, L., 2009. Carry trades and currency crashes. In: NBER Macroeconomics Annual 2008, vol. 23, pp. 313–347.

Burnside, C., Eichenbaum, M., Kleshchelski, I., Rebelo, S., 2006. The returns to currency speculation. NBER Working Paper No. 12489.

Burnside, C., Eichenbaum, M., Kleshchelski, I., Rebelo, S., 2011. Do peso problems explain the returns to the carry trade? Review of Financial Studies 24, 853–891.

Burnside, C., Eichenbaum, M., Rebelo, S., 2007. The returns to currency speculation in emerging markets. American Economic Review, Papers and Proceedings 97, 333–338.

Burnside, C., Eichenbaum, M., Rebelo, S., 2008. Carry trade: the gains of diversification. Journal of the European Economic Association 6, 581–588.

Burnside, C., Eichenbaum, M., Rebelo, S., 2011. Carry trade and momentum in currency markets. Annual Review of Financial Economics 3, 511–535.

Burnside, C., Han, B., Hirshleifer, D., Wang, T., 2011. Investor overconfidence and the forward premium puzzle. Review of Economic Studies 78, 523–558.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Chabot, B., Ghysels, E., Jagannathan, R., 2009. Momentum cycles and limits to arbitrage: evidence from Victorian England and post-depression us stock markets. NBER Working Paper No. W15591.

Chan, K., Hameed, A., Tong, W., 2000. Profitability of momentum strategies in the international equity markets. Journal of Financial and Quantitative Analysis 35, 153–172.

Chan, L., Jegadeesh, N., Lakonishok, J., 1996. Momentum strategies. Journal of Finance 51, 1681–1713.

Chinn, M.D., Ito, H., 2006. What matters for financial development? Capital controls, institutions, and interactions. Journal of Development Economics 81, 163–192.

Chordia, T., Shivakumar, L., 2002. Momentum, business cycle, and time-varying expected returns. Journal of Finance 62, 985–1019.

Chui, A., Titman, S., Wei, K., 2010. Individualism and momentum around the World. Journal of Finance 65, 361–392.

Cooper, M., Gutierrez, R., Hameed, A., 2004. Market states and momentum. Journal of Finance 59, 1345–1365.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. Journal of Finance 53, 1839–1885.

Deutsche Bank, 2010. Exchange Rate Perspectives: 30 Years of FX Investment Returns. February 8, 2010.

Dooley, M., Shafer, J., 1976. Analysis of short-run exchange rate behavior: March 1973 to September 1975. International Finance Discussion Papers, No. 76. Federal Reserve System, Washington, DC.

Dooley, M., Shafer, J., 1983. Analysis of short-run exchange rate behavior: March 1973 to November 1981. In: Bigman, D., Taya, T. (Eds.), Exchange Rate and Trade Instabilities: Causes, Consequences, and Remedies, Ballinger Publishing, Cambridge, MA, pp. 43–70.

Duffie, D., 2010. Presidential address: asset price dynamics with slow-moving capital. Journal of Finance 65, 1237–1267.

Eisdorfer, A., 2008. Delisted firms and momentum profits. Journal of Financial Markets 11, 160–179.

Fama, E., 1984. Forward and spot exchange rates. Journal of Monetary Economics 14, 319–338.

Fama, E., French, K., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.

Fama, E., French, K., 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55–84.

Fama, E., MacBeth, J., 1973. Risk, return and equilibrium: empirical tests. Journal of Political Economy 81, 607–636.

Gebhardt, W., Hvidkjaer, S., Swaminathan, B., 2005. Stock and bond market interaction: Does momentum spill over? Journal of Financial Economics 75, 651–690.

Gorton, G., Hayashi, F., Rouwenhorst, K., 2008. The fundamentals of commodity futures returns. Yale ICF Working Paper No. 07-08.

Goyal, A., Saretto, A., 2009. Cross-section of option returns and volatility. Journal of Financial Economics 94, 310–326.

Griffin, J., Martin, J., 2003. Momentum investing and business cycle risk: evidence from pole to pole. Journal of Finance 58, 2515–2547.

Grinblatt, M., Han, B., 2005. Prospect theory, mental accounting, and momentum. Journal of Financial Economics 78, 311–339.

Gromb, D., Vayanos, D., 2010. Limits of arbitrage: the state of the theory. Annual Review of Financial Economics 2, 251–275.

Gutierrez Jr, R., Kelley, E., 2008. The long-lasting momentum in weekly returns. Journal of Finance 63, 415–447.

Harris, R.D., Yilmaz, F., 2009. A momentum trading strategy based on the low frequency component of the exchange rate. Journal of Banking and Finance 33, 1575–1585.

Harvey, C., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55, 1263–1295.

Hong, H., Lim, T., Stein, J., 2000. Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies. Journal of Finance 55, 265–295.

Hong, H., Stein, J., 1999. A unified theory of underreaction, momentum trading and overreaction in asset markets. Journal of Finance 54, 2143–2184.

Hvidkjaer, S., 2006. A trade-based analysis of momentum. Review of Financial Studies 19, 457–491.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.

Jegadeesh, N., Titman, S., 2001. Profitability of momentum strategies: an evaluation of alternative explanations. Journal of Finance 56, 699–720.

Johnson, T., 2002. Rational momentum effects. Journal of Finance 57, 585–608.

Jostova, G., Nikolova, S., Philipov, A., Stahel, C., 2010. Momentum in corporate bond returns. Unpublished working paper, George Washington University, U.S. Securities and Exchange Commission, George Mason University, U.S. Securities and Exchange Commission.

King, M., Osler, C., Rime, D., 2012. Foreign exchange market structure, players and evolution. In: James, J., Marsh, I., Sarno, L. (Eds.), Handbook of Exchange Rates, Wiley Publishing Inc., New Jersey, pp. 1–48.

Korajczyk, R., Sadka, R., 2004. Are momentum profits robust to trading costs? Journal of Finance 59, 1039–1082.

Lequeux, P., Acar, E., 1998. A dynamic benchmark model for managed currencies funds using cme currency contracts. European Journal of Finance 4, 311–330.

Lesmond, D., Schill, M., Zhou, C., 2004. The illusory nature of momentum profits. Journal of Financial Economics 43, 1121–1142.

Levich, R., Thomas III, L., 1993. The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach. Journal of International Money and Finance 12, 451–474.

Liu, L., Zhang, L., 2011. A model of momentum. NBER Working Paper No. 16747.

Lustig, H., Roussanov, N., Verdelhan, A., 2011. Common risk factors in currency markets. Review of Financial Studies 24, 3731–3777.

Lustig, H., Verdelhan, A., 2007. The cross section of foreign currency risk premia and consumption growth risk. American Economic Review 97, 89–117.

Lyons, R., 2001. The Microstructure Approach to Exchange Rates. MIT Press, Cambridge, MA.

Melvin, M., Shand, D., 2011. Active currency investing and performance benchmarks. Journal of Portfolio Management 37, 46–59.

Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A., 2012. Carry trades and global foreign exchange volatility. Journal of Finance 67, 681–718.

Menkhoff, L., Taylor, M., 2007. The obstinate passion of foreign exchange professionals: technical analysis. Journal of Economic Literature 45, 936–972.

Mitchell, M., Pedersen, L., Pulvino, T., 2007. Slow moving capital. American Economic Review, Papers and Proceedings 97, 215–220.

Moskowitz, T., Ooi, Y., Pedersen, L., 2012. Time series momentum. Journal of Financial Economics 104, 228–250.

Neely, C., 1998. Technical Analysis and the Profitability of U.S. Foreign Exchange Intervention. Review, Federal Reserve Bank of St. Louis 80, 3–17.

Neely, C., 2002. The temporal pattern of trading rule returns and exchange rate intervention: intervention does not generate technical trading rule profits. Journal of International Economics 58, 211–232.

Neely, C., Weller, P., 1999. Technical trading rules in the European Monetary System. Journal of International Money and Finance 18, 429–458.

Neely, C., Weller, P., 2012. Technical analysis in the foreign exchange market. In: James, J., Marsh, I., Sarno, L. (Eds.), Handbook of Exchange Rates, Wiley Publishing Inc., New Jersey, pp. 369–402.

Neely, C., Weller, P., Dittmar, R., 1997. Is technical analysis in the foreign exchange market profitable? A genetic programming approach. Journal of Financial and Quantitative Analysis 32, 405–426.

Neely, C.J., Weller, P., Ulrich, J., 2009. The adaptive markets hypothesis: evidence from the foreign exchange market. Journal of Financial and Quantitative Analysis 44, 467–488.

Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Okunev, J., White, D., 2003. Do momentum-based strategies still work in foreign currency markets? Journal of Financial and Quantitative Analysis 38, 425–447.

Olson, D., 2004. Have trading rule profits in the currency markets declined over time? Journal of Banking and Finance 28, 85–105.

Pastor, L., Stambaugh, R., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

Patton, A., Ramadorai, T. On the high frequency dynamics of hedge fund risk exposures. Journal of Finance, in press.

Pojarliev, M., Levich, R., 2010. Trades of the living dead: style differences, style persistence and performance of currency fund managers. Journal of International Money and Finance 29, 1752–1775.

Pukthuanthong-Le, K., Levich, R., Thomas III, L., 2007. Do foreign exchange markets still trend? Journal of Portfolio Management 34, 114–118.

Pukthuanthong-Le, K., Thomas III, L., 2008. Weak-form efficiency in currency markets. Financial Analysts Journal 64, 31–52.

Rouwenhorst, K., 1998. International momentum strategies. Journal of Finance 53, 267–284.

Rouwenhorst, K., 1999. Local return factors and turnover in emerging stock markets. Journal of Finance 54, 1439–1464.

Sagi, J., Seascholes, M., 2007. Firm-specific attributes and the cross-section of momentum. Journal of Financial Economics 84, 389–434.

Sarno, L., Schneider, P., Wagner, C., 2012. Properties of Foreign Exchange Risk Premiums. Journal of Financial Economics 105, 279–310.

Sarno, L., Taylor, M., 2001. Official intervention in the foreign exchange market: Is it effective and, if so, how does it work? Journal of Economic Literature 39, 839–868.

Serban, A., 2010. Combining mean reversion and momentum trading strategies in foreign exchange markets. Journal of Banking and Finance 34, 2720–2727.

Shleifer, A., Vishny, R., 1997. The limits of arbitrage. Journal of Finance 52, 35–55.

Silber, W., 1994. Technical trading: when it works and when it doesn't. Journal of Derivatives 1, 39–44.

Sweeney, R., 1986. Beating the foreign exchange market. Journal of Finance 41, 163–182.

Verardo, M., 2009. Heterogeneous beliefs and momentum profits. Journal of Financial and Quantitative Analysis 44, 795–822.

Currency Premia and Global Imbalances[*†]

Pasquale DELLA CORTE Steven J. RIDDIOUGH Lucio SARNO

First submission: February 2014 - Revised: December 2014

[*]Acknowledgements: We are grateful for comments and suggestions to Geert Bekaert (Editor), three anonymous referees, Rui Albuquerque, Nicola Borri, Justinas Brazys, Craig Burnside, Antonio Diez de los Rios, Xavier Gabaix, Federico Gavazzoni, Leonid Kogan, Ralph Koijen, Matteo Maggiori, Lukas Menkhoff, David Ng, Maik Schmeling, Astrid Schornick, Giulia Sestieri, Andreas Stathopoulos, Ilias Tsiakas and Adrien Verdelhan. We also thank participants at the 2013 WFA meeting, 2013 EFMA meeting, 2013 First Annual Conference on "Foreign Exchange Markets", 2013 China International Conference in Finance, 2013 Asian Finance Association meeting, 2013 EFA conference, 2013 FMA meeting, 2013 Royal Economic Society Conference, Bank of Canada/Bank of Spain Workshop on "International Financial Markets," 2013 International Paris Finance Meeting, 2012 Annual Conference on "Advances in the Analysis of Hedge Fund Strategies," 2012 ECB-Bank of Italy Workshop on "Financial Determinants of Exchange Rates" and seminar participants at various institutions. We thank JP Morgan, Philip Lane, Gian Milesi-Ferretti and Jay Shambaugh for making data available which are used in this study. We gratefully acknowledge financial support from INQUIRE Europe. Steven Riddiough gratefully acknowledges financial support from the Economic and Social Research Council (ESRC), and Lucio Sarno acknowledges the gracious hospitality of the Cambridge Endowment for Research in Finance (CERF) of the University of Cambridge, where part of this research was conducted. The paper was the recipient of the Kepos Capital Award for the Best Paper on Investments at the 2013 WFA Annual Meeting. All errors remain ours.

[†]Pasquale Della Corte is with Imperial College Business School, Imperial College London, email: `p.dellacorte@imperial.ac.uk`. Steven Riddiough is with Warwick Business School, University of Warwick, email: `s.j.riddiough@warwick.ac.uk`. Lucio Sarno is with Cass Business School, City University London and CEPR, email: `lucio.sarno@city.ac.uk`.

Currency Premia and Global Imbalances

First version: February 2014 - Revised: December 2014

Abstract

We show that a global imbalance risk factor that captures the spread in countries' external imbalances and their propensity to issue external liabilities in foreign currency explains the cross-sectional variation in currency excess returns. The economic intuition is simple: net debtor countries offer a currency risk premium to compensate investors willing to finance negative external imbalances because their currencies depreciate in bad times. This mechanism is consistent with recent exchange rate theory based on capital flows in imperfect financial markets. We also find that the global imbalance factor is priced in the cross sections of other major asset markets.

Keywords: Currency Risk Premium; Global Imbalances; Carry Trade.

JEL Classification: F31; F37; G12; G15.

1 Introduction

Imbalances in trade and capital flows have been the centerpiece of much debate surrounding the causes and consequences of the global financial crisis. Therefore it would seem natural that, given the financial crisis consisted of collapsing asset prices worldwide, global imbalances may help shed light on our fundamental understanding of asset price dynamics. The foreign exchange (FX) market provides a logical starting point for testing this hypothesis as exchange rate fluctuations and currency risk premia are theoretically linked to external imbalances (e.g., Gabaix and Maggiori, hereafter GM, 2014), and recent events in the FX market provide a reminder of the potential importance of such a link. Following the US Federal Reserve's announcement on 22 May 2013 that they would taper the size of their bond-buying programme, emerging market currencies including the Indian rupee, Brazilian real, South African rand and Turkish lira all sold-off sharply. A common characteristic among these four countries is that they are some of the world's largest debtor nations. In fact, the Financial Times on 26 June 2013 attributed the large depreciation of the Indian rupee (which fell by 22 percent against the US dollar between May and August 2013) to investors' concerns over India being "one of the most vulnerable emerging market currencies due to its current account deficit" (Ross, 2013).

In this paper we provide empirical evidence that exposure to countries' external imbalances is key to understanding currency risk premia and thus supports a risk-based interpretation of the carry trade, a popular strategy that involves an investor borrowing in currencies with low interest rates (funding currencies) and simultaneously lending in currencies with high interest rates (investment currencies).[1] Our findings are consistent with the model of GM (2014), who provide a novel theory of exchange rate determination based on capital flows in imperfect financial markets. Specifically, GM (2014) propose a two-country model in which exchange rates are jointly determined by global imbalances and financiers' risk-bearing capacity. In their model, countries run trade imbalances and financiers absorb the resultant currency risk, i.e., financiers are long the debtor country and short the creditor country. Financiers, however,

[1] The carry trade strategy builds on the violation of uncovered interest rate parity. See Hansen and Hodrick (1980), Bilson (1981), Fama (1984), Engel (1996), Lustig and Verdelhan (2007), Della Corte, Sarno and Tsiakas (2009), Lustig, Roussanov and Verdelhan (2011), Burnside, Eichenbaum, Kleshchelski and Rebelo (2011), Koijen, Moskowitz, Pedersen and Vrugt (2013), Menkhoff, Sarno, Schmeling and Schrimpf (2012a), and Lettau, Maggiori and Weber (2014).

are financially constrained and this affects their ability to take positions. Intuitively, if there is little risk-bearing capacity financiers are unwilling to intermediate currency mismatches regardless of the excess return on offer. In contrast, when financiers have unlimited risk-bearing capacity they are willing to take positions in currencies whenever a positive excess return is available, and hence the currency risk premium is miniscule.

We focus the empirical analysis around two simple testable hypotheses, which are consistent with predictions of the GM (2014) theory. First, currency excess returns, i.e. the returns to a carry trade strategy, are higher when (i) interest rate differentials are larger, (ii) the funding (investment) country is a net foreign creditor (debtor), and (iii) the funding (investment) country has a higher propensity to issue liabilities denominated in domestic (foreign) currency. Condition (i) is standard since on average exchange rate movements do not offset interest rate differentials due to the risk premium. Condition (ii) is novel and captures the link between external imbalances and currency risk premia in the theory of GM (2014). Condition (iii) is not derived explicitly in GM (2014) but it reflects their argument, studied in one of their extensions of the model, that the currency denomination of external debt also matters for currency risk premia. One reason why this is the case is that countries which do not or cannot issue debt in their own currency are more risky.[2] In essence, this testable hypothesis distinguishes between the interest rate and the net foreign asset position as two different, even if related, sources of currency risk premia.

Second, the GM theory also predicts that, when there is a financial disruption (i.e., risk-bearing capacity is very low and global risk aversion is very high), net-debtor countries experience a currency depreciation, unlike net-creditor countries. This testable hypothesis makes clearer an important part of the mechanism that generates currency risk premia in imperfect capital markets: investors demand a risk premium for holding net debtor countries' currencies *because* these currencies perform poorly in bad times, which are times of large shocks to risk-bearing capacity and global risk aversion.

In our empirical analysis, we test and provide evidence in support of the two hypotheses described above. In relation to the first testable hypothesis, we document that a currency

[2] This mechanism is consistent with the literature on the 'original sin' (e.g. Eichengreen and Hausmann, 2005) and is akin to a valuation channel for the external adjustment of imbalances whereby the exchange rate facilitates the re-equilibration of external imbalances (see e.g. Gourinchas and Rey, 2007; Gourinchas, 2008; Lane and Shambaugh, 2010).

strategy that sorts currencies on net foreign asset positions and a country's propensity to issue external liabilities in domestic currency – termed the 'global imbalance' strategy – generates a large spread in returns. Then, we empirically test whether a risk factor that captures the combination of spread in external imbalances and the propensity to issue external liabilities in domestic currency can explain the cross-section of currency excess returns in a standard asset pricing framework. Our central result in this respect is that our global imbalance risk factor explains over 90 percent of currency excess returns, thus supporting a risk-based view of exchange rate determination that is based on macroeconomic fundamentals and, specifically, on net foreign asset positions. This result holds both for a broad sample of 55 currencies and for a subsample of the 15 most liquid currencies over the period from 1983 to 2014.[3]

The economic intuition of this factor is simple, in line with the GM (2014) model: investors demand a risk premium to hold the currency of net debtor countries, especially if funded principally in foreign currency. High interest rate currencies load positively on the global imbalance factor, and thus deliver low returns in bad times when there is a spike in global risk aversion and the process of international financial adjustment requires their depreciation. Low interest rate currencies are negatively related to the global imbalance factor, and thus provide a hedge by yielding positive returns in bad times. This result suggests that returns to carry trades are compensation for time-varying fundamental risk, and thus carry traders can be viewed as taking on global imbalance risk.

We also document how net foreign asset positions contain information that is (related but) not identical to interest rate differentials in the cross section of countries. Conditioning on information on external imbalances to form currency portfolios provides economic value to a currency investor who uses conventional strategies such as carry, momentum, and value. Notably, in a global minimum volatility portfolio that includes carry, momentum, value and our proposed global imbalance strategy, we find that the latter strategy receives a weight of 46 and 51 percent for the two samples of countries we examine. This supports the validity of GM's (2014) prediction that there is an effect of net foreign asset positions on currency excess

[3]Despite the existence of theoretical models that link exchange rates to external imbalances, there have hardly been any attempts to relate currency risk premia *cross-sectionally* to currencies' sensitivity to external imbalances. When the FX literature has investigated the empirical link between exchange rates and external imbalances, the analysis was carried out in a time series setting (e.g. Alquist and Chinn, 2008; Della Corte, Sarno and Sestieri, 2012). It thus seems quite natural to employ a cross-sectional perspective on the role of global imbalances to help understand currency risk premia in general, and carry trades in particular.

returns that is distinct from a pure interest rate channel.

In relation to the second testable hypothesis, we provide evidence using a battery of panel regressions that in bad times (defined as times of large global risk aversion shocks, as proxied by the change in the VIX) net-debtor countries experience a currency depreciation, whereas net-creditor countries experience an appreciation. This result is consistent with the risk premium story of GM (2014): investors demand a risk premium for holding net debtor countries' currencies *because* these currencies perform poorly in bad times.

Our paper builds on the growing literature searching for a risk-based explanation to currency premia. Lustig, Roussanov and Verdelhan (2011) and Menkhoff, Sarno, Schmeling and Schrimpf (2012a) have both found a global risk factor in currency excess returns. However, while these global risk factors provide valuable information on the properties of currency returns, the question as to what fundamental economic forces drive the factors and, hence, currency risk premia, remains unanswered, leaving us tantalizingly close to a more complete understanding of currency premia. This paper tackles this issue by shedding some light on the *macroeconomic* forces driving currency premia.[4]

In the empirical analysis we sort currencies into five portfolios according to their forward discounts as pioneered by Lustig and Verdelhan (2007). This is equivalent to using the interest rate differential relative to the US dollar to rank foreign currencies because no-arbitrage requires that forward discounts are equal to interest rate differentials. The first portfolio contains the funding currencies of a carry trade strategy (low-yielding currencies relative to the US dollar) while the last portfolio contains the investment currencies in a carry trade strategy (high-yielding currencies relative to the US dollar). We then show that carry trade returns can be understood as compensation for risk by relating their cross-section to the global imbalance factor. This factor is an easily constructed variable. We first split currencies into two baskets using the ratio of net foreign assets to GDP, and then sort currencies within each basket based on countries' percentage share of external liabilities denominated in domestic currency. The reordered currencies, beginning with creditors whose external liabilities are primarily denominated in domestic currency (the safest currencies) and moving to debtors whose external liabilities are primarily denominated in foreign currency (the riskiest currencies), are

[4] Other papers studying carry trade returns include Brunnermeier, Nagel and Pedersen (2009), Christiansen, Ranaldo and Söderlind (2011), Colacito and Croce (2013), Farhi and Gabaix (2014), Farhi, Fraiberger, Gabaix, Ranciere and Verdelhan (2014) and Jurek (2014).

grouped into quintiles. These quintiles form our five 'global imbalance' portfolios. The global imbalance factor is simply constructed as the difference between the excess returns on the extreme portfolios. It is equivalent to a high-minus-low strategy that buys the currencies of debtor nations with mainly foreign currency denominated external liabilities and sells the currencies of creditor nations with mainly domestic currency denominated external liabilities. We refer to the global imbalance risk factor as the IMB factor, or simply IMB.

It is important to note that, while the IMB factor contains information that is clearly related to the spread in interest rates across countries, its pricing power is not mechanical in the sense that it cannot be attributed simply to feedback effects from interest rates to net foreign assets. Although feedback effects may exist between interest rates and net foreign assets whereby higher interest rates attract more capital flows, global imbalances capture fundamental information related to currency risk premia that is not embedded in interest rates. This argument is key in GM (2014), who show theoretically that currency premia will be required even if both countries have the same interest rate as long as one is a debtor relative to the other.[5] In fact, recent anecdotal evidence emphasizes the fundamental importance of net foreign assets over and above interest rates in determining currency premia: the US Federal Reserve's announcement in May 2013 that it would scale-back its bond buying programme caused a spike in risk aversion – the VIX index rose from below 14 to over 20 during the subsequent month. In currency markets, following the Federal Reserve announcement, currencies with very similar interest rates behaved very differently, and only the currencies with large external deficit positions experienced sharp depreciations. The economic mechanism in GM (2014) and the empirical evidence in this paper make sense of this behavior, which would be hard to rationalize otherwise.[6]

Further analysis provides refinements and robustness of our main results, including the following: (i) We show that sorting currencies on their beta with the global imbalance factor

[5] This result is also consistent with the empirical work of Habib and Stracca (2012), who find that net foreign assets are more important for predicting exchange rate returns than interest rate differentials.

[6] Specifically, at the point of the Federal Reserve announcement, Australia, New Zealand, and South Korea – three of the most volatile currencies in the Asia Pacific region – had almost identical interest rates (2.50 percent in New Zealand and Korea, 2.75 percent in Australia). Yet, over the May to September period, the Australian dollar depreciated by 16 percent against the US dollar, the New Zealand dollar depreciated by 10 percent, while the Korean won fell by only 1 percent. The contrasting sizes of depreciation reflect the contrast in deficit positions at the end of the first quarter of 2013, when Australia and New Zealand both had external deficit positions relative to GDP of over 60 percent, while South Korea had a far more modest 6 percent deficit.

yields portfolios with a significant difference in returns. These portfolios are related, but not identical, to the base test assets of currency portfolios sorted on forward discounts. (ii) We test the pricing power of the global imbalance risk factor for currency excess returns sorted by momentum and value, as well as for cross-sections of returns in other markets, including equities, bonds and commodities. These tests are much more powerful given the larger cross-section of test assets, and the results suggest that the IMB factor also prices these portfolios. (iii) We depart from the base scenario of a US-based investor and run calculations using alternative base currencies, taking the viewpoint of a British, Japanese, Euro-based and Swiss investor. The results indicate that the IMB factor is priced in each case. (iv) We test the IMB risk factor on portfolios formed using only the most liquid developed and emerging currencies, showing that there are no qualitative changes in the results. (v) We measure the individual contribution to the currency risk premium of net foreign asset positions and the propensity to issue liabilities in domestic currency, and find that both matter. (vi) We also run cross-sectional asset pricing tests on individual currencies' excess returns, and again record that IMB is priced. Overall, we find that the further analysis corroborates the core finding that global imbalance risk is a key fundamental driver of risk premia in the FX market.

The remainder of the paper is organized as follows. Section 2 describes the theoretical background of our analysis and the hypotheses we take to the data. In Section 3 we describe the data and provide details of how portfolios are constructed. In Section 4 we describe the properties of the global imbalance strategy and its returns, while in Section 5 we report asset pricing tests and further analysis to understand the link between currency excess returns and global imbalances. Section 6 reports evidence on the behavior of exchange rate returns and global imbalances in bad times. In Section 7 we present a number of extensions and robustness exercises, before concluding in Section 8. A separate Internet Appendix provides robustness tests and additional analyses.

2 Theoretical Motivation and Testable Hypotheses

The contribution of this paper is purely empirical, but our empirical analysis has a clear theoretical motivation, which is based primarily on the recent theory of exchange rate determination proposed by GM (2014). This theory makes a substantial leap forward in considering the interaction between capital flows and financial intermediaries' risk-bearing capacity in a model of exchange rate determination with imperfect financial markets. In the most basic, two-period version of the model – termed the 'Gamma' model – each country borrows or lends in its own currency and global financial intermediaries absorb the exchange rate risk arising from imbalanced capital flows. Since financial intermediaries demand compensation for holding currency risk in the form of an expected currency appreciation, exchange rates are jointly determined by global capital flows and by the intermediaries' risk-bearing capacity, which GM (2014) refer to as 'broadly defined risk aversion shocks'.[7]

This theory has clear implications for the returns to a carry trade strategy that buys high-interest rate ("investment") currencies and sells low-interest rate ("funding") currencies. GM (2014, equation (27), Proposition 12) derive the return to the carry trade from the Gamma model as follows:

$$E\left(RX\right) = \Gamma \frac{\frac{i^*}{i} E\left(imp_1\right) - imp_0}{\left(i^* + \Gamma\right) imp_0 + \frac{i^*}{i} E\left(imp_1\right)} \tag{1}$$

where RX denotes currency excess returns, or the return to a carry trade strategy (higher RX means higher carry trade returns). The variable imp_t denotes the dollar value of US imports at time t; with exports normalized to unity in equation (1), the evolution of imports $E\left(imp_1\right) - imp_0$ determines net foreign asset positions in the model. i and i^* are the domestic and foreign riskless interest rates. Γ captures risk-bearing capacity of financiers. When risk-bearing capacity is low (Γ is high in equation (1)), financial intermediaries are unwilling to absorb any imbalances, regardless of the excess return available, and hence no financial flows are necessary as trade inflows and outflows will be equal in each period. As risk-bearing capacity increases (Γ decreases), excess returns fall but do not entirely disappear, except when Γ is extremely low and financial intermediaries are prepared to absorb any currency imbalance so that uncovered interest rate parity holds. GM (2014) show that during periods of financial distress, when risk-bearing capacity declines, debtor countries suffer a currency depreciation whereas creditor countries experience a currency appreciation.

[7] In the empirical work, we proxy Γ by changes in the VIX, a commonly used measure of global risk aversion. In the first version of their model, GM had Γ exogenous. In the latest draft of GM (2014), the model is solved under the constraint that Γ is directly related to conditional volatility, which nests as a special case the simpler model where Γ is exogenous. In the more general formulation an increase in volatility tightens the constraint both directly and indirectly via feedback effects. In short, GM (2014) refer to Γ as loosely proxying for global risk aversion shocks, and the latest version of their model also allows for the direct effect of conditional volatility.

Equation (1) shows that the returns of a carry trade will be higher when interest rate differentials are larger, and when the funding (investment) currency is issued by a net creditor (debtor) country. Put another way, carry traders require a premium to hold the currency of debtor nations relative to creditor nations.[8]

In the basic version of the Gamma model, by assumption each country borrows or lends in its own currency, but in practice most countries do not (or cannot) issue all their debt in their own currency.[9] Although GM (2014) do not provide a full analytical extension of their model that allows for currency mismatches between assets and liabilities, in Section 2.3 and in Proposition 7 (point 3), GM (2014) consider the impact of the currency denomination of external liabilities, illustrating how this generates a valuation channel to the external adjustment of countries whereby the exchange rate moves in a way that facilitates the re-equilibration of external imbalances.[10] GM highlight how this result is consistent with the valuation channel to external adjustment studied by e.g. Gourinchas and Rey (2007), Gourinchas (2008), and Lane and Shambaugh (2010), and gives a role to the currency denomination of external liabilities in their model. Thus in our empirical analysis we also allow for this effect by considering whether currencies of countries with a higher propensity to issue liabilities in foreign currency offer a higher currency risk premium, given that such countries require much sharper depreciations to correct their external imbalances.[11]

The mechanism described above implies the first testable hypothesis, which is a variant of Proposition 12 in GM (2014) with the additional condition that captures the effect of the currency denomination of liabilities.

Hypothesis 1 *The carry trade return is bigger when (i) the interest rate differential is*

[8] To clarify these effects analytically in equation (1), first consider the case when $i^*/i > 1$, i.e., the interest rate in the foreign (investment) country is higher than the one in the funding country (the US). GM show that $\frac{\partial E(RX)}{\partial(i^*-i)} > 0$, which means that the carry trade return increases with higher interest rate differentials. Second, set $E\left[imp_1\right] - imp_0 > 0$ (while setting $i^*/i = 1$), i.e., the funding country (the US) is a net foreign creditor. Given that imp is the value of US imports in US dollars, $E\left[imp_1\right] - imp_0 > 0$ implies that the US is expected to become a net importer at $t = 1$ in order to offset its positive external imbalance at $t = 0$, and clearly $\frac{\partial E(RX)}{\partial(E(imp_1)-imp_0)} > 0$. This establishes the result that the expected carry trade return is higher if the country of the funding currency is a net creditor, and viceversa.

[9] See the literature on the 'original sin' (e.g. Eichengreen and Hausmann, 2005, and the references therein).

[10] This is achieved by allowing for some pre-existing level of debt which is not necessarily fully denominated in a country's own currency.

[11] This is because the initial depreciation makes countries with foreign-currency denominated liabilities poorer, not richer, by increasing their debt burden (see e.g. the portfolio balance model in Gourinchas, 2008, Section 3.2.2, and the references therein).

larger, (ii) the funding (investment) country is a net foreign creditor (debtor), and (iii) the funding (investment) country has a higher propensity to issue liabilities denominated in domestic (foreign) currency.

This testable prediction suggests that there are two drivers of FX excess returns, the first driven purely by the carry component (condition (i) above), and the second driven by the evolution of external debt and its currency of denomination (conditions (ii) and (iii)). In our portfolio analysis, we combine the information in conditions (ii) and (iii) to capture both the spread in external imbalances and the propensity to issue external liabilities in foreign currency, although we also examine their separate effects in some of our tests and in the regression analysis.

We test Hypothesis 1 in several ways. Above all, we form portfolios sorted on external imbalances (net foreign assets to GDP ratio) and the share of foreign liabilities in domestic currency to examine whether they provide predictive information for the cross-section of currency excess returns. We show that this portfolio sort generates a sizable and statistically significant spread in returns: a currency strategy that buys the extreme net debtor countries with highest propensity to issue external liabilities in foreign currency and sells the extreme creditor countries with lowest propensity to issue liabilities in foreign currency – which we term the 'global imbalance' strategy – generates Sharpe ratios of 0.59 for a universe of major countries and 0.68 for a broader set of 55 countries. This confirms the essence of Hypothesis 1 that currency excess returns are higher for net-debtor countries with higher propensity to issue liabilities in foreign currency, which are also found to be countries with higher interest rates, i.e. typical investment currencies in the carry trade.

A central mechanism in the model of GM (2014) is that during periods of financial distress, when risk-bearing capacity declines, debtor countries suffer a currency depreciation whereas creditor countries experience a currency appreciation. This is indeed the logic that rationalizes why net debtor countries must offer a currency risk premium, implying the second empirical prediction we take to the data, which is Proposition 2 of GM (2014).

Hypothesis 2 *When there is a financial disruption (Γ increases), countries that are net external debtors experience a currency depreciation, while the opposite is true for net-creditor countries.*

This prediction follows naturally from the previous analysis and our empirical results provide supporting evidence on its validity through the estimation of a battery of panel regressions.

3 Data and Currency Portfolios

This section describes the main data employed in the empirical analysis. We also describe the construction of currency portfolios and the global imbalance risk factor.

Data on Spot and Forward Exchange Rates. We collect daily spot and 1-month forward exchange rates vis-à-vis the US dollar (USD) from Barclays and Reuters via Datastream. The empirical analysis uses monthly data obtained by sampling end-of-month rates from October 1983 to June 2014. Our sample comprises 55 countries: Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, China, Colombia, Croatia, Czech Republic, Denmark, Egypt, Estonia, Euro Area, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, India, Indonesia, Ireland, Israel, Italy, Japan, Kazakhstan, Latvia, Lithuania, Malaysia, Mexico, Morocco, Netherlands, New Zealand, Norway, Philippines, Poland, Portugal, Russia, Singapore, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Thailand, Tunisia, Turkey, Ukraine, United Kingdom, and Venezuela. We call this sample 'all countries'.

A number of currencies in this sample are pegged or subject to capital restrictions. In reality, investors may not easily trade some of these currencies in large amounts even though quotes on forward contracts (deliverable or non-deliverable) are available.[12] Hence, we also consider a subset of 15 countries which we refer to as 'developed countries'. This sample includes: Australia, Belgium, Canada, Denmark, Euro Area, France, Germany, Italy, Japan, Netherlands, New Zealand, Norway, Sweden, Switzerland, and the United Kingdom. After the introduction of the euro in January 1999, we remove the Eurozone countries and replace them with the euro.[13] As in Lustig, Roussanov, and Verdelhan (2011) and many subsequent studies, we remove data when we observe large deviations from the covered interest rate parity

[12] According to the Triennial Survey of the Bank for International Settlements (2013), the top 10 currencies account for 90 percent of the average daily turnover in FX markets.

[13] The sample of developed countries matches both Lustig, Roussanov, and Verdelhan (2011) and Menkhoff, Sarno, Schmeling and Schrimpf (2012). The full sample of countries, instead, comprises a wider set of countries than previous studies. We also consider a set of 35 countries as in Lustig, Roussanov, and Verdelhan (2011) and 48 countries as in Menkhoff, Sarno, Schmeling and Schrimpf (2012a), as well as a set of countries comprising the most liquid currencies (termed 'developed and emerging' sample). Qualitatively, the results remain the same; see e.g. Tables A.8 and A.9 of the Internet Appendix.

(CIP) condition.[14]

Data on External Assets and Liabilities. Turning to macroeconomic data, we obtain end-of-year series on foreign assets and liabilities, and gross domestic product (GDP) from Lane and Milesi-Ferretti (2004, 2007), kindly updated by Gian Maria Milesi-Ferretti. Foreign (or external) assets are measured as the dollar value of assets a country owns abroad, while foreign (or external) liabilities refer to the dollar value of domestic assets owned by foreigners. The data for all countries included in our study are until the end of 2012. For each country we measure external imbalances – the indebtedness of a country to foreigners – using the net foreign asset position (the difference between foreign assets and foreign liabilities) relative to the size of the economy (GDP), which we denote nfa. We retrieve monthly observations by keeping end-of-period data constant until a new observation becomes available.

We also use end-of-year series on the proportion of external liabilities denominated in domestic currency (denoted ldc) from Benetrix, Lane and Shambaugh (2014), which updates the data from Lane and Shambaugh (2010), kindly provided by Philip Lane. The data is available from 1990 to 2012. We construct monthly observations by keeping end-of-period data constant until a new observation becomes available. Note that we maintain the 1990 proportions back until 1983.[15]

Currency Excess Returns. We define spot and forward exchange rates at time t as S_t and F_t, respectively, and take into account the standard value date conventions in matching the forward rate with the appropriate spot rate (e.g., Bekaert and Hodrick, 1993). Exchange

[14] On the one hand, one may be concerned that times when CIP deviations occur are likely to be times of distress in the FX market that can be particularly informative about risk premia. On the other hand, some of those times can be characterized by extreme illiquidity and lack of tradability, so that prices are essentially uninformative. Moreover, the GM theory assumes no riskless arbitrage. We filtered the data as follows: for the Developed sample of 15 countries, we do not remove any observations. For the broader sample, we eliminate the following observations: Argentina from September 2008 to April 2009, and from May 2012 to June 2014; Egypt from November 2011 to August 2013; Indonesia from December 1997 to July 1998, and from February 2001 to May 2005; Malaysia from May 1998 to June 2005; Turkey from November 2000 to November 2001; South Africa for August 1985, and from January 2002 to May 2005; Russia from December 2008 to January 2009; Kazakhstan from November 2008 to February 2009. These are episodes where CIP deviations were very large (generally in excess of 25 percent) and likely not tradable. Note that the removal of CIP deviations does not affect any of our results in this paper with two exceptions: Turkey for the period around the 2001 devaluation, and Malaysia for the 1998-2005 period of capital controls.

[15] This assumption makes no qualitative difference to our findings as when we examine the sample period starting in 1990 (dropping the first 7 years of data altogether) our portfolio results are qualitatively identical. This is not surprising since ldc is a highly persistent variable (see Benetrix, Lane and Shambaugh, 2014).

rates are defined as units of US dollars per unit of foreign currency such that an increase in S_t indicates an appreciation of the foreign currency. The excess return on buying a foreign currency in the forward market at time t and then selling it in the spot market at time $t+1$ is computed as

$$RX_{t+1} = \frac{(S_{t+1} - F_t)}{S_t}, \tag{2}$$

which is equivalent to the spot exchange rate return minus the forward premium

$$RX_{t+1} = \frac{S_{t+1} - S_t}{S_t} - \frac{F_t - S_t}{S_t}. \tag{3}$$

According to the CIP condition, the forward premium approximately equals the interest rate differential $(F_t - S_t)/S_t \simeq i_t - i_t^*$, where i_t and i_t^* represent the US and the foreign riskless rates respectively, over the maturity of the forward contract. Since CIP holds closely in the data (e.g., Akram, Rime, and Sarno, 2008), the currency excess return is approximately equal to the exchange rate return (i.e., $(S_{t+1} - S_t)/S_t$) plus the interest rate differential relative to the US (i.e., $i_t^* - i_t$). As a matter of convenience, throughout this paper we refer to $fd_t = (S_t - F_t)/S_t = i_t^* - i_t$ as the forward discount or interest rate differential relative to the US.

We construct currency excess returns adjusted for transaction costs using bid-ask quotes on spot and forward rates. The net excess return for holding foreign currency for a month is computed as $RX_{t+1}^l \simeq (S_{t+1}^b - F_t^a)/S_t^a$, where a indicates the ask price, b the bid price, and l a long position in a foreign currency. The net excess return accounts for the full round-trip transaction cost occurring when the foreign currency is purchased at time t and sold at time $t+1$. If the investor buys foreign currency at time t but decides to maintain the position at time $t+1$, the net excess return is computed as $RX_{t+1}^l \simeq (S_{t+1} - F_t^a)/S_t^a$. Similarly, if the investor closes the position in foreign currency at time $t+1$ already existing at time t, the net excess return is defined as $RX_{t+1}^l \simeq (S_{t+1}^b - F_t)/S_t^b$. The net excess return for holding domestic currency for a month is computed as $RX_{t+1}^s \simeq (F_t^b - S_{t+1}^a)/S_t^b$, where s stands for a short position on a foreign currency. In this case, the investor sells foreign currency at time t in the forward market at the bid price F_t^b and offsets the position in the spot market at time $t+1$ using the ask price S_{t+1}^a. If the foreign currency leaves the strategy at time t and the short position is rolled over at time $t+1$, the net excess return is computed as $RX_{t+1}^s \simeq (F_t^b - S_{t+1})/S_t^b$. Similarly, if the investor closes a short position on the foreign currency at time $t+1$ already existing at time t, the net excess return is computed as $RX^s_{t+1} \simeq (F_t - S^a_{t+1})/S^b_t$. In short, excess returns are adjusted for the full round-trip transaction cost in the first and last month of the sample. The total number of currencies in our portfolios changes over time, and we only include currencies for which we have bid and ask quotes on forward and spot exchange rates in the current and subsequent period.

Carry Trade Portfolios. We construct five carry portfolios, rebalanced monthly, and use them as test assets in our empirical asset pricing analysis. At the end of each period t, we allocate currencies to five portfolios on the basis of their forward discounts (or interest rate differential relative to the US). This exercise implies that currencies with the lowest forward discounts (or lowest interest rate differential relative to the US) are assigned to Portfolio 1, whereas currencies with the highest forward discounts (or highest interest rate differential relative to the US) are assigned to Portfolio 5. We then compute the excess return for each portfolio as an equally weighted average of the currency excess returns within that portfolio. For the purpose of computing portfolio returns net of transaction costs, we assume that investors go short foreign currencies in Portfolio 1 and long foreign currencies in the remaining portfolios. The strategy that is long Portfolio 5 and short Portfolio 1 is referred to as CAR.[16]

Global Imbalance Portfolios. Motivated by the theoretical considerations discussed in Section 2, we construct the global imbalance risk factor as follows: at the end of each period t, we first group currencies into two baskets using the net foreign asset position relative to GDP (nfa), then reorder currencies within each basket using the percentage share of external liabilities denominated in domestic currency (ldc). Hence, we allocate this set of currencies to five portfolios. Portfolio 1 corresponds to creditor countries whose external liabilities are primarily denominated in domestic currency (safest currencies), whereas Portfolio 5 comprises debtor countries whose external liabilities are primarily denominated in foreign currency (riskiest currencies). We refer to these five portfolios as the global imbalance (or IMB) portfolios. We

[16] Lustig, Roussanov, and Verdelhan (2011) study these currency portfolio returns using the first two principal components. The first principal component is proxied using an equally weighted strategy across all portfolios, which is a strategy that borrows in the US money market and invests in foreign money markets. This zero-cost portfolio is called the dollar risk factor, abbreviated to DOL. The second principal component is proxied with a long position in Portfolio 5 and a short position in Portfolio 1, and is equivalent to a carry trade strategy that borrows in the money markets of low yielding currencies and invests in the money markets of high yielding currencies. This high-minus-low portfolio is called the slope factor.

then compute the excess return for each portfolio as an equally weighted average of individual currency excess returns within the portfolio. For the purpose of computing portfolio returns net of transaction costs, we assume that investors go short foreign currencies in Portfolio 1 and long foreign currencies in the remaining portfolios. We construct the global imbalance risk factor as the difference between Portfolio 5 and Portfolio 1. This is equivalent to a high-minus-low strategy that buys the currencies of debtor countries with mainly foreign currency denominated external liabilities and sells the currencies of creditor nations with mainly domestic currency-denominated external liabilities. We refer to this strategy as the global imbalance strategy, and to the global imbalance risk factor as the IMB factor.

Figure 1 clarifies the outcome of our sequential sorting procedure. Note that the procedure does not guarantee monotonicity in both sorting variables (nfa and ldc) because Portfolio 3 contains both low and high ldc countries. However, the corner portfolios contain the intended set of countries: specifically, P_1 contains the extreme 20% of all currencies with high nfa and high ldc (creditor nations with external liabilities mainly in domestic currency) whereas portfolio P_5 contains the top 20% of all currencies with low nfa and low ldc (debtor nations with external liabilities mainly in foreign currency). The global imbalance factor is constructed as the average return on P_5 minus the average return on P_1. We use 5 portfolios (as we do for carry) rather than 6 as we have a limited number of currencies in the sample of developed countries as well as at the beginning of the sample of all countries, and also because we want to have the same number of portfolios in both samples of countries. In the Internet Appendix we show that our core results are qualitatively identical if we form the IMB factor using 4 portfolios for developed countries and 6 portfolios for all countries; see Figure A.1, Tables A.12 and Table A.13 in the Internet Appendix. We also show that, if we use an independent double sort, the risk prices for both nfa and ldc are statistically significant; see Figure A.1, and Tables A.14 and A.15. Finally, we also show robustness using a simple double sort of interest rate differentials and nfa.

Currency Momentum Portfolios. At the end of each period t, we form five portfolios based on exchange rate returns over the previous 3 months. We assign the 20% of all currencies with the lowest lagged exchange rate returns to Portfolio 1, and the 20% of all currencies with the highest lagged exchange rate returns to Portfolio 5. We then compute the excess return for each portfolio as an equally weighted average of the currency excess returns within that portfolio. A strategy that is long in Portfolio 5 (*winner currencies*) and short in Portfolio 1 (*loser currencies*) is then denoted as MOM.[17]

Value Portfolios. At the end of each period t, we form five portfolios based on the lagged five-year real exchange rate return as in Asness, Moskowitz and Pedersen (2013). We assign the 20% of all currencies with the highest lagged real exchange rate return to Portfolio 1, and the 20% of all currencies with the lowest lagged real exchange rate return to Portfolio 5. We then compute the excess return for each portfolio as an equally weighted average of the currency excess returns within that portfolio. A strategy that is long in Portfolio 5 (*undervalued currencies*) and short in Portfolio 1 (*overvalued currencies*) is then denoted as VAL.

Risk Reversal Portfolios. At the end of each period t, we form five portfolios based on out-of-the-money options.[18] We compute for each currency in each time period the risk reversal, which is the implied volatility of the 25-delta call less the implied volatility of the 25-delta put, and assign the 20% of all currencies with the highest risk reversal to Portfolio 1, and the 20% of all currencies with the lowest risk reversal to Portfolio 5. We then compute the excess return for each portfolio as an equally weighted average of the currency excess returns within that portfolio. A strategy that is long in Portfolio 5 (*high-skewness currencies*) and short in Portfolio 1 (*low-skewness currencies*) is then denoted as RR.

4 The Global Imbalance Strategy

This section describes the properties of the net returns from implementing the global imbalance currency strategy and constructing the IMB factor. We also provide a comparison with the net returns from the traditional carry trade strategy (CAR). Specifically, Table A.1 in the Internet Appendix presents descriptive statistics for the five portfolios sorted on interest

[17] Consistent with the results in Menkhoff, Sarno, Schmeling, and Schrimpf (2012b), sorting on lagged exchange rate returns or lagged currency excess returns to form momentum portfolios makes no qualitative difference to our results below. The same is true if we sort on returns with other formation periods in the range from 1 to 12 months.

[18] The implied volatility quotes used for this exercise are an updated sample of the data used by Della Corte, Sarno and Tsiakas (2011), to which we refer the reader for a full description of the data.

rate differentials (forward discount), both for the full sample of countries and the subset of developed countries. Recall that CAR is a long-short strategy that is long in Portfolio 5 (the investment currencies) and short in Portfolio 1 (the funding currencies). Excess returns to Portfolio 1 are adjusted for transactions costs occurring in a short position and excess returns to Portfolio 5 are adjusted for transaction costs occurring in a long position. All excess returns are expressed in percentage per annum.

The CAR results show the properties recorded in several papers in the literature on carry trades. Average excess returns display an increasing pattern when moving from Portfolio 1 to Portfolio 5 for both samples, and the average excess return from a long-short strategy that buys Portfolio 5 and sells Portfolio 1 – the CAR portfolio – is 5.44 (4.67) percent per annum after transaction costs for all (developed) countries. The Sharpe ratio (SR) is equal to 0.65 for all countries, and 0.43 for developed countries.[19]

It is instructive to look at the last three rows in Table A.1, reporting the average interest rate differential (fd), net foreign asset position to GDP (nfa), and share of external liabilities in domestic currency (ldc) across portfolios P_1 to P_5. Clearly the spread in interest rate differentials (which by construction increase monotonically from P_1 to P_5) is very large, about 11 and 6 percent for all countries and developed countries, respectively. The last two rows reveal that there is a similar spread in nfa and ldc, which means that investment (funding) currencies tend to be currencies of net debtor (creditor) countries with a relatively higher (lower) propensity to issue external liabilities in foreign currency.

In Table 1 we present the same summary statistics for the five global imbalance portfolios sorted on nfa and ldc, as well as for the global imbalance factor IMB. The average excess return tends to increase from P_1 (0.92 and 0.67 percent per annum) to P_5 (5.32 and 4.65 percent per annum) for both samples, albeit non-monotonically for the sample of all countries. When we compare SRs, we observe that the global imbalance strategy has a higher Sharpe ratio than the carry trade strategy: 0.68 compared to 0.65 for all countries, and 0.59 compared to 0.43 for developed countries. This comparison suggests that the global imbalance strategy

[19] We also report the maximum drawdown (MDD) and the frequency of currency portfolio switches ($Freq$). The MDD, defined as the maximum cumulative loss from the strategy's peak to the following trough, is large in both samples, reflecting the large-scale unwinding of carry trade positions following the bankruptcy of Lehman Brothers in September 2008. $Freq$ is computed as the ratio between the number of portfolio switches and the total number of currencies at each date. Overall, there is little variation in the composition of these portfolios, which is not surprising given that interest rates are very persistent.

has appealing risk-adjusted returns in its own right, which is perhaps surprising given the information required to update the global imbalance strategy arrives only once a year.[20]

The last three rows in Table 1 report the average fd, nfa, and ldc across portfolios P_1 to P_5. The spread in interest rate differentials is about 7 and 3.5 percent for all countries and developed countries, which is a large spread but far less than the 11 and 6 percent reported for CAR in Table A.1. This suggests that part of the return from the global imbalance strategy is clearly related to CAR (interest rate information), but part of it is driven by a different source of predictability which is in external imbalances but not in interest rate differentials. The last two rows reveal that there is a sizable spread in nfa and ldc, which is monotonic for nfa for both samples of countries examined and is much larger than the corresponding spread for CAR portfolios.

Overall, the currencies of net debtor countries with a relatively higher propensity to issue external liabilities in foreign currency have higher (risk-adjusted) returns than the currencies of net creditor countries with higher propensity to issue liabilities in domestic currency, consistent with GM (2014) and Hypothesis 1 stated in Section 2.

In Figure 2 we present graphical evidence on the relation between carry trade returns and global imbalance risk by grouping carry trade returns into four baskets conditional on the distribution of the IMB factor. The first group comprises the 25 percent of months with the lowest realizations of the IMB factor whereas the last group contains the 25 percent of months with the highest realizations of the IMB factor. We then compute for each group the average carry trade return. Figure 2 shows that average excess returns for the carry trade strategy increase monotonically when moving from low to high global imbalance risk. The carry trade has its best overall performance when global imbalance risk is high and vice versa, suggesting a relation between currency excess returns and global imbalance risk. We now turn to a more rigorous investigation of this similarity using formal asset pricing tests.

[20] Specifically, we construct monthly excess returns but global imbalance portfolios are in practice rebalanced only at the end of each year when new information on nfa and ldc becomes available. In contrast, carry trade portfolios are rebalanced every month as information on forward discounts is available monthly. The impact of this difference is confirmed by the frequency of currency portfolio switches (Freq), which displays less variation for the global imbalance portfolios than the carry trade portfolios.

5 Does Global Imbalance Risk Price Carry Returns?

This section presents cross-sectional asset pricing tests for the five carry portfolios and the global imbalance risk factor, and empirically documents that carry trade returns can be thought of as compensation for time-varying global imbalance risk.

Methodology. We denote the discrete excess returns on portfolio j in period t as RX_t^j. In the absence of arbitrage opportunities, risk-adjusted excess returns have a price of zero and satisfy the following Euler equation:

$$E_t[M_{t+1}RX_{t+1}^j] = 0 \tag{4}$$

with a Stochastic Discount Factor (SDF), M_{t+1} linear in the pricing factors f_{t+1}, given by

$$M_{t+1} = 1 - b'\left(f_{t+1} - \mu\right) \tag{5}$$

where b is the vector of factor loadings, and μ denotes the factor means. This specification implies a beta pricing model where the expected excess return on portfolio j is equal to the factor risk price λ times the risk quantities β^j. The beta pricing model is defined as

$$E[RX^j] = \lambda'\beta^j \tag{6}$$

where the market price of risk $\lambda = \Sigma_f b$ can be obtained via the factor loadings b. $\Sigma_f = E\left[\left(f_t - \mu\right)\left(f_t - \mu\right)'\right]$ is the variance-covariance matrix of the risk factors, and β^j are the regression coefficients of each portfolio's excess return RX_{t+1}^j on the risk factors f_{t+1}.

The factor loadings b entering equation (4) are estimated via the Generalized Method of Moments (GMM) of Hansen (1982). To implement GMM, we use the pricing errors as a set of moments and a prespecified weighting matrix. Since the objective is to test whether the model can explain the cross-section of expected currency excess returns, we only rely on unconditional moments and do not employ instruments other than a constant and a vector of ones. The first-stage estimation (GMM_1) employs an identity weighting matrix. The weighting matrix tells us how much attention to pay to each moment condition. With an identity matrix, GMM attempts to price all currency portfolios equally well. The second-stage estimation (GMM_2) uses an optimal weighting matrix based on a heteroskedasticity and autocorrelation consistent (HAC) estimate of the long-run covariance matrix of the moment conditions. In this case, since currency portfolio returns have different variances and may be correlated, the optimal weighting matrix will attach more weight to linear combinations of moments about which the data are more informative (Cochrane, 2005). The tables report estimates of b and implied λ, and standard errors based on Newey and West (1987) with optimal lag length selection set according to Andrews (1991).[21] The model's performance is then evaluated using the cross-sectional R^2, the square-root of mean-squared errors $RMSE$, the χ^2 test statistics, and the HJ distance measure of Hansen and Jagannathan (1997). The χ^2 test statistic evaluates the null hypothesis that all cross-sectional pricing errors (i.e., the difference between actual and predicted excess returns) are jointly equal to zero. We report asymptotic p-values for the χ^2 test statistics. The HJ distance quantifies the mean-squared distance between the SDF of a proposed model and the set of admissible SDFs. To test whether the HJ distance is equal to zero, we simulate p-values using a weighted sum of χ^2_1-distributed random variables (see Jagannathan and Wang, 1996; Ren and Shimotsu, 2009).

The estimation of the portfolio betas β^j and factor risk price λ in equation (6) is also undertaken using a two-pass ordinary least squares regression following Fama and MacBeth (FMB, 1973). In the first step, we run time-series regressions of portfolio excess returns against a constant and the risk factors, and estimate the betas β^j. In the second step, we run cross-sectional regressions of portfolio returns on the betas, and estimate the factor risk prices λ as averages of all these slope coefficients. Note that in the second stage of FMB regressions we do not add any constant to capture the common over- or under-pricing in the cross section of returns. Our results, however, remain virtually identical when we replace the DOL factor with a constant in the second stage regression. This is because the DOL factor has no cross-sectional relation with currency returns, and it works as a constant that allows for a common mispricing. We report Newey and West (1987) and Shanken (1992) standard errors with optimal lag length selection set according to Andrews (1991).

Risk Factors and Pricing Kernel. The most recent literature on cross-sectional asset pricing in currency markets has considered a two-factor pricing kernel. The first risk factor

[21] We estimate μ and Σ_f using the sample average and the sample covariance matrix of the risk factors, respectively (e.g., Lustig, Roussanov, and Verdelhan, 2011). We also implement a first-stage GMM where μ and Σ_f are jointly estimated with the factor loadings b. In doing so, we account for estimation uncertainty associated with the fact that factor means and the factor covariance matrix have to be estimated (Burnside, 2011; Menkhoff, Sarno, Schmeling and Schrimpf, 2012a). The results remain qualitatively the same.

is typically the expected market excess return, approximated by the average excess return on a portfolio strategy that is long in all foreign currencies with equal weights and short in the domestic currency – essentially the DOL factor. For the second risk factor, the literature has employed several return-based factors such as the slope factor of Lustig, Roussanov, and Verdelhan (2011) or the global volatility factor of Menkhoff, Sarno, Schmeling and Schrimpf (2012a). Regardless of its parsimony and the likely omission of other potential factors, this simple empirical model has delivered important insights on the relation between global risk and expected currency returns. Following this literature, we employ a two-factor SDF with DOL as the first factor. For the second risk factor, we use the IMB factor to further assess the validity of the theoretical prediction in Hypothesis 1 that carry trade returns are linked to external imbalances, and that currencies more exposed to global imbalance risk offer a higher risk premium.

Cross-Sectional Regressions. Panel A of Table 2 presents the cross-sectional asset pricing results. The excess returns to portfolios sorted on forward discounts (RX^j_{CAR} for $j = 1, \ldots, 5$) serve as test assets whereas the dollar factor DOL and the global imbalance factor IMB enter as risk factors. Both test assets and risk factors are adjusted for transactions costs. The SDF is defined as

$$M_{t+1} = 1 - b_{DOL}\left(DOL_{t+1} - \mu_{DOL}\right) - b_{IMB}\left(IMB_{t+1} - \mu_{IMB}\right)$$

where μ_{DOL} and μ_{IMB} denote the factor means. Panel A reports estimates of factor loadings b, the market prices of risk λ, the cross-sectional R^2, the square-root of mean-squared errors $RMSE$, the χ^2 test statistics, and the HJ distance. Newey and West (1987) corrected standard errors with lag length determined according to Andrews (1991) are reported in parentheses, while Shanken corrected standard errors are in brackets. The p-values of the χ^2 test statistics and HJ distance measure are also reported in brackets. The results are reported for all countries (left panel) and developed countries (right panel) using GMM_1, GMM_2, and the FMB approach.

We focus our interest on the sign and the statistical significance of λ_{IMB}, the market price of risk attached to the global imbalance risk factor. We find a positive and significant estimate of λ_{IMB}. The global imbalance risk premium is 7 percent per annum for all countries, and 5 percent per annum for developed countries, and these point estimates are identical for all three estimation methods employed. A positive estimate of the factor price of risk implies higher risk premia for currency portfolios whose returns comove positively with the global imbalance factor, and lower risk premia for currency portfolios exhibiting a negative covariance with the global imbalance factor. The standard errors of the risk prices are approximately equal to 2 percent for all estimation methods. Overall, the risk price is more than two standard deviations from zero, and thus highly statistically significant. We also uncover strong cross-sectional fit with R^2s ranging between 74 and 87 percent for the full sample of countries, and between 88 and 91 percent for the subset of developed countries. Further support in favor of this model comes from the fact that we are unable to reject the null hypotheses that the cross-sectional pricing errors are jointly equal to zero and that the HJ distance is equal to zero.[22]

Time-Series Regressions. In Panel B of Table 2, we report the least squares estimates obtained from running time-series regressions of currency excess returns on a constant and risk factors for each of the five currency portfolios (for $j = 1, \ldots, 5$)

$$RX^j_{CAR,t+1} = \alpha^j + \beta^j_{DOL} DOL_{t+1} + \beta^j_{IMB} IMB_{t+1} + \varepsilon^j_{t+1}.$$

This exercise allows us to clearly identify which of the currency portfolios provide a hedge against global imbalance risk. As expected, the estimate of the betas for the DOL factor are essentially all equal to one as this factor does not capture any of the dispersion in average excess returns across currency portfolios. The estimates of β_{IMB} are positive for currencies with a high forward discount (high interest rate differential relative to the US), and negative for currencies with a low forward discount (low interest rate differential relative to the US). For example, these betas increase monotonically for the sample of all countries from -0.33 for the first portfolio to 0.46 for the last portfolio, and results for developed countries are comparable. Finally, the last column reports the time-series R^2s, which range from 74 to 85 percent for all countries, and from 74 to 86 percent for developed countries.

[22] Lewellen, Nagel and Shanken (2010) show that a strong factor structure in test asset returns can give rise to misleading results in empirical work. If the risk factor has a small (but non-zero) correlation with the 'true' factor, the cross-sectional R^2 could still be high suggesting an impressive model fit. This is particularly problematic in small cross sections, like in our case. For this reason, we carry out asset pricing tests that involve, in addition to the carry trade strategy, other currency strategies as well as equity, bond and commodity strategies. The results, reported later in the paper, corroborate the findings in this section.

These results suggest that carry trade returns are systematically related to global imbalance risk, and carry trade funding currencies are associated with net creditor nations whereas carry trade investment currencies are linked to net debtor nations. The unconditional time-series correlation between carry trade returns and the global imbalance risk factor is 60 percent, indicating a strong, albeit not perfect, positive correlation. This is consistent with Hypothesis 1 in the sense that there are two sources of risk premia driving carry trade returns – interest rate differentials and the evolution of net foreign assets.[23]

Note that we do not argue that these two channels are unrelated. On the contrary, it is well-documented that there is a cross-sectional correlation between interest rates (typically real interest rates) and net foreign asset positions (e.g. Rose, 2010). In Table 3, we present results from a cross-sectional regression of the nominal interest rate differentials used in our study on net foreign assets and the share of liabilities denominated in domestic currency. These results show clearly that net foreign assets enter the regression with a strongly statistically significant coefficient and with the expected sign: higher nfa is associated with lower interest rates. The R^2 is lower than one might expect, however, suggesting that there are likely to be important omitted variables in the regression. Indeed, when we add inflation differentials and output gap differentials to the regression, net foreign asset positions remain strongly significant, but the R^2 increases dramatically, mainly due to inflation differentials.[24] In short, even though the information in global imbalances is related to interest rate differentials, there is likely to be some independent information in global imbalances. Next, we provide some evidence on the value of this information.

Value added of global imbalance information. Taken together, the results reported till now suggest that the global imbalance strategy has creditable excess returns overall, and these returns are highly but imperfectly correlated with the returns from the carry trade. The lack of a perfect correlation is in line with the GM (2014) theory and Hypothesis 1,

[23]Table A.2 in the Internet Appendix reports details on the portfolio composition for both carry trade and global imbalance portfolios. Panel A (Panel B) reports the top six currencies for each of the carry (global imbalance) portfolios. Panel C reports the probability that a given currency enters simultaneously in the same carry and global imbalance portfolio. For corner portfolios, this probability ranges from 38 to 45 percent for all countries, and from 41 to 47 percent for developed countries, consistent with the notion that the carry trade and global imbalance strategies are similar, albeit not identical.

[24]Inflation and the output gap are the core variables in macro models of the short-term interest rate, commonly used in the 'Taylor rule' literature. Note that the regressions in Table 3 are run for 53, rather than 55, countries due to difficulties in obtaining reliable data for the full sample for Greece and Venezuela.

which states that interest rate differentials and the evolution of global imbalances are different sources of currency risk premia. This means that a currency investor would likely gain some diversification benefit from adding the global imbalance strategy to a currency portfolio to enhance risk-adjusted returns.

To better understand the value of global imbalance information for a currency investor, we compute the optimal currency portfolio for an investor who uses three common currency strategies – namely carry, momentum, and value – and adds to this menu of strategies the global imbalance strategy. Specifically, consider a portfolio of N assets with covariance matrix Σ. The global minimum volatility portfolio is the portfolio with the lowest return volatility, and represents the solution to the following optimization problem: min $w'\Sigma w$ subject to the constraint that the weights sum to unity $w'\iota = 1$, where w is the $N \times 1$ vector of portfolio weights on the risky assets, ι is a $N \times 1$ vector of ones, and Σ is the $N \times N$ covariance matrix of the asset returns. The weights of the global minimum volatility portfolios are given by $w = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}$. We compute the optimal weights for both samples of countries examined, and report the results graphically in Figure 3.

The results show that the optimal weight assigned to the global imbalance strategy is actually the highest across all four strategies, equal to 46 and 51 percent for the two sets of countries. The Sharpe ratio of the minimum volatility portfolio for the developed sample, for instance, is quite impressive, at 0.63. However, this number drops to 0.49 if the investor is not given access to the global imbalance strategy, and only employs the other three currency strategies. Similarly for the sample with all countries, the Sharpe ratio equals 0.75 when the global imbalance strategy is included and drops to 0.67 when it is excluded from the menu of currency strategies.[25] These findings confirm that there is independent information in global imbalances about currency risk premia which is not embedded in interest rate differentials.[26]

[25] In a further exercise, we also include in the menu of strategies the risk-reversal currency strategy. In this case the sample is reduced considerably as data for risk reversals only start in 1996. Nevertheless, for both samples of countries the weight on the global imbalance strategy is higher than 40 percent and the highest of all 5 strategies considered.

[26] Another way to assess the value added of the information in net foreign assets beyond interest rate differentials is to double sort on nfa and interest rate differentials. Ideally, one would want to sort on nfa, ldc and interest rate differentials, but our cross-section is simply too small to do this. However, for the sample of all countries, an independent double sort on nfa and interest rate differentials delivers a gross mean return of 2.19 (volatility of 5.05) for net foreign assets and 4.33 (volatility of 6.46) for interest rate differentials. The Sharpe ratios are 0.51 and 0.78. The results for the sample of developed countries are qualitatively identical. In short, these results confirm that there is additional information in the ratio of NFA to GDP that is not

Portfolios based on IMB Betas. We provide evidence of the explanatory power of the IMB factor for currency excess returns from a different viewpoint. We form portfolios based on an individual currency's exposure to global imbalance risk, and investigate whether these portfolios have similar return distributions to portfolios sorted on forward discounts. If global imbalance risk is a priced factor, then currencies sorted according to their exposure to global imbalance risk should yield a cross section of portfolios with a significant spread in average currency returns.

We regress individual currency excess returns at time t on a constant and the global imbalance risk factor using a 36-month rolling window that ends in period $t-1$, and denote this slope coefficient as $\beta^i_{IMB,t}$. This exercise provides currency i exposure to IMB only using information available at time t. We then rank currencies according to $\beta^i_{IMB,t}$ and allocate them to five portfolios at time t. Portfolio 1 contains the currencies with the largest negative exposure to the global imbalance factor (lowest betas), while Portfolio 5 contains the most positively exposed currencies (highest betas). Table 4 summarizes the descriptive statistics for these portfolios. We find that buying currencies with a low beta (i.e., insurance against global imbalance risk) yields a significantly lower return than currencies with a high beta (i.e., high exposure to global imbalance risk). The spread between the last portfolio and the first portfolio is in excess of 5 percent per annum for both sets of countries. Average excess returns generally increase, albeit not always monotonically, when moving from the first to the last portfolio. Moreover, we also find a clear monotonic increase in both average *pre*-formation and *post*-formation betas when moving from Portfolio 1 to Portfolio 5: they line up perfectly well with the cross-section of average excess returns in Table 1. Average *pre*-formation betas vary from -0.22 to 1.35 for all countries, and from -0.94 to 0.67 for developed countries. *Post*-formation betas are calculated by regressing realized excess returns of beta-sorted portfolio j on DOL and IMB. These figures range from -0.30 to 0.31 for all countries, and from -0.57 to 0.59 for developed countries. Overall, these results confirm that global imbalance risk is important for understanding the cross-section of currency excess returns, providing further support to Hypothesis 1.

contained in interest rates.

6 Exchange Rates and Net Foreign Assets in Bad Times

We now turn to testing Hypothesis 2, as stated in Section 2. In essence, the testable prediction from GM (2014) we take to the data is that exchange rates are jointly determined by global imbalances and financiers' risk-bearing capacity so that net external debtors experience a currency depreciation in bad times, which are times of large shocks to risk bearing capacity and risk aversion (Γ is high in the model). In contrast, net external creditors experience a currency appreciation in bad times.

We test this hypothesis in two different ways. First, we estimate a panel regression where we regress monthly exchange rate returns on a set of macro variables, allowing for fixed effects. As right-hand-side variables, we employ the net foreign assets to GDP ratio (nfa) lagged by 12 months, and the interest rate differential lagged by 1 month. In some specifications we also include the share of external liabilities in domestic currency (ldc), and the change in VIX on its own. Importantly, we also allow for an interaction term between nfa as well as the interest rate differential and the change in the VIX index (specification 1-2-3), or the change in VIX times a dummy that is equal to unity when the change in VIX is greater than one standard deviation and is zero otherwise (specifications 4-5-6).[27] The VIX index is commonly used as a proxy for global risk appetite and, in our exercise, we use the change in VIX (in two different formulations) to proxy what GM (2014) term 'global risk aversion shocks' in reference to Γ, i.e. shocks to the willingness of financiers to absorb exchange rate risk.[28]

The key variable of interest in these regressions is the interaction term between nfa and the VIX change. Given our variable definitions, Hypothesis 2 requires a positive coefficient on this variable, which would imply that at times when global risk aversion increases (as proxied

[27] We also add a constant, and the lagged exchange rate return as a control variable.

[28] Given the reference to shocks, we use the change in VIX rather than the VIX in level. It is well known that the change in VIX has negligible serial correlation, whereas the VIX in level is very persistent (e.g. Ang, Hodrick, Xing and Zhang, 2006). We use the change in VIX contemporaneously in these regressions in order to capture the effect of the shock on exchange rate returns predicted by Hypothesis 2, which states that net debtor countries' currencies depreciate on impact when global risk aversion increases. Presumably the impact of risk aversion shocks on exchange rate returns goes beyond this contemporaneous effect though, and indeed we find that results are similar when using 1 or 2 lags of the change in VIX. Finally, an alternative interpretation of Γ might be that it captures (changes in) the amount of capital available in financial markets to bear risk. In this case one would expect that returns of the carry trade strategy decline as the amount of capital increases, and in fact there is evidence in the literature that this is the case (e.g. Jylha and Suominen, 2011; Barroso and Santa-Clara, 2014). However, our interpretation of Γ is, much like GM (2014), that it reflects shocks to global risk aversion and hence the change in VIX seems a reasonable proxy.

by the VIX change) countries with larger net foreign asset positions to GDP experience a currency appreciation, whereas the currencies of countries with larger net debtor positions depreciate. The results, reported in Table 5, indicate that this is the case as the interaction term is positive and strongly statistically significant in all regression specifications, even when controlling for the interest rate differential, the change in VIX and the other control variables described above. It is instructive to note that the change in VIX also enters significantly and with the expected sign, meaning that increases in global risk aversion are associated with appreciation of the US dollar.[29]

For completeness, we also run similar panel regressions for excess returns rather than exchange rate returns, reported in Table A.5 of the Internet Appendix.[30] The results corroborate the results obtained for exchange rate returns but also provide one more interesting finding, namely that the share of foreign liabilities issued in domestic currency is now statistically significant, whereas it was not in Table 5. This indicates that this variable is likely to be related to currency excess returns (carry trade returns) via interest rate differentials rather than exchange rate returns.

The second test of Hypothesis 2 we carry out involves estimating time-series regressions of the returns from the five global imbalance portfolios on the change in VIX. Remember that the long (short) portfolio comprises the currencies with highest (lowest) net foreign liabilities and a higher (lower) propensity to issue external liabilities in foreign currency. Hence Hypothesis 2 requires that the return on the long portfolio is negatively related to global risk aversion shocks, proxied by the change in VIX; by contrast the return on the short portfolio should be positively related to the change in VIX. The results from estimating these regressions are reported in Table 6 (both for excess returns and just the spot exchange rate component), and show a monotonic decline in the coefficients on the change in VIX, as one would expect. However, the coefficients for Portfolio P_1 (the short portfolio) and P_2 are not statistically different from zero, implying that the currencies of net creditors do not respond to global risk

[29] Finally, it is also interesting to note that the net foreign asset position to GDP ratio (not interacted with the VIX) is either not statistically different from zero or, in two cases, enters the regression with a negative coefficient, implying that in normal times net debtor currencies experience appreciation. This seems plausible in the GM (2014) framework as in normal times there is strong demand for net debtor currencies by investors interested in capturing the risk premium offered by these currencies.

[30] The main difference is that we do not condition on the interest rate differential in these regressions as the interest rate differential is on the left-hand-side of the regression (in the currency excess return).

aversion shocks. The coefficients for portfolios P_3, P_4 and P_5 are negative and statistically significant, and they are largest for P_5, implying that the currencies in the long portfolio of the global imbalance strategy depreciate the most in bad times. Overall, the currencies issued by the extreme net debtor countries with the highest propensity to issue liabilities in foreign currency depreciate sharply in bad times relative to the currencies issued by the extreme net creditor countries with the lowest propensity to issue liabilities in foreign currency. This result constitutes further supportive evidence for Hypothesis 2.

7 Further Analysis

In this section, we present a battery of additional exercises that further refine and corroborate the results reported earlier.

Asset Pricing Tests on Other Cross-Sections of Returns. In our empirical asset pricing analysis, the pricing power of the global imbalance factor was tested using two cross-sections: the carry trade cross section, and the global imbalance cross section. While this is a very direct way to test the predictions of GM (2014), the cross-sectional regressions are based on a small number of observations (5 data points). Moreover, the two cross-sections are highly related in light of our empirical results, further reducing the number of test assets. Therefore, we test the pricing power of the global imbalance factor on a larger number of test assets.

First, we consider other currency strategies, both separately and jointly. In Table 7, we report estimates of factor loadings and risk prices using first-stage GMM for the carry trade (already reported earlier in Table 2), the global imbalance strategy, currency momentum, currency value, and risk reversal strategies. For all strategies the sample spans from 1983 to 2014, except for the risk reversal strategy, where the sample begins in 1996. In these individual tests we are still using only five data points in the cross-sectional regressions, but it is more likely that at least some of these cross-sections of returns do not have the same factor structure. The results indicate that global imbalance risk prices fairly well four of these cross-sections (with the price of risk being estimated in the range between 0.04 and 0.07, and a high R^2), with the exception being currency momentum – for which the price of risk is not significant, the R^2 is miniscule and the HJ rejects the null of zero pricing errors. This is not surprising given the well-documented difficulty in pricing momentum portfolios (e.g. Menkhoff, Sarno, Schmeling and Schrimpf, 2012b). We then increase test power by pooling these cross-sections of returns to form a cross-section of 20 currency portfolio returns (excluding the risk reversal portfolios) since 1983, and 25 currency portfolio returns since 1996 (including also the risk reversal portfolios). The results, reported in the last two columns of Table 7, indicate that global imbalance risk prices both these cross-sections of currency returns, with a reasonable R^2 (0.53 and 0.65 respectively), and insignificant HJ tests.

Second, in Table 8, we further expand the set of assets by adding to the cross-section of 20- or 25- currency returns also 25 equity portfolio returns (sorted either by size and book-to-market, or by size and momentum), 5 international bond portfolio returns, and 7 commodity portfolio returns.[31] Starting from Panels A and B, which report the results for the cross-sections of currency and equity portfolios, we find that the global imbalance risk factor is priced in this large cross-section of returns, even controlling for Fama-French factors.[32] The IMB risk price estimate is between 0.07 and 0.09 and highly statistically significant, the R^2 is in the range between 0.78 and 0.86, and the HJ test is statistically insignificant. Similarly, Panel C reports the results when we augment the currency cross-sections with 5 international bond portfolio returns, and we control for an international bond factor. In Panel D, we add to the currency cross-sections 7 commodity portfolio returns, and in this case we control for a commodity factor. Both in Panels C and D, the global imbalance risk factor is priced with comparable estimates of the risk price, and we find insignificant HJ tests.

Overall, these results suggest that global imbalance risk is priced broadly across currency strategies, and some of the most common equity, international bonds and commodity strategies.

Asset Pricing with a Constraint on the Price of Risk. The asset pricing exercise reported in Table 2 suggested that the IMB factor prices the cross-section of currency excess returns sorted on interest rate differentials, i.e. carry trades. However, IMB is a tradable

[31] For the equity cross-section, we collect the 25 global portfolios sorted on size and book-to-market, and size and momentum from Kenneth French's website. We use equally weighted portfolios that do not include the US in order to be consistent with currency portfolios that are dollar-neutral by construction. These portfolios include 22 countries. For the commodity portfolios, we take the seven commodity portfolios from Yang (2013). For international bonds, we sort bonds of different maturities (1-3y, 3-5y, 5-7y, 7-10y, >10y) for 19 countries into five portfolios depending on their redemption yield. We use total return indices denominated in US dollars from Datastream.

[32] We use 3 Fama-French factors in each regression. In Panel A, they are market, SMB and HML, while in Panel B HML is replaced by WML (momentum). Overall, we estimate 5 risk prices in both Panels A and B.

risk factor and thus its price of risk must equal its expected return. This means that the price of global imbalance risk cannot be a free parameter in estimation. To address this issue, we follow the suggestion of Lewellen, Nagel and Shanken (2010) and include the global imbalance factor as one of the test assets, alongside the interest rate differential-sorted portfolios, which effectively means we constrain the price of risk for IMB to be equal to the mean return of the traded global imbalance portfolio. The results, reported in Table A.3 in the Internet Appendix, provide evidence that the performance of the model is slightly improved and that the estimates of the price of risk are statistically identical to the returns from the global imbalance strategy reported in Table 1.[33]

This result is comforting since it implies that our factor price of risk makes sense economically, that the factor prices itself, and is thus arbitrage-free. We add the global imbalance risk factor to the test assets for *all* the asset pricing tests that follow.[34]

Individual Currencies. Ang, Liu and Schwarz (2010) argue that forming portfolios may potentially destroy information by shrinking the dispersion of betas. In Table A.4 we deal with this concern and present cross-sectional asset pricing tests with individual currency excess returns as test assets. Since the set of currencies is now unbalanced, we only report estimates of market prices of risk obtained via FMB regressions. Also, since country-level excess returns, especially for currencies with limited trading activity, may be contaminated by outliers, least square estimates can be severely distorted and fail to deliver unbiased estimates. We deal with this problem by using the least absolute deviation (LAD) estimator which is robust to thick-tailed errors and is not sensitive to atypical data points (Bassett and Koenker,

[33] Alternatively, if we calibrate the risk price to the mean returns reported in Table 1 and estimate a single-factor model, we obtain results which are virtually identical to the ones in Table A.3.

[34] Recall that data for ldc are only available since 1990 and we backfill the data to 1983 by keeping that constant at their 1990 values for all countries. One may be concerned about the impact of this choice, and therefore we check the robustness of this decision. Suppose we start in Jan 1991 (given that ldc is available at the end of Dec 1990) and stop the sample in December 2013 (we keep using forward filling up to 1-year ahead). Using the above sample, we construct the global imbalance portfolios and the IMB factor, and also carry out asset pricing tests on the carry portfolios while imposing the restriction on the price of IMB risk. For the developed sample, we find that the IMB factor has a mean return of 3.91 (t-stat = 2.54), and $SR = 0.53$. The one-step GMM estimate of the price of risk is 4 percent (t-stat = 2.53). For the sample of all countries, the mean return is 4.73 (t-stat = 2.84), and $SR = 0.69$, with the point estimate of the price of IMB risk from one-step GMM equal to 5.5 percent (t-stat = 3.31). For both samples of countries, the IMB factor prices well the test assets, and we cannot reject the null of zero pricing errors with large p-values. In short, the results are qualitatively identical when using a sample period that does not require backfilling the ldc data prior to 1990.

1978; Koenker and Bassett, 1982). In short, we use the FMB procedure with robust regressions in the first and second step to account for outliers in individual currency excess returns. We report bootstrapped standard errors in parentheses.[35]

In Panel A the test assets are excess returns constructed as long positions in foreign currencies irrespective of the level of interest rates. Note that these individual currency excess returns are not adjusted for transaction costs as ex-ante we ignore whether an investor should buy or sell the foreign currency. We refer to these excess returns as unconditional excess returns. The pricing kernel includes the DOL and IMB as risk factors. The market price of global imbalance risk is positive and statistically significant, and the estimate is very close to the estimates obtained in Table 2 (0.05 and 0.06 for our two samples of countries). The cross-sectional R^2 is reasonably high, 34 percent for all countries and 64 percent for developed countries, but is of course lower than the R^2 for portfolio returns. This is expected as individual excess returns are far more noisy than portfolio returns.

In Panel B we use as test assets excess returns managed on the basis of interest rate differentials: the US investor buys the foreign currency and sells the US dollar when the forward discount is positive (i.e., the foreign currency interest rate is higher than the US interest rate), and vice versa. Results remain largely comparable to the previous panel. In short, these results suggest that the global imbalance risk factor does a reasonably good job at pricing the cross-section of individual currency excess returns.

Alternative base currencies. We depart from the base scenario of a US-based investor and run calculations using alternative base currencies, taking the viewpoint of a British, Japanese, Euro-based and Swiss investor. The results indicate that, in each case, the global imbalance portfolio has similar return characteristics to the ones reported in Table 1 (see Table A.6 in the Internet Appendix), and the global imbalance risk factor prices the cross-section of carry trade returns (Table A.7 in the Internet Appendix).

This is comforting since it makes clear that the US is not playing a key role in driving

[35] To calculate bootstrapped standard errors, we simulate $y_{i,t} = \alpha_i + \beta_i f_t + \varepsilon_{i,t}$ and $f_t = \mu + \sum_{i=1}^{p} A_i f_{t-i} + u_t$, where $y_{i,t}$ is the excess return on the i-th currency, α_i is the constant, β_i is the vector of factor loadings, f_t denotes the risk factors following a p-order VAR process, $\varepsilon_{i,t}$ are idiosyncratic residuals, and $u_t \sim N(0, \Sigma)$. We estimate this system, and use the parameter estimates to generate $1,000$ time-series by jointly resampling $\varepsilon_{i,t}$ and u_t. Since the panel is unbalanced, we carefully resample the same dates across all individual currencies, and then remove the missing values before running FMB regressions.

our results, which are qualitatively identical regardless of whether the currency portfolios are dollar-neutral or not. Indeed, the US may be seen as an interesting exception to our story in this paper, especially during the recent crisis, because it is one of the largest external debtors in the world and yet it appreciated strongly during the crisis when instead the carry trade experienced a large drawdown. Part of the explanation may be that the US, which has a substantial currency mismatch on its balance sheet, borrows in domestic currency and is generally considered a safe reserve currency (see Maggiori, 2013 for a theoretical discussion on this 'reserve currency paradox'). In any event, we find it comforting that our results hold when using four alternative base currencies.

Removing Illiquid Currencies. Tables A.8 and A.9 display the results from building the global imbalance portfolio (and running asset pricing tests) using a sample where currencies with limited liquidity are removed. Specifically, using the latest *BIS Triennial Survey*, we select the most liquid currencies and name this sample 'developed and emerging countries,' which is an intermediate sample (in terms of size) between the two samples analyzed in the paper till now.[36] We hypothesize that while forward rates may be available for a large number of currencies, there would have been low liquidity in many of them. Additionally, the imposition of capital controls in a number of the emerging market nations might have made it impossible to engage in a carry trade strategy at some points in time. If this is the case, we would anticipate that the asset pricing results for a limited subset of the most liquid currencies would show an improvement over and above the full sample.

Table A.8 shows that there are no qualitative changes to the properties of the global imbalance strategy as reported in Table 1, but the performance of the strategy is enhanced, reaching a Sharpe ratio of 0.95. Table A.9 reports cross-sectional asset pricing results, which are highly comparable to the results in Table 2.

Portfolios Sorted on Real Interest Rate Differentials. We also find that the global imbalance risk factor prices portfolios of currencies sorted on real (as opposed to nominal) interest rate differential relative to the US; the results are reported in Tables A.10 and A.11 of the Internet Appendix. At time t, we allocate currencies to five portfolios according to

[36] This is the set of currencies employed by Deutsche Bank for its global carry trade (Global Currency Harvest) strategy.

their inflation-adjusted forward discount $fd_t - E_t(\pi^*_{t+1} - \pi_{t+1})$, where π^*_{t+1} and π_{t+1} denote the one-month foreign and domestic inflation rates at time $t+1$, respectively, and E_t is the conditional expectations operator given information at time t. This is equivalent to sorting currencies according to their real, rather than nominal, interest rate differential. Since π^*_{t+1} and π_{t+1} are not observed at time t, we construct inflation forecasts by simply using current inflation, that is we set $E_t(\pi^*_{t+1} - \pi_{t+1}) = \pi^*_t - \pi_t$.[37] Currencies with the lowest real interest rate differential are assigned to Portfolio 1, whereas currencies with the highest real interest rate differential are assigned to Portfolio 5.

Table A.10 reports descriptive statistics for the portfolios described above, while Table A.11 reports asset pricing tests where we use the same DOL and IMB risk factors as in the core analysis. The global imbalance risk premium remains positive and statistically different from zero, with estimates comparable to the ones reported in earlier tests. The cross-sectional R^2 remains high, and we cannot reject the null hypothesis that the pricing errors are zero as well as the null hypothesis that the HJ distance is zero. Overall, these results are largely comparable to our core findings in Table 2. We confirm higher risk premia for currency portfolios whose returns comove positively with the global imbalance factor, and lower risk premia for currency portfolios exhibiting a negative covariance with the global imbalance factor.

Independent Double Sort. Our global imbalance factor is constructed by sequentially sorting currencies first with respect to the net foreign asset positions to GDP (nfa), and then with respect to the percentage share of foreign liabilities in domestic currency (ldc). A natural question to ask is whether the information in the global imbalance factor is driven by nfa or ldc, or both. To address this point, we construct a factor that captures only the information arising from nfa and a factor that summarizes only the signal coming from ldc. We will refer to these factors as NFA and LDC, respectively. Figure A.1 in the Internet Appendix reports a visual description of how we construct these factors. We use 6 portfolios, except for the subset of developed countries where we are restricted to use only 4 portfolios. At the end of each month, currencies are first sorted in two baskets using the net foreign asset positions to GDP (nfa), and then in 3 baskets using the percentage share of foreign liabilities in domestic currency (ldc). The NFA factor is computed as the average return on the low nfa portfolios

[37] This assumption is empirically motivated since inflation is a very persistent process and current inflation is highly correlated with future inflation at the monthly frequency.

(P_3, P_4 and P_5) minus the average return on the high nfa portfolios (P_1, P_2 and P_3) whereas the LDC factor is computed as the average return on the low ldc portfolios (P_3, and P_6) minus the average return on the high ldc portfolios (P_1, and P_4). We use a similar procedure for the developed countries sample.

We report the summary statistics of these portfolios' excess returns along with the NFA and LDC factors in Table A.14 in the Internet Appendix. The excess return per unit of volatility risk on both factors tends to be comparable when we inspect the subset of developed countries (the SR equals 0.34 for NFA and 0.38 for LDC). When we move away from developed countries, the LDC factor tends to outperform the NFA factor: the LDC (NFA) factor displays an SR of 0.78 (0.59) when we add the most liquid emerging market currencies to the set of developed countries, and an SR of 0.71 (0.28) when we consider the full set of currencies.[38]

Table A.15 in the Internet Appendix presents asset pricing tests based on a linear three-factor model that comprise the DOL, NFA and LDC factors. As test assets, we continue to use the five carry trade portfolios used in the core analysis. Panel A reports cross-sectional tests. The market price of risk is positive and statistically significant for both NFA and LDC, regardless of the methodology used, when we focus on all countries and developed and emerging countries. Results are mixed for the subset of developed countries: the market price of risk is always positive and statistically significant for NFA, and positive and significant for LDC only for GMM_1 and FMB. In addition to relying on standard asset pricing estimates, we run a simple model comparison to understand whether one set of factors drives out another in the spirit of Cochrane (2005). Specifically, we compare the three-factor model above (unrestricted model) to a two-factor model that contains either DOL and NFA (restricted 1) or DOL and LDC (restricted 2). We compare the unrestricted model to each restricted model using a simple χ^2 difference test (*D-test*) that uses the same weighting matrix – the one resulting from the unrestricted model. We report in brackets the p-values for the null hypothesis that either restricted 1 or restricted 2 are correct when compared to the unrestricted model. We reject the null in each case for all countries and developed and emerging countries, thus suggesting

[38] Note that here we report summary statistics for portfolios gross of transaction costs. Otherwise, we would need to report both long and short net positions for the same portfolio as NFA and LDC require different combinations of long and short portfolios. In Table A.15, however, we use risk factors net of transaction costs in order to make results comparable with our core analysis.

that both NFA and LDC are important for these sets of countries. For developed countries, we reject the null hypothesis for restricted 1 (with a *p-value* of 0.04) but fail to reject the null hypothesis for restricted 2 (with a *p-value* of 0.12). This suggests that NFA may be more important than LDC for major economies.

Panel B of Table A.15 reports least square estimates obtained from running time-series regressions. Results show that NFA and LDC are both important for all countries and developed and emerging countries. For developed countries, NFA tends to be more important, although LDC remains statistically significant for low-yielding currencies.

Finally, we appreciate that the use of a three-factor model on such a small cross-section presents small sample problems, and for this reason we carry out the asset pricing test using the 20- and 25-currency portfolios used earlier as test assets. Estimation of the three-factor model with NFA and LDC confirms that in this larger cross-section of currency returns both NFA and LDC are priced (results are qualitatively identical to the ones in Table A.15).

Overall, the evidence in this section confirms that both sorting variables used in our global imbalance strategy and for the purpose of constructing the IMB factor contribute to the price of global imbalance risk, and reflect slightly different aspects of this source of risk. The sorting procedure used in the core analysis allows us to combine the information in nfa and ldc in a simple fashion, and to construct a single risk factor that captures these two different aspects of the evolution of global imbalances across countries. In fact, we also note that the pricing errors from the two-factor model used in the core analysis (e.g. Table 2) are lower than the pricing errors from the three-factor model used in Table A.15.

8 Conclusions

The large and sudden depreciation of high-interest currencies in the aftermath of the Lehman Brothers' collapse has revived interest in the risk-return profile of the carry trade, a popular strategy that exploits interest rate differentials across countries. If high-interest rate currencies deliver low returns when consumption is low, then currency excess returns simply compensate investors for higher risk exposure and carry trade returns reflect time-varying risk premia. While the recent empirical literature has established that there is systematic risk in carry trades, it is silent about the economic determinants underlying currency premia.

This paper tackles exactly this issue by shedding empirical light on the *macroeconomic* forces driving currency risk premia. Motivated by the theoretical insights of Gabaix and Maggiori (2014), we show that sorting currencies on net foreign asset positions and a country's propensity to issue external liabilities in domestic currency generates a large spread in returns. In fact, a risk factor that captures exposure to global imbalances and the currency denomination of external liabilities explains the bulk of currency excess returns in a standard asset pricing model. The economic intuition for this risk factor is simply that net debtor countries offer a currency risk premium to compensate investors willing to finance negative external imbalances. This means that carry trade returns are actually driven by two different, albeit related, sources of risk premia: the first is related to the familiar interest rate differentials, and the second is related to the evolution of net foreign asset positions and their currency of denomination.

We also show that, when global risk aversion spikes, net debtor nations experience a sharp currency depreciation, corroborating the notion that carry trade investors take on global imbalance risk. Moreover, global imbalance risk appears to be priced pervasively, in addition to carry portfolios, in other cross-sections of currency returns as well as in several cross-sections of returns in other major asset markets.

Overall, we provide empirical support for the existence of a meaningful link between exchange rate returns and macroeconomic fluctuations, uncovering a fundamental and theoretically motivated source of risk driving currency returns.

Table 1. Descriptive Statistics: Global Imbalance Portfolios

The table presents descriptive statistics of currency portfolios sorted on time $t-1$ net foreign asset position to gross domestic product (nfa), and the share of foreign liabilities in domestic currency (ldc). The first portfolio (P_1) contains the top 20% of all currencies with high nfa and high ldc (creditor nations with external liabilities mainly in domestic currency) whereas the last portfolio (P_5) contains the top 20% of all currencies with low nfa and low ldc (debtor nations with external liabilities mainly in foreign currency). IMB is a long-short strategy that buys P_5 and sells P_1. The table also reports the first order autocorrelation coefficient (ac_1), the annualized Sharpe ratio (SR), the maximum drawdown (mdd), the frequency of portfolio switches ($freq$), the average forward discount or interest rate differential relative to the US (fd), the average nfa, and the average ldc. t-statistics based on Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in brackets. Excess returns are expressed in percentage per annum and adjusted for transaction costs. The portfolio are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction, and Figure 1 for a detailed description of the portfolio construction.

	P_1	P_2	P_3	P_4	P_5	IMB	P_1	P_2	P_3	P_4	P_5	IMB
	All Countries						*Developed Countries*					
mean	0.92	3.51	1.40	3.57	5.32	4.40	0.67	2.45	3.06	3.46	4.65	3.98
t-stat	[0.60]	[2.18]	[1.10]	[2.39]	[2.73]	[3.51]	[0.37]	[1.31]	[1.77]	[2.00]	[2.38]	[3.26]
med	1.20	2.69	3.52	4.24	6.79	4.94	1.24	2.73	3.66	3.87	6.90	5.27
sdev	7.80	8.71	6.52	7.92	10.05	6.43	9.90	10.25	9.33	9.06	10.29	6.76
skew	−0.16	−0.03	−0.86	−0.48	−0.27	0.17	0.05	−0.07	−0.26	−0.16	−0.28	−0.53
kurt	3.56	3.95	6.42	5.49	4.36	6.17	3.56	3.27	3.90	6.08	3.66	5.17
ac_1	0.08	0.05	0.09	0.06	0.08	0.09	0.06	0.02	0.05	0.08	0.06	−0.01
SR	0.12	0.40	0.22	0.45	0.53	0.68	0.07	0.24	0.33	0.38	0.45	0.59
mdd	0.46	0.29	0.33	0.26	0.30	0.20	0.54	0.36	0.34	0.32	0.31	0.26
$freq$	0.03	0.04	0.04	0.04	0.03		0.02	0.02	0.02	0.02	0.03	
fd	−0.54	1.20	2.02	3.50	6.80		−1.32	−0.76	1.81	2.15	2.23	
nfa	0.43	0.14	0.10	−0.46	−0.56		0.41	0.31	0.04	−0.37	−0.37	
ldc	0.63	0.47	0.44	0.47	0.28		0.61	0.46	0.48	0.49	0.34	

Table 2. Asset Pricing Tests: Global Imbalance Risk

The table presents cross-sectional asset pricing results for the linear factor model based on the dollar (DOL) and the global imbalance (IMB) risk factor. The test assets are excess returns to five carry trade portfolios sorted on the one-month forward discounts. IMB is a long-short strategy that buys the currency (top 20%) of debtor nations with external liabilities mainly in foreign currency, and sells the currencies (top 20%) of creditor nations with external liabilities mainly in domestic currency. *Panel A* reports GMM (first and second-stage) and Fama-MacBeth (FMB) estimates of the factor loadings b, the market price of risk λ, and the cross-sectional R^2. Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in parentheses whereas Shanken (1992) standard errors are reported in brackets. χ^2 denotes the test statistics (with p-value in brackets) for the null hypothesis that all pricing errors are jointly zero. HJ refers to the Hansen and Jagannathan (1997) distance (with simulated p-value in brackets) for the null hypothesis that the HJ distance is equal to zero. *Panel B* reports least-squares estimates of time series regressions with Newey and West (1987) and Andrews (1991) standard errors in parentheses. Excess returns are in annual terms and adjusted for transaction costs. The portfolios are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction, and Figure 1 for a detailed description of the global imbalance risk factor.

Panel A: Factor Prices

	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ
	All Countries								*Developed Countries*							
GMM_1	−0.07	1.53	0.02	0.07	0.87	1.70%	4.92	0.14	0.16	0.85	0.02	0.05	0.91	1.12%	1.15	0.06
	(0.29)	(0.54)	(0.01)	(0.02)			[0.18]	[0.16]	(0.22)	(0.47)	(0.02)	(0.02)			[0.77]	[0.80]
GMM_2	−0.02	1.48	0.02	0.07	0.74	1.76%	4.89		0.19	0.97	0.02	0.05	0.88	1.13%	1.06	
	(0.29)	(0.54)	(0.01)	(0.02)			[0.19]		(0.21)	(0.46)	(0.02)	(0.02)			[0.79]	
FMB	−0.07	1.52	0.02	0.07	0.87	1.70%	4.92		0.16	0.85	0.02	0.05	0.91	1.12%	1.16	
	(0.25)	(0.46)	(0.01)	(0.02)			[0.18]		(0.19)	(0.39)	(0.02)	(0.02)			[0.77]	
	(0.24)	(0.44)	(0.01)	(0.02)					[0.18]	[0.37]	[0.02]	[0.02]				

Panel B: Factor Betas

	α	β_{DOL}	β_{IMB}	R^2	α	β_{DOL}	β_{IMB}	R^2
P_1	−0.01	1.00	−0.33	0.80	0.01	0.97	−0.46	0.74
	(0.01)	(0.05)	(0.04)		(0.01)	(0.04)	(0.06)	
P_2	−0.02	0.99	−0.17	0.83	−0.01	1.01	−0.16	0.82
	(0.01)	(0.04)	(0.03)		(0.01)	(0.04)	(0.04)	
P_3	0.01	1.05	−0.10	0.85	−0.01	0.97	0.01	0.86
	(0.01)	(0.03)	(0.02)		(0.01)	(0.03)	(0.03)	
P_4	−0.01	1.04	0.12	0.82	−0.01	0.97	0.14	0.83
	(0.01)	(0.04)	(0.05)		(0.01)	(0.03)	(0.04)	
P_5	0.01	0.90	0.46	0.74	0.01	1.02	0.52	0.77
	(0.01)	(0.05)	(0.08)		(0.01)	(0.04)	(0.06)	

Table 3. Forward Discounts and Global Imbalances

The table presents results from cross-sectional regressions of average forward discount (or interest rate differential relative to the US) on the average (i) net foreign asset position to gross domestic product (nfa), (ii) share of foreign liabilities in domestic currency (ldc), (iii) inflation differential relative to the US, (iv) output gap, and (v) a constant. White (1980) corrected standard errors are reported in parentheses. See Section 3 for a detailed description of data sources and data construction.

	Dependent variable: forward discount			
	(1)	(2)	(3)	(4)
nfa	−0.141	−0.075	−0.127	−0.072
	(0.036)	(0.017)	(0.037)	(0.017)
ldc	−0.221	0.089	−0.302	0.064
	(0.465)	(0.169)	(0.482)	(0.169)
inflation differential		0.969		0.959
		(0.043)		(0.045)
output gap			0.074	0.021
			(0.033)	(0.009)
constant	0.298	−0.209	0.349	−0.190
	(0.253)	(0.082)	(0.265)	(0.083)
Adjusted R^2	0.05	0.86	0.07	0.86

Table 4. Portfolios Sorted on Betas

The table presents descriptive statistics of β-sorted currency portfolios. Each β is obtained by regressing individual currency excess returns on the global imbalance risk factor using a 36-month moving window that ends in period $t-1$. The first portfolio (P_1) contains the top 20% of all currencies with the lowest betas whereas the last portfolio (P_5) contains the top 20% of all currencies with the highest betas. H/L denotes a long-short strategy that buys P_5 and sells P_1. Excess returns are expressed in percentage per annum. The table also reports the first order autocorrelation coefficient (ac_1), the annualized Sharpe ratio (SR), the maximum drawdown (mdd), the frequency of portfolio switches ($freq$), the average net foreign asset position to gross domestic product (nfa), the share of foreign liabilities in domestic currency (ldc), the *pre-* and *post*-formation forward discount or interest rate differential relative to the US (fd), the *pre*-formation βs (with standard deviations in parentheses) and the *post*-formation βs (with standard errors in parentheses). t-statistics based on Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in brackets. The sample runs from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction.

	P_1	P_2	P_3	P_4	P_5	H/L	P_1	P_2	P_3	P_4	P_5	H/L
	All Countries						*Developed Countries*					
mean	−0.54	2.18	3.85	3.10	4.67	5.21	−1.02	3.61	2.47	2.33	4.92	5.93
t-stat	[−0.38]	[1.49]	[2.39]	[1.59]	[2.38]	[2.83]	[−0.51]	[1.80]	[1.31]	[1.40]	[2.33]	[2.76]
med	−0.29	2.47	3.53	4.53	4.27	5.79	−1.23	3.18	5.25	3.51	6.77	6.94
sdev	6.62	7.62	8.18	9.10	9.61	9.11	9.74	10.29	9.25	8.51	10.59	10.79
skew	0.17	0.13	−0.59	−0.42	−0.43	−0.30	0.01	−0.06	−0.36	−0.26	−0.32	−0.31
kurt	3.90	3.98	5.56	4.18	4.77	3.55	3.65	3.98	3.71	4.07	5.28	4.37
ac_1	0.13	0.03	0.09	0.15	0.12	0.11	0.11	0.05	0.11	0.06	0.09	0.07
SR	−0.08	0.29	0.47	0.34	0.49	0.57	−0.10	0.35	0.27	0.27	0.46	0.55
mdd	0.49	0.35	0.18	0.30	0.26	0.20	0.65	0.36	0.30	0.27	0.33	0.42
freq	0.10	0.14	0.15	0.14	0.07	0.17	0.10	0.15	0.13	0.10	0.04	0.14
nfa	0.45	−0.03	−0.02	−0.11	−0.41		0.47	0.28	−0.04	−0.18	−0.49	
ldc	0.53	0.50	0.48	0.46	0.41		0.57	0.49	0.47	0.46	0.46	
pre-fd	−0.36	0.55	2.13	2.60	4.30		−1.53	−0.02	0.89	1.45	3.04	
post-fd	−0.35	0.56	2.11	2.59	4.24		−1.51	0.00	0.84	1.43	3.04	
pre-β	−0.22	0.14	0.51	0.78	1.35		−0.94	−0.50	−0.28	0.05	0.67	
	(0.35)	(0.47)	(0.66)	(0.76)	(0.76)		(0.97)	(0.96)	(0.88)	(0.67)	(0.57)	
post-β	−0.30	−0.31	−0.09	0.07	0.31		−0.57	−0.19	0.03	0.12	0.59	
	(0.05)	(0.04)	(0.04)	(0.05)	(0.06)		(0.05)	(0.03)	(0.04)	(0.04)	(0.05)	

Table 5. Determinants of Spot Exchange Rate Returns

The table presents results from fixed-effects panel regressions. We use discrete exchange rate returns at monthly frequency as dependent variable. Exchange rates are defined as units of US dollars per unit of foreign currency such that a positive return denotes a foreign currency appreciation. The set of independent variables includes the net foreign asset position to gross domestic product (nfa), the share of foreign liabilities in domestic currency (ldc), the forward discount or interest rate differential relative to the US (fd), the monthly change in the VIX index (ΔVIX), and a dummy variable that equals one if ΔVIX is greater than one standard deviation as estimated across the entire sample, and zero otherwise (ΔVIX *dummy*). Robust standard errors are clustered at country level and reported in parentheses. The superscripts a, b and c denote statistical significance at 10%, 5% and 1% level, respectively. The sample runs from January 1986 to June 2014. See Section 3 for a detailed description of data sources and data construction.

	Dependent variable: nominal exchange rate returns					
	(1)	(2)	(3)	(4)	(5)	(6)
nfa (lagged 12 months)	−0.043	−0.040	−0.015	-0.158^{b}	-0.159^{b}	−0.113
	(0.069)	(0.072)	(0.076)	(0.071)	(0.073)	(0.071)
ΔVIX	-0.143^{c}	-0.143^{c}	-0.126^{c}			
	(0.017)	(0.017)	(0.019)			
$\Delta VIX \times nfa$ (lagged 12 months)	0.069^{c}	0.069^{c}	0.058^{c}			
	(0.018)	(0.018)	(0.017)			
ldc (lagged 12 months)		−0.092	−0.327		0.027	−0.105
		(0.226)	(0.203)		(0.266)	(0.221)
fd (lagged 1 month)			-0.004^{b}			−0.001
			(0.002)			(0.003)
$\Delta VIX \times fd$ (lagged 1 month)			−0.001			
			($<$ .001)			
ΔVIX *dummy*				-1.119^{c}	-1.119^{c}	-0.903^{c}
				(0.243)	(0.244)	(0.268)
ΔVIX *dummy* $\times nfa$ (lagged 12 months)				0.731^{c}	0.731^{c}	0.563^{c}
				(0.265)	(0.264)	(0.247)
ΔVIX *dummy* $\times fd$ (lagged 1 month)						-0.012^{b}
						(0.006)
Additional Variables: Constant and lagged exchange rate returns	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
Adjusted R^2	0.08	0.08	0.08	0.02	0.02	0.02
Observations	8960	8960	8960	9112	9112	9112

Table 6. Risk Bearing Capacity and Global Imbalance Portfolios

This table presents results from time-series regressions. In *Panel A*, we regress monthly currency excess returns to the *global imbalance portfolios* (see Table 1) on a constant and the monthly changes in the VIX index. In *Panel B*, we regress the exchange rate return component to the *global imbalance portfolios* on a constant and the monthly changes in the VIX index. Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in parentheses. The sample runs from January 1986 to June 2014. See Section 3 for a detailed description of data sources and data construction, and Figure 1 for a detailed description of the portfolio construction.

	Panel A: Currency Excess Returns				
	P_1	P_2	P_3	P_4	P_5
ΔVIX	−0.037	−0.059	−0.144	−0.146	−0.148
	(0.036)	(0.044)	(0.034)	(0.034)	(0.054)
Constant	0.102	0.312	0.157	0.326	0.493^c
	(0.117)	(0.129)	(0.094)	(0.115)	(0.146)
Adjusted R^2	0.01	0.01	0.15	0.11	0.07
	Panel B: Spot Exchange Rate Returns				
	P_1	P_2	P_3	P_4	P_5
ΔVIX	−0.035	−0.057	−0.143	−0.143	−0.143
	(0.037)	(0.045)	(0.035)	(0.033)	(0.054)
Constant	0.127	0.187	−0.027	0.027	−0.092
	(0.166)	(0.128)	(0.094)	(0.116)	(0.147)
Adjusted R^2	0.00	0.01	0.15	0.10	0.07

Table 7. Asset Pricing Tests: Currency Strategies

The table presents asset pricing results for the *carry trade portfolios* (sorted on the one-month forward discounts), the *global imbalance portfolios* (sorted on the net foreign assets to gross domestic product ratio and the share of foreign liabilities in domestic currency), the *momentum portfolios* (sorted on the past three-month exchange rate returns), the *value portfolios* (sorted on the past five-years exchange rate returns), the *risk reversal portfolios* (sorted on the one-year 25 delta currency option risk reversal), the *20 currency portfolios* (all except the *risk reversal* portfolios), and the *25 currency portfolios* (all portfolios). These portfolios' excess returns are used as test assets whereas the dollar (DOL) and the global imbalance (IMB) act as risk factors. We report first-stage GMM estimates of the factor loadings b, the market price of risk λ, and the cross-sectional R^2. Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in parentheses. HJ is the Hansen and Jagannathan (1997) distance (with simulated p-value in brackets) for the null hypothesis that the HJ distance is equal to zero. Excess returns are in annual terms and adjusted for transaction costs. The portfolios are rebalanced monthly from October 1983 (or January 1996) to June 2014. See Section 3 for a detailed description of data sources and data construction.

Portfolios	Sample	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	HJ
5 Carry Trade	10/83 – 06/14	−0.07	1.53	0.02	0.07	0.87	0.14
		(0.29)	(0.54)	(0.01)	(0.02)		[0.16]
5 Global Imbalance	10/83 – 06/14	0.30	0.63	0.03	0.04	0.75	0.17
		(0.26)	(0.29)	(0.01)	(0.01)		[0.92]
5 Momentum	10/83 – 06/14	0.31	0.22	0.02	0.02	$< .01$	0.16
		(0.30)	(0.75)	(0.01)	(0.03)		[0.04]
5 Value	10/83 – 06/14	0.01	1.33	0.02	0.07	0.66	0.14
		(0.30)	(0.53)	(0.01)	(0.02)		[0.18]
5 Risk Reversal	01/96 – 06/14	−0.12	1.55	0.02	0.09	0.96	0.04
		(0.44)	(0.93)	(0.02)	(0.05)		[0.97]
20 Currency	10/83 – 06/14	0.12	0.97	0.02	0.05	0.53	0.81
		(0.27)	(0.36)	(0.01)	(0.01)		[0.17]
25 Currency	01/96 – 06/14	−0.08	1.38	0.02	0.08	0.65	0.93
		(0.41)	(0.38)	(0.01)	(0.01)		[0.86]

Table 8. Asset Pricing Tests: Bond, Currency, Commodity and Equity Strategies

The table presents asset pricing results for the *currency portfolios* defined in Table 6, the Fama and French (2012) *size and book-to-market (momentum) global portfolios*, the *international bond portfolios* (sorted on redemption yields), and the Yang (2013) *commodity portfolios*. These portfolios' excess returns are used as test assets whereas the set of risk factor includes the dollar (DOL), the global imbalance (IMB), the Fama and French (2012) factors (MKT, SMB, HML , and WML), the high-minus-low international bond factor (IB), and the high-minus-low commodity factor (COM). We report first-stage GMM estimates of the factor loadings b, the market price of risk λ, and the cross-sectional R^2. Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in parentheses. HJ is the Hansen and Jagannathan (1997) distance (with simulated p-value in brackets) for the null hypothesis that the HJ distance is equal to zero. Excess returns are in annual terms. See Section 3 for a detailed description of data sources and data construction.

Panel A: Size and Book-to-Market Global Portfolios

	Sample	b_{MKT}	b_{SMB}	b_{HML}	b_{DOL}	b_{IMB}	λ_{MKT}	λ_{SMB}	λ_{HML}	λ_{DOL}	λ_{IMB}	R^2	HJ
20 Currency + 25 Size and Book-to-Market	07/90-06/14	0.16	0.47	0.98	−0.22	1.19	0.06	0.02	0.07	0.02	0.07	0.78	0.96
		(0.21)	(0.36)	(0.43)	(0.50)	(0.39)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)		[0.38]
25 Currency + 25 Size and Book-to-Market	01/96-06/14	0.25	0.47	1.06	−0.61	1.45	0.07	0.01	0.08	0.02	0.09	0.84	1.02
		(0.34)	(0.40)	(0.50)	(0.74)	(0.35)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)		[0.20]

Panel B: Size and Momentum Global Portfolios

	Sample	b_{MKT}	b_{SMB}	b_{WML}	b_{DOL}	b_{IMB}	λ_{MKT}	λ_{SMB}	λ_{WML}	λ_{DOL}	λ_{IMB}	R^2	HJ
20 Currency + 25 Size and Momentum	11/90-06/14	0.41	0.77	0.63	−0.39	1.24	0.09	0.04	0.10	0.02	0.07	0.83	0.98
		(0.24)	(0.37)	(0.31)	(0.42)	(0.36)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)		[0.34]
25 Currency + 25 Size and Momentum	01/96-06/14	0.37	0.62	0.60	−0.48	1.52	0.10	0.04	0.11	0.02	0.09	0.86	1.02
		(0.31)	(0.42)	(0.34)	(0.57)	(0.37)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)		[0.22]

Panel C: International Bond Portfolios

	Sample	b_{IB}	b_{DOL}	b_{IMB}	λ_{IB}	λ_{DOL}	λ_{IMB}	R^2	HJ
20 Currency + 5 International Bond	10/83-06/14	2.03	0.01	0.90	0.06	0.02	0.05	0.63	0.92
		(0.87)	(0.38)	(0.40)	(0.01)	(0.01)	(0.01)		[0.48]
25 Currency + 5 International Bond	01/96-06/14	3.08	−0.43	1.37	0.08	0.02	0.07	0.64	1.02
		(0.92)	(0.47)	(0.45)	(0.01)	(0.01)	(0.01)		[0.26]

Panel D: Commodity Portfolios

	Sample	b_{COM}	b_{DOL}	b_{IMB}	λ_{COM}	λ_{DOL}	λ_{IMB}	R^2	HJ
20 Currency + 7 Commodity	10/83-12/98	0.17	0.16	0.86	0.09	0.02	0.06	0.37	0.89
		(0.14)	(0.51)	(0.43)	(0.04)	(0.01)	(0.01)		[0.59]
25 Currency + 7 Commodity	01/96-12/98	0.28	0.01	1.25	0.15	0.02	0.09	0.61	0.96
		(0.19)	(0.78)	(0.51)	(0.03)	(0.01)	(0.01)		[0.41]

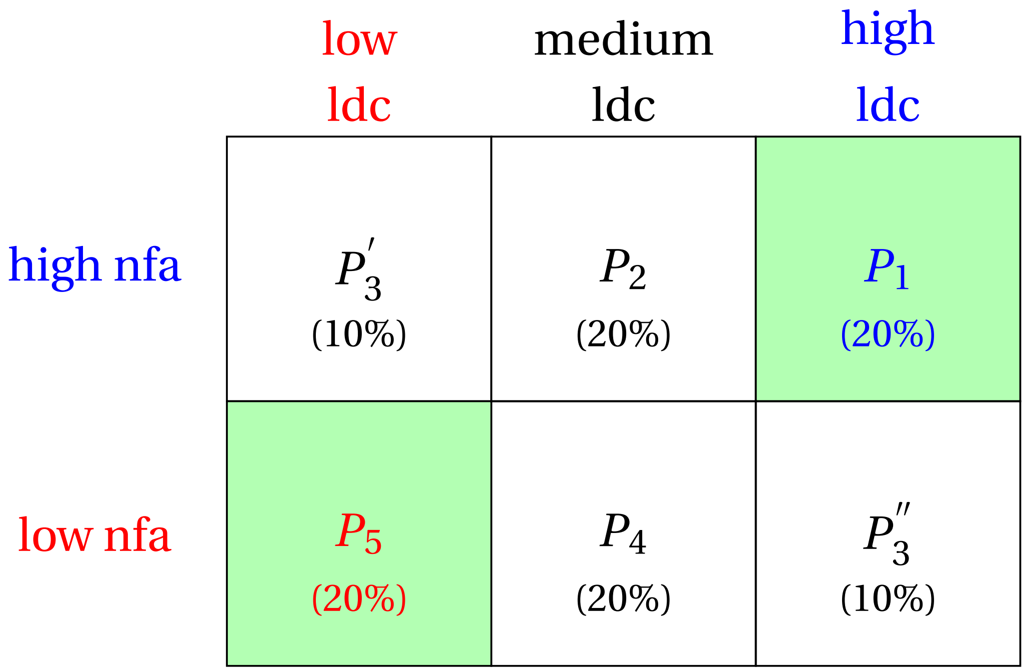


Figure 1. Global Imbalance Portfolios: Construction

This chart describes the construction of the global imbalance portfolios. At the end of each month, currencies are first grouped into two baskets using the net foreign asset position to gross domestic product (nfa), and then into 3 baskets using the share of foreign liabilities in domestic currency (ldc). The nfa breakpoint is the median value whereas the ldc breakpoints are the 40^{th} and 80^{th} percentiles. The first portfolio (P_1) contains the top 20% of all currencies with high nfa and high ldc (creditor nations with external liabilities mainly in domestic currency) whereas the last portfolio (P_5) contains the top 20% of all currencies with low nfa and low ldc (debtor nations with external liabilities mainly in foreign currency). The portfolios $P_3^{'}$ and $P_3^{''}$ are intermediate portfolios containing each 10% of all currencies, and are aggregated into the portfolio P_3. The global imbalance factor (IMB) is constructed as the average return on P_5 minus the average return on P_1. We use 5 portfolios rather than 6 portfolios as we have a limited number of currencies in *developed countries*. Figure A.1 in the Internet Appendix describes the construction of the IMB factor based on 6 (4) portfolios for *all countries* (*developed countries*).

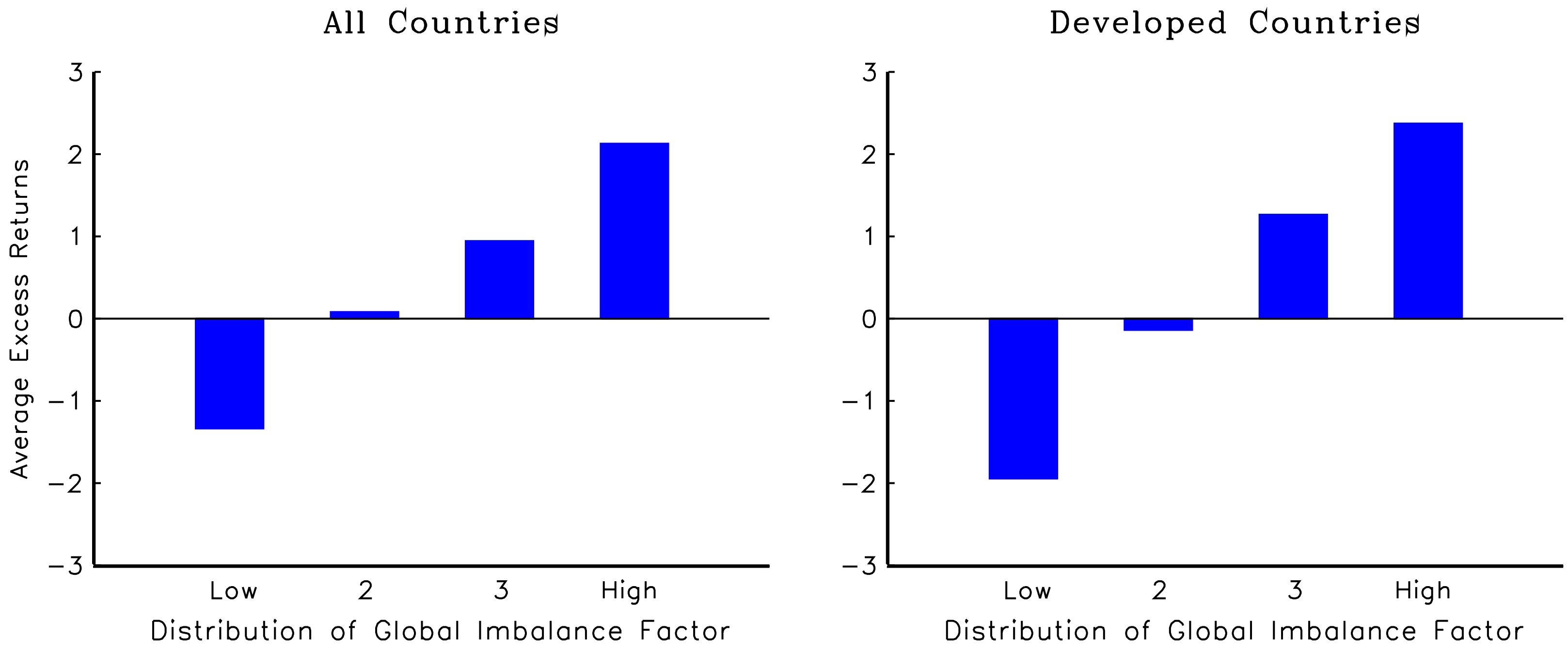


Figure 2. Carry Trade Returns and Global Imbalance Risk

The figure presents average excess returns for carry trade returns conditional on the global imbalance risk factor being within the lowest to highest quartile of its sample distribution. Excess returns are expressed in percentage per month. The sample runs from October 1983 to June 2014.

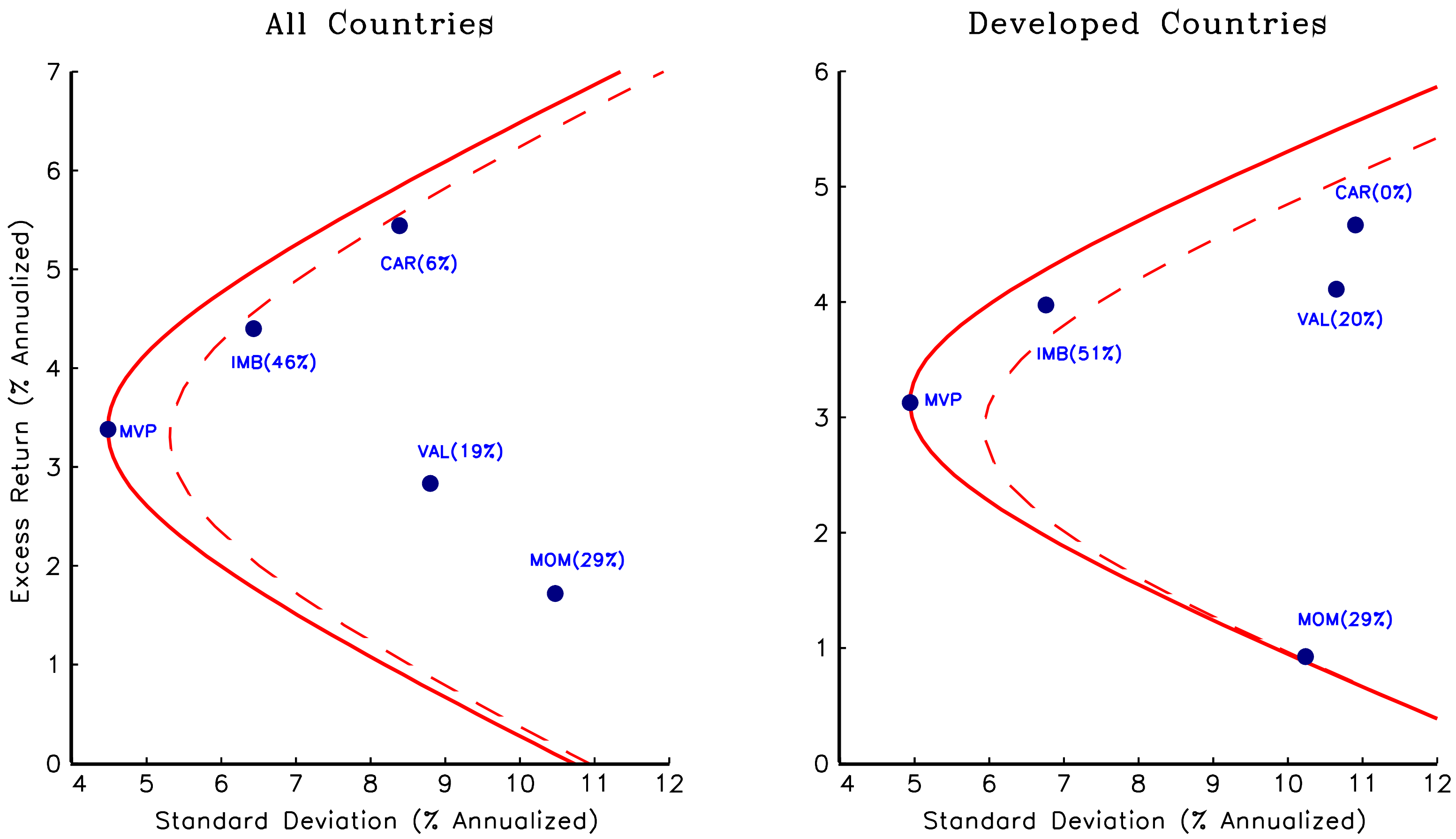


Figure 3. Global Minimum Variance Portfolio

The figure presents the global minimum variance portfolio (MVP) and the efficient frontier (*solid line*) built using a set of currency strategies formed using $t-1$ information. CAR is the carry strategy that buys (sells) the top 20% of all currencies with high (low) interest rate differential relative to the US dollar. Similarly, IMB is the global imbalance strategy that buys (sells) the currencies of debtor nations with external liabilities mainly in foreign currency (currencies of creditor nations with external liabilities mainly in domestic currency), MOM is the momentum strategy that buys (sells) currencies with high (low) past 3-month exchange rate return, and VAL is the value strategy that buys (sells) currencies with low (high) past 5-year exchange rate return. The portfolio weights are reported in parentheses and computed as $w = (\Sigma^{-1}\iota)/(\iota'\Sigma^{-1}\iota)$ where Σ is the $N \times N$ covariance matrix of the strategies' excess returns, ι is a $N \times 1$ vector of ones, and N denotes the number of strategies. The *dashed line* denotes the efficient frontier when we exclude the IMB strategy from the investment opportunity set. Excess returns are adjusted for transaction costs and expressed in percentage per annum. The strategies are rebalanced monthly from October 1983 to June 2014.

References

Alquist, R., and M.D. Chinn (2008). "Conventional and Unconventional Approaches to Exchange Rate Modelling and Assessment," *International Journal of Finance and Economics* **13**, 2–13.

Akram, Q.F., D. Rime, and L. Sarno (2008). "Arbitrage in the Foreign Exchange Market: Turning on the Microscope," *Journal of International Economics* **76**, 237–253.

Andrews, D.W.K. (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* **59**, 817–858.

Ang, A., R.J. Hodrick, Y. Xing, and X. Zhang (2006). "The Cross-Section of Volatility and Expected Returns," Journal of Finance **61**, 259–299.

Ang, A., J. Liu, and K. Schwarz (2010). "Using Individual Stocks or Portfolios in Tests of Factor Models," Working Paper, Columbia Business School.

Asness, C.S., T.J. Moskowitz, and L.H. Pedersen (2013). "Value and Momentum Everywhere," *Journal of Finance* **68**, 929-985.

Bank for International Settlements (2013). *Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity in 2013.* Basel: Bank for International Settlements Press.

Barroso, P., and P. Santa-Clara (2014). "Beyond the Carry Trade: Optimal Currency Portfolios," *Journal of Financial and Quantitative Analysis*, forthcoming.

Bassett, G. and R. Koenker (1978). "Asymptotic Theory of Least Absolute Error Regression," *Journal of American Statistical Association* **73**, 618–622.

Bekaert, G. and R.J. Hodrick (1993). "On Biases in the Measurement of Foreign Exchange Risk Premiums," *Journal of International Money and Finance* **12**, 115-138.

Benetrix, A., P.R. Lane, and J.C. Shambaugh (2014). "International Currency Exposures, Valuation Effects and the Global Financial Crisis," *Journal of International Economics*, forthcoming.

Bilson, J.F.O. (1981). "The 'Speculative Efficiency' Hypothesis," *Journal of Business* **54**, 435–451.

Brunnermeier, M.K., S. Nagel, and L.H. Pedersen (2009). "Carry Trades and Currency Crashes," *NBER Macroeconomics Annual 2008*, 313–347.

Burnside, C. (2011). "The Cross Section of Foreign Currency Risk Premia and Consumption Growth Risk: Comment," *American Economic Review* **101**, 3456–76.

Burnside, C., M. Eichenbaum, I. Kleshchelski, and S. Rebelo (2011). "Do Peso Problems Explain the Returns to the Carry Trade?" *Review of Financial Studies* **24**, 853–891.

Christiansen, C., A. Ranaldo, and P. Soderlind (2011). "The Time-Varying Systematic Risk of Carry Trade Strategies," *Journal of Financial and Quantitative Analysis* **46**, 1107–1125.

Cochrane, J.H. (2005). *Asset Pricing.* Princeton: Princeton University Press.

Colacito, R., and M.M. Croce (2013). "International Asset Pricing with Recursive Preferences," *Journal of Finance* **68**, 2651–2686.

Della Corte, P., L. Sarno, and G. Sestieri (2012). "The Predictive Information Content of External Imbalances for Exchange Rate Returns: How Much Is It Worth?" *Review of Economics and Statistics* **94**, 100–115.

Della Corte, P., L. Sarno, and I. Tsiakas (2009). "An Economic Evaluation of Empirical Exchange Rate Models," *Review of Financial Studies* **22**, 3491–3530.

Della Corte, P., L. Sarno, and I. Tsiakas (2011). "Spot and Forward Volatility in Foreign Exchange," *Journal of Financial Economics* **100**, 496–513.

Eichengreen, B. and Hausmann, R. (2005). *Other People's Money - Debt Denomination and Financial Instability in Emerging Market Economies*, University of Chicago Press: Chicago and London.

Engel, C. (1996). "The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence," *Journal of Empirical Finance* **3**, 123–192.

Fama, E.F. (1984). "Forward and Spot Exchange Rates," *Journal of Monetary Economics* **14**, 319–338.

Fama, E.F., and K.R. French (2012). "Size, Value, and Momentum in International Stock Returns," *Journal of Financial Economics* **105**, 457-472.

Fama, E.F., and J. MacBeth (1973). "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy* **81**, 607–636.

Farhi, E., and X. Gabaix (2014). "Rare Disasters and Exchange Rates," Working Paper, New York University.

Farhi, E., S.P. Fraiberger, X. Gabaix, R. Ranciere, and A. Verdelhan (2014). "Crash Risk in Currency Markets," Working Paper, MIT Sloan.

Gabaix, X., and M. Maggiori (2014). "International Liquidity and Exchange Rate Dynamics," Working Paper, New York University and Harvard University.

Gourinchas, P.O. (2008). "Valuation Effects and External Adjustment: A Review," in Cowan, K., Edwards, S., and R. Valdes (eds.), *Current Account and External Financing*. Santiago, Chile: Central Bank of Chile.

Gourinchas, P.O., and H. Rey (2007). "International Financial Adjustment," *Journal of Political Economy* **115**, 665–703.

Habib, M.M., and L. Stracca (2012). "Getting Beyond Carry Trade: What Makes a Safe Haven Currency?" *Journal of International Economics* **87**, 50–64.

Hansen, L.P. (1982). "Large Sample Properties of Generalized Methods of Moments Estimators," *Econometrica* **50**, 1029–1054.

Hansen, L.P., and R. Jagannathan (1997). "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance* **52**, 557–590.

Hansen, L.P., and R. Hodrick (1980). "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy* **88**, 829–853.

Jagannathan, R., and Z. Wang (1996). "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance* **51,** 3–53.

Jurek, J.W. (2014). "Crash-Neutral Currency Carry Trades," *Journal of Financial Economics* **113**, 325–347.

Jylha P., and M. Suominen (2011). "Speculative Capital and Currency Carry Trades," *Journal of Financial Economics* **99**, 60–75.

Koenker, R. and G. Bassett (1982). "Robust Tests for Heteroskedasticity Based on Regression Quantiles," *Econometrica*, **50**, 43–61.

Koijen, R.S.J., T.J. Moskowitz, L.H. Pedersen and E.B. Vrugt (2013), "Carry," NBER Working Paper 19325.

Lane, P.R., and G. M. Milesi-Ferretti (2004). "The External Wealth of Nations: Measures of Foreign Assets and Liabilities for Industrial and Developing Countries," *Journal of International Economics* **55**, 263–294.

Lane, P.R., and G.M. Milesi-Ferretti (2007). "The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities, 1970-2004," *Journal of International Economics* **73**, 223–250.

Lane, P.R., and J.C. Shambaugh (2010). "Financial Exchange Rates and International Currency Exposure," *American Economic Review* **100**, 518–540.

Lettau, M., M. Maggiori, and M. Weber (2014). "Conditional Risk Premia in Currency Markets and Other Asset Classes," *Journal of Financial Economics* **114**, 197–225.

Lewellen, J., S. Nagel, and J. Shanken (2010). "A Skeptical Appraisal of Asset Pricing Tests," *Journal of Financial Economics* **96**, 175–194.

Lustig, H., N. Roussanov, and A. Verdelhan (2011). "Common Risk Factors in Currency Markets," *Review of Financial Studies* **24**, 3731–3777.

Lustig, H., and A. Verdelhan (2007). "The Cross Section of Foreign Currency Risk Premia and US Consumption Growth Risk," *American Economic Review* **97**, 89–117.

Maggiori, M. (2013). "Financial Intermediation, International Risk Sharing, and Reserve Currencies," Working Paper, Harvard University.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012a). "Carry Trades and Global FX Volatility," *Journal of Finance* **67**, 681–718.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012b). "Currency Momentum Strategies," *Journal of Financial Economics* **106**, 620-684.

Newey, W.K., and K.D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* **55**, 703–708.

Ren, Y., and K. Shimotsu (2009). "Improvement in Finite Sample Properties of the Hansen-Jagannathan Distance Test," *Journal of Empirical Finance* **16**, 483-506.

Rose, D. (2010). "The Influence of Foreign Assets and Liabilities on Real Interest Rates," *Institute of Policy Studies Working Paper No. 10/09.*

Ross, A. (2013). "Indian Rupee Hits Record Low," *Financial Times*, 26 June.

Shanken, J. (1992). "On the Estimation of Beta-Pricing Models," *Review of Financial Studies* **5**, 1–34.

Yang, F. (2013). "Investment Shocks and the Commodity Basis Spread," *Journal of Financial Economics* **110**, 164–184.

Internet Appendix for:

Currency Premia and Global Imbalances

(not for publication)

December 2014

Table A.1. Descriptive Statistics: Carry Trade Portfolios

The table presents descriptive statistics of currency portfolios sorted on time $t-1$ forward discounts or interest rate differential relative to the US (fd). The first portfolio (P_1) contains the top 20% of all currencies with low fd (low-yielding currencies) whereas the last portfolio (P_5) contains the top 20% of all currencies with high fd (high-yielding currencies). CAR is a long-short strategy that buys P_5 and sells P_1. The table also reports the first order autocorrelation coefficient (ac_1), the annualized Sharpe ratio (SR), the maximum drawdown (mdd), the frequency of portfolio switches ($freq$), the average fd, the average net foreign asset position to gross domestic product (nfa), and the average share of foreign liabilities in domestic currency (ldc). t-statistics based on Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in brackets. Excess returns are expressed in percentage per annum and adjusted for transaction costs. The portfolio are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction.

	P_1	P_2	P_3	P_4	P_5	CAR	P_1	P_2	P_3	P_4	P_5	CAR
	All Countries						*Developed Countries*					
$mean$	0.47	0.05	2.38	2.27	5.91	5.44	0.65	0.93	1.93	2.59	5.31	4.67
t-$stat$	[0.33]	[0.04]	[1.55]	[1.34]	[3.10]	[3.20]	[0.36]	[0.49]	[1.11]	[1.47]	[2.41]	[2.24]
$median$	0.82	1.53	2.87	2.98	8.53	7.67	−0.67	1.91	3.55	3.33	5.30	7.48
$sdev$	7.79	7.66	8.10	8.67	9.28	8.39	9.98	9.64	9.12	9.51	11.39	10.90
$skew$	0.16	−0.16	−0.30	−0.59	−0.42	−0.86	0.33	0.01	−0.07	−0.42	−0.19	−0.95
$kurt$	4.22	3.73	4.12	5.66	4.67	5.02	3.76	3.60	3.86	4.69	4.28	5.38
ac_1	0.01	0.03	0.07	0.10	0.13	0.13	0.00	0.07	0.08	0.04	0.11	0.07
SR	0.06	0.01	0.29	0.26	0.64	0.65	0.06	0.10	0.21	0.27	0.47	0.43
mdd	0.42	0.37	0.37	0.35	0.31	0.33	0.55	0.45	0.38	0.29	0.37	0.38
$freq$	0.19	0.28	0.31	0.33	0.17		0.13	0.26	0.32	0.24	0.13	
fd	−1.69	−0.86	0.93	3.30	9.28		−2.05	−1.01	0.31	1.61	4.04	
nfa	0.53	0.06	−0.15	−0.35	−0.39		0.60	0.00	−0.10	−0.08	−0.39	
ldc	0.52	0.50	0.43	0.43	0.40		0.54	0.55	0.44	0.43	0.41	

Table A.2. Portfolio Composition

The table presents the composition of the five carry trade and global imbalance portfolios. In *Panel A* and *Panel B*, we report the top six currencies (with the frequency in parentheses) entering each portfolio. *Panel C* presents the probability that a given currency enters simultaneously the same carry trade and global imbalance portfolio. The portfolios are rebalanced monthly from October 1983 to June 2014.

All Countries					*Developed Countries*				
P_1	P_2	P_3	P_4	P_5	P_1	P_2	P_3	P_4	P_5
Panel A: Carry Trade Portfolios									
JPY [0.18]	CAD [0.08]	GBP [0.08]	AUD [0.07]	ZAR [0.13]	CHF [0.42]	NLG [0.14]	DKK [0.20]	GBP [0.17]	NZD [0.35]
CHF [0.16]	DKK [0.08]	NOK [0.06]	NZD [0.06]	TRY [0.10]	JPY [0.40]	EUR [0.14]	CAD [0.15]	NOK [0.16]	AUD [0.25]
SGD [0.12]	SGD [0.06]	CAD [0.05]	MXN [0.05]	BRL [0.06]	DEM [0.06]	CAD [0.10]	GBP [0.14]	SEK [0.15]	ITL [0.12]
HKD [0.09]	EUR [0.06]	KRW [0.05]	HUF [0.05]	MXN [0.05]	DKK [0.03]	DEM [0.10]	SEK [0.11]	AUD [0.14]	NOK [0.10]
CNY [0.05]	SEK [0.05]	HKD [0.05]	GBP [0.05]	HUF [0.05]	CAD [0.03]	SEK [0.08]	NOK [0.10]	CAD [0.12]	GBP [0.08]
MYR [0.04]	NLG [0.04]	NZD [0.05]	PHP [0.04]	NZD [0.05]	NLG [0.02]	DKK [0.07]	FRF [0.07]	DKK [0.08]	SEK [0.06]
Panel B: Global Imbalance Portfolios									
CHF [0.11]	GBP [0.10]	AUD [0.11]	HUF [0.09]	TRY [0.11]	CHF [0.26]	CHF [0.20]	NOK [0.29]	CAD [0.28]	DKK [0.30]
HKD [0.10]	JPY [0.09]	NOK [0.10]	NZD [0.09]	PHP [0.09]	JPY [0.26]	JPY [0.19]	AUD [0.25]	SEK [0.22]	NZD [0.22]
EUR [0.10]	CHF [0.08]	SGD [0.08]	PLN [0.07]	DKK [0.09]	DEM [0.16]	NLG [0.16]	GBP [0.17]	NZD [0.22]	SEK [0.19]
JPY [0.09]	NLG [0.07]	GBP [0.07]	MXN [0.06]	SEK [0.08]	EUR [0.12]	FRF [0.13]	EUR [0.11]	AUD [0.11]	GBP [0.18]
SGD [0.09]	FRF [0.06]	HKD [0.07]	TND [0.06]	NZD [0.08]	CAD [0.10]	DKK [0.13]	ITL [0.11]	NOK [0.09]	ITL [0.06]
CNY [0.08]	ILS [0.06]	CAD [0.06]	MYR [0.06]	ISK [0.07]	FRF [0.07]	GBP [0.09]	BEF [0.03]	ITL [0.04]	AUD [0.04]
Panel C: Joint Probability									
[0.45]	[0.25]	[0.24]	[0.26]	[0.38]	[0.47]	[0.28]	[0.19]	[0.25]	[0.41]

Table A.3. Asset Pricing: Global Imbalance Risk with Constraints

This table re-estimates the linear factor model of Table 2 while including the global imbalance factor as one of the test assets as suggested by Lewellen, Nagel and Shanken (2010). We report GMM (first and second-stage) and Fama-MacBeth (FMB) estimates of factor loadings b, market price of risk λ, and cross-sectional R^2. Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in parentheses whereas Shanken (1992) standard errors are reported in brackets. χ^2 denotes the test statistics (with p-value in brackets) for the null hypothesis that all pricing errors are jointly zero. HJ refers to the Hansen and Jagannathan (1997) distance (with simulated p-value in brackets) for the null hypothesis that the HJ distance is equal to zero. Excess returns are in annual terms and adjusted for transaction costs. The portfolios are rebalanced monthly from October 1983 to June 2014.

	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ
	Developed Countries								*All Countries*							
GMM_1	0.17	0.75	0.02	0.04	0.91	1.22%	1.56	0.07	0.06	1.03	0.02	0.05	0.79	2.31%	7.08	0.17
	(0.21)	(0.33)	(0.02)	(0.01)			[0.82]	[0.85]	(0.27)	(0.31)	(0.01)	(0.01)			[0.13]	[0.10]
GMM_2	0.20	0.73	0.02	0.05	0.96	1.36%	1.53		0.11	0.91	0.02	0.07	0.95	2.88%	6.85	
	(0.21)	(0.26)	(0.02)	(0.01)			[0.82]		(0.26)	(0.27)	(0.01)	(0.01)			[0.14]	
FMB	0.17	0.75	0.02	0.04	0.91	1.22%	1.56		0.06	1.03	0.02	0.05	0.79	2.31%	7.07	
	(0.19)	(0.23)	(0.02)	(0.01)			[0.82]		(0.24)	(0.28)	(0.01)	(0.01)			[0.13]	
	[0.17]	[0.25]	[0.02]	[0.01]					[0.22]	[0.26]	[0.01]	[0.01]				

Table A.4. Country-Level Asset Pricing Tests: Global Imbalance Risk

This table presents cross-sectional asset pricing results for individual currency excess returns. The linear factor model includes the dollar (DOL) and the global imbalance (IMB) risk factor. IMB is a long-short strategy that buys the currency (top 20%) of debtor nations with external liabilities mainly in foreign currency and sells the currencies (top 20%) of creditor nations with external liabilities mainly in domestic currency. The test assets are use country-level unconditional and conditional excess returns. The unconditional excess return for each currency pair is computed as $RX_{t+1} = (F_t - S_{t+1})/S_t$, where S_t denotes the spot exchange rate and F_t is the one-month forward rate. The conditional excess return is calculated as $RX_{t+1} = \gamma \times (F_t - S_{t+1})/S_t$, where $\gamma = 1$ when $F_t > S_t$ (foreign interest rate is higher than US interest rate) and $\gamma = -1$ when $F_t < S_t$ (foreign interest rate is lower than US interest rate). The table reports estimates of the market price of risk λ, the cross-sectional R^2 and the root mean squared error ($RMSE$) obtained via Fama-MacBeth procedure with robust regressions in the first and second step to account for outliers in individual currency excess returns. Bootstrapped standard errors are reported in parentheses. We use a block-bootstrap algorithm based on 1,000 repetitions. Excess returns are in annual terms and run at monthly frequency from October 1983 to June 2014.

Unconditional Excess Returns				Conditional Excess Returns			
λ_{DOL}	λ_{IMB}	R^2	$RMSE$	λ_{DOL}	λ_{IMB}	R^2	$RMSE$
All Countries							
0.04	0.06	0.34	28.4	0.04	0.07	0.25	35.9
(0.02)	(0.02)			(0.02)	(0.02)		
Developed Countries							
0.03	0.05	0.64	4.6	0.05	0.08	0.32	7.4
(0.02)	(0.02)			(0.03)	(0.02)		

A.5. Determinants of Excess Currency Returns

The table presents results from fixed-effects panel regressions. We use discrete currency excess returns at monthly frequency as dependent variable. The set of independent variables includes the net foreign asset position to gross domestic product (nfa), the share of foreign liabilities in domestic currency (ldc), the monthly change in the VIX index (ΔVIX), and a dummy variable that equals one if ΔVIX is greater than one standard deviation as estimated across the entire sample, and zero otherwise (ΔVIX *dummy*). Robust standard errors are clustered at country level and reported in parentheses. The superscripts a, b and c denote statistical significance at 10%, 5% and 1% level, respectively. The sample runs from January 1986 to June 2014. See Section 3 for a detailed description of data sources and data construction.

	Dependent variable: currency excess returns			
	(1)	(2)	(3)	(4)
nfa (lagged 12 months)	0.001	0.031	−0.110	−0.084
	(0.071)	(0.088)	[0.069)	(0.077)
ΔVIX	-0.145^c	-0.145^c		
	(0.017)	(0.017)		
$\Delta VIX \times nfa$ (lagged 12 months)	0.069^c	0.069^c		
	(0.018)	(0.018)		
ldc (lagged 12 months)		-0.751^b		-0.630^b
		(0.291)		(0.292)
ΔVIX *dummy*			-1.130^c	-1.127^c
			(0.243)	(0.243)
ΔVIX *dummy* $\times nfa$ (lagged 12 months)			0.719^c	0.718^c
			(0.261)	(0.260)
Additional Variables: Constant and lagged excess currency return	YES	YES	YES	YES
Adjusted R^2	0.08	0.08	0.02	0.02
Observations	8960	8960	9112	9112

Table A.6. Global Imbalance Portfolios and Other Pricing Currencies

The table presents descriptive statistics of currency portfolios sorted on time $t-1$ net foreign asset position to gross domestic product (nfa), and the share of foreign liabilities in domestic currency (ldc) when the pricing currency is not the US dollar. The first portfolio (P_1) contains the top 20% of all currencies with high nfa and high ldc (creditor nations with external liabilities mainly in domestic currency) whereas the last portfolio (P_5) contains the top 20% of all currencies with low nfa and low ldc (debtor nations with external liabilities mainly in foreign currency). IMB is a long-short strategy that buys P_5 and sells P_1. The table also reports the first order autocorrelation coefficient (ac_1), the annualized Sharpe ratio (SR), the maximum drawdown (mdd), the frequency of portfolio switches ($freq$), the average forward discount (fd), the average nfa, and the average ldc. t-statistics based on Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in brackets. Excess returns are expressed in percentage per annum and adjusted for transaction costs. The portfolio are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction, and Figure 1 for the portfolio construction.

	P_1	P_2	P_3	P_4	P_5	IMB	P_1	P_2	P_3	P_4	P_5	IMB
	CHF as pricing currency						*DEM-EUR as pricing currency*					
mean	0.48	1.73	0.38	2.45	4.34	3.85	−0.29	1.39	−0.14	1.92	3.91	4.20
t-stat	[0.40]	[1.47]	[0.19]	[1.50]	[2.65]	[3.04]	[−0.25]	[1.55]	[−0.08]	[1.40]	[2.82]	[3.06]
median	0.28	2.23	2.58	2.80	5.63	3.14	−1.83	1.50	1.68	1.72	2.66	4.38
sdev	6.75	6.65	10.24	9.03	8.52	6.11	6.31	4.95	8.88	7.60	6.92	6.56
skew	−0.06	−0.15	−0.19	−0.44	0.17	0.53	0.29	0.34	−0.02	−0.41	0.82	0.41
kurt	4.48	4.03	4.36	3.84	4.78	7.45	5.22	3.87	4.92	4.06	7.60	6.61
ac_1	−0.05	−0.04	0.07	0.01	0.07	0.15	0.01	−0.01	0.07	−0.01	0.13	0.14
SR	0.07	0.26	0.04	0.27	0.51	0.63	−0.05	0.28	−0.02	0.25	0.57	0.64
mdd	0.30	0.23	0.44	0.31	0.27	0.23	0.42	0.16	0.43	0.37	0.22	0.24
$freq$	0.02	0.03	0.03	0.04	0.03		0.02	0.03	0.04	0.04	0.03	
fd	1.95	3.31	3.77	5.36	8.33		0.55	2.04	2.56	4.18	7.17	
nfa	0.26	0.07	0.09	−0.47	−0.57		0.43	0.15	0.09	−0.47	−0.56	
ldc	0.62	0.45	0.44	0.47	0.28		0.61	0.45	0.44	0.47	0.28	

continued

Table A.6. Global Imbalance Portfolios and Other Pricing Currencies *(continued)*

	P_1	P_2	P_3	P_4	P_5	IMB	P_1	P_2	P_3	P_4	P_5	IMB
	GBP as pricing currency						*JPY as pricing currency*					
mean	−1.24	0.60	−0.80	1.25	3.17	4.41	1.27	3.41	1.41	3.61	5.64	4.37
t-stat	[−0.82]	[0.41]	[−0.46]	[0.80]	[2.09]	[3.38]	[0.69]	[1.69]	[0.60]	[1.65]	[2.38]	[3.43]
median	−2.25	−1.44	0.24	1.86	3.00	4.38	2.58	5.96	5.11	7.14	7.27	4.29
sdev	7.48	7.76	9.35	8.89	8.41	6.26	9.92	10.38	11.86	11.77	11.70	6.18
skew	0.57	0.87	0.13	0.28	0.91	0.39	−0.44	−0.50	−0.57	−0.68	−0.43	0.78
kurt	6.27	6.02	6.09	5.21	7.88	6.74	5.32	4.60	4.52	5.06	4.08	8.20
ac_1	0.12	0.08	0.05	−0.05	0.02	0.15	0.05	0.09	0.08	0.06	0.12	0.14
SR	−0.17	0.08	−0.09	0.14	0.38	0.70	0.13	0.33	0.12	0.31	0.48	0.71
mdd	0.56	0.39	0.55	0.43	0.30	0.24	0.37	0.29	0.49	0.48	0.38	0.23
$freq$	0.02	0.03	0.03	0.04	0.03		0.02	0.03	0.03	0.04	0.03	
fd	−2.30	−0.71	0.21	1.74	4.74		2.53	4.39	4.45	6.32	9.14	
nfa	0.39	0.15	0.12	−0.47	−0.57		0.41	0.13	0.09	−0.47	−0.57	
ldc	0.63	0.47	0.44	0.47	0.28		0.62	0.45	0.45	0.47	0.28	

Table A.7. Asset Pricing Tests and Other Pricing Currencies

The table presents cross-sectional asset pricing results for the linear factor model based on the dollar (DOL) and the global imbalance (IMB) risk factor when the pricing currency is not the US dollar. The test assets are excess returns to five carry trade portfolios sorted on the one-month forward discounts. IMB is a long-short strategy that buys the currency (top 20%) of debtor nations with external liabilities mainly in foreign currency, and sells the currencies (top 20%) of creditor nations with external liabilities mainly in domestic currency. *Panel A* reports GMM (first and second-stage) and Fama-MacBeth (FMB) estimates of the factor loadings b, the market price of risk λ, and the cross-sectional R^2. Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in parentheses whereas Shanken (1992) standard errors are reported in brackets. χ^2 denotes the test statistics (with p-value in brackets) for the null hypothesis that all pricing errors are jointly zero. HJ refers to the Hansen and Jagannathan (1997) distance (with simulated p-value in brackets) for the null hypothesis that the HJ distance is equal to zero. *Panel B* reports least-squares estimates of time series regressions with Newey and West (1987) and Andrews (1991) standard errors in parentheses. Excess returns are in annual terms and adjusted for transaction costs. The portfolios are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction, and Figure 1 for the portfolio construction.

Panel A: Factor Prices

	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ
	CHF as pricing currency								*DEM-EUR as pricing currency*							
GMM_1	−0.03	0.97	0.01	0.04	0.78	2.22%	4.08	0.12	0.13	0.91	0.00	0.05	0.81	2.33%	5.92	0.14
	(0.22)	(0.34)	(0.01)	(0.01)			[0.40]	[0.30]	(0.29)	(0.31)	(0.01)	(0.01)			[0.21]	[0.15]
GMM_2	−0.03	0.94	0.01	0.05	0.95	2.37%	4.06		0.04	0.87	0.00	0.05	0.96	2.56%	5.79	
	(0.22)	(0.29)	(0.01)	(0.01)			[0.40]		(0.29)	(0.27)	(0.01)	(0.01)			[0.22]	
FMB	−0.03	0.96	0.01	0.04	0.78	2.22%	4.08		0.13	0.91	0.00	0.05	0.81	2.33%	5.92	
	(0.21)	(0.28)	(0.01)	(0.01)			[0.40]		(0.28)	(0.27)	(0.01)	(0.01)			[0.21]	
	[0.21]	[0.26]	[0.01]	[0.01]					[0.27]	[0.24]	[0.01]	[0.01]				

Panel B: Factor Betas

	α	β_{DOL}	β_{IMB}	R^2	α	β_{DOL}	β_{IMB}	R^2
P_1	−0.01	0.94	−0.27	0.78	−0.01	0.86	−0.31	0.71
	(0.01)	(0.04)	(0.05)		(0.01)	(0.05)	(0.04)	
P_2	−0.02	0.93	−0.15	0.81	−0.02	0.88	−0.13	0.73
	(0.01)	(0.05)	(0.04)		(0.01)	(0.06)	(0.03)	
P_3	−0.01	0.89	−0.03	0.78	−0.01	0.82	−0.02	0.67
	(0.01)	(0.03)	(0.04)		(0.01)	(0.04)	(0.03)	
P_4	−0.02	0.95	0.21	0.78	−0.02	0.93	0.20	0.65
	(0.01)	(0.04)	(0.06)		(0.01)	(0.06)	(0.05)	
P_5	0.00	1.25	0.27	0.74	0.01	1.42	0.27	0.69
	(0.01)	(0.05)	(0.09)		(0.01)	(0.07)	(0.08)	

continued

Table A.7. Asset Pricing Tests and Other Pricing Currencies *(continued)*

Panel A: Factor Prices

	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ
	GBP as pricing currency								*JPY as pricing currency*							
GMM_1	−0.12	1.02	0.00	0.05	0.87	2.10	5.97	0.15	0.01	1.05	0.02	0.05	0.73	2.31	6.43	0.15
	(0.24)	(0.34)	(0.01)	(0.01)			[0.20]	[0.19]	(0.18)	(0.30)	(0.02)	(0.01)			[0.17]	0.16
GMM_2	−0.17	1.09	−0.01	0.05	0.95	2.12	5.89		0.08	0.95	0.02	0.05	0.89	2.39	6.08	
	(0.22)	(0.30)	(0.01)	(0.01)			[0.21]		(0.17)	(0.28)	(0.02)	(0.01)			[0.19]	
FMB	−0.12	1.02	0.00	0.05	0.87	2.10	5.97		0.01	1.05	0.02	0.05	0.73	2.31	6.42	
	(0.21)	(0.28)	(0.01)	(0.01)			[0.20]		(0.17)	(0.28)	(0.02)	(0.01)			[0.17]	
	[0.21]	[0.26]	[0.01]	[0.01]					[0.15]	[0.26]	[0.02]	[0.01]				

Panel B: Factor Betas

	α	β_{DOL}	β_{IMB}	R^2	α	β_{DOL}	β_{IMB}	R^2
P_1	−0.01	1.02	−0.32	0.81	−0.01	0.93	−0.23	0.89
	(0.01)	(0.03)	(0.04)		(0.01)	(0.02)	(0.04)	
P_2	−0.02	1.00	−0.19	0.84	−0.02	0.99	−0.14	0.92
	(0.01)	(0.03)	(0.03)		(0.01)	(0.01)	(0.03)	
P_3	−0.01	0.98	−0.07	0.81	−0.01	1.00	−0.07	0.90
	(0.01)	(0.03)	(0.03)		(0.01)	(0.02)	(0.03)	
P_4	−0.02	0.98	0.23	0.77	−0.02	1.02	0.17	0.89
	(0.01)	(0.03)	(0.06)		(0.01)	(0.02)	(0.05)	
P_5	0.00	1.00	0.35	0.62	0.01	1.07	0.27	0.80
	(0.01)	(0.06)	(0.08)		(0.01)	(0.03)	(0.09)	

Table A.8. Descriptive Statistics of Global Imbalance Portfolios: Subset of Currencies

The table presents descriptive statistics of currency portfolios sorted on time $t-1$ net foreign asset position to gross domestic product (nfa), and the share of foreign liabilities in domestic currency (ldc). The first portfolio (P_1) contains the top 20% of all currencies with high nfa and high ldc (creditor nations with external liabilities mainly in domestic currency) whereas the last portfolio (P_5) contains the top 20% of all currencies with low nfa and low ldc (debtor nations with external liabilities mainly in foreign currency). IMB is a long-short strategy that buys P_5 and sells P_1. The table also reports the first order autocorrelation coefficient (ac_1), the annualized Sharpe ratio (SR), the maximum drawdown (mdd), the frequency of portfolio switches ($freq$), the average forward discount (fd), the average nfa, and the average ldc. t-statistics based on Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in brackets. Excess returns are expressed in percentage per annum and adjusted for transaction costs. The sample *Developed & Emerging Currencies* comprises the currencies of developed economies plus the most liquid emerging market currencies. The portfolio are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction, and Figure 1 for a detailed description of the portfolio construction.

	P_1	P_2	P_3	P_4	P_5	IMB
	Developed & Emerging Currencies					
$mean$	1.18	2.58	2.08	3.49	7.62	6.45
t-$stat$	[0.69]	[1.46]	[1.36]	[2.01]	[3.89]	[4.21]
$median$	1.13	1.99	2.59	4.23	8.85	7.85
$sdev$	9.14	9.63	8.33	9.05	10.55	6.76
$skew$	0.04	−0.09	−0.40	−0.50	−0.71	−0.60
$kurt$	3.11	3.69	4.73	5.91	5.09	6.41
ac_1	0.05	0.04	0.04	0.08	0.06	0.18
SR	0.13	0.27	0.25	0.39	0.72	0.95
mdd	0.51	0.37	0.32	0.32	0.31	0.28
$freq$	0.02	0.03	0.03	0.03	0.02	
fd	−0.66	0.06	2.06	4.05	7.74	
nfa	0.26	0.26	0.10	−0.42	−0.41	
ldc	0.64	0.48	0.44	0.50	0.33	

Table A.9. Asset Pricing Tests and Global Imbalance Risk: Subset of Currencies

The table presents cross-sectional asset pricing results for the linear factor model based on the dollar (DOL) and the global imbalance (IMB) risk factor. The test assets are excess returns to five carry trade portfolios sorted on the one-month forward discounts. IMB is a long-short strategy that buys the currency (top 20%) of debtor nations with external liabilities mainly in foreign currency, and sells the currencies (top 20%) of creditor nations with external liabilities mainly in domestic currency. *Panel A* reports GMM (first and second-stage) and Fama-MacBeth (FMB) estimates of the factor loadings b, the market price of risk λ, and the cross-sectional R^2. Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in parentheses whereas Shanken (1992) standard errors are reported in brackets. χ^2 denotes the test statistics (with p-value in brackets) for the null hypothesis that all pricing errors are jointly zero. HJ refers to the Hansen and Jagannathan (1997) distance (with simulated p-value in brackets) for the null hypothesis that the HJ distance is equal to zero. *Panel B* reports least-squares estimates of time series regressions with Newey and West (1987) and Andrews (1991) standard errors in parentheses. Excess returns are in annual terms and adjusted for transaction costs. The portfolios are rebalanced monthly from October 1983 to June 2014. In *Developed & Emerging Currencies*, both test assets and risk factors are constructed using the currencies of developed economies as well as the most liquid emerging market currencies. See Section 3 for a detailed description of data sources and data construction.

Panel A: Factor Prices								
	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ
	Developed and Emerging Currencies							
GMM_1	0.09	1.12	0.03	0.06	0.90	1.69	2.91	0.11
	(0.24)	(0.37)	(0.02)	(0.02)			[0.57]	[0.67]
GMM_2	0.18	1.33	0.03	0.05	0.96	2.27	2.39	
	(0.23)	(0.32)	(0.02)	(0.02)			[0.66]	
FMB	0.09	1.11	0.03	0.06	0.90	1.69	2.92	
	(0.20)	(0.29)	(0.02)	(0.02)			[0.57]	
	[0.19]	[0.25]	[0.02]	[0.01]				
Panel B: Factor Betas								
		α	β_{DOL}	β_{IMB}	R^2			
P_1		−0.01	0.99	−0.26	0.78			
		(0.01)	(0.05)	(0.04)				
P_2		−0.02	1.00	−0.18	0.83			
		(0.01)	(0.04)	(0.03)				
P_3		0.00	1.04	−0.07	0.84			
		(0.01)	(0.03)	(0.03)				
P_4		−0.02	1.03	0.18	0.82			
		(0.01)	(0.04)	(0.05)				
P_5		0.01	0.93	0.39	0.72			
		(0.01)	(0.06)	(0.07)				

Table A.10. Descriptive Statistics: Currencies Sorted by Real Interest Rates

The table presents descriptive statistics of currency portfolios sorted on time $t-1$ real interest rate differentials (one-month forward discounts adjusted for inflation differentials). The first portfolio (P_1) contains the top 20% of all currencies with low real interest rate differentials whereas the last portfolio (P_5) contains the top 20% of all currencies with high interest rate differentials. H/L is a long-short strategy that buys P_5 and sells P_1. The table also reports the first order autocorrelation coefficient (ac_1), the annualized Sharpe ratio (SR), the maximum drawdown (mdd), the frequency of portfolio switches ($freq$), the average forward discounts (fd), the average net foreign asset position to gross domestic product (nfa), and the average share of foreign liabilities in domestic currency (ldc). t-statistics based on Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in brackets. Excess returns are expressed in percentage per annum and adjusted for transaction costs. The portfolio are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction.

	P_1	P_2	P_3	P_4	P_5	H/L	P_1	P_2	P_3	P_4	P_5	H/L
			All Countries						*Developed Countries*			
mean	−0.25	0.84	2.28	3.29	4.82	5.07	0.77	0.77	1.22	4.11	4.64	3.87
t-stat	[−0.17]	[0.58]	[1.47]	[2.06]	[2.55]	[3.02]	[0.43]	[0.43]	[0.68]	[2.28]	[2.04]	[1.79]
median	0.72	1.64	3.56	3.74	5.63	7.23	−0.71	0.96	2.56	5.28	6.39	8.28
sdev	8.01	7.77	8.13	8.30	9.15	8.36	9.90	9.83	9.18	9.80	11.27	11.19
skew	0.00	−0.22	−0.30	−0.39	−0.29	−0.55	0.32	0.03	−0.20	−0.30	−0.28	−0.94
kurt	4.08	3.90	4.22	4.64	5.11	4.97	3.80	3.65	4.12	4.39	4.29	5.16
ac_1	0.00	0.06	0.07	0.08	0.15	0.15	−0.01	0.00	0.10	0.04	0.11	0.07
SR	−0.03	0.11	0.28	0.40	0.53	0.61	0.08	0.08	0.13	0.42	0.41	0.35
mdd	0.41	0.42	0.32	0.29	0.31	0.35	0.50	0.50	0.38	0.26	0.38	0.41
freq	0.17	0.26	0.26	0.25	0.13		0.14	0.26	0.30	0.25	0.15	
fd	−1.35	−0.75	1.10	3.13	8.65		−1.98	−0.91	0.28	1.58	3.93	
nfa	0.40	0.05	−0.12	−0.29	−0.34		0.57	−0.01	−0.08	−0.10	−0.37	
ldc	0.53	0.49	0.44	0.42	0.39		0.55	0.51	0.49	0.40	0.42	

Table A.11. Asset Pricing Tests: Test Assets Sorted by Real Interest Rates

The table presents cross-sectional asset pricing results for the linear factor model based on the dollar (DOL) and the global imbalance (IMB) risk factor. The test assets are excess returns to five currency portfolios sorted on the one-month real interest rate differentials. IMB is a long-short strategy that buys the currency (top 20%) of debtor nations with external liabilities mainly in foreign currency, and sells the currencies (top 20%) of creditor nations with external liabilities mainly in domestic currency. *Panel A* reports GMM (first and second-stage) and Fama-MacBeth (FMB) estimates of the factor loadings b, the market price of risk λ, and the cross-sectional R^2. Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in parentheses whereas Shanken (1992) standard errors are reported in brackets. χ^2 denotes the test statistics (with p-value in brackets) for the null hypothesis that all pricing errors are jointly zero. HJ refers to the Hansen and Jagannathan (1997) distance (with simulated p-value in brackets) for the null hypothesis that the HJ distance is equal to zero. *Panel B* reports least-squares estimates of time series regressions with Newey and West (1987) and Andrews (1991) standard errors in parentheses. Excess returns are in annual terms and adjusted for transaction costs. The portfolios are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction.

Panel A: Factor Prices																
	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ
	All Countries								*Developed Countries*							
GMM_1	0.06	1.01	0.02	0.05	0.85	1.73%	5.48	0.13	0.18	0.72	0.02	0.04	0.74	2.09%	6.30	0.13
	(0.26)	(0.32)	(0.01)	(0.01)			[0.24]	[0.34]	(0.21)	(0.33)	(0.02)	(0.01)			[0.18]	[0.30]
GMM_2	0.07	0.71	0.02	0.08	0.98	3.29%	4.25		0.26	0.68	0.02	0.04	0.81	2.09%	6.09	
	(0.25)	(0.27)	(0.01)	(0.01)			[0.37]		(0.21)	(0.26)	(0.02)	(0.01)			[0.19]	
FMB	0.06	1.01	0.02	0.05	0.85	1.73%	5.46		0.17	0.72	0.02	0.04	0.74	2.09%	6.30	
	(0.24)	(0.28)	(0.01)	(0.01)			[0.24]		(0.19)	(0.22)	(0.02)	(0.01)			[0.18]	
	[0.22]	[0.26]	[0.01]	[0.01]					(0.17]	[0.25]	[0.02]	[0.01]				
Panel B: Factor Betas																
		α	β_{DOL}	β_{IMB}	R^2					α	β_{DOL}	β_{IMB}	R^2			
P_1		−0.02	1.03	−0.25	0.80					0.00	0.95	−0.49	0.74			
		(0.01)	(0.04)	(0.04)						(0.01)	(0.05)	(0.06)				
P_2		−0.01	1.03	−0.22	0.86					−0.02	1.02	−0.06	0.82			
		(0.01)	(0.03)	(0.03)						(0.01)	(0.04)	(0.04)				
P_3		0.00	1.06	−0.07	0.87					−0.01	0.97	−0.05	0.85			
		(0.01)	(0.04)	(0.04)						(0.01)	(0.03)	(0.04)				
P_4		0.00	0.99	0.11	0.81					0.01	1.00	0.14	0.83			
		(0.01)	(0.05)	(0.06)						(0.01)	(0.03)	(0.05)				
P_5		0.00	0.86	0.44	0.69					0.00	0.98	0.51	0.73			
		(0.01)	(0.05)	(0.07)						(0.01)	(0.05)	(0.06)				

Table A.12. Descriptive Statistics: Even Number of Global Imbalance Portfolios

The table presents descriptive statistics of currency portfolios sorted on time $t-1$ net foreign asset position to gross domestic product (nfa), and the share of foreign liabilities in domestic currency (ldc). The first portfolio contains the top 20% of all currencies with high nfa and high ldc (creditor nations with external liabilities mainly in domestic currency) whereas the last portfolio contains the top 20% of all currencies with low nfa and low ldc (debtor nations with external liabilities mainly in foreign currency). IMB is a long-short strategy that buys the last portfolio and sells the first portfolio. The table also reports the first order autocorrelation coefficient (ac_1), the annualized Sharpe ratio (SR), the maximum drawdown (mdd), the frequency of portfolio switches ($freq$), the average forward discount (fd), the average nfa, and the average ldc. t-statistics based on Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in brackets. Excess returns are expressed in percentage per annum and adjusted for transaction costs. The portfolio are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction. Figure A.1 provides a detailed description of the portfolio construction.

	P_1	P_2	P_3	P_4	P_5	P_6	IMB	P_1	P_2	P_3	P_4	IMB
	All Countries							*Developed Countries*				
$mean$	0.61	2.84	3.44	1.76	3.08	5.74	5.12	0.75	2.83	3.02	4.36	3.61
t-$stat$	[0.43]	[1.67]	[2.27]	[1.29]	[1.83]	[3.00]	[3.82]	[0.40]	[1.55]	[1.83]	[2.24]	[2.90]
$median$	0.88	2.33	3.70	4.71	4.64	6.39	6.04	0.67	3.34	4.54	5.58	4.30
$sdev$	7.38	9.18	8.16	7.17	9.31	9.82	6.54	9.75	9.98	8.91	10.18	6.90
$skew$	−0.13	−0.05	−0.21	−0.48	−1.33	−0.03	0.50	0.05	−0.17	−0.22	−0.26	−0.57
$kurt$	3.36	4.38	4.14	5.45	10.38	4.68	6.77	3.68	3.24	5.32	3.70	5.25
ac_1	0.08	0.05	0.06	0.07	0.01	0.10	0.13	0.07	0.03	0.05	0.07	0.01
SR	0.08	0.31	0.42	0.25	0.33	0.58	0.78	0.08	0.28	0.34	0.43	0.52
mdd	0.49	0.31	0.24	0.34	0.36	0.29	0.22	0.56	0.27	0.32	0.34	0.26
$freq$	0.03	0.04	0.04	0.04	0.04	0.03	0.06	0.02	0.02	0.02	0.03	0.05
fd	−0.71	0.79	1.73	2.47	3.97	6.89		−1.04	0.19	2.09	2.29	
nfa	0.42	0.19	0.26	−0.23	−0.45	−0.59		0.35	0.31	−0.38	−0.35	
ldc	0.64	0.51	0.37	0.52	0.44	0.27		0.60	0.40	0.55	0.35	

Table A.13. Asset Pricing Tests: Global Imbalance Risk

The table presents cross-sectional asset pricing results for the linear factor model based on the dollar (DOL) and the global imbalance (IMB) risk factor. The test assets are excess returns to five carry trade portfolios sorted on the one-month forward discounts, and are the same as in Table 2. IMB is a long-short strategy that buys the currency of debtor nations with external liabilities mainly in foreign currency, and sells the currencies of creditor nations with external liabilities mainly in domestic currency, and is based on six (four) portfolios for *All* (*Developed*) *Countries*. *Panel A* reports GMM (first and second-stage) and Fama-MacBeth (FMB) estimates of the factor loadings b, the market price of risk λ, and the cross-sectional R^2. Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in parentheses whereas Shanken (1992) standard errors are reported in brackets. χ^2 denotes the test statistics (with p-value in brackets) for the null hypothesis that all pricing errors are jointly zero. HJ refers to the Hansen and Jagannathan (1997) distance (with simulated p-value in brackets) for the null hypothesis that the HJ distance is equal to zero. *Panel B* reports least-squares estimates of time series regressions with Newey and West (1987) and Andrews (1991) standard errors in parentheses. Excess returns are in annual terms and adjusted for transaction costs. The portfolios are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction. Figure A.1 provides a detailed description of the portfolio construction.

Panel A: Factor Prices

	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ	b_{DOL}	b_{IMB}	λ_{DOL}	λ_{IMB}	R^2	$RMSE$	χ^2	HJ
	All Countries								*Developed Countries*							
GMM_1	−0.10	1.62	0.02	0.08	0.83	1.89	4.70	0.15	0.17	0.80	0.02	0.05	0.90	1.16	1.24	0.07
	(0.29)	(0.59)	(0.01)	(0.02)			[0.20]	[0.12]	(0.22)	(0.44)	(0.02)	(0.02)			[0.74]	[0.76]
GMM_2	−0.04	1.61	0.02	0.07	0.68	1.96	4.65		0.21	0.91	0.02	0.05	0.86	1.16	1.14	
	(0.29)	(0.58)	(0.01)	(0.02)			[0.20]		(0.21)	(0.43)	(0.02)	(0.02)			[0.77]	
FMB	−0.10	1.62	0.02	0.08	0.83	1.89	4.70		0.17	0.80	0.02	0.05	0.90	1.16	1.24	
	(0.26)	(0.48)	(0.01)	(0.02)			[0.19]		(0.19)	(0.37)	(0.02)	(0.02)			[0.74]	
	[0.24]	[0.47]	[0.01]	[0.02]					[0.18]	[0.35]	[0.02]	[0.02]				

Panel B: Factor Betas

	α	β_{DOL}	β_{IMB}	R^2	α	β_{DOL}	β_{IMB}	R^2
P_1	−0.01	0.99	−0.26	0.78	0.00	1.00	−0.48	0.76
	(0.01)	(0.05)	(0.04)		(0.01)	(0.04)	(0.06)	
P_2	−0.02	1.00	−0.18	0.83	−0.01	1.04	−0.17	0.82
	(0.01)	(0.04)	(0.03)		(0.01)	(0.04)	(0.04)	
P_3	0.00	1.04	−0.07	0.84	−0.01	0.99	0.00	0.86
	(0.01)	(0.03)	(0.03)		(0.01)	(0.03)	(0.03)	
P_4	−0.02	1.03	0.18	0.82	−0.01	0.99	0.15	0.82
	(0.01)	(0.04)	(0.05)		(0.01)	(0.03)	(0.05)	
P_5	0.01	0.93	0.39	0.72	0.01	1.06	0.50	0.78
	(0.01)	(0.06)	(0.07)		(0.01)	(0.04)	(0.05)	

Table A.14. Decomposing the Global Imbalance Risk Factor

The table presents descriptive statistics of currency portfolios sorted on time $t-1$ information. For *All Countries* (and *Developed & Emerging Countries*), currencies are first grouped into two baskets using the net foreign asset position to gross domestic product (nfa), and then into three baskets using the share of foreign liabilities in domestic currency (ldc). The NFA factor is constructed as the average return on the low nfa portfolios (P_3, P_4 and P_5) minus the average return on the high nfa portfolios (P_1, P_2 and P_3). The LDC factor is computed as the average return on the low ldc portfolios (P_3, and P_6) minus the average return on the high ldc portfolios (P_1 and P_4). For *Developed Countries*, currencies are first grouped into two baskets using nfa, and then into two baskets using ldc. The NFA factor is constructed as the average return on the low nfa portfolios (P_3 and P_4) minus the average return on the high nfa portfolios (P_1 and P_2). The LDC factor is computed as the average return on the low ldc portfolios (P_2, and P_4) minus the average return on the high ldc portfolios (P_1, and P_3). The table also reports the first order autocorrelation coefficient (AC_1), the annualized Sharpe ratio (SR), the maximum drawdown in percent (MDD), the frequency of portfolio switches ($freq$), the average forward discount (fd), the average nfa, and the average ldc. t-statistics based on Newey and West (1987) standard errors with Andrews (1991) optimal lag selection are reported in brackets. Excess returns are expressed in percentage per annum. The portfolio are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction. Figure A.1 provides a detailed description of the portfolio construction.

	P_1	P_2	P_3	P_4	P_5	P_6	NFA	LDC	P_1	P_2	P_3	P_4	NFA	LDC
	All Countries								*Developed Countries*					
mean	0.55	2.92	3.52	1.85	3.21	5.90	1.32	3.51	0.70	2.90	3.09	4.43	1.96	1.77
t-stat	[0.38]	[1.71]	[2.33]	[1.35]	[1.91]	[3.09]	[1.46]	[3.82]	[0.37]	[1.58]	[1.87]	[2.27]	[1.86]	[2.11]
median	0.88	2.35	3.70	4.72	4.64	6.39	1.55	3.72	0.67	3.34	4.63	5.76	2.24	2.26
sdev	7.37	9.18	8.17	7.17	9.31	9.82	4.80	4.96	9.75	9.98	8.91	10.19	5.83	4.63
skew	−0.13	−0.05	−0.21	−0.47	−1.33	−0.04	−0.44	0.10	0.05	−0.17	−0.22	−0.26	−0.40	−0.05
kurt	3.36	4.38	4.15	5.45	10.41	4.69	3.88	4.58	3.68	3.25	5.32	3.71	3.88	3.39
ac_1	0.09	0.05	0.05	0.07	0.01	0.09	0.06	0.05	0.07	0.03	0.05	0.07	−0.01	−0.01
SR	0.07	0.32	0.43	0.26	0.34	0.60	0.28	0.71	0.07	0.29	0.35	0.43	0.34	0.38
mdd	0.49	0.30	0.24	0.33	0.36	0.29	0.27	0.15	0.56	0.27	0.31	0.34	0.20	0.14
$freq$	0.03	0.04	0.04	0.04	0.04	0.03			0.02	0.02	0.02	0.03		
fd	−0.76	0.84	1.80	2.53	4.07	7.03			−1.09	0.24	2.15	2.34		
nfa	0.42	0.19	0.26	−0.23	−0.45	−0.59			0.35	0.31	−0.38	−0.35		
ldc	0.64	0.51	0.37	0.52	0.44	0.27			0.60	0.40	0.55	0.35		
	Developed & Emerging Countries													
mean	0.91	2.22	2.54	2.15	4.17	8.59	3.08	4.03						
t-stat	[0.54]	[1.18]	[1.57]	[1.30]	[1.95]	[4.34]	[3.04]	[4.03]						
median	1.31	1.77	2.87	5.33	5.32	10.33	2.89	4.44						
sdev	8.82	10.17	8.81	8.58	11.55	10.67	5.22	5.16						
skew	0.08	−0.06	−0.18	−0.61	−0.84	−0.68	−0.52	0.02						
kurt	3.31	3.80	3.91	6.55	6.39	4.93	4.03	3.96						
ac_1	0.06	0.05	0.04	0.08	0.05	0.05	0.09	0.10						
SR	0.10	0.22	0.29	0.25	0.36	0.81	0.59	0.78						
mdd	0.56	0.39	0.26	0.32	0.42	0.33	0.18	0.16						
$freq$	0.03	0.03	0.03	0.03	0.03	0.02								
fd	−0.68	−0.28	1.21	2.86	5.25	8.32								
nfa	0.22	0.32	0.35	−0.36	−0.39	−0.42								
ldc	0.65	0.51	0.36	0.55	0.44	0.30								

Table A.15. Asset Pricing and Independent Double Sort

The table presents cross-sectional asset pricing results for the linear factor model based on the dollar (DOL), the net foreign asset (NFA), and the share of foreign liability in domestic currency(LDC) risk factors. The test assets are excess returns to five carry trade portfolios sorted on the one-month forward discounts (the same as in Table 2). NFA and LDC are described in Table A.15. *Panel A* reports GMM (first and second-stage) and Fama-MacBeth (FMB) estimates of the factor loadings b, the market price of risk λ, and the cross-sectional R^2 (b and λ for the DOL are statistically insignificant, and we do not report them to save space). χ^2 denotes the test statistics for the null hypothesis that all pricing errors are jointly zero whereas HJ refers to the Hansen and Jagannathan (1997) distance for the null hypothesis that the HJ distance is equal to zero. We bold the statistic when fail to reject the null at 5% significance level. *D-test* denotes the χ^2 difference test for the null hypothesis that the restricted model ($b_{NFA} = 0$ or $b_{LDC} = 0$) is correct (with p-values based on Newey and West (1987) and Andrews (1991) in brackets). *Panel B* reports least-squares estimates of time series regressions. The superscript a, b, and c indicate statistical significance at 10%, 5% and 1% level, respectively, using Newey and West (1987) standard errors with Andrews (1991) optimal lag selection. Excess returns are in annual terms and adjusted for transaction costs. The portfolios are rebalanced monthly from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction, and Figure A.1 for the portfolio construction.

Panel A: Factor Prices

	NFA	LDC	R^2	χ^2	HJ	NFA	LDC	R^2	χ^2	HJ	NFA	LDC	R^2	χ^2	HJ
	All Countries					*Developed & Emerging*					*Developed Countries*				
b_{GMM1}	1.10^{b}	1.56^{c}	0.80	**6.28**	**0.17**	0.93^{b}	1.26^{c}	0.92	**1.64**	**0.08**	0.70^{a}	0.63	0.87	**2.51**	**0.09**
λ_{GMM1}	0.02^{b}	0.04^{c}				0.03^{c}	0.04^{c}				0.02^{b}	0.02^{b}			
b_{GMM2}	0.85^{b}	1.41^{c}	0.78	**5.65**		1.13^{c}	1.34^{c}	0.95	**1.27**		0.59^{a}	0.55	0.92	**2.35**	
λ_{GMM2}	0.02^{b}	0.05^{c}				0.02^{b}	0.04^{c}				0.03^{b}	−0.01			
b_{FMB}	1.09^{c}	1.56^{c}	0.80	**6.28**		0.93^{c}	1.25^{c}	0.92	**1.65**		0.69^{b}	0.63^{a}	0.87	**2.49**	
λ_{FMB}	0.02^{b}	0.05^{c}				0.03^{c}	0.04^{c}				0.02^{b}	0.02^{b}			
D-test	[0.03]	[< .01]				[< .01]	[< .01]				[0.04]	[0.12]			

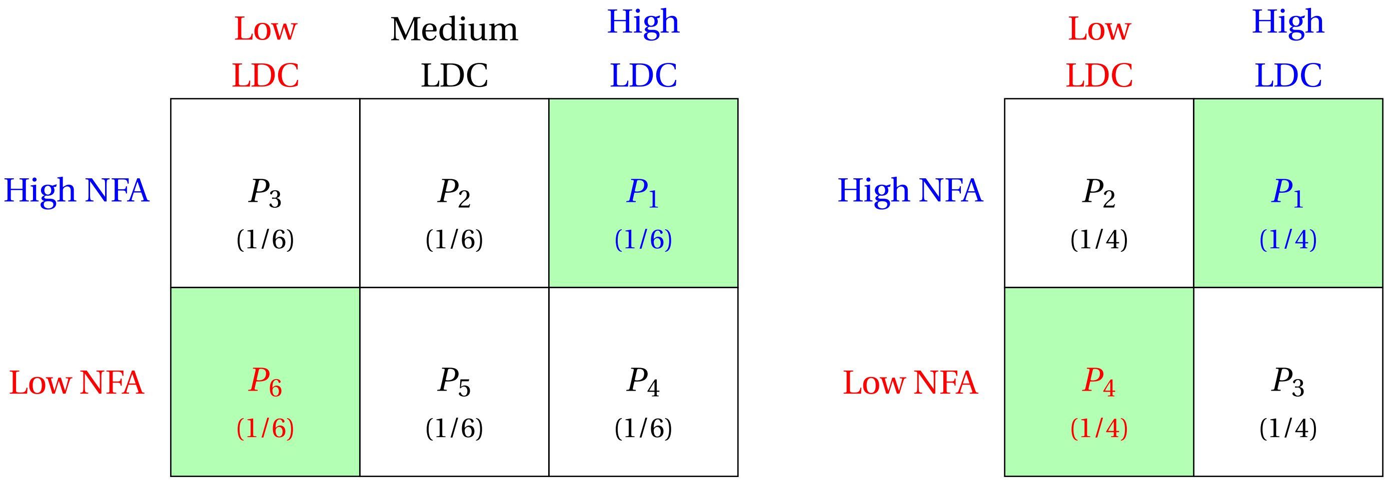
Panel B: Factor Betas

	α	β_{DOL}	β_{NFA}	β_{LDC}	R^2	α	β_{DOL}	β_{NFA}	β_{LDC}	R^2	α	β_{DOL}	β_{NFA}	β_{LDC}	R^2
	All Countries					*Developed Countries*					*Developed Countries*				
P_1	-0.01^{a}	0.95^{c}	-0.47^{c}	-0.17^{b}	0.81	0.01	0.94^{c}	-0.63^{c}	-0.14^{a}	0.82	0.01	0.90^{c}	-0.59^{c}	-0.24^{c}	0.76
P_2	-0.02^{c}	1.02^{c}	-0.12^{b}	-0.27^{c}	0.83	-0.01^{a}	1.00^{c}	-0.28^{c}	-0.12^{b}	0.84	-0.01^{a}	1.00^{c}	-0.18^{c}	−0.12	0.82
P_3	0.01	1.04^{c}	-0.14^{c}	-0.08^{a}	0.85	−0.01	1.00^{c}	-0.12^{c}	0.05	0.88	−0.01	0.95^{c}	−0.04	0.08	0.86
P_4	-0.01^{a}	1.02^{c}	0.08	0.18^{b}	0.82	-0.02^{a}	1.11^{c}	0.22^{c}	0.06	0.86	−0.01	1.00^{c}	0.22^{c}	0.04	0.84
P_5	0.01	0.97^{c}	0.65^{c}	0.29^{c}	0.76	0.00	0.96^{c}	0.87^{c}	0.15^{b}	0.79	0.01	1.14^{c}	0.74^{c}	0.08	0.81

Table A.16. Global Equity Realized Volatility

The table presents results from fixed-effects panel regressions. We use discrete exchange rate returns at monthly frequency as dependent variables. Exchange rates are defined as units of US dollars per unit of foreign currency such that a positive return denotes a foreign currency appreciation. The set of independent variables includes the net foreign asset position to gross domestic product (nfa), the share of foreign liabilities in domestic currency (ldc), the forward discount or interest rate differential relative to the US (fd), the monthly change in the realized volatility of the MSCI World index ($\Delta RVOL$), and a dummy variable that equals one if $\Delta RVOL$ is greater than one standard deviation as estimated across the entire sample, and zero otherwise ($\Delta RVOL$ *dummy*). Robust standard errors are clustered at country level and reported in parentheses. The superscripts a, b and c denote statistical significance at 10%, 5% and 1% level, respectively. The sample runs from October 1983 to June 2014. See Section 3 for a detailed description of data sources and data construction.

	Dependent variable: nominal exchange rate returns					
	(1)	(2)	(3)	(4)	(5)	(6)
nfa (lagged 12 months)	−0.057	−0.053	−0.032	−0.092	−0.091	−0.073
	(0.073)	(0.076)	(0.078)	(0.064)	(0.066)	(0.066)
$\Delta RVOL$	-0.069^c	-0.069^c	-0.061^c			
	(0.010)	(0.010)	(0.010)			
$\Delta RVOL \times nfa$ (lagged 12 months)	0.033^b	0.033^b	0.027^b			
	(0.012)	(0.012)	(0.013)			
ldc (lagged 12 months)		−0.089	−0.286		−0.027	−0.167
		(0.243)	(0.219)		(0.243)	(0.221)
fd (lagged 1 month)			−0.003			−0.002
			(0.002)			(0.002)
$\Delta RVOL \times fd$ (lagged 1 month)			−0.001			
			(< .001)			
$\Delta RVOL$ *dummy*				-1.236^c	-1.236^c	-1.175^c
				(0.212)	(0.213)	(0.234)
$\Delta RVOL$ *dummy* $\times nfa$ (lagged 12 months)				0.682^c	0.682^c	0.640^c
				(0.241)	(0.241)	(0.249)
$\Delta RVOL$ *dummy* $\times fd$ (lagged 1 month)						−0.003
						(0.004)
Additional Variables: Constant and lagged exchange rate returns	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
Adjusted R^2	0.03	0.03	0.03	0.02	0.02	0.02
Observations	9112	9112	9112	9112	9112	9112



(a) All Countries

(b) Developed Countries

Figure A.1. Global Imbalance Risk Factor with an Even Number of Portfolios

This chart describes the construction of the global imbalance risk factor (IMB) when we use an even number of portfolios. Figure a) reports the construction for *all countries*. At the end of each month, currencies are first sorted in two baskets using the net foreign asset position to gross domestic product (nfa), and then in 3 baskets using the share of foreign liabilities in domestic currency (ldc). The nfa breakpoint is the median value whereas the ldc breakpoints are the 33^{th} and 63^{th} percentiles. The IMB factor is constructed as the average return on P_6 - the portfolio with low nfa and low ldc (debtor nations with external liabilities mainly in foreign currency) - minus the average return on P_1 - the portfolio with high nfa and high ldc currencies (creditor nations with external liabilities mainly in domestic currency). The global imbalance risk factor is then decomposed into the NFA factor computed as the average return on the low nfa portfolios (P_3, P_4 and P_5) minus the average return on the high nfa portfolios (P_1, P_2 and P_3), and the LDC factor constructed as the average return on the low ldc portfolios (P_3 and P_6) minus the average return on the high ldc portfolios (P_1 and P_4). Figure b) reports the construction for *developed countries*. At the end of each month, currencies are first sorted first in two baskets using nfa, and then in two baskets using ldc. The breakpoints for both nfa and ldc are the median values. The IMB factor is constructed as the average return on P_4 minus the average return on P_1. The NFA factor is constructed as the average return on low nfa portfolios (P_3 and P_4) minus the average return on the high nfa portfolios (P_1 and P_2), and the LDC factor is computed as the average return on the low ldc portfolios (P_2 and P_4) minus the average return on the high ldc portfolios (P_1 and P_3).

Journal of Financial Economics 137 (2020) 659–678

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Business cycles and currency returns☆

Riccardo Colacito[a,b,*], Steven J. Riddiough[c], Lucio Sarno[d,e]

[a] Kenan-Flagler Business School, University of North Carolina at Chapel Hill, McColl Building 4100, Chapel Hill, NC 27599-3490, USA
[b] National Bureau of Economic Research (NBER), USA
[c] Faculty of Business and Economics, University of Melbourne, 198 Berkeley Street, Carlton, VIC 3010, Australia
[d] Cambridge Judge Business School, University of Cambridge, Trumpington Street, Cambridge CB2 1AG, UK
[e] Centre for Economic Policy Research (CEPR), UK

ARTICLE INFO

Article history:
Received 14 January 2019
Revised 26 July 2019
Accepted 21 August 2019
Available online 4 May 2020

JEL classification:
G12
G15
F31

Keywords:
Exchange rates
Currency risk premium
Business cycles
Long-run risk

ABSTRACT

We find a strong link between currency excess returns and the relative strength of the business cycle. Buying currencies of strong economies and selling currencies of weak economies generates high returns both in the cross-section and time series of countries. These returns stem primarily from spot exchange rate predictability, are uncorrelated with common currency investment strategies, and cannot be understood using traditional currency risk factors in either unconditional or conditional asset pricing tests. We also show that a business cycle factor implied by our results is priced in a broad currency cross section.

1. Introduction

A core issue in asset pricing is the need to understand the relationship between fundamental macroeconomic conditions and asset market returns (Cochrane, 2005; 2017). Nowhere is this more central, and yet consistently difficult to establish, than in the foreign exchange (FX) market, in which currency returns and country-level fundamentals are highly correlated in theory, and yet the empirical relationship is typically found to be weak (Rossi, 2013). A recent literature in macro-finance has shown,

☆ We are grateful for comments and suggestions to Bill Schwert (editor), an anonymous referee, Antje Berndt, Magnus Dahlquist, Pasquale Della Corte, Antonio Gargano, Federico Gavazzoni, Michael Halling, Kai Li, Matteo Maggiori, Nelson Mark, Thomas Maurer, Michael Moore, Anella Munro, Dagfinn Rime, Maik Schmeling, Thomas Stolper, Andrea Tamoni, Raman Uppal, Adrien Verdelhan, Mungo Wilson, Josef Zechner and to other participants at presentations held at: 2018 American Finance Association; 2017 Auckland Finance Conference; 2017 China International Conference in Finance; 2017 Wellington Finance Summit; 2017 FIRN Annual Conference; 6th Workshop on "Financial Determinants of Exchange Rates," Bank of England; 2016 Australian National University Annual Summer Camp; the 9th INQUIRE Business School Workshop, Cass Business School; BI Norwegian Business School; University of Geneva; and University of Melbourne. We also thank Sidharth Dube, Harvey Huang, and Rachel Quan for excellent research assistance. Steven Riddiough gratefully acknowledges financial support from the Faculty of Business and Economics, University of Melbourne. The paper was the recipient of the CFA Institute Best Paper Prize 2017 FIRN Annual Conference and the Runner-up Best Paper Prize 2017 Vienna Symposium on Foreign Exchange Markets. All errors are ours.

* Corresponding author at: Kenan-Flagler Business School, University of North Carolina at Chapel Hill, McColl Building 4100, Chapel Hill, NC 27599-3490, USA.
E-mail address: ric@unc.edu (R. Colacito).

https://doi.org/10.1016/j.jfineco.2020.04.005
0304-405X/© 2020 Published by Elsevier B.V.

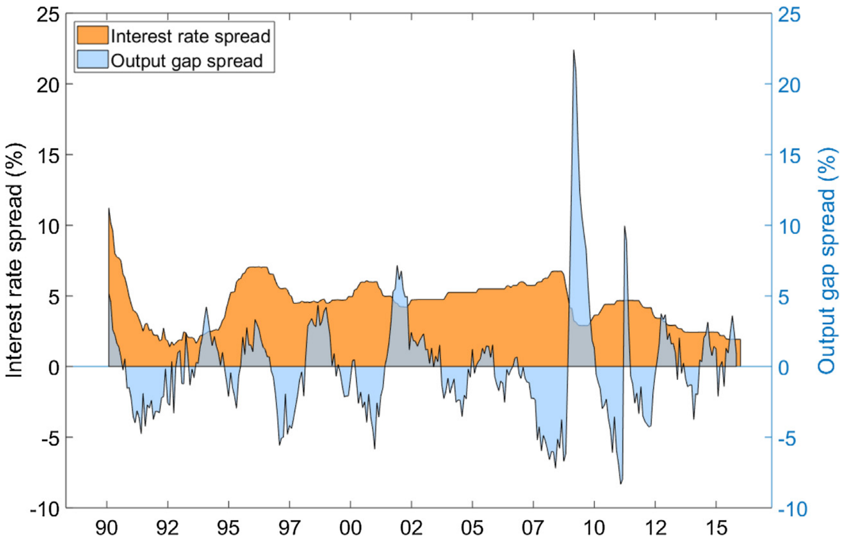


Fig. 1. Output gap and interest rate spreads between Australia and Japan. The figure plots the interest rate and output gap spread between Australia and Japan. The interest rates reflect one-month euro deposit rates, while the output gaps are calculated using the Hodrick-Prescott filter. When a series is above the origin, it indicates the Australian value is higher (i.e., either a higher interest rate or output gap). A description of the data and details on the construction of the output gap can be found in Section 3.

however, that the behavior of exchange rates becomes easier to explain once exchange rates are studied relative to one another in the cross-section rather than in isolation (Lustig et al., 2011; Verdelhan, 2018; Lustig and Richmond, 2019). This insight offers the tantalizing prospect that new empirical tests focusing on relative macroeconomic conditions across countries could reveal a stronger relationship between currency market returns and macroeconomic fundamentals.

In this paper, we take this empirical step by investigating the cross-sectional properties of currency returns to provide novel evidence on the relationship between currency returns and country-level macroeconomic conditions. We focus on the broadest measure of aggregate macroeconomic conditions—the business cycle—which constitutes a key building block in theoretical models of exchange rates. We find that business cycles are a key driver and powerful predictor of both currency excess returns and spot exchange rate fluctuations in the cross-section and this predictability can be understood from a risk-based perspective. This allows us to connect our findings to a broad literature that has analyzed the linkages between macro fundamentals and currency risk premia (see, inter alia, Colacito and Croce, 2011; Hassan, 2013; Gabaix and Maggiori, 2015; Ready et al., 2017; Berg and Mark, 2018; Colacito et al., 2018).

We measure macroeconomic conditions using the output gap, defined as the difference between a country's actual and potential level of output, for a broad sample of 27 developed and emerging market economies. Since it is not directly observable, we measure the output gap using industrial production data and apply several commonly adopted methods in the literature, including the filters proposed by Hodrick and Prescott (1980) and Baxter and King (1999), the quadratic time trend used by Clarida et al. (1998), and the linear projection method recently introduced by Hamilton (2018). We define the relative strength of the economy based on its position within the business cycle (i.e., whether it is nearer the trough (weak) or peak (strong) in the cycle). Using monthly data from 1983 to 2016, we find that sorting currencies into portfolios on the basis of the differential in output gaps relative to the US generates a monotonic increase in excess returns as we move from portfolios of "weak" to "strong" economy currencies. Thus our results imply that currency excess returns are higher for strong economies, a finding that we show to be robust to various ways of constructing currency portfolios.

Importantly, the predictability stemming from business cycles is quite different from other sources of cross-sectional predictability observed in the literature. Sorting currencies by output gaps is not equivalent, for example, to the currency carry trade that requires sorting currencies by their differentials in nominal interest rates. We highlight this point in Fig. 1 using two common carry trade currencies—the Australian dollar and Japanese yen. The interest rate differential is highly persistent and consistently positive between the two countries in recent decades. A carry trade investor would have thus always been long the Australian dollar and short the Japanese yen. In contrast the output gap differential varies substantially over time, and an output gap investor would have thus taken both long and short positions in the Australian dollar and Japanese yen as their relative business cycles fluctuated.

Moreover, we find that the cross-sectional predictability arising from business cycles stems primarily from the spot exchange rate component rather than from interest rate differentials. That is, currencies of strong economies tend to appreciate and those of weak economies tend to depreciate over the subsequent month. This feature makes the returns from exploiting business cycle information different from the returns delivered by most canonical currency investment strategies and most notably distinct from the carry trade, which generates a negative exchange rate return.

We initially calculate output gaps using the full time series of industrial production data observed in 2016. This exercise allows us to carefully show the relationship between relative macroeconomic conditions and exchange rates by exploiting the longest sample of data to formulate the most precise estimates of the output gap over time. Indeed, in the international economics literature it has been difficult to uncover a predictive link between macro fundamentals and exchange rates even when the econometrician is assumed to have perfect foresight of future macro fundamentals (Meese and Rogoff, 1983). However, the use of two-sided filters and revised data in the long sample also raises questions as to whether the relationship is exploitable in real time. We explore this question using a shorter sample of "vintage" data beginning in 1999 and find that the results are qualitatively identical. The vintage data mimics the information set available to investors, and thus sorting is conditional only on information available at the time. Between 1999 and 2016, a high-minus-low cross-sectional strategy that sorts on relative output gaps across countries, which we denote as GAP_{CS}, generates a Sharpe ratio of 0.72 before transaction costs and 0.50 after costs. The result also holds when assigning linear or rank weights (and thus trading all currencies simultaneously), which reassures us that the cross-sectional results are not driven by a few outlier currencies but apply generally to the broader cross-section. Moreover, a time-series strategy, which goes long (short) currencies issued by countries with output gaps above (below) the US, denoted as GAP_{TS}, generates a Sharpe ratio of 0.65 before costs and 0.50 after costs.

The two strategies, GAP_{CS} and GAP_{TS}, exhibit a correlation of around 35%, and thus the investment performance increases further once these strategies are combined.[1] Furthermore, the time-series correlations between the output gap strategies and the currency carry trade are found to be essentially zero, while the correlations with other canonical currency investment strategies—including "dollar carry" (Lustig et al., 2014), momentum (Menkhoff et al., 2012b), and value (Menkhoff et al., 2017)—are also low and close to zero.[2] This apparent lack of correlation implies that the output gap strategy offers useful diversification gains to an investor who adds it to a conventional menu of currency portfolios, and we quantify these gains in the empirical analysis.

We investigate whether the returns of output-gap-sorted portfolios reflect compensation for risk. Specifically, we test the pricing power of conventional risk factors using a battery of linear asset pricing models and do not find evidence that these pricing kernels can price the cross-section of currency returns sorted on output gaps. We then consider the possibility that business cycles proxy for a priced state variable as implied by many macro-finance models, giving rise to a "GAP risk premium." To do so, we consider the pricing power of a business cycle factor, taken to equal the returns on the GAP_{CS} strategy, and test whether it is priced in the cross-section of currencies. We find that the pricing power of the factor is strong and not confined to portfolios sorted on output gaps, extending to other popular currency cross-sections, including portfolios sorted on carry (interest rate differentials), momentum, and value.

We analyze these empirical findings in the context of the international long-run risk model of Colacito and Croce (2011). We make two assumptions concerning the correlation of the shocks in the model. First, we allow for an imperfect degree of correlation between shocks to volatility and shocks to the predictive components of consumption, which we take as our proxy for the cyclical components within each country. Second, we assume that the correlation of the output gap of any country with the output gap of the US is decreasing in the level of that country's output gap. This model delivers sharp predictions that allow us to speak to the novel empirical evidence that we set forward in our analysis. Indeed, we show that, in the model, sorting currencies by interest rates is not the same as sorting by output gaps. Furthermore, a currency GAP premium arises in equilibrium in this economy. While this setup abstracts away from trade in the consumption goods market, it illustrates the properties of the consumption process that are necessary in a successful fully fledged general equilibrium model, whose analysis we leave to future research.

The remainder of the paper is as follows. Section 2 discusses related literature. Section 3 describes the data and defines the currency portfolios studied in the empirical analysis. Section 4 reports results on the predictive information content of business cycles for currency excess returns and on the performance and diversification gains from incorporating information on relative business cycles. Section 5 reports the results for asset pricing tests designed to explore whether the returns to output-gap-sorted portfolios can be understood as compensation for risk and whether a business cycle risk factor implied by our results is priced in the cross-section of currencies. Section 6 provides a theory that can explain the returns from output-gap-sorted strategies in terms of compensation for business cycle risk. Section 7 concludes. The Internet Appendix reports additional results, theoretical proofs,

[1] Moskowitz et al. (2012) study the performance of a time-series momentum strategy, while Baz et al. (2015) consider combinations of time-series and cross-sectional strategies across asset classes using carry, value, and momentum signals. Goyal and Jegadeesh (2018) show that, unlike cross-sectional strategies, time-series strategies are not zero-cost and once scaled, cross-sectional portfolio performance is substantially stronger.

[2] Lustig et al. (2014) propose a "dollar carry trade" strategy that trades a basket of currencies against the US dollar on the basis of the average forward discount relative to the US. Their strategy is different from the standard carry trade, and the returns compensate US investors for taking on aggregate risk by shorting the dollar in bad times, when the US price of risk is high. Our strategy is distinct conceptually—in that it directly sorts on relative business cycles across all countries rather than on interest rate (forward discount) information relative to the US—and empirically we show that the returns of the dollar carry trade are only mildly correlated with the returns to our strategies. Furthermore, we also show that our strategy returns are virtually uncorrelated with the returns from strategies that sort on Taylor-rule-implied interest rates, which are instead highly correlated with carry trade returns.

and some technical details on the asset pricing tests and on the construction of the output gap measures.

2. Related literature

This paper contributes to several related strands of literature at the intersection of international macro-finance and empirical asset pricing. First, we contribute to the growing body of research documenting predictability in the cross-section of currency excess returns. This strand of the literature has shown that cross-sectional predictability in currencies can be exploited using various investment strategies, including carry (Lustig and Verdelhan, 2007; Lustig et al., 2011; Menkhoff et al., 2012a), momentum (Menkhoff et al., 2012b; Asness et al., 2013), value (Asness et al., 2013; Menkhoff et al., 2017), "good" carry (Bekaert and Panayotov, 2020), strategies that combine carry with other signals (Jordà and Taylor, 2012; Barroso and Santa-Clara, 2015), information in the volatility risk premium (Della Corte et al., 2016a), and optimal dynamic currency strategies (Maurer et al., 2018). Our contribution to this literature is to uncover an economically distinct source of predictive information for the cross-section of currency excess returns stemming from the relative state of business cycles across countries and to illustrate the economic mechanism through which this predictability arises as compensation for business cycle risk.

In related work, Dahlquist and Hasseltoft, 2020 propose a currency strategy based on economic momentum, defined on the basis of eight economic variables that capture interest rate, price, industrial production, and unemployment information. This broad information set is combined to generate signals of countries that are growing more strongly (long position) and countries that are growing the least (short position), based on past trends that range from 1 to 60 months. Their results suggest that the strategy generates high risk-adjusted returns and subsumes the carry trade. Instead, we focus on a single proxy for business cycles, the output gap, which allows us to connect our empirical findings to macro-finance models and thus provide a direct economic interpretation. The simple model we propose also allows us to understand why the predictive information in output gaps is different from the predictive information in interest rates, and thereby we provide an economic mechanism to account for the difference between our strategy and the carry trade.

A second strand of the literature attempts to explain cross-sectional currency return predictability by theoretically and empirically investigating whether the returns generated by these investment strategies are compensation for risk. A series of recent papers find evidence in support of a variety of risk factors, including "global" exchange rate risk (Lustig et al., 2011; Colacito et al., 2018), unanticipated global volatility risk (Menkhoff et al., 2012a), downside risk (Lettau et al., 2014), global imbalance risk (Della Corte et al., 2016b), and correlation risk (Mueller et al., 2017), among others. We contribute to this literature by further bridging the gap between the set of factors that explain the cross-section of currency risk premia and macroeconomic fundamentals.

Our results show that a business cycle risk factor is priced in the cross-section of currency excess returns and the factor can be rationalized in terms of an international macro-finance model with long-run risk, as in Colacito and Croce (2011), Bansal and Shaliastovich (2012), Lustig and Richmond (2019), and Kremens and Martin (2019). The model abstracts away from trade in the consumption goods' market and delivers closed form solutions for most equilibrium objects of interest. In the interest of space, we leave the analysis of our empirical findings in the context of a fully fledged general equilibrium model to future research.

A third strand of related literature focuses on predicting exchange rate changes using information on macro fundamentals (Meese and Rogoff, 1983; Mark, 1995; Engel et al., 2007; Molodtsova et al., 2008; Rossi, 2013). Our empirical approach in this paper is very different in that we move away from traditional forecasting of bilateral exchange rate movements using time-series regressions and statistical metrics of forecast evaluation. Instead, we focus on the role of business cycles in predicting the cross-section of currency excess returns in a multicurrency portfolio setting that is typical of the empirical asset pricing literature.[3]

3. Data and currency portfolios

This section describes the main data employed in the empirical analysis as well as the construction of output gaps.

3.1. Data on spot and forward exchange rates

We collect daily bid, mid, and ask spot and one-month forward exchange rates vis-à-vis the US dollar from Barclays and Reuters via Datastream. The empirical analysis uses monthly data obtained by sampling end-of-month rates from October 1983 to January 2016. Our sample comprises 27 countries: Australia, Austria, Belgium, Brazil, Canada, Chile, Czech Republic, Germany, Finland, France, Iceland, Ireland, Italy, Japan, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, South Korea, Spain, Sweden, Switzerland, Turkey, United Kingdom, and the United States. The sample period for each currency differs, and thus the number of countries in our sample fluctuates over time. We replace Germany with the euro area in January 1999, while countries that join the euro area drop out of the sample upon entry into the single currency. We provide full details of the source and availability of currency data in Table A1 of the Internet Appendix.

[3] While the output gap is a common measure of business cycle conditions in the macroeconomics literature, it has received comparatively little attention in financial economics. Cooper and Priestley (2009) provide a notable exception, finding that the output gap can help predict future stock returns for the US and other G7 countries both in sample and out of sample. In international macroeconomics, Molodtsova et al. (2008) show that Taylor rule models that incorporate output gap and inflation information display predictive power for spot exchange rate changes in time-series regressions for three major exchange rates, although this result was challenged by Rogoff and Stavrakeva (2008), who argue the predictability is not robust across different subsample periods.

3.2. Currency excess returns

We define spot and forward exchange rates at time t as $Spot_t$ and Fwd_t. Exchange rates are defined as units of US dollars per unit of foreign currency such that an increase in $Spot_t$ indicates an appreciation of the foreign currency. The excess return on buying a foreign currency in the forward market at time t and selling in the spot market at time $t+1$ is computed as

$$RX_{t+1} = \frac{(Spot_{t+1} - Fwd_t)}{Spot_t}, \tag{1}$$

which is equivalent to the spot exchange rate return minus the forward premium

$$RX_{t+1} = \frac{Spot_{t+1} - Spot_t}{Spot_t} - \frac{Fwd_t - Spot_t}{Spot_t}. \tag{2}$$

According to the covered interest parity (CIP) condition, the forward premium approximately equals the interest rate differential $(Fwd_t - Spot_t)/Spot_t \simeq i_t - i_t^*$, where i_t and i_t^* represent the US and foreign riskless rates, respectively, over the maturity of the forward contract. If CIP holds, then the currency excess return is approximately equal to the exchange rate return (i.e., $(Spot_{t+1} - Spot_t)/Spot_t$) plus the interest rate differential relative to the US (i.e., $i_t^* - i_t$). As a matter of convenience, throughout this paper we refer to $fd_t = (Spot_t - Fwd_t)/Spot_t \approx i_t^* - i_t$ as either the forward discount or interest rate differential relative to the US dollar. However, it is important to note that, while CIP held closely in the data prior to the global financial crisis (e.g., Akram et al., 2008), recent evidence has highlighted that post-crisis deviations from CIP have become more pronounced (e.g., Du et al., 2018; Andersen et al., 2019), in which case the forward discount captures both interest rate differentials and CIP deviations. As we show below, the results in this paper are driven largely by spot exchange rate predictability, whereas the forward discount plays a negligible role in accounting for the returns of the strategies we propose. This feature is in stark contrast with carry trade returns, which are entirely driven by the forward discount. Thus, our results do not depend on whether CIP holds.

3.3. The output gap and data on economic activity

The output gap is defined as the logarithm of the difference between the actual (y_t) and "potential" ($\bar{y}_t$) output: $gap_t = y_t - \bar{y}_t$. A country's potential output is not directly observable and must therefore be estimated. Numerous statistical methods have been proposed to measure potential output $\bar{y}_t$, with the principal aim being to decompose output into its trend and cyclical components. The trend component can be viewed as the economy's natural or potential growth path, from which growth cyclically deviates. The cyclical component is thus a measure of short-term deviations and serves as our empirical proxy for the output gap.

To measure economic activity, we collect industrial production data from the Organisation for Economic Co-operation and Development's (OECD's) Original Release Data and Revisions Database. The database provides monthly vintages, which reflect the precise time series available to market participants each month and is thus free of any subsequent revisions or forward-looking information.

The full sample analysis uses the April 2016 vintage of data. The full series of monthly industrial production data begin at various dates across countries. The earliest start date is January 1960, and the sample ends in January 2016 to coincide with the last industrial production data point that was available for the majority of countries in April 2016. We estimate output gaps using various statistical techniques to extract a cyclical component from macroeconomic data: (i) the linear projection method of Hamilton (2018), (ii) the Hodrick and Prescott (1980, 1997) filter, (iii) the Baxter and King (1999) filter, and the quadratic trend specification used by Clarida et al. (1998).[4] Hamilton (2018) provides a quantitative analysis of the main drawbacks of the HP filter and suggests an alternative procedure for detrending output and measuring the output gap. Although the focus is to improve on the HP filter out of sample, Hamilton's analysis and criticisms are relevant for all other filters commonly used in this literature. Therefore, we use the Hamilton procedure in our real-time analysis, implementing the procedure recursively conditioning only on data available at the time of sorting.

For the real-time analysis, we use the full set of monthly industrial production vintages from December 1999 until January 2016. While the first vintage is in December 1999, the industrial production series within each vintage span back to 1960 (i.e., in December 1999, an investor could observe US industrial production data from January 1960 to October 1999).[5] Each monthly vintage records the industrial production data available to an investor in that particular month. In the out-of-sample analysis, we construct output gap estimates using each monthly vintage in turn, applying the linear projection method of Hamilton (2018), and therefore the resulting estimate at time t is conditioned only on information available at that time.

The linear projection methodology requires the estimation of the following time-series regression:

$$y_{i,t} = \alpha_i + \sum_{s=0}^{11} \beta_{i,s} y_{i,t-24-s} + \varepsilon_{i,t}, \tag{3}$$

where $y_{i,t}$ is the (log) value of industrial production for country i available at time t. We regress time-t values on their corresponding value from two years (24 months) earlier and include 12 lags in total, following the suggestion of Hamilton (2018). If industrial production is only available at quarterly intervals, we use four lags beginning eight-quarters earlier. We measure the cyclical component as $c_t = y_t - \hat{y}_t$, where $\hat{y}_t$ is the fitted value from the regression in (3). Our real-time measure is therefore purely

[4] We provide further details of the parameters and functional forms of these statistical techniques in the Internet Appendix.

[5] The data are available from February 1999 onwards; however the early months in the data set have unusually short samples and missing observations. We therefore begin the analysis with the most complete data set starting in December 1999.

backward-looking, making no use of either revised data or forward-looking information.

3.4. Output gap portfolios

At the end of each month t, we sort currencies into five portfolios based on the difference between each country's output gap and the US output gap. Portfolio 5 corresponds to countries with the highest output gap relative to the US, whereas Portfolio 1 comprises countries with the lowest output gap relative to the US. We calculate portfolio returns as the equally weighted $(1/N_k)$ return across the N_k currencies within portfolio k. We refer to the zero-cost dollar-neutral strategy that takes a long position in Portfolio 5 (P_5) and a short position in Portfolio 1 (P_1) as the GAP_{CS} strategy, which is a tradeable investment portfolio that exploits the relative cross-sectional spread in business cycle conditions around the world.

In addition, we also report results for trading strategies that trade all currencies with linear weights equal to

$$w_{j,t+1} = c_t\left(x_{j,t} - \bar{x}_t\right), \tag{4}$$

where $x_{j,t}$ denotes the signal for currency j in month t (i.e., the output gap of country j minus the US output gap) and $\bar{x}_t = N_t^{-1}\Sigma_{j=1}^{N_t} x_{j,t}$ denotes the cross-sectional average of the signal (across countries, N_t). c_t is a scaling factor such that the absolute sum of all portfolio weights equals unity (i.e., $c_t = 1/\sum_j |x_{j,t} - \bar{x}_t|$). Currencies with a signal value above the cross-sectional mean receive positive portfolio weights, whereas currencies with a signal value below the average receive negative weights. The portfolio return is then given by $rx_{t+1}^p = \sum_{j=1}^{N_t} w_{j,t+1} rx_{j,t+1}$. In the implementation of this approach we rebalance the portfolios at the end of each month.

Finally, we report returns of rank portfolios, where weights are given by

$$w_{j,t+1} = c_t\left(\text{rank}(x_{j,t}) - \sum_{j=1}^{N_t}\text{rank}(x_{j,t})/N_t\right). \tag{5}$$

The scaling factor c_t is analogous to the case of linear weights above (but uses ranks of signals instead of actual signals) and ensures that we are one dollar long and one dollar short, as in Asness et al. (2013). The procedures based on linear weights and rank portfolios are useful for a comparison with the GAP_{CS} strategy because they trade all currencies in every period and thereby are less reliant on currencies with extreme output gaps. This translates into a lower cross-sectional standard deviation of the weights for linear and rank portfolios relative to the GAP_{CS} strategy. Given the relatively small number of assets in the corner portfolios traded with the GAP_{CS} strategy, these strategies provide some reassurance that results from the GAP_{CS} strategy are not driven by just a few currencies.

Finally, it is important to note that all of the cross-sectional strategies are unaffected by time-series trends in the US dollar since they take long and short positions in the US dollar of equal amount by construction.

3.4.1. Time-series portfolio

At the end of each month t, we form a $1/N_t$ (equally weighted) strategy that takes long positions in the currencies of countries with output gaps above the US output gap and short positions in the currencies of countries with output gaps below the US. The strategy thus invests in all currencies available at each point in time, under the expectation that countries with higher (lower) output gaps than the US should subsequently offer higher (lower) currency excess returns. In the out-of-sample analysis, we refer to this portfolio strategy as GAP_{TS}. Unlike the cross-sectional strategies described above, the GAP_{TS} strategy is not dollar neutral because the number of currencies with output gaps above the US varies over time. The strategy is therefore exposed to any (macro) factor that impacts the evolution of the US dollar over time.

4. Business cycles and currency returns

In this section we explore if business cycles can predict currency excess returns. Our benchmark approach is based on a cross-sectional portfolio sort in which currencies are sorted into five bins (P_1, P_2, P_3, P_4, P_5) based on quintiles of the cross-sectional distribution of relative output gaps from the weakest to the strongest economy currencies.

4.1. Full-sample performance

In Table 1, we present the average excess returns of the five output-gap-sorted portfolios, which display an increasing pattern from P_1 to P_5 for all four of the output gap measures. Furthermore, the spread in returns between P_1 and P_5 is sizable, ranging from 4.56% to 6.66% per annum, which are all statistically different from zero at the 1% level.

Further scrutiny of the results in Table 1 reveals that the predictability of the cross-section portfolio ($P5$ - $P1$) and the time-series portfolio are mainly driven by predicting spot exchange rate returns (see row denoted *fx*), whereas the return from the forward premium (equal to the interest rate differential under CIP) contributes comparatively little to the return (see row denoted *ir*). This finding contrasts with the currency carry trade strategy, in which returns are entirely driven by exploiting forward premia across countries—the exchange rate component of the excess return is typically negative.[6] The last three rows in Table 1 report the currency turnover and the spread in both forward premia and output gaps in each of the five portfolios. The turnover measure is slightly higher than that reported in the literature for carry trade strategies but lower than momentum strategies (see, e.g., Menkhoff et al., 2012a; 2012b). We note that a tendency exists for forward premia to increase as we move from P_1 to P_5, albeit non monotonically; however, the spread is low, consistent with the returns being driven largely by the spot exchange rate component.

The results in Table 1 are qualitatively identical for all four measures of the output gap considered, indicating that they lead to comparable portfolio sorts (i.e., similar rankings of countries by the state of the business cycle). In Ta-

[6] For comparison, we present the equivalent descriptive statistics for forward-premia-sorted (carry) portfolios in Table A2 of the Internet Appendix.

Table 1
Full sample business cycle currency portfolios.

The table presents descriptive statistics for five currency portfolios sorted by output gaps. The output gap is estimated as (log) industrial production minus the (log) trend in industrial production. The trend is estimated in four ways using a (i) Hodrick-Prescott filter, (ii) Baxter-King filter, (iii) linear projection, and (iv) quadratic time trend. Portfolios are rebalanced monthly with strong (weak) economy currencies entering P_5 (P_1). We report summary statistics for the annualized excess mean return and its decomposition between the exchange rate (*fx*) and interest rate (*ir*) components. We also report the Sharpe ratio (*Sharpe*), standard deviation (*std*), skewness (*skew*), kurtosis (*kurt*), maximum drawdown (*mdd*), average turnover (*t/o*), average forward premium (*fp*), and average output gap (*gap*) for each portfolio. The *Cross-section* portfolio is long P_5 and short P_1. The *Time series* portfolio takes a 1/N position in currencies, going long (short) currencies issued by countries with an output gap above (below) the US output gap. The superscripts *, **, *** represent significance of the *Cross-section* and *Time series* portfolios at the 10%, 5%, and 1% level using Newey and West (1987) standard errors. The sample is from October 1983 to January 2016.

	Hodrick-Prescott filter					Cross-section	Time series	*Baxter-King filter*					Cross-section	Time series
	P_1	**P_2**	**P_3**	**P_4**	**P_5**			**P_1**	**P_2**	**P_3**	**P_4**	**P_5**		
mean (%)	−0.25	0.96	2.77	4.00	6.41	6.66***	2.45***	−0.44	2.45	2.39	3.72	5.97	6.41***	3.83***
fx (%)	*−2.34*	*−1.03*	*0.88*	*1.58*	*2.72*	*5.06*	*2.03*	*−2.34*	*0.65*	*0.49*	*1.23*	*1.92*	*4.26*	*3.38*
ir (%)	*2.09*	*1.99*	*1.89*	*2.41*	*3.69*	*1.60*	*0.41*	*1.90*	*1.80*	*1.90*	*2.49*	*4.06*	*2.15*	*0.44*
Sharpe	−0.02	0.11	0.27	0.43	0.71	0.82	0.54	−0.04	0.26	0.26	0.39	0.68	0.77	0.74
std	10.18	9.09	10.12	9.32	9.05	8.14	4.57	10.15	9.50	9.33	9.51	8.82	8.31	5.21
skew	−0.06	−0.47	−0.28	−0.27	−0.28	0.01	−0.92	−0.08	0.06	−0.21	−0.29	−0.61	0.07	−0.39
kurt	4.49	4.72	4.75	4.39	3.97	4.32	10.89	3.94	3.83	4.32	4.22	5.00	4.25	5.70
mdd (%)	42.5	34.2	23.9	23.6	24.4	9.0	8.6	49.0	28.8	26.1	24.6	21.8	23.0	9.2
t/o (%)	44.8	58.2	67.2	60.6	44.8			10.0	21.8	29.8	23.1	11.6		
fp (t, %)	2.23	2.03	1.80	2.45	4.15			1.91	1.87	1.75	2.38	4.81		
gap (t, %)	−3.08	−0.96	0.11	1.17	3.01			−2.75	−0.84	0.18	1.30	3.01		
	Linear projection					**Cross-section**	**Time series**	*Quadratic time trend*					**Cross-section**	**Time series**
	P_1	**P_2**	**P_3**	**P_4**	**P_5**			**P_1**	**P_2**	**P_3**	**P_4**	**P_5**		
mean (%)	0.46	2.85	2.23	3.18	5.41	4.95***	3.72***	0.27	1.99	3.08	4.21	4.83	4.56***	2.14**
fx (%)	*−2.25*	*0.93*	*0.27*	*1.13*	*1.95*	*4.20*	*3.36*	*−1.04*	*−0.16*	*0.02*	*1.54*	*1.75*	*2.80*	*1.95*
ir (%)	*2.71*	*1.92*	*1.96*	*2.05*	*3.46*	*0.74*	*0.37*	*1.31*	*2.15*	*3.06*	*2.67*	*3.08*	*1.76*	*0.19*
Sharpe	0.05	0.32	0.23	0.33	0.56	0.66	0.72	0.03	0.21	0.32	0.49	0.51	0.60	0.41
std	9.80	8.91	9.59	9.65	9.63	7.47	5.18	10.05	9.58	9.76	8.66	9.53	7.56	5.27
skew	−0.24	−0.24	−0.50	−0.07	−0.54	−0.34	−1.23	−0.23	−0.21	−0.07	−0.51	−0.15	−0.68	−1.24
kurt	4.66	5.21	4.79	3.58	5.44	5.13	10.94	4.23	4.58	4.29	4.50	5.02	6.29	10.93
mdd (%)	40.4	28.4	31.8	29.4	19.1	35.3	11.6	38.9	28.5	31.1	24.8	21.5	18.3	16.5
t/o (%)	26.0	43.8	52.7	44.5	26.4			20.0	32.9	44.3	33.8	19.7		
fp (t, %)	2.68	2.01	1.90	2.17	3.97			1.17	2.04	3.15	2.87	3.35		
gap (t, %)	−1.33	−0.30	0.31	0.90	1.98			−8.41	−3.35	−0.22	2.59	7.78		

ble A3 of the Internet Appendix, we report evidence on the correlation across portfolio sorts obtained by the different output gap measures. While the correlations are not perfect, they are sizable and in a range between 0.41 to 0.65. Furthermore, we report the results from a principal component analysis applied to the four output gap estimates. The average percentage of cross-sectional variation explained by the first principal component is a hefty 86%, indicating that the output gap measures have a very strong common component.

In Table 2, we present the results from a principal component decomposition of the returns of the five portfolios sorted on relative output gaps. The results indicate a strong factor structure in currency portfolio returns sorted by relative output gaps. The first principal component accounts for most of the variation in portfolio returns, but the loadings appear to be almost identical across the five portfolios, suggestive of a "level" factor, as also shown by Lustig et al. (2011). The second principal component is instead a "slope" factor and the loadings of the five portfolios on this principal component display a tendency to increase (monotonically for two of the output gap measures) from negative values for P_1 to positive values for P_5. Therefore, it is the second principal component that is key to understanding the cross-sectional difference in excess returns.[7]

We also compare the results from sorting on output gaps to a strategy that sorts on Taylor rule fundamentals. Specifically, we consider the standard Taylor rule with coefficients of 1.5 on inflation and 0.5 on the output gap, respectively. We then use the differential in interest rates implied by the Taylor rule to sort currencies. The results, presented in Table 3, suggest that sorting on Taylor-rule-implied interest rates generates large Sharpe ratios both in the cross-section (Sharpe ratio of 0.90) and the time series (Sharpe ratio of 0.65).

However, sorting on Taylor rules also generates excess returns that display high correlations with the HML_{FX} factor of Lustig et al. (2011) (0.84 and 0.51), whereas sorting on relative output gaps provides excess returns that are not

[7] These features of the factor structure resemble the features displayed by carry portfolios sorted on interest rate differentials, in which the slope factor is key to understanding carry trade excess returns. However, we also find that the second principal component for portfolios sorted on output gaps is orthogonal to the analogous HML_{FX} factor shown by Lustig et al. (2011), confirming that sorting currencies on output gaps is very different from sorting currencies on interest rates.

Table 2
Principal component decomposition of business cycle portfolios.

The table presents results from a principal component decomposition of the returns to five currency portfolios sorted by output gaps. The output gap is estimated as (log) industrial production minus the (log) trend in industrial production. The trend is estimated in four ways using a (i) Hodrick-Prescott filter, (ii) Baxter-King filter, (iii) linear projection, and (iv) quadratic time trend. Portfolios are rebalanced monthly with strong (weak) economy currencies entering P_5 (P_1). We report the loading of each portfolio on all five principal components (*PCs*) and the percentage of total return variation explained by each *PC*. The sample is from October 1983 to January 2016.

	Hodrick-Prescott filter						*Baxter-King filter*				
	PC_1	**PC_2**	**PC_3**	**PC_4**	**PC_5**		**PC_1**	**PC_2**	**PC_3**	**PC_4**	**PC_5**
P_1	0.47	−0.71	0.46	−0.08	−0.23	P_1	0.48	−0.51	0.69	0.11	−0.16
P_2	0.43	−0.19	−0.33	0.00	0.82	P_2	0.46	−0.18	−0.24	−0.26	0.79
P_3	0.49	0.18	−0.25	0.75	−0.32	P_3	0.45	−0.18	−0.52	−0.39	−0.59
P_4	0.44	0.19	−0.47	−0.64	−0.38	P_4	0.46	0.25	−0.27	0.81	−0.02
P_5	0.40	0.63	0.63	−0.11	0.19	P_5	0.38	0.78	0.36	−0.34	−0.05
var explained	78%	8%	5%	5%	4%	*var explained*	79%	7%	6%	4%	4%
	Linear projection						*Quadratic time trend*				
	PC_1	**PC_2**	**PC_3**	**PC_4**	**PC_5**		**PC_1**	**PC_2**	**PC_3**	**PC_4**	**PC_5**
P_1	0.45	−0.85	0.00	0.04	−0.26	P_1	0.48	−0.08	0.70	−0.34	−0.40
P_2	0.42	−0.03	−0.39	−0.15	0.80	P_2	0.47	−0.10	0.24	0.21	0.82
P_3	0.45	0.40	−0.57	−0.19	−0.53	P_3	0.45	−0.52	−0.60	−0.40	−0.04
P_4	0.46	0.27	0.25	0.81	0.04	P_4	0.40	−0.10	−0.13	0.80	−0.41
P_5	0.45	0.21	0.68	−0.53	0.00	P_5	0.43	0.84	−0.29	−0.17	−0.02
var explained	78%	7%	6%	5%	4%	*var explained*	78%	7%	6%	5%	4%

Table 3
Taylor rule.

The table presents descriptive statistics for five currency portfolios, sorted by their Taylor-rule-implied interest rate. The Taylor rule is calibrated to equal $1.5\pi_t + 0.5y_t$, where π_t is inflation and y_t is the in-sample output gap calculated using a Hodrick-Prescott filter. Portfolios are rebalanced monthly with high (low) implied interest rate currencies entering P_5 (P_1). We report summary statistics for the annualized excess mean return and its decomposition between the exchange rate (*fx*) and interest rate (*ir*) components. We also report the Sharpe ratio (*Sharpe*), standard deviation (*std*), skewness (*skew*), kurtosis (*kurt*), maximum drawdown (*mdd*), average turnover (*t/o*), average forward premium (*fp*), and average output gap (*gap*) for each portfolio. The *Cross-section* portfolio is long P_5 and short P_1. The *Time series* portfolio takes a 1/N position in currencies, going long (short) currencies issued by countries with a Taylor-rule-implied interest rate above (below) the US Taylor-rule-implied interest rate. The superscripts *, **, *** represent significance of the *Cross-section* and *Time series* portfolios at the 10%, 5%, and 1% level using Newey and West (1987) standard errors. We also report the correlation of the *Cross-section* and *Time series* portfolios with the equivalent portfolios sorted on Hodrick-Prescott filtered output gaps (ρ_{GAP}) and interest rates ($\rho_{HML_{FX}}$). The sample is from October 1983 to January 2016.

	Taylor rule					Cross-	Time
	P_1	**P_2**	**P_3**	**P_4**	**P_5**	section	series
mean (%)	−1.52	1.19	4.25	2.68	7.45	8.97***	2.75***
fx (%)	*−0.56*	*1.07*	*2.74*	*0.16*	*−1.67*	*−1.11*	*0.32*
ir (%)	*−0.96*	*0.12*	*1.51*	*2.52*	*9.12*	*10.08*	*2.43*
Sharpe	−0.16	0.12	0.45	0.28	0.75	0.90	0.65
std	9.59	9.79	9.36	9.45	9.92	9.92	4.25
skew	0.16	−0.10	−0.33	−0.42	−0.55	−0.42	−1.04
kurt	4.05	3.88	4.79	4.35	4.57	4.85	9.99
mdd (%)	61.3	34.7	25.2	28.8	20.6	13.8	11.2
t/o (%)	24.9	40.9	43.5	33.2	12.9		
fp (t, %)	−0.85	0.19	1.37	2.50	9.87		
gap (t, %)	−1.59	−0.18	0.36	0.50	1.11		
ρ_{GAP}						0.25	0.36
$\rho_{HML_{FX}}$						0.84	0.51

(or only) mildly correlated with Taylor-rule-sorted portfolios. The difference between the Taylor-rule-sorted portfolios and the portfolios sorted on relative output gaps is also apparent by noticing that all of the predictability in the Taylor rule portfolios stems from the forward premia component, consistent with the returns from carry trade portfolios. The desirable correlation properties of portfolios sorted on output gaps are an important feature that we explore further in the paper, specifically when we investigate the diversification benefits of adding a trading strategy that sorts on output gaps to a conventional menu of currency investment strategies.[8]

[8] In Table A4 of the Internet Appendix, we also present statistics for portfolios sorted on the basis of deviations in interest rates from the Taylor-rule-implied rates. These deviations capture movements in short-

Table 4
Real-time business cycle currency portfolios.
The table presents investment performance for output-gap-based currency trading strategies. The output gap is estimated using monthly "vintages" of real-time industrial production data from the OECD's Real-Time Data and Revisions Database. To estimate the output gap, we follow the linear projection procedure in Hamilton (2018) by running the regression $y_{i,t} = \alpha_i + \sum_{s=0}^{11} \beta_{i,s} y_{i,t-24-s} + \varepsilon_{i,t}$ each month in which y is (log) industrial production. The output gap is constructed as the difference between the most recently available data point at time t (y_t) and the fitted value from the regression. GAP_{CS} is a high-minus-low portfolio formed as $P_5 - P_1$, after sorting currencies into five portfolios ranging from the lowest (P_1) to the highest (P_5) output gap. *LIN* and *RNK* take a position in all currencies with the weight determined by either the magnitude or relative size of the output gap. GAP_{TS} is a 1/N time-series strategy long (short) currencies issued by countries with an output gap above (below) the US output gap. The three *COM* portfolios take 50 – 50 weights in GAP_{TS} and the GAP_{CS}, *LIN*, and *RNK* strategies. We report summary statistics for the annualized mean, which is then further split between the exchange rate (*fx*) and interest rate (*ir*) components. We also report the Sharpe ratio (*Sharpe*), skewness (*skew*), kurtosis (*kurt*), and maximum drawdown (*mdd*). The superscripts *, **, *** represent significance of the strategies' mean excess returns at the 10%, 5%, and 1% significance levels using Newey and West (1987) corrected standard errors. The sample runs from December 1999 to January 2016.

Panel A: Investment performance excluding bid-ask spreads

	GAP_{CS}	LIN	RNK	GAP_{TS}	COM_{GAP}	COM_{LIN}	COM_{RNK}
mean (%)	4.92***	2.16***	3.88***	2.76**	3.82***	2.45***	3.30***
fx (%)	4.21	1.74	3.38	2.83	3.50	2.26	3.08
ir (%)	0.71	0.42	0.49	−0.07	0.33	0.19	0.22
Sharpe	0.72	0.74	0.72	0.65	0.82	0.82	0.82
skew	0.31	0.34	0.25	−0.70	−0.01	−0.46	−0.23
kurt	2.83	3.23	3.24	5.17	3.02	3.49	3.25
mdd (%)	6.88	2.91	5.40	4.27	4.66	2.97	4.03

Panel B: Investment performance including bid-ask spreads

	GAP_{CS}	LIN	RNK	GAP_{TS}	COM_{GAP}	COM_{LIN}	COM_{RNK}
mean (%)	3.46**	1.45**	2.50**	2.14*	2.79**	1.78**	2.31**
fx (%)	3.10	1.19	2.34	2.37	2.71	1.76	2.33
ir (%)	0.36	0.25	0.15	−0.22	0.08	0.03	−0.02
Sharpe	0.50	0.50	0.46	0.50	0.60	0.60	0.57
skew	0.31	0.33	0.24	−0.70	−0.01	−0.46	−0.24
kurt	2.81	3.20	3.22	5.19	3.01	3.50	3.26
mdd (%)	6.86	2.91	5.39	4.27	4.66	2.97	4.03

Overall, these initial results suggest a strong link exists between the relative state of the business cycle and future currency excess returns, which is mainly driven by spot exchange rate predictability and is thus distinct from the predictability found in either carry- or Taylor-rule-based currency portfolios.

4.2. Real-time performance

In Panel A of Table 4 we report the gross returns from implementing in real time the GAP_{CS} strategy as well as the linear and rank weight cross-sectional portfolios. We also report the returns from implementing the time-series variant of this strategy, GAP_{TS}, and combinations of the GAP_{TS} strategy with the cross-sectional strategies. We observe that each investment strategy generates a statistically significant return at conventional significance levels. The Sharpe ratio of GAP_{CS} is 0.72, the same as when using rank weights, while it is slightly higher at 0.74 when using linear weights.[9] GAP_{TS} generates a Sharpe ratio of 0.65 out of sample, and the Sharpe ratio always increases when the cross-sectional strategies are combined with the GAP_{TS} strategy, reaching 0.82. This higher Sharpe ratio is due to the fact that the correlation between GAP_{CS} and GAP_{TS} is positive but far from perfect.

Fig. 2 plots the cumulative returns to both GAP_{CS} and GAP_{TS} currency strategies. Specifically, the figure shows the out-of-sample cumulative returns (left-hand plot), the equivalent in-sample cumulative returns obtained using the Hamilton (2018) linear projection applied to revised data (middle plot), and combination strategies COM_{GAP}, COM_{LIN}, and COM_{RNK} (right-hand plot). Returns are normalized so that each series has the same volatility of 6% annualized, which makes it straightforward to compare the cumulative returns. The shaded bars in Fig. 2 are the National Bureau of Economic Research (NBER) recession periods, which occur twice during this sample. It seems natural to examine whether the returns from our strategy are different across booms and recessions in the US, particularly in light of the recent paper by Gómez-Cram (2018) that

term rates not captured by systematic monetary policy (Taylor rule reaction functions) including, for example, liquidity fluctuations in money markets or movements in interbank risk. We find these deviations do not generate a spread in returns that is statistically different from zero, suggesting that currency excess returns are not driven by the nonsystematic component of short-term rates.

[9] In Table A5 of the Internet Appendix we report evidence on the minimum and maximum weights for all three cross-sectional strategies. Clearly, for the high-minus-low strategy, they are -20% and 20% given that the corner portfolios have at most five currencies and this strategy employs equal weights. The cross-sectional standard deviation is 28%. The largest weights remain modest for linear and rank portfolios, with a minimum of -25% and a maximum of 25% but with considerably lower cross-sectional standard deviations (9% and 16%) given that these strategies trade all currencies and not just those in the extreme quintiles of the cross-sectional distribution of output-gap-sorted portfolios.

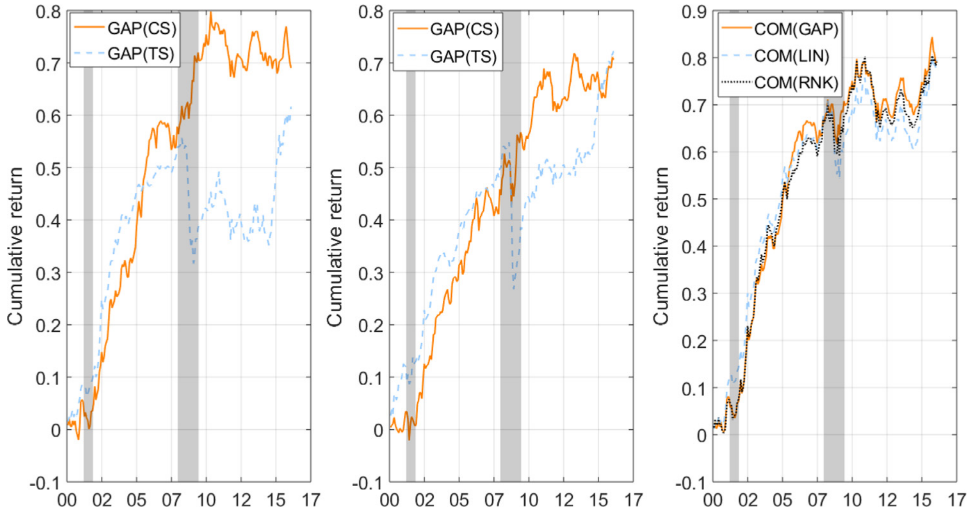


Fig. 2. Cumulative returns to *GAP* strategies. The figure plots the cumulative returns to *GAP* currency strategies. The left-hand plot shows the out-of-sample cumulative returns for the GAP_{CS} and GAP_{TS} strategies. The middle plot shows the equivalent in-sample cumulative returns using the Hamilton (2018) linear projection. The right-hand plot shows the three combination strategies, COM_{GAP}, COM_{LIN}, and COM_{RNK}. Returns are normalized so that each series has the same volatility of 6% annualized. The shaded bars reflect NBER recession periods..

shows expected stock returns are predictively negative in the first four to six months after the onset of a recession but are high thereafter. The graphs in Fig. 2 show the steady performance of the GAP_{CS} strategy, in which returns do not display different behavior across booms and recessions regardless of whether the cross-sectional strategy is implemented in or out of sample. This is not surprising given the GAP_{CS} strategy is dollar neutral, and hence it is not obvious why or how US business cycles should be related to the GAP_{CS} excess returns. However, for the GAP_{TS} strategy, which is not dollar neutral, there is evidence that its returns drop during the recession associated with the global financial crisis and take a long time before returning to their previous peak, therefore displaying a very different behavior to that recorded for US stock returns by Gómez-Cram (2018).[10]

By construction, the GAP_{TS} strategy is more likely to be long the US dollar during US expansions and short the US dollar during US recessions. The poor performance of the GAP_{TS} strategy during the recession induced by the global financial crisis likely occurs because of the sharp dollar appreciation against most currencies in the world (with the exception of other safe haven currencies such as the Japanese yen and the Swiss franc) observed during the crisis. This safety premium effect is typically explained by the special reserve currency status of the US dollar, which makes the US the recipient of large capital flows at times of major negative global shocks (e.g., Maggiori, 2017). Finally, the graphs of the combination strategies show a striking similarity, suggesting that the specific method for assigning weights to long and short positions in GAP_{CS} has little bearing on the final return outcome and, for each combination, the losses incurred by GAP_{TS} during the global financial crisis are substantially mitigated by combining it with the GAP_{CS} strategy.

In Panel B of Table 4 we report results in the same format as Panel A for returns net of transaction costs (i.e., accounting for bid-ask spreads). We incorporate transactions costs using bid-ask ($b - a$) spreads such that the return on a long position is calculated as

$$RX^{net}_{t+1} = \frac{Spot^b_{t+1} - Fwd^a_t}{Spot^m_t}. \tag{6}$$

The bid-ask spread data available are for quoted spreads and not effective spreads. Since it is known that quoted spreads are much higher than effective spreads, we follow earlier work (e.g., Goyal and Saretto, 2009; Menkhoff et al., 2012a; 2017) and employ 50% of the quoted bid-ask spread as the actual spread.[11]

The Sharpe ratio for GAP_{CS} goes down from 0.72 to 0.50 when adjusting for bid-ask spreads, and the Sharpe ratio for GAP_{TS} reduces from 0.65 to 0.50, while the combination of GAP_{CS} and GAP_{TS} generates a Sharpe ratio of about 0.60. In short, transaction costs do not wipe out the performance of strategies that sort on output gaps out of sample,

[10] In Table A6 of the Internet Appendix, we present summary statistics on the returns to the GAP_{CS} and GAP_{TS} portfolios during expansions and NBER recession periods. The GAP_{CS} strategy generates similar returns in both periods (4.75% during expansions and 5.94% during recessions), while the GAP_{TS} portfolio generates a positive return of 3.90% during expansions but a -3.99% return in recessions. This result is driven by taking cyclical positions in the US dollar, which results in a large negative return during recessions when the dollar appreciates. In the final row we present the US dollar exposure—measured as the over or under exposure to the dollar, in which a value of zero indicates a dollar-neutral portfolio, while a value of -1 indicates a dollar portfolio that is long all foreign currencies against the US dollar. During expansions, the GAP_{TS} portfolio has a positive dollar exposure of 0.26, but this flips to -0.27 during recessions.

[11] Even this number seems conservative: Gilmore and Hayashi (2011) find transaction costs due to bid-ask spreads are likely much lower than our 50% rule, while Cespa et al. (2019) find a rule of 25% is more appropriate since 2011.

Table 5
Correlations between trading strategies.
The table presents linear correlation coefficients between trading strategies. In Panel A, we report correlations between output gap currency trading strategies. The output gap is estimated using monthly "vintages" of real-time industrial production data from the OECD's Real-Time Data and Revisions Database. To estimate the output gap, we follow the linear projection procedure in Hamilton (2018) by running the regression, $y_{i,t} = \alpha_i + \sum_{s=0}^{11} \beta_{i,s} y_{i,t-24-s} + \varepsilon_{i,t}$ each month in which y is (log) industrial production. The output gap is constructed as the difference between the most recently available data point at time t (y_t) and the fitted value from the regression. GAP_{CS} is a high-minus-low portfolio formed as $P_5 - P_1$, after sorting currencies into five portfolios ranging from the lowest (P_1) to the highest (P_5) output gap. *LIN* and *RNK* take a position in all currencies with the weight determined by either the magnitude or relative size of the output gap. GAP_{TS} is a 1/N time-series strategy long (short) currencies issued by countries with an output gap above (below) the US output gap. The three *COM* portfolios take 50 − 50 weights in the GAP_{TS} and the GAP_{CS}, *LIN*, and *RNK* strategies. In Panel B, we present correlations between the output gap currency trading strategies and various currency and equity-based strategies. The sample runs from December 1999 to January 2016.

Panel A: Output gap currency trading strategies

	$\mathbf{GAP_{CS}}$	LIN	RNK	$\mathbf{GAP_{TS}}$	$\mathbf{COM_{GAP}}$	$\mathbf{COM_{LIN}}$	$\mathbf{COM}_{\mathrm{RNK}}$
High-minus-low $\mathrm{GAP_{CS}}$							
Linear weights (LIN)	0.86						
Rank weights (RNK)	0.88	0.93					
GAP_{TS}	0.36	0.34	0.38				
GAP model combo ($\mathrm{COM_{GAP}}$)	0.90	0.79	0.83	0.73			
LIN model combo ($\mathrm{COM_{LIN}}$)	0.68	0.74	0.73	0.89	0.91		
RNK model combo ($\mathrm{COM_{RNK}}$)	0.78	0.81	0.87	0.79	0.94	0.96	

Panel B: Alternative trading strategies in currency, equity, and interest rate markets

	$\mathbf{GAP_{CS}}$	**LIN**	**RNK**	$\mathbf{GAP_{TS}}$	$\mathbf{COM_{GAP}}$	$\mathbf{COM_{LIN}}$	$\mathbf{COM}_{\mathrm{RNK}}$
HML_{fx}	0.06	0.07	0.02	0.02	0.06	0.05	0.03
Dollar	0.24	0.21	0.22	0.23	0.28	0.27	0.27
Dollar carry	0.23	0.21	0.22	0.23	0.28	0.27	0.27
Value	0.15	0.03	0.13	0.04	0.13	0.05	0.11
Momentum	0.07	0.15	0.08	−0.06	0.02	0.03	0.02
Global imbalance	0.13	0.12	0.10	0.23	0.20	0.22	0.18
Foreign exchange trend strategy	0.08	0.08	0.08	0.07	0.09	0.09	0.09
Interest rate trend strategy	0.07	0.12	0.13	−0.22	−0.05	−0.10	−0.03
Illiquidity	0.08	0.12	0.04	0.21	0.15	0.21	0.14
US equity	0.00	−0.04	−0.01	0.10	0.05	0.06	0.05

and the Sharpe ratios remain attractive even after accounting for bid-ask spreads.[12]

The results in Table 4 also confirm that the predictive power stems mainly from spot rate predictability rather than interest rate differentials: approximately 90% of the total return is delivered from the spot component across all portfolios considered. Therefore the basic features of exchange rate predictability recorded in the full sample hold in the real-time analysis.[13]

4.2.1. Relationship with other strategies and diversification gains

In addition to analyzing the real-time performance of currency strategies that sort on output gaps, we also compare the returns arising from these strategies to a number of other portfolio strategies. This analysis is useful to assess whether sorting on output gaps simply recovers returns that can be obtained in other ways or whether they constitute a novel source of exchange rate predictability that can offer diversification gains to investors.

Table 5 reports a battery of correlation coefficients between the returns from the strategies sorting on output gaps and the returns from a variety of currency strategies and equity-based strategies that we provide full details of in the Internet Appendix. The main point arising from this table is that the returns of each output gap strategy are generally uncorrelated, or only mildly correlated, with any of the alternative strategies and factors considered. For example, for the GAP_{CS} strategy, the correlations range from zero (for the US equity market) to 0.24 (for the dollar factor), and they are of similar magnitude for each variant of the strategy. The results suggest strategies that sort on output gaps contain novel economic information and are not a mechanical relabeling of existing currency strategies or factors. In turn, large diversification benefits are potentially available to currency investors from adding an output-gap-sorted strategy to a broader currency portfolio.

Thus, to better understand the value of the GAP_{CS} strategy for a currency investor, we combine it with various canonical currency strategies and assess its value added in terms of performance. In Panel A of Table 6 we show the returns from carry, dollar carry, momentum, and value

[12] In Table A7 we again consider portfolios obtained from sorting currencies on the basis of interest rates implied by Taylor rules, exactly as for Table 3 but using real-time data for the out-of-sample period. While the results confirm the strong investment performance of the Taylor rule strategy, they also confirm the in-sample result that the strategy is very highly correlated with the HML_{FX} factor while being only modestly correlated with the GAP_{CS} strategy.

[13] Up to now, we have taken the perspective of a US investor by calculating excess returns and building dollar-neutral portfolios. As a robustness check, we depart from this base scenario and run calculations with four alternative base currencies. Specifically, we construct the output gap strategy out of sample from the separate perspectives of Eurozone, British, Japanese, and Swiss investors. The results, reported in Table A8 in the Internet Appendix, indicate no qualitative changes to our results based on a US perspective.

Table 6
Diversification benefits from the GAP_{CS} trading strategy.

The table presents the investment performance of common currency trading strategies and the impact on performance from adding the GAP_{CS} strategy. The output gap is estimated using monthly "vintages" of real-time industrial production data from the OECD's Real-Time Data and Revisions Database. To estimate the output gap, we follow the linear projection procedure in Hamilton (2018) by running the regression, $y_{i,t} = \alpha_i + \sum_{s=0}^{11} \beta_{i,s} y_{i,t-24-s} + \varepsilon_{i,t}$ each month in which y is (log) industrial production. The output gap is constructed as the difference between the most recently available data point at time t (y_t) and the fitted value from the regression. GAP_{CS} is a high-minus-low portfolio formed as $P_5 - P_1$, after sorting currencies into five portfolios ranging from the lowest (P_1) to the highest (P_5) output gap. In Panel A, we report the investment performance of popular currency investment strategies, in which HML_{FX} is the currency carry trade; *DCAR* is the "dollar carry" trade; *MOM* is a momentum trade; *VAL* is a value trade and *EW* is a 1/N portfolio that takes an equal position in each currency strategy. In Panel B, we add the GAP_{CS} strategy. We report summary statistics for the annualized excess mean return, the Sharpe ratio (*Sharpe*), standard deviation (*std*), skewness (*skew*), kurtosis (*kurt*), maximum drawdown, percentage increase in Sharpe ratio (%Δ Sharpe), and weight in the GAP_{CS} portfolio (w_{GAP}). The superscripts *, **, *** represent significance of the strategies' mean excess returns at the 10%, 5%, and 1% confidence levels using Newey and West (1987) corrected standard errors. The sample runs from December 1999 to January 2016.

Panel A: Excluding the GAP_{CS} strategy					
	$\mathbf{HML_{FX}}$	DCAR	MOM	VAL	EW
mean (%)	6.34**	2.60	1.41	0.05	2.60***
Sharpe	0.58	0.31	0.16	0.01	0.74
std	10.91	8.51	9.01	8.65	3.53
skew	−0.72	−0.49	0.28	0.47	−0.24
kurt	5.23	4.78	3.31	4.42	4.13
mdd (%)	0.16	0.17	0.23	0.39	0.07
Panel B: Including the GAP_{CS} strategy					
	$\mathbf{GAP_{CS}}$+ $\mathbf{HML_{FX}}$	$\mathbf{GAP_{CS}}$+ **DCAR**	$\mathbf{GAP_{CS}}$+ **MOM**	$\mathbf{GAP_{CS}}$+ **VAL**	**EW**
mean (%)	5.61***	3.74***	3.15**	2.47*	3.06***
Sharpe	0.85	0.62	0.54	0.42	0.87
std	6.64	6.07	5.87	5.93	3.53
skew	−0.11	−0.30	0.34	0.54	0.12
kurt	4.38	3.91	3.57	5.10	4.20
mdd (%)	0.07	0.09	0.08	0.19	0.06
%Δ Sharpe	45.6	102	243	7625	17.7
w_{GAP} *(%)*	50.0	50.0	50.0	50.0	20.0

strategies from 1999 and 2016. The value strategy performs the worst during this sample, with a Sharpe ratio of essentially zero, while carry performs the best with a Sharpe ratio of 0.58. We also consider a strategy that combines the above four canonical strategies with equal weights (*EW*), which generates a Sharpe ratio of 0.74—higher than each individual strategy by exploiting (albeit simplistically with equal weights) the imperfect correlation of returns across the individual strategies.

In essence, the results in Panel A of Table 6 provide us with a benchmark on performance of standard currency strategies, and we ask whether combining them with the GAP_{CS} strategy improves performance and to what extent. We report results in Panel B of Table 6, both when we combine each individual strategy with GAP_{CS} and when we add GAP_{CS} to the equally weighted strategy alongside carry, dollar carry, momentum, and value. The results indicate that adding the GAP_{CS} strategy to this menu of strategies delivers substantially higher Sharpe ratios. For example, the Sharpe ratio of the carry trade improves from 0.58 to 0.85, and the equally weighted strategy that includes all four benchmark strategies and GAP_{CS} delivers a Sharpe ratio of 0.87, in contrast to 0.74 that is obtained when GAP_{CS} is excluded.

Overall, we view these findings as a confirmation of the value that the GAP_{CS} strategy adds when included in a currency portfolio, driven by its desirable return and correlation properties with existing currency-based strategies.

5. Asset pricing and implications

In this section, we begin by investigating if a range of alternative pricing models can explain the returns generated by output-gap-sorted portfolios. The purpose of this analysis is to evaluate whether the relationship between currency returns and business cycles can be understood from a risk-return perspective. We go on to consider the role that business cycles could also play as a novel source of risk.

5.1. Methodology

We denote the discrete excess returns on portfolio j in period t as RX_t^j. In the absence of arbitrage opportunities, risk-adjusted excess returns have a price of zero and satisfy the following Euler equation:

$$E_t[M_{t+1} RX_{t+1}^j] = 0, \tag{7}$$

where M_{t+1} is the stochastic discount factor (SDF). A vast literature on currency asset pricing, starting from Lustig and Verdelhan (2007) and including Lustig et al. (2011), Menkhoff et al. (2012a), Della Corte et al. (2016b), and many others, considers an SDF specification that is linear in the pricing factors f_{t+1}, given by

$$M_{t+1} = 1 - b'(f_{t+1} - \mu), \tag{8}$$

where b is the vector of factor loadings and μ denotes the factor means. When relying on Eq. (8), the resulting asset pricing tests are unconditional because the SDF factor loadings in b are assumed to be time invariant (i.e., $b_t = b$). However, in a more general setting the Euler Eq. (7) implies SDF parameters that are time varying, which means the covariance between excess returns and the SDF is conditional (Hansen and Richard, 1987). If this is the case, the unconditional version of the model with constant b is misspecified, potentially resulting in biased alphas (see Boguth et al., 2011). We discuss and analyze the more general case of conditional asset pricing tests later while beginning our analysis on the basis of unconditional asset pricing tests in line with the relevant literature.

The above SDF specification implies a beta pricing model in which the expected excess return on portfolio j is equal to the factor risk price λ times the risk quantities β^j. The beta pricing model is defined as

$$E[RX^j] = \lambda' \beta^j, \tag{9}$$

where the market price of risk $\lambda = \Sigma_f b$ can be obtained via the factor loadings b. $\Sigma_f = E\left[(f_t - \mu)(f_t - \mu)'\right]$ is the

Table 7
Pricing business cycle portfolios.
The table presents cross-sectional asset pricing results. We construct various two-factor linear SDFs that include the *DOL* factor plus a second pricing factor, including “slope” risk (HML_{FX}), global imbalance risk (*IMB*), volatility risk (*VOL*), and the GAP_{CS} factor. In each model, we price five currency portfolios sorted on output gaps using real-time information. We report generalized method of moments (GMM) one-step estimates of factor loadings on the pricing kernel (*b's*) and prices of factor risk (λ*'s*). The superscripts *, **, *** represent significance of the coefficients at the 10%, 5%, and 1% significance levels using Newey and West (1987) corrected standard errors (reported in parentheses). We also report goodness-of-fit statistics for each model including the adjusted R^2 statistic, root mean squared pricing error (RMSE), and the Hansen-Jagannathan distance statistic (HJ_{dist}) with simulated *p*-values in brackets. The HJ_{dist} statistic measures the distance between the estimated pricing kernel and the efficient set of permissible pricing kernels. A *p*-value less than 5% indicates the null hypothesis that the pricing kernel is efficient and can be rejected at the 95% confidence level. We provide full details of the pricing factors in Section 5. The sample runs from December 1999 to January 2016.

	SDF **loadings** (b)		Risk **prices** (λ)		Model fit		
	DOL	FAC	DOL	FAC	Adj.R^2	RMSE	HJ_{dist}
DOL + HML_{FX}	0.22	0.19	0.02	0.03	−0.78	1.69	0.22
	(0.26)	(0.72)	(0.02)	(0.10)			[0.03]
DOL + IMB	−1.39	7.52	0.04*	0.26**	−0.15	1.59	0.19
	(1.43)	(5.78)	(0.02)	(0.12)			[0.61]
DOL + VOL	−3.21	−40.5	0.03	−0.03**	−0.19	1.51	0.21
	(2.86)	(31.6)	(0.02)	(0.02)			[0.46]
DOL + GAP_{CS}	0.08	0.83***	0.02	0.05***	0.44	0.95	0.13
	(0.26)	(0.29)	(0.02)	(0.02)			[0.34]

variance-covariance matrix of the risk factors, and β^j denotes the regression coefficients of each portfolio's excess return RX^j_{t+1} on the risk factors f_{t+1}.

5.2. Risk factors and pricing kernel

The recent literature on cross-sectional asset pricing in currency markets has considered a two-factor SDF. The first risk factor is the expected market excess return, approximated by the average excess return on a portfolio strategy that is long in all foreign currencies with equal weights and short in the domestic currency—the *DOL* factor. For the second risk factor, the literature has employed several return-based factors such as the slope factor (HML_{FX}) of Lustig et al. (2011) or the global volatility risk factor of Menkhoff et al. (2012a).

Following this literature, we start from a two-factor SDF with *DOL* as the first factor and then consider various second factors, including the slope factor (HML_{FX}) proposed by Lustig et al. (2011), the global imbalance factor (*IMB*) of Della Corte et al. (2016b), the volatility factor (*VOL*) of Menkhoff et al. (2012a) in its factor-mimicking version (i.e., the fitted values in a regression of global FX volatility risk on currency returns), and a *GAP* factor constructed simply as the excess return from the GAP_{CS} strategy. The *GAP* factor essentially measures the excess returns generated by sorting currencies on the output gap information and is increasing in the spread of output gaps across the world: it is therefore a measure of the return arising from divergences in business cycles such that the more business cycles diverge across countries, the more the currencies of fast-growing countries appreciate.

Later in the paper we provide a simple model of international financial markets with long-run risk that generates this risk factor in the pricing kernel. Our test assets are the five output-gap-sorted currency portfolios obtained using real-time conditioning information as described in Section 4. We later expand the test assets to consider larger cross-sections given that asset pricing tests tend to have low power in small cross-sections of assets.

5.3. Cross-sectional regressions

Table 7 presents the cross-sectional asset pricing results, including estimates of factor loadings *b* and the market prices of risk λ. The factor loadings *b* are estimated via the generalized method of moments (GMM) of Hansen (1982). To implement GMM, we use the pricing errors as a set of moments and the identity weighting matrix. Since the objective is to test whether the model can explain the cross-section of expected currency excess returns, we only rely on unconditional moments and do not employ instruments. By estimating the first-stage GMM using an identity-weighting matrix, we thus attempt to price all currency portfolios equally well.

We report estimates of *b* and λ and standard errors based on Newey and West (1987). The model's performance is evaluated using the cross-sectional R^2, the root mean squared error (RMSE), and the *HJ* distance measure of Hansen and Jagannathan (1997), which quantifies the mean-squared distance between the SDF of a proposed model and the set of admissible SDFs.[14] To test whether the *HJ* distance is statistically significant, we simulate *p*-values using a weighted sum of χ^2_1 distributed random variables (see, Jagannathan and Wang, 1996; Ren and Shimotsu, 2009). The *p*-values of the *HJ* distance measure are reported in brackets.

In Table 7 we report results for two-factor SDF models that include *DOL* and, in turn, the carry factor (HML_{FX}), the

[14] Note that the *HJ* calculation is essentially a GMM application with the important difference that the (non-optimal) weighting matrix is equal to the inverse of the second moment matrix of test asset returns, not the identity matrix; see the Internet Appendix for details on the technical aspects of the asset pricing methods employed in this section.

volatility risk factor (*VOL*), the global imbalance risk factor (*IMB*), and the *GAP* factor. The results suggest that none of the factor loadings are statistically significant at conventional significance levels with the exception of the *GAP* factor, which displays strong statistical significance both in terms of the factor loading and price of risk. The models involving HML_{FX}, *VOL*, and *IMB* also display poor explanatory power in terms of the adjusted R^2 (always negative), which is surprising considering the relative ease in achieving high R^2 statistics when test assets are characterized by a strong factor structure (Lewellen et al., 2010).

In contrast, the SDF involving *DOL* and *GAP* generates an adjusted R^2 of 44% and a substantially lower RMSE relative to other SDF specifications, indicating that the pricing errors are much lower. The *p*-values from the *HJ* distance measure are always above 5% with the exception of the SDF involving HML_{FX}. However, this finding is likely due to the low power of the *HJ* statistic in our small cross-section of five test assets, and in the absence of statistical significance of factor loadings for all risk factors other than *GAP*, this result cannot be viewed as supportive of these pricing models.

Overall, the results in Table 7 suggest that the only factor that can price the currency excess returns obtained from sorting on output gaps is *GAP* and conventional risk factors from the currency literature are not priced. This finding highlights the novelty of the returns and the need for alternative risk factors to account for this cross-section of asset returns.

5.4. A business cycle factor?

Next, we consider in more detail the possibility that currency excess returns reflect compensation for risk linked to the relative state of business cycle conditions. The theoretical link between aggregate macroeconomic conditions and asset prices is fundamental to the study of asset pricing, and most classes of risk-based models require the SDF to be a function of the business cycle (see Cochrane, 2017, for a comprehensive review and reconciliation of the link between business cycle variables and asset pricing models).[15] Specifically, we carry out asset pricing tests for two SDF specifications: a two-factor model including *DOL* and HML_{FX}, which is the most common benchmark in the literature since its introduction by Lustig et al. (2011), and a three-factor model that also includes the *GAP* factor. This allows us to gauge the incremental pricing power of a business cycle factor beyond the two-factor benchmark.

5.4.1. Test portfolios

We consider two sets of test portfolios, increasing in the number of portfolios. Recall that we first considered the five output-gap-sorted portfolios (as in Table 7), which constitute a small set of test assets for the purpose of asset pricing tests. However, Lewellen et al. (2010) show that a strong factor structure in test asset returns can give rise to misleading results in empirical work, and this outcome is especially the case in small cross-sections. Therefore, we now conduct asset pricing tests on the following two sets of portfolios: 10 portfolios sorted on currency value and momentum (i.e., out-of-sample test assets in which the sorting variable is neither carry nor the output gap), a larger cross-section of 20 portfolios that comprises the 5 portfolios sorted on output gap, 5 portfolios sorted on forward premia (carry), 5 portfolios sorted on momentum, and 5 portfolios sorted on value. We conduct the asset pricing tests excluding the pricing factors as test assets (Panel A of Table 8) and including them (Panel B of Table 8). Lewellen et al. (2010) advocate adding risk factors as test assets to ensure the factors price themselves (i.e., $\lambda \approx E[R_{factor}]$).

5.4.2. Cross-sectional regressions

Starting from Panel A of Table 8, we ask whether a two-factor model including *DOL* and HML_{FX} can price the two sets of test assets described above. We focus our interest on the sign and the statistical significance of the market price of risk λ attached to the HML_{FX} factor and of the associated factor loading *b*. We know from Table 7 that this SDF specification cannot price the returns from output-gap-sorted portfolios. We find that this SDF, which is known to be powerful at pricing carry portfolios, also does not explain satisfactorily the other cross-sections considered. Specifically, the factor loading on HML_{FX} is not statistically different from zero, and the adjusted R^2 is low. The *HJ* distance test does not indicate a rejection of the model in two cases, but with an insignificant factor loading and low R^2, the *HJ* distance result cannot be considered as supportive of the SDF.

When augmenting the SDF specification with the *GAP* factor, we find that both the loading and the price of risk for the *GAP* factor enter with positive and statistically significant coefficients. Moreover, the factor loading on HML_{FX} continues to be statistically insignificant. The adjusted R^2 for the three-factor model including the *DOL*, HML_{FX}, and *GAP* factors is substantially higher (in a range between 59% and 62%), and the RMSE is substantially lower than the two-factor specification that excludes *GAP*.

In Panel B of Table 8 we carry out the same tests while augmenting the test assets to include HML_{FX} in the two-factor SDF and both HML_{FX} and *GAP* in the three-factor SDF. The results are similar to those in Panel A. Specifically, while we observe statistically significant risk prices for HML_{FX}, this finding is due to the inclusion of HML_{FX} as a test asset. More importantly, the results from the three-factor model indicate that the addition of the *GAP* factor leads to strongly statistically significant factor loadings and risk prices on *GAP*, much higher cross-sectional R^2 statistics and far lower RMSEs than in the two-factor specification.[16]

[15] In complementary work, Maurer et al. (2019) use the currency market as a setting to empirically estimate country-specific SDFs and show a linear relationship with domestic output gaps.

[16] In Internet Appendix Tables A9 and A10 we present the equivalent results to those reported in Tables 7 and 8 when employing Fama-MacBeth estimation techniques, while in Tables A11 and A12 we present results in which we replace HML_{FX} with either the global imbalance factor (*IMB*) of Della Corte et al. (2016b) or the global volatility factor (*VOL*) proposed by Menkhoff et al. (2012a). In Table A13, we present the asset pricing results

Table 8

Asset pricing using DOL, HML_{FX}, and GAP_{CS} as pricing factors.

The table presents cross-sectional asset pricing results for two sets of test portfolios. The SDF is constructed as a linear combination of *DOL* and HML_{FX} (two pricing factors, left side) and *DOL*, HML_{FX}, and GAP_{CS} (three pricing factors, righ side). In Panel B, we also include HML_{FX} and GAP_{CS} as test assets. We report generalized method of moments (GMM) one-step estimates of factor loadings on the pricing kernel (*b's*) and prices of factor risk (λ*'s*). The superscripts *, **, *** represent significance of the coefficients at the 10%, 5%, and 1% significance levels using Newey and West (1987) corrected standard errors (reported in parentheses). In addition, we report goodness-of-fit statistics for each model including the adjusted R^2 statistic, root mean squared pricing error (RMSE), and the Hansen-Jagannathan distance statistic (HJ_{dist}) with simulated *p*-values in brackets. The HJ_{dist} statistic measures the distance between the estimated pricing kernel and the efficient set of permissible pricing kernels. A *p*-value less than 5% indicates the null hypothesis that the pricing kernel is efficient can be rejected at the 95% confidence level. The sample runs from December 1999 to January 2016.

Panel A: Excluding pricing factors as test portfolios

	2 pricing factors (DOL + HML_{FX})							**3 pricing factors** (DOL + HML_{FX} + GAP_{CS})								
	Loadings (b)		**Risk prices** (λ)		**Model fit**			**Loadings** (b)			**Risk prices** (λ)			**Model fit**		
	DOL	HML_{FX}	DOL	HML_{FX}	Adj.R^2	RMSE	HJ_{dist}	DOL	HML_{FX}	GAP_{CS}	DOL	HML_{FX}	GAP_{CS}	Adj.R^2	RMSE	HJ_{dist}
10 TPs	0.19	0.08	0.02	0.01	−0.35	1.32	0.22	−0.31	0.43	2.57**	0.02	0.07	0.14***	0.59	0.67	0.16
(val, mom)	(0.26)	(0.32)	(0.02)	(0.04)			[0.81]	(0.36)	(0.41)	(1.01)	(0.02)	(0.05)	(0.05)			[0.96]
20 TPs	0.18	0.36	0.02	0.05*	0.26	1.36	0.33	−0.02	0.35	1.05***	0.02	0.06*	0.06***	0.62	0.94	0.30
(gap, car,	(0.27)	(0.22)	(0.02)	(0.03)			[0.99]	(0.28)	(0.22)	(0.32)	(0.02)	(0.03)	(0.02)			[0.99]
val, mom)																

Panel B: Including pricing factors as test portfolios

	2 pricing factors (DOL + HML_{FX})							**3 pricing factors** (DOL + HML_{FX} + GAP_{CS})								
	Loadings (b)		**Risk prices** (λ)		**Model fit**			**Loadings** (b)			**Risk prices** (λ)			**Model fit**		
	DOL	HML_{FX}	DOL	HML_{FX}	Adj.R^2	RMSE	HJ_{dist}	DOL	HML_{FX}	GAP_{CS}	DOL	HML_{FX}	GAP_{CS}	Adj.R^2	RMSE	HJ_{dist}
10 TPs	0.16	0.41*	0.02	0.06**	0.31	1.37	0.23	−0.02	0.40*	0.97***	0.02	0.06**	0.06***	0.65	0.96	0.19
(val, mom)	(0.28)	(0.22)	(0.02)	(0.03)			[0.80]	(0.28)	(0.21)	(0.30)	(0.02)	(0.03)	(0.02)			[0.91]
20 TPs	0.17	0.41*	0.02	0.06**	0.46	1.34	0.69	0.00	0.40*	0.91***	0.02	0.06**	0.05***	0.74	0.92	0.69
(gap, car,	(0.28)	(0.22)	(0.02)	(0.03)			[0.91]	(0.28)	(0.21)	(0.30)	(0.02)	(0.03)	(0.02)			[0.98]
val, mom)																

5.5. Conditional asset pricing

In this section we consider conditional tests based on an SDF specification that allows for time-varying loadings b_t:

$$M_{t+1} = 1 - b_t'(f_{t+1} - \mu). \tag{10}$$

Other examples of conditional asset pricing tests in the currency asset pricing literature include Lustig et al. (2011) in the context of carry portfolios and Menkhoff et al. (2016) (Section III.C) for portfolios sorted on order flow.

Given Eq. (10), risk equals the conditional exposure to risk factors given the information available to investors, which implies that the covariance between excess returns and the SDF is conditional (Hansen and Richard, 1987). If this is the case, the unconditional version of the model with constant b is mispecified, and alphas are biased due to underconditioning on only a subset of the investors information. This problem can be mitigated, for example, by allowing the SDF loadings to equal realized betas estimated from rolling-window regressions. The key implication of underconditioning is that, in the presence of conditional risk exposures, unconditional asset pricing tests lead to a bias in the estimated alpha. An overconditioning bias can also arise, however, when the econometrician uses a conditional risk proxy that is not entirely in the information set of investors (see Boguth et al., 2011), for example, when using contemporaneous realized betas as proxies for conditional risk since these betas cannot be known in real time.

In the context of our paper, it could be that, for example, the excess returns from the high output gap currency portfolio (strong economies) are more risky in bad times, in which case the unconditional version of the model with constant b is mispecified because time variation in the loadings of the portfolios on the SDF is not permitted. More generally, it is possible that time-varying differences in risk exposures can account for the inability of asset pricing models, not including the business cycle factor, to explain output-gap-portfolio returns, as shown in Section 5.1. We investigate this possibility by carrying out conditional asset pricing tests that help in addressing both the underconditioning and overconditioning biases, using as test assets the five portfolios sorted on output gap and the excess returns from the GAP_{CS} strategy (i.e., P_5 - P_1).

Specifically, we follow the two-step instrumental variable (IV) procedure proposed by Boguth et al. (2011). In the first step we estimate the contemporaneous exposure of the excess returns to currency portfolio j to currency factors at time t. This step helps to overcome the underconditioning bias by allowing for a dynamic relationship between factors and returns. In the case of the two-factor pricing model, which includes DOL and HML_{FX}, we estimate

$$RX_t^j = \alpha + \beta_{1,t}^c DOL_t + \beta_{2,t}^c HML_{FX,t} + \varepsilon_t. \tag{11}$$

We consider different rolling windows over which to estimate the slope coefficients in the above regression, including 12, 18, 24, 30, and 36 months. The first step potentially generates an overconditioning bias, and hence we do not use the betas estimated in the first step when estimating alphas. Instead, in the second step, we forecast the betas estimated in the first step using a vector ($Z_{n,t-1}$) of predictor variables:

$$\beta_{k,t}^c = \gamma_k + \sum_{n=1}^{N} \gamma_n Z_{n,t-1}' + \epsilon_{k,t}. \tag{12}$$

We include as predictor variables the one-period lagged value of beta as well as variables that have been found to predict dollar and carry factor returns including global FX volatility (Bakshi and Panayotov, 2013), the commodity return on the Commodity Research Bureau's raw industrials index (Bakshi and Panayotov, 2013), the Treasury-Eurodollar (TED) spread (Brunnermeier et al., 2009), and the average forward discount (Lustig et al., 2014). The fitted value from this regression, $\hat{\beta}_{k,t}^c$, is used to estimate the time-series IV alpha (α^{IV}) that Boguth et al. (2011) show mitigates against biases stemming from both over and underconditioning:

$$RX_t^j = \alpha^{IV} + (\phi_0^{DOL} + \phi_1^{DOL}\hat{\beta}_{1,t}^c)DOL_t + (\phi_0^{HML} + \phi_1^{HML}\hat{\beta}_{2,t}^c)HML_{FX,t} + u_t. \tag{13}$$

In Panel A of Table 9 we report the IV alphas for the five output-gap-sorted portfolios and the GAP_{CS} strategy excess returns. In addition to the two-factor model with DOL and HML_{FX} (left-hand side), we report results from the two-factor model with DOL and GAP_{CS} as risk factors (middle) and the three-factor model with DOL, HML_{FX}, and GAP_{CS} (right-hand side). We also report the χ^2 test statistic and associated p-value for the test that the alphas are jointly equal to zero. A p-value below 0.05 indicates that the null hypothesis, which is that all alphas are jointly zero, can be rejected at the 5% level of significance. In Panel B of Table 9 we report the χ^2 test statistic for the equivalent test on a larger set of 20 test portfolios including portfolios sorted by output gaps, forward premia, momentum, and value.

Starting from Panel A, for the two-factor model including DOL and HML_{FX}, the χ^2 test statistic rejects the null hypothesis that the alphas are jointly zero for every one of the five rolling windows considered for the beta estimation. Examining the alphas of the individual portfolios, we can see that this result is driven by statistically significant alphas in P_5 and sometimes P_4 as well as significant alphas in the GAP_{CS} excess returns. However, the model specifications that include GAP_{CS} (middle and right-hand side of Table 9) generates alphas that are not statistically significantly different from zero, with the χ^2 test statistic becoming considerably smaller and its associated p-values comfortably above the 5% level. These results are consistent with the results from unconditional asset pricing tests reported earlier, suggesting that the most common benchmark for currency asset pricing, namely the two-factor model of Lustig et al. (2011), is unable to price the output-gap-sorted portfolios and a business cycle factor is needed in the SDF specification to price these portfolios.

The results are qualitatively identical when we employ all 20 test portfolios including portfolios sorted by output

for the same set of test portfolios but for which the SDF is a two-factor linear combination of DOL and GAP_{CS}. All of these additional tests confirm that the GAP factor is priced in each of the sets of test assets considered.

Table 9
Conditional asset pricing.
The table presents results from conditional asset pricing tests. In Panel A, we present annualized two-step instrumental variable (IV) alphas following the procedure of Boguth et al. (2011) on five output-gap-sorted portfolios and the GAP_{CS} factor. The contemporaneous exposure of the portfolios to the pricing factors (i.e., betas) are initially estimated over either 12, 18, 24, 30, or 36 months. The instruments used to predict betas include the lagged beta, global foreign exchange volatility, the Commodity Research Bureau's industrial metals index, the Treasury-Eurodollar (TED) spread, and the average forward discount. The superscripts *, **, *** represent significance of the alphas at the 10%, 5%, and 1% significance levels using Newey and West (1987) corrected standard errors. We also report the χ^2 test statistic and associated p-value in square brackets for the test that alphas are jointly equal to zero. A p-value less than 5% indicates the null hypothesis that the alphas are jointly equal to zero can be rejected at the 95% confidence level. In Panel B, we report the χ^2 test statistic and associated p-value in square brackets for the equivalent test when pricing 20 currency portfolios, which include portfolios sorted according to output gaps (x5), forward premia (x5), currency momentum (x5), and currency value (x5). The sample runs from December 1999 to January 2016.

Panel A: GAP portfolios

	2 pricing factors (DOL + HML_{FX})					2 pricing factors (DOL + GAP_{CS})					3 pricing factors (DOL + HML_{FX} + GAP_{CS})				
Months	12	18	24	30	36	12	18	24	30	36	12	18	24	30	36
P_1	−0.89	−1.27	−1.13	−1.16	−0.95	0.58	0.33	0.42	0.59	0.71	0.62	0.25	0.03	0.13	0.34
P_2	−0.85	−0.58	−0.40	−0.49	−0.87	−0.78	−0.40	−0.29	−0.41	−0.66	−1.14	−0.69	−0.50	−0.46	−0.69
P_3	−0.91	−0.76	−0.38	−0.83	−0.76	−0.83	−0.84	−0.77	−1.03	−0.89	−0.82	−0.94	−0.54	−0.89	−0.67
P_4	1.54	1.78*	1.87*	1.88*	1.53	0.37	0.93	0.85	0.98	0.66	1.11	1.19	1.31	1.57	1.19
P_5	2.18**	2.05**	1.62**	1.85**	1.97**	0.58	0.33	0.42	0.59	0.71	0.62	0.25	0.03	0.13	0.34
GAP_{CS}	2.96**	3.18**	2.74*	3.00**	3.12**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
χ^2	14.46	14.50	15.79	13.62	12.65	1.45	2.02	6.25	9.29	5.55	3.25	1.64	2.35	3.70	3.26
	[0.025]	[0.025]	[0.015]	[0.034]	[0.049]	[0.963]	[0.918]	[0.396]	[0.158]	[0.475]	[0.776]	[0.950]	[0.885]	[0.717]	[0.775]

Panel B: All portfolios

	2 pricing factors (DOL + HML_{FX})					**2 pricing factors** (DOL + GAP_{CS})					**3 pricing factors** (DOL + HML_{FX} + GAP_{CS})				
Months	**12**	**18**	**24**	**30**	**36**	**12**	**18**	**24**	**30**	**36**	**12**	**18**	**24**	**30**	**36**
χ^2	31.29	30.70	44.86	36.95	33.53	23.35	17.91	24.49	27.96	19.37	25.06	16.49	17.85	17.09	17.20
	[0.051]	[0.059]	[0.001]	[0.012]	[0.029]	[0.272]	[0.593]	[0.222]	[0.110]	[0.498]	[0.199]	[0.686]	[0.597]	[0.647]	[0.640]

gaps, forward premia, momentum, and value (Panel B of Table 9). Specifically, there is no evidence that the two-factor model with *DOL* and HML_{FX} can price these test assets, with the exception of two cases with short rolling windows of 12 and 18 months, which generate moderately higher *p*-values between 5% and 6%, whereas model specifications that include GAP_{CS} all generate large *p*-values.

Overall, the conditional asset pricing tests corroborate the earlier findings that conventional risk factors used in the currency asset pricing literature are unable to price portfolios sorted on relative output gaps and a business cycle factor implied by our results is priced in a broad currency cross-section.

6. A model for the GAP premium

The asset pricing results suggest that standard risk factors used in the literature cannot explain the returns from currency portfolios that sort on output gaps. In this section we present a simple model that can generate a risk premium associated with relative output gaps across countries, the GAP premium. While this setup abstracts away from trade in the consumption goods market, it constitutes a useful benchmark in the international finance literature, and it has been applied to the analysis of exchange rates' volatility (Colacito and Croce, 2011), international term structure of interest rates (Bansal and Shaliastovich, 2012), gravity in exchange rates' fluctuations (Lustig and Richmond, 2019), and quanto contracts (Kremens and Martin, 2019). We follow the literature and focus on this setup due to its ability to deliver closed form solutions for all the objects of interest and leave a fully fledged general equilibrium analysis to future research.

6.1. Setup of the economy

Preferences. The economy consists of *N* countries. Each country *i* is populated by a representative agent with recursive preferences

$$U_{i,t} = (1-\delta)\log(C_{i,t}) + \delta \cdot \frac{1}{1-\gamma}\log\left(E_t[\exp((1-\gamma)U_{i,t+1})]\right), \tag{14}$$

where $C_{i,t}$, δ, and γ denote the consumption in country *i* at date *t*, the subjective discount factor, and risk aversion, respectively. These preferences correspond to Epstein and Zin (1989) preferences for the case of unit intertemporal elasticity of substitution (henceforth IES).

Consumption dynamics. The logarithm of consumption growth in each country evolves according to the following law of motion:

$$\begin{aligned}\Delta c_{i,t+1} &= \mu_c + z_{i,t} + \sqrt{\sigma_{i,t}}\varepsilon^c_{i,t+1}\\ z_{i,t} &= \rho_z z_{i,t-1} + \varphi_z\sqrt{\bar{\sigma}}\varepsilon^z_{i,t}\\ \sigma_{i,t} &= (1-\rho_\sigma)\bar{\sigma} + \rho_\sigma\sigma_{i,t-1} + \sqrt{\sigma_\sigma}\varepsilon^\sigma_{i,t},\\ &\quad \forall i \in \{1,2,\ldots,N\},\end{aligned} \tag{15}$$

where $z_{i,t}$ is a consumption factor that proxies for the output gap and $\sigma_{i,t}$ is the conditional variance of consumption growth. We show in the Internet Appendix that the output gap estimated using the same methodology that we adopt in our empirical analysis is highly correlated with $z_{i,t}$. In what follows, we proxy the output gap with the $z_{i,t}$ factor.[17]

All shocks within a country are orthogonal to each other with the exception of

$$\rho^{z\sigma}_{i,t} = corr\left(\varepsilon^\sigma_{i,t}, \varepsilon^z_{i,t}\right) = corr\left(\varepsilon^z_{i,t} + \alpha\varepsilon^\nu_{i,t}, \varepsilon^z_{i,t}\right) = \frac{1}{\sqrt{1+\alpha^2}} \geq 0,$$

where $\varepsilon^z_{i,t}$ and $\varepsilon^\nu_{i,t}$ are two independent and identically distributed Normal random variables. To better highlight the key mechanism of the model, we set to zero all the correlations of international shocks with the exception of the international correlation of the shock $\varepsilon^z_{i,t}$ in each country with the shock $\varepsilon^z_{k,t}$ in the base country *k*. Specifically, we assume that this correlation is time varying according to the process:

$$corr\left(\varepsilon^z_{i,t}, \varepsilon^z_{k,t}\right) = \rho^z_{ik,t} = (2\bar{\rho}-1) + 2(1-\bar{\rho})\frac{1}{1+\exp\{z_{i,t}\}}, \tag{16}$$

from which it is easy to show that (i) the unconditional mean of the correlation is $\bar{\rho}$ (i.e., $E\left[\rho^z_{ik,t}\right] = \bar{\rho}$) and (ii) the correlation of each country *i* with the base country *k* declines with the level of the output gap in country *k* (i.e., $\frac{\partial\rho^z_{ik,t}}{\partial z_{i,t}} < 0$). In the Internet Appendix we provide some empirical support for this hypothesized comovement between correlations and output gap.

Financial markets. We assume that there is a complete set of state and date contingent bonds that each investor has access to in frictionless financial markets at each point in time.

6.2. Equilibrium outcomes

Risk-free rates. We show in Lemma 3 of the Internet Appendix that the risk-free rates are equal to

$$r^f_{i,t} = \mu_c - \log(\delta) + z_{i,t} - \left(\gamma - \frac{1}{2}\right)\sigma_{i,t}, \quad \forall i \in \{1,2,\ldots,N\}. \tag{17}$$

Eq. (17) provides an important connection between our model and the empirical evidence provided in the previous sections: given the imperfect degree of correlation between $z_{i,t}$ and $\sigma_{i,t}$, sorting on the level of the risk-free rate is not the same as sorting on the output gap.

[17] The explanation for why the correlation between output gap and the predictive component $z_{i,t}$ is positive is straightforward. Consider the case in which the output gap is measured as the residual from the regression $c_{i,t} = \alpha + \beta c_{i,t-1} + \xi_{i,t}$. Since the estimated output gap, $\hat{\xi}_{i,t}$, is equal to $z_{i,t-1} + \sqrt{\sigma_{i,t-1}}\varepsilon^c_{i,t}$, the positive correlation with the predictive component $z_{i,t}$ follows immediately. In the Internet Appendix, we extend this analysis to the case of a general number of regressors and show that this correlation is generally large. We focus on the correlation with the cyclical component extracted using the methodology proposed by Hamilton (2018) because it coincides with the empirical measure that we adopt for our core set of results. We cannot claim that any trend-cycle decomposition would give rise to a large and/or positive correlation and thank an anonymous referee for pushing us to illustrate this point.

Currency risk premia. The following proposition establishes the connection between sorting the cross-section of currencies according to the output gap and the existence of an excess return.

Proposition 1 (GAP premium). *Let i and j be two countries for which the following condition holds: $z_{i,t} \geq z_{j,t}$. Then the excess return of a strategy that is long the currency of country i and short the currency of country j is*

$$E_t\left[GAP^{ij}_{t+1}\right] = E_t\left[RX^i_{t+1} - RX^j_{t+1}\right] = 2(1-\bar{\rho})\kappa \times \left(\frac{1}{1+\exp\left\{z_{j,t}\right\}} - \frac{1}{1+\exp\left\{z_{i,t}\right\}}\right) > 0, \tag{18}$$

where the coefficient of proportionality κ is equal to

$$\kappa = \delta\left[\frac{\rho_z\sqrt{\bar{\sigma}}}{1-\delta\rho_z} + \frac{1}{2}\frac{(1-\gamma)\sqrt{\sigma_\sigma}}{1-\delta\rho_\sigma}\right]^2 > 0.$$

Proof. See Internet Appendix. □

The interpretation of Proposition 1 is intuitive. The representative investor of the base country k dislikes (likes) periods of low (high) output gap shocks. Given the assumption of complete markets, the log-growth rate of exchange rates is equal to the difference of logarithms of the SDFs of any two countries. The more correlated the output gap of a country with the output gap of country k is, the better hedge it provides against negative output gap shocks in country k. Indeed, in the extreme case in which two countries have perfectly correlated output gap shocks, the exchange rate of their currencies would have no exposure to output gap risk. Given our postulated negative relationship between the output gap of country i and its correlation with the output gap of the base country k, it follows that low output gap countries (high $\rho^z_{ik,t}$) have safe currencies, while high output gap countries (low $\rho^z_{ik,t}$) have risky currencies.

We note that in the absence of time-varying volatility, the GAP premium would be entirely driven by differentials in risk-free rates (this follows immediately from shutting down $\sigma_{i,t}$ in Eq. (17)). However, in the general case discussed here, the presence of time-varying macroeconomic uncertainty breaks this tight relationship. In fact, a country i could have a very strong economy (very positive $z_{i,t}$) and low risk-free rates as long as the level of uncertainty in country i is large and investors are risk-averse enough. In this case, the currency of country i would still command a large GAP premium despite the possibly low risk-free rate. Equivalently, the GAP premium depends crucially on expected currency appreciations, which are going to result in a positive premium notwithstanding a low or sometimes negative interest rate differential.

7. Conclusions

Understanding and measuring the sources of macroeconomic risk that drive asset prices is a fundamental challenge in asset pricing. In this paper, we provide robust empirical evidence that business cycles, proxied by output gaps, are an important determinant of the cross-section of expected currency returns. Our primary result is that the currencies of strong economies (high output gaps) command higher expected returns. The excess returns from a trading strategy that sorts currencies on relative output gaps generates high risk-adjusted returns that are uncorrelated with the excess returns from popular currency investment strategies, thereby providing tangible diversification gains to global investors.

Moreover, we find that a business cycle risk factor that captures the spread in output gaps across countries is priced in the cross-section of currency excess returns that includes portfolios sorted by carry, value, and momentum. In general, these findings are important for the broad theoretical literature seeking to explain the macroeconomic drivers of currency risk premia. We show that a currency GAP premium arises in equilibrium in an international macro-finance model with long-run risk, in which the correlations of the shocks vary over time with the predictive components of consumption. Extending this framework to a fully fledged general equilibrium analysis represents an important direction for future research.

In future work, researchers could explore alternative theoretical mechanisms that can explain the link between business cycles and the cross-section of currency excess returns reported in the paper. The model presented here is only one of potentially several frameworks that can predict these facts. Empirical researchers could also explore alternative ways to measure business cycles, using richer financial and economic data sets, as a fruitful avenue to break new ground in the empirical asset pricing of currency markets.

References

Akram, Q.F., Rime, D., Sarno, L., 2008. Arbitrage in the foreign exchange market: turning on the microscope. J. Int. Econ. 76, 237–253.
Andersen, L., Duffie, D., Song, Y., 2019. Funding value adjustments. J. Finance 74, 145–192.
Asness, C.S., Moskowitz, T.J., Pedersen, L.H., 2013. Value and momentum everywhere. J. Finance 68, 929–985.
Bakshi, G., Panayotov, G., 2013. Predictability of currency carry trades and asset pricing implications. J. Financ. Econ. 110, 139–163.
Bansal, R., Shaliastovich, I., 2012. A long-run risks explanation of predictability puzzles in bond and currency markets. Rev. Finan. Stud. 26, 1–33.
Barroso, P., Santa-Clara, P., 2015. Beyond the carry trade: optimal currency portfolios. J. Financ. Quant. Anal. 50, 1037–1056.
Baxter, M., King, R.G., 1999. Measuring business cycles: approximate band-pass filters for economic time series. Rev. Econ. Stat. 81, 575–593.
Baz, J., Granger, N. M., Harvey, C. R., Le Roux, N., Rattray, S., 2015. Dissecting investment strategies in the cross section and time series. Unpublished working paper. Man Group.
Bekaert, G., Panayotov, G., 2020. Good carry, bad carry. J. Financ. Quant. Anal. 55 (4), 1063–1094.
Berg, K.A., Mark, N.C., 2018. Global macro risks in currency excess returns. J. Empir. Finance 45, 300–315.
Boguth, O., Carlson, M., Fisher, A., Simutin, M., 2011. Conditional risk and performance evaluation: volatility timing, overconditioning, and new estimates of momentum alphas. J. Financ. Econ. 102, 363–389.
Brunnermeier Markus, K., Nagel, S., Pedersen, L.H., 2009. Carry trades and currency crashes. In: Acemoglu, D., Rogoff, K., Woodford, M. (Eds.), Proceedings of the NBER Macroeconomics Annual 2008. University of Chicago Press, Chicago, pp. 313–347.
Cespa, G., Gargano, A., Riddiough, S. J., Sarno, L., 2019. Foreign exchange volume. Unpublished working paper. City University London, University of Cambridge, and The University of Melbourne.

Clarida, R., Gali, J., Gertler, M., 1998. Monetary policy rules in practice: some international evidence. Eur. Econ. Rev. 42, 1033–1067.
Cochrane, J.H., 2005. Asset Pricing: Revised Edition. Princeton University, Princeton.
Cochrane, J.H., 2017. Macro-finance. Rev. Finance 21, 945–985.
Colacito, R., Croce, M., 2011. Risks for the long-run and the real exchange rate. J. Polit. Econ. 119, 153–181.
Colacito, R., Croce, M.M., Gavazzoni, F., Ready, R.C., 2018. Currency risk factors in a recursive multi-country economy. J. Finance 73, 2719–2756.
Cooper, I., Priestley, R., 2009. Time-varying risk premiums and the output gap. Rev. Financ. Stud. 22, 2801–2833.
Dahlquist, M., Hasseltoft, H., 2020. Economic momentum and currency returns. J. Financ. Econ. 136 (1), 152–167.
Della Corte, P., Ramadorai, T., Sarno, L., 2016. Volatility risk premia and exchange rate predictability. J. Financ. Econ. 120, 21–40.
Della Corte, P., Riddiough, S.J., Sarno, L., 2016. Currency premia and global imbalances. Rev. Financ. Stud. 29, 2161–2193.
Du, W., Tepper, A., Verdelhan, A., 2018. Deviations from covered interest rate parity. J. Finance 73, 915–957.
Engel, C., Mark, N.C., West, K.D., 2007. Exchange rate models are not as bad as you think. NBER Macroecon. Ann. 22, 381–441.
Epstein, L.G., Zin, S.E., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. Econometrica 57 (4), 937–969.
Gabaix, X., Maggiori, M., 2015. International liquidity and exchange rate dynamics. Q. J. Econ. 130, 1369–1420.
Gilmore, S., Hayashi, F., 2011. Emerging market currency excess returns. Am. Econ. J. Macroecon. 3, 85–111.
Gómez-Cram, R., 2018. Late to recessions: Stocks and the business cycle. Unpublished working paper. The Wharton School, University of Pennsylvania.
Goyal, A., Jegadeesh, N., 2018. Cross-sectional and time-series tests of return predictability: what is the difference? Rev. Financ. Stud. 31, 1784–1824.
Goyal, A., Saretto, A., 2009. Cross-section of option returns and volatility. J. Financ. Econ. 94, 310–326.
Hamilton, J.D., 2018. Why you should never use the Hodrick-Prescott filter. Rev. Econ. Stat. 100, 831–843.
Hansen, L.P., 1982. Large sample properties of generalized method of moments. Econometrica 50, 1029–1054.
Hansen, L.P., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factor models. J. Finance 52, 557–590.
Hansen, P., Richard, S.F., 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. Econometrica 55, 587–613.
Hassan, T.A., 2013. Country size, currency unions, and international asset returns. J. Finance 68, 2269–2308.
Hodrick, R. J., Prescott, E. C., 1980. Postwar U.S. business cycles: an empirical investigation. Unpublished working paper. Carnegie-Mellon University.
Hodrick, R.J., Prescott, E.C., 1997. Postwar U.S. business cycles: an empirical investigation. J. Money, Credit Bank. 29, 1–16.
Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. J. Finance 51, 3–53.
Jordà, Ò., Taylor, A.M., 2012. The carry trade and fundamentals: nothing to fear but FEER itself. J. Int. Econ. 88, 74–90.
Kremens, L., Martin, I., 2019. The quanto theory of exchange rates. Am. Econ. Rev. 109, 810–843.
Lettau, M., Maggiori, M., Weber, M., 2014. Conditional risk premia in currency markets and other asset classes. J. Financ. Econ. 114, 197–225.
Lewellen, J., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset pricing tests. J. Financ. Econ. 96, 175–194.
Lustig, H., Richmond, R.J., 2019. Gravity in the exchange rate factor structure. Rev. Financ. Stud. forthcoming.
Lustig, H., Roussanov, N., Verdelhan, A., 2011. Common risk factors in currency markets. Rev. Financ. Stud. 24, 3731–3777.
Lustig, H., Roussanov, N., Verdelhan, A., 2014. Countercyclical currency risk premia. J. Financ. Econ. 111, 527–553.
Lustig, H., Verdelhan, A., 2007. The cross-section of foreign currency risk premia and consumption growth risk. Am. Econ. Rev. 97, 89–117.
Maggiori, M., 2017. Financial intermediation, international risk sharing, and reserve currencies. Am. Econ. Rev. 107, 3038–3071.
Mark, N.C., 1995. Exchange rates and fundamentals: evidence on long horizon predictability. Am. Econ. Rev. 85, 201–218.
Maurer, T., Tô, T.-D., Tran, N.-K., 2018. Optimal factor strategy in FX markets. Unpublished working paper. Washington University in St. Louis and the University of New South Wales.
Maurer, T., To, T.-D., Tran, N.-K., 2019. Pricing risks across currency denominations. Manag. Sci. 65, 5308–5336.
Meese, R.A., Rogoff, K., 1983. Empirical exchange rate models of the seventies: do they fit out of sample? J. Int. Econ. 14, 3–24.
Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A., 2012. Carry trades and global foreign exchange volatility. J. Finance 67, 681–718.
Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A., 2012. Currency momentum strategies. J. Financ. Econ. 106, 660–684.
Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A., 2016. Information flows in foreign exchange markets: dissecting customer currency trades. J. Finance 71, 601–634.
Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A., 2017. Currency value. Rev. Financ. Stud. 30, 416–441.
Molodtsova, T., Nikolsko-Rzhevskyy, A., Papell, D.H., 2008. Taylor rules with real-time data: a tale of two countries and one exchange rate. J. Monetary Econ. 55, 167–180.
Moskowitz, T.J., Ooi, Y.H., Pedersen, L.H., 2012. Time-series momentum. J. Financ. Econ. 104, 228–250.
Mueller, P., Stathopoulos, A., Vedolin, A., 2017. International correlation risk. J. Financ. Econ. 126, 270–299.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Ready, R., Roussanov, N., Ward, C., 2017. Commodity trade and the carry trade: a tale of two countries. J. Finance 72, 2629–2684.
Ren, Y., Shimotsu, K., 2009. Improvement in finite sample properties of the Hansen-Jagannathan distance test. J. Empir. Finance 16, 483–506.
Rogoff, K. S., Stavrakeva, V., 2008. The continuing puzzle of short horizon exchange rate forecasting. Unpublished working paper. Harvard University and London Business School.
Rossi, B., 2013. Exchange rate predictability. J. Econ. Literature 51, 1063–1119.
Verdelhan, A., 2018. The share of systematic variation in bilateral exchange rates. J. Finance 73, 375–418.

How Constraining Are Limits to Arbitrage?

Alexander Ljungqvist
Stern School of Business, New York University and NBER

Wenlan Qian
NUS Business School

We document the existence of a strategy designed to circumvent limits to arbitrage. Faced with short-sale constraints and noise trader risk, small arbitrageurs publicly reveal their information to induce the target's shareholders ("the longs") to sell, thereby accelerating price discovery. Using data for 124 short-sale campaigns in the United States between 2006 and 2011, we show that investors respond strongly to the information, with spikes in SEC filing views, volatility, order imbalances, realized spreads, turnover, and selling by the longs. Share prices fall by an aggregate $14.8 billion. Our findings imply that even extreme short-sale constraints need not constrain arbitrage. (*JEL* G12, G14, G23, G2)

Received June 18, 2014; accepted March 11, 2016 by Editor Andrew Karolyi.

Imagine you have reason to believe that a particular stock is severely overvalued. Unfortunately, the market does not share your assessment: stock lenders are demanding high shorting fees, there are few stocks available for borrowing, and you have shallow pockets. What do you do?

Received wisdom tells you to walk away: the short-sale constraints you face are so severe that you cannot take a large enough short position to profitably correct the overvaluation, and even if you could, you would face the risk of the mispricing getting worse in the short-run, triggering margin calls, thereby forcing early liquidation of your position at a loss. This combination of short-sale constraints and what DeLong and others (1990) call noise trader risk results in mispricing persisting, which limits the market's informational efficiency.

The authors are grateful to Malcolm Baker, Utpal Bhattacharya (our HKUST discussant), Alon Brav, Lauren Cohen (our NBER discussant), Patricia Dechow, Stephen Figlewski, Robin Greenwood (our AEA discussant), David Hirshleifer, Andrew Karolyi (the Editor), Adam Kolasinski, Charles Lee (our ABFER discussant), David McLean (our Cavalcade discussant), Randall Morck, Abhiroop Mukherjee, Stefan Nagel, Ramana Nanda, Marco Pagano, Tarun Ramadorai, Anna Scherbina (our EFA discussant), Andrei Shleifer, and Erik Stafford, as well as to two anonymous reviewers and audiences at various seminars and conferences. Special thanks go to Scott Bauguess at the SEC for facilitating access to the EDGAR download statistics. The authors thank Zhou Li and Weibiao Xu for excellent research assistance. A. L. gratefully acknowledges the generous hospitality of NUS Business School while working on this project. W. Q. gratefully acknowledges financial support from a CAMRI applied finance research grant. All data used in this research come from publicly available sources. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to Alexander Ljungqvist, NYU Stern School of Business, Suite 9-160, 44 West Fourth Street, New York, NY 10012-1126; telephone: (212)998-0304. E-mail: aljungqv@stern.nyu.edu.

doi:10.1093/rfs/hhw028 Advance Access publication June 7, 2016

Received wisdom needs reevaluating. We document the existence of an increasingly popular informational arbitrage strategy that is designed to circumvent limits to arbitrage of this kind. The strategy turns conventional arbitrage on its head. Instead of identifying a suitable target and quietly shorting its stock until the price adjusts, the arbitrageur publicly reveals her information. The aim is to engage the one group of investors who are not constrained: the target's current shareholders ("the longs"). If the longs can be persuaded to sell, this will not only correct the mispricing but, in the process, will also accelerate price discovery and so reduce the duration (and hence risk) of the arbitrage gap.[1] Prominent examples of this arbitrage strategy include Muddy Waters' June 2011 report on Sino-Forest, a Chinese company listed in Toronto, which a year later went bankrupt, and Citron Research's January 2011 report on China MediaExpress, a Chinese company delisted from NASDAQ 6 weeks later.[2]

For the strategy to work, it is critical that the arbitrageurs (or "arbs" for short) provide information that is both credible and cannot be ignored. They do so by way of detailed reports, which they make available for free and to which they draw maximum media attention. The reports contain a wealth of new facts, often assembled with the help of forensic accountants and professional investigators, and tend to focus on questionable governance practices and aggressive accounting (sometimes bordering on fraud). They often include "smoking guns" in the form of recorded phone calls, video surveillance, and photographs. By presenting new facts that are impossible to ignore or dismiss out of hand, the arbs hope to induce a stampede (similar to a bank run), in which no long investor wants to be the last to sell.

Our sample consists of 31 arbitrageurs who are either individuals or small boutique hedge funds. We show that the arbs target companies with the most potential for overpricing, namely those with high idiosyncratic volatility (Pontiff 2006) or severe short-sale constraints in the form of a low and inelastic supply of shortable stock, high lending fees, and expensive put options. Despite these constraints, the arbs manage to correct mispricing. On average, the prices of the 124 target companies in our sample fall by around 7.5% when a report is released to the public and then continue to drift lower as further negative information comes to light: down by 21.4% to 26.2% over 3 months and by 42.3% to 47.3% over 12 months (depending on the benchmark used). Based on

[1] A few words on terminology: it is useful to distinguish between informational arbitrage, which is risky because there is no perfectly correlated asset with which to hedge the short position in the overvalued stock, and statistical arbitrage as defined by Bondarenko (2003). To avoid clutter, we refer to the former as simply *arbitrage*. We follow the literature on risky arbitrage and use the term *arbitrageur* rather than the more generic term *short seller*. This helps distinguish our arbitrageurs from other short sellers (discussed in Section 4.4) who might mimic their trades in the spirit of Abreu and Brunnermeier (2002).

[2] Around half of the companies in our sample are Chinese firms listed in the United States; the other half are American.

the 3-month change in market value, we estimate that the average target was overvalued by an economically meaningful $119.7 million.

The observed price corrections appear sufficiently large to make arbitrage profitable. We estimate that the arbs earn cumulative abnormal profits averaging 24.1% over 3 months, net of short-sale fees, and risk-adjusted using the three Fama-French factors and a momentum factor. In dollar terms, these trading profits amount to $241,000 in gains for every $1 million in shorts. Given their shallow pockets and the unusually "tight" shorting conditions for their targets, the arbs can only short a few million dollars, which we estimate is nonetheless likely sufficient to cover the information production costs they incur in identifying and investigating their targets.

Much of our analysis focuses on empirically identifying the mechanism that allows these small information producers to move prices and thereby make profits on their short positions. Consistent with our argument that their aim is to induce unconstrained long investors to trade on their behalf, we find that investors appear to pay attention. Investors show abnormally high interest in targets' SEC filings, with filing views up by an average of 80.7% the day the arbs release their reports. Volatility spikes up by 236% on average, as market participants process the information and share turnover spikes by even more, up by 339%.

We find that very little of the price correction comes from the short side of the market, consistent with our argument that the arbs seek to induce the longs to trade on their behalf. While we see evidence of the arbs building significant (albeit relatively small) short positions *before* releasing their reports, once the reports are out, there is no further abnormal shorting activity. This reflects a dramatic worsening in short-sale constraints: fees for initiating new shorts jump by 50% on average when a report is released, putting them in the 79th percentile of the distribution of all stocks in CRSP. Over the next 3 months, fees continue to climb, reaching the 86th percentile (up by 174% compared to the prerelease period). At the same time, the supply of stock available for borrowing falls, and put options become unusually expensive.

The price falls instead come about because of the trading behavior of the one group of investors who are unconstrained: investors with long positions in the targets. Our data show that turnover involving longs spikes by 524% on average when an arb releases her report; with the average institutional investor who is long, the target's stock (according to ANcerno) selling 43.6% of its holding over the course of the report day and fully 65% of the average target's long investors being net sellers that day. These patterns imply a massive increase in liquidity demand, which, coupled with the fact that widespread selling by the longs causes order imbalances to build up, results in liquidity providers increasing realized bid-ask spreads by 94.8% on average.

Clearly, the reports should only induce the longs to sell if the information they contain is credible. When we condition investors' responses on an arb's track record, we find that only arbs with a history of making credible claims (those with reports that have proved profitable in the past) are able to induce the longs to sell and thereby put pressure on a target's share price. Moreover, it is only reports by credible arbs that generate profits, net of shorting fees: absent credibility, prices do not fall significantly.

Credibility matters, but not absolutely. To be listened to, an arb also has to have something new to say. We show that reports that present new facts previously unknown to investors result in longs selling and rapid price corrections, whereas reports that merely reinterpret known data do not. As a result, only reports that contain "scoops" generate significant profits for the arbs.

There is no indication that the arbs in our sample are "bear raiders" intent on manipulating share prices by disseminating falsehoods (a strategy sometimes called "short and distort"). In fact, subsequent events usually prove the arbs factually right. For example, 50% of targets are later delisted; 47% replace their auditors or see their auditors resign; and 23% restate earnings. Investigations by third parties such as the SEC, the Department of Justice, or a stock exchange come to similar conclusions as the reports in fully 90% of the cases. This is remarkable given that our sample is unbiased: we have a complete, *ex ante* list of target companies (rather than a self-reported selected list of only those that made money for the arbs).

Our study contributes to the asset-pricing and behavioral-finance literatures by showing that arbitrage need not be limited even when short-sale constraints are formidable, noise trader risk is high, and the arbitrageur has only limited capital and so little hope to correct mispricing on her own. In fact, the arbs in our sample are attracted to firms with extreme short-sale constraints (such as the targets in our sample) precisely because they are most likely to be overpriced. But the short-sale constraints are not meaningfully binding (beyond limiting the size of the arbs' short position): it is by inducing the unconstrained longs to sell that the arbs correct the mispricing and turn a profit.

Our study also sheds light on how short sellers produce and transmit information. There is little prior evidence on what short sellers know or how they acquire information. Our data allow us to observe the information-discovery process at the level of individual information producers and to study how the information the arbs discover is then incorporated in security prices.[3]

Our findings highlight the economic importance of small arbitrageurs, such as those in our sample, to the market's informational efficiency: while the extent of overvaluation—an aggregate of $14.8 billion across the 124 targets—is large, the magnitude of the arbitrage opportunity is relatively small, owing

[3] Previous research on short sales finds that they predict lower subsequent returns (Figlewski 1981; Desai et al. 2002; Diether, Lee, and Werner 2009) and target overvalued firms (Dechow et al. 2001; Hirshleifer, Teoh, and Yu 2011), and often precede negative corporate events such as negative earnings announcements, analyst downgrades, or financial misconduct (Christophe, Ferri, and Angel 2004; Desai, Krishnamurthy, and Venkataraman 2006; Christophe, Ferri, and Hsieh 2010; Karpoff and Lou 2010). However, Engelberg, Reed, and Ringgenberg (2012) find that the short sellers in their sample are more likely to trade *after* corporate events. Another strand of the literature documents that short-sale constraints lead to overvaluation by withholding negative information from the market (Lamont and Thaler 2003; Nagel 2005; Beneish, Lee, and Nichols 2013).

to the difficulty of shorting these particular targets. If it is too small to attract the attention of deep-pocketed arbitrageurs, overvaluation will persist unless small information producers, such as the arbs in our sample, can go after it profitably.

What can and cannot be arbitraged this way? Because persuading the longs to sell is key, arbs are unlikely to target companies for which new and persuasive facts cannot easily be discovered at reasonable cost. This may seem to rule out the possibility that the arbs can help prick asset-pricing bubbles. Nonetheless, they may inadvertently contribute to doing so. As Kovbasyuk and Pagano (2014) show in a contemporaneous theory paper, "advertising" arbitrage opportunities can be particularly effective when the mispricing is caused by limited attention: by focusing unconstrained investors' attention on a mispriced stock, mispricing can be reduced. This, in turn, can help deflate broader bubbles. When confronted with new information about specific targets, investors may start to pay closer attention to similar companies whose characteristics make them unsuitable targets for the arbs. We present suggestive evidence consistent with such informational spillovers. Specifically, we show that as a critical mass of negative reports about Chinese companies listed in the United States accumulated, other U.S.-listed Chinese stocks eventually saw steep price falls.

1. Conceptual Framework

To frame our empirical analysis, consider an informational arbitrageur who is risk averse and has limited capital. The arb's strategy is to identify overvalued stocks and trade against them. Doing so is both risky and costly. The main risk is that targets could become even more overvalued, resulting in potentially ruinous margin calls. DeLong and others (1990) label this noise trader risk. The main cost of trading against overvalued stocks is the cost of initiating and maintaining a short position.[4] The more risk averse the arb, the more limited her capital, or the higher the shorting costs are, the smaller the short position the arb can take. And the smaller the short position is, the less likely it is that the arb can exert sufficient downward pressure through shorting alone to correct the overpricing.

To identify overvalued targets, the arb invests in the production of information about the target. (We provide specific examples in Section 3.)

---

[4] Short selling involves a shorting fee, which the short seller pays the stock lender in return for borrowing the stock and the opportunity cost of the capital tied up through margin calls. Short sales tie up considerable amounts of capital. Under Regulation T, the short seller is required to deposit the proceeds of the short sale in her margin account, along with an additional 50% of the value of the short sale. In addition to this 150% initial margin requirement, the short seller faces margin calls when the price of the stock moves against him. Under Regulation T, this maintenance margin requirement amounts to 100% of the current value of the short sale plus a minimum of 25%, though many brokers set higher additional margins. The short seller is required to post additional capital when the stock price rises such that the maintenance margin exceeds the initial margin. When the stock price falls such that the maintenance margin is below the initial margin, the excess margin capital is released to the short seller.

Naturally, the arb conditions her choice of how much to short on the signal she acquires, subject to the exogenous short-sale constraints she faces: all else equal, a stronger signal will prompt the arb to short more aggressively.

Maintaining the assumption that the arb cannot induce convergence to fundamental value through her trading alone, she remains exposed to noise trader risk however strong her signal. To counteract this risk, the arb may disclose her signal to the market. Her aim in doing so is to convince the long investors that the target is indeed overvalued and thus to persuade the longs to sell. If successful, disclosing her signal accelerates the price adjustment (thanks to the longs selling) and so reduces her exposure to noise trader risk. It also sidesteps the short-sale constraints that prevent her from correcting the mispricing through shorting alone: the longs, after all, face no constraints in selling the stock.

In equilibrium, there will be circumstances in which noise trader risk is so low that the arb need not disclose her signal publicly and instead waits for the market to discover the signal independently. This is the classic arbitrage strategy: short and keep mum. Because U.S. regulations do not require short positions to be disclosed, this strategy is difficult to capture in the data.

Our empirical focus is instead on circumstances in which it is optimal for the arb to disclose her signal publicly given her information production cost and the short-sale constraints she faces. The expected information production cost may be so high relative to the expected trading gain given short-sale constraints that the arb will refrain from investigating a particular target. Information production costs thus constrain the arb's choice of target.

Credibility is key for the short-and-disclose strategy to work: the arb can only hope to induce the longs to sell if the signal she discloses is considered credible. Credibility in turn is a function of the arb's reputation (i.e., her past track record of disclosing signals that were *ex post* verified to be true) and the quality of her evidence about the current target. We thus expect the longs to respond more strongly to the arb disclosing her signal, and the share price to fall by more, the better the arb's reputation and the more compelling the signal she discloses.

We leave a richer model of arbs' strategic choices to future work and focus instead on providing empirical support for the basic building blocks of such a model outlined above. For example, it would be interesting to explore the optimal timing of the arb's disclosure. Early disclosure may be motivated by the arb's signal being short-lived (for example, if it is likely that other arbs may discover the same signal). At the same time, given short-sale constraints, early disclosure may limit the size of the arb's short position. A fuller treatment of these and other trade-offs requires a richer model than we are able to test, not least because we do not observe the delay between acquiring and disclosing the signal in our data.

2. Sample and Data

2.1 Arbitrageurs and reports

The arbitrage strategy we describe relies on publicity to induce the longs to sell and thereby generate a return on identifying overvalued companies. This makes finding the relevant arbitrageurs relatively straightforward: they are in the habit of drawing attention to their reports via the media and popular investor websites such as seekingalpha.com.

We search news sources (accessed via Factiva), as well as the Internet, for information producers who satisfy three main criteria: they target what they claim are overvalued companies listed in the United States, they disclose having a short position, and they share their information freely with the investing public in the form of written reports. The second criterion filters out bloggers who post casual comments on Internet forums and so restricts the sample to investors who make a living from information production. The third criterion filters out larger hedge funds that restrict access to their information to their own investors via password-protected websites or that "talk their book" at invitation-only investor conferences not open to the public. Our sample of targets starts in July 2006 (when DataExplorers first makes daily shorting data available). It ends in December 2011 to all allow us to track their subsequent performance through 2015.

Our search yields 31 arbs, listed in Table 1. The pioneer is Asensio & Co., which was founded in 1992 and started publishing reports on overvalued companies in 1994. The most prolific arb in our sample is Citron Research, which has been in business since 2001 and which describes itself as an "activist short seller." A firm that has come to prominence in the media is Muddy Waters, which describes itself as a "pioneer in on-the-ground, freely published investment research."

Because private firms in the United States do not generally have to make public disclosures, we are unable to provide summary statistics on the 31 arbs. Inspection of their websites suggests that they are either one-man-bands or small hedge funds. Except for Asensio & Co., Bronte Capital, Kerrisdale, and Spruce Point, none of the arbs currently is, or has ever been, registered as an investment adviser or broker-dealer with the SEC or FINRA, suggesting that the arbs generally do not manage money on behalf of external clients (at least not in the United States).

What is unique about our setting is that these small and presumably shallow-pocketed arbs not only trade against overvaluation but also share the information they acquire with investors. For each arb, we collect *every* report for *every* target, so there is no selection bias.[5] Specifically, from the arbs'

[5] To check how comprehensive our list of reports is, we use the Internet Archive (also known as the Wayback Machine), which stores historical web pages. We find no instance of an arb removing any reports.

Table 1
Summary statistics: Arbitrageurs

	Year started	Number of first reports	Number of firms covered	Total number of reports	Mean abnormal return on report date (first reports only) %	Mean CAR from report date to trading day (first 60 reports only) %	Mean cumulative abnormal profit from trading day −5 to trading day 60 (first reports only) %	Fraction of all reports coded as "more credible" %
Citron Research	2001	43	46	106	−7.2	−27.4	23.5	87
Bronte Capital	2008	9	12	33	−5.8	7.4	−5.2	0
GeoInvesting	2011	8	10	16	−12.1	−46.9	48.2	56
Ian Bezek	2009	7	9	14	0.5	−27.3	28.9	57
Shareholder Watchdog	2009	7	8	9	−6.0	−38.9	30.4	44
Alfred Little	2010	6	13	37	−17.9	−28.5	29.7	73
Muddy Waters	2010	5	6	13	−17.3	−20.1	19.2	54
Kerrisdale Capital	2009	4	8	11	−7.5	−28.9	34.9	73
Asensio & Co.	1994	4	5	34	−9.3	−42.2	42.8	76
Spruce Point	2010	4	4	6	−2.7	−40.5	42.8	50
Chimin Sang	2009	3	9	18	1.0	−8.6	21.2	61
Prescience Investment	2011	2	4	5	−14.5	13.9	−25.4	20
Absaroka Capital Management	2011	2	4	6	−5.5	−48.1	40.2	33
Chinese Company Analyst	2010	2	4	11	1.0	−10.4	9.2	18
The Forensic Factor	2011	2	2	6	−10.4	−28.6	11.6	33
Glaucus Research	2011	2	3	4	−5.0	−0.5	1.7	0
OLP Global	2010	2	2	3	−3.9	−12.6	14.3	n/a
Average (across the 17 repeat arbs)		7	9	20	−7.2	−22.8	21.6	46
14 one-time arbs		14	14	26	−9.5	−31.8	28.3	n/a

The sample contains 31 arbitrageurs who target 124 firms with 358 reports over the period from July 2006 to December 2011. Note that there are 126 first reports on 124 target companies, as two arbs release first reports on the same day in the case of two target companies. The table presents (for each of the 17 repeat arbs individually and for the 14 one-time arbs as a group), summary statistics on the number of reports and target firms, postrelease returns and profits, and the credibility of the arbs' reports. Year started is the year in which the arb first released a report on an overvalued target. (Citron Research and Asensio & Co. started before the beginning of our sample period.) For variable definitions and details of their construction, see Appendix A.

websites, we download all 401 reports the arbs have published since July 2006.[6] We ignore reports published after December 2011, so that we have sufficient postreport share prices in CRSP to identify subsequent price corrections and measure the arbs' trading profitability. We remove 31 reports on firms that are traded over-the-counter or on the Pink Sheets (for which we have no share price or other trading data) and 12 reports on firms that are listed outside the United States (for which we have no short selling data; this filter removes perhaps the most famous target firm, Sino-Forest, which was listed in Toronto). This process leaves a set of 358 reports.

Of the 31 arbs in our sample, 14 initiate coverage on a single company over our sample period. The remaining 17 "repeat" arbs publish an average of 20 reports each. As Table 1 shows, Citron Research accounts for 106 of the reports,

[6] Four of the arbs (Chimin Sang, Ian Bezek, Shareholder Watchdog, and The Forensic Factor) do not have websites of their own and disseminate their research solely via third-party websites such as SeekingAlpha. As we show in the Online Appendix, our results are qualitatively unchanged if we exclude these four arbs from the sample.

followed by Alfred Little with 37 reports, and Asensio & Co. with 34 reports. (The Online Appendix shows that all of our results are qualitatively unchanged if we exclude Citron from the sample.)

In total, the 31 arbs target 124 U.S.-listed companies over our sample period. This means there are 2.9 reports per company on average. Of the 358 reports in our sample, 126 are "first" reports in which a company is targeted by one of the arbs for the first time. (In two cases, two arbs initiated coverage of the same company on the same day.) The remaining 232 are follow-on reports, usually written by the same arb, though in 25 cases authored by one or more other arbs. Citron Research publishes the most first reports (43), followed by Bronte Capital (9). Given its longer history, Citron also covers the largest number of targets (46).

To dispel the notion that the patterns we document are driven by Chinese companies listed in the United States, which have received bad press over the last few years, we note that far from all sample companies are Chinese: 60 of the 124 companies come from China (48.4% of the sample) and the remaining 64 come from the United States (51.6%). As the Online Appendix shows, our results are qualitatively unchanged if we exclude Chinese targets from the sample.[7]

For each target firm in our sample, we know the arb who first targeted it and the exact date of each report. (We describe the target companies in Section 2.3 and the contents of the reports in Section 3.) We also extract information on material events before and after a report is released from SEC filings (such as 10-Ks, 10-Qs, and 8-Ks) and from Factiva. This gives us a complete timeline of all material events surrounding each report through October 2015.

Table 1 presents, for each of the 31 arbs, summary statistics for abnormal share-price returns on the day a company first becomes a target, cumulative abnormal returns from the report date to 60 trading days later, and an estimate of the arb's trading profit. (For all variable definitions and details of their construction, see Appendix A.) For 14 of the 17 repeat arbs, share prices fall on average when a report is released. Alfred Little has the largest immediate market impact, averaging −17.9%, followed by Muddy Waters at −17.3% and Prescience Investment at −14.5%. The average repeat arb's average report is associated with a 7.2% price fall. One-time arbs have a similar market impact, averaging a −9.5% price fall when sharing their information.

Measured over 3 months from the release date, 15 of the 17 repeat arbs see significant price falls (adjusted for the three Fama-French factors and momentum). Over this timeframe, Absaroka Capital Management has the largest price correction, averaging −48.1%, followed by GeoInvesting at −46.9% and Asensio & Co. at −42.2%. In the 3 months following a first

[7] This finding mirrors Lee, Li, and Zhang's (2014) conclusion that Chinese firms listing in the United States through reverse takeovers do not differ much from their U.S. counterparts, despite the bad press a minority of them have received.

report, prices fall by 22.8% and 31.8% on average for the average repeat and one-time arb, respectively. Only two repeat arbs see prices move against them on average over this timeframe.

As an estimate of the arbs' trading profits, Table 1 reports returns to a marked-to-market borrow-and-hold strategy that shorts the target stock 5 days before the report day and closes out the short position 3 months after the report is released.[8] The returns are net of shorting costs and risk-adjusted using the three Fama-French factors and momentum. By this measure, GeoInvesting's reports yield the highest returns, averaging 48.2% over 3 months (not annualized), followed by Asensio & Co. at 42.8% and Spruce Point at 42.8%. For the average repeat and one-time arb, this borrow-and-hold strategy yields an average return of 21.6% and 28.3%, respectively.[9]

2.2 Other data sources

Daily price and trading data for the target companies are obtained from CRSP, accounting data from Compustat, intraday share-price data from Trade and Quote (TAQ), option data from OptionMetrics, and institutional trading data from ANcerno. Equity lending data come from DataExplorers, a research company that collects lending data directly from the security lending desks of leading financial institutions. The database contains comprehensive information on the supply of shares available for borrowing, the number of shares out on loan, and loan fees for over 85% of the global equity lending market (though our subscription covers only the U.S. market).[10] Following convention, we proxy for actual short sales using shares out on loan.[11]

2.3 Target companies

Table 2 characterizes the 124 target companies by providing a snapshot of firm characteristics and shorting conditions as of 1 month before a first report is released. At that time, the average target is a midcap stock; its market capitalization of $969.3 million puts it in the 54th percentile of the distribution of CRSP firms. It has a book-to-market ratio of 0.38 (equal to the 28th percentile) and so comes from the growth part of the value-growth spectrum. Its daily share turnover averages 1.13% of shares outstanding (70th percentile). And it is fairly liquid, with an Amihud (2002) illiquidity measure of 0.06 (44th percentile).

Shorting conditions are relatively tight. One month before the first report is released, the average target company has a shorting fee of four basis points a

---

8 Results are not sensitive to when we assume the arb initiates her short position.

9 The difference in trading profits between one-time and repeat arbs largely reflects variability associated with the small sample of one-time arbs. It disappears if we winsorize away outlier returns.

10 A detailed description of these data can be found in Saffi and Sigurdsson (2011).

11 During our sample, naked short selling is generally discouraged, and short selling must be associated with stock borrowing, suggesting that the number of shares out on loan is a good indicator for actual short selling.

Table 2
Summary statistics: Target firms

	Mean	Std. dev.	Lower quartile	Median	Upper quartile	Percentile in CRSP universe
Firm characteristics						
Market capitalization ($ million)	969.3	2,152.8	162.6	332.7	792.3	0.54
Book/market ratio	0.38	0.30	0.16	0.31	0.58	0.28
Daily turnover (%)	1.13	2.28	0.26	0.50	0.96	0.70
Monthly Amihud illiquidity measure	0.06	0.25	0.00	0.01	0.03	0.44
Monthly idiosyncratic volatility (%)	3.61	3.11	1.95	2.87	4.31	0.76
Shorting conditions						
Daily shorting fee (%)	0.04	0.06	0.00	0.01	0.05	0.74
Lendable (%)	5.43	7.80	0.22	1.45	7.46	0.40
Utilization (%)	67.48	36.08	33.01	87.77	99.17	0.83
Put-call implied volatility ratio	1.15	0.26	1.02	1.07	1.19	0.69

The sample contains 124 firms targeted by 31 arbs over the period from July 2006 to December 2011. The table reports summary statistics for key firm characteristics, measured as of the most recent calendar month before the release of the first report on the target. For each characteristic, the final column reports the percentile rank of the average target firm in the CRSP universe 1 month before the report release date. For variable definitions and details of their construction, see Appendix A.

day (10.1% annualized), which is in the right tail of the CRSP distribution (74th percentile). Consistent with Beneish, Lee, and Nichols (2013), this reflects a relatively tight supply of lendable stock: on average, only 5.43% of shares outstanding are available for borrowing (40th percentile), and 67.5% of the available shares are utilized, i.e., out on loan (83rd percentile). Putting downward pressure on a target's share price via the options market would also be difficult because its put options are unusually expensive: the ratio of the implied volatility of puts to the implied volatility of calls on the target's shares—which absent short-sale constraints would equal one—averages 1.15 (69th percentile).[12] Finally, idiosyncratic volatility is high, averaging 3.61% per month (76th percentile). As Pontiff (2006) argues, high idiosyncratic volatility imposes a high holding cost on short sellers.

In sum, targets are growth companies of average size whose shares are heavily traded and quite liquid. But they are also difficult to arbitrage, as shorting fees are high, the supply of lendable stock is tight, put options are pricey, and volatility is high. And yet, as we will show, arbitrageurs manage to systematically correct mispricing in spite of these short-sale constraints.

3. Discovering Mispricing

3.1 Identifying mispriced firms

While we do not know how the arbs identify targets, there are some telltale signs. The arbs typically pick up suspicious signals from publicly observable information that the market, arguably, had simply missed or misinterpreted. As Table 3 shows, sell-side securities analysts had positive ratings on target stocks

[12] Data limitations prevent us from measuring option liquidity.

Table 3
Sell-side analysts' views of companies targeted by sample arbs

	Event window				
	(−60,−1)	(0,60)	(0,252)	(0,60) minus (−60,−1)	(0,252) minus (−60,0)
Mean recommendation score	2.01	2.04	2.10	0.03	0.09
Median recommendation score	2.01	2.01	2.08	0.01	0.07
Number of recommendations	6.88	6.91	6.78	0.03	−0.10
% buy recommendation	71.68	70.10	64.84	−1.59	−6.84**
% hold recommendation	21.96	23.00	27.93	1.04	5.97**
% sell recommendation	6.36	6.90	7.23	0.55	0.87

Table 3 reports sell-side analyst recommendations accessed through the I/B/E/S summary files in the 60 trading days before the release of a first arb report on a target firm, as well as the 60 and 252 trading days after. For variable definitions and details of their construction, see Appendix A. We use ** to denote significance at the 5% level.

before the arbs released their reports. Based on data from the I/B/E/S summary files, the consensus recommendation for the average target was 2.01 (i.e., a buy) measured over the 60 trading days before the report date, with 72% of analysts rating the average target a buy or a strong buy. There is little immediate evidence of analysts incorporating the arbs' information in their recommendations. In the 60 trading days following the report day, the consensus recommendation for the average target remains 2.01 (i.e., a buy), with virtually the same fraction of analysts (70%) maintaining a buy or strong buy recommendation. Only over a 1-year window after a report do we see any significant updating, with the fraction of analysts on a buy falling from 72% before to 65% after, and the fraction of analysts with a hold recommendation increasing from 22% before to 28% after. (There is virtually no change in the fraction with a sell recommendation.) This finding suggests that analysts are slow to update their recommendations. Such inertia is consistent with the well-known tendency among analysts to avoid annoying the top management at the companies they cover.

Similarly, institutions filing 13f reports were net buyers of target stocks before the arbs released their reports, increasing their holdings in the average target from 33.8% in quarter −2 to 34.9% in quarter −1. These findings are consistent with analysts and institutional investors not paying attention to the public signals that prompted the arbs to investigate these firms.

Several targets caught the arbs' attention because of unusual patterns of behavior (for example, constantly raising equity from shareholders while claiming to have large unused cash balances), a sequence of unexpected management changes, or implausible, too-good-to-be-true margins. The following example illustrates the latter:

> *"[The company] boasts an unjustifiable 40%+ gross margin in the domestic business and reports operating margins 46% higher than its strongest competitor, which is over 8x its size. With all of the major competitors being much larger publicly traded companies with manufacturing facilities and cost structure similar or superior,*

> *I see no validity to the Company's explanation of its high margins due to its purportedly lower cost base and greater economy of scale."*

Other telltale signs include the individuals behind the firms. Some Chinese targets, for example, used the services of a particular "promoter" to obtain a listing in the United States, and once some of the promoter's companies became involved in regulatory investigations, his other clients also came into the arbs' crosshairs. In other cases, target firms shared a little-known boutique auditor that had become the subject of an investigation by the Public Company Accounting Oversight Board for violating quality control and auditing standards.

In sum, the arbs in our sample conform well to patterns of behavior commonly assumed in the investment literature: they typically identify mispricing based on publicly observable signals such as company financials (Hirshleifer, Teoh, and Yu 2011) or accounting irregularities (Desai, Krishnamurthy, and Venkataraman 2006).

Why the market failed to notice these signals is an interesting question. One potential explanation, emphasized by Kovbasyuk and Pagano (2014), is limited attention: investors may simply have insufficient resources to process all value-relevant information.[13] Another is that some investors may deliberately ignore information that does not conform to their beliefs—a form of confirmation bias (Shefrin 1999).

3.2 Investigating their targets

Once they have identified a potential target using publicly observable signals, the arbs follow up with in-depth investigations. Investigations may start with an extensive document review, not only of SEC filings but also of harder-to-access documents such as purchase agreements, customer orders, auditor reports, or tax returns, as well as the filings of key competitors. (A favorite, in the context of Chinese companies listed in the United States, is a comparison of Chinese-language filings with local regulators and English-language filings with the SEC, which often reveals aggressive accounting that flatters the company's U.S. earnings.)

In a process that can take weeks, many investigations involve poking holes in claims the target made in public disclosures (such as its SEC filings) or conference calls. To this end, arbs may contact target firms' management to clarify doubts (while secretly recording the conversations); consult industry experts for independent opinions; arrange authorized or unauthorized site visits, accompanied by industry experts or private investigators; or put a target's

13 See Peng and Xiong (2006) for a theoretical treatment and Barber and Odean (2008) and DellaVigna and Pollet (2009) for evidence consistent with limited attention affecting asset prices.

production facilities under video surveillance. To illustrate,

> *"Our on-site due diligence confirmed our thesis that the company is nowhere near the scale that it claims to be. On our visit, we saw a very small operation that appeared to be formerly government owned, and probably privatized for very little money. [The company] claims to have six legacy paper production lines, but despite our prior agreement to see all lines, it showed us only two. […] The equipment is clearly dated …"*

Arbs may also visit a target's distributors or customers to gauge the reliability and strength of its revenue prospects, or contact the target's business partners and competitors to verify specific claims made by the target. For example, one arb conducted an extensive 10-city, 60-store channel check:

> *"[T]he investigators were instructed to count the number of small kitchen appliance brands, note the prices each brand was selling for, and ask the store/department managers and at least two different sales clerks a short list of questions about their experience selling products manufactured by the company and its competitors. For purpose of verification, the investigators were also instructed to record the name, address, phone # of the stores, as well as the name and cell phone # of the managers they spoke to."*

This particular channel check revealed that the target's sales were suspiciously slow given the firm's reported revenue growth.

The costs and risks of this information-discovery process are not trivial. Besides the difficulty of obtaining evidence to support their suspicions, the arbs often face open resistance (and occasionally hands-on obstruction) from target companies. The following example illustrates:

> *"Surveillance efforts are costly and difficult to conduct under very threatening conditions. […] Agents first must spend a few weeks watching and evaluating factory operations to determine the production cycle, factory entrances, and security surrounding the facility. The expensive cameras must be hidden so that the company does not find them, typically quite some distance from the factory and requiring use of a good zoom lens. […] Sometimes the cameras get stolen, in which case a backup camera is always on hand. Each day the local operative replaces the camera batteries (usually in the darkness of night) and memory card. […] The local operatives […] have been detained, questioned, and beaten by company security."*

Though their reports enjoy first-amendment protection as free speech, the arbs also face the risk of being sued by their targets.[14] On a more positive note, the risk of lawsuits will, to some extent, keep the arbs from making claims they cannot substantiate (and deter manipulators mimicking the arbs' strategy with the intention of shorting and distorting). This, in turn, will make it likelier that their claims will be believed in the first place.

3.3 Making their case

Once an investigation is completed, the evidence is assembled into a detailed report that is subsequently disclosed to the investing public. To attract investors' attention, reports often have catchy titles such as "*Credibility is like virginity; once you lose it, you can never get it back.*" (See Table A1 in the Appendix for further examples.)

Each report prominently discloses that the arb has a short position in the target stock. Effectively, therefore, arbs (legally) front run the publication of their reports. However, given how costly the targets are to arbitrage, there is a substantial risk that the arbs' short positions are insufficient to correct the mispricing on their own—and thus that prices will move against the arbs, resulting in potentially unlimited losses. We argue that to counteract this risk, the arbs share their information with the market in an effort to convince the long investors to sell.

To be as convincing as possible, the reports include in-depth coverage of the issues identified during the investigation, often supplemented with scanned copies of original company documents, photos (of production facilities or distribution channels), and links to videos taken during site visits or to audio recordings of conversations with target executives. In several cases, reports provide "smoking gun" evidence in the form of audio clips of employees admitting misrepresentation, video clips contradicting production claims, or irreconcilable discrepancies between foreign and U.S. filings. One particularly egregious example involves apparent evidence of fraud (the target was subsequently investigated by the SEC and delisted by NASDAQ):

> *"We recorded a telephone conversation that contains an admission that [the company] is engaging in securities fraud."*

3.4 Types of reports

All sample reports claim that targets are overpriced, but they differ in the grounds for their claims and the evidence they can marshal. Based on a careful reading of the reports, we divide the sample into those reports that convey the

---

[14] We count four lawsuits filed by targets against arbs in our sample, none successful (though one was settled). Lamont (2012) identifies other antishorting actions. In our sample, 46 targets make what Lamont calls "belligerent statements," typically denying the arbs' allegations. Two targets demand that regulators take action against the arbs, and another five announce that they "consider" taking legal action.

Table 4
Range of allegations

	Fraction
Panel A: Reports producing new information (n = 295)	
Concerns regarding financial reporting or governance	
Accounting irregularities	
Questionable performance	0.34
Misrepresentation of financials	0.22
Questionable balance sheet	0.11
Other misrepresentation	0.30
Disclosure problems	0.22
Management	0.19
Auditor quality	0.14
Internal controls	0.12
Red flag events	
Questionable business practice	0.26
Self-dealing/relatedparty transactions	0.22
Questionable acquisition	0.18
Questionable insider sales	0.06
Questionable capital raise	0.06
Outstanding legal actions	0.03
Questionable stock repurchase	0.01
Panel B: Reports reinterpreting known information (n =63)	
Concerns regarding valuation	0.18

Table 4 provides a breakdown of the 358 sample reports according to whether they convey the results of information production (i.e., the discovery of facts previously unknown to investors) or whether they result from information processing (i.e., the reinterpretation of already known data). The former type of report contains concerns regarding financial reporting, governance, or "red flag" events. Panel A presents a frequency breakdown of the main concerns based on our reading of the reports. The latter type of report, tabulated in Panel B, essentially claims a stock is overvalued based on a different interpretation of known data.

results of information production (i.e., the discovery of hard and verifiable information previously overlooked or ignored by investors) and those that result from information processing (i.e., the reinterpretation of already known data).

The vast majority of the reports (295 of 358) reveal new and hard information. Panel A of Table 4 tabulates the kinds of new information they contain. We distinguish between allegations regarding financial reporting (such as accounting irregularities or misleading disclosure) and questionable corporate governance practices (such as forgivable loans to executives) on the one hand, and concerns that arise from "red flag" events (such as suspicious acquisitions, self-dealing, undisclosed related-party transactions, and questionable insider trades) on the other. Accounting irregularities are particularly prominent.

The 63 sample reports without new information argue that a target is overvalued based on reinterpretations of known facts, claiming, for example, that a particular business model is unsustainable or expressing disagreements about industry trends or macroeconomic forecasts. These reports are thus essentially opinions, given the lack of new evidence, and so are more similar to "sell" recommendations issued by Wall Street analysts. As such, we might expect them to affect share prices a little more than Wall Street analysts tend to do when they downgrade a stock. As we will see, our estimates are consistent with this prediction.[15]

4. Correcting Mispricing

4.1 Reaction in the equity market: Prices

Table 5, Panel A, shows abnormal returns over various time windows around the release of a first report on a target company. We use three abnormal-return metrics: Fama-French/ momentum four-factor cumulative abnormal returns (CARs), characteristics-adjusted abnormal returns computed as in Daniel and others (1997) (DGTW returns), and calendar-time abnormal portfolio returns (alphas). For further details on their construction, see Appendix A.

Between trading days −20 and −6, target companies' shares perform roughly in line with our benchmarks.[16] Over the week leading up to the release, there is some evidence of price falls, perhaps as a result of an arb building (or adding to) her short positions. The median CARs and DGTW returns of −2.64% and −3.25% are reliably statistically significant, though the means are smaller and insignificant, as is the calendar-time portfolio alpha.

Once an arb releases her report to the public, investors react strongly. On the release day, prices fall by an average (median) of 7.51% (4.7%) relative to the four-factor model, with fully 93 of the 124 targets falling in price. The DGTW returns and alpha are even more negative.[17] These patterns suggest that the reports contain relevant and novel information, which investors take seriously. The reaction to a follow-on report, shown in Panel B, is smaller, with price falls averaging 3.1% relative to the four-factor model and 2.91% relative to characteristics-matched non-targets, and an alpha of −2.96%. Each of these statistics is significant at the 1% level.

Figure 1 plots the release-day price impact of each first report arranged in chronological order, along with a linear trend line. This reveals that the average price impact has increased over time, perhaps reflecting learning among investors and the establishment of credible track records by some of the arbs.

When a report is released, investors appear to respond rapidly. In five cases, we know the exact time the report was released. Figure 2 shows average continuously compounded raw returns over 5-minute intervals during the 8 trading hours before and after the release. At least for these five reports, prices fall by more than 10% within 2 hours of release.

[15] Using I/B/E/S data to identify 1,366 downgrades to strong sells in the universe of CRSP stocks (rather than in our sample), we estimate that analyst downgrades trigger price falls averaging 2.98% during our sample period.

[16] Even though CARs average a significant 3.16% (p=0.047) over this timeframe, the median CAR is negative at −2.34% and statistically insignificant. Neither the DGTW returns nor the calendar-time alphas are statistically different from zero.

[17] Removing potential outliers by trimming 5% in the tails makes little difference to any of our results; see Table IA.1 in the Online Appendix.

Table 5
Share-price changes around report releases

	Cumulative abnormal returns (CARs)			Characteristics-adjusted abnormal returns (DGTW returns)			Calendar-time abnormal portfolio returns (alphas)
	Mean	Median	Positive: negative	Mean	Median	Positive: negative	Mean
Panel A: Abnormal returns							
First reports:							
Trading days -20 to -6	3.16**	−2.34	57 : 65	1.69	−2.26	54 : 69	3.43
Trading days -5 to -1	−1.05	−2.64**	49 : 73***	−0.59	−3.25**	50 : 74***	−0.17
Trading day 0 (report date)	−7.51***	−4.70***	29 : 93***	−7.62***	−5.00***	26 : 98***	−7.64***
Trading days 0 to +60	−26.15***	−27.83***	24 : 98***	−24.44***	−27.39***	23 : 101***	−21.35***
Trading days 1 to +60	−19.83***	−23.46***	33 : 87***	−17.62***	−18.93***	31 : 91***	−13.11***
Trading days 0 to +252	−42.86***	−55.19***	16 : 106***	−47.27***	−54.24***	9 : 115***	−42.28***
Trading days 1 to +252	−38.31***	−52.15***	20 : 100***	−42.90***	−51.39***	11 : 111***	−34.37***
Follow-on reports							
Trading day 0 (report date)	−3.10***	−1.67***	78 : 140***	−2.91***	−1.62***	86 : 133***	−2.96***

	Change in market value (unadjusted)		
	Mean	Median	Positive: negative
Panel B: Changes in market value of equity			
Trading days 0 to +60	−$119.7m	−$38.5m***	32 : 92***
Trading days 0 to +252	−$133.5m	−$84.1m***	27 : 97***

Panel A reports three measures of abnormal returns (in percent): Fama-French/momentum four-factor cumulative abnormal returns (CARs), characteristics-adjusted abnormal returns computed as in Daniel and others (1997), and calendar-time abnormal portfolio returns. Each abnormal-return metric is measured over different event windows around the release of a report. Panel B reports raw changes in the market value of equity of target firms. For variable definitions and details of their construction, see Appendix A. The number of target companies for which we can compute CARs is lower than for DGTW returns because of the lack of sufficient prereport trading data for a few companies targeted shortly after their IPOs. We perform a two-sided t-test for means, a Wilcoxon test for medians, and a Wilcoxon sign test for the number of negative observations. We use ***, **, and * to denote significance at the 1%, 5%, and 10% level, respectively.

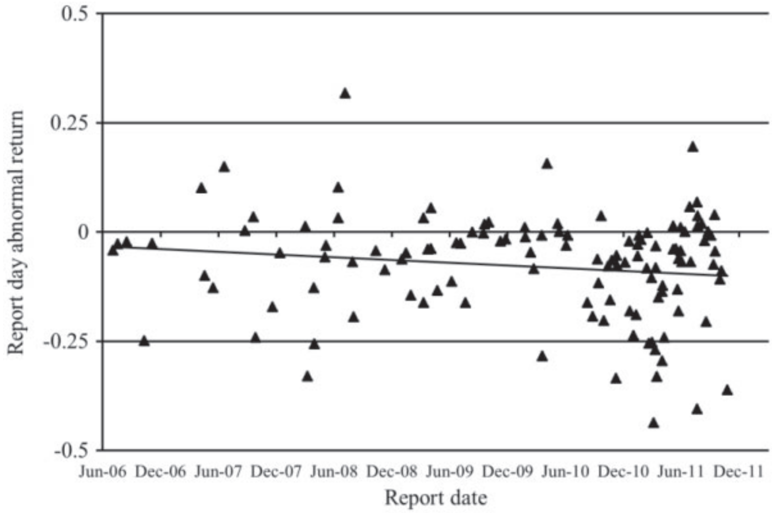


Figure 1
Report-day price effect over time

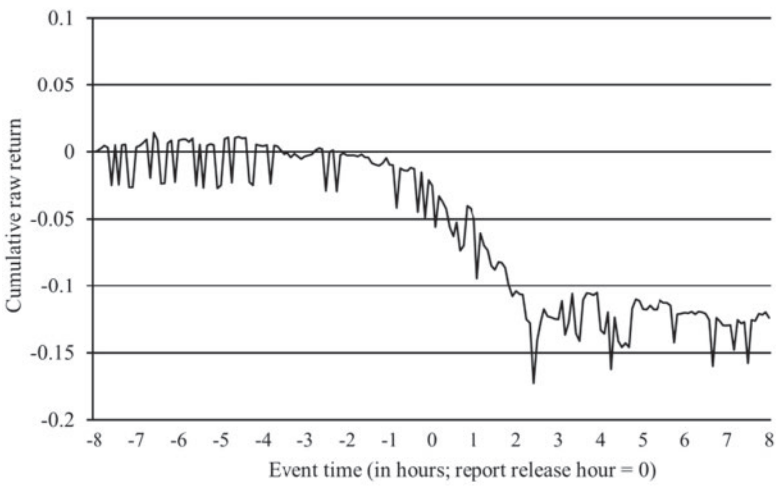


Figure 2
Price adjustment on the report day

Interestingly, the price adjustment is not complete on the release day. In the full sample, prices fall by an additional 13.1% to 19.8% on average over the next 3 months, depending on the benchmark used, leaving them between 21.4% and 26.2% below the closing price on day -1. The drift reflects the back and

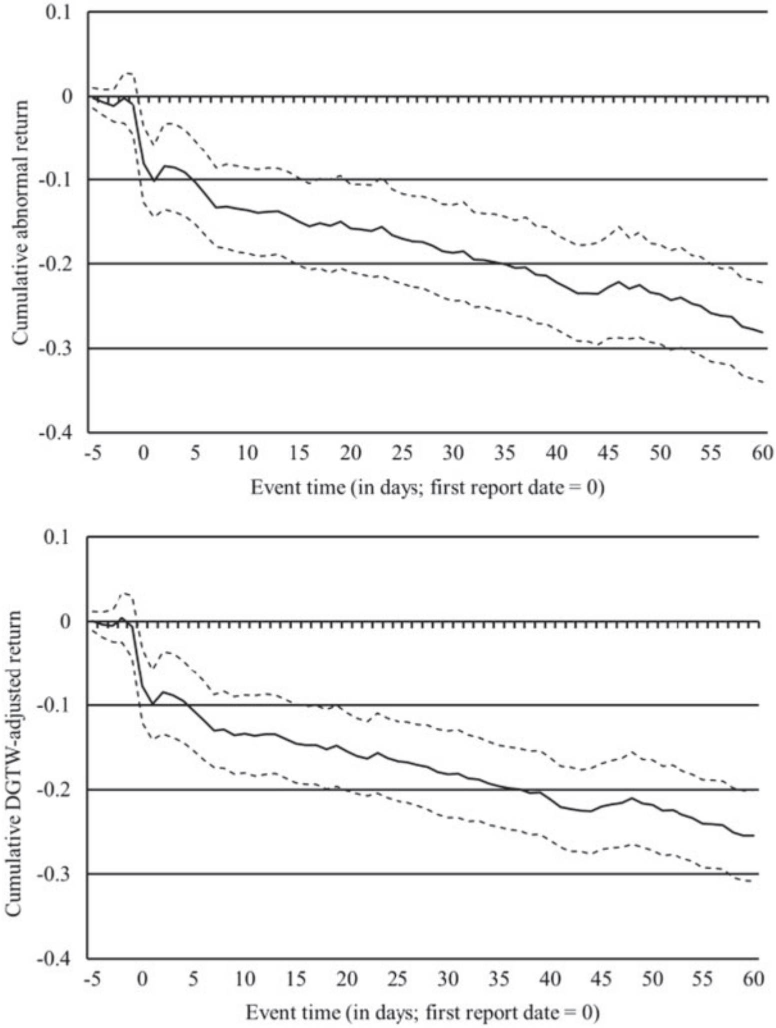


Figure 3
CARs and DGTW returns around report releases

forth between the arb and target management: typically, the target denies the arb's claims, prompting the arb to respond, while investors update their beliefs of the likelihood that the arb's claims are true. As Figure 3 shows for CARs and DGTW returns, at no time over the 3 months after becoming a target does the average firm's share price recover. These patterns suggest that the arbs' information usually proves correct.[18]

[18] Over a 1-year window, prices fall by 42.9% on average relative to the four-factor model, by 47.3% relative to characteristics-matched non-targets, and by 42.3% using the calendar-time portfolio approach. Depending

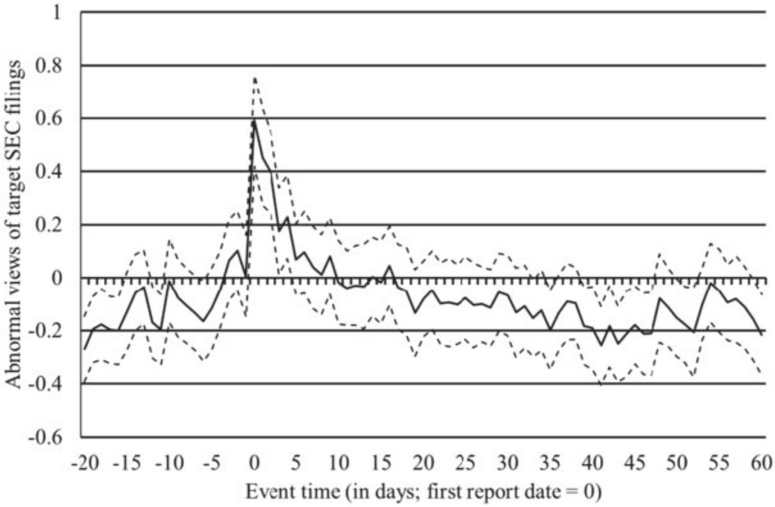


Figure 4
SEC filing views around report releases

To provide an estimate of the extent of overvaluation that the arbs help correct, we compute dollar changes in market value (not adjusting for market movements). As Table 5, Panel B, shows, the average (median) target's market value falls by $119.7 million ($38.5 million) over the 3 months starting the day before a report is released, and by $133.5 million ($84.1 million) over 1 year. In dollar terms, therefore, aggregate market values fall by $14.8 billion over 3 months and by $16.6 billion over 1 year. These numbers highlight the economic importance of small arbitrageurs to the market's informational efficiency.

4.2 Reaction in the equity market: SEC filing views

Another indication—besides falling share price—that investors pay close attention to the arbs' reports comes from SEC filing views. The SEC publishes statistics on the number of times per day that a company's regulatory filings are accessed through EDGAR. (See Drake, Roulstone, and Thornock [2005] for further details.) As Figure 4 shows, investors show abnormally high interest in a target's SEC filings when a report is released: on average, filing views increase by 80.7% on the release day, compared to the firm's baseline viewing pattern estimated in a 3-month window ending 21 trading days before the report day.[19]

on the benchmark, between 106 and 115 of the 124 targets experience negative abnormal returns over this timeframe.

19 In this and subsequent figures, we compute percentage changes relative to a baseline estimated over a 3-month period ending 1 month before a report date. Specifically, the "abnormal" value of variable X is computed as the log difference between X on day t and average daily X in a 3-month period beginning 80 trading days and ending 21 trading days before a report date (the "baseline period"). The log increase in filing views in Figure 4 shown

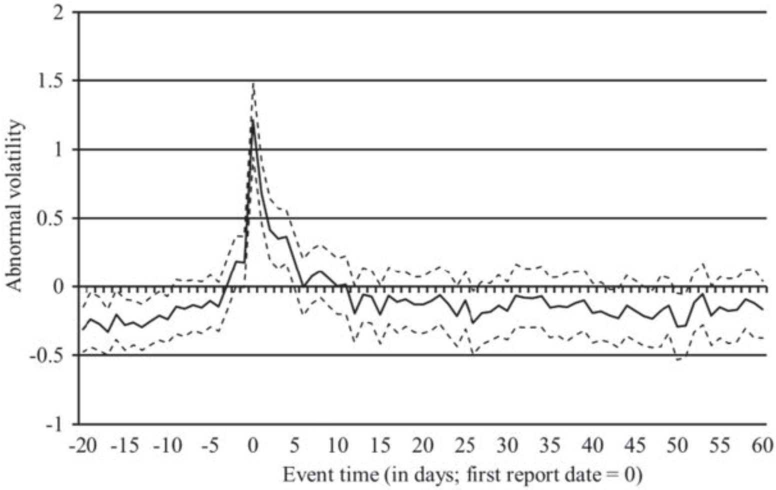


Figure 5
Average abnormal volatility

The dashed lines, which represent 95% confidence intervals, indicate that this increase is highly statistically significant. Interest remains significantly elevated on day 1 (up 57.2%), day 2 (up 48.1%), day 3 (up 19.2%), and day 4 (up 25.7%), before returning to baseline levels.

Importantly, the spike does not coincide with targets filing new disclosures with the SEC that could explain the increase in investor interest. On day 0, only 5 of the 124 targets file an event-driven disclosure (an 8-K or 6-K) in response to the arbs; 9 targets file disclosures in connection with other events; and 1 target files a periodic disclosure (a 10-K). Removing these targets leaves the spike in Figure 4 unchanged. Instead, it is the target's historic filings that investors are showing an increased interest in when the arbs release their reports.

4.3 Reaction in the equity market: Volatility, turnover, and liquidity

Just how big a shock to investors' information sets the reports represent can be seen in Figure 5, which shows that volatility spikes when a report is released: on average, volatility increases by 236% on the release day compared to the firm's baseline (i.e., by exp(1.211) – 1). Volatility remains significantly elevated for the next 5 days, suggesting that investors take up to a week to process the information revealed in the average report.

The reports also trigger a massive increase in trading activity. Figure 6 shows that share turnover (i.e., number of shares traded scaled by shares outstanding)

---

for day 0 is 0.592, meaning filing views are exp(0.592) – 1 = 80.7% higher on day 0 than during the baseline period.

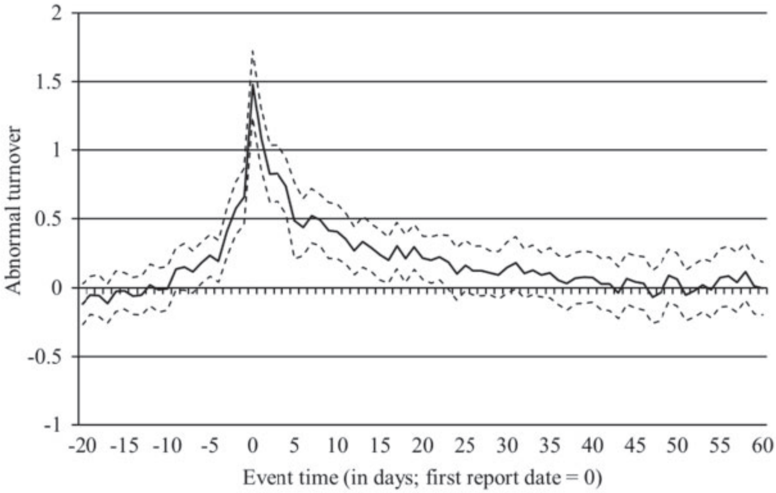


Figure 6
Average abnormal turnover

begins to rise significantly relative to the baseline approximately 4 days before the report day (perhaps because the arbs build their short positions) and then spikes at 339% above the baseline on the release day (i.e., by exp(1.480) – 1). In dollar terms, this represents a jump from an average baseline turnover of $8.9 million per day to an average of $26.8 million on the release day. Turnover stays significantly higher than normal for the next 23 trading days before returning to baseline levels.

All this extra trading affects liquidity in two ways. First, disclosure of the arbs' information reduces information asymmetry in the market and thereby alleviates adverse selection concerns. As a result, trades have lower price impact than before disclosure. Figure 7a illustrates this effect using Amihud (2002) illiquidity measure as a proxy for price impact. Relative to the prereport baseline, price impact falls dramatically on the release day and remains significantly subdued for 31 trading days.[20]

Second, we find evidence of sizable, albeit temporary, order imbalances on report days. When absorbing large trade volumes on one side of the bid-ask spread, liquidity providers' inventories become ever more unbalanced, eventually necessitating costly inventory-adjustment trades (Ho and Stoll 1983). To recover these expected inventory costs, liquidity providers quote wider spreads. A standard spread decomposition allows us to isolate the component of the spread that is due to inventory and order processing costs, called the realized spread. Figure 7b shows a significant increase in realized

20 We obtain similar results using the adverse selection component of the spread, an alternative measure of price impact.

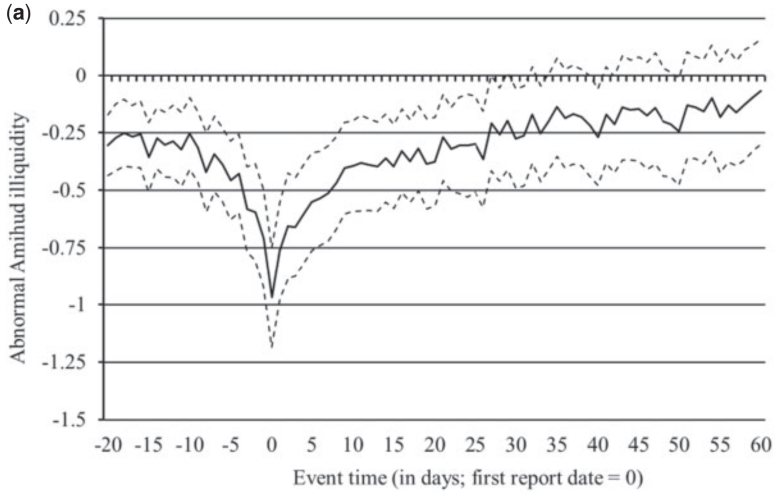


Figure 7a
Abnormal Amihud illiquidity measure

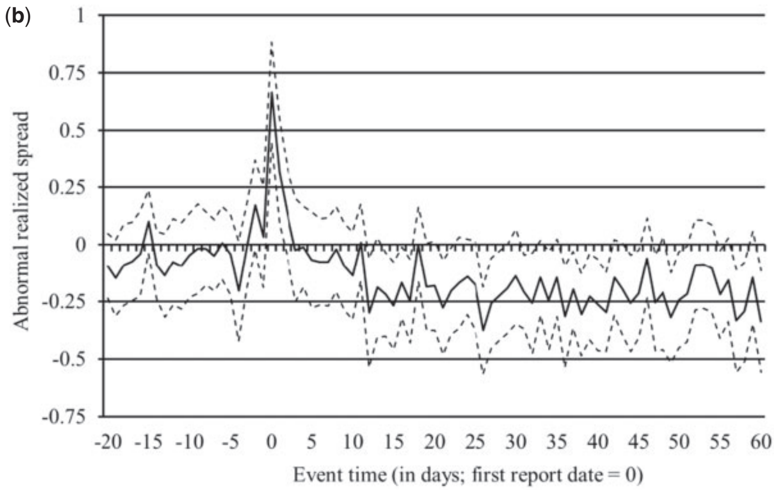


Figure 7b
Abnormal realized spread

spreads on report days, up by 94.8% on average relative to the prereport baseline. This increase in realized spreads is consistent with liquidity providers requiring greater compensation for providing liquidity at a time when, as Figure 6 suggests, liquidity demand is abnormally high. The increase is

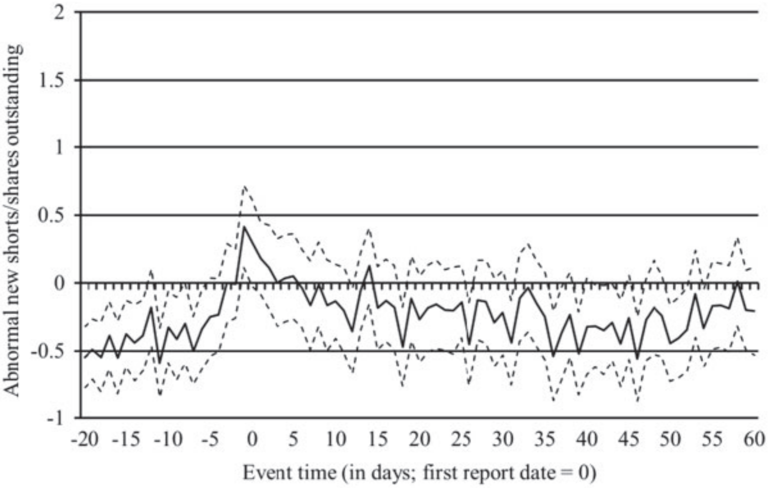


Figure 8
Average abnormal short trades

relatively short-lived, suggesting that liquidity demand spikes in response to the information release rather than being permanently higher.

4.4 Reaction in the short-sale and options markets

Who is responsible for the additional trading? Clearly, each trade involves a buyer and a seller, and only short sellers and current shareholders (the "longs") can sell the stock. Therefore, the price pressure we see on the report day comes either from increased short selling or from increased selling by the longs. We first consider the short side of the market. Figure 8 tracks the contribution of short sellers to the trading spike by plotting the average daily number of new shorts (scaled by outstanding shares) relative to the preevent baseline. One day before the release, new shorts spike at $\exp(0.417)-1=52\%$ above the prereport baseline ($p=0.008$), perhaps as the arbs build or add to their short positions.

The spike is, however, economically small: on average, only 0.45% of outstanding shares are newly shorted on day -1. Given an average day -1 market capitalization of \$951 million, this means that new shorts amount to only around \$4.28 million in trading (i.e., 0.0045*\$951m), not all of which will involve the arb in question. These small numbers underscore the fact that the arbs in our sample likely do not have particularly deep pockets. And yet, as Table 5 shows, they have a large price impact and eventually manage to help correct a substantial amount of misvaluation: shorts amounting to at most a few million dollars can help correct more than a hundred million dollars in overvaluation on average.

The spike in new shorts on day -1 is not only small but also short-lived. Beginning on the release day itself, new shorts are no more numerous than during the prereport baseline period. Thus, new shorts do not appear to be responsible for the massive increase in trading we saw in Figure 6. This is somewhat unexpected. Abreu and Brunnermeier (2002) argue that arbitrageurs face "synchronization risk," meaning that they do not know when other arbs will start targeting an overvalued firm. If a critical mass of arbs is required for a shorting strategy to be profitable, synchronization risk can lead to a coordination problem and so to insufficient arbitrage. However, by publishing their information, arbs remove the synchronization risk; the publication is essentially a coordination device. Yet we see no increase in shorting activity. Why not?

The reason is simple. The lack of unusual activity in the shorting market, once a report has been released, reflects a drastic increase in the cost of shorting and a concomitant fall in the supply of lendable stock available for shorting. Figure 9a below shows that the cost of initiating new shorts rises significantly 1 day before a report is released (perhaps in response to the arbs building their short positions) and then jumps to 50% above the baseline on the release day on average. It continues to drift higher, to a level 174% above the baseline by trading day 60 (i.e., exp(1.009) – 1). As Figure 9b shows, this puts the cost of initiating new shorts in the 79th percentile of the universe of stocks traded in the United States on the release day, drifting up to the 86th percentile over the next 3 months.

Part of the reason for the cost increase is presumably an increase in demand by short sellers, but part of it appears to be the result of a fall in the supply of lendable stock. Anecdotally, some targets put pressure on their shareholders

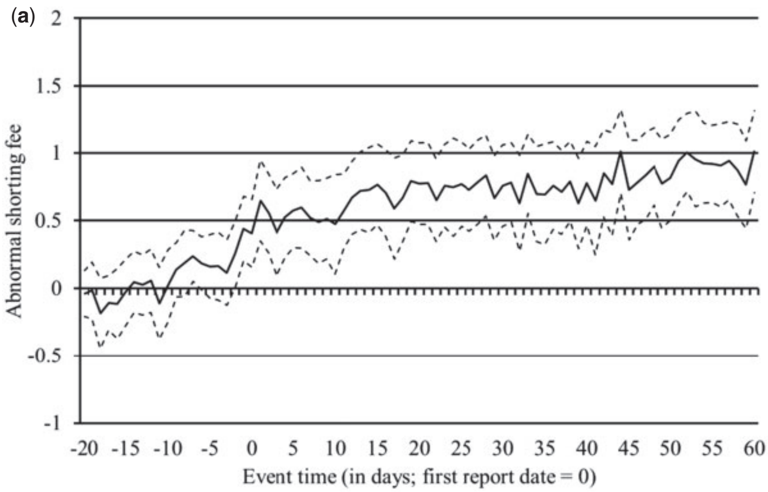


Figure 9a
Average abnormal shorting fee

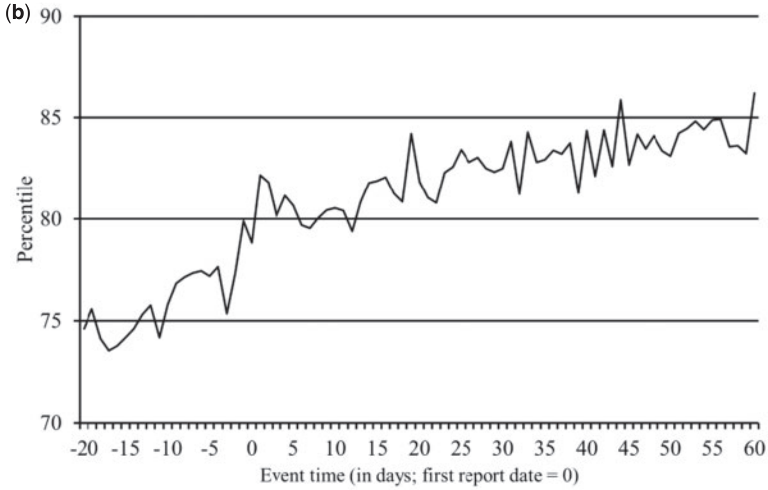


Figure 9b
Mean percentile rank

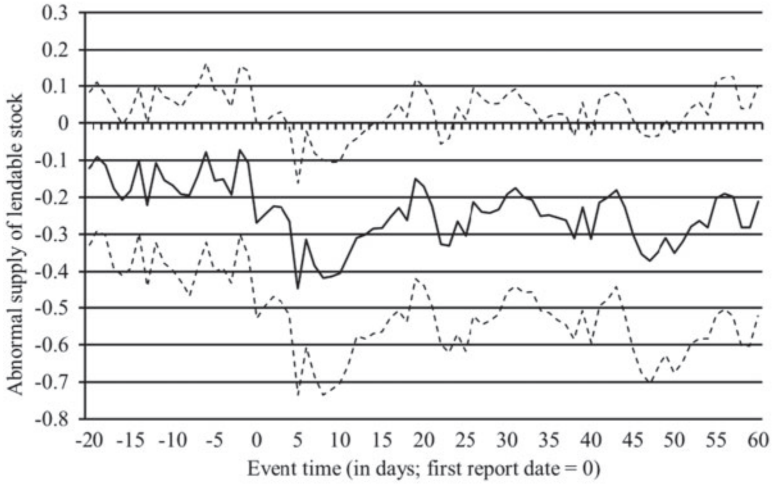


Figure 10
Average abnormal supply of lendable shares

to recall stock out on loan, to put a squeeze on short sellers.[21] Figure 10, though noisy, shows that the supply of lendable stock becomes unusually low,

[21] Twenty targets in our sample attempt to disrupt the arbs' ability to borrow shares, by asking shareholders not to lend out stock, having insiders purchase stock, or announcing share buybacks. To illustrate, a target issued the following press release in response to the release of a report: "The Company believes that short sellers' attempts to drive down the stock price and harm the Company's shareholders are likely to increase [...] In this context, the

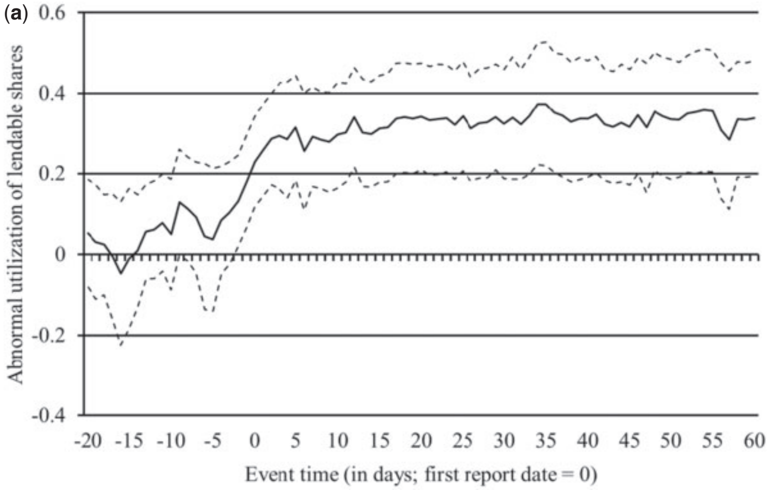


Figure 11a
Average abnormal utilization

relative to the baseline, on the report date and remains at 20% to 30% below the baseline for the next 3 months. Notably this occurs even while lending fees are exceptionally high.

While the available supply contracts, the supply utilization rate (i.e., the fraction of shares available for lending that are out on loan) jumps. Figure 11a shows utilization relative to the preevent baseline. The utilization rate increases significantly starting on day −2, consistent with the arbs building or adding to their short positions ahead of releasing their reports. On the release day, utilization is 26% above baseline; by day 60, it is 40% higher. Figure 11b shows that utilization is at the 86th percentile on the release day, climbing to the 90th percentile by day 60.

The fact that utilization increases significantly after report release (from what is already an exceptionally high level) implies that higher shorting demand cannot easily be met with the available supply. According to DataExplorers, utilization rates as high as those we see, both before and even more strongly after report release, are a strong signal of binding constraints in the equity lending market. The main reason is search costs: when the utilization rate hits 70%, it is said to become very difficult (if not impossible) to locate shares in this opaque over-the-counter market. Moreover, many would-be borrowers have margin accounts with only one or two brokers and cannot easily borrow stock from lenders outside their relationship networks.

Company believes that an important way to protect shareholder value is to limit short sellers' ability to borrow stocks and shareholders can contribute by reviewing whether their custodians or brokers are lending their shares to third parties." (PR Newswire, August 3, 2011)

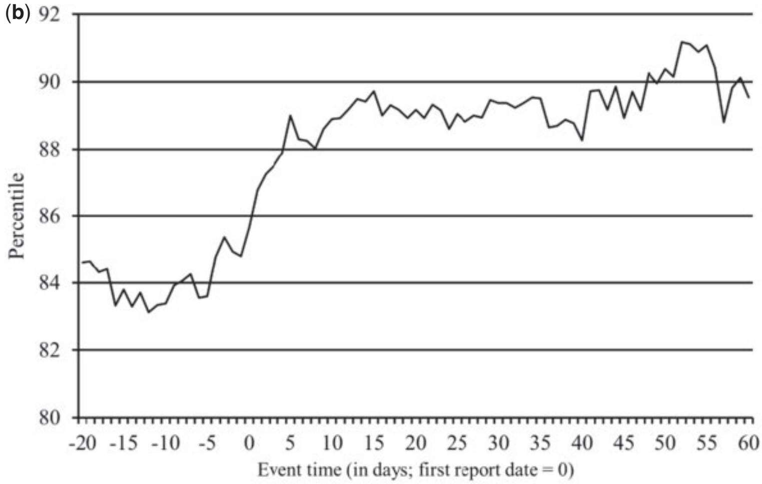


Figure 11b
Mean percentile rank

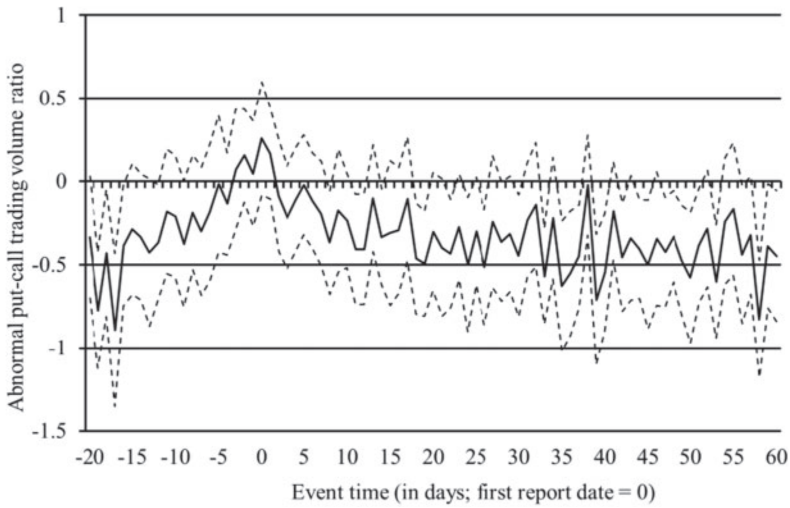


Figure 12
Average abnormal put-call option trading volume ratio

In light of this tightening in the short-sale market, investors might try to trade on the information revealed in the arbs' reports via the option markets instead. However, this too appears to be difficult. Figure 12 shows an uptick in put-option trading on the report date (relative to trading in call options with

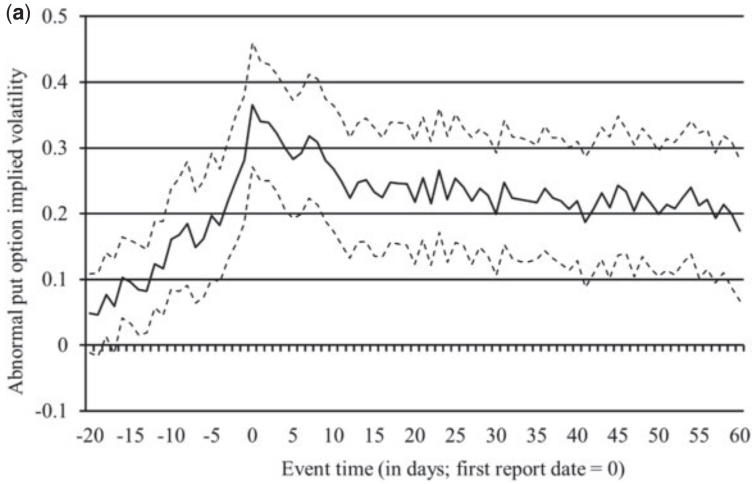


Figure 13a
Average abnormal cost of put options

the same strike price) among the 77 targets with traded options.[22] However, the uptick is not statistically significant and anyway short-lived.

Figure 13a shows why the put trading volume does not increase significantly: puts become unusually expensive. The implied volatility of puts drifts up significantly in the 3 weeks before the report date (consistent with arbs buying puts to profit from the negative information they are about to release) and spikes at 44% above baseline on the report day. Part of this spike reflects the volatility increase shown in Figure 5, but whereas volatility quickly reverts to the baseline, implied put-option volatility remains significantly elevated, at around 25% above baseline, for the next 3 months. This suggests that puts are in unusually high demand.

Another way to see that puts become expensive—beyond what is reasonable given the (temporary) increase in volatility shown in Figure 5—is to compare the implied volatilities of puts to the implied volatilities of calls with the same strike price and exercise date. By put-call parity, the implied volatilities of puts and calls must be identical and thus the ratio of put and call implied volatilities should be one—unless there are significant costs of carry, such as short-sale constraints. Figure 13b shows that puts begin to depart from parity 3 days before the report date on average, with the ratio settling at 10% above the preevent baseline once the report has been released and remaining there for the next 3 months. This is consistent with short-sale constraints becoming even tighter, making puts unusually expensive. As Figure 13c shows, once the arbs release

22 For details on how we construct this metric, see Appendix A.

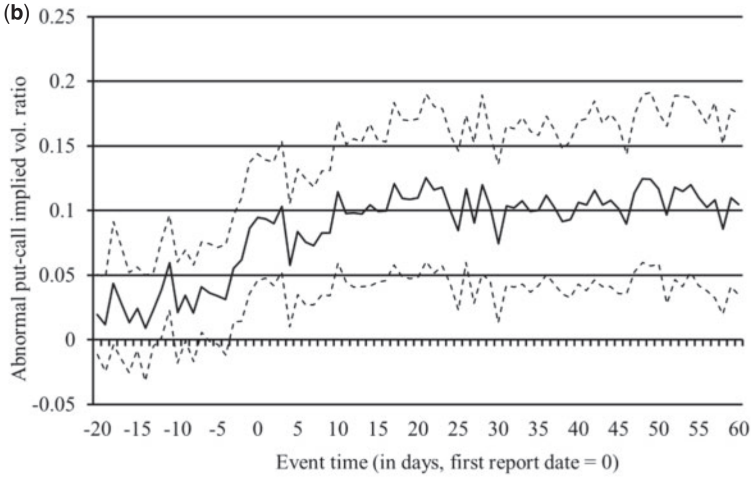


Figure 13b
Abnormal cost of puts relative to calls

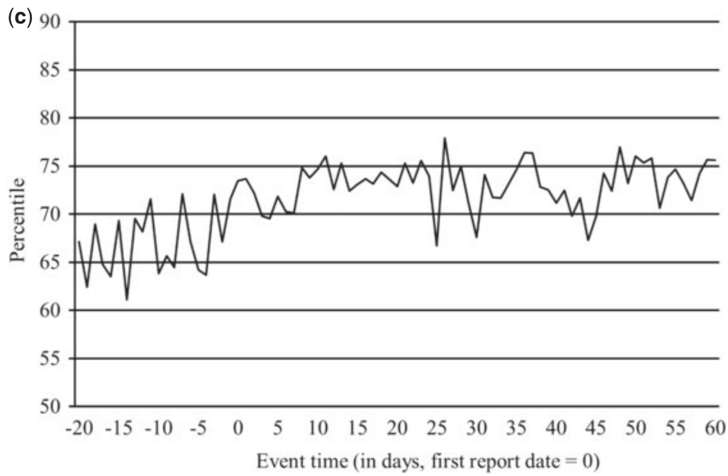


Figure 13c
Mean percentile rank

their reports, target companies' put prices move into the top quartile of the CRSP distribution.

In summary, the companies targeted by the arbs in our sample are simply too difficult to arbitrage directly—through shorting or put options—for publication of the reports to act as a coordination device in the sense of Abreu and Brunnermeier (2002). Another mechanism is therefore needed.

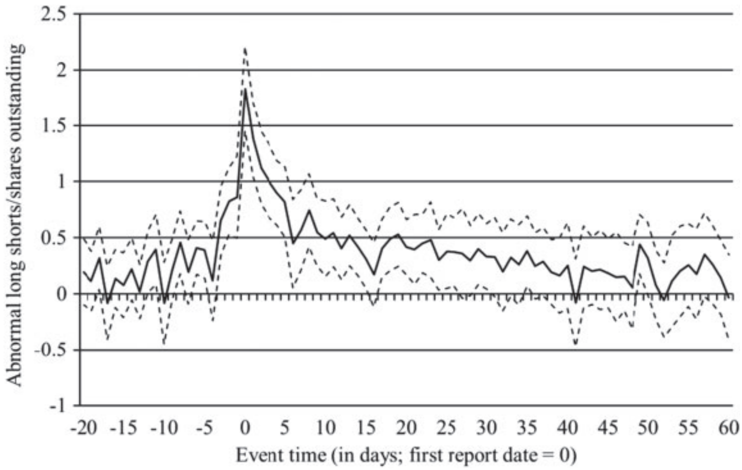


Figure 14
Average abnormal long trades

4.5 Reaction by long investors

The patterns in Figures 8 to 13 show that neither new shorts nor trading in the options market are likely to be the main cause of the report-day price falls shown in Table 5. As noted earlier, this is not surprising, given that targets exhibit many of the characteristics traditionally associated with short-sale constraints and limits to arbitrage.

What then explains the price falls? The answer, we argue, is trading by the one group of investors who are unconstrained: investors with long positions in the target companies' shares. We offer three pieces of evidence in support of this conclusion.

First, we disaggregate overall trading volume into long and short volume. *Long volume* is defined as the difference between one-way volume and new shorts. Figure 14 shows that the spike in overall trading volume we saw in Figure 6 involves a massive increase in long trading. On the report day, long trading is 524% above the baseline (= exp(1.831) – 1) and stays significantly elevated for the next month. Note that long trading necessarily involves selling by the targets' existing investors: after removing new shorts from overall volume, the only investors who can sell and thereby contribute to the volume spike in Figure 14 are those who already own the stock.

Second, we use ANcerno data, which allow us to study signed trades for a large set of pension plan sponsors and mutual funds.[23] For each trade, ANcerno

[23] Goldstein, Irvine, and Puckett (2011) and Puckett and Yan (2011) report that ANcerno institutions account for around 8% of CRSP trading volume and 10% of institutional trading volume. Puckett and Yan (2011) show that ANcerno institutions on average do not differ from 13F institutions in stock holdings, return characteristics, and stock trades, although they are larger in size.

reports the type (but not the name) of the institution, the stock ticker, the date and time the trade is executed, the number of shares traded, the execution price, and whether the trade is a buy or sell. Following Goldstein, Irvine, and Puckett (2011) and Puckett and Yan (2011), we separately aggregate sell and buy trades in ANcerno to the daily level. This gives us measures of aggregate selling and aggregate buying on a given day. The difference between the two measures gives us a third measure: daily net sales. We scale these three measures by baseline trading volume, estimated as the average daily one-way CRSP trading volume in our baseline preevent period (trading days −80 to −21). Of our 124 target stocks, 110 have trading data in ANcerno during the (−20, 60)-day event window around a report release.

Figure 15 shows aggregate abnormal sales, buys, and net sales by ANcerno institutions. There is a clear and significant spike in sales on the day the arb releases her report. (Table 6 tabulates the point estimates and reports significance levels.) For the average target, aggregate sales amount to 133% of baseline trading volume. For comparison, in the 20 days before a report release, ANcerno institutions' aggregate sales account for only 18% of trading volume on average. In other words, ANcerno institutions are more than seven times more active sellers when a report is released than normal. Net sales tell a similar story. These results confirm the central mechanism that the arbs' strategy builds on: by releasing their information to the public, the arbs can induce the long investors to sell. We see no corresponding spike in buys, implying that the buyers on the other side of the trades are predominantly retail investors or non-ANcerno institutions, including short sellers who close out their positions.

Third, for a subset of targets, we can use the ANcerno data to estimate the fraction of each long institution's prereport holding of a target company's stock that is sold on the day that a report is released.[24] This shows that the average (median) long institutional shareholder sells 43.6% (15.1%) of its holding over the course of the day on which a report is released. Fully 65% of the average target's long investors as of day −1 are net sellers on the report day.

4.6 How accurate are the reports?

The reports not only make an impression on investors; they also influence the SEC, the stock exchanges, and (perhaps inevitably) trial lawyers. Using SEC filings and Factiva searches, we find that through the end of October 2015, 69% of targeted companies are involved in class-action lawsuits; 50% are delisted by an exchange (usually out of public-interest concerns); and 36% are formally investigated by the SEC or the Department of Justice (DoJ) (occasionally for

[24] To estimate this, we restrict the sample to the 64 companies targeted before 2011. (From 2011 onward, ANcerno ceases to disclose institution-specific identifiers, so we cannot credit trades made in 2011 to specific institutions.) Of the 1,008 institutions covered in ANcerno pre-2011, 213 (21.1%) are long holders of our target stocks on the day before the first report.

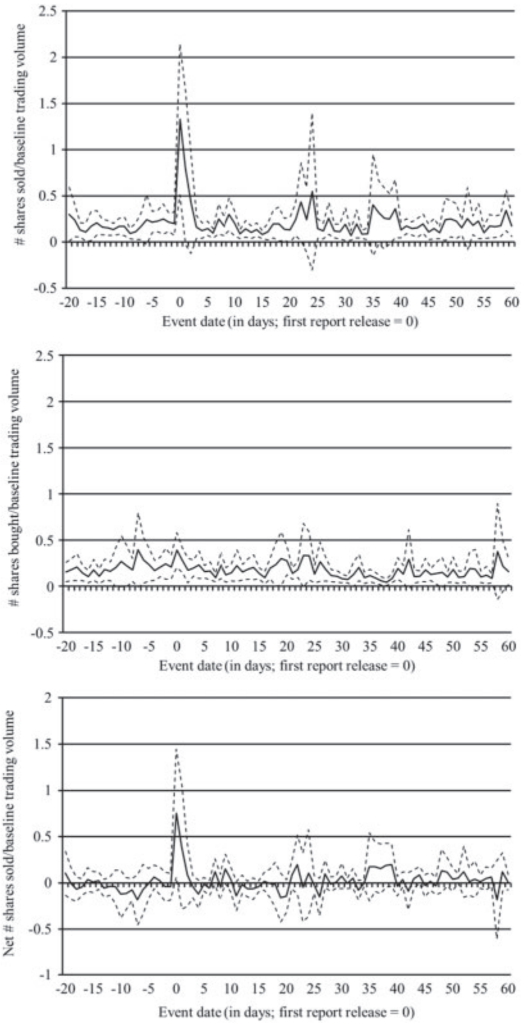


Figure 15
Average abnormal daily sales, buys, and net sales by ANcerno institutions

fraud).[25] Investigations by regulators such as the SEC or the DoJ or by an exchange back up the reports in fully 90% of the cases.

Moreover, subsequent actions taken by the targets indirectly confirm that the arbs' information is usually accurate rather than manipulative. Through

[25] The SEC and the DoJ launch fraud investigations in 21 cases and bring fraud charges in 17 of these. Fraud is not an exclusively Chinese phenomenon; 6 of the 17 fraud charges involve U.S. firms.

Table 6
Abnormal sales, buys, and net sales by ANcerno institutions

	Mean	Std. errs.	*p*-value
Institutional sells/baseline turnover on the first-report date	1.330***	0.406	0.002
Institutional buys/baseline turnover on the first-report date	0.392***	0.095	0.000
Institutional net sales/baseline turnover on the first-report date	0.753**	0.346	0.033

Table 6 reports abnormal sales, buys, and net sales by ANcerno institutions on the first-report date. Of our 124 target stocks, 110 have trading data in ANcerno during the (-20, 60)-day event window around a report release. For variable definitions and details of their construction, see Appendix A. We use *** and ** to denote significance at the 1% and 5%, level, respectively.

Table 7
Timing of subsequent events

	Calendar days since first report				
			percentile		
	Mean	Std. dev.	25th	50th	75th
Regulatory intervention	733	602	231	640	1,177
Delisting	680	594	208	533	1,034
Class-action lawsuit	540	669	265	457	845
Bankruptcy	1,409	770	609	1,415	2,169

Table 7 reports the distribution of the length of time that has elapsed between the first short-seller report on a target and the first regulatory intervention by the SEC, the Department of Justice, the Federal Trade Commission, the FDIC, the Consumer Bureau, or the FBI; a delisting action by an exchange; a class-action lawsuit being filed; or a target declaring bankruptcy.

October 2015, 92% of targets change management; 47% experience auditor turnover; and 23% restate earnings.

Table 7 shows that these subsequent events occur, for the most part, too long after our 60-day or 252-day event windows to be responsible for the price corrections seen in Table 5 and Figure 3. There are long delays between reports coming out and third-party actions. For example, it takes regulators on average 733 calendar days to intervene. This suggests that the market responses to these postreport regulatory interventions are unlikely to be responsible for the price declines we see in the immediate aftermath of the release of the arbs' reports.[26]

4.7 Do the arbs make money on their information production?

Our data allow us to estimate the arbs' trading profits, gross of the costs involved in identifying and investigating targets. Assuming that the arbs take a short position 5 days before the report day (which looks consistent with the patterns found in Figures 8 and 9) and then follow a marked-to-market borrow-and-hold strategy, their cumulative abnormal profit equals the negative of the CARs shown in Figure 3 minus the cumulative shorting fee. Figure 16 shows that this

[26] Our results are robust to excluding the five targets that experience a regulatory intervention, the four firms that are delisted, or the 35 firms subject to class-action lawsuits, in the first 60 trading days after the first report release.

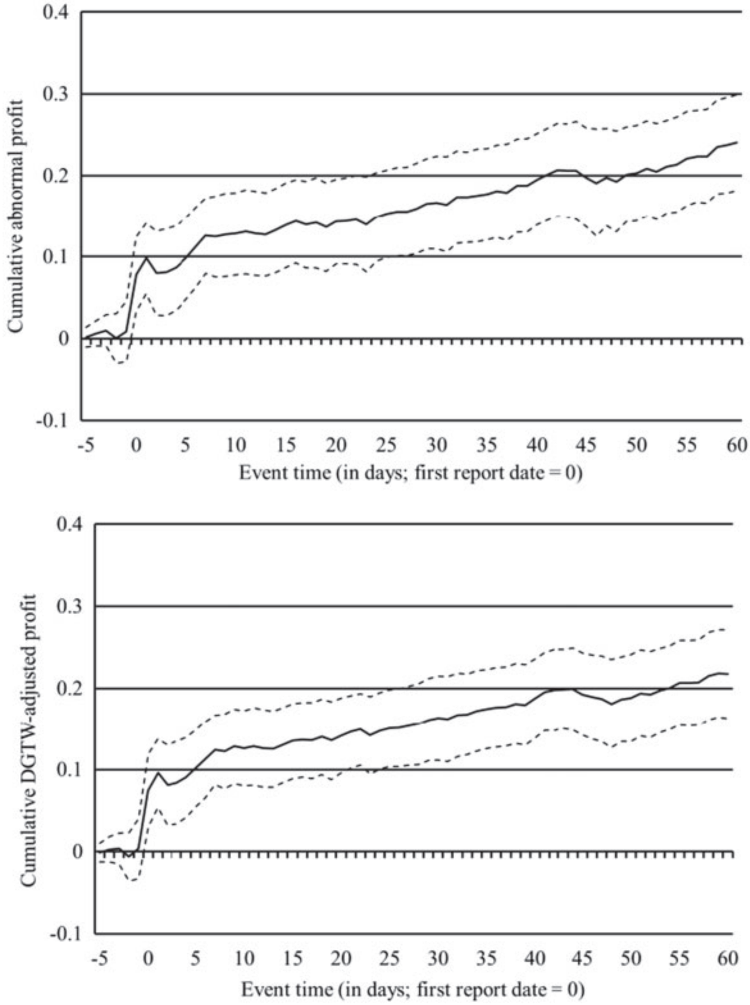


Figure 16
Average cumulative abnormal and DGTW-adjusted profit around report releases

strategy turns profitable as soon as the report is released. Relative to the four-factor model, shown on the left, the strategy makes an average return of 7.8% on the report day and a cumulative abnormal profit of 24.1% over 3 months. Relative to characteristics-matched non-targets, shown on the right, average trading profits are slightly smaller (7.5% on the report day and 21.8% over 3 months). Using a four-factor calendar-time approach (not shown in a graph) gives an average abnormal 3-month profit of 22.0%. Each of these estimates is highly statistically significant.

These profit estimates are conservative to the extent that they ignore the potential for additional—and leveraged—returns using put options.[27] They also ignore that the arbs, knowing that their reports will cause short-term spikes in volatility, could potentially set up profitable trading strategies in the options market (such as zero-beta straddles or butterflies) designed to capitalize on the volatility spike shown in Figure 5.

Whether the estimated trading profits are likely to cover the arbs' information production costs depends on three factors: the dollar size of their short position, the cost of investigating each target, and the "yield," i.e., how many companies have to be investigated to produce a viable target. Precise data on these factors are not publicly available, but with the help of interviews we have conducted with the arbs, it is possible to estimate ballpark numbers. We stress that these are crude estimates and that we cannot determine whether the trading profits simply represent "fair compensation" for the time, effort, and risks involved in this strategy.

We know that the arbs have shallow pockets and that the targets are expensive to short. This implies relatively small short positions, averaging a few million dollars.[28] (Figure 8 suggests abnormal shorting on the day before a report is released amounts to no more than \$5 million on average.) Using the results for the four-factor model to illustrate, the average 24.1% risk-adjusted return net of shorting fees seen in Figure 16 implies dollar gains of \$241,000 on a short position of \$1m, \$482,000 on a short position of \$2m, \$723,000 on a short position of \$3m, and so on—plus whatever profits the arbs can make through option-trading strategies. Anecdotally, the arbs tell us that an investigation typically costs between around \$10,000 and \$100,000. One of the early adopters of the short-and-disclose strategy claims a close to 100% yield; a later entrant claims to find three viable targets for every four he investigates. These numbers, if representative, suggest that the trading profits are large enough to cover analyst salaries, private investigators, and so on.

Another data point comes from Kerrisdale Capital, one of the arbs in our sample, whose performance data are available online.[29] Kerrisdale reports a cumulative return since inception in 2009 of 753.6% (net of fees to investors), which the firm compares to cumulative returns of 90.2% on the S&P500 Index and 31.8% on the Barclays Hedge Fund Index.

It is unlikely that the short-and-disclose strategy will continue to generate returns of this magnitude indefinitely. Eventually, entry will reduce returns by reducing the yield. As Table 1 shows, the strategy has already attracted increasing numbers of entrants. While we know of only five arbs who practiced it up to 2008, the strategy was adopted by five new arbs in 2009, seven in

[27] Anecdotally, we are told that the arbs tend to buy puts first, where possible, and then short the underlying stock.

[28] Consistent with a prediction in Kovbasyuk and Pagano (2014), we are told that the arbs in our sample tend to target one company at a time and so take concentrated short positions, resulting in underdiversified portfolios.

[29] See http://www.scribd.com/doc/156970121/Kerrisdale-Quarterly-Letter-6-30-13.

2010, and fifteen in 2011. According to the developer of a recently launched commercial database tracking "activist shorts," more than 200 U.S.-listed companies were targeted in 2012 and 2013, nearly twice as many as over our 2006–2011 sample period.

5. Pinpointing the Mechanism

The results so far suggest that the arbs do not have deep enough pockets to correct the mispricing on their own, given the short-sale constraints surrounding their targets. Our conjecture is that the arbs attempt to circumvent these constraints by making their information public, in an effort to persuade the longs to sell. This would be consistent with the observed massive increase in trading by the longs. If successful, the strategy will not only result in a price correction that translates into gains on the arbs' short positions but also reduce noise trader risk by making it less likely that prices diverge even further from fundamentals in the short-run and thereby put a squeeze on the arbs' short positions.

For the short-and-disclose strategy to work, the reports need to contain credible information. In Section 5.1, we show that only credible reports result in price corrections and profits to the arbs. Section 5.2 adds nuance to this result by showing that only reports containing new information that is costly to acquire move prices. Section 5.3 asks what kind of mispricing can be arbitraged away using the short-and-disclose strategy.

5.1 Credibility

Reports should only induce the longs to sell if the information they contain is credible. The arbs clearly understand this: many prominently post their past performance on their websites. To illustrate, Citron Research maintains a list of its targets that have subsequently been targeted by regulators: more than 50 as of January 2014 (see http://en.wikipedia.org/wiki/Andrew_Left).

Rather than relying on these posts, we construct our own measure of credibility. Specifically, to determine whether a report is likely to be considered credible, we examine each arb's prior track record, on the assumption that arbs with a stronger track record are more readily believed when they target a stock. We measure an arb's track record at time t as the rolling mean of the 3-month CARs of all her previous reports (issued at least 3 months before time t, to avoid look-ahead bias).[30] Using all 358 reports in our sample, we then code a report issued at time t as more credible if the arb's prior track record produced profits (a negative rolling mean CAR), and as less credible otherwise. This approach assumes that trading profits are a sufficient statistic for the market's assessment of the credibility of an arb's previous reports.

[30] We require each arb to have issued at least two reports before we compute a track record. Results are not sensitive to the choice of a 3-month window.

We obtain 202 reports coded as more credible and 35 reports coded as less credible.[31] Note that an arb's track record evolves over time such that she can gain or lose credibility depending on how accurate her reports prove to be. The final column of Table 1 reports what fraction of each arb's reports is coded as more credible.

Table 8, Panel A, splits the sample by this measure of credibility.[32] When a more credible arb is the first to issue a report on a target, the target's share price falls by an average of 9% relative to the four-factor model on the report day ($p < 0.001$). This is a significantly larger than the 2.2% average price fall for less credible reports, which in turn is not significantly different from zero ($p = 0.539$). If we include follow-on reports by this or other arbs, the pattern is similar: reports that are more credible are greeted with a significant price fall averaging 5.3% on the release day ($p < 0.001$), compared to a price fall of only 1.4% for less credible reports ($p = 0.315$). Turnover tells a similar story. Trading involving longs, defined as previously, responds significantly more strongly on the report day to more credible reports (up by 384%) than to less credible reports (up by 152%) ($p = 0.055$).

Consistent with these turnover patterns, prices converge faster in response to more credible reports. As Figure 17 shows, a borrow-and-hold strategy initiated 5 days before the release day of a more credible report becomes profitable immediately upon release and generates abnormal profits, relative to the four-factor model and net of shorting fees, averaging 16.5% by day +60. The rapid speed with which the price correction occurs implies a much reduced noise trader risk. Less credible reports, in contrast, do not move prices significantly, either on the release day or with any kind of lag. By day +60, for example, cumulative abnormal profits average an economically small and statistically insignificant 6.6% net of shorting fees ($p = 0.226$).

While the 10-point difference in profitability is economically large, it is not statistically significant. One reason is likely the small number of reports classified as less credible. Another becomes evident when we disaggregate the net trading profit into the market impact and the cumulative shorting fee. As Table 8 shows, reports that are more credible have significantly larger market impact, with CARs of -23.3% for more credible reports versus -10.4% for less credible reports. This suggests that investors do pay more attention to more credible reports. At the same time, we also see that the shorting fees are significantly greater for more credible reports: they average 6.8% over the assumed 3-month mark-to-market holding period (26% annualized), considerably more than the 3.9% average among less credible reports (14.9%

[31] We lose 121 of the 358 reports, in part because we require at least two reports to compute a track record; and in part, because we require that prior reports are at least 60 trading days old before we can classify the current report as more or less credible.

[32] The table reports CARs, DGTW characteristics-adjusted returns, and calendar-time portfolio alphas, but since the results are nearly identical, we discuss only the CAR estimates in the text.

Table 8
Abnormal returns and trading by type of report

Panel A:	More credible ($n=202$)	Less credible ($n=35$)	Difference in means
Abnormal return on report date (first reports only)			
Based on four-factor CARs	−9.0***	−2.2	−6.8*
Based on DGTW-adjusted returns	−9.3***	−2.3	−7.0**
Based on calendar-time abnormal portfolio returns	−9.3***	−2.1	−7.2***
Abnormal return on report date (all reports)			
Based on four-factor CARs	−5.3***	−1.4	−3.9**
Based on DGTW-adjusted returns	−5.2***	−1.5	−3.7**
Based on calendar-time abnormal portfolio returns	−5.3***	−1.6	−4.4***
3-month abnormal borrow-and-hold shorting profit			
Based on four-factor CARs	16.5***	6.6	9.9
Based on DGTW-adjusted returns	18.1***	9.8*	8.3
Based on calendar-time abnormal portfolio returns	14.5**	10.3	10.3
Disaggregating shorting profits			
Four-factor CARs (days -5,+60, not annualized)	−23.3***	−10.4*	−12.9**
Cumulative shorting fees (days -5,+60, not annualized)	6.8***	3.9***	2.9**
Abnormal turnover			
Total	1.156***	0.950***	0.205
Long side	1.578***	0.923***	0.655*

Panel B:	Evidence-based reports ($n=295$)	Opinion-based reports ($n=63$)	Difference in means
Abnormal return on report date (first reports only)			
Based on four-factor CARs	−9.0***	−2.2	−6.8***
Based on DGTW-adjusted returns	−9.7***	−2.5*	−7.2***
Based on calendar-time abnormal portfolio returns	−9.8***	−2.0	−6.4***
Abnormal return on report date (all reports)			
Based on four-factor CARs	−5.1***	−2.6**	−2.5
Based on DGTW-adjusted returns	−5.1***	−2.6**	−2.5*
Based on calendar-time abnormal portfolio returns	−5.1***	−1.7	−3.9***
3-month abnormal borrow-and-hold shorting profit			
Based on four-factor CARs	18.8***	9.5*	9.2*
Based on DGTW-adjusted returns	19.2***	10.8**	8.4*
Based on calendar-time abnormal portfolio returns	23.0***	3.9	20.1**
Abnormal turnover			
Total	1.081***	0.682***	0.399**
Tong side	1.357***	1.142***	0.215

Table 8 reports average abnormal returns, as well as shorting profits and trading statistics in subsamples sorted by the credibility of the report (Panel A), and the nature of the information discovery (Panel B). Because the subsamples in Panels A and B include follow-up reports for the same target, we also report abnormal returns on the day the first report is released. All returns, shorting profits, and fees are shown in percent. For variable definitions and details of their construction, see Appendix A. We use ***, **, and * to denote significance at the 1%, 5%, and 10% level (two-sided), respectively.

annualized). A plausible explanation is that stock lenders, who likely have market power given the tight shorting conditions shown in Table 5, charge higher fees in response to borrowing demand from the more credible arbs. This would allow stock lenders to extract part of the value these arbs create when they release their reports. Such behavior would tend to reduce the profitability gap between more and less credible reports.

5.2 Report content

Credibility appears to be necessary but is probably not sufficient to ensure the longs will listen. What likely also matters is what the arb has to say. Table 8,

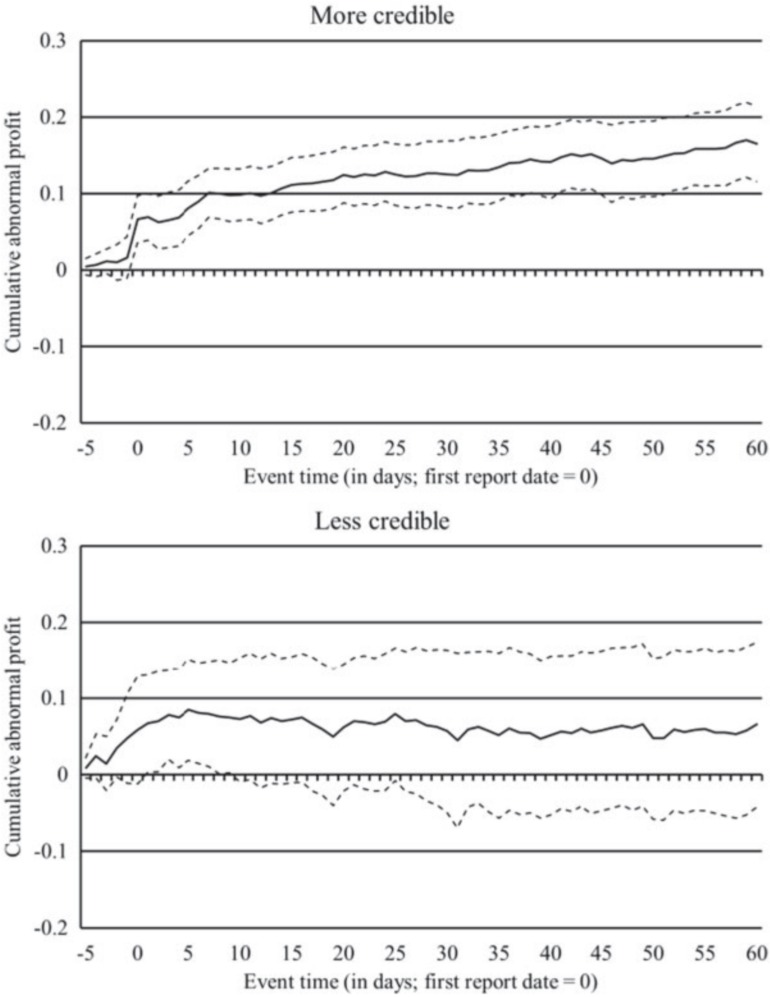


Figure 17
Cumulative abnormal profits sorted by credibility

Panel B, splits the sample of reports into those that reveal new evidence and those that merely reinterpret already known data. On first-report days, CARs average −9% for evidence-based reports ($p < 0.001$), compared to a statistically insignificant −2.2% for opinion-based reports ($p = 0.151$).[33] In other words, investors largely ignore claims of overvaluation if not backed up by new facts. The difference between the two cases is highly statistically

[33] In fact, many of the 29 reports in Table 5 that failed to result in a share-price fall upon release are opinion-based.

significant ($p=0.004$).[34] Turnover tells a similar story: both the overall increase in turnover and the reaction by the longs is stronger for reports that reveal new evidence, with the former difference being statistically significant ($p=0.024$).

These patterns imply that reports containing new facts should make higher profits than simple claims a stock is overvalued without the support of new information. Consistent with this prediction, Table 8 shows that evidence-based reports generate significant abnormal trading profits as soon as they are released while opinion-based reports have little immediate price impact and generate significantly lower abnormal trading profits, regardless of how we compute abnormal profits. Using four-factor CARs to illustrate, Figure 18 shows that average abnormal trading profits peak at 18.8% by day +60 for reports containing new evidence, twice as much as the 9.5% average abnormal trading profit for opinion-based reports. The difference is marginally statistically significant ($p=0.072$). Similar (indeed, stronger) results obtain when we compute abnormal trading profits using DGTW returns or calendar-time portfolio alphas.

5.3 What can be arbitraged using the short-and-disclose strategy?

Our results suggest that arbs who credibly reveal novel information about their targets are able to persuade long shareholders to sell and thereby contribute to a price correction, which in turn generates an economically meaningful return on the arbs' information production. What are the limits to this short-and-disclose strategy?

To have an incentive to engage in information discovery, two conditions must hold: the arbs need to be able to take a sufficiently large short position, via the cash or the options markets, to cover their expected information-discovery costs; and they need to expect to find sufficiently compelling hard evidence with which to induce the longs to sell. This implies that they are unlikely to target companies whose potential for misvaluation is too expensive to investigate or for which hard facts are unlikely to be discovered. We explore these limits to the strategy by sorting targets on various measures of arbitrage costs. The higher the arbitrage costs are, the more difficult it should be for the arbs to make profits.

Our first measure sorts targets by the average daily shorting fee during the trading month ending 1 month before the first report and splits the sample at the median. Table 9, Panel A, shows that the arbs make money regardless of how high the shorting fees are. For high-shorting-fee stocks the cumulative abnormal 3-month profits average 24.9% based on four-factor CARs ($p<0.001$), slightly more in fact than the average profit of 22.9% on low-shorting-fee stocks ($p<0.001$). (The profit difference is marginally wider when using DGTW returns or calendar-time portfolio alphas to compute abnormal profits.)

[34] Results are nearly identical using DGTW returns or calendar-time portfolio alphas.

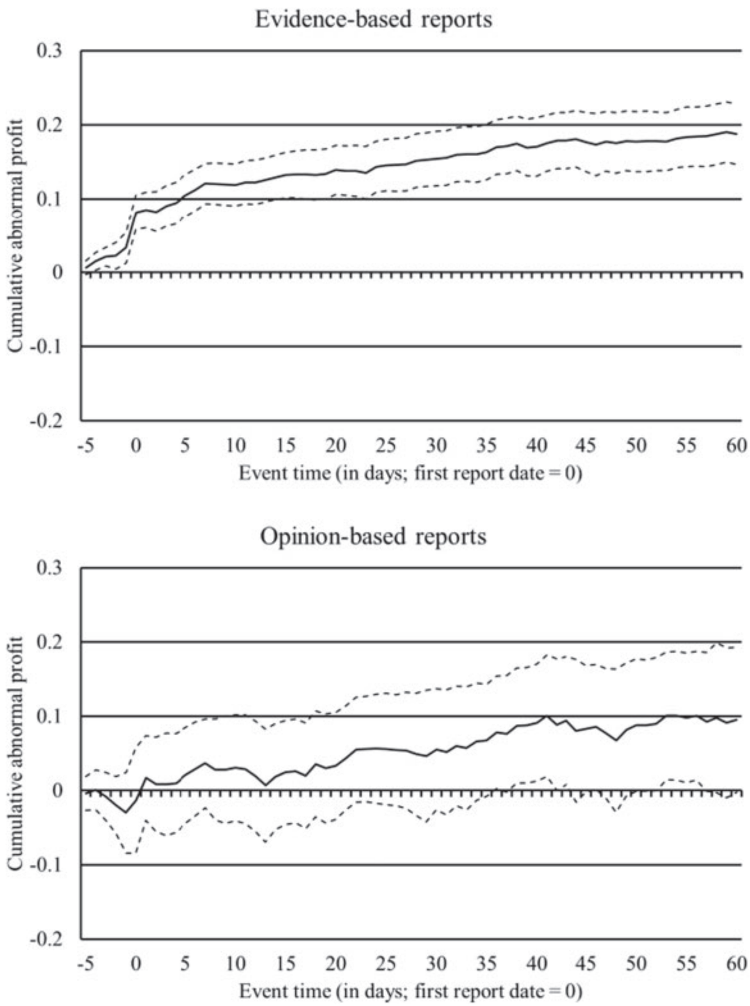


Figure 18
Cumulative abnormal profits sorted by type of report

We obtain similar results when we sort on the supply of lendable stock available for borrowing or on the utilization rate. Again splitting the sample at the median, Panel B shows that the arbs make money, net of shorting fees, regardless of how tight the supply of lendable stock is. If anything, they make greater returns in the harder-to-borrow targets (those with a low supply of lendable stock), though the differences are not statistically significant for any of our three measures of abnormal trading profits. In Panel C, the arbs make significantly greater profits in high-utilization stocks for two of the three measures.

Table 9
Abnormal returns and trading by arbitrage cost and firm size

Panel A:	High shorting fee (N=51)	Low shorting fee (N=51)	Difference in mean
Abnormal return on report date			
Based on four-factor CARs	−7.3***	−6.7***	−0.7
Based on DGTW-adjusted returns	−7.9***	−6.6***	−1.3
Based on calendar-time abnormal portfolio returns	−8.3***	−7.0***	−0.5
3-month abnormal borrow-and-hold shorting profit			
Based on four-factor CARs	24.9***	22.9***	2.0
Based on DGTW-adjusted returns	24.1***	19.9***	4.2
Based on calendar-time abnormal portfolio returns	25.9***	19.0**	3.3
Abnormal turnover			
Total	1.490***	1.380***	0.110
Long side	1.887***	1.777***	0.110

Panel B:	Low lendable (n=51)	High lendable (n=51)	Difference in means means
Abnormal return on report date			
Based on four-factor CARs	−7.8***	−6.2***	−1.6
Based on DGTW-adjusted returns	−8.4***	−6.0***	−2.4
Based on calendar-time abnormal portfolio returns	−8.3***	−5.9***	−1.2
3-month abnormal borrow-and-hold shorting profit			
Based on four-factor CARs	28.9***	18.8***	10.1*
Based on DGTW-adjusted returns	27.0***	17.0***	10.0
Based on calendar-time abnormal portfolio returns	31.5***	15.0**	9.3
Abnormal turnover			
Total	1.636***	1.235***	0.401
Long side	1.845***	1.814***	0.031

(*continued*)

Pontiff (2006) argues that idiosyncratic volatility is a key measure of arbitrage holding costs. Panel D shows that idiosyncratic volatility does not prevent the arbs from making money. In fact, their trading profits net of shorting fees increase in idiosyncratic volatility, significantly so based on four-factor CARs: cumulative abnormal profits average 34.1% after 60 days for high-volatility stocks ($p < 0.001$) versus 13.8% for low-volatility stocks ($p < 0.001$)

Panel E, finally, splits the sample by the targets' prereport market capitalization. While trading profits are larger in smaller target stocks, they are, perhaps remarkably large and statistically significant even for large targets: when targeting larger firms, the arbs make average trading profits of 20.9% relative to the four-factor model and 13.8% relative to characteristics-matched non-targets, net of shorting costs (both significant at the 0.001 level).

Table 9 further reveals that each subsample experiences similar spikes in turnover and in long trading when a report is released to the public, whether we sort on shorting fees, the supply of lendable stock, utilization rate, idiosyncratic volatility, or firm size. This finding is consistent with our proposed mechanism: if, as we propose, the price correction emanates not from the short side of the market but from long shareholders' responses to the negative information

Table 9
Continued

Panel C:	High utilization ($n=51$)	Low utilization ($n=51$)	Difference in means
Abnormal return on report date			
Based on four-factor CARs	−6.4***	−7.6***	1.2
Based on DGTW-adjusted returns	−7.5***	−7.7***	0.2
Based on calendar-time abnormal portfolio returns	−7.4***	−8.1***	0.0
3-month abnormal borrow-and-hold shorting profit			
Based on four-factor CARs	19.4***	28.1***	−8.7
Based on DGTW-adjusted returns	29.7***	13.8***	15.9***
Based on calendar-time abnormal portfolio returns	40.2***	7.9	31.3***
Abnormal turnover			
Total	1.410***	1.460***	−0.050
Long side	1.822***	1.841***	−0.020

Panel D:	High idiosyncratic volatility ($n=61$)	Low idiosyncratic volatility ($n=61$)	Difference in means
Abnormal return on report date			
Based on four-factor CARs	−9.0***	−6.2***	−2.8
Based on DGTW-adjusted returns	−9.0***	−6.3***	−2.7
Based on calendar-time abnormal portfolio returns	−8.9***	−6.1***	−1.6
3-month abnormal borrow-and-hold shorting profit			
Based on four-factor CARs	34.1***	13.8***	20.3***
Based on DGTW-adjusted returns	25.1***	17.1***	8.0
Based on calendar-time abnormal portfolio returns	26.5***	15.3**	18.6*
Abnormal turnover			
Total	1.712***	1.247***	0.466*
Long side	1.908***	1.760***	0.148

Panel E:	Small market capitalization ($n=62$)	Large market capitalization ($n=62$)	Difference in means
Abnormal return on report date			
Based on four-factor CARs	−7.5***	−7.5***	0.0
Based on DGTW-adjusted returns	−7.5***	−7.7***	0.2
Based on calendar-time abnormal portfolio returns	−7.4***	−8.1***	0.0
3-month abnormal borrow-and-hold shorting profit			
Based on four-factor CARs	27.2***	20.9***	6.4
Based on DGTW-adjusted returns	29.7***	13.8***	15.9***
Based on calendar-time abnormal portfolio returns	40.2***	7.9	31.3***
Abnormal turnover			
Total	1.523***	1.434***	0.089
Long side	1.782***	1.900***	−0.118

Table 9 reports average abnormal returns, as well as shorting profits and trading statistics, on the first-report date in subsamples sorted by various measures of arbitrage costs: shorting fees (Panel A), the supply of lendable stock available for borrowing (Panel B), the utilization rate (Panel C), idiosyncratic volatility (Panel D), and the target's market capitalization as of one month before a first-report date (Panel E). Panels A and B focus on the 102 targets for which data on shorting fees and the supply of lendable stock are available as of the baseline period (1 month before the release of the first report). In Panel D, we lose two observations with insufficient preevent trading data to compute idiosyncratic volatility. All returns and shorting profits are shown in percent. For variable definitions and details of their construction, see Appendix A. We use *** and ** to denote significance at the 1% and 5% level (two-sided), respectively.

revealed in the reports, it should make no difference how severe a set of short-sale constraints a target company happens to have: after all, the longs are not constrained.

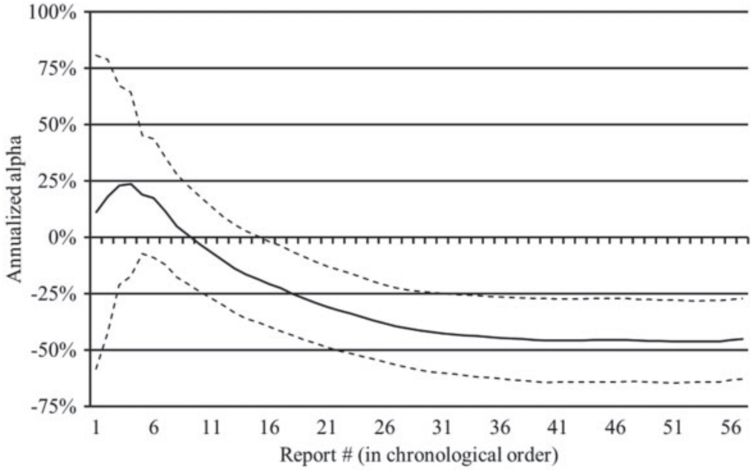


Figure 19
Calendar-time portfolio alphas going long non-target China stocks

5.4 Informational spillovers

The short-and-disclose strategy appears well suited to instances in which hard information can be discovered, such as when a company is overvalued because of aggressive accounting or fraud. This is consistent with Kovbasyuk and Pagano's (2014) model of traders who "advertise" arbitrage opportunities. In their model, advertising helps correct mispricing that is caused by limited investor attention: by focusing unconstrained investors' attention on a mispriced stock, mispricing can be reduced. We argue that advertising, inadvertently, may also help prick bubbles: on the one-rotten-apple-spoils-the-barrel principle, investors may start to pay closer attention to similar companies when confronted with negative information about specific targets.

Chinese companies listed in the United States provide an opportunity to test for such informational spillovers. In 2010, Chinese stocks were in high demand in the United States. The number of Chinese companies with a listing in the United States increased from 401 to 462 and the Bloomberg China-U.S. index rose in value by 29%. Over the same period, Chinese companies were much less popular in China (the Shanghai SSE Composite index closed down 13%), as were U.S. firms in the United States (the S&P500 index closed up by only 13%).

During 2010, and especially 2011, many U.S.-listed Chinese companies were targeted by arbs. As the Online Appendix shows, these Chinese targets suffered substantial share price falls. We now examine whether the arbs' reports may have changed sentiment about China stocks more generally. Figure 19 shows what happens to the stock prices of other Chinese firms listed in the United States (i.e., those not targeted by the short sellers) as reports are released. The graph shows 12-month alphas from calendar-time portfolios formed when the

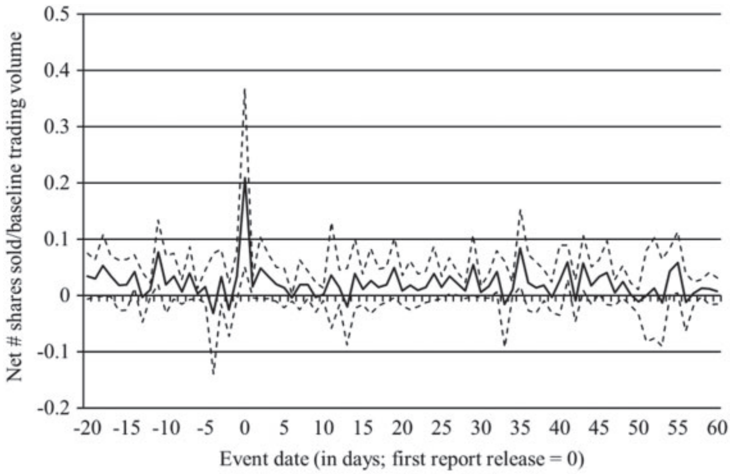


Figure 20
Average abnormal daily net sales of non-target Chinese stocks around first-report dates

first, second, third, and subsequent report on a Chinese stock came out. The portfolios only include Chinese stocks that were not themselves targets. (See Appendix A for details of the portfolio formation strategy.)

The first nine reports appear not to have influenced investors' views of China stocks in general. Starting with the 10th report, however, which came out on February 12, 2010, the prices of non-target China stocks started to be dragged down as well—and more so with every additional report that came out. This suggests that as a critical mass of negative reports about specific U.S.-listed Chinese companies accumulated, investors began to reevaluate U.S.-listed Chinese stocks in general. If so, we would expect institutional investors to sell non-target China stocks. Using ANcerno data, Figure 20 shows a clear and significant spike in net sales by institutional investors of non-target China stocks on the day an arb releases a report on a Chinese target.

6. Conclusions

Whether financial markets price securities efficiently depends on limits to arbitrage and on traders' incentives to engage in costly information production. The former limit informed traders' ability to correct mispricing *ex post* (Harrison and Kreps 1978). The latter affect whether mispricing is likely to be discovered *ex ante* (Nezafat and Wang 2014). Without limits to arbitrage, behavioral finance would not be possible (Brav, Heaton, and Rosenberg 2004).

In this study, we present evidence showing that even small arbitrageurs can help make prices efficient and that they can do so in situations characterized by what otherwise looks like formidable limits to arbitrage. To move prices in their favor despite their limited capital and the severe short-sale constraints they face, they induce target-company shareholders to trade on their behalf. The arbs do so by revealing their information. When the information is credible, the unconstrained investors (i.e., the longs) sell, driving down the price. This not only helps sidestep short-sale constraints but also accelerates the price correction and thereby reduces the risk of noise traders moving prices even further from fundamentals in the short-run, which might otherwise force the arbs to liquidate their short positions at a loss.

Credibility is key: as our evidence shows, arbs who lack (or lose) a track record of producing reliable evidence are ignored by investors and so cannot move prices by publishing their reports. Producing evidence that is new also is key: arbs who simply express the opinion that a stock is overvalued, based purely on existing information, are similarly ignored by investors.

In principle, revealing the information creates the potential for coordinated action by multiple arbitrageurs to overcome the synchronization problem modeled in Abreu and Brunnermeier (2002). But our results show that not even that is enough, on its own, to correct mispricing in our setting. The reason is that the arbs deliberately target companies with often-severe short-sale constraints, limiting the scope for coordinated action by a group of short sellers.

The strategy we describe is reminiscent of large investors such as Carl Icahn, George Soros, or Warren Buffett revealing their positions in certain companies, a practice called "talking one's book." The difference is that the arbs in our sample have limited capital, and so revealing their (necessarily small) position would, on its own, have no price impact. To be listened to, the arbs also have to reveal their information. The effectiveness of their message thus depends less on who they are or how much capital they have and more on their track record and credibility. The main barrier to entry into informational arbitrage for a small arbitrageur is thus not so much a lack of trading capital but the funding required to produce credible information.

While the primary audience is the longs, without whom prices would not adjust (or at least not sufficiently quickly), a reputation for credibly identifying overvalued targets might eventually allow some arbs to raise funds from institutional and other investors. This would help overcome the limit to arbitrage identified by Shleifer and Vishny (1997): the inability to persuade investors to commit capital to an investment strategy aimed at correcting mispricing.

An important implication of our findings is that neither limited arbitrage capital nor severe short-sale constraints need constrain informational arbitrage in practice: as long as the mispricing is sufficiently large, these obstacles can be overcome by the arbitrage mechanism we describe. It may even help prick behavioral "bubbles," by confronting overly optimistic investors with cold hard facts that are difficult to ignore.

Our evidence also illustrates why financial markets need short sellers to function well. We find no support for the widely held view that short sellers are speculators who do little more than manipulate and destabilize share prices. The fact that investors ignore reports that do not contain new, hard information suggests that it is quite difficult to manipulate share prices, at least in our setting. Instead, the short sellers in our sample are information producers who help correct mispricing and thereby help make markets more efficient.[35] This is all the more remarkable given that many targets in our sample were held by highly sophisticated investors who apparently did not spot the mispricing until it was too late.[36]

Appendix

A. Variable Definitions

A.1 Outcome Variables

A.1.1 Abnormal returns. ***CAR*** is the cumulative abnormal return over a specified event window. It is constructed using the Fama-French-Carhart benchmark. Factor loadings are estimated in a one-year preevent window ending 21 trading days before the report date. CARs for a given event window are then calculated as the difference between the realized return and the benchmark return computed using the estimated factor loadings. Standard errors are corrected for autocorrelation.

Characteristics-adjusted abnormal returns (DGTW returns) follow the methodology in Daniel and others (1997) and Wermers (2004). For each month, we sort firms into five size quintiles based on NYSE cutoffs. Within each size quintile, we then sort firms into five book-to-market quintiles based on data for the preceding December. Finally, we sort the firms in each of the 25 size/book-to-market portfolios into quintiles based on their monthly return. The resulting size, book-to-market, and momentum ranks for each stock are used in the following month to compute the DGTW-adjusted return by subtracting from a target's daily return the value-weighted return of all non-target stocks with the same size, book-to-market, and momentum ranks. Because many of our targets are Chinese firms that are listed in the United States, we relax the sampling criterion of Daniel and others (1997) and Wermers (2004) that stocks must be U.S. firms (i.e., we do not require firms to have a CRSP share code of 10 or 11).

Calendar-time abnormal portfolio returns (alpha) are computed as follows. Assuming an equal initial dollar investment in the portfolio, the portfolio return on day t is given by $\sum_{i=1}^{n_t} R_{it}x_{it}/\sum_{i=1}^{n_t} x_{it}$, where R_{it} is the day-$t$return on stock i, n_t is the number of stocks in the portfolio at day t, and x_{it} is the cumulative buy-and-hold return of stock i from the beginning of the date stock i enters the portfolio through the close of day t-1. (For a stock that enters the portfolio on day t, $x_{it}=1$.) The abnormal portfolio return (alpha) is the intercept from a weighted-least-squares regression of the daily portfolio return (in excess of the risk-free rate) on the daily Fama-French-momentum factors, using the number of stocks in the portfolio as weights (Fama 1998).

A.1.2 Abnormal profits. ***Cumulative abnormal profit*** measures the return to a borrow-and-hold strategy that goes short the stock on day t, marks the position to market on a daily basis, and closes out the position on day T. It is measured as the negative of the cumulative abnormal return (CAR)

[35] Consistent with this view, Carpenter, Lu, and Whitelaw (2014) find that stock prices became more informative about future earnings after China introduced short selling in 2006.

[36] According to media reports, investors that lost substantial amounts when arbs revealed their information include Paulson & Co. (with a $468 m loss on Sino-Forest) and C.V. Starr & Co. (with a $6.5 m loss on ChinaMedia Express). Other prominent investors that have suffered from the price fall upon the information release include Blackrock, Vanguard, Hartford Investment Management, Apollo Global Management, and Henderson Global Investors.

net of the cumulative daily shorting fee (the daily cost of borrowing new shares from equity lenders according to DataExplorers).

Characteristics-adjusted abnormal profit (DGTW profit) measures the return to a borrow-and-hold strategy that goes short the stock on day t, marks the position to market on a daily basis, and closes out the position on day T. It is measured as the negative of the cumulative *DGTW return* net of the cumulative daily shorting fee (the daily cost of borrowing new shares from equity lenders according to DataExplorers).

Calendar-time abnormal profit measures the return to a borrow-and-hold strategy that goes short the stock on day t, marks the position to market on a daily basis, and closes out the position on day T. It is computed as follows. Each trading day t, we measure the profit on target stock i as the negative of i's return net of the daily shorting cost. Assuming an equal initial dollar investment in the portfolio, the portfolio profit on day t is $\sum_{i=1}^{n_t} \pi_{it} x_{it} / \sum_{i=1}^{n_t} x_{it}$, where π_{it} is the day-t profit on stock i, n_t is the number of stocks in the portfolio at day t, and x_{it} is the cumulative buy-and-hold return of stock i from the beginning of the date stock i enters the portfolio through the close of day t-1. (For a stock that enters the portfolio on day t, $x_{it} = 1$.) The abnormal portfolio profit is the intercept from a weighted-least-squares regression of the daily portfolio profit (in excess of the risk-free rate) on the daily Fama-French-momentum factors, using the number of stocks in the portfolio as weights (Fama 1998).

A.1.3 Daily trading and shorting variables. ***Turnover*** is defined as the one-way number of shares traded in a day (CRSP variable *vol* divided by 2) divided by the number of shares outstanding (CRSP variable *shrout*).

New shorts/shares outstanding is the number of new shorts initiated (as proxied by the number of new shares on loan) on a given day divided by the number of shares outstanding (CRSP variable *shrout*). The number of daily new shares on loan is obtained from DataExplorers.

Long trades/shares outstanding is equal to the difference between one-way *turnover* and *new shorts/shares outstanding* (as defined above). In other words, it is the number of (one-way) traded shares on a given day that are not attributable to short sellers, divided by the total number of shares outstanding (CRSP variable *shrout*).

Shorting fee is the daily cost of initiating new short positions (i.e., the daily cost of borrowing new shares from equity lenders), using data obtained from DataExplorers.

Supply of lendable stock is defined as the number of shares available for borrowing on a given day divided by the number of shares outstanding (CRSP variable *shrout*). Data on shares available for borrowing are obtained from DataExplorers.

Utilization is defined as the number of shares out on loan divided by the number of shares available for borrowing on a given day. Data on shares available for borrowing are obtained from DataExplorers.

Institutional sells/baseline turnover measures the selling activity by ANcerno institutions at the daily level. Specifically, we follow Goldstein, Irvine, and Puckett (2011) and Puckett and Yan (2011) and aggregate sales in ANcerno to the daily level. We scale aggregate sales by the baseline trading volume, estimated as the average daily one-way CRSP trading volume in our baseline preevent period (trading days −80 to −21).

Institutional buys/baseline turnover measures the buying activity by ANcerno institutions at the daily level. Specifically, we follow Goldstein, Irvine, and Puckett (2011) and Puckett and Yan (2011) and aggregate purchases in ANcerno to the daily level. We scale aggregate purchases by the baseline trading volume, estimated as the average daily one-way CRSP trading volume in our baseline preevent period (trading days −80 to −21).

Institutional net sales/baseline turnover is defined as the difference between *Institutional sells/baseline turnover* and *Institutional buys/baseline turnover.*

A.1.4 Daily variables constructed using intraday data from TAQ. ***Amihud (2002) illiquidity measure*** is defined as the average, over a day, of the absolute value of the 5-minute continuously compounded return multiplied by 1,000,000 which then is divided by the dollar trading volume within the same 5-minute interval. We compute the continuously compounded returns using TAQ prices sampled every 5 minutes. Results using quote prices are very similar. The variable is winsorized at the 1% and 99% levels to reduce the effect of outliers.

Realized spreads is a measure of the compensation liquidity providers require to cover their inventory and order processing costs. It is obtained from a spread decomposition. This requires estimates of the value of the stock before and after a given trade. We estimate these following Huang and Stoll (1996), who proxy for the pretrade value using the midpoint of the most recent bid-ask spread before the trade and for the posttrade value using the transaction price 5 minutes after the trade. The realized spread then is the difference between the current transaction price and the posttrade value.

A.1.5 Daily options-related variables (from OptionMetrics). ***Put-call trading volume ratio*** is constructed as follows. First, for each pair of traded put and call options with the same strike price (OptionMetrics variable *strike_price*) and exercise date (OptionMetrics variable *exdate*), we compute the put-call ratio of the daily trading volume (OptionMetrics variable *volume*). Second, we compute the weighted average of the daily trading volume ratios for all the put and call option pairs, weighted by open interest (OptionMetrics variable *open_interest*) on the corresponding put-call option pair for a given stock.

Put option–implied volatility is the weighted average of the implied volatility of all traded put options on a day for a given stock, weighted by open interest (OptionMetrics variable *open_interest*) on each of the traded put options for that stock. The implied volatility measure for each traded put option is obtained directly from OptionMetrics (variable *impl_volatility*).

Put-call implied volatility ratio is constructed as follows. First, for each pair of traded put and call options with the same strike price (OptionMetrics variable *strike_price*) and exercise date (OptionMetrics variable *exdate*), we compute the put-call ratio of the daily implied volatilities (OptionMetrics variable *impl_volatility*). Second, we compute the weighted average of the daily implied volatility ratios for all the put and call option pairs, weighted by open interest (OptionMetrics variable *open_interest*) on the corresponding put-call option pair for a given stock.

A.1.6 Analyst recommendation variables (from I/B/E/S summary files). ***Mean (median) recommendation score*** is defined as the average (median) of the recommendation scores on a target stock by sell-side analysts. A score of 1 equals a strong buy and a score of 5 equals a strong sell.

Number of recommendations is defined as the number of sell-side analyst recommendations on a target stock.

% buy recommendation is defined as the percentage of sell-side analyst recommendations that have a buy rating.

Table A1
Illustrations of report titles

1.	"*Credibility is like virginity; once you lose it, you can never get it back,*" January 24, 2008 (Citron Research)
2.	"*Arthrocare—Something is rotten in the state of Denmark,*" May 2, 2008 (Citron Research)
3.	"*Emcore…Nothing plus nothing = nothing,*" September 9, 2008 (Citron Research)
4.	"*Citron exposes Apollo's big dirty secret – All new docs,*" March 2, 2009 (Citron Research)
5.	"AOB deal questionable even without undisclosed relationship between chairman and seller," August 5, 2009 (Asensio & Co.)
6.	"*SinoCoking: Follow the money!*" March 11, 2010 (Citron Research)
7.	"*Orient Paper's top supplier: An empty shell owned by ONP's CEO*," August 10, 2010 (Chinese Company Analyst)
8.	"*China New Borun (BORN)—You are cold busted: Now it is time to come clean*," November 15, 2010 (Citron Research)
9.	"*A stock only a trading robot could love*," December 28, 2010 (Citron Research)
10.	"*Another stock only a computer could love: The sequel*," January 24, 2011 (Citron Research)
11.	"*CCME: The China reverse merger stock that is 'too good to be true,'*" January 30, 2011 (Citron Research)
12.	"*ChinaCast Education Corporation: Show me the money! Questions to management regarding acquisition #1*," February 16, 2011 (OLP Global)
13.	"*Irrefutable evidence of fraud*," March 2, 2011 (Muddy Waters)
14.	"*China Biotics: The best research you haven't seen*," March 12, 2011 (Citron Research)
15.	"*DEER: Was the $22.3 million land use right certificate a forgery?*" March 18, 2011 (Alfred Little)
16.	"*Sino Clean Energy: Who lied about the weather*?" May 12, 2011(Chimin Sang)
17.	"*Gulf Resources: Financial claims are beyond reason*" May 19, 2011 (Kerrisdale Capital)
18.	"*ZAGG: Is it in the covers business, or covering up its real business?*" July 13, 2011 (Citron Research)
19.	"*Sino Clean Energy is a complete hoax and its shares are worthless*," April 28, 2011 (Alfred Little)
20.	"*The Harbin Land shuffle: A classic bait and switch*," September 23, 2011 (GeoInvesting)

Table 1 lists 20 examples of report titles used by sample arbitrageurs to attract attention to their reports.

% hold recommendation is defined as the percentage of sell-side analyst recommendations that have a hold rating.

% sell recommendation is defined as the percentage of sell-side analyst recommendations that have a sell rating.

A.2 Firm Characteristics (Measured as of the Most Recent Calendar Month-End before a First Report)

Market capitalization is defined as the product, reported in millions, of the end-of-month share price (CRSP variable *prc*) and the total number of shares outstanding (CRSP variable *shrout*).

Book/market ratio is measured as the ratio of a firm's book value of equity (Compustat items *seq* + *txditc* – *pstkrv*) to its market value (Compustat item *prcc* multiplied by Compustat item *csho*).

Monthly idiosyncratic volatility is defined as the monthly average of the standard deviation of residuals from adjusted daily Fama-French regressions specified as in Jiang, Xu, and Yao (2009).

Monthly Amihud (2002) illiquidity measure is constructed as follows. We use daily CRSP data (CRSP items *ret*, *prc*, and *vol*) to calculate the ratio of absolute stock return (multiplied by 1,000,000) to dollar trading volume for each day, after which we average these daily ratios over a month.

A.3 Short-Seller Characteristics

Credible is defined as follows. We examine the arb's prior record of accomplishment, on the assumption that arbs with a stronger record of accomplishment are more readily believed when they target a stock. We measure an arb's record of accomplishment at time *t* as the rolling mean of the 3-month cumulative abnormal returns (CARs) of all her previous reports (issued at least 3 months before time t, to avoid look-ahead bias). We require each arb to have issued at least two reports before we compute a track record. A report issued at time t is coded as more credible if the arb's prior track record produced profits (a negative rolling mean CAR) and as less credible otherwise.

A.4 Portfolio Formation Strategy (Figure 19)

For each first report targeting a Chinese firm and released on day t, we estimate the abnormal return to a trading strategy that buys all non-target U.S.-listed Chinese firms when reports $1, 2, \ldots t$ are released and sells the Fama-French and momentum portfolios. Non-targets are firms that have not themselves been targeted by sample arbs through report t. Abnormal returns are obtained from monthly calendar-time portfolio regressions assuming a 12-month holding period.

References

Abreu, D., and M. Brunnermeier. 2002. Synchronization risk and delayed arbitrage. *Journal of Financial Economics* 66:341–60.

Amihud, Y. 2002. Illiquidity and stock returns: Cross-section and time series effects. *Journal of Financial Markets* 5:31–56.

Barber, B., and T. Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21:785–818.

Beneish, M., C. Lee, and C. Nichols. 2013. In short supply: Equity overvaluation and short selling. Working Paper, Stanford University.

Bondarenko, O. 2003. Statistical arbitrage and securities prices. *Review of Financial Studies* 16:875–919.

Brav, A., J. Heaton, and A. Rosenberg. 2004. The rational-behavioral debate in financial economics. *Journal of Economic Methodology* 11:393–409.

Carpenter, J., F. Lu, and R. Whitelaw. 2014. The real value of China's stock market. Working Paper, New York University.

Christophe, S., M. Ferri, and J. Angel. 2004. Short-selling prior to earnings announcement. *Journal of Finance* 59:1845–75.

Christophe, S., M. Ferri, and J. Hsieh. 2010. Informed trading before analyst downgrades: Evidence from short sellers. *Journal of Financial Economics* 95:85–106.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52:1035–58.

Dechow, P., A. Hutton, L. Meulbroek, and R. Sloan. 2001. Short sellers, fundamental analysis and stock returns. *Journal of Financial Economics* 61:77–106.

DeLong, B., A.Shleifer, L.Summers, and R. Waldmann. 1990. Noise trader risk in financial markets. *Journal of Political Economy* 98:703–38.

DellaVigna, S., and J. Pollet. 2009. Investor inattention and Friday earnings announcements. *Journal of Finance* 64:709–49.

Desai, H., K. Ramesh, S. Thiagarajan, and B. Balachandran. 2002. An investigation of the information role of short interest in the NASDAQ market. *Journal of Finance* 52:2263–87.

Desai, H., S. Krishnamurthy, and K. Venkataraman. 2006. Do short sellers target firms with poor earnings quality? Evidence from earnings restatements. *Review of Accounting Studies* 11:71–90.

Diether, K., K.-H. Lee, and I. Werner. 2009. Short-sale strategies and return predictability. *Review of Financial Studies* 22:575–607.

Drake, M., D. Roulstone, and J. Thornock. 2015. The determinants and consequences of information acquisition via EDGAR. *Contemporary Accounting Research* 32:1128–61.

Engelberg, J., A. Reed, and M. Ringgenberg. 2012. How are shorts informed? Short sellers, news, and information processing. *Journal of Financial Economics* 105:260–78.

Fama, E. 1998. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics* 49:283–306.

Figlewski, S. 1981. The informational effects of restrictions on short sales: Some empirical evidence. *Journal of Financial and Quantitative Analysis* 16:463–76.

Goldstein, M., P. Irvine, and A. Puckett. 2011. Purchasing IPOs with commissions. *Journal of Financial and Quantitative Analysis* 46:1193–225.

Harrison, M., and D. Kreps. 1978. Speculative investor behavior in a stock market with heterogeneous expectations. *Quarterly Journal of Economics* 92:323–336.

Hirshleifer, D., S. Teoh, and J. Yu. 2011. Short arbitrage, return asymmetry, and the accrual anomaly. *Review of Financial Studies* 24:2429–61.

Ho, T., and H. Stoll. 1983. On dealer markets under competition. *Journal of Finance* 35:259–67.

Huang, R., and H. Stoll. 1996. Dealer versus auction markets: A paired comparison of execution costs on NASD and the NYSE. *Journal of Financial Economics* 41:313–57.

Jiang, G., D. Xu, and T. Yao. 2009. The information content of idiosyncratic volatility. *Journal of Financial and Quantitative Analysis* 44:1–28.

Karpoff, J., and X. Lou. 2010. Short sellers and financial misconduct. *Journal of Finance* 65:1879–913.

Kovbasyuk, S., and M. Pagano. 2014. Advertising arbitrage. Working Paper, University of Naples.

Lamont, O. 2012. Go down fighting: Short sellers vs. firms. *Review of Asset Pricing Studies* 2:1–30.

Lamont, O., and R. Thaler. 2003. Can the market add and subtract? Mispricing in tech stock carve-outs. *Journal of Political Economy* 111:227–68.

Lee, C., K. Li, and R. Zhang. 2014. Shell games: Are Chinese reverse merger firms inherently toxic? *Accounting Review* 90:1547–89.

Nagel, S. 2005. Short sales, institutional investors, and the cross section of stock returns. *Journal of Financial Economics* 78:277–309.

Nezafat, M., and Q. Wang. 2014. Short sale constraints, information acquisition, and asset prices. Working paper, Georgia Institute of Technology.

Peng, L., and W. Xiong. 2006. Investor attention, overconfidence, and category learning. *Journal of Financial Economics* 80:563–602.

Pontiff, J. 2006. Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics* 42:35–52.

Puckett, A., and X. Yan. 2011. The interim trading skills of institutional investors. *Journal of Finance* 66:601–33.

Saffi, P., and K. Sigurdsson. 2011. Price efficiency and short selling. *Review of Financial Studies* 24:821–52.

Shefrin, H. 1999. *Beyond greed and fear: Understanding behavioral finance and the psychology of investing*. Boston: Harvard Business School Press.

Shleifer, A., and R. Vishny. 1997. The limits of arbitrage. *Journal of Finance* 52:35–55.

Wermers, R. 2004. Is money really “smart”? New evidence on the relation between mutual fund flows, manager behavior, and performance persistence. Working Paper, University of Maryland.

The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXIII, NO. 3 • JUNE 2018

Margin Requirements and the Security Market Line

PETRI JYLHÄ*

ABSTRACT

Between 1934 and 1974, the Federal Reserve changed the initial margin requirement for the U.S. stock market 22 times. I use this variation to show that investors' leverage constraints affect the pricing of risk. Consistent with earlier theoretical predictions, I find that tighter leverage constraints result in a flatter relation between betas and expected returns. My results provide strong empirical support for the idea that the constraints investors face may help explain the empirical failure of the capital asset pricing model.

ONE OF THE FIRST AND MOST persistent anomalous findings in finance is that the return difference between high-beta and low-beta stocks is significantly smaller than predicted by the capital asset pricing model (CAPM) of Sharpe (1964), Lintner (1965), and Mossin (1966). An early explanation for the empirical flatness of the security market line, originally documented by Black, Jensen, and Scholes (1972), is given by Black (1972), who shows that investors' inability to borrow at the risk-free rate results in a lower cross-sectional price of risk than in the unconstrained CAPM. Black's version of the CAPM is unrealistic, however, as in the real world investors are able to borrow, just not in infinite amounts as assumed in the CAPM. This idea is further developed by Frazzini and Pedersen (2014), who present a model in which investors face a limit on their leverage. In their model, the slope of the security market line, that is, the return difference between high-beta and low-beta stocks, depends on the tightness of investors' leverage constraints: a tighter leverage constraint results in a flatter security market line.

Despite the theoretical and intuitive appeal, convincing empirical evidence of leverage constraints affecting the security market line is lacking. This is due in part to the difficulty of identifying an appropriate measure of the tightness

*Petri Jylhä is with Aalto University School of Business. I thank Kenneth Singleton (the Editor), two anonymous referees, Stéphane Chrétien, Stefanos Delikouras, Darrell Duffie, Samuli Knüpfer, Ralph Koijen, Ľuboš Pástor, Lasse Pedersen, Joshua Pollet, Oleg Rytchkov, Mungo Wilson, and participants at the American Economic Association 2015, European Summer Symposium in Financial Markets 2016, Financial Intermediation Research Society 2014, the Western Finance Association 2016 meetings, and seminar participants at Aalto University, Copenhagen Business School, Imperial College London, Luxembourg School of Finance, and Manchester Business School for helpful comments. I have read the *Journal of Finance*'s disclosure policy and have no conflicts of interest to disclose.

DOI: 10.1111/jofi.12616

of these constraints. A commonly used measure in the funding constraints literature is the spread between Eurodollar and Treasury bill rates, also known as the TED spread. However, such an interest rate spread is endogenous to investors' portfolio choice problem and does not directly measure the constraint on maximum leverage, but rather can be seen as a proxy for the cost of that leverage. Also, the empirical evidence in Cohen, Polk, and Vuolteenaho (2005) and Frazzini and Pedersen (2014) shows that a higher interest rate spread—typically used in the literature to indicate a tighter funding constraint—does not result in a flatter security market line as the theory would predict.

In this paper, I use a direct measure of investors' leverage constraints and find strong and robust empirical evidence in support of the theoretical prediction that tighter leverage constraints result in a flatter security market line. My measure of leverage constraints is based on the active management of the minimum initial margin requirement by the Federal Reserve. Pursuant to the Securities Exchange Act of 1934, the Federal Reserve, in its Regulation T, sets the minimum level of initial margin required when purchasing common stock on credit on U.S. stock exchanges.[1] Between October 1934 and January 1974, this margin requirement was changed 22 times and ranged between 40% and 100%. This frequent and sizable variation in a federally mandated leverage constraint provides an excellent setting for testing whether such constraints affect asset prices.

The main results of this paper are as follows. First, I show that the margin requirement significantly affects investors' leverage but is largely uncorrelated with other prevailing and future financial market and macroeconomic conditions. These findings establish that the federally set margin requirement is a useful measure of investors' leverage constraints and not merely a proxy for the overall state of the economy.

Second, and more importantly, I find that the slope of the security market line is negatively related to the prevailing margin requirement. Similarly, the intercept of the beta-return relation is positively related to the margin requirement. These findings are in line with the theoretical prediction of Black (1972) and Frazzini and Pedersen (2014). Figure 1 provides a simple illustration of these main results. Specifically, it separately plots the security market lines for periods of low, medium, and high initial margin requirements. The difference between the security market lines for low- and high-margin requirements is striking. When the margin requirement is low (between 40% and 55%), the empirical security market line runs very near its CAPM prediction. In contrast, during periods of a high (75% to 100%) initial margin requirement, the empirical security market line differs significantly from that predicted by the CAPM, and actually has a negative slope. This figure thus provides a simple but powerful summary of the main result of this paper.

[1] The initial margin requirement dictates the minimum value of collateral needed when purchasing stock. For example, a 40% initial margin requirement means that an investor can borrow up to 60% of the cost of a new stock purchase.

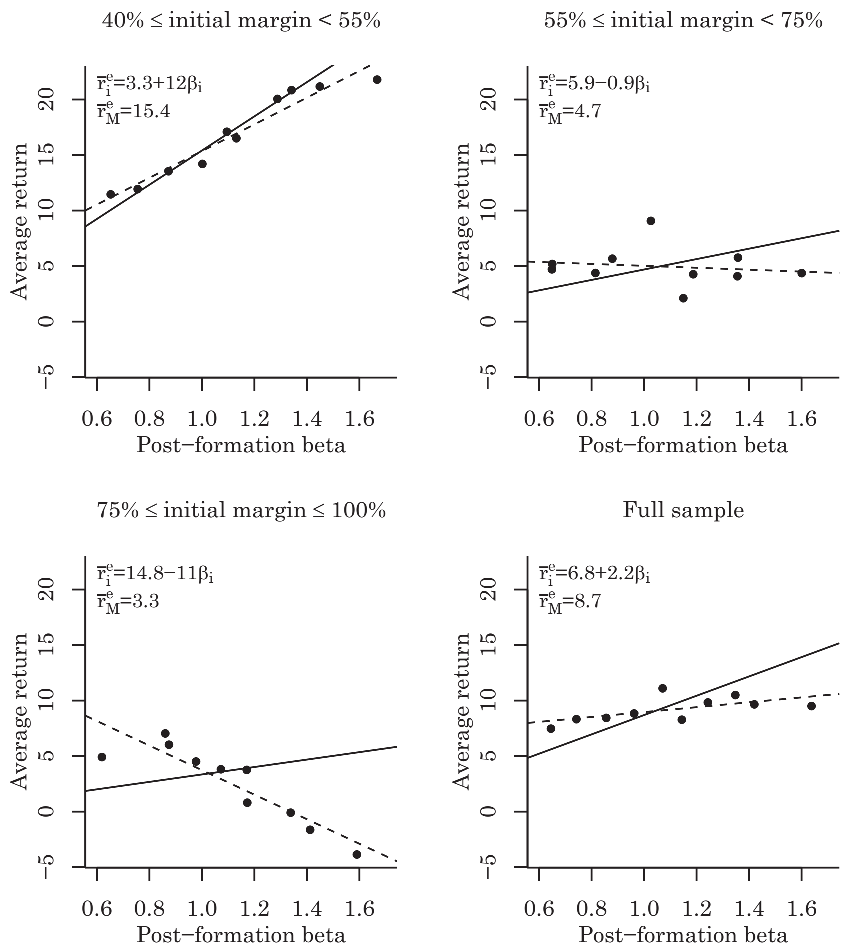


Figure 1. Initial margin requirement and security market line. This graph plots the empirical relation between beta and average excess return in subsamples with different initial margin requirements. The test assets are 10 beta-sorted value-weighted portfolios. The solid line gives the theoretical security market line predicted by the CAPM, and the dashed line gives the empirical security market line. The top left panel includes the 197 months for which the initial margin requirement is between 40% and 55%, the top right panel includes months for which the requirement is between 55% and 75% (183 months), and the bottom left panel includes months for which the requirement is above 75% (112 months). The bottom right panel presents the security market lines for the full sample of 492 months from 10/1934 to 9/1975.

Third, I show that the main result above is highly robust to using different test assets, control variables, and estimation techniques. In all specifications, the effects of leverage constraints on the security market line take the predicted sign and are statistically significant. In the most interesting specification, which is closely related to the analysis of Frazzini and Pedersen (2014), I include the difference between investors' borrowing and lending rates—a measure of the cost of leverage—as an additional explanatory variable. The results confirm that the interest rate spread does not explain the slope of the security market line, nor does it affect the explanatory power of the margin requirement.

These findings are somewhat at odds with those reported by Frazzini and Pedersen (2014). They use the empirically documented flatness of the security market line to motivate a trading strategy of betting-against-beta (BAB), which delivers high and consistent risk-adjusted returns across asset classes. To connect the profitability of this strategy to leverage constraints, they regress the BAB returns on their proxy for such constraints, the TED spread. Since the theory predicts that a tighter leverage constraint results in a flatter security market line, which, in turn, results in higher BAB returns, one would expect the strategy returns to be positively correlated with the TED spread. But this is not the case: the authors find a strong negative correlation between the strategy returns and the TED spread, which would imply that the security market line is steeper during more binding leverage constraints, not flatter as predicted by their model.[2] My results show that tighter leverage constraints do indeed result in a flatter security market line, consistent with the theoretical model of Frazzini and Pedersen (2014).

In an interesting recent paper, Bali et al. (2017) argue that the flatness of the security market line is driven by investors' demand for lottery-like stocks rather than by leverage constraints. They arrive at this conclusion by showing that the BAB returns disappear after controlling for lottery demand, as measured by the average of the five highest daily returns of the stock during the previous month. My results provide an alternative perspective on this discussion. Margin requirements should have little to do with investors' lottery preferences but are a key contributor to leverage constraints. Hence, the finding that a higher margin requirement results in a flatter security market line provides strong evidence in support of the leverage constraint explanation.[3]

This work is related to a stream of empirical papers examining the factors that affect the shape of the security market line. Cohen, Polk, and Vuolteenaho (2005) test whether investors suffer from a "money illusion" (Modigliani and Cohn (1979)) by examining the effect of inflation on the security market line. Consistent with the Modigliani-Cohn hypothesis, they find a negative relation between the level of inflation and the slope of the security market line. The empirical design of this paper closely follows that of Cohen, Polk, and Vuolteenaho (2005). Hong and Sraer (2016) show that aggregate disagreement affects both

[2] Cohen, Polk, and Vuolteenaho (2005) also show that a higher interest rate spread does not produce a flatter security market line.

[3] It should also be noted that investors' lottery demand spikes in January (Doran, Jiang, and Peterson (2012)), and stocks with strong lottery characteristics—such as a high maximum daily return (Bali, Cakici, and Whitelaw (2011)) or high idiosyncratic volatility (Ang et al. (2006))—tend to have high January returns. If lottery demand fully explained the flatness of the security market line, the BAB returns should be low in January when investors flock to buy high-beta stocks. Empirically, this is not the case—the strategy's average January return does not differ from the average return of the other months.

the slope and the concavity of the beta-expected return relation, and Antoniou, Doukas, and Subrahmanyam (2016) find that the security market line is flatter during periods of high investor sentiment. Huang, Lou, and Polk (2016) construct a measure of speculative capital committed to betting against beta and show that when this measure is high, the security market line tends to have a low (or even negative) slope.[4] Savor and Wilson (2014) show that the security market line is much steeper on macroeconomic announcement days than on nonannouncement days.

In providing empirical support for a constrained version of the CAPM, this paper also connects to the recent literature that provides non-return-based evidence in support of a single-factor model. Barber, Huang, and Odean (2016) and Berk and van Binsbergen (2016) find that mutual fund flows are better explained by the single-factor CAPM than more sophisticated models or more factors. The results in this paper show that the empirical rejection of the unconstrained CAPM in return-based tests is due at least in part to the leverage constraints faced by investors.

Finally, this paper relates to the literature examining the effects of the federal margin regulation on stock market volatility and on the use of credit to purchase securities. Starting with Moore (1966) and Officer (1973), the nearly unanimous conclusion in the literature is that margin regulation has no impact on market volatility (see Ferris and Chance (1988), Kupiec (1989), Schwert (1989), and Hsieh and Miller (1990)). Kupiec (1997) provides an extensive review of this literature. Moore (1966) and Hsieh and Miller (1990) report that an increase in the margin requirement leads to a decrease in the amount of margin credit. To my knowledge, this is the first paper to study the asset pricing implications of the Regulation T margin requirements.

The rest of the paper is organized as follows. Section I provides the theoretical motivation for the paper. Section II describes the margin regulation in the U.S. stock market. Section III introduces the estimation of the security market line. Section IV presents the empirical results. Section V concludes.

I. Asset Pricing with Margin Constraints

The idea that portfolio constraints affect the equilibrium relation between risk and expected returns is not new. Black (1972) shows that in a model with no risk-free borrowing, the security market line is flatter than in the CAPM. A more realistic assumption is that at least some investors can borrow, but their maximum level of leverage is exogenously restricted. This is the key assumption in the model of Frazzini and Pedersen (2014), which serves as the main theoretical motivation for this paper. Their model features overlapping generations of investors who face a realistic margin constraint: an investor's total investment in risky securities cannot exceed the fraction $1/m$ of her total

[4] In the regressions below, I control for these and other potential explanations for security market line flatness and show that my results are robust to the inclusion of additional control variables.

wealth. Frazzini and Pedersen (2014) consider a model in which the margin requirement, m, can vary across investors. Here, I focus on the simpler case in which all investors are subject to the same m. This framework is consistent with the empirical setup of the paper, which uses a market-wide margin requirement as a measure of leverage constraints.

If $m = 0$, the portfolio constraint will never bind and the model reduces to the CAPM. If $m = 1$, the investor cannot borrow the risk-free assets, as in Black (1972). If $m > 1$, investors are forced to hold part of their wealth in the risk-free asset. The most interesting, and realistic, case is $0 < m < 1$, which implies that investors can use leverage but face a margin requirement that limits the maximum amount of leverage. A margin requirement of $m \times 100\%$ means that the investor's own capital must make up at least $m \times 100\%$ of her total investment in risky securities. Hence, her total investment is limited to $1/m$ times her total wealth and the m in the model is equivalent to a real-world margin requirement. For example, if the margin requirement is $m = 40\%$, an investor with \$1 to invest can deposit the \$1 in a margin account, borrow up to \$1.5, and invest the total amount in risky securities. The margin (\$1) then equals 40% of the total invested amount (\$2.5) and the investor's investment in risky assets is $1/0.4 = 2.5$ times her wealth.

Intuitively, the margin requirement affects asset prices in the following manner. In an unconstrained CAPM world (where $m = 0$ and $1/m = \infty$), an investor with very low risk-aversion borrows heavily in the risk-free asset and invests in the market portfolio of risky assets. However, in the constrained world, she is not able to pursue this strategy, as the maximum amount of leverage is limited by the margin requirement. As she is no longer able to achieve her desired level of risk by leveraging her investment, she does so by investing in a portfolio with beta greater than one rather than in the market portfolio. Such behavior of constrained investors seeking higher portfolio risk creates higher demand for high-beta stocks than in the unconstrained CAPM case. In equilibrium, this results in higher prices and lower expected returns for high-beta stocks. Similarly, in the constrained case, the demand for low-beta stocks will be lower and their expected return will be higher than in the unconstrained case. This makes the security market line flatter in the presence of leverage constraints.

Formally, the relation between expected excess returns and beta, or the security market line, under margin constraints is given by

$$E\left(r_s^e\right) = \psi m + \beta_s \left[E\left(r_M^e\right) - \psi m\right], \tag{1}$$

where ψ is the average shadow price of the margin constraint in investors' portfolio optimization problem.[5] In the absence of a margin requirement, when $m = 0$, the security market line reduces to its CAPM form. Since $\psi \geq 0$, it is straightforward to see from equation (1) that, other things being equal, a higher margin requirement results in a higher intercept and a lower slope for the security market line. Below, I use time-variation in the federally mandated

[5] The Appendix provides the derivation of the security market line.

minimum initial margin requirement, a direct measure of m in the model, to show that this prediction holds empirically.

Frazzini and Pedersen (2014) use this model mainly to motivate the BAB trading strategy that goes long a portfolio of low-beta assets and short a portfolio of high-beta assets. The long and short legs are weighted by the reciprocals of their betas to make the resulting portfolio ex ante beta-neutral. In their empirical work, Frazzini and Pedersen show that such a strategy yields positive risk-adjusted returns in a number of asset markets.

To connect the profitability of the trading strategy to the leverage constraints, the authors regress the BAB strategy returns on the TED spread. The TED spread is often used in the literature as a measure of leverage constraints.[6] Since the BAB returns are higher when the security market line is flatter, which according to the theory is a result of tighter leverage constraints, one should expect to find a positive correlation between the TED spread and the profitability of BAB. But this is not what Frazzini and Pedersen (2014) find. Rather, they find a negative and statistically significant correlation between the lagged level of the TED spread and the BAB returns, which seems to be in direct contrast with the theoretical prediction. They also find a negative and significant correlation between the contemporaneous change in the TED spread and the BAB returns. They interpret these findings as suggesting that both explanatory variables proxy for the change in funding conditions, that is, a higher lagged level of the TED spread and a contemporaneous increase in the TED spread imply tightening leverage constraints. This may be a reasonable interpretation of the change in the TED spread, but not necessarily for the level of the spread, as it is commonly used as a measure of funding conditions, not as a measure of a change in funding conditions.

The TED spread, however, is not an optimal measure of the exogenously imposed leverage constraint, which is the key component of the theoretical model. First, rather than capture the maximum limit on investors' leverage, the TED spread more likely measures investors' costs of obtaining leverage. Frazzini and Pedersen (2014) implicitly posit that higher and increasing funding costs lead a broker to increase the margin requirements that it sets for its customers. A more plausible conjecture is that the broker increases the interest rate that it charges on margin lending rather than reduces the amount it lends. Hence, the TED spread is more plausibly a proxy for the spread between the borrowing and lending rates faced by investors. However, in the model, borrowing carries the same interest rate as lending but leverage is capped. Thus, what affects the security market line in the model is the constraint on maximum leverage, not the cost of obtaining that leverage.[7]

[6] Papers using the TED spread as a measure of leverage constraints include Asness, Moskowitz, and Pedersen (2013), Moskowitz, Ooi, and Pedersen (2012), Cornett et al. (2011), Ranaldo and Söderlind (2010), Gârleanu and Pedersen (2011), Brunnermeier and Pedersen (2009), and Brunnermeier, Nagel, and Pedersen (2008).

[7] In Section IV.C, I show that the relation between my measure of leverage constraints and the security market line is robust to including a direct measure of investors' leverage cost as a control variable.

Second, as a difference between two yields, the TED spread is derived from asset prices and hence is an outcome of investors' portfolio choice problem. Thus, for example, changes in investors' risk preferences or expectations could affect both the TED spread and the shape of the security market line, without any mechanism involving leverage constraints. This is especially critical when contemporaneous changes in the TED spread are used to explain asset returns. Third, Nagel (2016) shows that the TED spread measures the liquidity premium in Treasury bills, which has little to do with the limit on investors' maximum leverage. In this paper, I study the relation between leverage constraints and the security market line using a direct measure of leverage constraints that is not based on asset prices but rather on the minimum margin requirements set by the Federal Reserve.

II. Margin Regulation in the U.S. Stock Market

A. History of Margin Regulation

The Federal Reserve's control over margin requirements is based on the Securities Exchange Act of 1934. The Act bestows the responsibility for regulating the amount of credit that can be used for purchasing and carrying securities on the Board of Governors of the Federal Reserve System.[8] This move reflected the widely held view that unregulated stock market credit resulted in excessive leverage that fueled the stock market boom in the 1920s, and that the subsequent margin calls had exacerbated the market crash in 1929 (Hsieh and Miller (1990)). There was also a concern that loans extended to investors could crowd out loans to businesses and farmers.

Pursuant to the Securities Exchange Act, the Fed Board regulates margin borrowing by setting a minimum level for the initial margin that lenders must require. The margin requirements are set in Regulation T, which governs lending by brokers and dealers.[9] Analogous to m in the model presented above, an initial margin corresponds to the amount of cash or other collateral that investors must put down when purchasing stocks. For example, a 45% initial margin requirement means that investors can borrow up to 55% of the cost of a new investment in stock. Hence, a higher margin requirement translates into a tighter borrowing constraint. The Fed Board does not regulate the maintenance margin, which dictates the minimum amount of collateral required to carry the position. Also, lenders are allowed to require higher initial margins than what is set in the federal regulation.

Importantly from an econometrician's point of view, the initial margin requirement in Regulation T has not remained constant since its inception. Following the guidelines set forth in the Securities Exchange Act, the Fed Board changed the requirement 22 times between 1934 and 1974. The margin

[8] In line with common conventions, I use the terms "Fed Board" and "Board" instead of the formal "Board of Governors of the Federal Reserve System."

[9] Regulations U and X, and formerly G, apply similar margin requirements to borrowing from banks and other nonbroker-dealer lenders.

Table I

Margin Regulation Changes

This table lists the changes to the minimum margin requirement in the Federal Reserve's Regulation T. The first column gives the date when the new margin requirement was decided by the Fed Board, and the second column gives the date when the new requirement became effective. The following two columns give the change and the new level of the margin requirement. The four last columns indicate the reasons provided by the Board for changing the margin requirement, where the reasons are categorized as relating to developments in margin credit, stock prices, stock market activity, and consumer prices. The reasons are collected from the summary minutes of the Fed Board meetings available in the Board's annual reports.

Date		Margin		Reason for Change			
Decision (1)	Effective (2)	Change (3)	Level (4)	Credit (5)	Prices (6)	Activity (7)	Inflation (8)
	October 1, 1934		45%				
January 24, 1936	February 1, 1936	+10%	55%	×	×	×	
October 27, 1937	November 1, 1937	−15%	40%	×	×		
February 2, 1945	February 5, 1945	+10%	50%	×	×	×	
July 3, 1945	July 5, 1945	+25%	75%	×	×	×	×
January 17, 1946	January 21, 1946	+25%	100%				×
January 17, 1947	February 1, 1947	−25%	75%	×	×		×
March 28, 1949	March 30, 1949	−25%	50%	×			
January 16, 1951	January 17, 1951	+25%	75%	×	×	×	×
February 20, 1953	February 20, 1953	−25%	50%	×			×
January 4, 1955	January 4, 1955	+10%	60%	×		×	
April 22, 1955	April 23, 1955	+10%	70%	×		×	
January 15, 1958	January 16, 1958	−20%	50%	×	×		
August 4, 1958	August 5, 1958	+20%	70%	×	×	×	
October 15, 1958	October 16, 1958	+20%	90%	×	×	×	
July 27, 1960	July 28, 1960	−20%	70%	×	×	×	
July 9, 1962	July 10, 1962	−20%	50%	×			
November 5, 1963	November 6, 1963	+20%	70%	×		×	
June 7, 1968	June 8, 1968	+10%	80%	×			
May 5, 1970	May 6, 1970	−15%	65%	×	×		
December 3, 1971	December 6, 1971	−10%	55%	×			
November 22, 1972	November 24, 1972	+10%	65%	×	×		
January 2, 1974	January 3, 1974	−15%	50%	×			

requirement has remained unchanged at 50% since January 1974. Table I and Figure 2 present the time series of the initial margin requirement changes and levels.[10] As can be seen, the initial margin requirement shows frequent and substantial variation over time. At its lowest, from November 1937 to February 1945, the initial margin requirement was 40%, implying that, for every

[10] Between October 1, 1934 and March 31, 1936, the margin requirement depended countercyclically on the stock's price development over the preceding three-year period. Hence, over this period, the minimum margin requirement was a range rather than a single number. In accordance with previous literature and both Federal Reserve and NYSE statistics, I use the highest value of the range for this time period. The countercyclical margin requirement was abandoned in March 1936 due to it being unnecessarily complicated for lenders to manage. Since April 1, 1936, the margin requirement has been expressed as a single number.

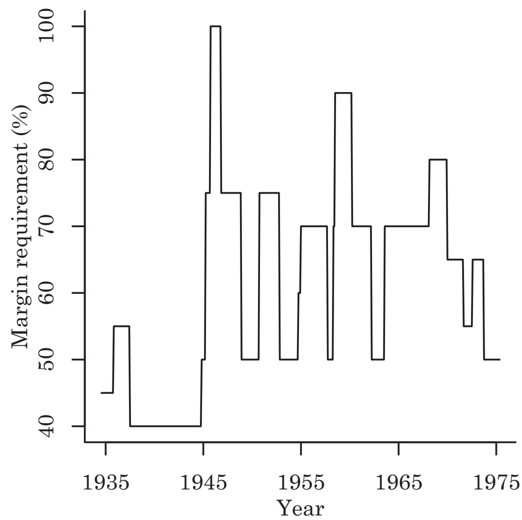


Figure 2. Initial margin requirement. This graph plots the level of the initial margin required on positions in listed U.S. equities. The initial margin requirement is set by the Board of Governors of the Federal Reserve System via Regulation T. The sample period is 10/1934 to 9/1975.

$1 of capital, an investor could purchase up to $2.5 of stocks. At its highest, from January 1946 to January 1947, the requirement was 100% completely forbidding new borrowing to purchase stocks. This range of variation indicates that the changes in Regulation T significantly affected the leverage constraints faced by investors.

During the sample period of this paper—from October 1934 to September 1975—stocks in the United States were held predominantly by households. The average level of household ownership of stocks between 1945 and 1975 was 84.7% while domestic institutions—mainly private pension funds, insurance companies, and mutual funds—held 12.8% and foreigners only 2.5% of U.S. stocks.[11] In comparison, the same figures for 2015 are 37.3%, 46.7%, and 16.0%, respectively. The high household ownership of stocks makes the Regulation T initial margin a highly relevant leverage constraint for the whole market during the sample period, as households do not face other exogenous leverage constraints like many institutional investors do.

Today, the overall picture of investors' leverage constraints is complicated by, for example, access to derivatives, regulations that apply only to some market participants (such as limits on mutual funds' use of leverage in the Investment Advisers Act of 1940), the use of offshore or joint back-office arrangements to circumvent the Regulation T margins (especially in the case of hedge funds), and the use of portfolio margining, which allows for lower margin requirements especially in portfolios that hold derivatives. These complications, however,

[11] The data come from Table L.223 of the flow of funds accounts and begin in 1945. The data are available online at https://www.federalreserve.gov/releases/z1/current/data.htm.

either did not exist or did not have a sizable impact during the sample period considered in this paper. Mutual funds were relatively small, holding, on average, only 3.2% of stocks, compared to 20.5% in 2015. Portfolio margining was introduced in 2005, joint back-office arrangements in 1998, and any offshore accounts were marginal as they are included in the 2.5% average foreign ownership. Simply put, during the sample period, investors' leverage constraints were more straightforward than presently and were dictated predominantly by the initial margin requirement in Regulation T.

It is important to note that the purpose of this paper is not to argue that the Regulation T initial margin requirement is the only relevant leverage constraint faced by investors. Rather the purpose is to use the variation in the Regulation T margin requirements to identify the effect of leverage constraints on the pricing of risk. In today's financial markets, with different investors facing very different leverage constraints, such identification would be very difficult if not impossible.

Consistent with the theories of Black (1972) and Frazzini and Pedersen (2014), in Section IV below, I show empirically that a higher level of the initial margin requirement results in a flatter security market line. However, before using changes in the minimum initial margin requirements to identify the effects of leverage constraints on asset prices, it is important to understand the reasons behind the margin requirement changes themselves. Specifically, the margin requirement should have two particular empirical properties to be a useful measure of the leverage constraints faced by investors. First, changes in the margin requirement should not strongly reflect economic or market conditions that also affect investors' expectations or preferences and, in turn, the security market line. The potential concern here is that any correlation between the margin requirement and the security market line could merely be a reflection of the former being a proxy for prevailing economic or market conditions. Second, the margin requirement should have a significant impact on investors' ability to obtain leverage. To show that these two conditions hold, Section II.B shows that changes in the margin requirement do not depend heavily on past market and macroeconomic variables, and Section II.C shows that changes in the margin requirement forecast future margin credit but no other market or economic variables.

B. Determinants of Margin Regulation Changes

To understand the relation between the margin requirement and economic and market conditions, it is useful to start from the reasons provided for changing the margin requirement. According to Section 7 of the Securities Exchange Act of 1934, the Federal Reserve should adjust the margin requirements from time to time for the purpose of preventing the excessive use of credit for the purchase and carrying of securities. The same section states that higher margin requirements should be prescribed when necessary or appropriate to prevent the excessive use of credit to finance transactions in securities but provides no guidance regarding what level of credit should be considered excessive, nor does it describe the potential adverse effects of excessive credit. Lower margin requirements should be prescribed when "necessary or appropriate for the accommodation of commerce and industry, having due regard to the general credit situation of the country."

To better understand the motivations behind the 22 margin requirement changes over the 1934 to 1974 period, I review the Records of Policy Actions provided in the annual reports of the Fed Board.[12] These records provide detailed descriptions of and rationales for various policy actions, including changes to the margin requirements. The reasons provided for the margin requirement changes can be broadly grouped as relating to changes in stock market credit, changes in the market prices of stocks, changes in speculative activity, or overall inflationary pressure. Table I lists the reasons provided for each margin requirement change by the Board.

Developments in stock market credit are mentioned as a reason for 21 of the 22 changes in the minimum initial margin requirement in Regulation T. This is not surprising given the explicit focus on the excessive use of credit in the Securities Exchange Act. Indeed, a strict reading of the Act would imply that the Board has a mandate to increase the margin requirement only to prevent margin credit from growing to excessive levels—not for any other reason. However, in 12 cases, the Board uses market returns to justify margin changes made in a countercyclical manner: higher margins are applied following increases in stock prices. The Board also mentions changes in speculative activity as a contributing factor behind 10 margin requirement changes.[13] Interestingly, the Board never gives any indication of how it measures speculative activity or why that activity should be curbed by higher margin requirements. Finally, overall inflationary pressure in the economy is cited as a reason for changing margin requirements in five cases.

To quantify the relation between the change in the minimum initial margin requirement and prevailing market and economic conditions, I regress the former on a number of measures of the latter and report the results in Table II. In column (1), the dependent variable is simply the change in the margin requirement in month t. Columns (2) and (3) present results of a multinomial logit regression for an increase and a decrease in the margin requirement, respectively. The explanatory variables are motivated by the above analysis of the Fed Board's reasons for changing the margin requirement. These are the changes in the logarithm of margin credit from month $t-13$ to $t-1$, stock market returns from month $t-13$ to $t-1$ and from month $t-37$ to $t-13$, the volatility and skewness of daily stock market returns measured from month $t-13$ to $t-1$, the value-weighted average daily turnover of NYSE-listed stocks over the period from month $t-13$ to $t-1$, the price-to-dividend ratio of the

[12] The annual reports of the Board of Governors of the Federal Reserve System are available online at https://fraser.stlouisfed.org/title/117.

[13] References to price developments and market activity in decisions to change the margin requirement caused occasional disputes among the Board, with governors dissenting from decisions and legal counsel advising the Board to adhere to the standards of the Securities Exchange Act (Meltzer (2003)).

Table II
Determinants of Regulation T Changes

This table presents results of regressing the change in the Regulation T minimum margin requirement in month t on the lagged financial market and macroeconomic variables. The explanatory variables are the change in the logarithm of the aggregate margin credit from month $t-13$ to month t, the stock market returns from month $t-13$ to month $t-1$ and from month $t-37$ to month $t-13$, the standard deviation and skewness of the daily stock market returns measured over months $t-13$ to $t-1$, the average share turnover measured over months $t-13$ to $t-1$, the stock market price-dividend ratio measured at the end of month $t-1$, and the changes in the logarithms of consumer prices, M1 money supply, and industrial production from month $t-13$ to $t-1$. Newey and West (1987) t-statistics with 12 lags are reported in parentheses and the R^2s are adjusted for degrees of freedom. The sample period is 10/1934 to 9/1975 with 492 monthly observations.

		Multinomial Logit	
	OLS Change (1)	Increase (2)	Decrease (3)
Constant	0.000	−5.402	−4.775
	(0.07)	(−7.26)	(−6.81)
Credit growth	0.007	1.702	−0.418
	(2.14)	(4.49)	(−0.49)
Market return 1-12	0.005	0.393	−1.360
	(2.90)	(0.85)	(−2.25)
Market return 13-36	0.003	0.713	−0.057
	(1.25)	(2.19)	(−0.13)
Market volatility	0.004	0.039	−0.662
	(1.78)	(0.06)	(−1.24)
Market skewness	0.002	0.816	−0.138
	(1.36)	(1.85)	(−0.41)
Share turnover	0.002	0.751	0.065
	(1.12)	(1.36)	(0.21)
Market P/D	−0.001	0.044	0.405
	(−0.61)	(0.12)	(0.75)
Inflation	0.001	0.324	0.072
	(0.40)	(0.76)	(0.18)
M1 growth	0.003	0.216	−0.830
	(1.55)	(0.48)	(−1.59)
IP growth	−0.004	−0.604	0.282
	(−2.24)	(−3.20)	(0.89)
R^2	0.037	0.012	

S&P Composite Index in month $t-1$, and the changes in the logarithms of the consumer price index, the M1 money supply, and industrial production from month $t-13$ to $t-1$.[14] The explanatory variables are standardized for ease of comparison.

[14] Margin credit data are collected from the Federal Reserve Board (1976a, 1976b) and the NYSE Facts and Figures database. The margin credit time series is constructed by chaining the following time series: "customers' debit balances (net)" in Table 143 of Federal Reserve Board (1976a) from October 1934 to December 1941,"customer credit, net debit balances with NYSE

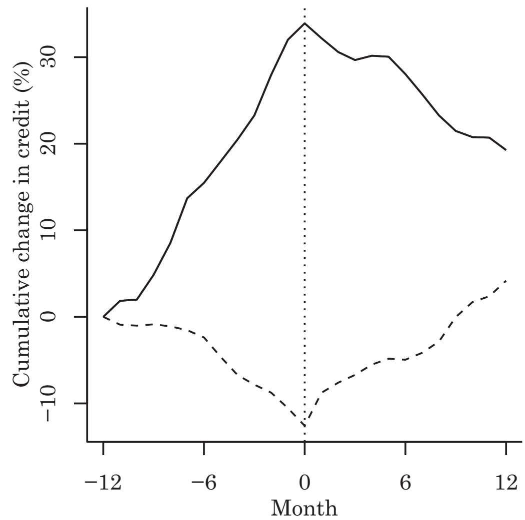


Figure 3. Margin requirement changes and margin credit. This graph plots the average cumulative change in margin credit 12 months before and after an increase (solid line) or a decrease (dashed line) in the minimum initial margin requirement in Regulation T. Month 0 is the margin change month. There are 12 margin requirement increases and 10 decreases during the sample period from 10/1934 to 9/1975.

Three important observations can be made from the results in Table II. First, changes in the margin requirement are significantly positively affected by changes in margin credit. This is consistent with the requirement of the Securities Exchange Act that the Fed Board uses the margin requirement to prevent excessive use of margin credit, as well as the fact that almost every change to Regulation T was motivated by developments in the amount of credit. In column (1), the coefficient on the credit change is 0.007, implying that a one-standard-deviation change in stock market credit (24%) results in, on average, a 0.7% increase in the margin requirement. This corresponds to 17% of a standard deviation of margin requirement changes (3.9%). In the multinomial logit model, the change in margin credit also has a statistically significant effect on the probability of a margin increase. However, credit growth does not significantly affect the probability of a margin decrease, though it takes the predicted negative sign.

Figure 3 illustrates the relation between changes in margin credit and the changes in the minimum initial margin requirement. The figure plots

firms" in Table 12.23 of Federal Reserve Board (1976b) from January 1938 to December 1967, and "margin debt" from NYSE Facts and Figures online database (http://www.nyxdata.com/nysedata/asp/factbook/main.asp) from January 1959 to September 1975. These three data sources partially overlap each other, which allows me to check that the data are consistent across sources. Stock return and volume data are from CRSP. The price-to-dividend ratio for the S&P Composite Index is from Robert Shiller's website (http://www.econ.yale.edu/shiller/data.htm). The macroeconomic data come from the Fred database maintained by the Federal Reserve Bank of St. Louis (https://research.stlouisfed.org/fred2/).

the average cumulative change in margin credit 12 months before and after margin requirement increases (solid line) and decreases (dashed line). In the 12 months preceding an increase in the margin, the amount of credit grows by 33.9% on average, whereas an average margin decrease is preceded by a 12.6% decrease in credit. Following the margin changes, the trend in credit growth strongly reverts. In Section II.C, I provide further evidence on the effects of margin changes on future developments in margin credit and other market and macroeconomic variables.

Second, changes in the margin requirement are also significantly affected by past returns. In the OLS model (column (1)), the market return measured over the previous 12 months has a positive and statistically significant effect on the margin changes. This implies that the Board has practiced a countercyclical policy whereby margins are increased in response to increasing stock prices. Third, no other variable (with the exception of the growth in industrial production) affects the Board's margin requirement decisions.

Overall, the above results indicate that the margin requirement is not merely a projection of prevailing financial market and macroeconomic conditions. In the empirical analysis below, however, I include these market and macro variables as controls when studying the effect of margin requirements on the security market line.

C. Effects of Margin Regulation Changes

As mentioned above, for it to be a useful measure of investors' leverage constraints, the federally mandated margin requirement should affect investors' ability to borrow to finance their stock purchases. Given that the explicit goal of margin regulation is to control the amount of credit, it is not surprising that Hsieh and Miller (1990) find that an increase in the Regulation T margin requirement results in a decrease in margin credit. Columns (1) and (5) of Table III confirm their finding, reporting results from a regression of the change in margin credit on the lagged change in the margin requirement:

$$\Delta credit_{t:t+k} = a + b\,\Delta margin_{t-1:t} + e. \tag{2}$$

The coefficient b in this regression is negative and statistically significant both for $k = 1$ and $k = 12$. This implies that, consistent with the spirit of the Securities Exchange Act and the findings of Hsieh and Miller (1990), an increase in margin requirement lowers the amount of credit used to purchase and carry stocks both in the short term and in the long term. This effect is also visible in Figure 3, where margin increases (decreases) are followed on average by a 1.7% decrease (3.8% increase) in credit over the next month, and a 14.6% decrease (16.8% increase) over the next year. In addition to statistically significant, these results are also economically large.

The strong relation between the margin requirement and credit is also echoed in the commentary on the Fed Board's decisions to change Regulation T. For example, with respect to its February 1953 action to decrease the margin, the

Table III

Determinants of Margin Credit Changes

This table presents results of regressing the change in margin credit on the lagged changes in the Regulation T minimum margin requirement and the call spread. The dependent variable is the change in the logarithm of the aggregate margin credit measured over one month (first four columns) and 12 months (last four columns). The call spread is the difference between brokers' call rate and the three-month Treasury bill rate, and proxies for the difference between investors' borrowing and lending rates. The control variables are defined in Table II. Newey and West (1987) t-statistics with 12 lags are reported in parentheses and the R^2s are adjusted for degrees of freedom. The sample period is 10/1934 to 9/1975 with 492 monthly observations.

	1 Month				12 Months			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.003	0.003	0.003	0.003	0.035	0.031	0.035	0.035
	(0.96)	(0.96)	(0.97)	(1.29)	(1.05)	(0.91)	(1.05)	(1.16)
Margin change	−0.178		−0.179	−0.232	−0.892		−0.893	−0.714
	(−2.75)		(−2.82)	(−4.12)	(−2.87)		(−2.88)	(−3.19)
Call spread change		−0.933	−0.997	−0.620		−0.935	−0.674	−1.945
		(−1.51)	(−1.55)	(−0.94)		(−0.29)	(−0.22)	(−0.75)
Controls	No	No	No	Yes	No	No	No	Yes
R^2	0.025	0.001	0.026	0.107	0.019	−0.002	0.017	0.200

Board wrote that "Stock market credit expanded immediately following the relaxation of margin requirements and stabilized thereafter. Such credit has not been large in amount for more than two decades."[15] Similarly, the July 1945 decision to increase the margin was later called "an important factor in restraining increase in credit."[16]

This result confirms that the federal margin regulation is a useful measure of leverage constraints, as it significantly affects the availability of leverage for investors. A finding of no relation between the margin requirement and margin credit would imply that the leverage constraint does not bind for sufficiently many investors for the regulation to have a significant impact. However, the result here shows that the leverage constraint does bind, and that changes in the margin requirement affect investors' access to credit.

One potential concern regarding the use of the Regulation T margin requirement as a measure of leverage constraints is that, as an initial margin requirement, it affects only new levered purchases of stocks, not existing positions. Consequently, the effects of Regulation T may take time to materialize. It should be noted, however, that Regulation T contains a component—the so-called retention requirement—that makes the initial margin requirement changes also affect existing positions. To understand the retention requirement, it is useful to note that there are three types of margin accounts. Accounts whose equity ratio is above the initial margin are known as unrestricted accounts. More stocks can be purchased into an unrestricted account

[15] *Annual Report of The Board of Governors of the Federal Reserve System*, 1953.
[16] *Annual Report of The Board of Governors of the Federal Reserve System*, 1945.

using the buying power of the excess equity until the account's equity ratio is equal to the Regulation T initial margin. An account whose equity value is below the maintenance margin receives a margin call and has to be refinanced so that the equity ratio is at least equal to the Regulation T margin. Between these two are the so-called restricted accounts. The equity ratio of a restricted account is higher than the maintenance margin, but lower than the initial margin.

The retention requirement affects the restricted accounts. Specifically, the retention requirement stipulates that, whenever stocks are sold from a restricted account, a certain fraction of the proceeds must be used to pay back the margin debt. The retention requirement was included in Regulation T on March 21, 1938 and is still in place. Until June 15, 1959, the retention requirement was equal to the initial margin requirement. This implies that if the initial margin was 60%, a holder of a restricted account had to spend at least 60% of any sales proceeds to pay back her margin debt. Since June 1959, the retention requirement has been set separately by the Board. The retention requirement makes initial margin changes affect existing accounts in two ways. First, increasing the initial margin also increases the fraction that holders of restricted accounts must use to pay back their debt, directly affecting their portfolio leverage. Second, increasing the initial margin makes new accounts restricted, as the equity ratio cutoff for restricted accounts is increased. Due to these mechanisms, changes in Regulation T initial margins affect not only new levered purchases, but also part of the existing levered portfolios.

As the Regulation T margin requirement is offered here as an alternative to using an interest rate spread as a measure of leverage constraints, it is useful to compare the effects of these two leverage constraint measures on margin credit. As mentioned above, studies that focus on more recent data typically use the TED spread as a proxy for the cost of leverage. Nowadays, brokers finance their positions primarily through repos and commercial paper issues (Adrian, Etula, and Shin (2015)), for which the TED spread is a reasonable measure of the interest rate spread over the risk-free rate. However, the Eurodollar market became central for interbank lending only after the sample period used in this paper (October 1934 to September 1975). Also, data on the Eurodollar rate do not extend back beyond the mid-1980s. Hence, it is impossible to include Regulation T margin requirement changes and TED spread changes in the same regression.

During the period when the Fed Board actively managed the Regulation T margin requirement, brokers obtained financing primarily from commercial banks in the form of call loans (Rappoport and White (1993)). The benchmark rate on these loans was brokers' call money rate. This call money rate was the interest rate paid by brokers for their funding. Hence, the spread between the call money rate and the Treasury bill rate, dubbed here "call spread," provides the counterpart of the TED spread for the 1934 to 1975 period.[17] To study the

[17] Data on brokers' call money rate come from Table 120 of Federal Reserve Board (1976a) from 1934 to 1941, Table 12.23 of Federal Reserve Board (1976b) from 1942 to 1970, and Survey of

effect of the cost of leverage on margin credit, I include the change in the call spread as an explanatory variable in equation (2).

Columns (2) and (6) of Table III show that changes in the call spread do not significantly predict changes in margin credit. The coefficients on the spread have the predicted negative sign but lack statistical significance. In regressions where both margin requirement changes and call spread changes are used to predict future margin credit changes (columns (3), (4), (7), and (8)), only the margin requirement has a statistically significant effect. This finding holds when forecasting one month or one year ahead, and when additional control variables are included in the regression. These results confirm the conjecture above that the margin requirement is a more appropriate measure of leverage constraints than the interest rate spread.

In addition to its effect on margin credit, another important empirical feature of margin regulation is that it does not affect the riskiness of stocks. A number of authors have studied the effect of the margin requirement on stock market volatility. Moore (1966), Officer (1973), and Ferris and Chance (1988) find no evidence that the Fed Board's margin regulation affects stock market volatility. However, the 1987 stock market crash reinvigorated discussion as to whether the Federal Reserve should again take a more active stance in managing the margin requirement. This discussion was further fueled by the finding of Hardouvelis (1990) that a higher margin requirement does result in lower stock market volatility. This finding has been disputed by a number of authors, including Kupiec (1989), Schwert (1989), and Hsieh and Miller (1990), who attribute the finding of Hardouvelis (1990) to flaws in the empirical tests. An extensive review by Kupiec (1997) concludes that there is no undisputed evidence that margin regulation affects stock market volatility.

To shed additional light on this result, I regress the one-month and 12-month changes in the volatility of daily stock market returns on the changes in margin requirements and report the results in Table IV. In line with the consensus in the literature, I find no statistically significant impact of margin requirement changes on stock market volatility in the short term or the long term. Also, the skewness of market returns is unaffected by the margin policy changes.

Theoretically, a higher margin requirement could increase or decrease stock return volatility. On the one hand, a higher margin requirement limits speculators' ability to provide liquidity and hence increases volatility. This concern was voiced by the *Wall Street Journal* during the preparation of the Securities Exchange Act: "Rigid fixation of minimum requirements threatens to produce disastrous consequences at a time of crisis. Such requirements are most likely to produce the effect of a series of stop-loss selling orders with the absence of any effective demand to meet them. The result can be easily imagined. The effect of the margin provisions in the bill will tend to accentuate in high degree the extent and the violence of these disturbances and cause large losses

Current Business from 1971 to 1975. The Survey of Current Business is a monthly publication by the Bureau of Economic Analysis and is available online at http://www.bea.gov/scb/.

Table IV

Effects of Margin Regulation

This table presents results of regressing the change in financial market and macroeconomic variables on the lagged change in the Regulation T minimum margin requirement. The dependent variables are the stock market return, the standard deviation and skewness of the daily stock market return, the average share turnover, and the changes in the logarithms of consumer prices, M1 money supply, and industrial production. The dependent variables are measured over one month (first three columns) and 12 months (last three columns). Newey and West (1987) t-statistics with 12 lags are reported in parentheses and the R^2s are adjusted for degrees of freedom. The sample period is 10/1934 to 9/1975 with 492 monthly observations.

	1 Month			12 Months		
	Constant (1)	ΔMargin (2)	R^2 (3)	Constant (4)	ΔMargin (5)	R^2 (6)
Market return	0.007	0.065	0.001	0.093	0.020	−0.002
	(3.13)	(1.80)		(3.59)	(0.12)	
Market volatility	0.000	0.015	−0.002	0.002	0.477	0.001
	(−0.03)	(0.18)		(0.05)	(1.08)	
Market skewness	0.001	0.578	−0.001	0.014	−1.353	0.001
	(0.11)	(0.45)		(0.13)	(−1.33)	
Share turnover	0.000	−0.006	0.003	−0.002	−0.017	−0.002
	(0.02)	(−1.16)		(−0.12)	(−0.20)	
Inflation	0.084	−0.211	0.001	0.034	0.060	0.002
	(5.28)	(−1.08)		(6.07)	(0.98)	
M1 growth	0.408	−0.363	−0.001	0.029	0.000	−0.002
	(6.16)	(−0.62)		(8.52)	(0.02)	
IP growth	0.077	−0.227	−0.001	0.049	0.039	−0.002
	(3.01)	(−0.29)		(3.15)	(0.29)	

to the public."[18] Similar skepticism toward margin regulation was echoed by the chairman of the New York Curb Exchange, Edwin Posner, in January 1947 when he commented that the decrease in the margin requirement from 100% to 75% "will have a beneficial effect on broadening the base of the securities markets introducing stability and narrowing the range within which stock prices move."[19]

On the other hand, a higher margin requirement could reduce unhealthy and excessive speculation and the buildup of highly levered positions whose deleveraging in a market downturn could increase market volatility. By limiting such volatility-increasing activities, a higher margin requirement could reduce volatility. This is the view held by the Fed Board prior to ending the management of margin requirements in 1974. In his statement at a congressional hearing in 1955, Fed chairman William McChesney Martin Jr. noted that "The task of the Board, as I see it, is to formulate regulations with two principal objectives. One is to permit adequate access to credit facilities for security

[18] "Public Interest Requires Further Changes in Bill," *Wall Street Journal*, March 23, 1934.

[19] "New York Bankers, Brokers Hail 75% Margins as Step in the Right Direction; Say It Should Be 50%," *Wall Street Journal*, January 18, 1947.

markets to perform their basic economic functions. The other is to prevent the use of stock market credit from becoming excessive. The latter helps to minimize the danger of pyramiding credit in a rising market and also reduces the danger of forced sales of securities from undermargined accounts in a falling market."[20] The fact that no empirical relation is discovered between the margin requirement and volatility could be a result of the two opposing effects canceling each other out, or both of the effects being too weak to be detected empirically.

The fact that the margin regulation does not have an effect on the fundamental riskiness of the stock market is important for the current study. Below, I show that a higher margin requirement results in a flatter security market line. If a higher margin were also associated with a less risky stock market, this finding could be justified by investors requiring a lower risk premium during less risky times. However, as the margin regulation has no impact on the riskiness of the market, the findings below support the hypothesis that investors' leverage constraints have an impact on the security market line. Also, I control for contemporaneous market returns in all of the regressions below. This captures the time-variation in the slope of the security market line resulting from the time-variation in the market risk premium.[21]

Finally, Table IV reports the effect of margin requirement changes on other market and macroeconomic variables. The results show that changes in the margin have no impact on market returns, trading activity, inflation, money growth, or industrial production. The result of no stock price impact is in line with contemporary commentary on the policy changes. Fed chairman William McChesney Martin Jr. opened his above-mentioned statement in the 1955 congressional hearing with, "Let me say at the outset that this responsibility of the Board of Governors relates to stock market credit and not to the price of stocks." Similarly, in its commentary on the Board's decision on October 15, 1958 to increase the margin, the *Wall Street Journal* wrote that "The financial community appeared to take the margin increase in stride, saying that it would have only a slight and shortlived effect on stock prices. Edward T. McCormick, president of the American Stock Exchange, said, 'I think the change is completely meaningless. I said at the time of the last increase (August 5) that it would have no impact on the market. I believe that has been proved by subsequent events.'"[22]

Overall, the fact that the margin changes are uncorrelated with future market and macroeconomic conditions is favorable for this study. Below, I show that the slope of the security market line is significantly correlated with the prevailing margin. Given that the margin is not correlated with general

[20] Statement of William McChesney Martin, Jr., Chairman, Board of Governors of the Federal Reserve System, at hearings on the study of the stock market before the Senate Committee on Banking and Currency, Monday, March 14, 1955, available online at https://fraser.stlouisfed.org/title/448.

[21] Table VII shows that adding contemporaneous market volatility as a control variable does not affect the key results of this paper.

[22] "Stock Margins Hiked to 90% from 70%, High Since '47; Record Stock Market Credit at End of September Cited," *Wall Street Journal*, October 16, 1958.

economic conditions, the reported results are unlikely to be an outcome of the margin acting as a proxy for market or macroeconomic conditions. However, I include the market and macro variables as controls in the regressions below.

III. Estimating the Security Market Line

The empirical strategy of this paper closely follows the efficient methodology developed by Cohen, Polk, and Vuolteenaho (2005) for a similar setup. I first sort stocks into portfolios based on their historical betas. Then, for every month, I estimate the cross-sectional relation between the portfolios' ex ante betas and realized returns. This yields a monthly series of security market line intercepts and slopes. Finally, in time series, I regress the intercept and the slope on the prevailing initial margin requirement and controls. The results clearly indicate that a high initial margin requirement results in a low security market line slope and a high intercept, consistent with the theoretical predictions of Black (1972) and Frazzini and Pedersen (2014) that leverage constraints flatten the security market line.

As the goal is to study the relation between CAPM betas and returns, I first construct a set of test assets that has a large spread in terms of betas. For every month, I calculate betas for all of the NYSE-listed common stocks of U.S.-domiciled corporations in the CRSP file by regressing the stocks' monthly returns in excess of the risk-free rate over the past three years on the value-weighted CRSP index return.[23] I then rank the stocks on the basis of the estimated betas and form 20 equally sized portfolios. The first portfolio contains the 5% of stocks with the lowest betas, and the 20th portfolio contains the 5% of stocks with the highest betas.[24]

Second, I estimate monthly betas for the 20 beta-sorted portfolios by regressing the value-weighted portfolio returns over the past 36 months on the value-weighted CRSP index return. The portfolios provide a set of test assets that has a wide range of postformation betas. The estimated beta of the first portfolio ranges between 0.2 and 1.0, while that of the 20th portfolio takes values between 1.3 and 2.2. The difference between the highest and lowest betas has an average of 1.2, and ranges between 0.6 and 2.2.

Third, I estimate the cross-sectional relation between the ex ante betas and the realized returns each month. I do this by regressing the portfolio excess returns during month t on the portfolio betas estimated using data from month $t-36$ to $t-1$. This way, there is no mechanical connection between the

[23] I exclude NASDAQ- and Amex-listed stocks as data availability is limited during the sample period, October 1934 to September 1975. CRSP has data on Amex stocks starting in July 1962 and on NASDAQ stocks starting in December 1972. To avoid any jumps in the number and types of stocks covered, I focus on NYSE-listed stocks, for which CRSP data begin in December 1925. Focusing on NYSE has the added benefit of excluding some of the smallest and most illiquid stocks.

[24] In Table VIII, I show that the results reported below are robust to using 10 or 40 portfolios instead of 20, constructing portfolios with equal market capitalization rather than equal number of stocks, and excluding the smallest 30% of stocks.

Table V
Descriptive Statistics

This table presents descriptive statistics for key variables used in the paper. *Margin* is the minimum initial margin requirement set by the Federal Reserve's Regulation T. *Intercept* and *slope* are the monthly security market line intercept and slope, respectively, which are constructed by regressing monthly the cross section of excess returns of 20 beta-sorted portfolios on the lagged estimated portfolio betas. *Market return* is the excess return of the CRSP value-weighted index. The sample period is 10/1934 to 9/1975 with 492 monthly observations.

	Margin (1)	Intercept (2)	Slope (3)	Market return (4)
Mean	0.613	0.006	0.002	0.007
Standard deviation	0.157	0.041	0.060	0.047
Skewness	0.286	−0.198	0.956	−0.377
Excess kurtosis	−0.691	7.395	6.312	3.570
25%	0.500	−0.014	−0.033	−0.019
Median	0.650	0.007	0.000	0.010
75%	0.700	0.023	0.027	0.031
Correlation with				
Intercept	0.100			
Slope	−0.134	−0.607		
Market return	−0.077	0.107	0.722	

dependent variable and the independent variable in the regression.[25] These regressions yield monthly time series of security market line slope and intercept coefficients. More specifically, every month I run the regression

$$r_t^e = intercept_t + slope_t\, \beta_{t-1} + e_t, \tag{3}$$

where *intercept* and *slope* are the estimated parameters. The estimates of the intercept and slope can also be seen as excess returns on two portfolios—the intercept represents the return on a portfolio that is a unit investment with a zero ex ante beta, whereas the slope is the return on a zero-investment portfolio with unit beta.

Table V presents descriptive statistics for the key variables used in this study: the initial margin requirement, the security market line intercept and slope, and the market excess return. A few interesting observations are worth noting. First, the average security market line intercept is positive and large, and the average slope (0.2%) is far smaller than the average market excess return (0.7%), indicating that the security market line over the period in question is flatter than predicted by the CAPM. Second, the security market line intercept is positively correlated with the initial margin, whereas the slope has a negative correlation with the margin. These univariate results are consistent with the prediction that stricter leverage constraints result in a flatter security market line. Third, the correlation between the margin requirement and the market

[25] Table VIII shows that the results are robust to using full-sample betas instead of the rolling-window betas.

return is relatively low, so there should be no concerns of multicollinearity in regressions where both are included as explanatory variables.

IV. Results

A. Margin Requirements and the Security Market Line

To test whether margin constraints have an effect on the relation between betas and expected returns, I regress the time series of the slope and intercept coefficients from equation (3) on the lagged initial margin requirement. Depending on the specification, I also include the excess return of the CRSP value-weighted index ($r^e_{M,t}$) and other additional controls (X_t) as explanatory variables. Formally, I run the following set of regressions:

$$intercept_t = a_1 + b_1\, margin_{t-1} + c_1\, r^e_{M,t} + d_1\, X_t + u_{1,t}, \tag{4}$$

$$slope_t = a_2 + b_2\, margin_{t-1} + c_2\, r^e_{M,t} + d_2\, X_t + u_{2,t}. \tag{5}$$

This pair of equations directly resonates with equation (1) above. Theoretically, the intercept of the security market line is given by the nonnegative shadow price of the leverage constraint times the required margin, and the slope is equal to the expected market return minus the shadow price times the margin. Hence, the coefficient b_1 should be nonnegative and the coefficient b_2 should be nonpositive. If b_1 and b_2 are not significantly different from zero, this would imply that the leverage constraint is nonbinding and does not affect the cross-sectional pricing of risk. In contrast, a significantly positive b_1 and a significantly negative b_2 would be direct empirical evidence in support of leverage constraints flattening the security market line as predicted by Black (1972) and Frazzini and Pedersen (2014). The theory also predicts that $b_1 = -b_2$. The results of estimating equations (4) and (5) are presented in Table VI.

Columns (1) and (4) of Table VI provide the main result of this paper. The coefficient of the lagged initial margin requirement in the intercept regression (b_1, column (1)) is equal to 0.024 and is statistically significant with a t-statistic of 2.2. The effect of the initial margin on the security market line slope (b_2, column (4)) is statistically significantly negative: -0.053 with a t-statistic equal to -3.7. These findings show that a higher margin requirement results in the security market line having a higher intercept and a lower slope in line with the theoretical predictions in equation (1).

The result that a higher margin requirement flattens the security market line could potentially be driven by a common factor affecting both the margin levels and the shape of the security market line. More specifically, the margin requirement could reflect general market conditions that also affect the security market line. I control for the most obvious candidate for such a common factor, the contemporaneous market return, in the regressions presented in columns (2) and (5) of the table. The market return is not significantly

Table VI

Margin Regulation and Security Market Line

This table presents results of regressing the monthly security market line intercept and slope on the lagged Regulation T minimum initial margin requirement, the contemporaneous market excess return, and controls. The security market line intercept and slope are constructed by regressing monthly the cross section of excess returns of beta-sorted portfolios on the lagged betas. The control variables are defined in Table II. Newey and West (1987) t-statistics with 12 lags are reported in parentheses and the R^2s are adjusted for degrees of freedom. The sample period is 10/1934 to 9/1975 with 492 monthly observations.

	Intercept			Slope		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	−0.009	−0.012	−0.032	0.035	0.013	0.035
	(−1.27)	(−1.50)	(−3.39)	(3.65)	(1.63)	(3.51)
Margin	0.024	0.027	0.060	−0.053	−0.029	−0.064
	(2.15)	(2.32)	(4.14)	(−3.73)	(−2.42)	(−4.25)
Market return		0.102	0.107		0.921	0.917
		(0.99)	(1.04)		(8.41)	(8.47)
Market return 1-12			−0.006			0.006
			(−2.24)			(2.21)
Market return 13-36			0.000			0.001
			(−0.19)			(0.45)
Credit growth			0.015			−0.015
			(3.91)			(−3.96)
Market volatility			0.007			−0.007
			(2.92)			(−2.70)
Market skewness			0.004			−0.004
			(1.74)			(−1.77)
Share turnover			−0.005			0.005
			(−2.37)			(2.34)
Market P/D			0.003			−0.002
			(1.03)			(−0.79)
Inflation			0.003			−0.003
			(1.26)			(−1.20)
M1 growth			−0.006			0.006
			(−2.77)			(2.82)
IP growth			−0.002			0.002
			(−1.16)			(0.95)
R^2	0.007	0.018	0.059	0.018	0.526	0.544

correlated with the security market line intercept, and adding it as an explanatory variable in the intercept regression has no discernible impact on the coefficient of the margin, which remains statistically significant. In contrast, the market return is (quite naturally) highly correlated with the security market line slope. Adding the market return to the slope equation affects the coefficient on the margin, but it still remains statistically significantly negative (t-statistic −2.4).

Since the theory predicts that $b_1 = -b_2$, I estimate equations (4) and (5) jointly in a seemingly unrelated regressions framework and run a Wald test to empirically test the coefficient restriction. The Wald test F-statistic is equal to 0.007 with an associated p-value of 0.93, which implies that the restriction cannot be rejected. This provides further support for the model.[26]

In addition to the market return, other measures of market or macroeconomic conditions could be correlated with the margin requirement and affect the security market line. To confirm that the main result here is not driven by such omitted variables, I add a number of control variables that could correlate with both the Federal Reserve's margin policy and investors' preferences or expectations. These variables are the stock market return over the previous 12 months and over the previous 36 to 12 months; the standard deviation and skewness of the stock market return over the previous 12 months; the change in margin credit over the previous 12 months; the average monthly turnover of NYSE-listed stocks over the previous 12 months; the price-dividend ratio of the S&P 500 index; and the changes in consumer prices, M1 money supply, and industrial production over the past 12 months. These variables are the same as those used above to explain the Fed Board's decisions to change the margin requirement. Also, previous literature has established that the security market line's shape depends on inflation (Cohen, Polk, and Vuolteenaho (2005)) and investor sentiment (Antoniou, Doukas, and Subrahmanyam (2016)), both of which might be correlated with the Fed Board's decisions to alter the minimum initial margin requirement.[27]

Columns (3) and (6) of Table VI present the results with the controls in place. Many of the control variables do affect the security market line. For example, high growth in margin credit predicts a flatter security market line. This is consistent with the findings of Antoniou, Doukas, and Subrahmanyam (2016), who document that high investor sentiment results in a flatter beta-return relation. Most importantly, however, the results show that the findings above are not driven by a confounding market or economic factor affecting both the margin requirement and investors' preferences or expectations. Indeed, including the controls actually results in a stronger effect of the margin requirement on the security market line. The coefficient on the margin requirement is 0.060 (t-statistic 4.1) in the intercept equation and -0.064 (t-statistic -4.2) in the slope equation. All of the results below are presented with the control variables to ensure robustness.

It is important to note that these results differ from those presented by Frazzini and Pedersen (2014) regarding the relation between leverage constraints and the slope of the security market line. They find that the lagged level of the TED spread has a negative effect on the returns of their BAB strategy. A high TED spread is usually taken to be an indication of tighter leverage

[26] The F-statistics of the Wald tests are very small for all of the specifications below, and the restriction is never rejected. This is driven, to a large extent, by the high negative correlation between the estimates of the security market line slope and intercept.

[27] The ideal measure of sentiment would be the Baker and Wurgler (2007) sentiment index. Unfortunately, monthly data on this index only go back to 1965. Stock turnover is the only component of the index for which monthly data extend back to the beginning of the sample period used here, 1934. Also, the change in margin credit should capture investor sentiment, as investors are plausibly more likely to lever their portfolios when sentiment is high.

constraints, which, according to the theory, should result in a flatter security market line and higher returns to the strategy. Hence, their result seems to be in contrast with the model prediction. They rationalize this result as the lagged TED spread possibly being a proxy for the change in credit conditions, in which case the negative correlation would be expected. As argued above, the margin requirement is a better and more direct measure of the leverage constraints faced by investors than the TED spread, and hence it is not surprising that the results here are more in line with the predictions of the theoretical model.

These results also have direct consequences for the empirical testing of the CAPM. As binding margin constraints make the beta-expected return relation flatter, they also help to reject the CAPM hypothesis that the security market line has a zero intercept and a slope equal to expected market excess return. Following Cohen, Polk, and Vuolteenaho (2005), one can define the conditional excess slope and intercept of the security market line as

$$\frac{a_2}{c_2} + \frac{b_2}{c_2}\, margin_{t-1} \tag{6}$$

and

$$a_1 - \frac{a_2\, c_1}{c_2} + \left(b_1 - \frac{b_2\, c_1}{c_2}\right) margin_{t-1}, \tag{7}$$

respectively. Using the coefficients in columns (2) and (5) of Table VI shows that an initial margin of around 44% would result in the excess slope and intercept being equal to zero. This simple calculation indicates that at relatively low levels of initial margin, the CAPM might not be rejected empirically. The sample period average margin requirement, however, is higher (61%) than the low level required to match the CAPM predictions, and the CAPM is rejected in the data.

This result raises an interesting question regarding the years after the Fed Board ended active management of the Regulation T margin requirement. The requirement has remained at 50% since 1974. During the period from 1975 to 2012, the security market line has an average monthly slope equal to 0.34%, whereas the average market excess return is 0.58%. Clearly, the 50% margin requirement does not result in the slope being equal to the market excess return in this sample, even though it is not too much higher than the 44% estimated above to result in a zero excess slope. This is most likely driven by the large changes in stock ownership over the past decades. As mentioned above, households, whose only leverage constraint comes from the Regulation T margin requirements, held on average 84.7% of the U.S. stocks during the sample period. This number has decreased to 37.3% in 2015. At the same time, institutions have increased their ownership stake significantly. For example, mutual funds held, on average, 3.2% of the stocks during the sample period, and 20.5% by 2015. Many institutions face leverage constraints beyond those in Regulation T, such as the limits in the Investment Advisers Act of 1940 and internal rules in the case of mutual funds (Almazan et al. (2004)). Such an increase in the importance of leverage-constrained institutions has likely

Table VII

Controlling for Additional Risk Factors

This table presents results of regressing the monthly security market line intercept and slope on the lagged Regulation T minimum initial margin requirement, the contemporaneous market excess return, and controls. The security market line intercept and slope are constructed by regressing monthly the cross section of excess returns of beta-sorted portfolios on the lagged betas. *Market volatility* is the standard deviation of the daily market returns measured over the month, *SMB* and *HML* are the Fama and French (1993) size and value factors, and *UMD* is the momentum factor (Carhart (1997)). The other control variables are defined in Table II. Newey and West (1987) t-statistics with 12 lags are reported in parentheses and the R^2s are adjusted for degrees of freedom. The sample period is 10/1934 to 9/1975 with 492 monthly observations.

	Intercept			Slope		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	−0.018	−0.019	−0.005	0.020	0.021	0.008
	(−1.30)	(−2.34)	(−0.52)	(1.42)	(2.51)	(0.74)
Margin	0.059	0.041	0.041	−0.064	−0.044	−0.044
	(4.11)	(3.14)	(3.25)	(−4.22)	(−3.27)	(−3.38)
Market return	0.052	0.271	0.211	0.973	0.743	0.802
	(0.42)	(3.10)	(2.15)	(7.50)	(8.18)	(7.90)
Market volatility	−0.118		−0.117	0.120		0.115
	(−1.73)		(−2.84)	(1.71)		(2.78)
SMB		−0.498	−0.519		0.517	0.538
		(−4.31)	(−4.36)		(4.46)	(4.51)
HML		−0.242	−0.220		0.250	0.229
		(−2.20)	(−2.12)		(2.30)	(2.20)
UMD		−0.032	−0.068		0.014	0.049
		(−0.29)	(−0.67)		(0.12)	(0.46)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.078	0.189	0.207	0.554	0.612	0.620

contributed to the flatter security market line in the more recent times, as suggested by Karceski (2002), Buffa, Vayanos, and Woolley (2014), Frazzini and Pedersen (2014), Christoffersen and Simutin (2017), and Boguth and Simutin (2018).[28]

B. Robustness Checks

Tables VII and VIII provide robustness checks of the main results by using alternative sets of test assets and by controlling for additional factors. The first candidate for an additional control variable is contemporaneous market volatility. The lagged change in market volatility is already included in the tests reported in columns (3) and (6) of Table VI. However, one might argue

[28] Extending the sample period of the regressions reported in Table VI to include the 1975 to 2012 period does not change the results qualitatively or quantitatively. As the margin requirement remained constant for this period, extending the data does not improve identification of the effect of margin requirements on the security market line. Importantly, however, the margin requirement affects the security market line in a statistically significant way also in the 1934 to 2012 sample.

Table VIII

Alternative Test Assets

This table presents results of regressing the monthly security market line intercept and slope on the lagged Regulation T minimum initial margin requirement, the contemporaneous market excess return, and controls using alternative test assets. Columns (1) and (2) present the results using 10 or 40 beta-sorted portfolios. In column (3), the 20 beta-sorted portfolios are constructed by first excluding from the sample the smallest 30% of stocks each month. In column (4), the 20 beta-sorted portfolios are constructed such that they all have the same total market capitalization each month. In column (5), the portfolio betas are estimated using the full sample of data, rather than a rolling window. In columns (6) and (7), the test assets are the 25 size and book-to-market portfolios and the 41 industry portfolios, respectively. The security market line intercept and slope are constructed by regressing monthly the cross section of excess returns of test assets on the lagged betas. The control variables are defined in Table II. Newey and West (1987) t-statistics with 12 lags are reported in parentheses and the R^2s are adjusted for degrees of freedom. The sample period is 10/1934 to 9/1975 with 492 monthly observations.

Panel A: Intercept

	Beta-Sorted Portfolios					Other Portfolios	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
N	10	40	20	20	20	25	41
Constant	−0.035	−0.027	−0.030	−0.028	−0.030	−0.032	−0.023
	(−3.49)	(−2.92)	(−3.14)	(−2.78)	(−2.92)	(−2.70)	(−2.82)
Margin	0.064	0.051	0.057	0.053	0.059	0.062	0.045
	(4.21)	(3.66)	(3.88)	(3.43)	(3.62)	(3.21)	(3.65)
Market return	0.074	0.149	0.107	0.166	0.010	0.231	0.387
	(0.69)	(1.51)	(1.04)	(1.63)	(0.08)	(2.10)	(5.74)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.048	0.071	0.060	0.080	0.046	0.047	0.243

Panel B: Slope

	Beta-Sorted Portfolios					Other Portfolios	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
N	10	40	20	20	20	25	41
Constant	0.037	0.030	0.031	0.027	0.030	0.039	0.028
	(3.60)	(3.14)	(3.21)	(2.71)	(2.87)	(3.43)	(3.56)
Margin	−0.068	−0.056	−0.060	−0.052	−0.057	−0.072	−0.052
	(−4.31)	(−3.87)	(−3.95)	(−3.36)	(−3.58)	(−3.90)	(−4.30)
Market return	0.948	0.878	0.912	0.837	0.990	0.791	0.636
	(8.43)	(8.48)	(8.44)	(7.98)	(7.56)	(7.07)	(10.02)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.525	0.561	0.539	0.522	0.556	0.331	0.482

that the margin requirement could affect volatility—even though the empirical evidence in existing literature and Table IV is not consistent with this view—and that volatility affects the security market line with investors requiring a lower price of risk during periods of lower volatility. This effect of a lower risk premium resulting from lower volatility should already be captured by the contemporaneous market return. The results presented in columns (1) and (4) of Table VII show that contemporaneous market volatility has some explanatory power over the security market line: higher volatility results in a lower intercept and higher slope of the line. However, since the volatility is uncorrelated with the margin requirement, the key result remains unchanged by the inclusion of volatility as a control: the margin requirement still has a strong impact on the security market line.

As the estimates of the security market line intercept and slope can also be interpreted as portfolio returns, it is natural to check that any regularity concerning them does not arise simply from exposures to standard risk factors. To do so, columns (2) and (5) of Table VII include the size and value factors (Fama and French (1993)) and the momentum factor (Carhart (1997)) as additional controls. Both SMB and HML have a significant positive (negative) correlation with the security market line slope (intercept). This is not surprising, as small firms and value firms are known to have higher betas than large firms and growth firms (e.g., Fama and French (1993), Novy-Marx (2014)). Importantly, however, the results presented in columns (2) and (5) show that the inclusion of these risk factors does not alter the main result. In the presence of all the controls (columns (3) and (6)), the coefficient on the margin requirement is statistically significantly positive in the intercept equation, and significantly negative in the slope equation. This implies that the main result of the paper is not driven by the security market line being affected by contemporaneous market volatility or other risk factors.

In Table VIII, I confirm that the main result is also robust to using alternative test assets. So far, the results have been presented with the security market line intercept and slope estimated based on 20 beta-sorted portfolios. In columns (1) and (2), the security market line is estimated from 10 and 40 beta-sorted portfolios, respectively. These changes have no impact on the main result.

A potentially valid concern is that, even though the portfolios are value-weighted, the results could be affected by relatively small stocks. This is especially so given that extreme beta values are more common among smaller stocks. Columns (3) and (4) of Table VIII confirm that the results above are not driven by small stocks. In column (3), I exclude micro-cap stocks—defined as stocks falling in the three smallest size deciles—from the sample and sort the remaining stocks into 20 value-weighted portfolios to estimate the security market line. In column (4), I construct 20 beta-sorted portfolios so that each month each portfolio has the same total market capitalization, rather than the same number of stocks. The key result of the paper remains unchanged using these alternative portfolio construction methods to alleviate the concern that the extreme beta portfolios are populated by very small stocks. The margin requirement has a statistically significant negative (positive) impact on the security market line slope (intercept), and the coefficients of interest are very close to those reported in Table VI above.

To alleviate the potential concern that time-variation in the estimation error of betas is correlated with the margin requirement, column (5) presents the basic result with the monthly security market lines estimated using full-sample betas of the portfolios rather than rolling-window betas. The results here are very similar to those reported in Table VI, confirming that those results are not driven by time-variation in beta estimates.

Finally, columns (6) and (7) of Table VIII report the key results using other than beta-sorted portfolios. In column (6), the test assets are the 25 size and book-to-market portfolios of Fama and French (1993), and in column (7), the test assets comprise 41 of the 49 Fama and French industry portfolios for which full return history are available for the October 1934 through September 1975 sample period.[29] The main result of the paper also holds for these alternative sets of test assets. Overall, the result that the margin requirement affects the security market line is robust to using alternative test assets and portfolio construction techniques.

A potential concern is that the level of the margin requirement is highly persistent, which could affect statistical inference. To confirm that the results are not materially affected by this, I perform three alternative tests: I cluster the standard errors, I simulate the standard errors, and I collapse the data so that one margin requirement regime is represented by a single observation.[30] Clustering the standard errors by regime increases the standard error estimates only marginally, and does not affect the conclusions. In the slope (intercept) equation, the clustered t-statistic of the margin requirement is -4.02 (3.94), whereas the unclustered Newey and West (1987) t-statistic is -4.25 (4.14).

As the second way of confirming that the results are not driven by persistence of the margin requirement, I simulate a time series of random margin requirements, and use that in the regressions instead of the true margin. The simulated margin requirement (m^s) is generated by the following system:

$$m_t^s = m_{t-1}^s + I_t x_t, \tag{8}$$

$$I_t \sim Bernoulli(p = 22/492), \tag{9}$$

$$x_t \sim Normal(\mu = 0, \sigma = 0.189). \tag{10}$$

Every month there is a 22/492 probability that the simulated margin requirement changes. This matches the data, as there are 22 margin requirement changes during the 492-month sample period. If a change happens, the size of the change is drawn from a Normal distribution with standard deviation equal to that of the margin requirement changes in the data. This yields a time series that has statistical properties similar to the true margin requirement but, by construction, has no explanatory power over the security market line.

Using this simulated margin requirement, I estimate regressions (4) and (5) using the same controls as in columns (3) and (6) of Table VI. I repeat the simulation and regressions 10,000 times and collect the Newey and West (1987) t-statistics of the simulated margin requirements in the regressions. This gives

[29] The data for the size and book-to-market portfolios and for the industry portfolios are from Kenneth French's website.

[30] By a regime, I refer to the period between two consecutive changes in margin requirement.

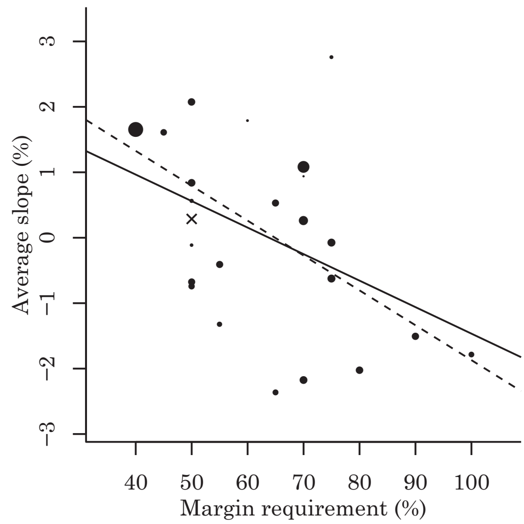


Figure 4. Margin requirement regimes. The dots in this figure plot the margin requirement and the average security market line slope for the 23 Regulation T margin requirement regimes during the sample period from 10/1934 to 9/1975. The area of the dots is proportional to the length of the regimes. The solid line gives the ordinary least squares fit of the data, whereas the dashed line gives the weighted least squares fit. The cross depicts the margin requirement and average security market line slope for the period from 10/1975 to 12/2012.

the distribution of the t-statistics under the null hypothesis of no relation between the margin requirement and the security market line. In 14 of the 10,000 simulations, the absolute value of the t-statistic of the simulated margin requirement in the intercept equation exceeds that reported in column (3) of Table VI. This means that the simulated p-value of the margin requirement coefficient is 0.0014. In the slope equation, the absolute simulated t-statistic exceeds the estimated t-statistic 16 times in 10,000 simulations, resulting in a p-value of 0.0016. These results show that the statistical significance of the margin requirement in explaining the security market line intercept and slope is not a mere artifact of the persistent nature of the margin.

Finally, I collapse the data into 23 observations, one observation for each margin requirement regime, and regress the average security market line slope during the period on the margin requirement. The results are best summarized in Figure 4, which plots the average slope against the margin requirement. The lengths of the margin requirement regimes vary from just 2 months to 88 months. To illustrate the amount of data each point in the figure is based on, the area of each dot in the plot is proportional to the length of the regimes. It is evident from the figure that there is a negative relation between the margin requirement and the security market line slope. The solid line in the figure plots the OLS fit, whereas the dotted line gives the weighted least squares (WLS) fit using the regime lengths as weights. An OLS (WLS) regression of the average security market line slope on the margin requirement and the average market excess return yields a coefficient of −0.022 (−0.036) with an associated t-statistic of −1.69 (−2.73) for the margin. These coefficients are close to the values reported in Table VI for the monthly regression, and the OLS coefficient is significant at the 10% level, whereas the WLS coefficient, which underweights the observations based on only a few months of data, is significant at the 5% level.[31]

Overall, the checks presented above show that the effect of the margin requirement on the security market line is strongly robust to various alternative estimation methods and control variables. The next two sections show that the main result of the paper is not driven by investors' costs of obtaining leverage or the short sales constraint implied by Regulation T margin requirements.

C. Cost of Leverage

Besides the limits on borrowing, another leverage-related factor that theoretically results in a flatter beta-return relation is the difference between the lending and borrowing interest rates faced by investors. When the two rates differ, the efficient portfolios involving borrowing lie on a line that is flatter than the line of portfolios involving lending. In this section, I confirm that the results above are not driven by correlation between margin requirements and the spread between the borrowing and lending rates.

My measure for the spread is the difference between the so-called brokers' call money rate and the three-month Treasury bill rate. As discussed above in Section II.C, during the sample period, the brokers' call money rate was the benchmark rate on the brokers' primary source of funding. Customers' margin borrowing typically carried an interest rate defined as the brokers' call rate plus a spread (Statman (1987)). Assuming that customers' spread remained constant, the difference between the brokers' call rate and the three-month Treasury bill rate provides a good approximation of the time-series variation of the difference between the borrowing and lending rates faced by investors. Table IX presents the results of estimating regression equations (4) and (5) including the spread between the brokers' call rate and the Treasury bill rate—the call spread—as an explanatory variable. Theoretically, a higher spread should make the security market line flatter, so one would expect a negative (positive) correlation between the call spread and the security market line slope (intercept).

Columns (1) and (3) of the table provide the results without including the margin requirement as an explanatory variable. The coefficients on the call spread have the expected signs—positive in the intercept equation and negative in the slope equation—but are not statistically significant. Columns (2) and (4) confirm that the results presented above are not driven by a correlation between the margin requirement and the call spread. Including the call

[31] In a regression of the regime-to-regime change in the security market line slope on the change in margin requirements, the OLS coefficient on the margin requirement change is −0.053 with a t-statistic of −2.45. This result is also in line with those reported using the monthly data in Table VI.

Table IX

Controlling for Cost of Leverage

This table presents results of regressing the monthly security market line intercept and slope on the lagged Regulation T minimum initial margin requirement, investors' cost of leverage, the contemporaneous market excess return, and controls. The security market line intercept and slope are constructed by regressing monthly the cross section of excess returns of beta-sorted portfolios on the lagged betas. *Call spread* is the difference between brokers' call rate and the three-month Treasury bill rate, and proxies for the difference between investors' borrowing and lending rates. The control variables are defined in Table II. Newey and West (1987) t-statistics with 12 lags are reported in parentheses and R^2s are adjusted for degrees of freedom. The sample period is 10/1934 to 9/1975 with 492 monthly observations.

	Intercept		Slope	
	(1)	(2)	(3)	(4)
Constant	0.005	−0.035	−0.004	0.038
	(1.14)	(−3.46)	(−0.98)	(3.69)
Margin		0.061		−0.066
		(4.32)		(−4.48)
Call spread	0.011	0.167	−0.054	−0.222
	(0.03)	(0.48)	(−0.14)	(−0.61)
Market return	0.103	0.107	0.921	0.917
	(0.99)	(1.04)	(8.39)	(8.45)
Controls	Yes	Yes	Yes	Yes
R^2	0.035	0.058	0.531	0.544

spread in the baseline regression (columns (3) and (6) of Table VI) does not alter the coefficient on the margin requirement. This is not surprising, given that the correlation between the margin requirement and the call spread is only 0.02.[32]

D. Short Sales Constraints

In addition to leverage constraints and the cost of leverage, short sales constraints can also result in a flatter security market line (see, e.g., Schnabel (1984) and Hong and Sraer (2016)). Since November 1937, Regulation T has also dictated the minimum initial margin on short sales, and since February 1945, the initial margin on short sales has been identical to the margin on stock purchases.[33] In this section, I provide some evidence that the results presented

[32] Cohen, Polk, and Vuolteenaho (2005) also conclude that the variation in the difference between investors' borrowing and lending rates does not appear to explain the variation in the security market line slope. They use interest rates on car loans, personal loans, and credit cards as measures of borrowing rates. The brokers' call rate used here is a more direct measure of the rate paid on margin borrowing.

[33] Prior to November 1937, the initial margin on short sales was defined as "the margin customarily required by the brokers and dealers." Between November 1937 and February 1945, the short sales margin was 50%, whereas the margin requirement on purchases was 40%.

above are not significantly driven by the short sales constraints affecting the security market line.

First, short interest was very low throughout the sample period. The aggregate short interest ratio varied between 0.03% and 0.19%, with an average value of 0.08%.[34] There are two reasons why such a low short interest ratio cannot result from the regulation T requirements alone. First, margin credit, which faces an identical margin requirement, was on average 1.4% of the total NYSE market capitalization during the sample period, which is 18 times larger than the short interest. Second, the short interest ratio has grown steadily to about 5% in 2015, while the initial margin requirement has remained constant at 50%. However, margin credit relative to market capitalization does not exhibit such massive growth from its 1934 to 1975 average, being about 2% in 2015. The volume of shorting was also never mentioned by the Fed Board as a motive for altering the margin requirements. All of this evidence points to the conclusion that short selling was nowhere near as common a practice as it is today. Hence, it is unlikely that the results presented above could be driven by short sales constraints imposed by Regulation T alone.

The main empirical tests examining the effect of short sales constraints on the above results rely on the role of disagreement. As short sales arise from investor disagreement about the stock, the stocks that investors disagree about more should be more affected by any constraints on short sales. Hence, a higher initial margin requirement on short sales should affect the beta-expected return relation more for stocks with high disagreement and during times of high aggregate disagreement. I use these predictions in the cross section and the time series to test how strongly the results above are affected by the short sales constraints implied by Regulation T.

First, I divide the sample of stocks into three groups (low, medium, and high) based on measures of disagreement. A natural and commonly used measure is the dispersion in analyst forecasts. However, the I/B/E/S data on analyst forecasts extends back only to 1982, eight years after the Fed Board last changed the margin requirement in Regulation T. Instead, I use idiosyncratic volatility and share turnover as proxies for disagreement and estimate security market lines monthly for high-disagreement and low-disagreement stocks separately. Diether, Malloy, and Scherbina (2002) show that volatility and turnover are very significantly correlated with analyst forecast dispersion. Table X presents the results of the baseline regression for the two categories, as well as the difference between the coefficients and the *p*-value of a Wald test of equality of the coefficients.

If the above-reported relation between the margin requirement and the security market line were driven by the short sales constraint implied by Regulation T, the effect should be stronger for high-disagreement (high-idiosyncratic-volatility or high-turnover) stocks. Empirically, the effect of the margin requirement on the security market line intercept is actually

[34] The data on aggregate short interest are from the NYSE Facts and Figures online database (http://www.nyxdata.com/nysedata/asp/factbook/main.asp).

Table X

Low- and High-Disagreement Stocks

This table presents results of regressing the monthly security market line intercept and slope on the lagged Regulation T minimum initial margin requirement, the contemporaneous market excess return, and controls separately for low- and high-disagreement stocks. Every month stocks are grouped in three categories based on a measure of disagreement. The security market line intercept and slope are constructed by regressing monthly the cross section of excess returns of beta-sorted portfolios on the lagged betas separately within each of the three disagreement categories. Disagreement is measured using idiosyncratic volatility and share turnover. The control variables are defined in Table II. Columns (3) and (6) provide the difference in coefficient between the high- and low-disagreement stocks and the Wald test p-value for the test of the coefficients being equal for the high- and low-disagreement stocks. In columns (1), (2), (4), and (5), the Newey and West (1987) t-statistics with 12 lags of the coefficients are reported in parentheses and the R^2s are adjusted for degrees of freedom. The sample period is 10/1934 to 9/1975 with 492 monthly observations.

Panel A: Intercept

	Idiosyncratic Volatility			Turnover		
	Low (1)	High (2)	H-L (3)	Low (4)	High (5)	H-L (6)
Constant	−0.023	−0.009	0.014	−0.020	−0.005	0.015
	(−2.24)	(−0.57)	(0.48)	(−1.93)	(−0.41)	(0.39)
Margin	0.044	0.027	−0.016	0.039	0.015	−0.024
	(2.75)	(1.11)	(0.61)	(2.47)	(0.75)	(0.37)
Market return	0.099	0.290	0.191	0.161	0.216	0.054
	(0.98)	(2.27)	(0.17)	(1.38)	(1.73)	(0.69)
Controls	Yes	Yes		Yes	Yes	
R^2	0.054	0.033		0.082	0.011	

Panel B: Slope

	Idiosyncratic Volatility			Turnover		
	Low (1)	High (2)	H-L (3)	Low (4)	High (5)	H-L (6)
Constant	0.022	0.022	−0.001	0.018	0.020	0.002
	(1.97)	(1.71)	(0.97)	(1.57)	(2.04)	(0.91)
Margin	−0.043	−0.047	−0.004	−0.036	−0.039	−0.004
	(−2.42)	(−2.31)	(0.89)	(−1.99)	(−2.52)	(0.88)
Market return	0.907	0.788	−0.119	0.825	0.852	0.028
	(8.28)	(6.99)	(0.43)	(6.59)	(7.45)	(0.87)
Controls	Yes	Yes		Yes	Yes	
R^2	0.455	0.314		0.379	0.371	

somewhat weaker, not stronger, for the high-disagreement stocks. In the slope equation, the coefficient on the margin requirement is nearly identical for low-disagreement and high-disagreement stocks. None of the differences between the coefficients is statistically significant. Hence, these results do not support the hypothesis that the main result of this paper would be driven by the short sales constraints imposed by Regulation T.

Table XI

Conditioning on Aggregate Disagreement

This table presents results of regressing the monthly security market line intercept and slope on the lagged Regulation T minimum initial margin requirement, a measure of aggregate disagreement, the interaction of margin and disagreement, the contemporaneous market excess return, and controls. The security market line intercept and slope are constructed by regressing monthly the cross section of excess returns of beta-sorted portfolios on the lagged betas. The measures of disagreement are average idiosyncratic volatility (*IV*), cross-sectional standard deviation of stock returns (*Disp*), aggregate share turnover (*TO*), and aggregate short interest ratio (*Short*). The control variables are defined in Table II. Newey and West (1987) *t*-statistics with 12 lags are reported in parentheses and the R^2s are adjusted for degrees of freedom. The sample period is 10/1934 to 9/1975 with 492 monthly observations.

	Intercept				Slope			
	IV (1)	Disp (2)	TO (3)	Short (4)	IV (5)	Disp (6)	TO (7)	Short (8)
Constant	−0.034	−0.031	−0.033	−0.031	0.036	0.033	0.036	0.034
	(−2.55)	(−2.67)	(−3.18)	(−3.02)	(2.62)	(2.75)	(3.33)	(3.12)
Margin ×	0.003	0.001	0.006	0.002	−0.003	−0.003	−0.008	−0.001
Disagreement	(0.14)	(0.07)	(0.36)	(0.13)	(−0.12)	(−0.13)	(−0.47)	(−0.11)
Margin	0.063	0.059	0.062	0.059	−0.066	−0.062	−0.067	−0.063
	(2.85)	(3.10)	(3.84)	(3.54)	(−2.89)	(−3.17)	(−3.97)	(−3.63)
Disagreement	−0.001	−0.001	−0.003	−0.001	0.001	0.003	0.005	0.001
	(−0.09)	(−0.11)	(−0.33)	(−0.15)	(0.10)	(0.20)	(0.47)	(0.13)
Market return	0.107	0.107	0.107	0.108	0.917	0.917	0.917	0.916
	(1.04)	(1.04)	(1.04)	(1.03)	(8.48)	(8.46)	(8.41)	(8.26)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.055	0.055	0.056	0.055	0.543	0.543	0.543	0.543

Second, I make use of the result of Hong and Sraer (2016) that short sales constraints make the security market line flatter during periods of high aggregate disagreement. If the relation between the Regulation T margin requirement and the security market line is a result of short sales constraints, the relation should be stronger during periods of high aggregate disagreement. Hong and Sraer (2016) measure disagreement using analyst forecasts, which, again, are not available during the 1934 to 1975 sample period. I use four different measures of aggregate disagreement: average idiosyncratic volatility, cross-sectional standard deviation of monthly stock returns, aggregate share turnover, and aggregate short interest ratio. As noted above, Diether, Malloy, and Scherbina (2002) show that volatility and turnover are positively related to analyst forecast dispersion. Cross-sectional return dispersion and the short interest ratio are also natural candidates for disagreement proxies. I augment the baseline regressions (4) and (5), by including an interaction term between the margin requirement and disagreement and present the results in Table XI.

Again, if the reported relation between the Regulation T margin requirement and the security market line were driven by the short sales constraint, this relation should be stronger during periods of high aggregate disagreement. In the results presented in Table XI, this would imply a negative (positive) coefficient on the interaction between the margin and disagreement in the slope (intercept) equation. The empirical results do not support the hypothesis that the short sales constraints significantly drive the relation between the margin requirement and the security market line. Though they have the predicted signs, the interaction coefficients are small and lack statistical significance.

Overall, Tables X and XI provide evidence that the main result presented in this paper is not significantly affected by the short sales constraints imposed by Regulation T. Thus, while the possibility of the short sales constraints having some effect cannot be ruled out, that effect is likely to be rather minor compared to the effect of the leverage constraint.

V. Conclusions

In this paper, I study the effect of leverage constraints on the relation between CAPM betas and expected returns. Using sizeable variation in the minimum initial margin requirement in the U.S. stock market, I show that, during periods of tighter leverage constraints, the empirical security market line has a lower slope and a higher intercept than at times of looser constraints. This result is robust to controlling for additional factors and to using different test assets, portfolio construction rules, and estimation methods. Taken together, the results provide strong empirical evidence in support of the hypothesis that tighter leverage constraints result in a flatter security market line, as predicted by Black (1972) and Frazzini and Pedersen (2014).

Some of the results presented in the paper, however, indicate that leverage constraints cannot fully explain the empirical flatness of the security market line. In many of the observations in Figure 4, the security market line has a negative slope, which is not consistent with the model presented in Section I. In the model, investors are risk-averse and have homogeneous beliefs. Hence, in equilibrium, the price of risk—the slope of the security market line—is positive. As leverage constraints alone cannot explain the empirical patterns, other factors must also play a role. Potential other factors affecting the security market line include the unobservability of the true portfolio of risky assets (Roll (1977)), the estimation errors in betas, the interaction of short sales constraints and heterogeneous expectations (e.g., Hong and Sraer (2016)), investor sentiment (Antoniou, Doukas, and Subrahmanyam (2016)), and investors' preferences for lottery-like stocks (Bali et al. (2017)). Many of these features could also be introduced into a model with leverage constraints to simultaneously capture multiple determinants of the security market line.

Initial submission: April 8, 2015; Accepted: March 25, 2017
Editors: Bruno Biais, Michael R. Roberts, and Kenneth J. Singleton

Appendix

This appendix provides a short derivation of equation (1), the security market line when all investors face the same margin requirement. The derivation presented here is a simplification of the overlapping generations model presented by Frazzini and Pedersen (2014) and is also closely related to the models presented by Ashcraft, Gârleanu, and Pedersen (2010) and Gârleanu and Pedersen (2011).

Securities. There are S risky securities, indexed by s. Security s pays a random periodic dividend $\delta_{s,t}$, has X_s shares outstanding, and trades at the price $P_{s,t}$. The risky payoffs are correlated, with Ω_t representing the covariance matrix of $P_{s,t+1} + \delta_{s,t+1}$. There is also a risk-free security with return r_f.

Investors. Each period I investors, indexed by i, are born with wealth $W_{i,t}$. Investors invest their wealth at birth, and in the next period, they sell their securities to the next generation to finance their final consumption. The portfolio of investor i contains $x_i = (x_{i,1}, \ldots, x_{i,S})$ shares of the risky securities, and the rest of her wealth is invested in the risk-free asset. Investor i has a risk aversion coefficient of γ_i and her expected utility is given by

$$U = x_i' \, \mathrm{E}_t(P_{t+1} + \delta_{t+1}) + (1 + r_f)(W_{i,t} - x_i' \, P_t) - \frac{\gamma_i}{2} \, x_i' \, \Omega_t \, x_i. \tag{A1}$$

All investors face the same margin requirement: an investor can borrow at the risk-free rate, but needs to post an initial margin of m. This results in a constraint on the amount of shares an investor can buy. With wealth $W_{i,t}$, the maximum investment in the risky securities is $W_{i,t}/m$.

Portfolio choice. Given the above, investor i's portfolio choice becomes

$$\begin{aligned} \max \quad & x_i'[\mathrm{E}_t(P_{t+1} + \delta_{t+1}) - (1 + r_f)P_t] - \frac{\gamma_i}{2} \, x_i' \, \Omega_t \, x_i \\ \text{s.t.} \quad & m x_i' \, P_t \leq W_{i,t}. \end{aligned}$$

The Lagrangian of the portfolio choice program is given by

$$\mathfrak{L} = x_i'[\mathrm{E}_t(P_{t+1} + \delta_{t+1}) - (1 + r_f)P_t] - \frac{\gamma_i}{2} \, x_i' \, \Omega_t \, x_i - \psi_i(m x_i' \, P_t - W_{i,t}), \tag{A2}$$

where ψ_i is the shadow price of investor i's margin constraint. The first-order condition for investor i is then

$$\frac{\partial \mathfrak{L}}{\partial x} = \mathrm{E}_t(P_{t+1} + \delta_{t+1}) - (1 + r_f)P_t - \gamma_i \, \Omega_t x_i - \psi_i \, m P_t = 0 \tag{A3}$$

and her optimal portfolio is given by

$$x_i = \frac{1}{\gamma_i} \Omega_t^{-1} [\mathrm{E}_t(P_{t+1} + \delta_{t+1}) - (1 + r_f + \psi_i \, m)P_t]. \tag{A4}$$

Equilibrium. Equilibrium prevails when the market clears and the sum of all investors' positions equals the number of shares outstanding:

$$X = \frac{1}{\gamma} \Omega_t^{-1} [\mathrm{E}_t(P_{t+1} + \delta_{t+1}) - (1 + r_f + \psi \, m)P_t], \tag{A5}$$

where $\gamma = (\sum_{i=1}^{I} \gamma_i^{-1})^{-1}$ is the aggregate risk aversion and $\psi = \sum_{i=1}^{I}(\gamma/\gamma_i)\psi_i$ is the weighted-average shadow price of the margin constraint. Rearranging the market-clearing condition yields the equilibrium prices

$$P_t = \frac{\mathrm{E}_t(P_{t+1} + \delta_{t+1}) - \gamma\, \Omega_t\, X}{1 + r_f + \psi\, m}. \tag{A6}$$

Price of risk. Focusing on a single risky security s and defining its return as $r_{s,t+1} = (P_{s,t+1} + \delta_{s,t+1})/P_{s,t} - 1$, the equilibrium price equation yields the equilibrium expected return as

$$\mathrm{E}_t(r_{s,t+1}) = r_f + \psi\, m + \gamma\, \frac{1}{P_{s,t}}\, \mathbf{1_s}'\, \Omega_t\, X, \tag{A7}$$

where $\mathbf{1_s}$ is a vector with a one on row s and zeros elsewhere. Defining market portfolio M as the value-weighted average of the risky securities gives $\frac{1}{P_{s,t}}\mathbf{1_s}'\Omega_t = \mathrm{cov}_t(r_{s,t+1}, r_{M,t+1})P_t$, which results in

$$\mathrm{E}_t(r_{s,t+1}) = r_f + \psi\, m + \gamma\, \mathrm{cov}_t(r_{s,t+1}, r_{M,t+1})P_t'\, X. \tag{A8}$$

The expected return of the market portfolio is

$$\mathrm{E}_t(r_{M,t+1}) = r_f + \psi\, m + \gamma\, \mathrm{var}_t(r_{M,t+1})P_t'\, X, \tag{A9}$$

which gives

$$\gamma\, P_t'\, X = \frac{\mathrm{E}_t(r_{M,t+1}) - r_f - \psi\, m}{\mathrm{var}_t(r_{M,t+1})}. \tag{A10}$$

Plugging (A10) into (A8) and defining beta in the standard manner as

$$\beta_{s,t} = \frac{\mathrm{cov}_t(r_{s,t+1}, r_{M,t+1})}{\mathrm{var}_t(r_{M,t+1})}, \tag{A11}$$

and the excess return as $r^e = r - r_f$ yields the security market line as

$$\mathrm{E}_t\left(r^e_{s,t+1}\right) = \psi\, m + \beta_{s,t}\left[\mathrm{E}_t\left(r^e_{M,t+1}\right) - \psi\, m\right]. \tag{A12}$$

REFERENCES

Adrian, Tobias, Erkko Etula, and Hyun Song Shin, 2015, Risk appetite and exchange rates, Federal Reserve Bank of New York Staff Report No. 361. Available at https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr361.pdf.

Almazan, Andres, Keith C. Brown, Murray Carlson, and David A. Chapman, 2004, Why constrain your mutual fund manager? *Journal of Financial Economics* 73, 289–321.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Antoniou, Constantinos, John A. Doukas, and Avanidhar Subrahmanyam, 2016, Investor sentiment, beta, and the cost of equity capital, *Management Science* 62, 347–367.

Ashcraft, Adam, Nicolae Gârleanu, and Lasse Heje Pedersen, 2010, Two monetary tools: Interest rates and haircuts, in Daron Acemoglu and Michael Woodford, eds.: *NBER Macroeconomics Annual 2010* (University of Chicago Press, Chicago, IL).

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen, 2013, Value and momentum everywhere, *Journal of Finance* 68, 929–985.

Baker, Malcolm, and Jeffrey Wurgler, 2007, Investor sentiment in the stock market, *Journal of Economic Perspectives* 21, 129–152.

Bali, Turan G., Stephen Brown, Scott Murray, and Yi Tang, 2017, A lottery demand-based explanation of the beta anomaly, *Journal of Financial and Quantitative Analysis* 52, 2369–2397.

Bali, Turan G., Nusret Cakici, and Robert F. Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* 99, 427–446.

Barber, Brad M., Xing Huang, and Terrance Odean, 2016, Which factors matter to investors? Evidence from mutual fund flows, *Review of Financial Studies* 29, 2600–2642.

Berk, Jonathan B., and Jules H. van Binsbergen, 2016, Assessing asset pricing models using revealed preference, *Journal of Financial Economics* 119, 1–23.

Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, *Journal of Business* 45, 444–455.

Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael C. Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger Publishing, NY).

Boguth, Oliver, and Mikhail Simutin, 2018, Leverage constraints and asset prices: Insights from mutual fund risk taking, *Journal of Financial Economics* 127, 325–341.

Brunnermeier, Markus K., Stefan Nagel, and Lasse Heje Pedersen, 2008, Carry trades and currency crashes, in Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, eds.: *NBER Macroeconomics Annual 2008* (University of Chicago Press, Chicago, IL).

Brunnermeier, Markus K., and Lasse Heje Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial Studies* 22, 2201–2238.

Buffa, Andrea M., Dimitri Vayanos, and Paul Woolley, 2014, Asset management contracts and equilibrium prices, Working paper, SSRN.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Christoffersen, Susan E. K., and Mikhail Simutin, 2017, On the demand for high-beta stocks: Evidence from mutual funds, *Journal of Financial Economics* 30, 2596–2620.

Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho, 2005, Money illusion in the stock market: The Modigliani-Cohn hypothesis, *Quarterly Journal of Economics* 120, 639–668.

Cornett, Marcia Millon, Jamie John McNutt, Philip E. Strahan, and Hassan Tehranian, 2011, Liquidity risk management and credit supply in the financial crisis, *Journal of Financial Economics* 101, 297–312.

Diether, Karl B., Christopher J. Malloy, and Anna Scherbina, 2002, Differences of opinion and the cross section of stock returns, *Journal of Finance* 57, 2113–2141.

Doran, James S., Danling Jiang, and David R. Peterson, 2012, Gambling preference and the New Year effect of assets with lottery features, *Review of Finance* 16, 658–731.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Federal Reserve Board, 1976a, *Banking and Monetary Statistics, 1914-1941* (U.S. Government Printing Office, Washington, DC). Available at https://fraser.stlouisfed.org/title/38.

Federal Reserve Board, 1976b, *Banking and Monetary Statistics, 1941-1970* (U.S. Government Printing Office, Washington, DC). Available at https://fraser.stlouisfed.org/title/41.

Ferris, Stephen P., and Don M. Chance, 1988, Margin requirements and stock market volatility, *Economic Letters* 28, 251–254.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Gârleanu, Nicolae, and Lasse Heje Pedersen, 2011. Margin-based asset pricing and deviations from the Law of One Price, *Review of Financial Studies* 24, 1980–2022.

Hardouvelis, Gikas A., 1990, Margin requirements, volatility, and the transitory component of stock prices, *American Economic Review* 80, 736–762.

Hong, Harrison, and David Sraer, 2016, Speculative betas, *Journal of Finance* 71, 2095–2144.

Hsieh, David A., and Merton H. Miller, 1990, Margin regulation and stock market volatility, *Journal of Finance* 45, 3–29.

Huang, Shiyang, Dong Lou, and Christopher Polk, 2016, The booms and busts of beta arbitrage, Working paper, Centre for Economic Policy Research.

Karceski, Jason, 2002, Returns-chasing behavior, mutual funds, and beta's death, *Journal of Financial and Quantitative Analysis* 37, 559–594.

Kupiec, Paul H., 1989, Initial margin requirements and stock return volatility: Another look, *Journal of Financial Services Research* 3, 287–301.

Kupiec, Paul H., 1997, Margin requirements, volatility, and market integrity: What have we learned since the crash? Board of Governors of the Federal Reserve System Research Paper Series, 97-22. Available at https://www.federalreserve.gov/pubs/feds/1997/199722/199722pap.pdf.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.

Meltzer, Allan H., 2003, *A History of the Federal Reserve, Volume 1: 1913-1951* (The University of Chicago Press, Chicago, IL).

Modigliani, Franco, and Richard A. Cohn, 1979, Inflation, rational valuation and the market, *Financial Analysts Journal* 35, 24–44.

Moore, Thomas Gale, 1966, Stock market margin requirements, *Journal of Political Economy* 74, 158–167.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of Financial Economics* 104, 228–250.

Mossin, Jan, 1966, Equilibrium in a capital asset market, *Econometrica* 34, 768–783.

Nagel, Stefan, 2016, The liquidity premium of near-money assets, *Quarterly Journal of Economics* 131, 1927–1971.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Novy-Marx, Robert, 2014, Understanding defensive equity, Working paper, National Bureau of Economic Research.

Officer, Robert R., 1973, The variability of the market factor of the New York Stock Exchange, *Journal of Business* 46, 434–453.

Ranaldo, Angelo, and Paul Söderlind, 2010, Safe haven currencies, *Review of Finance* 14, 385–407.

Rappoport, Peter, and Eugene N. White, 1993, Was there a bubble in the 1929 stock market? *Journal of Economic History* 53, 549–574.

Roll, Richard, 1977, A critique of the asset pricing theory's tests, Part I: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129–176.

Savor, Pavel, and Mungo Wilson, 2014, Asset pricing: A tale of two days, *Journal of Financial Economics* 113, 171–201.

Schnabel, J. A., 1984, Short sales restrictions and the security market line, *Journal of Business Research* 12, 87–96.

Schwert, G. William, 1989, Margin requirements and stock volatility, *Journal of Financial Services Research* 3, 153–164.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Statman, Meir, 1987, How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis* 22, 353–363.

Does Algorithmic Trading Reduce Information Acquisition?

Brian M. Weller
Duke University

I demonstrate an important tension between acquiring information and incorporating it into asset prices. As a salient case, I analyze algorithmic trading (AT), which is typically associated with improved price efficiency. Using a new measure of the information content of prices and a comprehensive panel of 54,879 stock-quarters of Securities and Exchange Commission (SEC) market data, I establish instead that the amount of information in prices decreases by 9% to 13% per standard deviation of AT activity and up to a month before scheduled disclosures. AT thus may reduce price informativeness despite its importance for translating available information into prices. (*JEL* G10, G12, G14)

Received May 21, 2016; editorial decision October 25, 2017 by Editor Itay Goldstein. Authors have furnished an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

Financial markets aggregate information to determine asset values and allocate resources and risk. Traders contribute to price discovery by (1) acquiring new information and (2) incorporating existing information into prices. In this paper, I investigate a striking conflict between these components of price discovery. Mechanisms or traders that improve price efficiency with respect to *existing* information themselves deter information acquisition and diminish price efficiency with respect to *acquirable* information. This trade-off is important because innovations improving market efficiency classically defined (by, e.g., Fama (1970)) may nonetheless distort asset prices and worsen allocations.

I investigate this tension within price discovery by considering the epochal technological changes associated with algorithmic trading. Algorithmic trading, or the use of computer systems to execute trading strategies, has

I thank the Editor, Itay Goldstein; two anomymous referees; Snehal Banerjee, Eric Budish, Anna Cieslak, Nicolas Crouzet, Ian Dew-Becker, Austin Gerig, Andrei Kirilenko, Robert Korajczyk, Jia Li, Matt Lyle, Joshua Mollner, James Naughton, Ryan Riordan, Matt Spiegel, Jiang Wang, and Mao Ye; and seminar participants at Cornell University, Northwestern Kellogg, Duke University, Temple Fox, UBC Sauder, UIUC College of Business, the NBER Microstructure Meeting, the Texas Finance Festival, the Financial Intermediation Research Society Conference, the Frontiers in Finance Conference, the Western Finance Association Conference, and the 7th Conference on Current Topics in Financial Regulation for valuable comments and discussion. Xu Lu provided excellent research assistance. Supplementary data can be found on *The Review of Financial Studies* Web site.
Send correspondence to Brian M. Weller, Duke University, 213 Social Sciences Building, Box 90097, Durham, NC 27708; telephone: (919) 660-1720. E-mail: brian.weller@duke.edu.

doi:10.1093/rfs/hhx137 Advance Access publication December 7, 2017

come to dominate stock, futures, and Treasury markets, among others, in the United States and globally. Algorithmic traders (AT) rapidly incorporate public information into prices (e.g., Zhang 2017; Chakrabarty, Moulton, and Wang 2017), and AT orders are more likely to translate information into prices through heightened quoting efficiency (Hendershott, Jones, and Menkveld 2011) and greater permanent price impacts (Brogaard, Hendershott, and Riordan 2014). However, these studies exclusively speak to one part of the price discovery equation: AT enhances market efficiency with respect to public information *conditional on that information being revealed by other sources*. It is left unasked whether AT-derived improvements in the extent to which prices reflect acquired information ("informational efficiency") come at the expense of discouraging the acquisition of new information. I focus on algorithmic trading precisely because prior work in this literature overlooks this important component of price discovery in evaluating AT's effects on the information content of prices ("price informativeness") and because AT represents a transformative innovation in information processing in markets.[1]

Several potential mechanisms connect algorithmic trading to information acquisition in addition to information incorporation. Algorithmic liquidity provision and smart execution algorithms both reduce costs for typical trades, thereby increasing the potential trading profits due to an informational edge. However, as noted by Stiglitz (2014) and Han, Khapko, and Kyle (2014), these lower average costs may result from a greater ability of algorithmic liquidity providers to screen order flow and avoid adverse selection.[2] Improved screening of informed order flow in turn deters information acquisition by transferring prospective rents away from information acquirers. More directly, "back-running" algorithms may erode information rents and deter use of costly information sources like in Yang and Zhu (2017), or more colorfully, in Michael Lewis' *Flash Boys*.[3]

My core analysis evaluates the net effect of these mechanisms around 54,879 quarterly earnings announcements from January 2012 through September 2016. Evaluating the determinants of the information content of prices poses an empirical challenge because information sets are difficult to observe. I address this challenge by developing a new measure of relative information acquisition,

1 Brunnermeier (2005) distinguishes between "informational efficiency" and "informativeness," and I adopt his terminology to cleanly separate the components of price discovery. Prices may reflect acquired information well (high informational efficiency), but nonetheless summarize a low absolute level of information (low informativeness).

2 Stiglitz (2014) and Han et al. (2014) consider HFT rather than AT, but their screening and stale-quote avoidance mechanisms are not specific to AT. For example, Foucault, Röell, and Sandås (2003) explain how market makers use software to monitor different types of news to avoid being "picked off" by fast traders. This paper predates HFT, and the liquidity providers described are human-algorithm hybrids rather than HFT.

3 Front- and back-running strategies fall under the umbrella of "order anticipation" strategies and are described in the Securities and Exchange Commission Concept Release on Equity Market Structure U.S. Securities and Exchange Commission (2010). Order anticipators attempt to trade ahead of large and/or informed traders to benefit from near-term price movements in the direction of trade.

the price jump ratio. The price jump ratio divides the return at the time of information's public disclosure to the total return over the pre-announcement period. Large impacts at announcement indicate that information is not discovered until publicly revealed, and this feature is reflected in a high price jump ratio. Unlike most alternative measures, such as the absolute cumulative abnormal return, the price jump ratio captures how much information enters prices early *relative to how much is potentially acquirable*.

I combine the price jump ratio measure with new data from the United States. Securities and Exchange Commission's Market Information Data Analytics System (MIDAS) to analyze the effects of algorithmic trading on information acquisition. Developed in response to analytics deficiencies revealed by the 2010 Flash Crash, MIDAS captures all public messages across all major stock exchanges for most U.S. common stocks and exchange-traded products since January 2012.[4]

The MIDAS data significantly improve identification of AT activity relative to alternative data sources. MIDAS facilitates construction of algorithmic trading proxies including odd lot volume ratios (or the fraction of volume associated with abnormally small trades), trade-to-order volume ratios, cancel-to-trade count ratios, and average trade sizes. With the (recent) exception of odd lot ratios and average trade sizes, these measures cannot be constructed using the standard NYSE Trades and Quotes database, and most alternative sources provide data for a short time series, a small sample of securities, or a single, potentially unrepresentative exchange.[5]

The price jump ratio measure and MIDAS data enable tests of the relationship between algorithmic trading and information acquisition for the first time. The baseline panel regressions of the price jump ratio on algorithmic trading proxies establish a strong negative association between algorithmic trading and information acquisition around earnings announcements. However, these results are potentially biased by causality running in the opposing direction. Algorithmic traders might sort into stocks with less information acquisition, or they may detect informed traders in the market and scale back their activities.

I use lagged log stock prices as a simple instrument to work around this concern. Lagged stock prices shift the incentives to use trading algorithms. The "sub-penny" rule (SEC Rule 612) imposes a minimum tick size of one cent for securities covered by Reg NMS, and this minimum price increment translates into variation in the fineness of the price grid as a function of price. High price stocks have a relatively fine price mesh, which favors algorithms

4 Messages include transactions; quote additions, modifications, and cancellations; order imbalances; and order-book level updates. MIDAS maps these messages—numbering in the hundreds of billions—into publicly available aggregates by stock, date, and exchange.

5 The New York Stock Exchange Trade and Quotes (TAQ) data only deliver order book changes at the best bid and offer rather than for the entire order book; NASDAQ TotalView-ITCH, NYSE OpenBook, and other feeds cover single exchanges in isolation; and proprietary resources, such as HFT-identified NASDAQ data, are typically quite limited in scope or in time.

over humans for continually updating limit orders for liquidity provision or monitoring for stale limit orders in liquidity taking. When the price mesh is relatively sparse, for example, 20 basis points for a $5 stock as compared to 2 basis points for a $50 stock, the underlying value of a security can vary more without requiring traders to update their order strategies, all else equal. High prices hence favor highly attentive algorithms over human traders. At the same time, lagged stock prices focus the empirical analysis on the ex ante effects of algorithmic trading on information acquisition rather than the endogenous response of algorithmic activity to informed traders. Conditional on other controls, such as market capitalization and institutional ownership, differences between high-priced stocks with few shares outstanding versus low-priced stocks with many shares outstanding are unlikely to affect the incentives to acquire information outside of the AT channel.[6]

Armed with this instrument, I find that algorithmic trading powerfully undermines pre-announcement information acquisition: a one standard deviation increase in algorithmic trading decreases information acquisition before earnings announcements by 9% to 13%.[7] These estimates hold across market capitalization and time strata, and for both individual securities and portfolios. Consistent with strategic trading models with informed traders, the "information gap" between high- and low-AT stocks grows monotonically as earnings announcements approach. Importantly, AT measurably discourages information acquisition up to *a month* in advance, by contrast with post-announcement price efficiency gains on horizons of (milli-)seconds to minutes typically considered in the literature. To the extent that they identify similar quantities, other measures of information in prices considered in robustness analyses arrive at the same conclusions.

Algorithmic traders impede information acquisition with potentially significant welfare consequences. First, reduced information acquisition facilitates trade at imprecise prices, in the sense that average absolute pricing errors are larger than they would be under a counterfactual low-AT regime. The welfare reduction associated with negative AT effects on price discovery may be quite severe if price uncertainty aggregates to systematic risks (like in O'Hara 2003). Among other channels, less informative prices may undermine efficient capital allocation through noisy costs of capital or budget constraints (Merton 1974; Baker, Stein, and Wurgler 2003) or reduce managerial learning from prices (Baumol 1965; Dow and Gorton 1997; Chen, Goldstein, and Jiang 2007).

Although such welfare effects may be muted around earnings announcements, the technologies and strategies relating AT to reduced information

6 Section 5.2 discusses and rules out potential contamination by other determinants of nominal stock price levels.

7 Incidentally, slightly greater magnitudes of the OLS coefficients offers fresh evidence for the endogenous withdrawal of algorithmic traders when confronted by informed traders, as theorized by Han et al. (2014) and Baldauf and Mollner (2017).

acquisition should be active in other settings for which the information content of prices cannot be readily estimated. Indeed, an analysis of an alternative measure of information in prices around earnings announcements, product releases, executive appointments and resignations, and merger and acquisition (M&A) events suggests that my findings extend to a wide range of economically important firm events.

This paper also contributes to the debate on the role of technology in financial markets. The same technological advances that improve price efficiency reduce the informativeness of prices in the medium run. This finding complements recent work by Bai, Philippon, and Savov (2016) relating the growth of the financial sector to stagnant price informativeness at horizons of a year or less.

1. Related Literature

1.1 Algorithmic trading and price discovery

Empirical work on algorithmic trading focuses primarily on whether algorithmic trading affects trading costs, market resilience, or price discovery.[8] A lively literature suggests that algorithmic trading contributes to faster incorporation of information into prices conditional on information acquisition having taken place. Zhang (2017) finds that high-frequency traders (HFT) are particularly skilled in incorporating hard information, and Chakrabarty, Moulton, and Wang (2017) echo this result in the context of earnings announcements temporarily overlooked by inattentive human traders. Carrion (2013), Hirschey (2013), and Brogaard, Hendershott, and Riordan (2014) demonstrate that HFT trades forecast price changes several seconds ahead and are more likely to have permanent price effects. Although these papers focus on the HFT subset of algorithmic trading, the consensus that HFT improve incorporation of acquired signals carries through to algorithmic trading generally.

By contrast, there is comparatively little work considering whether algorithmic trading encourages information acquisition by human and computerized agents. Grossman and Stiglitz (1980) suggest that changes to the profitability of acquiring information should change the amount of information acquired, and Stiglitz (2014) questions whether high-frequency traders dissuade information acquirers by reducing information rents. More positively, Foucault, Hombert, and Roşu (2016) theoretically exposit a mechanism by which HFT directly contribute to information gathering by trading on Brownian news surprises.

Baldauf and Mollner (2017) develop a model of the market quality trade-off between tight bid-ask spreads and effective price discovery. Their paper

[8] Several papers focus on high-frequency trading as a specific category of algorithmic trading. High-frequency trading is distinguished by ultra-low latencies and extensive automation of decision-making and execution.

is close in spirit to this work. Rather than balancing the desirable features of low transactions costs and informative prices, I focus instead on the conflict between information acquisition and information incorporation components of price discovery. Brunnermeier (2005) describes a similar potential conflict between informational efficiency and price informativeness in the context of theoretical model of information leakage, with an emphasis on the long-run negative effects of information leakage on both quantities rather than on a trade-off between them.

1.2 Measuring information acquisition

A rich literature on the role of traders and trade in incorporating information into prices precedes my work. To isolate my contribution to this literature, I briefly discuss selected related empirical measures. Because (absolute) cumulative abnormal returns (ACAR) are central to the construction of my price jump ratio measure, I revisit ACAR separately in Section 3.

The price jump ratio measure can also be understood as a stock-level variant of the intraperiod timeliness (IPT) measure of McNichols (1984), Freeman (1987), and Alford et al. (1993). IPT compares the performance of two zero-cost "perfect-foresight" portfolios comprised of stocks with different characteristics, such as the level of participation by algorithmic traders. Higher perfect-foresight returns earlier in the pre-event period indicate more pre-event information acquisition for stocks with that characteristic. Because the IPT measure compares two portfolios rather than many assets, it is not suitable for examining several characteristics simultaneously or for controlling for covariates when examining a single characteristic. Nevertheless, I adopt the IPT approach without controls in Section 6.1 to confirm the robustness of my results to the use of this portfolio-level measure with established properties.

The jump ratios of Morse (1981) and Meulbroek (1992) are the most closely related to my price jump ratio measure, and they are applied in a similar context of studying price changes prior to earnings announcements.[9] Both measures quantify the extent to which announcement returns are especially large in absolute value relative to a benchmark return, with the distinction that Meulbroek's variant enables analysis of an additional determinant of abnormal returns (such as illegal insider trading in her study). Like the price jump ratio, these measures differ from the other informational measures in part because they gauge information acquisition without hinging on a particular model of how information is incorporated into prices. This distinction is important because tests of potential determinants of information acquisition using other measures are in fact joint tests with an auxiliary hypothesis that informed traders behave in a particular way.

[9] Heflin, Subramanyam, and Zhang (2003) use similar measures to evaluate the effects of SEC Regulation Fair Disclosure ("Reg FD").

Recent work by Collin-Dufresne and Fos (2015) emphases the importance of such "model-free" measures in finding that measures of informed trading do not reflect informed trades. For example, if informed agents signal their intentions suboptimally via their order behavior, liquidity providers should respond by adjusting their quotes rather than suffering losses in trading. Trade-based measures such as trade informativeness (Hasbrouck 1991a, 1991b) or the weighted price contribution (Barclay and Warner 1993) thus may be quite low even as prices increasingly reflect the knowledge of an informed trader. As I detail in Section 3, the price jump ratio and Morse's and Meulbroek's measures sidestep these issues by assuming only that some set of market participants moves prices in response to information acquired before its public disclosure.

My price jump ratio departs from Morse's and Meulbroek's measures in two important respects. First, I extend application of jump ratios to a panel context by constructing a variant that is estimated separately for each stock-quarter event. This simple extension greatly expands jump ratios' empirical applicability. Second, my measure normalizes by a different benchmark, the total abnormal return associated with each announcement event. This benchmark facilitates analysis of the determinants of the information content of prices *relative to acquirable information*, the primary aim of my study.

2. Hypothesis Development and Empirical Setting

Several potential mechanisms link algorithmic trading to information acquisition, and the net effect of many different algorithms acting in concert is uncertain a priori. First, algorithmic traders may be better than human traders at extracting price-relevant information from public signals. For example, Brogaard et al. (2014) find that liquidity-taking algorithmic traders are more responsive to information in order books and public announcements. Foucault et al. (2016) generate a similar theoretical prediction that fast liquidity takers may improve pre-announcement price discovery via superior signal processing. More generally, algorithmic "analysts" may augment or supplant their human counterparts, potentially skewing the type or quantity of information acquired (Zhang 2017).

Algorithmic traders also may encourage or discourage human traders from acquiring information. From a Grossman and Stiglitz (1980) perspective in which information rents motivate information acquisition, lower trading costs associated with algorithmic trading—either from "smart" algorithmic execution or through lower bid-ask spreads—may encourage the collection and incorporation of information. However, this mechanism may be reversed if the decline in spreads results from algorithmic traders' improved ability to avoid being adversely selected (Han et al. 2014; Baldauf and Mollner 2017). Likewise, if aggressive algorithmic traders anticipate or "piggyback" on informed orders, as Stiglitz (2014) and Yang and Zhu (2017) suggest, profits accruing to information acquisition decline, and AT decreases equilibrium information acquisition.[10]

These competing theories motivate my main empirical tests of how and when algorithmic trading contributes to information acquisition, and I investigate these relations in the context of quarterly earnings announcements. Earnings announcements provide a near-ideal setting for this investigation for three reasons. First, earnings announcements are among the recurring firm events with the greatest average effects on stock prices. The importance of earnings announcement news relative to day-to-day price movements boosts the signal-to-noise ratio of my measure of relative information acquisition. Second, earnings announcements are scheduled, public events. Kyle (1985) and other models of informed trading offer crisp predictions for how prices should evolve in this setting. Third, quarterly earnings announcements are ubiquitous, which contributes to rich cross sections and minimal selection effects over which information events are observed.

Notwithstanding these advantages, the earnings announcements setting has two notable drawbacks related to its external validity. For one, market participants may be able to mitigate the effects of changes in pre-announcement information acquisition. For example, firms may avoid conditioning their decisions on market prices shortly before announcements in favor of waiting until after public disclosures. Consequently, the economic cost of delayed information incorporation around earnings announcements is likely to be muted relative to other events I cannot analyze. For this reason, I only take limited steps toward assessing the potential welfare impact of AT on information acquisition.

Second, I can only indirectly assess whether the effects of algorithmic trading around earnings announcements apply across a wider set of events. However, given the prominence of earnings as a focus of analyst and market attention, earnings announcements (1) are important in their own right and (2) represent a best case for information acquisition by market participants. Moreover, the mechanisms linking algorithmic trading and information acquisition apply equally well for the informational events that I cannot observe as a researcher.

3. The Price Jump Ratio as a Measure of Information Acquisition

3.1 Defining the price jump ratio

The absolute cumulative abnormal return (ACAR) of Fama et al. (1969) and Ball and Brown (1968) is a standard measure of the incorporation of information into prices. This measure constructs pre-event price drift net of a predicted

[10] Harris (2013) emphasizes that this channel concerns algorithmic trading generally rather than high-frequency trading specifically: “the successful implementation of [order anticipation] strategy depends less on low-latency communications than on high-quality pattern-recognition algorithms.”

returns from a factor model, for example,

$$CAR_{it}^{(k_1,k_2)} = \sum_{t=k_1}^{k_2} \left(r_{it} - \alpha_i - \sum_{m=1}^{M} \beta_{im} r_{mt} \right) = \sum_{t=k_1}^{k_2} \epsilon_{it}, \tag{1}$$

where r_{it} is the log return of stock i on date t and α_i and β_i are estimated from a M-factor model of realized returns. The cumulative abnormal return (CAR) cumulates the abnormal return from dates k_1 to k_2 around the announcement date T, and the ACAR is the absolute value of this number.

The ACAR answers the question of how much pre-announcement information enters prices prior to an earnings announcement *without consideration for how much information might be available to acquire*. Rather than using the ACAR, I proceed along the lines of Meulbroek (1992) and normalize the CAR by a measure of total announcement-related variation. Specifically, I use the ratio of post-announcement price variation to total variation before and including the earnings announcement:

$$jump_{it}^{(a,b)} = \frac{CAR_{it}^{(T-1,T+b)}}{CAR_{it}^{(T-a,T+b)}}, \tag{2}$$

with $a > 1$ to capture pre-announcement variation and $b \geq 0$ to capture post-announcement drift.

I use a normalized price jump measure for two reasons. First, the normalized measure naturally accounts for differences in the expected magnitude of abnormal returns among stocks and over time. For example, panel analysis using the ACAR measure implicitly assign large weights to stocks in the "dusty corners" of the market because market capitalization and variance of earnings announcement impact are strongly negatively correlated.[11] Second, the price jump ratio has a different economic interpretation from the ACAR. The price jump ratio quantifies the *share* of information acquired and incorporated into prices pre-announcement. If acquirable return information is proportional to the total realized return—which holds, for example, when market participants can learn all price-relevant information before the announcement—this ratio directly measures the fraction of acquirable information incorporated into prices before its public disclosure.

To build intuition for the price jump ratio measure, consider the price path of a hypothetical stock with and without information acquisition before an earnings announcement or other scheduled public news event. Without an informed trader, the price may be efficient with respect to all information acquired by non-insiders, but no market participant is aware of the information content of the impending disclosure event. By contrast, in the presence of

[11] Chakrabarty, Moulton, and Wang (2017) note precisely this feature of ACAR in a concurrent study of high-frequency trading and investor attention after public announcements.

an informed trader, the price drifts toward the post-announcement asset value on account of her order submission and trading strategy: strategic informed trading models have a common implication of smooth convergence of prices to post-announcement values *before* information is publicly revealed (e.g., Kyle 1985; Back 1992). The difference in price paths with and without informed traders represents an "information gap," or the extent to which prices better approximate the post-announcement asset value when informed traders are active. Importantly, several measures of price efficiency like pricing error variance (Hasbrouck 1993), serial autocorrelation in returns, and variance ratios speak only to price efficiency with respect to acquired information rather than to price informativeness and the information gap.

The price jump ratio by contrast is the empirical counterpart to the (unobservable) information gap. A high price jump ratio corresponds to a large announcement-date jump relative to pre-announcement drift, whereas a low price jump ratio corresponds to a small announcement-date jump. Aggressive informed trading drives the price jump ratio toward 0, and the absence of informed trading precipitates a price jump ratio close to 1. Higher values of the price jump ratio thus represent less information in prices relative to the post-announcement information set.

Strictly speaking, the price jump ratio measures the pre-announcement information content of prices rather than information acquisition. Market participants could acquire information but not push stock prices appreciably toward their post-announcement values. However, such trading behavior around scheduled announcements is dominated by an alternative trading strategy in which informed traders enforce near or complete convergence of prices to the post-announcement values. Otherwise, the residual discrepancy of prices an instant before the announcement date represents foregone profits from trading against uninformed resting orders. For this reason, the optimal interim price path may vary with liquidity conditions, but the anticipated post-announcement value should remain an informed trader's target. Moreover, from a welfare perspective, reduced information acquisition and more "smart trading" have similar distorting effects for the allocation of risk and resources because prices remain uninformative with respect to acquirable information.

3.2 Implementing the price jump ratio measure

I estimate abnormal returns relative to a Fama and French (1992) three-factor model using daily returns over a 365-calendar day window ending 90 days before the earnings announcement. Observations with estimation windows with fewer than 63 valid preceding trading days (one calendar quarter) are dropped. I select a 21-day pre-announcement window ($a=21$ in Equation (2)) to balance resolution on earnings announcement price effects against the possibility of earnings-related information entering earlier than the period considered, although results are robust to this choice.[12] I define the cumulative variation associated with the earnings announcement to include an additional two trading days after the announcement ($b=2$) because prices may exhibit post-earnings announcement drift.

In practice noise in informed traders' signals on the ex post value of an asset moves the realized price jump ratio away from the 0 or 1 outcomes delivered by a standard Kyle model. This noise introduces potentially sizable deviations in the price jump ratio's numerator and denominator. For this reason an individual price jump ratio realization reveals little about information acquisition around any particular announcement. However, relating conditional (truncated) means or medians to covariates reveals the determinants of information acquisition, and this is the path my analysis takes.

Unlike the unnormalized (A)CAR, the price jump ratio has an undefined conditional mean because its denominator may be close to zero. Throughout the empirical analysis, I resolve this problem by dropping observations with small values of the price jump ratio denominator, namely nonevents from the market's perspective. I define the cutoff for retaining an announcement event in terms of return volatility over the preceding month, estimated simply as the square root of the daily return variance. Each event retained satisfies

$$\left|CAR_{it}^{(T-21,T+2)}\right| > \sqrt{24}\hat{\sigma}_{it}. \tag{3}$$

Events exceeding the value on the right represent large announcement period returns (relative to scaled daily volatility) indicative of material earnings announcement information. This filter is strict, and only 45.5% of observations survive. Appendix A plots the distribution of the price jump ratio and compares stock-quarters that do and do not survive this volatility cutoff. Overall the observable characteristics of included and excluded observations are very much alike. Likewise, Appendix B verifies that results are not driven by differences in market capitalization or relative importance of earnings news $\left|CAR_{it}^{(T-21,T+2)}\right|/\hat{\sigma}_{it}$ among stocks with different levels of AT.

I choose this approach for economic and statistical reasons. Economically, near-zero price impact events have limited potential informational distortions because prices with or without informed traders are similar throughout the pre-announcement period. From a statistical perspective, these small-denominator events also have very low signal-to-noise ratios. All coefficient estimates for this restricted sample nonetheless are virtually identical to those estimated by median regression or weighted least squares on the full sample of stock-quarter events. The Online Appendix tabulates results and provides additional discussion.

12 I also conduct a separate analysis in which I control for price changes in the more distant past. Although the main results are effectively unchanged, AT appear to counteract behavioral or strategic disclosure responses to negative news. I present these results with additional discussion in the Online Appendix.

4. Data

This paper combines several data sources for the main analysis: the Center for Research in Securities Prices (CRSP); NYSE Trades and Quotes (TAQ); the Thomson Reuters Institutional Brokers' Estimate System (I/B/E/S) and SEC Form 13F Institutional Holdings Database; and the SEC Market Information Data Analytics System (MIDAS). CRSP provides abnormal and cumulative abnormal returns[13] for each earnings event as well as daily prices, market capitalization, and return volatility. Monthly TAQ (2012–2014) and Daily TAQ (2015–2016) contribute quoted spreads via derived national best bid and offer (NBBO) tables. I/B/E/S provides quarterly earnings announcement dates and times, as well as the number of analyst estimates for each earnings announcement, and the 13F database delivers institutional holdings information. SEC MIDAS enables construction of algorithmic trading proxies and order volume shares by exchange, stock, and date.

In the months following the May 6, 2010, Flash Crash, the Securities and Exchange Commission sought to enhance its ability to monitor markets in real time and rapidly reconstruct market past events. The Market Information Data Analytics System (MIDAS), launched in January 2013, provides the SEC with order book information from all major U.S. stock exchanges with microsecond timestamps, as well as the capacity to process these data quickly. To the best of my knowledge, this paper is the first to use SEC MIDAS data for academic research.

The SEC provides limited summary statistics by security and exchange for public use. The summary metrics include lit trade and order volume,[14] hidden volume, odd lot volume, and counts of trades and cancellations (full or partial), as well as several quantities derived from these metrics. MIDAS also provides distributions of quote survival lifetimes by market capitalization terciles for stocks and exchange traded products.

MIDAS collects order data across all major U.S. stock exchanges to present a comprehensive view of lit market activity. By contrast with the now-standard market data found in TAQ, the MIDAS data incorporates quotes and cancellations from all levels of the order book. It also reports odd lot trading, which only recently entered the TAQ data. Most relevant for this paper, public availability of order book summary data for the construction of algorithmic trading proxies at the stock-day frequency is an important step forward for avoiding potential sample biases endemic to the specialized or proprietary data typically used by researchers.[15]

---

13 Factor return information is provided by Kenneth French through his website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

14 "Lit" orders are visible to other market participants before a trade, and lit trades are executions resulting from lit orders. "Hidden" orders are not lit, and hidden volume references shares exchanged using hidden orders.

15 U.S. Securities and Exchange Commission (2014) surveys data sources used in the analysis of algorithmic and high-frequency trading and their respective limitations.

The main analysis includes all CRSP common stocks (CRSP share code 10 or 11) matched in the I/B/E/S and MIDAS databases.[16] Stock events include all quarterly earnings announcements from January 1, 2012, through September 30, 2016. Announcements recorded by I/B/E/S as occurring after market close are assigned to the next trading day, as returns are computed from market closing prices. I exclude observations for which earnings announcements are less than 45 days or more than 150 days from the time of the previous earnings announcement. The resulting sample has 3,839 unique securities and 54,879 security-quarter observations.

For each stock-earnings announcement and date pair it, I construct rolling 21-day averages of price, market capitalization, and quoted spreads over the interval $[T-42, T-22]$ (measured in days relative to the earnings announcement date T). I then take the log of price and market capitalization values. Log return volatility is computed as the log of the standard deviation of daily returns over the same 21-day horizon.

Daily quoted spread values from TAQ are "time-in-force" or duration-weighted averages from 9:35 a.m. to 4:00 p.m., where the first five minutes of trading are dropped to avoid incorporating abnormal values associated with market open. Bid-ask spreads use daily TAQ when available (2015–2016) and monthly TAQ for 2012–2014. Daily TAQ NBBOs are coarsened to seconds rather than milliseconds to equate observation frequencies across datasets. Quoted spreads less than zero or greater than 25% of the concurrent price are excluded from rolling averages over the same $[T-42, T-22]$ interval.

I also include two controls for institutional attention and activity. Analyst coverage for each observation is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio is the total number of shares held by 13F filing institutions at the end of the preceding calendar quarter divided by the total number of shares outstanding at that time.

4.1 Construction of algorithmic trading proxies

I construct four algorithmic trading proxies using the MIDAS data: the odd lot volume ratio, or the total volume executed in quantities smaller than 100 shares divided by total volume traded;[17] the trade-to-order volume ratio, or the total volume traded divided by the total volume across all orders placed; the cancel-to-trade ratio, or the number of full or partial cancellations divided by the number of trades; and the average trade size, or the number of shares traded divided by the number of trades. Existing literature suggests that higher odd lot and cancel-to-trade ratios indicate more algorithmic trading, and higher

[16] To facilitate matching, I use the WRDS IBES to CRSP linking table and retain observations with matching 6-digit CUSIPs and date ranges (99.3% of the data).

[17] Three tickers have round lot sizes different from 100: BH has a round lot size of 10, and BRK.A and SEB have a round lot size of 1.

trade-to-order volume ratios and average trade sizes indicate less algorithmic trading.

I follow MIDAS's prescription for constructing the odd lot, trade-to-order, and cancel-to-trade ratios. Of the 12 stock exchange feeds captured by MIDAS,[18] NYSE and NYSE MKT are excluded from the odd lot ratio, the cancel-to-trade ratio, and average trade size measures because the level-book reporting method of these two exchanges does not allow for comparable accounting of trades with the order-based feeds used by the other 10 exchanges.[19] For each stock, I average each algorithmic trading proxy across dates $[T-21, T-1]$ and take the log of the result to eliminate significant right skew in the averaged values. I drop observations with ratios less than the 1st percentile or greater than the 99th percentile of AT proxies to guard against potential reporting errors in the MIDAS data.[20]

Table 1 presents summary statistics for AT proxies and controls. All AT proxies exhibit significant dispersion across stocks. The top portion of Panel A quantifies this distributional information for the pooled sample, and the bottom portion of panel A reports similar statistics for the set of controls. Panel B provides pairwise correlations among key independent variables (top) and controls (bottom). Of note on top are the strong correlations among algorithmic trading measures. As previous studies suggest, odd lot and cancel-to-trade ratios covary positively with each other and covary negatively with trade-to-order volume ratios and average trade sizes. These correlations are nonetheless far from one, and the measures appear to embed separate information content.

5. Main Results

5.1 Algorithmic trading and the information content of prices

My main empirical analysis investigates whether algorithmic trading is associated with differences in pre-announcement information acquisition around quarterly earnings announcements. I start with a simple panel regression relating price jump ratios to algorithmic trading proxies,

$$jump_{it}^{(21)} = \alpha + \beta x_{it} + \gamma \times controls_{it} + \epsilon_{it}, \tag{4}$$

where $jump^{(21)}$ is shorthand for $jump^{(21,2)}$ and x_{it} stands in for the odd lot ratio, the trade-to-order volume ratio, the cancel-to-trade ratio, or the average trade

---

[18] I exclude IEX because it starts trading all stocks as an exchange in September 2016, the last month of my sample.

[19] NYSE and NYSE MKT report the trade size of the initiating order, whereas the other exchanges separate trades by initiating and contra orders. For this reason the number of trades and trade size distributions for NYSE stocks are not comparable to their counterparts on other exchanges. Notwithstanding these differences, the pairwise correlations between log odd lot ratios, cancel-to-trade ratios, and average trade sizes for NYSE/NYSE MKT and all other exchanges are quite high at 88.0%, 55.6%, and 76.7%, respectively, and using only NYSE and NYSE MKT variables on the left-hand side delivers similar results.

[20] Additional MIDAS details and discussion of exchange exclusions are provided on the MIDAS website at http://www.sec.gov/marketstructure/mar_methodology.html.

Table 1
Summary statistics for algorithmic trading proxies and controls

A. Summaries by stock-event observations

AT proxies	Odd lots	Trades/Orders	Cancels/Trades	Trade size
Mean	−2.43	−3.76	3.31	4.71
SD	0.80	0.63	0.58	0.45
10%	−3.59	−4.61	2.62	4.31
Median	−2.26	−3.72	3.25	4.59
90%	−1.58	−3.00	4.09	5.33
N	53,796	53,844	53,844	54,879

Controls	Market cap.	Price	Ret. vol.	Quoted spr.	#Analysts	*IOR*
Mean	20.71	2.90	−3.96	0.50	1.71	0.57
SD	1.89	1.21	0.58	0.84	0.96	0.28
10%	18.28	1.22	−4.64	0.05	0.00	0.00
Median	20.65	3.06	−4.00	0.23	1.79	0.64
90%	23.20	4.27	−3.22	1.15	2.94	0.88
N	54,879	54,879	54,879	54,595	54,879	54,439

B. Pairwise correlations by stock-event observations

AT proxies	Odd lots	Trades/Orders	Cancels/Trades
Trade-to-order ratio	−0.56		
Cancel-to-trade ratio	0.42	−0.83	
Trade size	−0.94	0.52	−0.34

Controls	Market cap.	Price	Ret. vol.	Quoted spr.	#Analysts
Price	0.77				
Ret. vol.	−0.50	−0.56			
Quoted spr.	−0.60	−0.48	0.33		
#Analysts	0.74	0.51	−0.25	−0.53	
IOR	0.42	0.44	−0.24	−0.40	0.43

This table presents descriptive statistics (top panel) and correlations (bottom panel) among key explanatory and control variables. The first table in each panel summarizes algorithmic trading proxies (all in logs). The second table in each panel presents corresponding summaries for dollar market capitalization, share price, and return volatility (all in logs) from the CRSP data. Additional summaries include the median end-of-minute quoted bid-ask spreads in percent from the TAQ NBBO files (one-second version); the number of analysts as the log of the maximum number of reporting analysts for each stock-event pair in Thomson Reuters I/B/E/S; and the institutional ownership ratio (*IOR*) as the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. All proxy variables are rolling 21-day mean values from $[T-21, T-1]$, and control variables are lagged an additional 21 days prior to the announcement date ($[T-42, T-22]$).

size for stock i at date t. As Section 4 describes, these proxies alternate between strongly increasing in algorithmic trading activity (odd lot and cancel-to-trade ratios) and strongly decreasing in algorithmic trading activity (trade-to-order volume ratio and average trade size). Throughout, standard errors are two-way clustered by stock and month to account for correlated errors arising from differences in firm characteristics, such as disclosure policies, or timing within the sample period.

Table 2 presents results from estimation of Equation (4). The first specification for each algorithmic trading measure includes the AT proxy and lagged log market capitalization. I include lagged log market capitalization as a control in all specifications because market capitalization is among the strongest covariates with AT, and it also has the potential to influence the amount of information that enters prices pre-announcement through channels other than AT.

Table 2
Determinants of announcement price impact

	x = Odd lot ratio			x = Trade-to-order ratio			x = Cancel-to-trade ratio			x = Avg. trade size		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
x	0.0741***	0.0898***	0.144***	−0.0725***	−0.0889***	−0.142***	0.0718***	0.113***	0.144***	−0.152***	−0.150***	−0.263***
	(0.00683)	(0.00937)	(0.0111)	(0.00863)	(0.0103)	(0.0107)	(0.00858)	(0.00988)	(0.0113)	(0.0122)	(0.0176)	(0.0248)
Market cap.	0.00244	−0.0183***	−0.0478***	0.00783***	−0.0280***	−0.0131	0.0135***	−0.0210***	−0.00424	0.000119	−0.0220***	−0.0423***
	(0.00318)	(0.00542)	(0.0121)	(0.00290)	(0.00572)	(0.0118)	(0.00314)	(0.00556)	(0.0119)	(0.00318)	(0.00529)	(0.0124)
Price		−0.0290***	X		−0.00510	X		−0.0105	X		−0.0151*	X
		(0.00867)			(0.00876)			(0.00835)			(0.00857)	
Ret. vol.		−0.0163	−0.000812		−0.0106	0.00528		−0.00193	0.00933		−0.0194	−0.00387
		(0.0124)	(0.0123)		(0.0117)	(0.0122)		(0.0115)	(0.0120)		(0.0120)	(0.0120)
Quoted spr.		−0.0324***	−0.0169		−0.0572***	−0.0363***		−0.0713***	−0.0372***		−0.0222***	−0.00209
		(0.00745)	(0.0123)		(0.00779)	(0.0120)		(0.00871)	(0.0129)		(0.00725)	(0.0115)
#Analysts		0.0521***	0.0326**		0.0546***	0.0361***		0.0564***	0.0334**		0.0505***	0.0305**
		(0.00754)	(0.0136)		(0.00742)	(0.0133)		(0.00746)	(0.0133)		(0.00728)	(0.0132)
IOR		0.123***	0.0239		0.123***	0.0421		0.139***	0.0550		0.106***	0.00781
		(0.0231)	(0.0358)		(0.0222)	(0.0349)		(0.0222)	(0.0347)		(0.0228)	(0.0338)
Constant	0.599***	X	X	0.0403	X	X	−0.0426	X	X	1.179***	X	X
	(0.0729)			(0.0692)			(0.0736)			(0.107)		
Month FEs	No	Yes	Yes	No	Yes	Yes	No	No	Yes	No	No	Yes
Stock FEs	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
R^2	0.0132	0.0337	0.0174	0.00904	0.0330	0.0198	0.00762	0.0352	0.0190	0.0157	0.0331	0.0158
N	24,107	23,814	23,517	24,068	23,800	23,503	24,120	23,842	23,539	24,512	24,201	23,910

$^{*}p<.10$, $^{**}p<.05$, $^{***}p<.01$

This table presents results from a regression of price jump ratios on a set of algorithmic trading proxies:

$$jump_{it}^{(21)} = \alpha + \beta x_{it} + \gamma \times controls_{it} + \epsilon_{it}.$$

For each stock i and quarterly earnings announcement t from January 2012 through September 2016, the price jump ratio ($jump_{it}^{(21)}$) is measured as the ratio of the announcement response divided by the total variation in the pre- and post-announcement period: $CAR_{it}^{(T-1,T+2)}/CAR_{it}^{(T-21,T+2)}$. Cumulative abnormal returns (in logs) are net of Fama and French (1992) three-factor implied returns over the same interval. Observations with cumulative net price impact not exceeding a minimum multiple of prior return volatility (described in the main text) are dropped. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42, T-22]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio (*IOR*) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading proxies (x_{it}) derived from SEC MIDAS data are described in the main text. All standard errors are clustered by security and month and are reported in parentheses. The log price is dropped from stock fixed effects regression on account of near-collinearity with market capitalization. Reported R^2 are between-group values in stock fixed effects specifications. The number of observations for the within-stock specifications is slightly smaller on account of dropping securities with only one observation.

The key entry in each column is the coefficient on x_{it}. Interpreting the coefficient in the upper left of the table, a one unit increase in the log odd lot ratio increases the price jump ratio by 0.074 (relative to a median price jump ratio of 0.455). Equivalently, a one standard deviation increase in the log odd lot ratio (0.80) is associated with a 13% decrease in the fraction of earnings announcement price impact that occurs pre-announcement.

The second and third specifications for each algorithmic trading proxy include additional controls. The second specification augments lagged log market capitalization with log share price, log return volatility, average quoted bid-ask spreads, institutional ownership share (all lagged), as well as the number of reporting analysts and month fixed effects. All nondummy variables except for analyst coverage clearly vary in response to information acquisition in the pre-announcement period and must be lagged to ensure predetermination. Notwithstanding these additions, the coefficient on the odd lot ratio typically strengthens slightly across specifications (1) and (2).

The final column adds stock fixed effects. This control absorbs potentially relevant fixed differences in disclosure policy and investor base not captured by the other controls. I drop log price from this specification because of the near-collinearity between log price and log market capitalization within stocks (99.6% correlation net of fixed effects).[21] The resulting within-stock coefficient is larger in absolute value, indicating that more algorithmic trading activity is associated with diminished information acquisition both across securities and within a given security.

The estimated effects of algorithmic trading on information acquisition have comparable economic and statistical magnitudes across all four algorithmic trading proxies. To illustrate, I multiply the point estimates for the first regression specification by each measure's respective standard deviation. By comparison with the 13% decrease in the pre-announcement share of earnings announcement price impacts estimated using odd lot ratios, I estimate a 10% decrease using trade-to-order volume ratios, a 9% decrease using cancel-to-trade ratios, and a 15% decrease using average trade sizes. Similar comparability is achieved across all regression specifications.

5.2 Identifying exogenous variation in algorithmic trading

Interpreting these estimates causally requires exogenous variation in AT participation because the preceding regressions potentially suffer from reverse causality and omitted variable problems. Algorithmic traders may respond to information entering prices, for example, by detecting the presence of informed traders and withdrawing from the market to avoid adverse selection. This

[21] The within-stock coefficient is identified using deviations from stock means, which raises the question of what causes the residual variation in algorithmic trading. Residual drivers of algorithmic trading may include time-varying substitutability with other securities (e.g., convertible debt and highly correlated stocks), index inclusion status, and institutional hedging demands, to name a few.

explanation accords with Hendershott, Jones, and Menkveld's (2011) finding of lower adverse selection with more AT, and such a response would contribute to the observed negative relationship between algorithmic trading and the price jump ratio. Likewise, trading algorithms may be deployed more or less frequently for stocks with less information to acquire. Such selection is another source of bias for the estimated AT-information acquisition relation.

Although I control for lagged market capitalization and bid-ask spreads, other liquidity variables not in the regression also may contribute to an omitted variable bias. A decrease in uninformed trading shortly before an announcement—arising, for example, from fear of adverse selection—increases the price impact of informed trades and decreases algorithmic market making revenues. Mechanisms such as these can generate a spurious relation between AT and the price jump ratio measure unrelated to AT's effects on information acquisition.

To address these concerns, I use the log of the average stock price from 42 to 22 days before each earnings announcement as an instrument for algorithmic trading activity. Controlling for covariates such as market capitalization and institutional ownership, variation in lagged stock prices should relate little to the incentives of market participants to acquire information or to the amount of information available to acquire.[22] I make this argument below. The use of lagged information also focuses the empirical tests squarely on the ex ante effects of algorithmic traders on information acquisition rather than on the responses of algorithmic traders to information already acquired.

The exclusion restriction requires that the component of lagged share prices orthogonal to market capitalization and other covariates must not affect the incentives of market participants to acquire information. For example, market capitalization increases the potential profits associated with acquiring information, but the component of prices orthogonal to market capitalization should have no clear direct effect on the desirability of learning about a stock. The literature on stock splits offers several theories for why firms may manage their share prices, and their potential impact on the exclusion restriction merits discussion.

I explicitly control for bid-ask spreads and institutional ownership to address potential liquidity or catering rationales for controlling nominal share prices (see, e.g., Conroy, Harris, and Benet 1990 and Angel 1997 for liquidity rationales; Baker, Greenwood, and Wurgler 2009 for a catering explanation). Under norm-based theories of choosing a target price range such as Rozeff (1998) and Weld, Michaely, Thaler, and Benartzi (2009), changing the numeraire for trading does not affect economic fundamentals or the incentives to acquire information.

[22] The sample universe excludes pink sheets and other issues for which low stock prices may translate into different disclosure regimes. All coefficient estimates are comparable with minimum stock price thresholds of \$1, \$20, and \$50.

Signaling theories like in Brennan and Copeland (1988) and Asquith, Healy, and Palepu (1989) conjecture that numeraire changes are a costly signal of firm value. Empirical evidence suggests that firm earnings and profitability significantly increase before stock splits, but not thereafter (Lakonishok and Lev 1987; Asquith et al. 1989). Because the change in numeraire does not translate into future shifts in fundamentals, it is not likely that split-induced variation in prices alters the incentives to acquire information. Moreover, Easley, O'Hara, and Saar (2001) find no evidence of reductions in information asymmetries around stock splits, as would be consistent with firms signaling private information. Because information asymmetries do not decline, either (1) the incentives to acquire information remain the same (and the exclusion restriction is satisfied) or (2) signaling by the firm is precisely counterbalanced by a change in the set of information available to acquire. In the Online Appendix I confirm that possible signaling through changes in price levels does not affect my inference by discarding all observations within three months of a stock split declaration or effective date.

The "sub-penny" rule (SEC Rule 612) mandates a minimum price increment of one cent for displayed orders in stocks covered by Reg NMS, and this minimum increment plays a critical role in creating dispersion in algorithmic trading activity across stocks. As stock prices increase, this minimum increment tightens the grid of tradable prices as a fraction of the midpoint price. For example, the minimum price increment on a \$10 stock equates to 10 basis points, whereas the minimum price increment on a \$100 stock equates to 1 basis point. For a given percent change in the underlying value of the asset, the \$100 stock is more likely to require quote revisions than the \$10 stock. A 5-basis-point change in the \$10 stock's underlying value may require no action by traders—after all, the movement is half of the smallest price increment—whereas the \$100 stock's value moves five price ticks. Because algorithmic liquidity providers are better equipped to continually update quotes than human traders are, and algorithmic liquidity takers are similarly better able to take advantage of momentarily "stale" quotes (in the sense of Foucault, Röell, and Sandås 2003), we would expect algorithmic traders to comprise a greater share of trading in stocks with higher prices all else equal.

Yao and Ye (2017) argue for the opposite relation between relative tick size and high-frequency trading, a subset of AT. Because tick sizes are larger than the equilibrium price of liquidity for low-priced stocks, inframarginal units of quoted depth are profitable. HFT have an advantage in obtaining time priority and enjoying these inframarginal rents, whereas slower traders tend toward the back of the queue at the same price or at a less competitive price. Using HFT-identified NASDAQ data, they find evidence for this opposing relation.

In my data, the quote-updating channel dominates the time-priority channel, and the positive relationship between algorithmic trading proxies and the log price is quite strong. Table 3 reports correlations of algorithmic trading proxies and the lagged log stock price instrument. All correlations are similar

Table 3
Correlations with lagged log price instrument

Sample	Odd lots	Trades/orders	Cancels/trades	Trade size	AT PCF_1
Full sample	0.730***	−0.432***	0.251***	−0.766***	0.640***
Regression sample	0.721***	−0.456***	0.293***	−0.752***	0.640***
Net of mkt. cap.	0.787***	−0.587***	0.579***	−0.763***	0.779***
Net of all controls	0.769***	−0.550***	0.560***	−0.750***	0.752***

$^{*}p<.10$, $^{**}p<.05$, $^{***}p<.01$
This table reports raw correlations of the lagged log stock price and algorithmic trading proxies for the full sample and for the regression sample, as well as correlations of these measures net of variation spanned by market capitalization and other control variables for each stock i and quarterly earnings announcement t from January 2012 through September 2016. Construction of algorithmic trading proxies (x_{it}) derived from SEC MIDAS data is described in the main text. PCF_1 denotes the first principal component factor of the AT proxies.

between full and regression samples, and the correlations only strengthen when orthogonalizing the log price with respect to the obvious confound of market capitalization. Importantly, these relations are not driven by extreme prices in the sample: averages of all AT proxies vary monotonically with deciles of the lagged log price and the lagged log price orthogonalized with respect to the controls (Figure 1).

Having established the necessary conditions for the lagged log stock price instrument, I now augment Equation (4) with a first stage for x_{it},

$$x_{it} = \zeta + \eta lprice_{it} + \theta \times controls_{it} + \delta_{it},$$

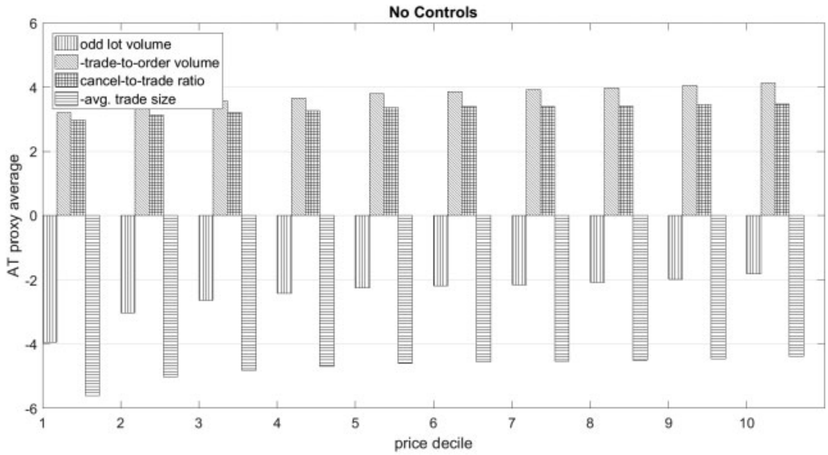
$$jump_{it}^{(21)} = \alpha + \beta \hat{x}_{it} + \gamma \times controls_{it} + \epsilon_{it}. \tag{5}$$

Specifications parallel the noninstrumented regressions, with the exception that lagged log price is necessarily excluded as an independent variable in the second-stage regression. I omit stock fixed effect specifications because stock-level averages absorb much of the variation in my instrument (stock prices are persistent across quarters).

Table 4 presents results from this analysis.[23] These IV estimations reinforce the OLS results: an increase in algorithmic trading causes a sharp reduction in the information content of prices around earnings announcements. Moreover, the estimated magnitudes are in line with the OLS estimates of Table 2 in the baseline specification, and coefficients are effectively unchanged with the addition of other controls in the IV.

The specifications with multiple controls differ in the use of lagged log stock price as an instrument rather than a control variable, and this distinction is revealing about how algorithmic traders may endogenously respond to information entering stock prices. Comparing coefficients in Tables 2 and 4, the part of AT captured by log prices (net of the other controls) has a weaker

[23] Throughout I use the (cluster-robust) Kleibergen and Paap (2006) *rk* Lagrange multiplier (LM) and Wald F statistics to evaluate instrument strength for each two-stage regression. All LM tests reject underidentification at the 1% significance level, and all specifications feature first-stage F statistics exceeding the critical value corresponding to a 10% maximal IV size (16.38).



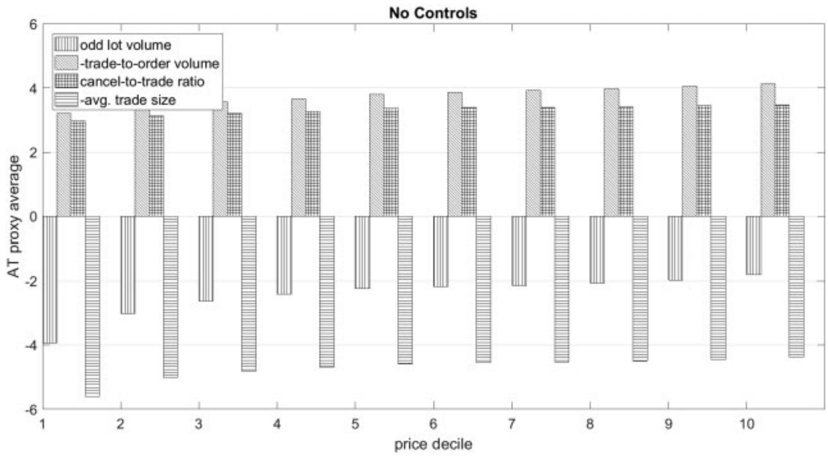


Figure 1
Algorithmic trading proxies by log price decile
This figure presents averages of the four algorithmic trading proxies for each decile of the log price and the component of log price orthogonal to all controls. "No controls" specification (top figure) consists of the algorithmic trading proxy averaged within log price deciles, where deciles are assigned by calendar month. "All controls" specification (bottom figure) averages AT proxies by residual log price decile, where the residual is constructed by regressing the log price on return volatility, quoted spread, number of analysts, institutional ownership ratio, and month fixed effects. AT proxies are normalized to be increasing in AT. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42, T-22]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio (*IOR*) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading proxies derived from SEC MIDAS data is described in the main text.

Table 4
Determinants of announcement price impact with lagged log price instruments

	x = Odd lot ratio		x = Trade-to-order ratio		x = Cancel-to-trade ratio		x = Avg. trade size	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
x	0.0605***	0.0507***	−0.0920***	−0.0778***	0.104***	0.0874***	−0.123***	−0.108***
	(0.00775)	(0.00841)	(0.0127)	(0.0138)	(0.0149)	(0.0158)	(0.0154)	(0.0175)
Market cap.	0.00422	−0.0281***	0.00719**	−0.0296***	0.0146***	−0.0254***	0.00269	−0.0272***
	(0.00320)	(0.00443)	(0.00302)	(0.00439)	(0.00311)	(0.00435)	(0.00321)	(0.00430)
Ret. vol.		−0.0202		−0.0120		−0.00644		−0.0209*
		(0.0127)		(0.0124)		(0.0126)		(0.0123)
Quoted spr.		−0.0395***		−0.0560***		−0.0660***		−0.0292***
		(0.00787)		(0.00792)		(0.00881)		(0.00815)
#Analysts		0.0485***		0.0534***		0.0537***		0.0488***
		(0.00764)		(0.00789)		(0.00775)		(0.00747)
IOR		0.128***		0.123***		0.136***		0.113***
		(0.0227)		(0.0222)		(0.0218)		(0.0227)
Constant	0.529***	X	−0.0193	X	−0.170**	X	0.991***	X
	(0.0747)		(0.0678)		(0.0814)		(0.120)	
Month FEs	No	Yes	No	Yes	No	Yes	No	Yes
Stock FEs	No	No	No	No	No	No	No	No
N	24,107	23,814	24,068	23,800	24,120	23,842	24,512	24,201
K-P rk LM	42.67***	41.92***	42.03***	41.14***	41.44***	41.00***	42.12***	41.44***
K-P rk Wald F	4,226.2	4,554.4	1,493.9	1,189.7	1,001.3	899.2	2,765.0	2,808.9

$^{*}p<.10$, $^{**}p<.05$, $^{***}p<.01$

This table reports results from an instrumental variables regression of price jump ratios on a set of algorithmic trading proxies:

$$x_{it} = \zeta + \eta lprice_{it} + \theta \times controls_{it} + \delta_{it},$$

$$jump_{it}^{(21)} = \alpha + \beta \hat{x}_{it} + \gamma \times controls_{it} + \epsilon_{it}.$$

For each stock i and quarterly earnings announcement t from January 2012 through September 2016, the first-stage regression instruments algorithmic trading measures using the log of the average end-of-day stock price from $T-42$ to $T-22$. The second-stage regression for which results are reported relates predicted algorithmic trading proxies to the price jump ratio ($jump_{it}^{(21)}$). The price jump ratio is measured as the ratio of the announcement response divided by the total variation in the pre- and post-announcement period: $CAR_{it}^{(T-1,T+2)}/CAR_{it}^{(T-21,T+2)}$. Cumulative abnormal returns (in logs) are net of Fama and French (1992) three-factor implied returns over the same interval. Observations with cumulative net price impact not exceeding a minimum multiple of prior return volatility (described in the main text) are dropped from both stages. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42, T-22]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio (*IOR*) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading proxies (x_{it}) derived from SEC MIDAS data is described in the main text. All standard errors are clustered by security and month and are reported in parentheses. K-P refers to the Kleibergen and Paap (2006) rk LM and Wald F statistics. In these specifications, a 10% maximal IV size corresponds to a critical value of 16.38.

association with reduced information acquisition than does the residual part of AT unspanned by log prices. One interpretation of this difference is that some algorithmic traders endogenously reduce their participation in stocks with greater information acquisition (like in Han et al. 2014; Baldauf and Mollner 2017), or conversely, that firm-quarters with larger price jump ratios tend to attract algorithmic traders.

5.3 When does AT deter information acquisition?

Pinpointing the timing of deterred information acquisition is critical for assessing whether informational distortions in financial markets might translate into suboptimal production or investment decisions. To assess the timing of information distortions, I decompose the price jump ratio for the entire pre-announcement period into daily *price response ratios* for earnings information. The price response ratio divides the cumulative abnormal return through date $T-k$ by the total abnormal return variation in the pre- and post-announcement period,

$$responseratio_{it}^{(k,21)} = \frac{CAR_{it}^{(T-21,T-k)}}{CAR_{it}^{(T-21,T+2)}}. \tag{6}$$

Lower price response ratios are associated with less pre-announcement information acquisition.

For each pre-announcement date k, I run the regressions of Equation (5) using the price response ratio for date k as the dependent variable in the second stage:

$$x_{it}^{(k)} = \zeta + \eta lprice_{it}^{(k)} + \theta \times controls_{it}^{(k)} + \delta_{it},$$

$$responseratio_{it}^{(k,21)} = \alpha + \beta^{(k)} \hat{x}_{it}^{(k)} + \gamma \times controls_{it}^{(k)} + \epsilon_{it}, \ \forall k = 1, \ldots, 21. \tag{7}$$

$x_{it}^{(k)}$ is the k-day lagged algorithmic trading proxy for stock i at date t. I instrument each lagged AT proxy using the lagged log stock price with corresponding additional lags, that is, the date k lagged log price is the log of the average price over $[T-42-k, T-22-k]$ measured in days relative to the earnings announcement date T. I simultaneously estimate the system (7) for all $k=1,\ldots,21$ using two-step GMM to test hypotheses involving βs across multiple dates. In particular, I test hypotheses using Wald tests on $\beta^{(k)}$s for several pre-announcement dates k. Coefficients jointly different from zero indicate that algorithmic trading changes information acquisition as of a given point in the pre-announcement period.

Figure 2 plots regression coefficients by date for each of the four algorithmic trading proxies. To facilitate comparison across measures of algorithmic trading, I scale all coefficients by the standard deviation of the predicted AT proxy $\hat{x}^{(k)}$ from the first-stage regression of $x^{(k)}$ on shifted log price and controls. The top plot's regressions include a single control (lagged

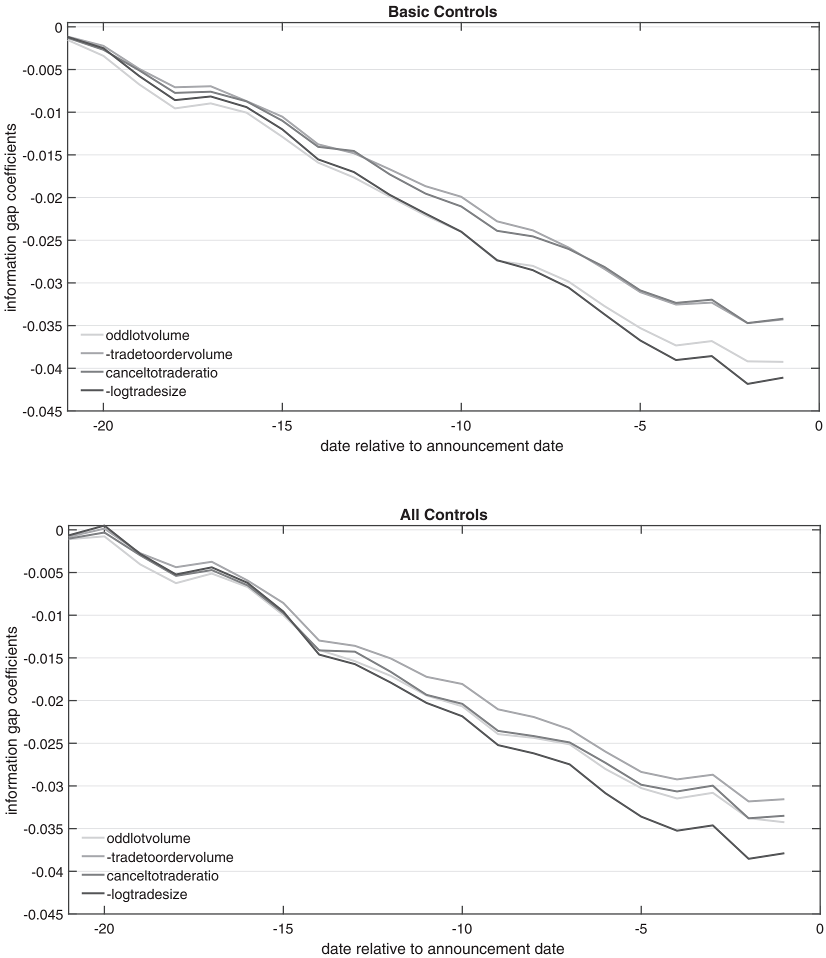


Figure 2
Price response ratios for algorithmic trading proxies
This figure presents coefficients from regressions of price response ratios on the four algorithmic trading proxies using the methodology of Table 5. To facilitate comparison across AT measures, I standardize all coefficients by normalizing the instrumented AT proxy $\hat{x}^{(k)}$ to have a unit standard deviation and to be increasing in AT. For each stock i and quarterly earnings announcement t from January 2012 through September 2016, the price response ratio is measured as the ratio of the cumulative price response through date k prior to the announcement date divided by the total variation over the information incorporation window: $CAR_{it}^{(T-21,T-k)}/CAR_{it}^{(T-21,T+2)}$. Cumulative abnormal returns (in logs) are net of Fama and French (1992) three-factor implied returns over the same interval. Observations with cumulative net price impact not exceeding a minimum multiple of prior return volatility (described in the main text) are dropped. "Basic" specification (top figure) consists of the algorithmic trading proxy and market capitalization. "All" specification (bottom figure) adds return volatility, quoted spread, number of analysts, institutional ownership, and month fixed effects. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42-k, T-22-k]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio (*IOR*) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading measures derived from SEC MIDAS data is described in the main text.

market capitalization), and the bottom plot's regressions include all controls of specification (2) from Table 2.

The figure demonstrates that the relation between algorithmic trading and information acquisition becomes increasingly negative throughout the pre-announcement period. Coefficients for all algorithmic trading proxies follow very similar paths, both within each plot and between plots. The stable downward slope from 21 days prior to announcement is consistent with a near-constant difference between high- and low-AT stocks in the share of information being acquired and impounded into prices each day. Smooth incorporation of information is a cornerstone of strategic informed trading models (e.g., Kyle 1985; Back 1992), so it should be expected that deterring information acquirers manifests as ever-widening pre-announcement information gaps.[24]

Table 5 supplements these graphical results with formal tests of the hypothesis that information gaps are present at various stages of the pre-announcement period. The top panel uses price response ratios as the dependent variable in the second stage and tests whether total information gaps within a period are distinguishable from zero. Tests reject the hypothesis of no information gap for both two week subperiods two to four weeks and zero to two weeks pre-announcement, as well as for the whole pre-announcement period (not tabulated). Rejection of no information gap is comparably strong for each algorithmic trading proxy independently as well as for the first principal component factor of the four AT proxies.

The bottom panel examines the incremental information gap by day for each subperiod. I again test the hypothesis that sets of $\beta^{(k)}s$ are jointly zero in Equation (7), where the second-stage dependent variable is now replaced by the *daily change* in the cumulative price response ratio. In both 2-week subperiods, as for the whole sample, the incremental information gaps are distinguishable from zero on average, indicating that algorithmic trading is associated with continual reductions in relative information acquisition for the entire pre-announcement month.

6. Alternatives to the Price Jump Ratio Measure

6.1 Price nonsynchronicity

This section considers two alternative measures of information acquisition—price nonsynchronicity and intraperiod timeliness—to confirm that findings are robust to how information acquisition is measured. Roll (1988) proposes an R^2-based measure of firm-specific information in prices, and this idea is further developed into a "price nonsynchronicity" measure by Morck, Yeung, and Yu (2000) and Durnev, Morck, Yeung, and Zarowin (2003).

[24] Interpretation of the price response ratio coefficients is unchanged by allowing for multiple informed traders like in Holden and Subrahmanyam (1992) or Foster and Viswanathan (1993). The Online Appendix provides additional discussion of this point.

Table 5
Algorithmic trading and the timing of information acquisition

		Δ: 2–4 weeks pre		Δ: 0–2 weeks pre	
		(3)	(4)	(5)	(6)
Controls		Basic	All	Basic	All
Degrees of freedom		10	10	10	10
x = Odd lot ratio	χ^2	53.62***	35.47***	98.61***	54.38***
	p	0.000	0.000	0.000	0.000
x = Trade-to-order ratio	χ^2	45.79***	34.89***	93.80***	56.27***
	p	0.000	0.000	0.000	0.000
x = Cancel-to-trade ratio	χ^2	46.73***	35.56***	91.18***	56.19***
	p	0.000	0.000	0.000	0.000
x = Avg. trade size	χ^2	46.56***	35.11***	97.44***	55.66***
	p	0.000	0.000	0.000	0.000
x = First PCF of AT measures	χ^2	48.78***	34.61***	89.39***	53.79***
	p	0.000	0.000	0.000	0.000
		Δ: 2–4 weeks pre		Δ: 0–2 weeks pre	
		(3)	(4)	(5)	(6)
Controls		Basic	All	Basic	All
Degrees of freedom		10	10	10	10
x = Odd lot ratio	χ^2	51.18***	35.00***	42.01***	22.93**
	p	0.000	0.000	0.000	0.011
x = Trade-to-order ratio	χ^2	44.03***	34.89***	41.34***	24.44***
	p	0.000	0.000	0.000	0.007
x = Cancel-to-trade ratio	χ^2	45.69***	35.96***	41.32***	25.42***
	p	0.000	0.000	0.000	0.005
x = Avg. trade size	χ^2	45.89***	35.47***	50.22***	28.08***
	p	0.000	0.000	0.000	0.002
x = First PCF of AT measures	χ^2	46.30***	34.06***	35.27***	22.20**
	p	0.000	0.000	0.000	0.014

$*p<.10$, **p<.05, $***\ p<.01$

This table presents results from IV regressions of price response ratios on a set of algorithmic trading proxies:

$$x_{it}^{(k)} = \zeta + \eta lprice_{it}^{(k)} + \theta \times controls_{it} + \delta_{it},$$

$$responseratio_{it}^{(k,21)} = \alpha + \beta^{(k)} \hat{x}_{it}^{(k)} + \gamma \times controls_{it} + \epsilon_{it},\ \forall k = 1,\ldots,21.$$

For each stock i and quarterly earnings announcement t from January 2012 through September 2016, the price response ratio is measured as the ratio of the cumulative price response (first subtable) through date k prior to the announcement date divided by the total variation over the information incorporation window: $CAR_{it}^{(T-21,T-k)}/CAR_{it}^{(T-21,T+2)}$. Second subtable replaces the cumulative price response with the concurrent price response $\left(CAR_{it}^{(T-21,T-k)} - CAR_{it}^{(T-21,T-k-1)}\right)/CAR_{it}^{(T-21,T+2)}$, where I set $CAR_{it}^{(T-21,T-22)}$ equal to zero. Cumulative abnormal returns (in log) are net of Fama and French (1992) three-factor implied returns over the same interval. Observations with cumulative net price impact not exceeding a minimum multiple of prior return volatility (described in the main text) are dropped.

Table entries correspond with cross-equation Wald test statistics and p-values for hypotheses on sets of $\beta^{(\cdot)}$ estimated using two-step GMM. "Basic" specification consists of algorithmic trading proxy and market capitalization. "All" specification adds share price, return volatility, quoted spread, number of analysts, and month fixed effects. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42, T-22]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with each stock-quarter announcement. The institutional ownership ratio (IOR) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading proxies ($x_{it}^{(k)}$) derived from SEC MIDAS data is described in the main text, and the first principal component factor of these proxies is denoted by "first PCF." All standard errors are clustered by security and month.

Intuitively, the less idiosyncratic information is produced and incorporated into the stock's price, the better a stock's return is approximated by aggregate information. Following Morck et al. (2000) and Durnev et al. (2003), I estimate price nonsynchronicity for each stock-quarter pair as $1-R^2$ from the regression

$$r_{it} = \alpha + \beta_{im} r_{mt} + \gamma r_{ind,t} + \epsilon_{it}, \tag{8}$$

where stock i's industry return is the value-weighted average return for stocks sharing stock i's three-digit SIC code.

One advantage of the price nonsynchronicity measure is the ability to examine a broader set of events for which the price jump ratio is not well-suited. In particular, I no longer require a large and prescheduled information shock because the price nonsynchronicity measure does not require an estimate of event impact, instead normalizing idiosyncratic shocks by the sum of firm plus aggregate shocks. By contrast, the price jump ratio normalizes by total (firm-specific) event impact, which is key for the price jump ratio's interpretation as information acquired relative to information available.

I use RavenPack News Analytics data to construct a wider range of economically important firm-specific events. RavenPack provides real-time collection and analysis of thousands of news sources for tens of thousands of companies in the United States and internationally. Critically for this analysis, RavenPack algorithmically links news items to companies, identifies distinct news events, and classifies items into categories ranging from earnings news to labor issues. Table 6 summarizes the set of events comprising more than 1% of RavenPack equity news coverage, excluding reports on trading or prices (technical analysis signals, stock price movements, order imbalance reports) and announcements of future disclosure dates (investor relations items). Of these items, 17.7% are earnings news, and more than 30% are earnings, dividend, or revenue items (often announced together). The remaining events that exceed the 1% threshold are distributed mostly among product releases, executive appointments and resignations (broadly, "labor issues"), M&A activity, analyst updates, and insider trading reports.

Informed by Table 6, I add three sets of common firm events that provide material information on firms' scope, inputs, or outputs: M&A activity, executive appointments and resignations, and product releases. I construct the set of stock-event pairs as follows. I first build a list consisting of each event with RavenPack "group" equal to one of these categories, relevance score exceeding 90 (highly relevant), and global novelty score equal to 100 (first occurrence of reporting within 24 hours).[25] Then, for each stock, I roll backward through the

[25] From the RavenPack Version 4.0 User Guide, "a relevance value of at least 90 indicates that the entity is referenced in the main title or headline of the news item." I exclude events with relevance scores lower than 90 on a 0–100 scale to ensure that news items have bearing on the algorithmically tagged companies. I also impose a novelty filter using RavenPack's proprietary "global novelty score" (GNS), and I require that the item be the first instance of reporting (GNS=100). The novelty filter prevents double counting of single news events reported in multiple publications.

Table 6
RavenPack news items

Rank	Group	Type	N	%	Cumulative %
1	Earnings	Earnings	3,903,482	17.7	17.7
2	Products-services	Product-release	1,475,898	6.7	24.4
3	Labor-issues	Executive-appointment	1,235,878	5.6	30.1
4	Products-services	Business-contract	1,184,949	5.4	35.4
5	Revenues	Revenue	876,281	4.0	39.4
6	Analyst-ratings	Analyst-ratings-change	875,978	4.0	43.4
7	Acquisitions-mergers	Scquisition	865,047	3.9	47.3
8	Marketing	Conference	766,278	3.5	50.8
9	Credit-ratings	Credit-rating-change	670,356	3.0	53.8
10	Earnings	Earnings-per-share	595,781	2.7	56.6
11	Investor-relations	Conference-call	577,830	2.6	59.2
12	Insider-trading	Insider-sell	571,758	2.6	61.8
13	Dividends	Dividend	541,002	2.5	64.2
14	Partnerships	Partnership	490,267	2.2	66.5
15	Earnings	Earnings-guidance	480,604	2.2	68.6
16	Insider-trading	Sell-registration	472,851	2.1	70.8
17	Acquisitions-mergers	Unit-acquisition	393,994	1.8	72.6
18	Insider-trading	Insider-buy	370,203	1.7	74.3
19	Price-targets	Price-target	354,128	1.6	75.9
20	Labor-issues	Executive-resignation	302,980	1.4	77.2
21	Assets	Facility	268,681	1.2	78.5
22	Revenues	Revenue-guidance	254,041	1.2	79.6
23	Analyst-ratings	Analyst-ratings-set	251,252	1.1	80.8

This table presents all RavenPack company news categories with shares exceeding 1% of the January 2012–March 2016 RavenPack sample. Technical analysis signals, stock price movements, order imbalance reports, and investor relations items (typically announcements of future information revelation dates) are excluded.

data to find any of these events within the last week and take the first event in the string of RavenPack reports. This step is necessary because a 24-hour gap in coverage generates a new, but often redundant event in the RavenPack data. Finally, I move each news announcement date to the first date in which the market is open after the announcement. Each element of this collapsed list represents a potential information shock with a tilt toward idiosyncratic or industry news relative to a typical day in the life of a stock. RavenPack contributes 72,544 events in total.[26] I stack these events with the 54,879 earnings announcement events to obtain a consolidated set of important firm events with which to assess the representativeness of my earnings announcement results.

Equipped with these data, one may be tempted to simply regress price nonsynchronicity around each event on algorithmic trading proxies and controls,

$$nonsynch_{it} = \alpha + \beta x_{it} + \gamma \times controls_{it} + \epsilon_{it}. \tag{9}$$

However, such an approach faces two challenges. First, high R^2 indicates that there is little idiosyncratic information to acquire (case 1) or that idiosyncratic innovations are not discovered and incorporated into stock returns (case 2). AT may prefer stocks with more or less potentially acquirable information so that

26 The RavenPack sample is available only through March 2016, so the date range is slightly shorter than that accommodated by MIDAS, CRSP, and other resources.

an association of AT with price nonsynchronicity reflects case 1 rather than case 2. This concern requires that AT proxies be cleansed of their correlation with the amount of information acquirable. This problem is amenable to an instrumental variables approach like in Section 5.2. For the exclusion restriction to be satisfied, the lagged log stock price must now be unrelated to the prevalence of potentially acquirable *idiosyncratic* information relative to factor information (conditional on other covariates). As before, there is no obvious channel by which the lagged price should relate to market-industry R^2s holding fixed covariates such as lagged market capitalization (e.g., large firms may be more representative of their industries) and institutional ownership ratios.

There is also a second challenge specific to my setting that requires the price nonsynchronicity measure to be treated with additional care. Algorithmic traders may better enforce true factor relationships than their human counterparts (by, e.g., statistical arbitrage), thereby increasing the R^2 of a market-industry model. In this case, it would be wrong to interpret this increase in R^2 as less idiosyncratic information entering prices because it may well reflect more industry or index information entering prices instead. As noted by Kirilenko and Lo (2013), index arbitrage is indeed quite common among algorithmic traders, so we would expect to encounter a negative relation between price nonsynchronicity and AT in estimating Equation (9).[27] For this reason, a significant negative association between AT and price nonsynchronicity is informative only insofar as a strong positive association would be difficult to reconcile with my previous results.

I address this problem by exploiting the fact that the period before major firm news may feature more *firm-specific* acquirable information than the typical span over the preceding year. If factor and industry betas do not change much from typical days to news periods, then subtracting price nonsynchronicity around "typical" periods from price nonsynchronicity around "event" periods nets out the effects of AT statistical arbitrage. A high *price nonsynchronicity difference* reflects larger increases in return responsiveness to increases in firm-specific news. Correspondingly, I define the price nonsynchronicity difference as $1-R^2$ of regression (8) for the $[T-21, T-1]$ period minus $1-R^2$ of regression (8) for the $[T-252, T-64]$ period.

Table 7 reports the results of IV regressions of price nonsynchronicity differences on algorithmic trading proxies,

$$x_{it} = \zeta + \eta lprice_{it} + \theta \times controls_{it} + \delta_{it},$$

$$nonsynch_{it}^{event} - nonsynch_{it}^{ave} = \alpha + \beta \hat{x}_{it} + \gamma \times controls_{it} + \epsilon_{it}. \tag{10}$$

Netting out the effects on the price nonsynchronicity measure arising from index arbitrage algorithms, more algorithmic trading decreases firm-specific

[27] Indeed, Gerig (2015) finds evidence of an association between HFT and increased price synchronization across securities.

Table 7
Determinants of announcement price impact: Difference of price nonsychronicities

	x = Odd lot ratio		x = Trade-to-order ratio		x = Cancel-to-trade ratio		x = Avg. trade size	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
x	−0.0170***	−0.00736***	0.0286***	0.0115***	−0.0312***	−0.0121***	0.0332***	0.0141***
	(0.00294)	(0.00227)	(0.00492)	(0.00385)	(0.00567)	(0.00434)	(0.00576)	(0.00456)
Market cap.	0.00630***	0.00688***	0.00515***	0.00698***	0.00318	0.00639***	0.00696***	0.00671***
	(0.00192)	(0.00200)	(0.00189)	(0.00198)	(0.00194)	(0.00197)	(0.00194)	(0.00199)
Ret. vol.		−0.0106**		−0.0117**		−0.0124**		−0.0103**
		(0.00481)		(0.00502)		(0.00505)		(0.00481)
Quoted spr.		−0.00309		−0.000796		0.000119		−0.00464
		(0.00362)		(0.00413)		(0.00440)		(0.00352)
#Analysts		−0.00131		−0.00171		−0.00158		−0.00104
		(0.00278)		(0.00276)		(0.00273)		(0.00271)
IOR		−0.00311		−0.00463		−0.00579		−0.00123
		(0.00728)		(0.00754)		(0.00752)		(0.00721)
Constant	−0.239***	X	−0.0674	X	−0.0313	X	−0.368***	X
	(0.0397)		(0.0411)		(0.0448)		(0.0510)	
Month FEs	No	Yes	No	Yes	No	Yes	No	Yes
Stock FEs	No	No	No	No	No	No	No	No
N	53,679	53,004	53,728	53,113	53,727	53,108	54,757	54043
K-P rk LM	43.05***	42.24***	42.54***	41.68***	41.57***	40.90***	42.39***	41.72***
K-P rk Wald F	4,527.7	5,021.6	1,442.0	1,123.1	1,193.1	974.3	3,027.8	3,077.3

* $p<.10$, ** $p<.05$, *** $p<.01$

This table presents results from a regression of price nonsynchronicity differences on a set of algorithmic trading proxies:

$$x_{it}=\zeta+\eta lprice_{it}+\theta\times controls_{it}+\delta_{it},$$

$$nonsynch_{it}^{event}-nonsynch_{it}^{ave}=\alpha+\beta\hat{x}_{it}+\gamma\times controls_{it}+\epsilon_{it}.$$

For each stock i and event date t from January 2012 through September 2016, the average price nonsynchronicity ($nonsynch_{it}^{ave}$) is measured as the average of one minus the R^2 of a regression of returns of stock i on market and sector returns for the year preceding the date of the news event (trading days $T-252$ through $T-1$). The event-period price nonsynchronicity ($nonsynch_{it}^{event}$) is the same quantity measured over the pre-event period of $T-21$ to $T-1$ days. Market capitalization, share price, and return volatility are the log of daily averages over $[T-42, T-22]$, and the quoted spread is average of the time-weighted bid-ask spread over the same interval (reported in percent). The number of analysts is the log of the largest number of reporting analysts in the Thomson Reuters I/B/E/S database associated with the earnings announcement in the same stock-quarter. The institutional ownership ratio (*IOR*) is the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. Construction of algorithmic trading proxies (x_{it}) derived from SEC MIDAS data is described in the main text. All standard errors are clustered by security and month and are reported in parentheses. The log price is dropped from stock fixed effects regression on account of near-collinearity with market capitalization. K-P refers to the Kleibergen and Paap (2006) rk LM and Wald F statistics. In these specifications, a 10% maximal IV size corresponds to a critical value of 16.38.

news in prices before major firm events.[28] This reduction in the discovery of idiosyncratic information around a range of important news events generalizes the evidence obtained before earnings announcements using the price jump ratio. Notably, these estimates are conservative because statistical arbitrage if anything becomes less attractive in the presence of more idiosyncratic news (e.g., information ratios fall as idiosyncratic risk increases). Consequently, algorithmic traders would be expected to pare back such strategies before idiosyncratic events, thereby pushing the price nonsynchronicity difference upward and biasing β up for proxies increasing in AT.

6.2 Intraperiod timeliness

I construct "perfect foresight" returns for the IPT measure by investing $+1/N_{up}$ dollars in all N_{up} stocks that increase in value around the earnings announcement and $-1/N_{down}$ dollars in all N_{down} stocks that decrease in value around the earnings announcement. These zero-cost or "hedge" portfolios are constructed separately for stocks with high and low values of a specific characteristic, namely algorithmic trading activity or lagged log stock prices in this application. I then compare (1) price jump ratio analogues at the portfolio level and (2) "areas under the curve" plotting cumulative abnormal returns relative to total abnormal returns for each pre-announcement date. Large areas under the curve represent earlier incorporation of information.

Figure 3 and Table 8 present results from portfolio price jump ratio and intraperiod timeliness tests, where I split AT measures into quintiles in constructing hedge portfolios. The figure plots scaled return differences between high AT and low AT quintiles using annual hedge portfolios. For every algorithmic trading measure and annual cut, increased algorithmic trading is associated with reduced portfolio-level price response ratios and intraperiod timeliness. Differences between high- and low-lagged price portfolios are comparably large. The table reports tests of the hypothesis of equal responsiveness in the high and low AT portfolios with quarterly observations, and equality is easily rejected for all measures for both CARs and intraperiod timeliness.

The IPT approach circumvents the undefined conditional mean problem of the price jump ratio if all portfolio abnormal returns are "large enough" in absolute value, as they are in this application. However, the comparison of the share of absolute cumulative abnormal returns accrued up to date $T-k$ (or its integral) is not valid if the portfolios are not "all else equal." A major drawback to this analysis is the inability to add multiple control variables to enforce this condition. Unfortunately, adjusting the methodology to overcome this issue is not straightforward because the IPT approach necessarily examines a small

[28] If the effect of additional index arbitrage on price nonsynchronicity is multiplicative rather than additive, the price nonsynchronicity difference should be replaced by $nonsynch_{it}^{event}/nonsynch_{it}^{ave}$ in regression (10). Using this ratio delivers the same qualitative relationships with similar levels of statistical significance.

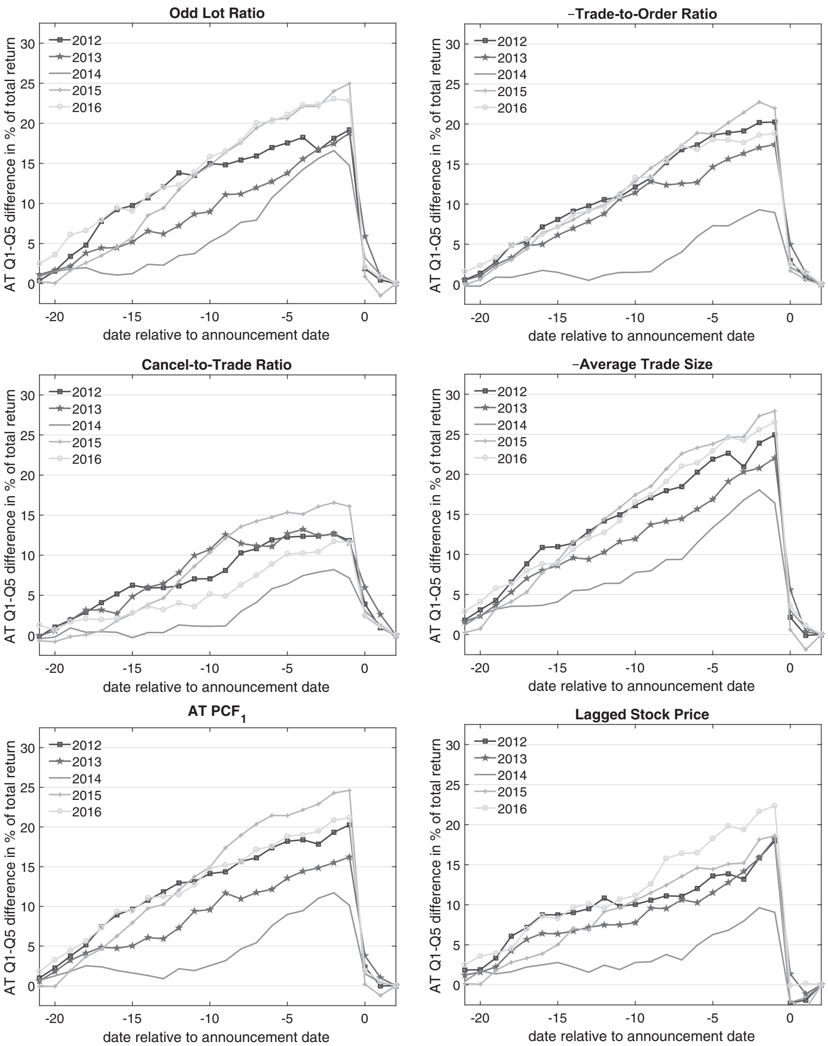


Figure 3
Differences in intraperiod timeliness
This figure presents differences in intraperiod timeliness in the mode of McNichols (1984) and Alford et al. (1993). I sort stocks into quintiles for each algorithmic trading proxy for each calendar year. For securities in quintiles one or five, I construct the "perfect foresight" return on a zero-cost portfolio with scale of one dollar. At date $T-21$ relative to announcement date T, I invest $+1/N_{up}$ dollars in each stock that earns a positive return through $T+2$, and I invest $-1/N_{down}$ dollars in each stock that earns a negative return through $T+2$, where N_{up} and N_{down} are the counts of securities in each group. For each date, I then compute the abnormal portfolio return for each quintile (net of Fama and French 1992 three-factor returns) divided by the total abnormal portfolio return as of $T+2$. Figures plot the difference between the relative returns as of date k preceding the announcement on the low-AT portfolio net of the high-AT portfolio. PCF$_1$ denotes the first principal component factor of the AT proxies.

Table 8
Intraperiod timeliness tests

A. Test of equality of cumulative abnormal portfolio return as of T

$x=$	Odd lots (1)	$\frac{\text{Trades}}{\text{Orders}}$ (2)	$\frac{\text{Cancels}}{\text{Trades}}$ (3)	Trade size (4)	AT PCF_1 (5)	log price (6)
Q5–Q1	−20.36***	−16.63***	−11.71***	−23.40***	−18.71***	−16.47***
	(14.19)	(11.38)	(8.85)	(15.06)	(11.62)	(10.84)

B. Test of equality of IPT as of T

$x=$	Odd lots (1)	$\frac{\text{Trades}}{\text{Orders}}$ (2)	$\frac{\text{Cancels}}{\text{Trades}}$ (3)	Trade size (4)	AT PCF_1 (5)	log price (6)
Q5–Q1	−230.21***	−195.23***	−136.23***	−268.61***	−221.78***	−175.57***
	(11.51)	(10.22)	(7.33)	(13.13)	(10.05)	(8.39)

$*p<.10$, $**\ p<.05$, $***\ p<.01$

This table presents results from intraperiod timeliness tests in the style of McNichols (1984) and Alford et al. (1993). I sort stocks into quintiles for each algorithmic trading proxy for each calendar quarter. For securities in quintiles one or five, I construct the "perfect foresight" return on a zero-cost portfolio with scale of one dollar. At date $T-21$ relative to announcement date T, I invest $+1/N_{up}$ dollars in each stock that earns a positive return through $T+2$, and I invest $-1/N_{down}$ dollars in each stock that earns a negative return through $T+2$, where N_{up} and N_{down} are the counts of securities in each group. For each date, I then compute the abnormal portfolio return for each quintile (net of Fama and French (1992) three-factor returns) divided by the total abnormal portfolio return as of $T+2$. Panel A tests the hypothesis that the average relative returns (in %) are equal in portfolios 1 and 5 as of date T across calendar quarters. Panel B supplements this t test with an IPT comparison by cumulating the area under each curve and comparing average areas as of date T between portfolios 1 and 5. PCF_1 denotes the first principal component factor of the AT proxies. Standard errors are reported in parentheses.

number of portfolios in place of the price jump ratio's richer cross section of stock-quarter observations.[29]

7. Conclusion

Friction between the building blocks of price discovery—acquiring new information and impounding existing information into prices—is more than a theoretical curiosity. I demonstrate a conflict between these forces in the context of major technological innovations in trading. Algorithmic trading simultaneously increases price efficiency with respect to acquired information while reducing the available information to which prices respond. These competing effects of algorithmic trading complicate the evaluation of technological advances in financial markets.

On balance, my findings reflect a large negative effect of algorithmic trading on information acquisition. Two channels are consistent with this result. First, AT order anticipation or learning from order flow may erode rents to information acquirers (Yang and Zhu 2017; Stiglitz 2014). Second, more sophisticated signal processing by algorithmic liquidity providers may reduce adverse selection costs, but at the same time, it may also reduce the profits

[29] In principle adding sorting dimensions or reweighting observations could accommodate additional controls. However both approaches face a curse of dimensionality with the number of sorting or matching dimensions, and adding more than one or two controls is not feasible given my sample size.

accruing to an informational advantage (Baldauf and Mollner 2017). I offer one attempt to quantify the relative importance of these channels in the Online Appendix, but more work is needed to assess the precise mechanisms by which improved trading technology reduces the information content of prices.

In general, weighing the trade-off between information acquisition and price efficiency requires additional structure from models of trader behavior. However, in the specific case of algorithmic trading, the net effect on social welfare through informational channels is almost surely a loss. With homogenous households, production or investment decisions must be affected for information acquisition to affect social welfare, and these decisions do not depend on split-second price changes. Unlike improvements in price efficiency that occur over horizons of seconds or milliseconds, deterrence of information acquisition persists over (at least) a 1-month horizon prior to the earnings announcements. In light of the allocative externalities associated with information-rich prices, formalizing the welfare trade-offs associated with information acquisition and market quality remains an important challenge for future research and policy guidance.

Appendix A. Sample Selection

Figure A1 plots the distribution of price jump ratios for the set of sufficiently economically important announcements for several choices of the minimum event size cutoff, $\left|CAR_{it}^{(T-21,T+2)}\right| > \sqrt{24}\hat{\sigma}_{it} \times$ cutoff. The plot reveals the trade-off between retaining a larger portion of the sample and reducing the fraction of price jump ratios in the $[0,1]$ interval. For the most permissive cutoff considered, more than 63% of retained observations are in the unit interval in which the "true" price jump ratio should lie absent idiosyncratic noise. I select a cutoff of one to balance retaining data and identifying coefficients off of important earnings announcements, but results do not vary much across a range of cutoffs exceeding 1/2 (or less).

A significant fraction of the sample is thus omitted from the main analysis. I now assess whether the observations included in the main analysis systematically differ from the observations excluded by the volatility cutoff rule. Table A1 presents summary statistics for the included and the excluded groups for every control variable in my analysis. Distributions of included and excluded observations are very similar along the dimensions of market capitalization, analyst coverage (excepting the left tail), and institutional ownership. Excluded observations have slightly higher volatility, as would be expected given the volatility-based cutoff. Log price distributions are shifted leftward by 10% to 20% for excluded observations, and correspondingly, relative bid-ask spreads are 20% to 30% larger for that set. If anything, the analysis seems to focus on a set of securities with marginally more predicted AT than the general set of stocks. Notwithstanding this difference, the overlap between price and bid-ask spread distributions is still quite substantial, for example, the 10th (25th) [50th] percentile of price among the included observations is between the 10th (25th) [50th] and 25th (50th) [75th] percentiles of price among the excluded observations.

Figure A2 presents corresponding breakdowns by industry. I classify stocks by SIC code into Fama and French's 49 industries (industry definitions are provided at Ken French's website). Industry membership is broadly similar, with a few notable exceptions. Included firm-quarters are somewhat more likely than excluded firm-quarters to be in the retail or software industries, and they are somewhat less likely than excluded firm-quarters to be in the banking, pharmaceutical, or financial industries. Dropping these five industries from the analysis strengthens my results slightly.

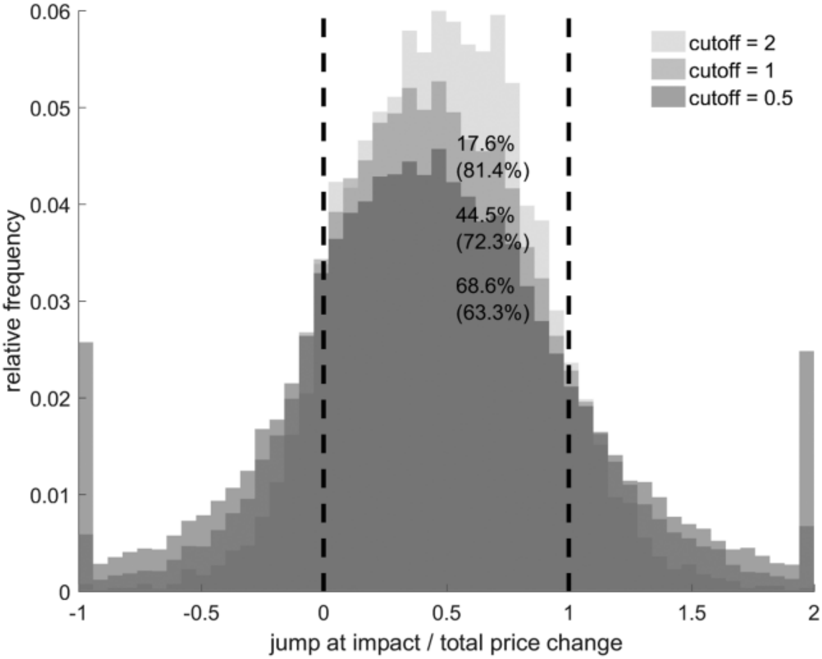


Figure A1
Range of announcement impact price jump ratio
This figure depicts the empirical distribution of earnings announcement price jump ratios for several minimum thresholds for the total price movement over the interval. The price jump ratio is given by $CAR_{it}^{(T-1,T+2)}/CAR_{it}^{(T-21,T+2)}$ for Fama and French (1992) three-factor residualized changes in log price. Stock-quarter observations enter the sample for $\left|CAR_{it}^{(T-21,T+2)}\right| > \sqrt{24}\hat{\sigma}_{it} \times$ cutoff, where $\hat{\sigma}_{it}$ is the daily return volatility from $T-42$ to $T-22$ days before an earnings announcement. For each choice of cutoff in $\{1/2, 1, 2\}$, the top number in the histogram is the proportion of the sample that exceeds the relevance cutoff. The bottom number is the fraction of the conditional ratio of post-announcement date variation to total variation that is bounded in the $[0,1]$ interval. Values exceeding -1 on the left and 2 on the right are winsorized at these respective values for visual clarity.

Table A1
Summary statistics of key variables for included and excluded data

Included	Market cap.	Price	Ret. vol.	Quoted spr.	#Analysts	*IOR*
Mean	20.89	36.55	−4.06	0.40	1.81	0.60
SD	1.81	56.73	0.52	0.67	0.92	0.27
10%	18.61	4.27	−4.68	0.05	0.69	0.16
25%	19.65	10.48	−4.41	0.09	1.10	0.44
Median	20.84	23.92	−4.09	0.21	1.95	0.66
75%	22.04	46.47	−3.73	0.43	2.48	0.79
90%	23.27	75.39	−3.39	0.87	2.94	0.88
Excluded	**Market cap.**	**Price**	**Ret. vol.**	**Quoted spr.**	**#Analysts**	***IOR***
Mean	20.56	31.52	−3.88	0.58	1.63	0.55
SD	1.94	50.20	0.60	0.96	0.98	0.28
10%	18.11	2.81	−4.59	0.05	0.00	0.10
25%	19.15	7.88	−4.30	0.10	1.10	0.35
Median	20.46	19.32	−3.93	0.26	1.79	0.62
75%	21.85	40.29	−3.51	0.61	2.40	0.77
90%	23.13	68.55	−3.10	1.41	2.89	0.87

Table presents descriptive statistics for all control variables in my study. The top panel summarizes dollar market capitalization, share price, and return volatility (all in logs) from CRSP data for the observations satisfying the $\left|CAR_{it}^{(T-21,T+2)}\right| > \sqrt{24}\hat{\sigma}_{it}$ cutoff ("included"). Additional summaries include the median end-of-minute quoted bid-ask spreads in percent from the TAQ NBBO files (one-second version); the number of analysts as the log of the maximum number of reporting analysts for each stock-event pair in Thomson Reuters I/B/E/S; and the institutional ownership ratio (*IOR*) as the fraction of shares held by 13F filing institutions at the end of the preceding calendar quarter. All control variables are lagged an additional 21 days prior to the announcement date relative to the AT proxies ($[T-42, T-22]$). The bottom table summarizes the corresponding quantities for observations not satisfying the CAR cutoff ("excluded").

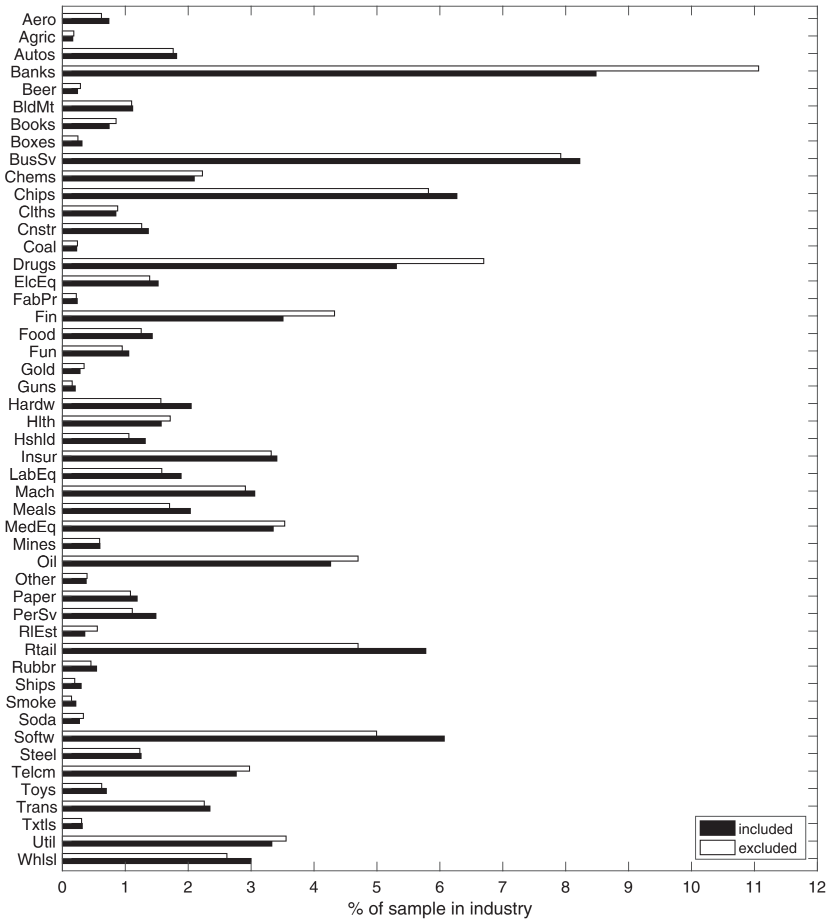


Figure A2
Industry membership for included and excluded data
The figure presents industry classifications for the included and the excluded observations. Black bars represent the share of observations in each Fama-French 49 industry category for "included" observations satisfying the $\left|CAR_{it}^{(T-21,T+2)}\right| > \sqrt{24}\hat{\sigma}_{it}$ cutoff. White bars represent the corresponding shares of "excluded" observations not satisfying this cutoff.

Appendix B. Nonrandom Variation in the Importance of Earnings Announcements

Idiosyncratic returns unrelated to earnings announcements add noise to the price jump ratio measure. This noise biases the estimated relation between AT and information acquisition if the relative importance of earnings news is correlated with algorithmic trading activity, as may be the case, for example, if algorithmic trading affects the steady-state information content of prices.

I address this potential bias using a matched pairs approach to control for the relative scale of announcement information innovations across stocks. The key grouping variables are (1) the first principal component factor of the four AT measures and (2) and the lagged log price AT instrument. For each calendar quarter, I generate pairs from these groups by minimizing the Mahalanobis distance between observations matched on lagged market capitalization and cumulative announcement impact relative to lagged stock return volatility, $\left|CAR_{it}^{(T-21,T+2)}/\hat{\sigma}_{it}\right|$. Matching on these characteristics provides a nonparametric control for the relative variation associated with earnings announcements as well as for market capitalization, which itself subsumes several liquidity-related variables.

I then conduct matched pairs analyses using (1) all stocks in the regression sample and (2) all stocks in the highest and lowest algorithmic trading activity quintiles (subject to the same restriction). In the first analysis, I match without replacement observations in the top half of algorithmic trading activity with observations in the bottom half of algorithmic trading activity. Observations with higher (lower) than median loadings on the first principal component factor of the AT measures are assigned to the "high AT" ("low AT") group. I retain the top 300 matches by distance in each quarter. In the second analysis, I assign matches between the top and bottom quintiles of the first factor loading and retain the top 100 matches by distance for each quarter. I then repeat both analyses using matches between observations within halves and quintiles of the lagged log stock price distribution. By sorting on this exogenous determinant of algorithmic trading, this second analysis provides an "instrumental variables" counterpart to the AT-sorted matched pairs studies.

Table B1 reports results from these matched pairs analyses. Panels A.1 and A.2 report summary statistics to confirm similarity between groups on the matching variables. Distributions of matched variables appear virtually identical across all specifications, and sorting variables exhibit large dispersion between groups. Importantly, stocks with a higher lagged log stock price in panel A.2 also have higher concentrations of AT activity—this relationship is the matched-pairs counterpart to the requirement of instrument relevance in IV analysis.

Panels B.1 and B.2 present pairwise differences in means and comparisons of distributions of the price jump ratio between high and low algorithmic trading groups and between high and low lagged log stock price groups, respectively. I estimate differences in means using a pairwise t-test and differences in distributions using a Wilcoxon signed rank test. To account for error in the matching step, I follow Abadie and Spiess (2016) and cluster or resample by security, month, and match index for t-tests and Wilcoxon signed-rank tests, respectively.[30] Consistent with Table 2, the high AT group has significantly larger price jump ratios on average, both statistically and economically. These differences are especially large for pairings between extreme AT quintiles, which indicates that extreme AT activity has an even larger negative association with information acquisition. Sorting on the lagged stock price instrument delivers similar results with the interpretation of the

[30] I construct three-way clustered variance-covariance matrices for the Wilcoxon signed rank test following Cameron, Gelbach, and Miller (2011) and in the spirit of Rosner, Glynn, and Lee (2006). Specifically, I first resample 1,000 times with match index clusters, month clusters, and security clusters. I then resample 1,000 times with match index-month clusters, match index-security clusters, and month-security clusters. Finally I resample 1,000 times clustering on all three variables. I calculate the three-way clustered variance matrix of the test statistic as the sum of variances on the single clusters and triple cluster minus the sum of variances on the double clusters.

Table B1
Determinants of announcement price impact: Matched pairs

A.1. Matched sample properties

Matched pairs (halves 2-1)

	Low AT				High AT			
Group Variable	Impact	Market cap.	AT PCF$_1$	*lprice*	Impact	Market cap.	AT PCF$_1$	*lprice*
Mean	1.75	20.95	−0.62	2.64	1.75	20.95	0.78	3.51
SD	0.68	1.42	0.64	0.89	0.68	1.42	0.53	0.77
25%	1.25	19.90	−0.94	2.05	1.25	19.91	0.38	3.00
50%	1.57	20.89	−0.47	2.70	1.57	20.88	0.70	3.54
75%	2.06	21.88	−0.15	3.29	2.06	21.88	1.09	4.04

Matched pairs (quintiles 5-1)

	Low AT				High AT			
Group Variable	Impact	Market cap.	AT PCF$_1$	*lprice*	Impact	Market cap.	AT PCF$_1$	*lprice*
Mean	1.75	20.54	−1.28	2.02	1.75	20.55	1.26	3.57
SD	0.68	1.34	0.55	0.80	0.69	1.33	0.42	0.78
25%	1.26	19.54	−1.58	1.48	1.25	19.54	0.95	3.03
50%	1.57	20.41	−1.16	2.06	1.57	20.42	1.19	3.57
75%	2.04	21.45	−0.87	2.58	2.04	21.45	1.50	4.11

B.1. Matched pairs comparison tests (means and medians)

Matched pairs (halves 2-1)		High AT	Low AT	High-Low AT test
jump$^{(21)}$	Mean	0.50	0.44	0.07*** (6.07)
	Median	0.49	0.42	6.90*** (6.62)

Matched pairs (quintiles 5-1)		High AT	Low AT	High-Low AT test
jump$^{(21)}$	Mean	0.49	0.39	0.10*** (5.06)
	Median	0.48	0.37	6.14*** (5.80)

A.2. Matched sample properties

Matched pairs (halves 2-1)

	Low lagged log stock price				High lagged log stock price			
Group Variable	Impact	Market cap.	AT PC$_1$	*lprice*	Impact	Market cap.	AT PC$_1$	*lprice*
Mean	2.08	20.83	−0.43	2.55	2.07	21.02	0.69	3.72
SD	1.06	1.01	0.85	0.55	1.05	0.95	0.69	0.45
25%	1.33	20.17	−0.93	2.27	1.32	20.47	0.23	3.39
50%	1.77	20.66	−0.40	2.68	1.74	20.97	0.69	3.65
75%	2.47	21.34	0.13	2.94	2.46	21.45	1.16	3.98

Matched pairs (quintiles 5-1)

	Low lagged log stock price				High lagged log stock price			
Group Variable	Impact	Market cap.	AT PC$_1$	*lprice*	Impact	Market cap.	AT PC$_1$	*lprice*
Mean	2.13	20.02	−1.08	1.63	2.03	21.44	0.95	4.29
SD	1.12	0.85	0.80	0.46	1.09	0.67	0.61	0.35
25%	1.34	19.47	−1.59	1.38	1.29	20.99	0.56	4.09
50%	1.81	19.84	−1.04	1.73	1.70	21.53	0.96	4.22
75%	2.56	20.35	−0.56	1.95	2.36	21.91	1.35	4.42

(continued)

Table B1
Continued

B.2. Matched pairs comparison tests (means and medians)

Matched pairs (halves 2-1)		High *lprice*	Low *lprice*	High-Low *lprice* Test
$jump^{(21)}$	Mean	0.51	0.48	0.03*** (2.73)
	Median	0.50	0.46	3.24*** (3.65)
Matched pairs (quintiles 5-1)		**High *lprice***	**Low *lprice***	**High-Low *lprice* Test**
$jump^{(21)}$	Mean	0.52	0.45	0.07*** (3.88)
	Median	0.51	0.43	4.21*** (3.65)

This table presents results from four matched pairs comparisons. For each calendar quarter, I generate matched pairs of observations by minimizing the Mahalanobis distance between lagged market capitalization and cumulative announcement impact, $\left|CAR_{it}^{(T-21,T+2)}/\hat{\sigma}_{it}\right|$. In the first set of matched pairs analyses, I construct matches between "high AT" and "low AT," groups, where high (low) AT observations feature above-median (below-median) factor loadings on the first principal component factor of the AT measures. Of matched pairs, I retain the 300 pairs in each quarter with the smallest distances. The bottom subpanel restricts the matching procedure to all observations in quintiles five ("very high AT") and one ("very low AT") and retains the 100 observations in each group with the minimum matching distances. For both matching groups, I consider only observations with cumulative announcement impacts exceeding the cutoff value established in the main text.

Panel A.1 reports summary statistics for matching variables and the algorithmic trading proxy for each matched pair group. Panel B.1 presents pairwise comparisons of means and medians between price jump ratios for high and low AT proxy groups. "High-Low AT" estimates differences in means by regressing pairwise differences on a constant and differences in distributions using the Wilcoxon signed-rank test (bootstrapped with 1,000 replications). Panels A.2 and B.2 repeat these analysis using matches between high lagged price and low lagged price groups. Standard errors are clustered by match index, security, and month and are reported in parentheses.

lagged stock price now acting as an exogenous treatment. High lagged stock price securities in the top-and-bottom-half matches have bigger price jump ratios holding fixed for market capitalization and cumulative announcement impact, and much higher lagged stock price securities in the quintile matches deliver even larger differences in information acquisition.

References

Abadie, A., and J. Spiess. 2016. Robust post-matching inference. Working Paper.

Alford, A., J. Jones, R. Leftwich, and M. Zmijewski. 1993. The relative informativeness of accounting disclosures in different countries. *Journal of Accounting Research* 31:183–223.

Angel, J. J. 1997. Tick size, share prices, and stock splits. *Journal of Finance* 52:655–81.

Asquith, P., P. Healy, and K. Palepu. 1989. Earnings and stock splits. *Accounting Review* 64:387–403.

Back, K. 1992. Insider trading in continuous time. *Review of Financial Studies* 5:387–409.

Bai, J., T. Philippon, and A. Savov. 2016. Have financial markets become more informative? *Journal of Financial Economics* 122:625–54.

Baker, M., R. Greenwood, and J. Wurgler. 2009. Catering through nominal share prices. *Journal of Finance* 64:2559–90.

Baker, M., J. C. Stein, and J. Wurgler. 2003. When does the market matter? Stock prices and the investment of equity-dependent firms. *Quarterly Journal of Economics* 118:969–1005.

Baldauf, M., and J. Mollner. 2017. High-frequency trading and market performance. Working Paper.

Ball, R., and P. Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6:159–78.

Barclay, M. J., and J. B. Warner. 1993. Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics* 34:281–305.

Baumol, W. J. 1965. *The stock market and economic efficiency*. New York, NY: Fordham University Press.

Brennan, M. J., and T. E. Copeland. 1988. Stock splits, stock prices, and transaction costs. *Journal of Financial Economics* 22:83–101.

Brogaard, J., T. Hendershott, and R. Riordan. 2014. High-frequency trading and price discovery. *Review of Financial Studies* 27:2267–306.

Brunnermeier, M. K. 2005. Information leakage and market efficiency. *Review of Financial Studies* 18:417–57.

Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2011. Robust inference with multiway clustering. *Journal of Business & Economic Statistics* 29:238–49.

Carrion, A. 2013. Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets* 16:680–711.

Chakrabarty, B., P. C. Moulton, and X. F. Wang. 2017. Attention effects in a high-frequency world. Working Paper.

Chen, Q., I. Goldstein, and W. Jiang. 2007. Price informativeness and investment sensitivity to stock price. *Review of Financial Studies* 20:619–50.

Collin-Dufresne, P., and V. Fos. 2015. Do prices reveal the presence of informed trading? *Journal of Finance* 70:1555–82.

Conroy, R. M., R. S. Harris, and B. A. Benet. 1990. The effects of stock splits on bid-ask spreads. *Journal of Finance* 45:1285–95.

Dow, J., and G. Gorton. 1997. Stock market efficiency and economic efficiency: Is there a connection? *Journal of Finance* 52:1087–129.

Durnev, A., R. Morck, B. Yeung, and P. Zarowin. 2003. Does greater firm-specific return variation mean more or less informed stock pricing? *Journal of Accounting Research* 41:797–836.

Easley, D., M. O'Hara, and G. Saar. 2001. How stock splits affect trading: A microstructure approach. *Journal of Financial and Quantitative Analysis* 36:25–51.

Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25:383–417.

Fama, E. F., L. Fisher, M. C. Jensen, and R. Roll. 1969. The adjustment of stock prices to new information. *International Economic Review* 10:1–21.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47:427–65.

Foster, F. D., and S. Viswanathan. 1993. The effect of public information and competition on trading volume and price volatility. *Review of Financial Studies* 6:23–56.

Foucault, T., J. Hombert, and I. Roşu. 2016. News trading and speed. *Journal of Finance* 71:335–82.

Foucault, T., A. Röell, and P. Sandås. 2003. Market making with costly monitoring: An aalysis of the SOES Controversy. *Review of Financial Studies* 16:345–84.

Freeman, R. N. 1987. The association between accounting earnings and security returns for large and small firms. *Journal of Accounting and Economics* 9:195–228.

Gerig, A. 2015. High-frequency trading synchronizes prices in financial markets. Working Paper.

Grossman, S. J., and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70:393–408.

Han, J., M. Khapko, and A. S. Kyle. 2014. Liquidity with high-frequency market making. Research Paper, Swedish House of Finance.

Harris, L. 2013. What to do about high-frequency trading. *Financial Analysts Journal* 69:6–9.

Hasbrouck, J. 1991a. Measuring the information content of stock trades. *Journal of Finance* 46:179–207.

———. 1991b. The summary informativeness of stock trades: An econometric analysis. *Review of Financial Studies* 4:571–95.

———. 1993. Assessing the quality of a security market: A new approach to transaction-cost measurement. *Review of Financial Studies* 6:191–212.

Heflin, F., K. R. Subramanyam, and Y. Zhang. 2003. Regulation FD and the financial information environment: Early evidence. *Accounting Review* 78:1–37.

Hendershott, T., C. M. Jones, and A. J. Menkveld. 2011. Does algorithmic trading improve liquidity? *Journal of Finance* 66:1–33.

Hirschey, N. 2017. Do high-frequency traders anticipate buying and selling pressure? Working Paper.

Holden, C. W., and A. Subrahmanyam. 1992. Long-lived private information and imperfect competition. *Journal of Finance* 47:247–70.

Kirilenko, A. A., and A. W. Lo. 2013. Moore's law versus Murphy's law: Algorithmic trading and its discontents. *Journal of Economic Perspectives* 27:51–72.

Kleibergen, F., and R. Paap. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133:97–126.

Kyle, A. S. 1985. Continuous auctions and insider trading. *Econometrica* 53:1315–35.

Lakonishok, J., and B. Lev. 1987. Stock splits and stock dividends: Why, who, and when. *Journal of Finance* 42:913–32.

Lewis, M. 2014. *Flash boys: A Wall Street revolt*. New York, NY: W. W. Norton & Company.

McNichols, M. F. 1984. *The anticipation of earnings in securities markets*. PhD Thesis, University of California, Los Angeles.

Merton, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29:449–70.

Meulbroek, L. K. 1992. An empirical analysis of illegal insider trading. *Journal of Finance* 47:1661–99.

Morck, R., B. Yeung, and W. Yu. 2000. The information content of stock markets: Why do emerging markets have synchronous stock price movements? *Journal of Financial Economics* 58:215–60.

Morse, D. 1981. Price and trading volume reaction surrounding earnings announcements: A closer examination. *Journal of Accounting Research* 19:374–83.

O'Hara, M. 2003. Presidential address: Liquidity and price discovery. *Journal of Finance* 58:1335–54.

Roll, R. 1988. R^2. *Journal of Finance* 43:541–66.

Rosner, B., R. J. Glynn, and M.-L. T. Lee. 2006. The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics* 62:185–92.

Rozeff, M. S. 1998. Stock splits: Evidence from mutual funds. *Journal of Finance* 53:335–49.

Stiglitz, J. 2014. Tapping the brakes: Are less active markets safer and better for the economy? Working Paper.

U.S. Securities and Exchange Commission. 2010. Concept release on equity market structure. Technical Report.

U.S. Securities and Exchange Commission. 2014. Equity market structure literature review, part II: High frequency trading. Technical Report, Securities and Exchange Commission.

Weld, W. C., R. Michaely, R. H. Thaler, and S. Benartzi. 2009. The nominal share price puzzle. *Journal of Economic Perspectives* 23:121–42.

Yang, L., and H. Zhu. 2017. Back-running: Seeking and hiding fundamental information in order flows. Working Paper.

Yao, C., and M. Ye. 2017. Why trading speed matters: A tale of queue rationing under price controls. Working Paper.

Zhang, S. 2017. Need for speed: An empirical analysis of hard and soft information in a high frequency world. Working Paper.

MUTUAL FUND FLOWS AND SEASONALITIES IN STOCK RETURNS

Moritz Wagner and John Byong-Tek Lee
University of Auckland Business School
Auckland, New Zealand

This Version: February 2015

ABSTRACT

In this paper, we propose a flow-based explanation for a long-standing anomaly in empirical finance – the Sell in May effect. We find that aggregate mutual fund flow exhibits a similar seasonality as stock returns. Given that flow can affect contemporaneous stock returns, the Sell in May effect becomes insignificant in standard statistical tests after controlling for flow. Flow explains about 54% of the variation in excess returns over the winter months. We also find that flow helps explaining the abnormally high returns of small-cap stocks in January.

I Introduction

Numerous seasonalities in stock returns have been documented in previous research. Among the widely cited anomalies are the January or turn-of-the year effect, the turn-of-the-month and the day-of-the week effect. The old saw "Sell in May and go away", known as the Halloween effect, represents probably the most pervasive calendar anomaly. It suggests that stock returns during the winter months should be higher than during the summer months. Bouman and Jacobsen (2002) find higher returns in 36 out of 37 markets in the November-April period than in the May-October period. Jacobsen et al. (2005), Jacobsen and Visaltanachoti (2009), Jacobsen and Zhang (2012) and Andrade et al. (2013) show that this return pattern is also present out of sample, is unrelated to other anomalies, and if anything, has become more pronounced over the recent past. As put forward by Jacobsen and Visaltanachoti, the Sell in May effect has survived all the usual controls and robustness checks up to the present day. It represents a puzzle yet to be explained. The explanations offered in literature such as general investor behaviour and a change in risk aversion due to vacations, Seasonal Affective Disorder (SAD) or temperature can at best partially explain this effect (see Bouman and Jacobsen, 2002; Kamstra et al., 2003, 2009; Cao and Wei, 2005; Jacobsen and Marquering, 2008, 2009; Hong and Yu, 2009). We argue that this empirical pattern is driven by a simple mechanism: mutual fund flows.

The body of literature on mutual fund flows and institutional trading documents that stock returns are contemporaneously correlated with flows into funds (see Chan and Lakonishok, 1993, 1995; Warther, 1995; Edelen and Warner, 2001; Rakowski and Wang, 2009). Coval and Stafford (2007) and Lou (2012) show strong price-pressure effects from flow-induced trading. In addition to the claim that flows cause returns, there are other competing hypotheses to explain the co-movement such as feedback trading, sentiment or simply information revelation. Without favouring one of these hypotheses, given that flow can affect contemporaneous stock returns, it is natural to ask whether flows can also cause seasonalities in stock returns. A visual examination of monthly net flow into US mutual funds reveals a distinct pattern that clearly supports the Sell in May wisdom. On average, realised flow into mutual funds is substantially larger during the winter months compared to the summer months. And in most years over the sample period, stock returns are higher during winter months than during summer months. However, in years with summer flow exceeding winter flow, the Sell in May effect is negative. The usual statistical tests first confirm a Sell in May effect separated from a potential January effect and after controlling for common risk factors. But the seasonal dummy variable drops out when we control for contemporaneous and lagged flow. Average net flow during the winter months in excess over the average flow during the remainder of the year explains about half of the variation of excess returns during the winter months. In addition, we find that flow provides a stronger explanation for the January effect than other explanations discussed in prior research.

The paper is organized as follows. Section II describes the data and methodology. Section III presents the empirical findings and we conclude in section IV.

II Data and Methodology

To estimate the seasonal return pattern, we use CRSP value- and equal-weighted stock market index returns (NYSE + AMEX + NASDAQ), as well as total returns from the S&P 500 index available from Wharton Research Data Services. We obtain monthly net flow for US-based mutual funds that invest in domestic equities and have more than 50 million USD in assets under management from Morningstar. Net flow is estimated from a fund's prior month assets, current month assets and the monthly total return.

$$Flow_{it} = TNA_{it} - TNA_{i,\,t-1}(1+r_{it}) \tag{1}$$

TNA_{it} is a fund's monthly total net asset and r_{it} is the fund's total return. Hence, equation (1) is simply the difference between current and prior month's assets that is not accounted for by monthly total return. The sample period is from January 1995 through December 2014.

We use standard regression analysis to test for seasonalities in stock returns:

$$r_t = \mu + \beta_1 Jan_t + \beta_2 Hal_t + \varepsilon_t \tag{2}$$

where r_t is the return on the stock index for month t, μ is a constant and ε_t is the usual error term. β_1 and β_2 estimate the January and the Halloween effect. Jan_t and Hal_t are seasonal dummy variables that take the value of 1 for January and November to April periods respectively, and 0 otherwise. This equation is equivalent to a simple *t*-test for differences between means. Using this regression however allows us to include other variables, which is vital for the claim of this paper. Table 1 reports summary statistics of the funds in our sample. There is a clear increasing trend visible in the number of funds and the percentage of the stock market held by funds.[1] If flow is truly the underlying force of the Sell in May effect, these trends might provide an explanation why the anomaly has become more rather than less pronounced in recent years.

[1] Percentage of the stock market held by mutual funds is slightly overstated here because we do not consider cash holdings separately. But the figures are mainly comparable to those reported in prior research.

III Results

A. The Sell in May Effect

Panel A of Figure 1 shows average net flow of US mutual funds by month. We first sum net flow over all funds before calculating the monthly average. Hence, Figure 1 presents a market-wide aggregate or the monthly average of one giant fund.

Insert Figure 1 here

Average net flow in the period November-April is substantially larger than in the period May-October, with more money being withdrawn on average than invested during the month September. In Panel B we plot net flow together with average monthly returns. As can be seen, average returns tend to be higher in months with higher average flow and vice versa. This plot provides a graphical depiction of the co-movement between fund flows and returns that has been documented in literature. Returns over the period May-October tend to be rather modest between -1% and 1%. In the same months, average flow is hardly over one billion USD. However, during the period November-April returns are approximately 2% plus and average flow is about three and up to seven billion USD. Corresponding with the January effect being predominantly found among small cap stocks, the equally-weighted market index peaks in January which gives more weight to small firms. During the other months of the year the market proxies are fairly close. January, together with April, is also the month with the highest flow measure.

Panel C plots winter excess stock returns and fund flows over time. The solid line is the cumulative return on the CRSP value-weighted market index during November-April minus the return during May-October. This proxy for the Sell in May effect seems to vary from year to year which mainly supports the findings of Jacobsen and Zhang (2013). However, in 15 out of 20 years returns during the winter months are higher than during the summer months. And the same is true for fund flows. The dashed line shows normalised flow over the winter months in excess of the summer months. In most years, flow during winter is higher than during summer, as indicated by a positive value. Remarkably, in all years where this is not the case, i.e. summer flow is higher than winter flow, the Sell in May effect is also negative. The only exception is 2005. The correlation between the two series is 0.73 (p-value = 0.0002).

Turning now to statistical tests, Table 2 reports estimation results from equation (2). As in previous research we find strong seasonalities in stock returns that are statistically and economically significant. The somewhat hefty turn-of-the-year effect is only present in the equally-weighted index, and thus predominantly among small-cap stocks. The dummy variables in columns 5 and 6 treat the January and Sell in May effect as separate seasonalities, i.e. the Sell in May dummy is 1 in the period November-April, except January and 0 otherwise. The last column contains the results of a regression with only the Sell in May dummy defined as 1 in the period November-April including January. In a nutshell, Table 2 resembles the empirical regularity documented in earlier research we are trying to explain in the following section.

Insert Table 2 here

If mutual fund flows can affect contemporaneous stock returns and given that fund flows exhibit a certain pattern (Figure 1), it is natural to ask whether fund flows can help explaining well-known seasonalities in stock returns. To test this possibility, Panel A of Table 3 report estimation results for different variants of equation (2). More specifically, the regression including all explanatory variables is as follows:

$$r_t = \mu + \beta_1 Hal_t + \beta_2 PE_{t-1} + \beta_3 Flow_t + \beta_4 Flow_{t-1} + \beta_5 Flow_{t-2} + \beta_6 Flow_{t-3} + \varepsilon_t \quad (3)$$

where r_t is the monthly return on the S&P 500 index. The first two variables on the right-hand side are the same as described above. PE_{t-1} is the price-earnings ratio of the market index at the end of the previous month. We also use DY_{t-1} below, which is the dividend yield of the market index at the end of the previous month instead of PE_{t-1}. Both have been found to be helpful predicting stock returns but are highly correlated (-0.83). Hence we cannot include both at the same time to avoid multicollinearity issues. This is our attempt to test variables that are related with stock returns but have so far not been considered for the Sell in May effect. $Flow_t$ is the aggregate mutual fund flow in month t estimated by equation (1) and normalised by the value of the stock market (NYSE + AMEX + NASDAQ) at the end of the previous month. In an attempt to capture both effects associated with the price pressure argument, that flow drives stock prices away from their fundamental values and a corresponding but lagged reversal, we also include lagged flow.[2] We do not re-examine temperature and the Onset/Recovery (aka SAD) variable from Kamstra et al. here. Both have been widely debated in literature as a potential cause for the seasonal anomaly in stock returns driven by mood

[2] In unreported tests, regressions of market returns on flow show that contemporaneous flow is positively and the first lag is negatively related to returns (on the one and five percent significance level respectively). We include up to three lags to capture as much as possible of the flow effects.

changes of investors because of the variation in daylight and temperature. However, the evidence in favour for these two explanations is not convincing (see Jacobsen and Marquering, 2008, 2009). This is partly due to their high correlation with the Halloween indicator which is 0.88 and -0.68 for temperature and SAD respectively, which makes it difficult to test the joint effects. By contrast, the correlation between flow and the Halloween indicator is only 0.18.[3] Column two of Panel A in Table 3 shows the strong and positive relation between stock returns and concurrent flow on the macro level that is known since Warther (1995). The coefficient on *Flow* is 1.67 with a *t*-statistic of 3.71. What is new is that the Sell in May dummy becomes insignificant, i.e. after accounting for flow there is no winter-summer seasonality in stock returns left in a statistical sense.

Insert Table 3 here

Furthermore, column three shows that contemporaneous flow is positively related to stock returns (*t*-statistic = 5.29) and consistent with a reversal of the price pressure effect, lagged flow is negatively related to stock returns (*t*-statistic = -2.43 for $Flow_{t-1}$). The estimate on flow is mainly unaffected when we include the price-earnings ratio or the dividend yield as shown in columns four and five. However, only the former is more than two standard errors away from zero in our tests. The coefficient on PE_{t-1} is -0.17 with a *t*-statistic of -2.71 and the coefficient on DY_{t-1} is 1.50 with a *t*-statistic of 1.62.

Panel B reports results of a regression in which the dependent variable is the six-month return over the period November-April in

[3] The variance inflation factors are 1 or very close to 1 in all regressions that include flow.

excess of the six-month return over May-October. The explanatory variable is the six-month flow during the same winter period in excess of the summer months. This univariate test explains 54% of the variation in the Sell in May effect. The coefficient estimate of 4.21 with a *t*-statistic of 4.89 implies that for an average excess winter flow of 1.39% (normalised) or USD 24 billion (absolute), the six-month excess return is about 5.84%. This estimate is very close to the average difference between November-April and May-October returns reported in Jacobsen and Zhang (2012). Over the past 50 years they find an average difference of 6.25%.

To address the question of causation, the first three columns of Table 4 report results of regressing market returns on expected and unexpected concurrent flow and the Halloween indicator. Expected flow is estimated in a first step using three lags of flow and three lags of returns. Unexpected flow is actual flow minus expected (predicted) flow.[4] The Sell in May dummy becomes only insignificant when we account for the unexpected component of flow. The coefficient on unexpected flow is 3.37 and highly significant with a *t*-statistic of 5.45, while expected flow is not significant in statistical terms. We have hoped to see the opposite as causation is generally accepted if the expected rather the unexpected component of a variable is driving the results. If only the unexpected component seems to matter, doubts are left because both the dependent and independent variable could be affected by an unknown variable causing simple correlation. Hence, more tests are required to clearly separate between the two possibilities.

[4] The results reported in Table 4 are insensitive to variations in the first step, i.e. the number of lags or if we just include flow, e.g. using a simple AR(3) model.

Regressions four and five shed a bit more light on the flow-return relationship by regressing expected and unexpected flow on concurrent and lagged returns. These results highlight why the coefficient on the expected component of flow in the first and third column is statistically insignificant. Expected flow lags return, while concurrent return is only related to unexpected flow. Based on this we can further infer that only the expected component of flow is consistent with the feedback-trader hypothesis which predicts that flows must lag returns.[5]

B. The January Effect

Since flow spikes in January (Figure 1) and January falls into the winter period, we address the obvious question whether flow helps to explain the January effect next. This empirical regularity refers to abnormally high stock returns in January, first documented by Wachtel (1942). Keim (1983), Rozeff and Kinney (1976) and Reinganum (1983) find it to be mainly present among small-cap firms. Schwert (2003) shows that the effect might have become smaller since its discovery, but the effect has not disappeared. And thus, the debate continues to date. Several explanations have been proposed to explain this anomaly, but empirical results are mixed. For example, Rozeff and Kinney (1976), Chang and Pinegar (1988, 1989, 1990), Rogalski and Tinic (1986), Keamer (1994) and Sun and Tong (2010) suggest that the January effect is due to the seasonality in risk or in the compensation for risk. Tinic and West (1984) find the mean-variance trade-off is only present in January. Haugen and

[5] This lag could be anything from picking up the phone or the order of months but there must be a nonzero lag between flows and returns, Warther (1995).

Lakonishok (1987) and Lakonishok et al. (1991) propose a window dressing hypothesis in which institutional investors try to make their portfolios look better by selling stocks with large losses at the end of the year. Branch (1977), Dyl (1977), Reinganum (1983), Jones et al. (1991) and Poterba and Weisenbenner (2001) attribute the effect to tax-loss selling in December and corresponding purchase activities in January. Chen and Singal (2004) demonstrate that tax-loss selling is the main driver behind this anomaly. We do not discuss prior research in more detail here, because the literature is large and surveys can be found elsewhere (e.g. Singal, 2004).

To test our flow-based explanation we begin by repeating the analysis from above but include a January indicator. Since we do not find a January effect in the value-weighted CRSP stock market index, all tests below are based on the equally weighted index. The first column in Table 5 reports an average January effect of 2.73% with a *t*-statistic = 1.95. If we include our flow variables the January indicator is completely subsumed. The coefficient on *Flow* is 4.93 with a *t*-statistic of 6.66. Again, consistent with a lagged reversal of the price pressure effect, lagged flow is negatively related to stock returns. The *t*-statistics on the first, second and third lags of flow are -1.77, -0.76 and -2.63 respectively. The estimates on flow are essentially unaffected when the price-earnings ratio and the dividend yield are included, as shown in columns three and four. Column five shows again that it is the unexpected component of flow that is driving the results with a coefficient of 4.69 and a *t*-statistic of 6.23. The coefficient on expected flow is not statistically significant due to the same reasons as discussed above.

Insert Table 5 here

Next, we provide a more direct test whether flow helps explaining the January regularity alongside other alternatives. For this, we first estimate abnormal return and flow in January with the following regressions:

$$r_t = \mu_t + \beta_1 r_{t-1} + \beta_2 Jan_{1995} + \beta_3 Jan_{1996} + \ldots + \beta_{21} Jan_{2014} + \varepsilon_t \qquad (4)$$

$$Flow_t = \mu_t + \beta_1 Flow_{t-1} + \beta_2 Jan_{1995} + \ldots + \beta_{21} Jan_{2014} + \varepsilon_t \qquad (5)$$

where r_t is the return on the equally-weighted CRSP stock market index in month t. $Flow_t$ is the aggregate net flow of our sample funds standardised by the value of the stock market (NYSE + AMEX + NASDAQ) in the previous month. Lagged values are included to take care of serial correlation. β_2 - β_{21} represent the abnormal return and flow in January estimated for each year over the sample period.

Insert Table 6 here

Table 6 shows the January effect is positive and statistically significant in 12 out of 20 years. With few exceptions (e.g. 2002 and the GFC), we also see that in years with strong and positive (negative) abnormal flow, the January effect is large and positive (negative). Table 7 reports results of a regression in which the dependent variable is the January excess return estimated with equation (4). The only explanatory variable in the first test is the estimated abnormal flow in January. In this univariate test, flow is positively related to the January effect with a coefficient of 2.87 and a t-statistic = 4.97. The diagnostics identify two influential points, the year 2001, which has an extreme January return, and 2009 the height of the global financial crisis. Adjusting for these events would increase the variation explained by abnormal flow to over 27%, but due to the relatively short time period we only report unadjusted results. The conclusion about flow is not affected if we take corrective measures. If anything, the estimate on flow becomes more significant. The second model includes proxies for alternative explanations for the January effect suggested in previous research. PTS_{t-1} is the maximum potential tax-loss selling at the end of a year. It is defined as the percentage decrease in stock price from the highest price during the year to mid-December, usually December 15. If there was no trading on this day, we take the price from the previous trading day. This is close enough to the end of the year and allows sufficient time for tax-related selling. This measure mainly follows prior research and enables us to make meaningful comparisons (e.g. see Reinganum, 1983; Chen and Singal, 2004).

$$PTS = \frac{\sum_{i=1}^{n} \left(\frac{price_{it,Dec.}}{price_{it,High}}\right) - 1}{n} \qquad (6)$$

By design, *PTS* might also pick up window dressing activities from institutional investors. Since we are not interested to disentangle the two competing explanations, we rather appreciate that *PTS* sort of captures both possibilities. We include liquidity and volume as a general source for the January seasonality. Abnormally high volume usually occurs with informed trading and as such, is consistent with the information release hypothesis. However, the entry of noise traders may also affect volume. Another reason to include volume here is to avoid the possibility that flow is just volume in disguise. Standard deviation takes care of the risk argument. Both of these measures are estimated in relative terms, i.e. January dollar volume and January standard deviation relative to volume and standard deviation over the previous six months.

The choice of the time period, whether it is the previous six or eleven months, does not affect the results.

Insert Table 7 here

The results reported in Table 7 show that excess flow helps explaining the January effect alongside alternative explanations discussed in literature. The coefficient of *Flow* is 3.17 with a *t*-statistic of 3.55. As is suggested, *PTS* is positively related to the January effect with a coefficient of -16.72, but it is not significant in statistical terms.[6] The estimates on *Vol* and *Std* are also not more than two standard errors away from zero. However, if we control for the GFC the *t*-statistic of *PTS* is -2.29, while the signs for *Vol* and *Std* change and become more in line with the general risk-return trade-off. Yet, both variables remain insignificant while the amount of variation in the January effect explained increases to 30%. Regardless if we control for potential data issues or not the estimate on flow hardly changes, with a persistent *t*-statistic of more than 3.0. Our results indicate, that flow is related to the January anomaly in stock returns. Even if flow may not be the sole driver behind the effect, it certainly is one element of the explanation.

IV Conclusion

Consistent with prior research we find a statistically and economically significant difference between returns during the winter and the summer months. We provide a flow-based explanation for this long-standing anomaly that challenges basic financial theory. Specifically, we show the Sell in May effect is

[6] PTS is between -1 and 0 by construction.

positive (negative) in years where flow during the winter months is higher (lower) than during the summer months. After controlling for mutual fund flows the Sell in May effect becomes insignificant. Excess fund flow explains almost half of the variation in the Sell in May effect. We also find that flow helps explaining the well-known January effect.

Our results build on the contemporaneous relationship between returns and flow. If flow provides an explanation for seasonalities in stock returns an interesting question remains, what drives seasonalities in fund flows?

REFERENCES

Andrade, Sandro C., Chhaochharia, Vidhi, & Fuerst, Michael E. (2013). "Sell in May and Go Away" Just Won't Go Away. *Financial Analysts Journal, 69*(4), 94-105.

Bouman, Sven, & Jacobsen, Ben. (2002). The Halloween Indicator, "Sell in May and Go Away": Another Puzzle. *American Economic Association, 92*(5), 1618-1635.

Cao, Melanie, & Wei, Jason. (2005). Stock Market Returns: A Note on Temperature Anomaly. *Journal of Banking & Finance, 29*(6), 1559-1573.

Chan, Louis, & Lakonishok, Josef. (1993). Institutional Trades and Intraday Stock Price Behavior. *Journal of Financial Economics*, *33*, 173-199.

Chan, Louis, & Lakonishok, Josef. (1995). The Behavior of Stock Prices Around Institutional Trades. *Journal of Finance*, *50*, 1147-1174.

Chang, Eric, & Pinegar, Michael. (1988a). Does the Market Reward Risk in Non-January Months? *Journal of Portfolio Management*, *15*(1), 55-57.

Chang, Eric, & Pinegar, Michael. (1988b). A Fundamental Study of the Seasonal Risk-Return Relationship: A Note. *Journal of Finance, 43*(4), 1035-1039.

Chang, Eric, & Pinegar, Michael. (1989). Seasonal Fluctuations in Industrial Production and Stock Market Seasonal. *Journal of Financial and Quantitative Analysis, 24*, 59-74.

Chang, Eric, & Pinegar, Michael. (1990). Market Seasonal and Prespecified Multifactor Pricing Relations. *Journal of Financial and Quantitative Analysis, 25*, 517-533.

Chen, Honghui, & Singal, Vijay. (2004). All Things Considered, Taxes Drive the January Effect. *Journal of Financial Research, 27*(3), 351-372.

Coval, Joshua, & Stafford, Erik. (2007). Asset Fire Sales (and Purchases) in Equity Markets. *Journal of Financial Economics, 86*, 479-512.

Dyl, Edward A. (1977). Capital Gains Taxation and Year-end Stock Market Behaviour. *Journal of Finance, 32*(1), 165-175.

Edelen, Roger M., & Warner, Jerold B. (2001). Aggregate Price Effects of Institutional Trading: A Study of Mutual Fund Flow and Market Returns. *Journal of Financial Economics, 59*, 195-220.

Fama, Eugene F., & French, Kenneth R. (1993). Common Risk Factors in the Return on Bonds and Stocks. *Journal of Financial Economics, 33*, 3-53.

Haugen, Robert A., & Lakonishok, Josef. (1987). *The Incredible January Effect: The Stock Market's Unsolved Mystery*: Dow Jones-Irwin.

Hong, Harrison, & Yu, Jialin. (2009). Gone fishin': Seasonality in Trading Activity and Asset Prices. *Journal of Financial Markets, 12*(4), 672-702.

Jacobsen, Ben, Mamun, Abdullah, & Visaltanachoti, Nuttawat. (2005). Seasonal, Size and Value Anomalies. *Working Paper, Massey University, Auckland New Zealand*.

Jacobsen, Ben, & Marquering, Wessel. (2008). Is it the Weather? *Journal of Banking & Finance, 32*(4), 526-540.

Jacobsen, Ben, & Marquering, Wessel. (2009). Is it the Weather? Response. *Journal of Banking & Finance, 33*(3), 583-587.

Jacobsen, Ben, & Visaltanachoti, Nuttawat. (2009). The Halloween Effect in U.S. Sectors. *The Financial Review, 44*, 437-459.

Jacobsen, Ben, & Zhang, Cherry Y. (2012). The Halloween Indicator: Everywhere and all the Time. *Working Paper, Massey University, Auckland New Zealand*.

Jones, Steven L., Lee, Winson, & Apenbrink, Rudolf. (1991). New Evidence on the January Effect before Personal Income Taxes. *Journal of Finance, 46*(5), 1909-1924.

Kamstra, Mark J., Kramer, Lisa A., & Levi, Maurice D. (2003). Winter Blues: A SAD Stock Market Cycle. *American Economic Review, 93*(1), 324-343.

Kamstra, Mark J., Kramer, Lisa A., & Levi, Maurice D. (2009). Is it the Weather? Comment. *Journal of Banking & Finance, 33*, 578-582.

Keim, Donal B. (1983). Size-related Anomalies and Stock Return Seasonality: Further Empirical Evidence. *Journal of Financial Economics, 12*(1), 13-32.

Lakonishok, Josef, Shleifer, Andrei, Thaler, Richard, & Vishny, Robert. (1991). Window Dressing by Pension Fund Managers. *American Economic Review, 81*, 227-231.

Lou, Dong. (2012). A Flow-Based Explanation for Return Predictability. *The Review of Financial Studies, 25*, 3457-3489.

Poterba, James M., & Weisenbenner, Scott J. (2001). Capital Gains Tax Rules, Tax-loss Trading, and the Turn-of-the-year Returns. *Journal of Finance, 56*(1), 353-368.

Rakowski, David, & Wang, Xiaoxin. (2009). The Dynamics of Short-term Mutual Fund Flows and Returns: A Time-series and Cross-sectional Investigation. *Journal of Banking & Finance, 33*, 2102-2109.

Reinganum, Marc R. (1983). The Anomalous Stock Market Behaviour of Small Firms in January. *Journal of Financial Economics, 12*, 89-104.

Rogalski, Richard, & Tinic, Seha M. (1986). The January Size Effect: Anomaly or Risk Mismeasurement? *Financial Analysts Journal, 42*(6), 63-70.

Rozeff, Michael S., & Kinney, William R. (1976). Capital Market Seasonality: The Case of Stock Returns. *Journal of Financial Economics, 3*(4), 379-402.

Schwert, G. William, Harris, Milton, & Stulz, Rene M. (2003). Anomalies and Market Efficiency. In George Constantinides (Ed.), *Handbook of the Economics of Finance* (pp. 937-972): Elsevier North-Holland.

Singal, Vijay. (2004). *Beyond the Random Walk: A Guide to Stock Market Anomalies and Low-Risk Investing*. New York: Oxford University Press, Inc.

Sun, Qian, & Tong, Wilson H.S. (2010). Risk and the January Effect. *Journal of Banking & Finance, 34*(5), 965-974.

Tinic, Seha M., & West, Richard R. (1984). Risk and Return: January vs. the Rest of the Year. *Journal of Financial Economics, 13*(4), 561-574.

Wachtel, Sidney. (1942). Certain Observations on Seasonal Movement in Stock Prices. *Journal of Business, 15*, 184-193.

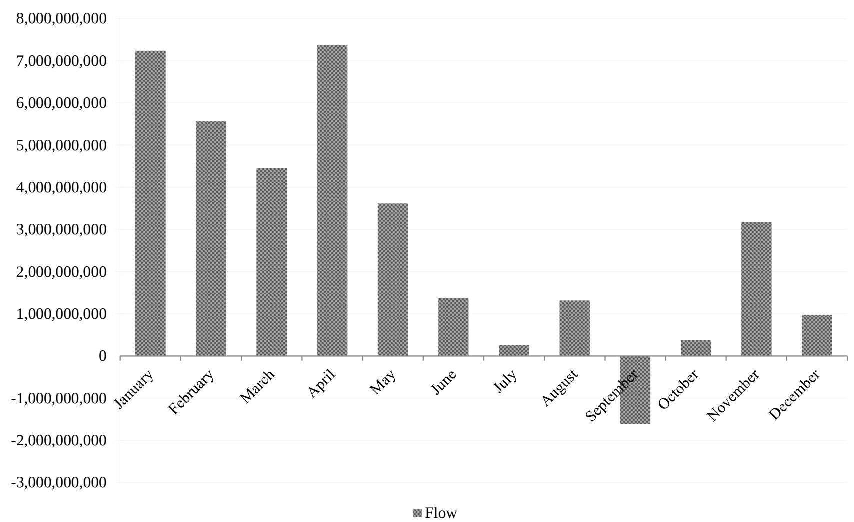
Warther, Vincent A. (1995). Aggregate Mutual Fund Flows and Security Returns. *Journal of Financial Economics, 39*, 209-235.

Zhang, Cherry Y., & Jacobsen, Ben. (2013). Are Monthly Seasonals Real? A Three Century Perspective. *Review of Finance, 17*, 1743-1785.

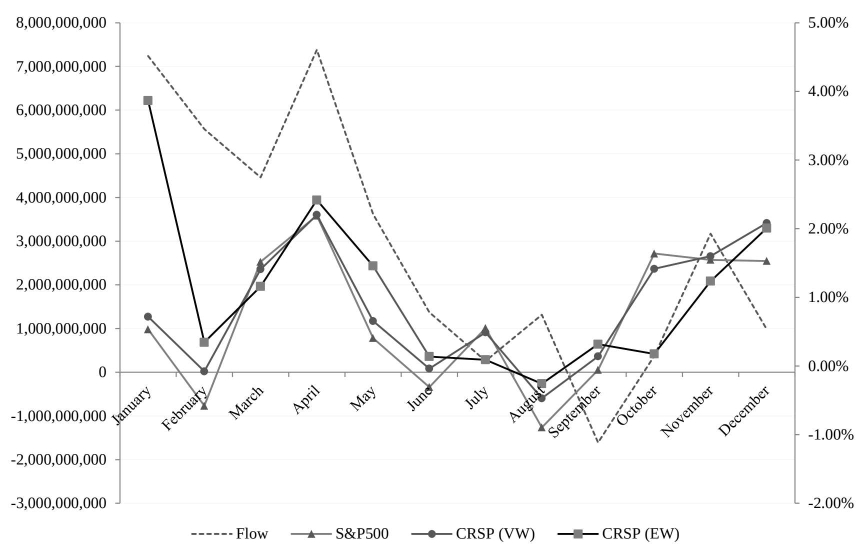
Figure 1
Mutual Fund Flows and Stock Returns

Panel A of this figure reports the average monthly flow of our sample funds (market-wide aggregate) by months. Panel B plots the same flow measure together with average monthly returns on the CRSP value- and equally-weighted stock market indices (NYSE + AMEX + NASDAQ) and the S&P 500 index. Panel C reports six-month returns on the CRSP value-weighted stock market index of the period November-April in excess over May-October and the same for mutual fund flows, normalised by the value of the market (NYSE + AMEX + NASDAQ) and scaled by 1000. The sample period is January 1995 to December 2014.

Panel A – Average monthly flow



Panel B – Average monthly flow and average monthly returns



Panel C – Winter excess fund flows and market returns

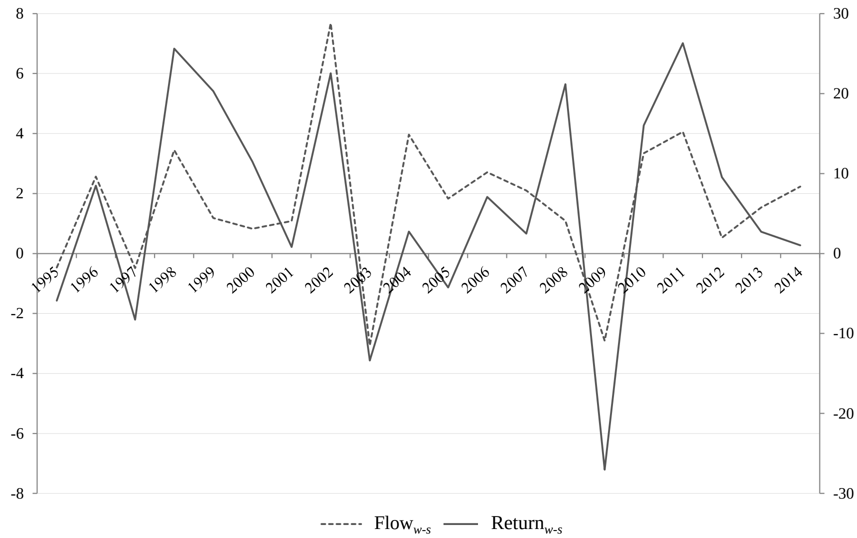


Table 1
Summary Statistics of US Equity Mutual Funds

This table reports summary statistics of all US-based mutual funds that invest in domestic equities as of the end of December in each year. The only filter we apply is that they have more than 50 million dollars in assets under management based on the most recent portfolio date. The number of funds is given in share classes. We calculate percent of market value as total net assets divided by total value of the stock market (NYSE + AMEX + NASDAQ). The sample period is from 1995 to 2014.

Year	Number of Funds	Total Net Assets ($M)	% of Market Value
1995	908	778,391	11.47%
1996	1,071	1,058,615	12.75%
1997	1,268	1,473,680	13.65%
1998	1,425	1,842,371	13.86%
1999	1,640	2,452,570	14.41%
2000	1,938	2,375,616	15.20%
2001	2,240	2,184,445	15.78%
2002	2,414	1,634,410	14.82%
2003	2,737	2,450,488	16.81%
2004	2,989	2,953,683	17.95%
2005	3,284	3,226,563	18.57%
2006	3,505	3,745,166	19.10%
2007	3,728	4,095,311	20.28%
2008	3,998	2,447,566	20.18%
2009	4,160	3,237,890	20.49%
2010	4,325	3,751,258	20.29%
2011	4,434	3,593,521	20.09%
2012	4,498	4,013,517	19.72%
2013	4,495	5,378,888	20.47%
2014	4,364	5,739,679	19.82%

Table 2
The Sell in May and January Effect

This table reports estimation results of the Sell in May or Halloween effect and the January effect. The first two rows are the value- and equally-weighted indices of all stocks listed on the NYSE, AMEX and NASDAQ. The S&P 500 index represents the 500 largest publicly traded corporations in the US. The January dummy is 1 for returns that fall into January and 0 otherwise. The Sell in May dummy (not) adjusted for January is 1 for the period November-April (including) excluding January and 0 otherwise. Mean and adjusted R^2 are reported from the regression including the January and the adjusted Sell in May dummy. The sample period is January 1995 to December 2014. *t*-statistics are reported in parentheses based on Newey-West corrected standard errors.

Market Index	Adj-R^2	Obs.	Mean	January Dummy	Sell in May Dummy (adjusted for January)	Sell in May Dummy (not adjusted for January)
CRSP (VW)	0.01	240	0.37	0.04	**1.22**	**1.03**
			(0.82)	(0.04)	**(2.19)**	**(1.88)**
CRSP (EW)	0.02	240	0.31	**3.30**	**1.26**	**1.60**
			(0.54)	**(2.38)**	**(1.76)**	**(2.36)**
S&P 500	0.01	240	0.24	-0.05	**1.17**	**0.96**
			(0.57)	(-0.05)	**(2.23)**	**(1.85)**

Table 3

Mutual Fund Flows and the Halloween Seasonality in Stock Returns

Panel A reports estimation results for different variants of equation (2) with *t*-statistics in parentheses based on Newey-West corrected standard errors. The dependent variable is the monthly return on the S&P 500 index. The Sell in May dummy, Hal_t, is 1 for the period November-April and 0 otherwise. PE_{t-1} and DY_{t-1} are the price-earnings ratio and the dividend yield of the market index at the end of the previous month. $Flow_t$ is the estimated monthly net flow (market-wide) of our sample funds normalised by the value of the market (NYSE + AMEX + NASDAQ) and scaled by 1000. The dependent variable in Panel B is the six-month return over the period November-April minus the six-month return May-October, a proxy for the Sell in May effect. The explanatory variable is the half-year flow over the winter months November-April minus the flow during the remainder of the year ($Flow_{w-s}$), normalised by the value of the market at the previous month and scaled by 1000. The sample period is January 1995 to December 2014.

Panel A

	(1)	(2)	(3)	(4)	(5)
Obs.	240	240	240	240	240
Adj-R^2	0.01	0.08	0.13	0.16	0.15
Intercept	0.24	0.00	0.41	**3.69**	-2.45
	(0.57)	(0.01)	(0.99)	**(2.77)**	(-1.44)
Hal	**0.96**	0.50	0.14	0.14	0.12
	(1.85)	(1.03)	(0.30)	(0.30)	(0.27)
PE_{t-1}				**-0.17**	
				(-2.71)	
DY_{t-1}					1.50
					(1.62)
Flow		**1.67**	**3.30**	**3.37**	**3.31**
		(3.71)	**(5.29)**	**(5.50)**	**(5.38)**
$Flow_{t-1}$			**-1.59**	**-1.41**	**-1.46**
			(-2.43)	**(-2.24)**	**(-2.33)**
$Flow_{t-2}$			-0.06	0.09	0.04
			(-0.07)	(0.12)	(0.06)
$Flow_{t-3}$			-0.73	-0.53	-0.58
			(-1.20)	(-0.88)	(-0.96)

Panel B

	Dependent Variable: Market $Return_{w-s}$			
	Obs.	R^2	Intercept	$Flow_{w-s}$
Coef.	20	0.54	-1.02	**4.21**
(*t* -stat.)			-0.46	**(4.89)**

Table 4
Expected and Unexpected Mutual Fund Flows and Stock Returns

The first three columns of this table reports estimation results from regressing market returns on expected and unexpected fund flow and the Halloween indicator, *Hal*. This is based on a two-step estimation procedure where expected and unexpected flow are generated from estimates of a first-step regression. In the first step we regress flow on three lags of flow and three lags of returns. Expected flow is the fitted value, while unexpected flow is the residual. In columns four and five we regress expected and unexpected flow on lagged market returns. The sample period is January 1995 to December 2014. *t*-statistics are reported in parentheses based on Newey-West corrected standard errors.

	Dependent Variable				
	Market Return			Expected Flow	Unexpected Flow
	(1)	(2)	(3)	(4)	(5)
Obs.	240	240	240	240	240
Adj-R^2	0.01	0.14	0.14	0.17	0.13
Intercept	0.13	**0.65**	0.53	**0.22**	-0.03
	(0.25)	**(1.79)**	(1.20)	**(3.39)**	(-0.90)
Hal	**0.95**	0.15	0.14		
	(1.90)	(0.33)	(0.29)		
Expected Flow	(0.44)		(0.46)		
	(0.73)		(0.77)		
Unexpected Flow		**3.37**	**3.38**		
		(5.45)	**(5.42)**		
Return				0.00	**0.04**
				(0.32)	**(4.99)**
$Return_{t-1}$				**0.05**	0.00
				(7.34)	(-0.50)
$Return_{t-2}$				**0.02**	0.00
				(3.10)	(0.23)
$Return_{t-3}$				**0.02**	0.00
				(2.77)	(-0.62)

Table 5
Mutual Fund Flows and the January Effect

This table reports estimation results of abnormal returns in January. The dependent variable is the monthly return on the EW CRSP stock market index (NYSE + AMEX + NASDAQ). The January dummy, Jan_t, is 1 for returns that fall into January and 0 otherwise. $Flow_t$ is the estimated monthly net flow (market-wide) of our sample funds. SMB_t and HML_t are Fama and French's (1993) firm size (small minus big) and value (high minus low book-to-market ratio) factors. MOM_t is Carhart's (1997) momentum (winner minus loser) factor. Standard errors are corrected for heteroskedasticity and autocorrelation. *t*-statistics are reported in parentheses. The sample period is January 1995 to December 2014.

	(1)	(2)	(3)	(4)	(5)
Obs.	240	240	240	240	240
Adj-R^2	0.02	0.20	0.21	0.20	0.18
Intercept	**0.89**	0.74	2.86	**-1.86**	0.79
	(2.02)	(1.63)	(1.52)	**(-0.82)**	(1.56)
Jan	**2.73**	0.42	0.47	0.44	0.74
	(1.95)	(0.29)	(0.33)	(0.30)	(0.53)
Flow		**4.93**	**4.97**	**4.94**	
		(6.66)	**(6.76)**	**(6.70)**	
$Flow_{t-1}$		**-1.52**	**-1.40**	**-1.40**	
		(-1.77)	**(-1.72)**	**(-1.77)**	
$Flow_{t-2}$		-0.58	-0.49	-0.49	
		(-0.76)	(-0.63)	(-0.63)	
$Flow_{t-3}$		**-1.56**	**-1.42**	**-1.42**	
		(-2.63)	**(-2.26)**	**(-2.20)**	
PE_{t-1}			**-0.11**		
			(-1.70)		
DY_{t-1}				1.35	
				(1.08)	
Expected Flow					0.92
					(1.23)
Unexpected Flow					**4.69**
					(6.23)

Table 6
Abnormal Return and Flow in January

Column two of the table reports abnormal return on the EW CRSP stock market index in January estimated with equation (4). Abnormal flow of the sample funds in January based on equation (5) is reported in column four. Standard errors are corrected for heteroskedasticity and *t*-statistics are reported in parentheses. The sample period is January 1995 to December 2014.

	January Excess Return		January Excess Flow	
	Coef.	*(t-stat.)*	Coef.	*(t-stat.)*
Obs.	240		240	
Adj-R^2	0.07		0.54	
1995	**2.41**	**5.28**	**0.59**	**19.14**
1996	**2.61**	**7.21**	**1.16**	**20.73**
1997	**5.78**	**14.40**	**1.24**	**40.56**
1998	**1.39**	**2.65**	**0.36**	**8.27**
1999	**5.41**	**15.43**	**0.61**	**18.56**
2000	**2.56**	**3.66**	**0.23**	**7.38**
2001	**21.99**	**52.10**	**0.50**	**16.09**
2002	-0.10	-0.19	**0.65**	**21.06**
2003	0.81	1.26	0.00	-0.02
2004	**5.07**	**12.78**	**1.17**	**26.23**
2005	**-4.74**	**-10.07**	**0.10**	**3.13**
2006	**6.68**	**18.88**	**-0.09**	**-2.90**
2007	**1.24**	**3.53**	**0.31**	**8.71**
2008	**-4.93**	**-11.13**	**-1.48**	**-34.51**
2009	**-3.98**	**-9.67**	**0.76**	**12.26**
2010	**-2.99**	**-6.19**	**0.48**	**5.83**
2011	-0.71	-1.23	**0.96**	**12.24**
2012	**7.95**	**19.61**	**0.47**	**4.89**
2013	**5.21**	**14.81**	**1.26**	**13.42**
2014	**-1.34**	**-3.76**	**0.08**	**1.73**

Table 7

Relationship between $Flow_t$, PTS_t, Vol_t, Std_t and the January Effect

This table reports results of regressions in which the dependent variable is the January effect estimated with equation (4). $Flow_t$ is the abnormal flow in January estimated with equation (5). PTS_t is the equally weighted year's end potential tax-loss selling over all stocks listed on the NYSE, AMEX and NASDAQ. It is defined as the percentage decrease from the highest price attained during a year to December 15. If there was no trading on December 15, we take the price of the previous day. Vol_t is the natural logarithm of dollar volume in January relative to the average monthly dollar volume over the previous six months (July – December). Monthly volume is calculated from the numbers of shares traded on day *t* times closing price of day *t* of each stock listed on the NYSE, AMEX and NASDAQ. Data on prices and number of shares are obtained from CRSP via WRDS. Std_t is the natural logarithm of the standard deviation of the EW CRSP stock market index in January relative to the standard deviation of the index over the previous six months (July – December). Standard deviation is calculated from daily returns. Standard errors are corrected for heteroskedasticity and autocorrelation. *t*-statistics are reported in parentheses. The sample period is 1995 to 2014.

	Flow	PTS	Vol	Std	R^2
Coef.	**2.87**				0.09
(t-stat.)	**(4.97)**				
Coef.	**3.17**	-16.72	4.48	-2.35	0.20
(t-stat.)	**(3.55)**	(-0.93)	(1.28)	(-0.63)	

Journal of Financial Economics 139 (2021) 138–161

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Are return seasonalities due to risk or mispricing?[☆]

Matti Keloharju[a,b,c], Juhani T. Linnainmaa[d,e,*], Peter Nyberg[a]

[a] Aalto University School of Business, Ekonominaukio 1, Espoo 02150, Finland
[b] Center for Economic Policy Research, 33 Great Sutton Street, London, EC1V 0DX, United Kingdom
[c] Research Institute of Industrial Economics, Grevgatan 34, Stockholm SE-102 15, Sweden
[d] Dartmouth College, Tuck School of Business, 100 Tuck Hall, Hanover NH 03755, USA
[e] National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge MA 02138, USA

ARTICLE INFO

Article history:
Received 11 April 2019
Revised 28 October 2019
Accepted 26 November 2019
Available online 15 July 2020

JEL classification:
G12

Keywords:
Cross-sectional seasonalities
Reversals
Risk
Mispricing

ABSTRACT

Stocks tend to earn high or low returns relative to other stocks every year in the same month (Heston and Sadka, 2008). We show these seasonalities are balanced out by seasonal reversals: a stock that has a high expected return relative to other stocks in one month has a low expected return relative to other stocks in the other months. The seasonalities and seasonal reversals add up to zero over the calendar year, which is consistent with seasonalities being driven by temporary mispricing. Seasonal reversals are economically large and statistically highly significant, and they resemble, but are distinct from, long-term reversals.

1. Introduction

Stocks that are winners in a given month tend to continue to outperform stocks that are losers in that same month, for up to 20 years (Heston and Sadka, 2008). For example, if a stock has previously performed well (poorly) relative to other stocks in March, we expect it to also offer a high (low) return relative to other stocks next March. Return seasonalities are not confined to stocks or monthly frequency. They also exist in commodity, country index, and anomaly returns and at daily and intraday frequencies.[1] In this paper, we show a striking pattern in the cross section of expected returns: seasonal reversals. The existence of these reversals suggests that return seasonalities likely emanate from mispricing.

Return seasonalities could stem from risk or mispricing. An asset that is systematically riskier in one period than in others could earn a disproportionate share of its risk premium in that period. Alternatively, systematic variation in investors' demand for risky assets from one period to the next could dislocate asset prices from fundamental values. For example, if investors tend to favor small stocks over big

☆ We thank Chris Hrdlicka, Mark Kamstra, Owen Lamont, Jon Lewellen, an anonymous referee, and the seminar and conference participants at the Australian National University, Chinese University of Hong Kong, City University of Hong Kong, Hong Kong Polytechnic University, Nanyang Technological University, National University of Singapore, Singapore Management University, University of New South Wales, University of Technology Sydney, 2018 Citi Quant Research Conference, and 2019 Chicago Quantitative Alliance Conference for insightful comments. Financial support from the Academy of Finland, Inquire Europe, and Nasdaq Nordic Foundation is gratefully acknowledged.

* Corresponding author at: Dartmouth College, Tuck School of Business, 100 Tuck Hall, Hanover, NH 03755, USA.

E-mail addresses: matti.keloharju@aalto.fi (M. Keloharju), Juhani.T.Linnainmaa@tuck.dartmouth.edu (J.T. Linnainmaa), peter.nyberg@aalto.fi (P. Nyberg).

[1] See, for example, Heston and Sadka (2010), Heston et al. (2010), Keloharju et al. (2016), Bogousslavsky (2016), Hirshleifer et al. (2020), and Birru (2018) for studies of return seasonalities.

https://doi.org/10.1016/j.jfineco.2020.07.009
0304-405X/

stocks at the beginning of the year, their excess demand could predictably push small stocks' valuations higher every January.

We propose the following test to disentangle the two competing explanations for return seasonalities. If they are due to mispricing, they should be offset by seasonal reversals. To illustrate our point, suppose that some stocks typically earn an additional return of 5% in January because of excess demand by investors. Mispricing, however, does not persist forever. As long as there are no infinitely lived bubbles, asset prices converge back to fundamental values. Therefore, if some stocks' January return is systematically 5% "too high," the cumulative non-January return should be 5% "too low."[2] Otherwise, prices would drift even further away from fundamental values.

This prediction of seasonal reversals, i.e., a high seasonal return in one month is offset by low seasonal returns in the other months, is specific to the mispricing explanation for seasonalities. A risk factor's premium could be higher in one month because the underlying risk matters more, or is perceived as being more costly, in that month than others. The risk-based explanation makes no predictions about seasonal reversals. If an asset earns an above-average return in one month, no economic reason exists that it should earn a below-average return in the other months.

Return seasonalities and seasonal reversals are both about the association between past and future returns. If there are return seasonalities, past same-month returns positively predict the cross section of returns. A high December return predicts high December returns in the future. If there are seasonal reversals, past other-month returns negatively predict returns: high non-December return predicts low December returns. We show that seasonal reversals are economically and statistically significant. Sorting stocks by their average same-month returns, the difference between the top and bottom deciles has an alpha of 1.09% (t-value = 7.19) in the Carhart (1997) model augmented with the long-term reversals factor. Sorting the stocks by the average other-month returns, the difference has an alpha of −0.53% (t-value = −4.74).

We cannot reject the null hypothesis that the seasonalities in US equities reverse perfectly. We gauge this adding-up constraint by computing the correlation between a stock's expected return in one month, proxied by its historical average return in that month, and the sum of its expected returns in the other months. If the adding-up constraint holds perfectly and if expected returns can be observed without noise, the correlation is −1. In reality, the noise in average returns biases the correlation toward zero. We correct for this bias by simulating data from a model in which the adding-up constraint holds perfectly and the simulated returns are as noisy as the actual stock returns. The correlation estimate is the same, −0.06, in both the simulations and the data. The data are thus consistent with a pure mispricing explanation for return seasonalities.

If seasonalities are balanced by seasonal reversals, we would expect to find seasonal reversals wherever seasonalities are found. Following Keloharju et al. (2016), we measure seasonalities and seasonal reversals not only in monthly US equity returns but also in daily stock returns and monthly and daily country equity index and commodity returns. Everywhere we find return seasonalities, we also find seasonal reversals. In addition, we find seasonalities and seasonal reversals in international stock returns. Return seasonalities across asset classes and frequencies are always offset or moderated by seasonal reversals.

Our insights on seasonal reversals improve the predictive power of seasonal trading strategies. Given that realized returns are noisy, both same- and other-month returns contain independent information about future expected returns. A factor that sorts stocks based on the same-month minus other-month difference earns an average return of 67 basis points per month with a t-value of 9.93, a notable increase from the seasonality factor's average return of 61 basis points per month (t-value = 8.37). Neither seasonalities nor seasonal reversals subsume each other, which is consistent with them containing independent information about expected returns.

Seasonalities and seasonal reversals are unrelated to short-term reversals, momentum, and long-term reversals. Although seasonal reversals resemble long-term reversals, different mechanisms drive them. The average return on the long-term reversal factor is 29 basis points per month (t-value = 2.95), but its correlations with size and value render its three-factor model alpha statistically insignificant (Fama and French, 1996; Asness et al., 2013). The seasonal reversal factor's three-factor model alpha is significant with a t-value of 6.17. The addition of the momentum and long-term reversal factors lower this t-value, but only to 5.33. That is, the seasonal reversal factor is more than just another version of the long-term reversal factor. This last finding is important. Heston and Sadka (2008) and Keloharju et al. (2016), for example, report non-annual return estimates that betray the existence of seasonal reversals. These papers appear to interpret reversals as a manifestation of long-term reversals. Apart from showing they are not, we demonstrate they are closely linked and explain how this link relates to mispricing.

Our results help explain why seasonal strategies require active trading. Suppose we identify a stock that earns reliably higher returns in December. We can therefore predict that this stock's return in December in 20 years will be high as well. From the viewpoint of buy-and-hold returns, seasonal reversals render this predictability inconsequential. The December return cannot be earned 20 years later without also earning the January-through-November returns that year. The point of the adding-up constraint is that these other returns perfectly offset the high expected December return. A buy-and-hold investor thus cannot harvest return seasonalities without also succumbing to seasonal reversals.

[2] This adding-up constraint is not a tautology. If a stock's expected return relative to the other stocks is high in one month, it does not have to have a low expected return relative to other stocks in the other months. It would be tautological to state that a stock with a high expected return in one month relative to its own time series mean must earn a low expected return in the other months relative to this mean. The adding-up constraint is a statement about cross-sectional differences in expected returns, not about time series differences in them.

Our findings point to temporary mispricing being the most plausible explanation for return seasonalities. The more the seasonalities reverse, the greater the role the mispricing channel likely plays in generating the seasonalities. Seasonalities could be driven by predictable trading by investors. This predictability can emanate from, for example, sentiment (Birru, 2018; Hirshleifer et al., 2020) or recurring in- and outflows. Heston et al. (2010) find seasonalities in intraday returns and attribute this seasonality to traders consistently trading in the same direction at the same time of the day, a pattern Murphy and Thirumalai (2017) observe for institutional traders. Bogousslavsky (2016) builds a model in which part of the investor population trades infrequently. Resulting predictable changes in supply and demand are imperfectly absorbed by the rest of the market, which generates seasonal patterns similar to those in Heston et al. (2010). Our results suggest that lower frequency seasonalities, including those at the monthly frequency, can also be due to temporary mispricing induced by the correlated trading by individuals or institutions, or both.[3]

While we interpret our evidence as supporting the mispricing interpretation for return seasonalities, we cannot rule out all risk-based explanations. To see why, consider how seasonalities in risk would affect prices. Suppose that all information arrives only once a year in December and that there is no asymmetric information or possibility of leakage of information. In this world, risky assets earn all of their risk premiums in December. Outside December, risky assets are effectively riskless. Risk premiums would not reverse: the high December risk premium is not offset by a negative non-December risk premium. At the same time, risk-based examples that generate reversals can be reverse-engineered. Suppose, for example, that stocks partition in 12 bins such that each bin earns all of its risk premium in a different month of the year. If all stocks earn the same risk premium, then both seasonalities and seasonal reversals would be evident in the cross section of stock returns. This example is, however, contrived. Our point is that no obvious reason exists that assets across different asset classes would satisfy this type of a constraint. By Occams razor, the greater number of assumptions makes the risk-based hypothesis a less plausible explanation to the seasonality phenomenon.

The rest of the paper is organized as follows. Section 2 describes how different assumptions about the nature of the cross-sectional variation in expected returns alter the predictive relation between returns and lagged returns. Then, it calibrates a model of return seasonalities to quantify the extent to which seasonalities add up to zero. Section 3 measures seasonalities and seasonal reversals using Fama and MacBeth (1973) regressions in monthly and daily US stock returns, international stock returns, country equity indices, and commodity returns. Section 4 presents additional empirical results and robustness tests. Section 5 constructs seasonality and seasonal reversal factors, and it examines their relation to short-term reversals, momentum, and long-term reversals. Section 6 concludes.

2. Model and calibration

2.1. Implications of seasonalities and seasonal reversals on cross-sectional predictability

We analyze how different assumptions about the nature of the cross-sectional variation in expected returns alter the predictive relation between past and contemporaneous returns. We compare estimates from Fama and MacBeth (1973) regressions computed using actual data and data simulated under different assumptions. In Panel A of Fig. 1, we plot the average coefficients from cross-sectional regressions of month t returns against month $t-k$ returns,

$$r_{it} = a + b \times r_{i,t-k} + \varepsilon_{it}, \tag{1}$$

where r_{it} is stock i's return in month t. Panel B is similar to Panel A except that it predicts returns using past annual returns. We estimate all regressions in Fig. 1 using lags up to ten years.

The Data graph uses the monthly Center for Research in Security Prices (CRSP) return data from January 1963 to December 2016 on stocks listed on the NYSE, Amex, and Nasdaq.[4] The negative coefficient at the first lag is about short-term reversals; the positive coefficients up to the year mark are about momentum; and the spikes denote the seasonalities in stock returns (Heston and Sadka, 2008).

Model 1 in Panel A has persistent differences in expected returns but no seasonalities. We draw stock returns from the following process:

$$r_{it} = \mu_i + \epsilon_{it}, \tag{2}$$

where ϵ_{it} is independent and identically distributed (i.i.d.). In this model, a stock's expected return could be 5% per year every month of the year; for another stock, it could

[3] Both retail investors and institutions herd, and the price pressures imparted by these herds can influence prices. Lakonishok et al. (1992) and Wermers (1999) show that both pension and mutual funds herd, in particular when trading small stocks, and that stocks bought by herds of mutual funds significantly outperform the stocks they sell. Ritter (1988), Kumar and Lee (2006), and Barber et al. (2009a,b) find that individual investors in the US trade in concert. Feng and Seasholes (2004), Dorn et al. (2008), and Grinblatt et al. (2012) show that retail investors in China, Germany, and Finland also herd. Kumar and Lee (2006) show that individuals' herding behavior explains return co-movement among stocks widely held by retail investors, and Barber et al. (2009a) find that the aggregate imbalances of retail investors forecast returns. Seasonal variation in sentiment can, in part, drive these herds, thereby inducing seasonalities into asset prices. Hirshleifer et al. (2020) review experimental, survey, and empirical evidence and find this evidence supportive of the notion that investor mood varies systematically across calendar months and weekdays.

[4] We exclude securities other than ordinary common shares. We use CRSP delisting returns. If a delisting return is missing and the delisting is performance-related, we impute a return of −30% (Shumway, 1997) for NYSE and Amex stocks and −55% for Nasdaq stocks. Later, we include book-to-market as a control variable. We use the book values of equity from the annual Compustat files, supplemented with the Davis et al. (2000) data, and follow the standard conventions to time this information.

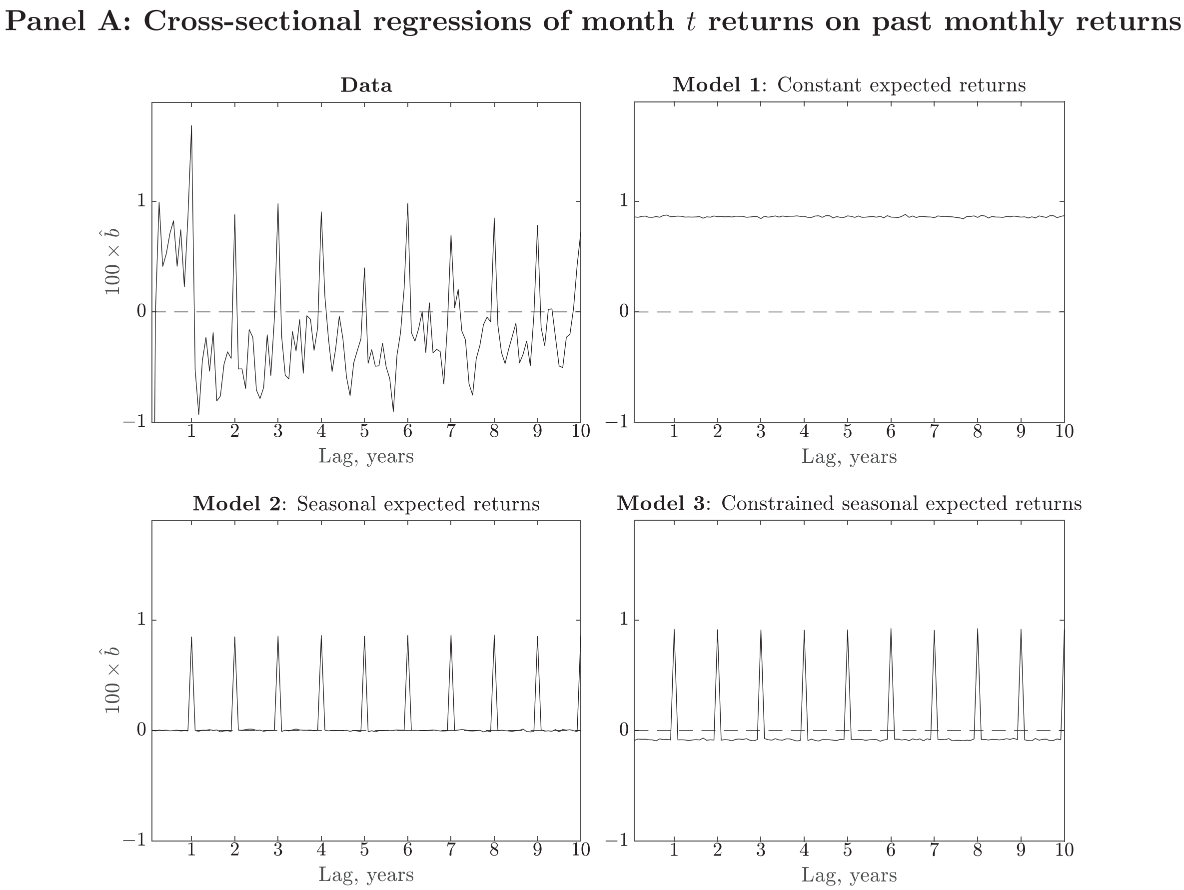


Fig. 1. Fama-MacBeth regressions: data versus theory. This figure reports estimates from cross-sectional regressions that predict the cross section of monthly returns either with past monthly (Panel A) or annual (Panel B) returns using lags up to ten years. The Data graph uses return data on NYSE, Amex, and Nasdaq stocks from January 1963 to December 2016. Models 1–3 simulate data under different assumptions about expected returns. In Model 1, expected stock returns are constant. In Model 2, expected stock returns vary by calendar month. In Model 3, expected stock returns vary by calendar month and satisfy the adding-up constraint. This adding-up constraint restricts the sum of each stock's expected returns to zero.

be 8% per year. With persistent differences in expected returns, realized returns exhibit "poor man's momentum." Month $t-k$ return predicts the cross section of month t returns because we explain $\mu_i + \epsilon_{it}$ with $\mu_i + \epsilon_{i,t-k}$, that is, the same expected return (μ_i) appears on both the left- and right-hand sides of the regression.[5] The theoretical regression coefficient at any lag k therefore equals

$$\hat{b}_k = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\epsilon^2}. \quad (3)$$

This positive predictive relation holds when today's returns are explained with both monthly (Panel A) and annual (Panel B) returns. The model's predictions profoundly contradict the data. Setting aside the short-term reversals and momentum, the main difference between the actual data and the simulated data is that, in Model 1, the past monthly returns predict today's returns at all lags. In the data, this positive predictive relation holds only at annual lags in the monthly regressions. In Panel B, past returns after year 1 are typically negatively correlated with the cross section of monthly returns.

In Model 2, stocks' expected returns display seasonal variation. The stock return process is

$$r_{it} = \mu_{i,m(t)} + \epsilon_{it}, \quad (4)$$

where ϵ_{it} is i.i.d. and $\mu_{i,m(t)}$ is stock i's expected return in calendar month $m(t) = 1, \ldots, 12$. In this model, a stock's expected return in October could differ from its expected return in November. Because we predict the cross section of returns, we do not specify the level of expected returns. It washes out from the regression estimates. In Model 2, we assume that a stock's expected return in one month is independent of its expected returns in the other months. That is, we do not impose any constraint on the sum of the expected returns, $\mu_{i,1} + \cdots + \mu_{i,12}$. Panel A's Fama-MacBeth regression coefficient then equals

$$\hat{b}_k = \begin{cases} \dfrac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\epsilon^2} & \text{at annual lags and} \\ 0 & \text{at non-annual lags.} \end{cases} \quad (5)$$

This model is consistent with the data with respect to the seasonal spikes. Past same-month returns positively pre-

[5] See, for example, Lo and MacKinlay (1990), Conrad and Kaul (1998), and Berk et al. (1999) for discussions of this mechanism.

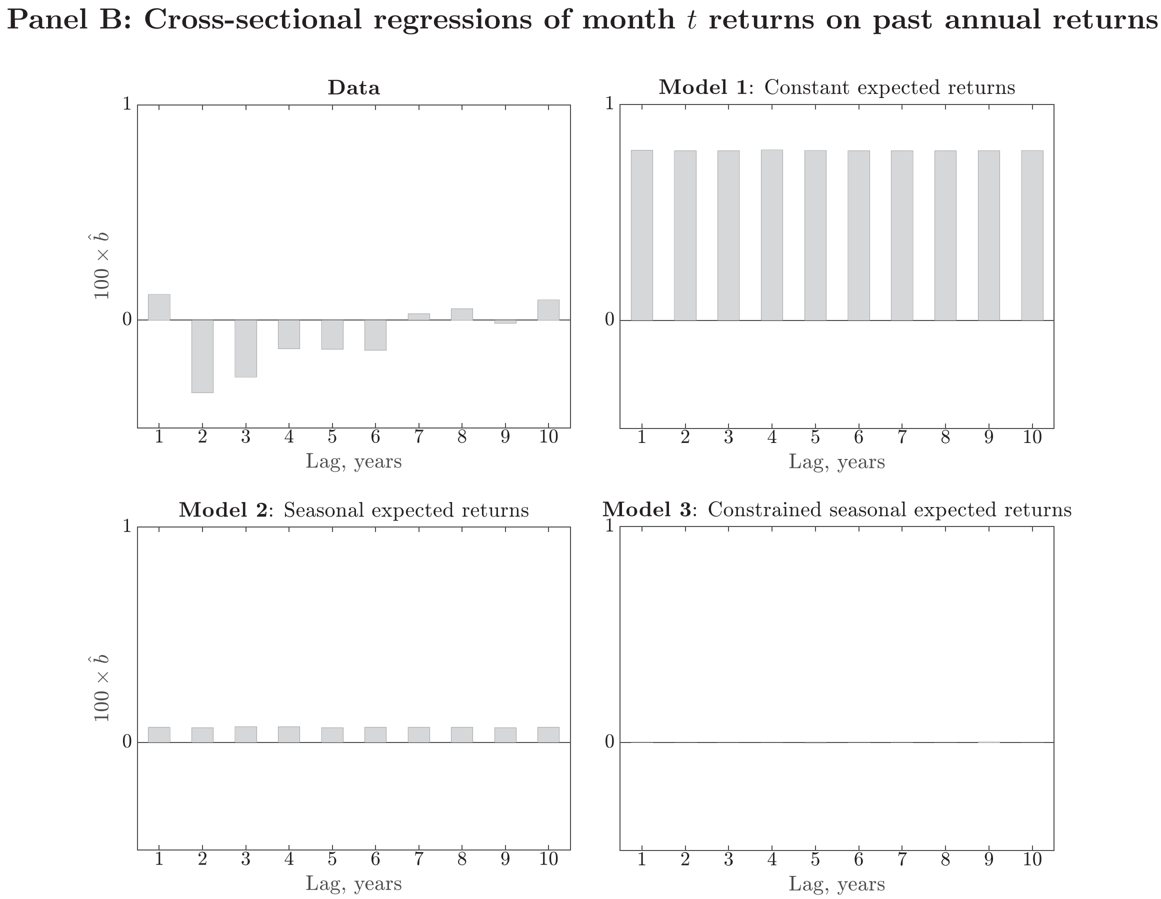


Fig. 1. Continued

dict today's return because both contain the same expected return component, $\mu_{i,m}$. At non-annual lags, past returns have no predictive power on expected returns today.

Panel B of Fig. 1 shows that, in Model 2, historical annual returns still positively predict today's returns. The reason is that some stocks have predominantly positive seasonalities ($\mu_{i,1} + \cdots + \mu_{i,12} > 0$), while others have predominantly negative seasonalities ($\mu_{i,1} + \cdots + \mu_{i,12} < 0$). A stock could, for example, have high expected returns for six months of the year and expected returns close to zero for the rest of the year. Annual returns provide information about these sums of seasonalities. A stock with a high realized annual return is more likely a stock with more positive than negative seasonalities. Therefore, without a constraint on the seasonalities in expected returns, past annual returns positively predict today's returns, just as they would in a model with constant expected returns. This positive correlation between today's returns and past annual returns contradicts the data.

Model 3 imposes an adding-up constraint on the seasonalities,

$$\mu_{i,1} + \mu_{i,2} + \cdots + \mu_{i,12} = 0. \tag{6}$$

In the monthly regression results in Panel A, the regression coefficient at annual lags is the same in Model 3 and Model 2, which does not have the constraint. Because of the adding-up constraint in Eq. (6), a stock's realized return in, say, January is informative about its expected returns both in January and in all other months. A stock with an unusually high January expected return must have unusually low expected returns throughout the rest of the year. A stock's expected return in January, therefore, negatively relates to its expected return in, for example, February:

$$\mu_i^{\text{Jan}} = -\left(\mu_i^{\text{Feb}} + \cdots + \mu_i^{\text{Dec}}\right) = -\mu_i^{\text{Feb}} + \text{noise}. \tag{7}$$

The Fama-MacBeth regression coefficient under Model 3 equals

$$\hat{b}_k = \begin{cases} \dfrac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\epsilon^2} & \text{at annual lags and} \\ \dfrac{-\frac{1}{11}\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\epsilon^2} & \text{at non-annual lags.} \end{cases} \tag{8}$$

This model is consistent with several features of the data. First, similar to Model 2, which does not have a constraint, the seasonalities in expected returns generate annual spikes in the monthly regression coefficients. Second, because the seasonalities add up to zero, the non-annual regression coefficients are pushed downward. These negative troughs are the seasonal reversals. Third, because every stock's annual expected return equals zero, annual realized returns in Panel B are not predictive of differences in future returns.

The model also is inconsistent with some aspects of the data. The short-term reversals and momentum are short-run, autocorrelation-like effects, and a model with only persistent variation in expected returns cannot match these features. Similarly, the long-term reversals of De Bondt and Thaler (1985) cannot be only about seasonal reversals. Panel B of Fig. 1 shows that these negative coefficients also are present in annual regression results. These negative coefficients cannot emanate from seasonalities alone. In a model with just seasonal variation in expected returns, the coefficients are either positive, as in Model 2, or zero, as in Model 3. They cannot be negative. Negative correlations must emanate from negative serial correlations or from positive cross-serial correlations (Lo and MacKinlay, 1990).

2.2. Model

We calibrate a model to the data to assess the extent to which return seasonalities cancel out over the year. We simulate data from a model and choose the parameters to fit the annual spikes and non-annual troughs of the Data graph of Fig. 1, Panel A. In this model, we continue with the assumption that a stock's realized return equals its seasonal expected return plus noise,

$$r_{it} = \mu_{i,m(t)} + \epsilon_{it}. \tag{9}$$

We generate the $\mu_{i,m(t)}$s as follows. We begin by generating 12 draws from a normal distribution,

$$\mu^e_{i,m} \sim N\left(0, \frac{12}{11}\sigma^2_\mu\right) \text{ for } m = 1, \ldots, 12, \tag{10}$$

and then demean the resulting draws,

$$\mu_{i,m} = \mu^e_{i,m} - \frac{1}{12}\sum_{m=1}^{12}\mu^e_{i,m}. \tag{11}$$

These expected returns $\mu_{i,m}$s thus perfectly satisfy the adding-up constraint: A high expected return in one month is offset by correspondingly lower expected returns in the other months.[6]

Because stock returns exhibit long-term reversals, and long-term reversals induce a negative cross-sectional correlation between stock returns and lagged stocks returns, we let the return innovations, ϵ_{it}, exhibit such reversals after the one-year mark. We assume that this return innovation is

$$\epsilon_{it} = \sum_{k=13}^{120}\delta_k \xi_{i,t-k} + \xi_{i,t}, \tag{12}$$

where $\xi_{it} \sim N(0, \sigma^2_\xi)$. With $\sum_{k=13}^{120}\delta_k < 0$, this assumption builds in long-term reversals. Because our interest is calibrating the model to the seasonalities and reversals in the data, we do not model the short-term reversals and momentum. For simplicity, we assume that δ_k is non-positive and that it changes linearly in k, that is, $\delta_k = \min(\bar{\delta} + k \times \gamma, 0)$. This assumption permits the possibility that the reversals strengthen or weaken over time.

2.3. Calibration

The model is characterized by four parameters: σ^2_μ generates the seasonalities in expected returns; $\bar{\delta}$ and γ generate the long-term reversals; and σ^2_ξ determines the amount of noise in individual stock returns. We simulate data by taking the full CRSP database as the starting point. We then replace the true returns with returns simulated from the model. This procedure ensures that the number of firms each month matches the number of firms in the actual data. We choose the parameters to match the autocorrelation patterns shown in Panel A of Fig. 1. We match, between the simulations and the data, the cross-sectional variance of stock returns and the coefficients from regressions that predict the cross section of monthly returns with past returns. The explanatory returns consist of the annual returns in months $t-12$, $t-24$,..., $t-120$. In addition, we include the average non-annual returns over the prior ten years, skipping a year. That is, in the first of these non-annual regressions we use the average return from month $t-23$ to month $t-13$; in the second, the average return from month $t-35$ to month $t-25$; and so forth. We find the parameters with the simulated method of moments, using the identity matrix as the weighting matrix to match these 20 moments (one cross-sectional variance, ten annual regression coefficients, and nine non-annual regression coefficients) between the data and the model.

Fig. 2 shows the average coefficients from Fama-MacBeth regressions that predict the cross section of monthly returns with lagged returns. The black line represents the estimates that use the actual data. These estimates are the same as those reported in Panel A of Fig. 1. The red and blue lines represent estimates that are based on one simulation each. The red line simulates one set of data from the full model with both seasonalities and long-term reversals. The blue line shows data with the same parameters except that long-term reversals are shut down by setting $\bar{\delta} = \gamma = 0$.

The model is not designed to match short-term reversals and momentum, so the red line (simulation) substantially differs from the black line (data) up to the one-year mark. After one year, the model matches the key features of the return data. Both the seasonal spikes and the non-seasonal troughs are of the same magnitude. This similarity indicates that the real data are consistent with a model in which seasonal reversals completely balance out seasonalities.

2.4. Correlations in average calendar-month returns: data versus simulations

A correlation between a stock's expected return in one month and the sum of its expected returns in the other months is a measure of the extent to which the seasonalities satisfy the adding-up constraint. This correlation is

[6] We multiply the variance σ^2_μ in Eq. (10) by $\frac{12}{11}$ to ensure that the variance of $\mu_{i,m}$s in the simulated data has this same variance, σ^2_μ. Because we demean $\mu^e_{i,m}$s in Eq. (11) to get $\mu_{i,m}$s, we lose one degree of freedom. The variance of the draws demeaned by the sample average is less than the variance of the original draws.

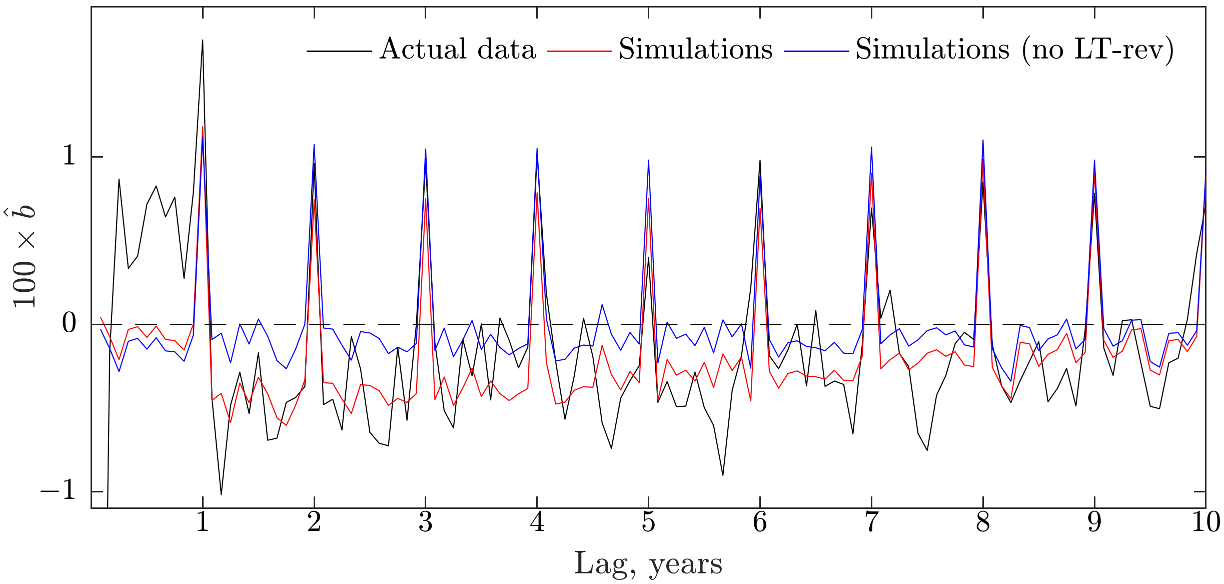


Fig. 2. Seasonalities, seasonal reversals, and long-term reversals. The black line represents the estimates from univariate Fama-MacBeth regressions used to predict the cross section of monthly returns with past monthly returns using lags up to ten years. The red and blue lines represent the same coefficients computed using simulated data with the same dimensions as in the actual data. In the model, expected returns vary by calendar month and add up to zero, and return innovations display long-term reversals from month $t-120$ to $t-13$. These long-term reversals dampen over time. The model is calibrated to match the cross-sectional variance of stock returns and the annual spikes and non-annual troughs in the Fama-MacBeth regressions. The red line represents simulated data from the full model. The blue line shuts down long-term reversals, leaving only seasonalities and seasonal reversals in the model.

-1 if this constraint holds perfectly. We measure the correlation from average returns. We first take all stocks with at least ten years of data over the entire sample period. We then cross-sectionally demean the data and compute, for each stock, the average return in each calendar month. We reorganize the data so that we have 12 observations for each stock: a stock's average January return aligned with the sum of its average February-December returns, and similarly for the other months. We then estimate the following regression:

$$\bar{r}_{i,m} = a + b \times \sum_{m' \neq m} \bar{r}_{i,m'} + e_{i,m}. \tag{13}$$

The slope coefficient estimate, $\hat{b}$, from this regression is -0.057, with a t-value of -33.1.[7] This regression is not a predictive regression. We estimate this regression to measure the degree to which average returns in one month are related to the sum of the average returns in the other months.

The negative slope coefficient indicates that a stock that earned, on average, high returns in one month earned, on average, lower returns in the other months. Because we demean the data, this negative correlation is not due to the seasonal patterns in market-wide returns (Kamstra et al., 2003). The fact that $\hat{b}$ is statistically significantly negative alone indicates that expected returns exhibit at least some amount of seasonal reversal.

How closely do seasonalities in expected returns satisfy the adding-up constraint? The estimate of -0.057 is substantially higher than -1, but this estimate is biased toward zero because of an errors-in-variables problem. The explanatory variable $\sum_{m' \neq m} \bar{r}_{i,m'}$ is noisy, so the -0.057 estimate does not quantify the degree to which this constraint holds. To get a sense of how noisy signals the realized returns are of expected returns, consider the Fama-MacBeth slope coefficients from annual lags. In Fig. 2, the estimates are, on average, just below 0.01. This indicates, by Eq. (8), that just under 1% of the cross-sectional variance of stock returns emanates from differences in expected returns. The bias in $\hat{b}$ in Eq. (13) is therefore substantial.

Fig. 3 shows how the noise in realized returns, induced through idiosyncratic shocks, draws the coefficient estimate from Eq. (13) toward zero. We take the calibrated model from Fig. 2 and then vary the amount of noise in returns from $\sigma_\xi = 0$ (no noise) to 20%. We simulate data from the model, again preserving the dimensions of the actual data, and estimate Eq. (13) using these simulated data. If returns have no noise, $\sigma_\xi = 0$, a stock's average return in one calendar month perfectly negatively correlates with the sum of its average returns in the other calendar months. The slope estimate is -1 because there is no noise in returns. Moving along the x-axis, noise draws the estimated coefficient toward zero. Using the calibrated value from Fig. 2, which sets $\sigma_\xi = 13.7\%$, the coefficient is -0.061, which is close to its value in the actual data, -0.057.

[7] We compute the standard error by block bootstrapping the data by calendar month. We draw calendar months in blocks with replacement, recompute stocks' average returns, and repeatedly reestimate the regression in Eq. (13). This bootstrapping procedure uses only time series variation in returns to quantify the amount of estimation uncertainty about b in Eq. (13).

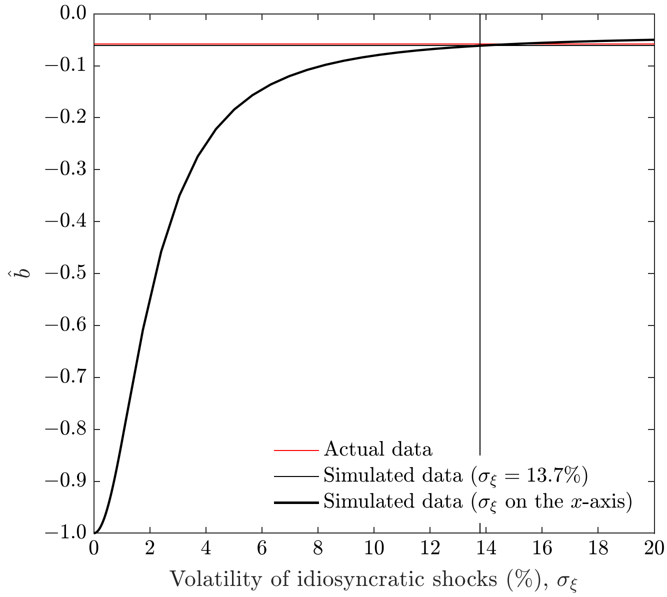


Fig. 3. Seasonal reversals: correlations between the average same- and other-month returns. We compute each stock's average return in each calendar month using monthly return data from January 1963 through December 2016. The red horizontal line at −0.057 indicates the estimated slope from a regression of $\bar{r}_{i,m}$ on $\sum_{m' \neq m} \bar{r}_{i,m'}$, where $\bar{r}_{i,m}$ is stock i's average return in month m. The black lines represent the same coefficients estimated using simulated data. These simulated data have the same dimensions as the actual return data, but the returns are generated through Eqs. (9)–(12). The parameter values in the simulations are calibrated to match the cross-sectional regression coefficients and return volatilities between the actual and simulated data. The thin black line denotes $\sigma_\xi = 13.7\%$, which is its value in the calibration. The thick black line simulates data using the value of σ_ξ shown on the x-axis.

The similarity in the estimates between the actual and simulated data is noteworthy. We do not calibrate the model to the data to match this moment. The negative slope estimate of −0.057 in the actual data is consistent with a world in which the seasonalities in expected returns perfectly cancel out. The empirical estimate is far above −1 due to the noise in stock returns.

3. Seasonalities and seasonal reversals in Fama-MacBeth regressions

3.1. Monthly data

3.1.1. Multivariate regressions

Table 1 reports the estimates from Fama-MacBeth regressions that predict the cross section of monthly stock returns. We estimate these regressions for all stocks, for all stocks except microcaps, and for 48 value-weighted Fama-French industry portfolios. Microcaps are stocks with market values of equity below the 20th percentile of the NYSE market capitalization distribution as of the end of month $t-1$.

Regressions 1, 4, and 7 predict returns using log size, log book-to-market, past-month return, prior one-year return skipping a month, and average same-month return. We compute this average return from cross-sectionally demeaned returns using up to 20 years of historical data.[8] We find that average returns decrease in size and increase in both book-to-market and momentum. These three effects are statistically significant both for all stocks and for the sample excluding microcaps. The estimated slope on the average same-month return is positive and statistically significant. Its t-value is 9.88 in the regressions that include all stocks and 8.70 in the sample excluding microcaps. This effect, which is consistent with the estimates in Heston and Sadka (2008) and Keloharju et al. (2016), is economically large. The coefficient estimate of 5.5 in the full sample, for example, implies that a 1% difference in past average same-month returns between two stocks predicts a 0.055% difference in these stocks' returns this month.

In Regressions 2, 5, and 8, we add a variable denoting the average other-month return to the model. We find that the estimated slope on this variable is negative and statistically significant. Its t-value is −6.57 in the full sample and −4.50 in the sample excluding microcaps. This effect is economically even larger. A 1% difference in the average other-month returns in Regression 2 translates into a −0.19% difference in monthly returns today.

The fact that both the same- and other-month returns remain significant is consistent with seasonalities being balanced by seasonal reversals. Although both variables measure the same underlying quantity, the stock's ex-

[8] If all stocks have the same amount of historical data, the cross-sectional demeaning does not change the estimates because the demeaning shifts all averages up or down by the same amount. Demeaning ensures that the average same-month returns of stocks with different amounts of historical data are comparable (Keloharju et al., 2016).

Table 1

Average same-month and other-month returns in Fama-MacBeth regressions.

This table presents average Fama and MacBeth (1973) regression slopes and their t-values from cross-sectional regressions that predict monthly returns. The regressions use data from January 1963 to December 2016 for all stocks (Regressions 1–3), all-but-microcaps (Regressions 4–6), and 48 value-weighted Fama-French industries (Regressions 7–9). Microcaps are stocks with market values of equity below the 20th percentile of the NYSE market capitalization distribution. The sample includes stocks with at least five years of historical same-month returns. We cross-sectionally demean the data before computing the same- and other-month returns, $\bar{r}_{\text{same-month}}$ and $\bar{r}_{\text{other-month}}$. Both averages use up to 20 years of historical data. The average other-month return skips a year; that is, to predict the cross section month t return, the first term in the other-month average is the month $t-13$ return. Regression estimates are multiplied by one hundred.

Explanatory	All stocks			All-but-microcaps			Industries		
variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(ME)	−0.07	−0.07	−0.06	−0.06	−0.07	−0.07	−0.04	−0.03	−0.03
	(−2.25)	(−2.01)	(−1.95)	(−1.97)	(−2.41)	(−2.42)	(−1.61)	(−1.16)	(−1.09)
log(BE/ME)	0.30	0.20	0.18	0.22	0.12	0.10	0.15	0.02	0.01
	(5.84)	(4.33)	(4.00)	(3.91)	(2.45)	(2.06)	(1.78)	(0.26)	(0.06)
r_1	−5.54	−5.58	−5.62	−3.39	−3.52	−3.56	4.74	4.80	4.24
	(−15.54)	(−15.82)	(−16.00)	(−8.07)	(−8.56)	(−8.76)	(4.48)	(4.49)	(4.07)
$r_{12,2}$	0.46	0.44	0.43	0.42	0.41	0.40	1.24	1.21	1.17
	(2.98)	(2.92)	(2.86)	(2.21)	(2.16)	(2.13)	(3.49)	(3.39)	(3.35)
$r_{60,13}$			−0.06			−0.05			−0.05
			(−2.10)			(−2.10)			(−0.60)
$\bar{r}_{\text{same-month}}$	5.47	4.67	4.93	6.52	6.14	6.56	19.10	17.98	19.22
	(9.88)	(8.17)	(8.56)	(8.70)	(8.10)	(8.66)	(5.92)	(5.46)	(5.87)
$\bar{r}_{\text{other-month}}$		−18.51	−16.05		−16.36	−12.84		−42.94	−35.25
		(−6.57)	(−5.71)		(−4.50)	(−3.46)		(−3.33)	(−2.64)

pected return $\mu_{i,m(t)}$, they are incrementally informative because returns are noisy signals of expected returns.

Because the average same- and other-month returns closely relate to long-term reversals, we add these reversals in Regressions 3, 6, and 9 as controls. We use the usual definition of long-term reversals, measuring stock returns over the prior five-year period and skipping a year. We find that the addition of these long-term reversals has only a modest effect on the slope estimates for the average same- and other-month returns. The coefficients and t-values on the average same-month return increase slightly, and those on the average other-month return decrease slightly. The long-term reversal variable itself is significant, with a t-value of −2.10 in both the full sample and in the sample excluding microcaps.

The industry estimates in Regressions 7–9 show that seasonal reversals are also present in the returns of value-weighted industry portfolios. Because these portfolios are well diversified, this significance suggests that seasonal reversals are unlikely to emanate from stock-specific effects. Although some patterns in industry returns differ from those in stock returns [most important, industries display significant momentum already in month 1 (Moskowitz and Grinblatt, 1999) and the cross-industry value effect is, at best, weak (Cohen and Polk, 1996; Novy-Marx, 2013)], the patterns related to seasonalities and seasonal reversals are similar. The industry results also further highlight the difference between seasonal and long-term reversals. While the long-term reversals estimate is within just one standard error from zero in Regression 9, the t-value associated with seasonal reversals is −2.64.

In Table A1 in the Appendix, we sort stocks into deciles by average same- and other-month returns and form strategies that are long the top portfolio and short the bottom portfolio. A strategy that selects stocks based on their average same-month returns earns an average return of 1.03% per month (t-value = 7.19), and this strategy's alpha from the Carhart (1997) four-factor model augmented with the long-term reversals factor is 1.09% (t-value = 7.19). A strategy that selects stocks based on their other-month returns earns an average return of −0.87% (t-value = −5.55) and has an alpha of −0.53% (t-value = −4.74). Consistent with the cross-sectional regressions of Table 1, these portfolio-based estimates suggest that seasonal reversals are economically and statistically significant and largely unrelated to long-term reversals.

3.1.2. Visualizing seasonalities, seasonal reversals, and the adding-up constraint

Panel A of Fig. 4 plots average cross-sectional covariances between month t and $t+k$ returns, $\text{cov}^{cs}(r_{i,t}, r_{i,t+k})$. Seasonalities in this panel are the positive spikes that occur at every 12-month horizon. These positive spikes indicate that a high return today predicts high returns exactly 12, 24,... months hence. This pattern is very similar to that in, for example, Heston and Sadka (2008).

In Panel B, we plot cumulative sums of the covariances reported in Panel A. The resulting pattern summarizes the relation between past and future returns. Returns at one- and two-month lags negatively predict returns (short-term reversals); returns up to 12 months positively predict returns (momentum); and returns from 12 months to approximately five to seven years negatively predict returns (long-term reversals). After seven years, the association between past and future returns is not economically significant. Despite the continuing annual spikes in Panel A, Panel B displays no drift in the cumulative sums of covariances after the end of the long-term reversal. If $\mu^s_{i,m(t)}$ is the seasonal component of stock i's expected return, Panel B suggests that, for example, $\mu^s_{i,\text{dec}} \approx -(\mu^s_{i,\text{jan}} + \cdots + \mu^s_{i,\text{nov}})$. That is, the fact that the cumulative sums flatten

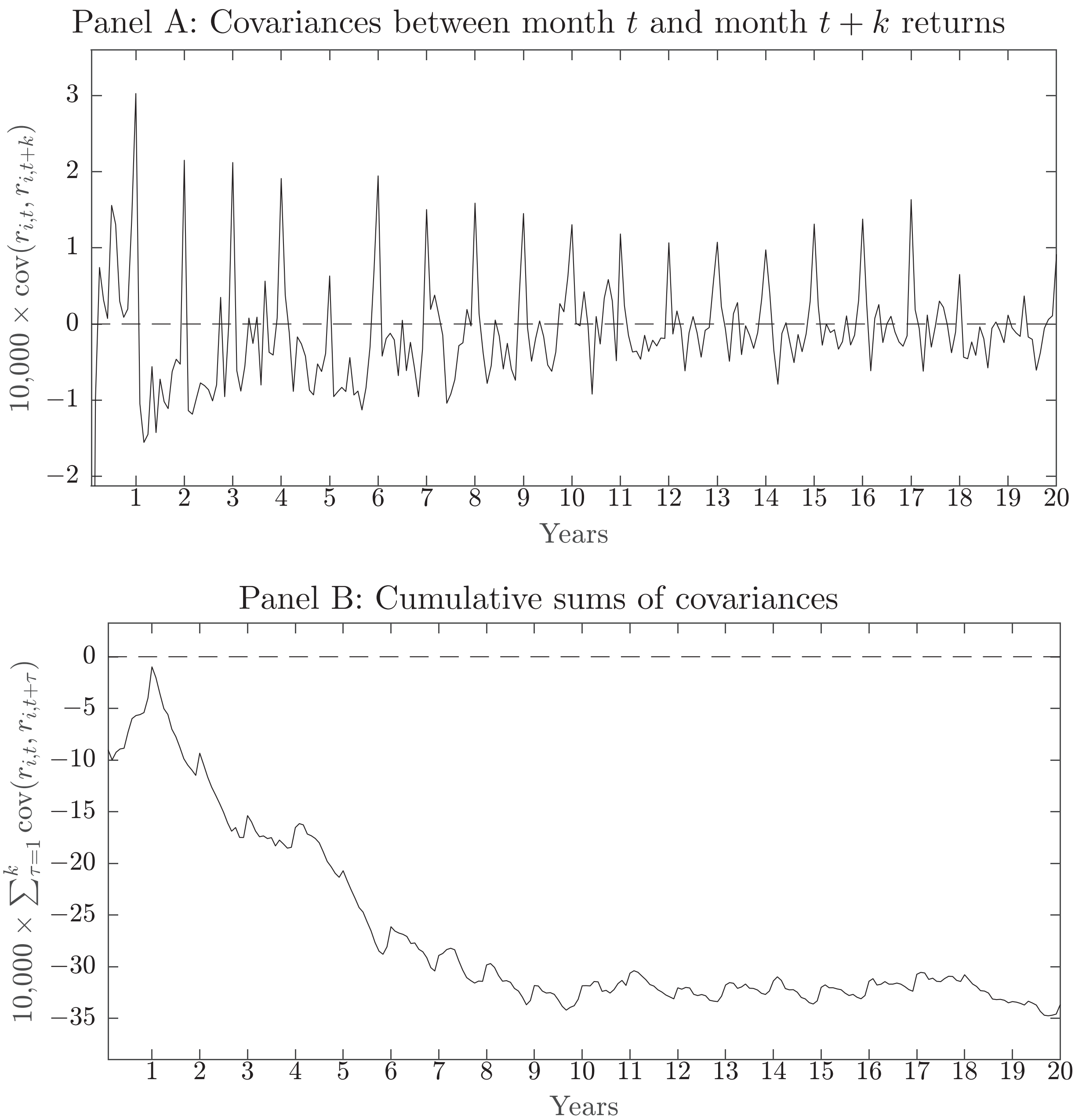


Fig. 4. Seasonalities and seasonal reversals. Panel A plots average cross-sectional covariances between month t and month $t+k$ returns, where k ranges from one to 240 months. Panel B plots cumulative sums of these average cross-sectional variances. The line at time $k = 10$, for example, is the sum of the first ten covariances from Panel A. The sample begins in January 1963 and ends in December 2016 and consists of all common stocks listed on NYSE, Amex, and Nasdaq.

out indicates that seasonal reversals almost perfectly balance out seasonalities.

Panel B of Fig. 4 also illustrates that, depending on horizon, measuring the extent to which seasonalities and seasonal reversals cancel each other out could be difficult. Depending on the horizon, past returns can predict returns today not only because of seasonalities and seasonal reversals, but also because of autocorrelations. Past returns up to a year positively predict returns because of momentum; and past returns from one year up to perhaps seven years negatively predict returns because of long-term reversals. In the tests that follow, we address the complications arising from these autocorrelations in two ways. First, we compare the same- and other-month regression coefficients for different formation periods. This difference is unaffected by the overall level of the autocorrelation. Second, if we assume that autocorrelations die out at some point, for example, if long-term reversals are unlikely to last for more than ten years, we can examine the extent to which seasonalities and seasonal reversals balance out after we

Table 2

Average same-month and other-month returns in Fama-MacBeth regressions: alternative formation periods.

This table presents the average Fama and MacBeth (1973) regression slopes and their *t*-values from cross-sectional regressions that predict monthly returns. The regressions predict returns using the average of all past returns, average of same-month returns, average of other-month returns, and the difference between the average same- and other-month returns. We estimate each specification as a univariate regression. Row "1" uses data from $t-12$ through $t-1$; "2–5" returns from $t-60$ to $t-13$; and so forth. The regressions use data from January 1963 through December 2016 for all stocks, all stocks except microcaps, and the 48 value-weighted Fama-French industries. Microcaps are stocks with market values of equity below the 20th percentile of the NYSE market capitalization distribution. We cross-sectionally demean returns when computing the averages of past returns. Regression estimates are multiplied by one hundred.

	Construction of historical average return							
	All		Same-month return		Other-month return		Same-month -other-month	
Years	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$
All stocks								
1	2.10	1.14	1.70	6.88	0.20	0.11	1.46	5.65
2–5	−13.63	−4.17	3.07	5.83	−17.45	−5.73	3.91	7.83
6–10	−7.58	−2.78	4.19	7.23	−12.27	−5.02	4.60	8.64
11–15	−0.06	−0.02	4.41	6.93	−4.97	−2.06	4.27	7.10
16–20	−3.94	−1.57	3.73	5.30	−8.36	−3.35	3.79	5.71
All-but-microcaps								
1	7.59	3.26	1.65	4.47	6.25	2.76	1.01	2.74
2–5	−10.93	−3.21	2.67	4.31	−14.18	−4.36	3.63	6.16
6–10	−5.21	−1.68	3.97	6.24	−9.07	−3.19	4.35	7.40
11–15	−0.89	−0.35	3.40	5.20	−4.00	−1.63	3.40	5.42
16–20	−3.66	−1.44	3.31	4.45	−7.05	−2.84	3.50	4.97
Industries								
1	24.93	5.54	4.94	4.47	22.54	5.22	2.06	1.89
2–5	−2.20	−0.33	1.10	0.55	−3.58	−0.55	2.93	1.59
6–10	−19.04	−2.67	6.83	3.37	−25.13	−3.64	8.52	4.30
11–15	−6.05	−0.98	6.04	3.21	−9.68	−1.64	6.91	3.70
16–20	−14.14	−2.21	5.01	2.30	−16.13	−2.68	6.00	2.94

skip over the first ten years. Panel B of Fig. 4 already suggests that seasonalities and seasonal reversals balance out. Below, we test for the association between the two.

3.1.3. Comparing same-month and other-month regressions

To test the extent to which seasonalities reverse, we compare the slope coefficients of same-month and other-month returns. Table 2 reports the results from regressions that predict the cross section of monthly returns with all, same-month, and other-month returns. For example, when we predict returns with the average returns from five to two years prior (row 2–5 in the table), the "All" regression uses the average of all 48 returns from month $t-60$ to $t-13$; the "Same-month" regression uses the average of the four returns in months $t-60$, $t-48$, $t-36$, and $t-24$; and the "Other-month" regression uses the average return over the other 44 months. We choose not to include the control variables such as the momentum from Table 1, because we do not want to use regressors relating to the other return windows.

Differences between the all, same-month, and other-month coefficients represent seasonal reversals. To see why, assume, as in Eq. (4), that stock returns take the form of $r_{it} = \mu_{i,m(t)} + \epsilon_{it}$. The regression coefficient that explains the cross section of monthly returns with lagged same-month returns is then $\hat{b}_{\text{same-month}} = \frac{\text{cov}(r_{i,t}, r_{i,t-k})}{\text{var}(r_{i,t-k})} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\epsilon^2}$ for lags k that are multiples of 12. If seasonalities in expected returns do not reverse, the coefficients from the all and same-month regressions are equal. If there are seasonal reversals, the value of the same-month coefficient exceeds that of the all coefficient. If seasonalities in expected returns completely reverse, the coefficient from the all regression is zero. These cases correspond to Models 2 and 3 in Fig. 1. We measure the extent to which Panels A and B of this figure differ in the data at different horizons.

Absent any complications, seasonal reversals should leave clear markers in the data. The coefficient from the same-month regression should be positive, that from the other-month regression should be negative, and that from the all regression coefficient should be zero. We must address two concerns. First, past returns can predict returns today because stock returns are autocorrelated. Fig. 4 shows how short-term reversals, momentum, and long-term reversals are intertwined with seasonalities and seasonal reversals. Because of such autocorrelations, the slope from the all regression is not a clear test of the extent to which seasonalities reverse. Second, our statistical tests yield the inconvenient prediction that the all-months coefficient is zero. An estimate that is not statistically significantly different from zero should not be construed as evidence for accepting the null hypothesis. A comparison of the same- and other-month regression coefficients circumvents both issues. In the presence of seasonalities and seasonal reversals, these two coefficients are predicted to differ from zero with opposite signs.

Table 2 reports these coefficients from regressions that predict the cross sections of monthly stock and industry returns. We explain returns with average returns in years 1, 2–5, 6–10, 11–15, and 16–20. We compare different periods of prior returns with separate seasonal reversals from long-term reversals. The negative all coefficient of −13.6 (t-value = −4.17) on row 2–5 for all stocks is consistent with these long-term reversals. When we regress today's stock returns against the five-year average return from years 11–15, the average all coefficient is close to zero. Therefore, by skipping ten years, we appear to skip over most, or all, of the long-term reversals. Over the same 11–15-year period, the average same-month return coefficient is significant with t-values of 6.93; the other-month return is significant with a t-value of −2.06; and the difference between the two has a t-value of 7.10. The t-values associated with all returns, same-month returns, and other-month returns are −1.57, 5.30, and −3.35, respectively, when we skip 15 years before we begin measuring average returns.

These estimates suggest that the seasonalities in individual stock returns reverse completely. The significantly positive same-month coefficient, as before, indicates that there are seasonalities in expected stock returns. The fact that the average same-month coefficient exceeds the all coefficient indicates that some of these seasonalities reverse. Assuming that autocorrelation patterns such as momentum and long-term reversals tend to zero as we lengthen the formation period, the finding that the all coefficient at long horizons is close to zero is consistent with the perfect reversal of seasonalities in individual stock returns. This result is consistent with the flattening of the cumulative sums of autocovariances in Panel B of Fig. 4.

The estimates for the value-weighted industry portfolios are consistent with the seasonalities in expected industry returns reversing perfectly as well. In the regression with the 11–15 years formation period, for example, the t-values associated with the same-month and other-month returns are 3.21 and −1.64, respectively, and the difference between the two has a t-value of 3.70. Seasonal reversals in industry portfolios relate to those in individual stock returns. Keloharju et al. (2016) show that a substantial part of seasonalities in individual stock returns stems from seasonalities in industry returns. At the same time, expected returns do not appear to vary significantly across industries (Moskowitz and Grinblatt, 1999). The existence of seasonal reversals reconciles these two sets of findings: long-term predictability in monthly returns in the absence of persistent differences in average returns.

If seasonalities reverse, both the same- and other-month average returns are predictive of the cross section of stock returns through the same mechanism. A high average December return, for example, predicts high December returns, but so must a low average non-December return. The regressions in the last column of Table 2 predict returns using the difference between the same- and other-month average returns. If both the same- and other-month average returns are noisy versions of the same economic signal (the seasonal return component), their combination should better predict returns than either of the two proxies in isolation. Consistent with this prediction, outside the one-year momentum effect, the t-values associated with the estimates in the last column always exceed those in the other columns.

These results are not sensitive to inclusion of the month of January in the analysis. Table A2 in the Appendix shows that our results remain qualitatively similar even if we confine the predicted returns to the months of February through December.

3.2. Daily data

Keloharju et al. (2016) show that stocks also have strong seasonalities at the day-of-the-week level. For example, a stock that has historically done well on Mondays has higher expected returns on Mondays. We now show that, similar to monthly seasonalities, daily seasonalities are balanced by seasonal reversals.

Table 3 reports estimates from Fama-MacBeth regressions that predict the cross section of daily returns. We predict returns using the average return computed over all days, the same weekdays, the other weekdays, and the difference between the two. We estimate these averages using historical data in years 1, 2–5, 6–10, 11–15, and 16-20.

The estimates for all days are similar to those in Table 2, except that the dependent variable is a daily instead of a monthly return. Consistent with Keloharju et al. (2016), the daily returns are highly seasonal. Same-weekday returns have significantly positive coefficients and the other-weekday returns have significantly negative coefficients beyond the first year. Because these two effects cancel each other out, that is, seasonal reversals balance out seasonalities, the estimates in the "All days" column are close to zero beyond the first year.

In the context of daily returns, the idea that seasonalities emanate from temporary mispricing is not controversial. Birru (2018) finds daily seasonalities are associated with stocks that are hard to value or have the greatest impediments to arbitrage. Lakonishok and Maberly (1990) find that individuals trade more frequently on Mondays and are more frequently net sellers, and that this pattern correlates with the weekend effect (French, 1980). Chan et al. (2004) add to this evidence by showing that the Monday seasonal is stronger in stocks with low institutional holdings and that the weakening of the Monday seasonal in the 1990s coincided with the increase in institutional holdings. These lower-than-average Monday returns could stem from retail investors' correlated selling pressure that is not absorbed by the rest of the market without a price impact. If so, we would expect non-Monday returns to be higher by an offsetting amount. Monday returns are low only because excess supply from retail investors temporarily depresses share prices.

3.3. Countries and commodities

Heston and Sadka (2010) show that return seasonalities exist not only in the cross section of US stock returns (Heston and Sadka, 2008) but also within international stock markets. Keloharju et al. (2016) show that seasonalities are not confined to equity returns. They find seasonalities also in stock market indices and commodity returns.

Table 3
Daily Fama-MacBeth regressions.
This table presents average Fama and MacBeth (1973) regression slopes and their t-values from cross-sectional regressions that predict daily returns. The first model predicts returns with the average return computed over all days; the second model, with the same-weekday average; the third model, with the other-weekday average; and the fourth model, with the difference between the two. The regressions use data from January 1963 to December 2016 for all stocks. The same-weekday average is, e.g., the average Monday return during the period reported in the "Years" column when the regression predicts the cross section of Monday returns; in regressions that predict Tuesday returns, it is the average Tuesday return and so forth. "Same-other" is the average same-weekday return minus the average other-weekday return. Regression estimates are multiplied by one hundred.

	All days		Same weekday		Other weekday		Same-other	
Years	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$
1	12.73	9.75	7.39	22.05	3.02	3.02	4.99	19.57
2–5	−3.31	−1.65	10.08	16.23	−15.28	−9.60	9.33	18.31
6–10	0.54	0.26	9.91	13.55	−13.11	−7.47	8.77	13.91
11–15	3.91	1.73	8.70	10.40	−9.25	−4.79	7.47	10.30
16–20	−1.91	−0.72	8.16	8.93	−13.42	−6.05	7.48	9.46

If seasonalities in different markets and asset classes emanate from the same mechanism, we would expect to find seasonal reversals in other markets and asset classes as well. Following the studies listed above, we measure seasonal reversals in the cross sections of international stock returns, stock market indices, and commodities.

Panel A of Table 4 reports the coefficients from cross-sectional regressions that predict the cross section of international stock returns.[9] These regressions are the same as those for Table 2 for the US stock market. To create the sample, we pool stocks from 25 developed countries, excluding the US. We remove the 20% of the smallest firms from each country each month. Stocks from a country are included in the sample in month t if the country's month-t cross section has at least 50 stocks for which we can compute the independent variables over the horizon specified in the "Years" column. The left-hand side return data in the cross-sectional regressions start in January 1987 and end in December 2016.

The international results in Panel A are similar to the US results. The same-month coefficients show that the same-month stock return is significantly predictive of future stock returns up to 20 years into the future. This seasonality result is consistent with the results in Heston and Sadka (2010). These seasonalities are again offset by seasonal reversals: the other-month regressions are negative and statistically significant at the 5% level for years 2–15 and at the 10% in the years 16–20 specification. Because of these reversals, the average of all prior returns in column 1 is either not predictive of the cross section of stock returns (years 6–10 and 16–20) or significantly negatively predictive of these returns.

Panel B of Table 4 reports estimates from cross-sectional regressions that predict the cross section of international stock index and commodity returns. Seasonalities are balanced out by seasonal reversals in both asset classes. For example, the all coefficient for equity indices has a t-value very close to zero, −0.16, and the same-month coefficient has a t-value of 2.33 and the other-month coefficient a t-value of −1.29. The statistical evidence of seasonal reversals is not particularly impressive due to the small number of assets (15 country equity indices and 25 commodities), but the results are in line with our other results.

In Panel C, we measure seasonalities and seasonal reversals in daily stock index and commodity returns. We estimate cross-sectional regressions in which the dependent variable is day d return, and the explanatory variable is the average return computed using all, same-weekday, or other-weekday returns or the difference between the same- and other-weekday returns. We account for momentum by dividing the look-back period into two parts, the prior year and the prior 20 years skipping a year. In the prior-20 year specification, we expand the window as we accumulate more data.

The evidence of daily seasonalities and reversals is weak for stock indices. We find momentum up to a year, and the regressions largely attribute this momentum to the same-weekday returns. The same-weekday coefficient is significant with a t-value of 3.36, and the other-weekday coefficient has a t-value of 0.49. We find evidence of neither seasonalities nor reversals after the one-year mark.

Commodities exhibit strong and long-lasting seasonalities. The first row shows a strong momentum effect in this market, consistent with Asness et al. (2013). The average commodity return over the prior year predicts the cross section of daily commodity returns with a t-value of 3.16. This predictability again concentrates to the same-weekday returns. The t-values for the same- and other-weekday returns are 5.42 and 0.62, respectively. The estimates on the second row show that seasonalities remain after momentum subsides. The same-weekday return now predicts returns with a t-value of 4.10. At the same time, seasonal reversals remain as well to balance out return seasonalities. The other-weekday return predicts returns with a t-value of −2.91. Because of these two offsetting effects, the association between returns and the average of all past returns in the "All" column is negative and statistically insignificant with a t-value of −0.74.

The results on seasonalities and seasonal reversals in monthly US equity returns could be dismissed as a chance finding. Despite the high t-values, seasonal reversals could offset seasonalities in one asset class (US equities) and at one frequency (monthly) just by luck. The results in this

[9] The stock return data are from 25 developed countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Hong Kong, Ireland, Israel, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Singapore, South Korea, Spain, Sweden, Switzerland, and United Kingdom.

Table 4
Seasonalities and seasonal reversals in international stock returns, country-level stock indices, and commodities.
This table reports results from Fama and MacBeth (1973) regressions that predict the cross section of international stock returns (Panel A) and country equity indices and commodities (Panels B and C). We compute average returns using all, same-month, and other-month past returns and the difference between the same- and other-month returns. International stock market data in Panel A are from Datastream and pool stocks from 25 developed countries. The first cross-sectional regression has January 1987 returns on the left-hand side. We remove the 20% of the smallest firms from each country each month. Stocks from a country are included in the sample in month t if the country's month t cross section has at least 50 stocks for which we can compute the independent variables over the horizon specified in the "Years" column. The regressions include country fixed effects. We report the results separately for five different horizons. The country equity index and commodities data in Panels B and C are from Datastream and Bloomberg. Table A3 in the Appendix lists the 15 countries and 24 commodities included in this analysis. The monthly country index and commodities data begin in January 1970, and we run the first cross-sectional regressions in Panel B with January 1975 returns on the left-hand side. These first regressions use five years of historical data to construct the right-hand-side variables, and this window expands up to 20 years as we accumulate more historical data. The country index regressions have the same set of countries throughout the sample because all countries enter the sample in January 1970. The number of commodities increases over time. We add a new commodity to the sample when it has at least five years of historical data. We have four commodities in 1975, ten in 1983, and the full set of 24 in 2007. In both the stock index and commodity regressions, we skip over the most recent year when computing the other-month return. The daily stock index regressions in Panel C start in January 1987 and the daily commodity regressions in January 1975. The last regression in each panel uses December 2016 return data. In all panels, the regression estimates are multiplied by one hundred.

Panel A: International stock returns

	All		Same-month return		Other-month return		Same-month -other-month	
Years	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$
1	3.39	3.51	1.16	5.08	2.96	2.89	0.59	3.17
2–5	−7.90	−5.58	1.39	4.64	−9.02	−6.43	2.09	7.39
6–10	−1.67	−1.32	2.76	7.64	−4.08	−3.51	2.87	8.91
11–15	−4.24	−2.44	2.25	5.03	−7.22	−4.16	2.49	5.77
16–20	−0.11	−0.05	2.44	4.86	−3.42	−1.72	2.32	4.70

Panel B: Monthly regressions using country-level stock indices and commodities

	All		Same-month return		Other-month return		Same-month -other-month	
Asset class	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$
Stock indices	−3.44	−0.16	11.00	2.33	−24.68	−1.29	11.89	2.68
Commodities	10.43	0.46	9.72	1.40	−18.26	−0.99	13.52	2.21

Panel C: Daily regressions using country-level stock indices and commodities

	All		Same-day return		Other-day return		Same-day -other-day	
Asset class, years	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$
Stock indices, 1	60.16	1.56	11.59	3.36	4.07	0.49	7.65	2.42
Stock indices, 2–20	28.53	0.80	−3.32	−0.32	−0.77	−0.03	0.03	0.00
Commodities, 1	20.45	3.16	14.36	5.42	3.48	0.62	9.06	3.75
Commodities, 2–20	−13.84	−0.74	30.56	4.10	−47.68	−2.91	28.39	4.47

section suggest that seasonalities are almost certainly intertwined with seasonal reversals. It seems that wherever return seasonalities exist, so do offsetting seasonal reversals.

4. Additional empirical tests

4.1. Do seasonal reversals relate to seasonality in sentiment?

Seasonal reversals are consistent with seasonalities stemming from mispricing. Mispricing, in turn, can relate to sentiment, that is, investors' excessive optimism or pessimism could systematically affect asset valuations.[10] Hirshleifer et al., 2020 and Birru (2018) suggest that seasonalities in sentiment can induce return seasonalities. An asset that performs well during a mood state, a period during which sentiment is abnormally high or low, tends to perform well during future congruent mood states and poorly during future noncongruent mood states. These states are determined based on research on, e.g., the seasonal affective disorder (SAD) effect and its relation to historical average stock returns (Kamstra et al., 2003).

In Table 5, we examine the relation between seasonalities, seasonal reversals, and mood states at monthly (Panel A) and daily (Panel B) frequencies. Hirshleifer et al., 2020 identify January and March as high-mood months and September and October as low-mood months. Panel A shows that monthly seasonalities and seasonal reversals exist in the data both during these high- and low-mood months (Regressions 2 and 6) and outside these four months (Regressions 3 and 7). In the full sample with all stocks (Regression 1), average other-month return predicts returns with a t-value of −5.71. In the regressions that split the sample into mood and non-mood months, the t-values are −4.00 and −4.16, respectively.

[10] See, for example, Baker and Wurgler (2006).

Table 5

Seasonalities and seasonal reversals in monthly and daily Fama-MacBeth regressions: association with high- and low-mood betas.

This table presents average Fama and MacBeth (1973) regression slopes and their t-values from cross-sectional regressions that predict monthly (Panel A) or daily (Panel B) returns. The samples and the main explanatory variables are the same as in Tables 1 and 3. In Panel A, Regressions 2 and 6 predict the cross sections of high- and low-mood month returns; Regressions 3 and 7 predict returns outside these months; and Regressions 4 and 8 control for the historical mood betas of Hirshleifer et al. (2020). In Panel B, we report full-sample estimates and estimates from regressions that predict only the cross sections of Monday and Friday or Tuesday, Wednesday, and Thursday returns. Regression estimates in both panels are multiplied by one hundred.

Panel A: Monthly regressions

Explanatory variable	All stocks (1)	(2)	(3)	(4)	All-but-microcaps (5)	(6)	(7)	(8)
log(ME)	−0.06	−0.25	0.03	−0.06	−0.07	−0.09	−0.07	−0.07
	(−1.95)	(−4.21)	(0.78)	(−2.17)	(−2.42)	(−1.60)	(−1.81)	(−2.67)
log(BE/ME)	0.18	0.11	0.22	0.13	0.10	0.14	0.07	0.07
	(4.00)	(1.36)	(3.92)	(2.97)	(2.06)	(1.73)	(1.30)	(1.54)
r_1	−5.62	−8.13	−4.37	−5.82	−3.56	−5.32	−2.69	−4.24
	(−16.00)	(−11.09)	(−11.97)	(−16.03)	(−8.76)	(−6.57)	(−5.95)	(−10.52)
$r_{12,2}$	0.43	−0.13	0.71	0.44	0.40	0.02	0.59	0.37
	(2.86)	(−0.43)	(4.10)	(2.81)	(2.13)	(0.07)	(2.52)	(1.92)
$r_{60,13}$	−0.06	−0.21	0.02	−0.07	−0.05	−0.19	0.02	−0.06
	(−2.10)	(−3.59)	(0.71)	(−2.55)	(−2.10)	(−3.57)	(0.62)	(−2.22)
$\bar{r}_{\text{same−month}}$	4.93	5.67	4.56	6.60	6.56	9.03	5.32	7.55
	(8.56)	(5.62)	(6.50)	(9.64)	(8.66)	(6.32)	(6.06)	(9.16)
$\bar{r}_{\text{other−month}}$	−16.05	−19.67	−14.23	−14.03	−12.84	−11.70	−13.41	−14.32
	(−5.71)	(−4.00)	(−4.16)	(−3.98)	(−3.46)	(−1.80)	(−2.96)	(−3.72)
$\hat{\beta}_{\text{mood}}$				0.15				0.12
				(1.27)				(0.93)
Sample	All	Jan, Mar, Sep, Oct	Other	All	All	Jan, Mar, Sep, Oct	Other	All

Panel B: Daily regressions

Regression specification	Independent variable	Full sample $\hat{b}$	$t(\hat{b})$	Sample with only Monday and Friday $\hat{b}$	$t(\hat{b})$	Sample without Monday and Friday $\hat{b}$	$t(\hat{b})$
1	Total return	−3.31	−1.65	−9.37	−2.93	0.57	0.22
2	Same-day return	10.08	16.23	17.90	17.86	5.06	6.44
3	Other-day return	−15.28	−9.60	−31.29	−11.97	−5.01	−2.51
4	Same-other	9.33	18.31	17.04	19.69	4.39	7.08

Regressions 4 and 8 control for mood betas we estimate following the Hirshleifer et al. (2020) procedure. At the end of each year, a stock's mood beta is estimated from the capital asset pricing model regression using at most ten years of monthly data. To isolate a stock's sensitivity to seasonal variation in mood, the sample is restricted to the pre-defined high-mood (January and March) and low-mood (September and October) months and, in addition, to the CRSP equally weighted index's realized high- and low-mood months.[11] In the regressions with all stocks, the average other-month return predicts returns with a t-value of −3.98 when we control for mood betas.[12] In the all-but-microcaps sample, the other-month return enters with a t-value of −3.72. In Table A4 in the Appendix, we report the regression estimates separately for the high- and low-months. We find that the average slope for the mood beta is significantly positive in high-mood months (t = 3.02) and significantly negative in low-mood months (t = −2.92). Both of these estimates are consistent with the results in Hirshleifer et al. (2020). These high- and low-mood slopes offset each other in the pooled sample and, therefore, the slopes on the mood beta are close to zero and statistically insignificant in Regressions 4 and 8 of Table 5. This adding-up behavior is akin to our finding that the seasonalities and seasonal reversals add up to zero.

Panel B of Table 5 reports univariate regressions similar to those in Table 3. Birru (2018) identifies Fridays as high-mood and Mondays as low-mood days. We find that seasonalities and seasonal reversals are more pronounced in the regressions that predict the cross sections of Monday and Friday returns. In these regressions, the average other-day return predicts returns with a t-value of −11.97. In the regressions that predict the Tuesday through Thursday returns, this t-value is −2.51.

The results in Table 5 should not be interpreted as suggesting that seasonal reversals are distinct from seasonalities in sentiment. Instead, we believe that systematic seasonal variation in sentiment is likely an important source

[11] We replicate the key result of Hirshleifer et al. (2020) to confirm that our beta estimates are correct. In their Table 4, Hirshleifer et al. predict the cross sections of January, March, September, and October returns with mood betas. The authors use β^{mood} to predict returns in the high-mood months and $-\beta^{\text{mood}}$ in the low-mood months. Using the average estimated betas from years 2 through 5, Hirshleifer et al. find an estimate of 1.47 with a t-value of 4.83. In our replication we get an estimate of 1.63 with a t-value of 5.17.

[12] We follow Hirshleifer et al. (2020) and control in the regression for the average mood beta from year 2 through 5.

Table 6
Seasonalities and proxies for limits to arbitrage.
This table presents average Fama and MacBeth regression slopes and their t-values from cross-sectional regressions that predict monthly returns. The regressions are similar to those in Table 1 except that we add seasonalities and seasonal reversals interacted with firm size, idiosyncratic volatility, and institutional ownership. Idiosyncratic volatility is the standard deviation of the residuals from the Fama and French (1993) three-factor model regression, estimated using daily returns over month t. Institutional ownership is size-adjusted. It is the residual from a cross-sectional logit regression in which the dependent variable is the share of institutional ownership and the independent variable is the square of the logarithm of firm size. In these regressions all independent variables are converted into z-scores separately in each cross section. Each variable is mean zero with unit standard deviation. Regressions 1 and 2 use data for all stocks; Regressions 3 and 4, for all-but-microcaps. Microcaps are stocks with market values of equity below the 20th percentile of the NYSE market capitalization distribution. The sample includes stocks with at least five years of historical same-month returns. We cross-sectionally demean the data before computing the same- and other-month returns, $\bar{r}_{\text{same-month}}$ and $\bar{r}_{\text{other-month}}$. Both averages use up to 20 years of historical data. The average other-month return skips a year; that is, to predict the cross section month t return, the first term in the other-month average is the month $t-13$ return. Regression estimates are multiplied by one hundred. The regressions use data from April 1980 through December 2016.

	All stocks				All-but-microcaps			
Explanatory	(1)		(2)		(3)		(4)	
variable	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$
Book-to-market	0.12	2.75	0.13	2.87	0.03	0.76	0.03	0.89
Short-term reversals	−0.65	−10.37	−0.64	−10.39	−0.28	−5.15	−0.29	−5.33
Momentum	0.18	2.69	0.17	2.64	0.09	1.22	0.08	1.15
Long-term reversals	−0.03	−0.98	−0.03	−1.14	−0.04	−1.31	−0.02	−0.68
$\bar{r}_{\text{same-month}}$	0.18	6.04	0.24	7.30	0.22	7.05	0.27	7.89
$\bar{r}_{\text{other-month}}$	−0.19	−4.77	−0.15	−3.60	−0.12	−2.91	−0.10	−2.46
Size	−0.08	−1.20	−0.10	−1.44	−0.09	−2.03	−0.09	−1.97
$\times\bar{r}_{\text{same-month}}$			0.06	1.79			0.08	4.16
$\times\bar{r}_{\text{other-month}}$			0.06	1.92			0.01	0.69
Idiosyncratic volatility	−0.03	−0.37	−0.08	−0.90	−0.15	−2.38	−0.16	−2.41
$\times\bar{r}_{\text{same-month}}$			−0.03	−1.08			−0.05	−2.45
$\times\bar{r}_{\text{other-month}}$			−0.04	−1.40			−0.05	−1.90
Institutional ownership	0.13	3.95	0.13	3.82	0.05	1.53	0.04	1.23
$\times\bar{r}_{\text{same-month}}$			0.04	2.10			0.03	1.84
$\times\bar{r}_{\text{other-month}}$			−0.01	−0.35			0.02	1.14

of mispricing and, therefore, by extension, of seasonalities and seasonal reversals. In this sense, our analyses can be viewed as complementary to those in Hirshleifer et al. (2020) and Birru (2018). The explanation we offer to the return seasonalities and seasonal reversals is more general, as it also includes mechanisms likely unrelated to sentiment such as predictable trading by individuals or institutions, or both. Regardless of what is the underlying mechanism behind the seasonalities, seasonal reversals would go hand in hand with any seasonal mispricings.

4.2. Limits to arbitrage

If seasonalities and seasonal reversals reflect the emergence and correction of mispricings, the strength of these effects can vary depending on the extent to which arbitrage is limited. In Table 6, we report estimates from regressions that predict the cross section of returns using same- and other-month returns, as well as the interactions between these variables and three measures of limits to arbitrage.

The three measures of limits to arbitrage are firm size, idiosyncratic volatility, and size-adjusted institutional ownership (IO).[13] We follow Ang et al. (2006) and define idiosyncratic volatility as the standard deviation of the residuals from the Fama and French (1993) three-factor model regression, estimated using daily returns over the prior month. We compute size-adjusted institutional ownership using Thomson Financial Institutional Holdings data from April 1980 through December 2016.[14] We follow Nagel (2005) and remove the correlation between firm size and institutional ownership by taking the residual from a cross-sectional logit regression in which the dependent variable is the share of institutional ownership and the independent variable is the square of the logarithm of firm size. Because of the start date of the holdings data, we begin the regressions in Table 6 in April 1980.

Unlike in our other analyses, we convert all explanatory variables in Table 6 into z-scores separately in each cross section. This transformation makes it easier to compare the magnitudes of the interactions with those of the main effects. Of the additional variables, idiosyncratic volatility predicts returns negatively and institutional ownership positively.[15] The interactions between these variables and the same- and other-month returns, reported in Regres-

[13] Fama and French (2018) and Hou et al. (2020), for example, discuss the extent to which return anomalies are stronger among microcaps and small stocks. Stambaugh et al. (2015) suggest that idiosyncratic volatility is a risk that makes arbitrageurs reluctant to trade against mispricings. Nagel (2005) and Hirshleifer et al. (2011) show that anomalies such as the value and accruals effects are stronger in stocks with low institutional ownership, perhaps because short-sale constraints are more likely to bind when the shares are dispersed among individual investors.

[14] The raw Thomson Financial Institutional Holdings data appear to contain systematic errors in the number of shares information. Different firms, for example, report this information in different units. We use a corrected version of these data provided by Jon Lewellen. These corrected data are described in Lewellen and Lewellen (2018).

[15] Idiosyncratic volatility, when converted into a z-score, does not predict returns as negatively in the sample with all stocks as it does when it is not transformed. When we estimate Regression 2 in Table 6 with untransformed variables, idiosyncratic volatility predicts returns with a t-value of −2.69. Other than for this one variable, the transformations do not meaningfully affect the t-values of the average regression slopes.

sions 2 and 4, measure how these additional variables correlate with the strength of return seasonalities.

The associations between seasonalities and all three measures of limits to arbitrage are weak and the signs are inconsistent. We would expect seasonalities to decrease in size and institutional ownership and to increase in idiosyncratic volatility. In terms of the regression specifications, the interaction between size and same-month return should be negative and that between size and other-month return should be positive. These signs would mean that seasonalities and seasonal reversals attenuate moving from small stocks toward big stocks. In Table 6 only firm size's interaction with the other-month return has the predicted sign. While the interaction between firm size and the average same-month return in the all-but-microcaps sample is statistically significant, this interaction is positive. This positive sign suggests that seasonalities are, if anything, more pronounced among large-cap stocks. This effect, which runs counter to the limits-to-arbitrage mechanism, is consistent with the findings of Keloharju et al. (2016). The other two proxies for limits to arbitrage behave similarly. Their interactions with the average same-month return have the wrong sign, those with the average other-month return typically have the correct sign, but most of these interactions are, at best, marginally significant.

In Table A5 in the Appendix, we examine limits to arbitrage using an alternative methodology. We first sort stocks by the limits-to-arbitrage proxy and then, conditional on this sort, by past return seasonality. We sort stocks into microcaps, small stocks, and large stocks by firm size; with the other two variables, we sort firms into quintiles. We define past return seasonality in these sorts as the difference between same- and other-month returns. Table A5 shows some evidence supportive of the limits-to-arbitrage mechanism. The seasonality return spread among microcap stocks, for example, is 101 basis points per month, that among large stocks is 76 basis points, and the 25-basis point difference has a t-value of 2.01. Similarly, moving from the low-idiosyncratic volatility quintile to the highest, the return spread from conditional sorts increases from 73 basis points to 111 basis points. The difference between the extremes has a t-value of 1.56. These estimates, although weak in terms of statistical significance, are consistent with some association between return seasonalities, seasonal reversals, and limits to arbitrage. The return spread associated with return seasonalities does not vary meaningfully in the degree of institutional ownership. In these conditional portfolio sorts, the return spread is 101 basis points per month in the low-IO sample and 112 basis points in the high-IO sample.

Birru (2018) considers the association between sentiment, limits to arbitrage, and seasonalities in daily factor returns. He defines a speculative portfolio as "the subset of stocks predicted to be most strongly affected by investor sentiment (small, young, high volatility, unprofitable, non-dividend paying, lottery-like, distressed, extreme growth, low-priced, lottery-like)" (p. 183) and finds stronger daily seasonalities for those factors in which one of the legs is speculative. If a portfolio's price more easily sways with sentiment when limits to arbitrage bind, then these results suggest that limits to arbitrage could amplify seasonalities (and seasonal reversals). At the same time, Birru's results suggest that this relation is not significant in every specification. In an analysis of the interactions between institutional ownership, liquidity, and seasonalities, Birru finds that neither the level of institutional ownership nor liquidity significantly alters the Monday versus Friday return differential. The evidence on the association between seasonalities and limits to arbitrage can be mixed because clear proxies that correlate only with limits to arbitrage and nothing else are lacking.

4.3. Measuring the speed of seasonal reversals

In the Section 2 model, we assume that seasonalities add up to zero over the calendar year. This assumption implies that seasonal reversals are evenly distributed over a year, that is a high, seasonal return in one month is offset, on average, by lower returns over the other 11 months. Thus far, we have not examined the temporal distribution of seasonal reversals. Do seasonalities reverse quickly or with some delay?

In Fig. A1 in the Appendix, we show that decomposing the other-month component into parts is informative about the speed of seasonal reversals. If seasonalities reverse quickly (a high seasonal return in April could, for example, typically reverse in full already in May), then the other-month returns nearest to the current month should predict returns the strongest with a negative sign. If seasonalities sometimes last for longer than a month (a stock's return could, for example, be seasonally high from March through May), then the other-month returns farthest away from the current month should predict returns stronger than those nearest to it.

In Fig. 5, we study the speed of reversals by running Fama-MacBeth regressions in which we decompose the other-month return into 11 monthly return components. Our analysis follows the same structure as that in Fig. A1. For example, in a regression that predicts March returns, the 11 explanatory variables are the average January, February, April,... returns. We order these return variables relative to the month of the dependent variable. In a March regression, month +1 is the average historical April return; month +2 is the average historical May return; and so forth and "Same" is the historical average March return. Our analysis also controls for (but does not separately report estimates for) log size, log book-to-market, short-term reversals, momentum, and prior five-year return skipping a year.

In Panel A of Fig. 5, we show the t-values associated with the average coefficients for the full sample. Panel B limits the sample to all-but-microcaps. In both samples, the average returns associated with the months nearest to the dependent variable, months +1 and +11, predict returns with a positive sign. The months more removed from the dependent variable, e.g., month +6 and those around it, are more significant negative predictors of returns.[16]

[16] The pattern of the month +1 through +11 coefficients in Panel A is almost perfectly symmetric. This symmetry is consistent with the simu-

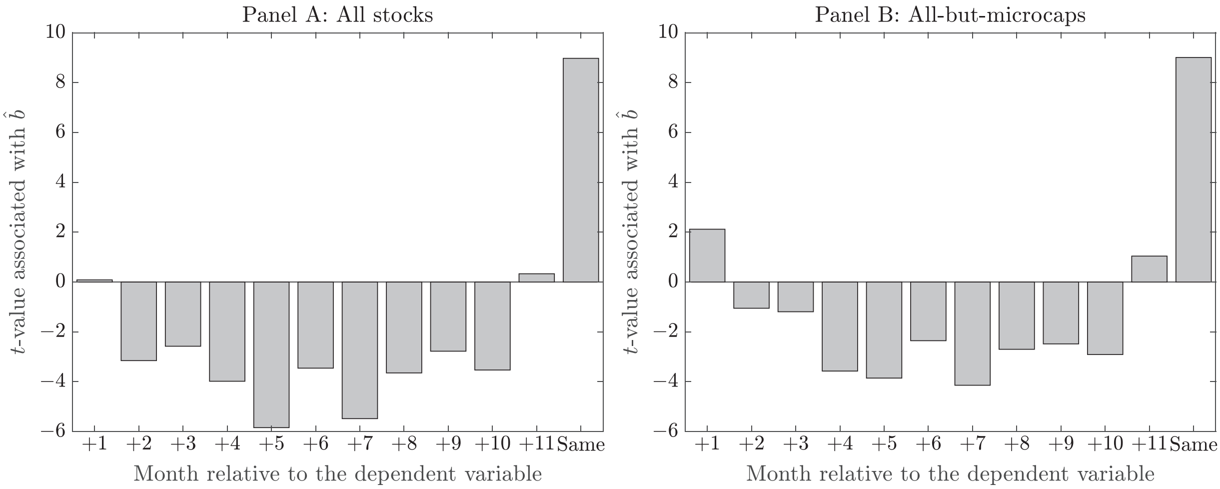


Fig. 5. The timing of seasonal reversals. This figure plots t-values associated with the average slope coefficients from Fama-MacBeth regressions that predict the cross section of monthly returns. The independent variables are log size, log book-to-market, prior one-month return, prior one-year return skipping a month, prior five-year return skipping a year, and average historical return in each month: average January return, average February return, and so forth. We order the 11 right-hand-side average-return variables relative to the month of the dependent variable. For example, in a regression that predicts March returns, month +1 is the average historical April return; month +2 is the average historical May return; and "Same" is the historical average March return. We do not report t-values associated with the coefficients of the control variables. We estimate the regressions using data for all stocks (Panel A) and for all-but-microcaps (Panel B).

The patterns in Fig. 5 suggest that seasonalities and seasonal reversals do not follow a precise monthly schedule. A stock that earns a high seasonal return in, say, February does not typically earn the offsetting low returns in the surrounding months of January and March. Whatever the mechanism that causes the seasonal dislocation of prices thus bleeds over to the surrounding months.

5. Seasonality, seasonal reversal, and long-term reversal factors

5.1. Average monthly factor returns and correlations

The coefficients of the Fama-MacBeth regressions in Table 1 suggest that average same- and other-month returns are informative about future returns. We measure the usefulness of these signals from the investment perspective by constructing high-minus-low (HML)-like factors that select stocks by their average past returns. We also construct a seasonality factor (ANN) by sorting stocks into six portfolios by size and the average same-month return using monthly rebalancing. The return on this factor is the difference between the value-weighted returns on the two high-average portfolios and the two low-average portfolios. We construct a seasonal reversals factor (NANN) using the same methodology, except that we sort stocks by their average other-month returns. We skip over the most recent year, as before, when computing the average other-month return. Because this is a reversal factor, we compute the return on the factor as the difference between the two low-average and the two high-average portfolios. Finally, we construct an annual minus non-annual factor by sorting stocks by the difference between the average same- and other-month returns.

Table 7 reports the monthly percentage returns for these factors and their correlations. We also report, for comparison, the same statistics and correlations for the market, size, value, momentum, and long-term reversals factors. The long-term reversals factor is another HML-like factor that is used to select stocks by their five-year-skip-a-year return (Fama and French, 1996).

The average returns on the seasonality and seasonal reversal factors are economically large and statistically significant. The seasonality factor earns an average return of 61 basis points per month (t-value = 8.37); the seasonal reversal factor earns 45 basis points (t-value = 4.89); and the combination of the two (the annual minus non-annual factor) earns 67 basis points (t-value = 9.93).[17]

lations reported in Fig. A1, and it is the result of the returns-on-returns nature of these regressions. To see why the pattern should be close to symmetric, consider a time series regression that predicts month t returns with month $t-k$ returns. If we were to find that month $t-k$ return positively predicts month t returns, then we would also expect to find that month t return positively predicts month $t+k$ returns. Covariances are symmetric. The pattern in Panel A is not perfectly symmetric because the sample is finite. Returns early in the sample are used only to predict future returns and returns late in the sample are used only as dependent variables. The pattern in Panel B is not symmetric because of the sample restriction: the regression includes only all-but-microcaps. A firm that today belongs to the all-but-microcaps sample perhaps did not belong to the all-but-microcaps sample in the past. By definition, the returns firms earned when they were below the all-but-microcaps threshold never enter the left-hand side of the regressions, thereby breaking the symmetry.

[17] The estimates in Fig. 5, which show that seasonalities can last for longer than a month, suggest that we could improve the seasonal reversals factor by skipping the surrounding months when computing the average non-annual return. For example, when predicting the cross section of June returns, we could define the non-annual return as the average of August to April returns, thereby cutting out the three-month segment around June. In an unreported analysis, we find that this version of

Table 7
Monthly percentage returns on factors and correlations.
This table reports the average monthly percentage returns, standard deviations, and t-values for various factors (Panel A) and the monthly return correlations (Panel B). The first four factors are those of the Carhart (1997) four-factor model: market, size, value, and momentum. The other factors are high-minus-low (HML)–like factors that first sort stocks into six portfolios by market capitalization and the sorting variable. The long-term reversal factor (LTREV) sorts stocks by their five-year return skipping a year; the seasonality factor (ANN) sorts stocks by their average same-month return; the seasonal reversal factor (NANN) sorts stocks by their average other-month return; and the annual minus non-annual factor (AMN) sorts stocks by the difference between the average same-month and other-month return. The average other-month return is computed after skipping over the most recent year. The data are demeaned in each cross section before computing the average same- and other-month returns. Both averages use up to 20 years of historical data. The factor return data are from January 1963 through December 2016.

Panel A: Monthly percentage returns

Factor	Name	Average return	Standard deviation	t-value
MKTRF	Market	0.52	4.41	3.00
SMB	Size	0.23	3.08	1.86
HML	Value	0.38	2.82	3.46
UMD	Momentum	0.66	4.21	4.02
LTREV	Long-term reversal	0.29	2.49	2.95
ANN	Seasonality	0.61	1.85	8.37
NANN	Seasonal reversal	0.45	2.33	4.89
AMN	Annual − non-annual	0.67	1.71	9.93

Panel B: Monthly return correlations

Factor	MKTRF	SMB	HML	UMD	LTREV	ANN	NANN	AMN
MKTRF	1							
SMB	0.29	1						
HML	−0.26	−0.21	1					
UMD	−0.13	0.00	−0.19	1				
LTREV	−0.02	0.26	0.45	−0.07	1			
ANN	0.18	0.03	−0.23	−0.05	−0.13	1		
NANN	−0.51	−0.25	0.72	0.00	0.45	−0.15	1	
AMN	−0.02	−0.07	0.06	−0.03	0.03	0.89	0.20	1

The difference in the predictive powers of the seasonality and annual minus non-annual factors is also large. The moderate difference in the levels of t-values (8.37 versus 9.93) is not the right comparison because these estimates are so far out in the tails of the distributions. In terms of likelihoods, a move from a t-value of 8.37 to 9.93 is equivalent to a move from a t-value of 1.96 to 5.56; that is, the difference between an estimate being statistically significant at the 5% level versus there being overwhelming evidence against the null hypothesis.[18]

The seasonality and seasonal reversal factors correlate differently with the market, value, and long-term reversals factors. The seasonality factor correlates positively with the market and negatively with both the value and the long-term reversal factors. These correlations have the opposite signs for the seasonal reversal factor. The seasonal reversal factor's correlation with the market is −0.51; with the value factor, 0.72; and with the long-term reversal factor, 0.45. As a consequence of these offsetting correlations, the annual minus non-annual factor is nearly uncorrelated with these other factors, with correlations ranging from −0.07 to 0.06.

a non-annual factor earns an average return of 53 basis points per month (t-value = 5.96).

[18] This comparison is based on the following computation. The t-values of 8.37 and 9.93 correspond to p-values of 5.762×10^{-17} and 3.083×10^{-23}. The latter event is therefore less probable than the first event by a factor of 1.9 million. If we start from a p-value of 0.05 (t-value = 1.96), an event that is proportionally as improbable has a p-value of 0.000000027 or a t-value of 5.56.

5.2. Incremental information

In Table 8, we examine the incremental information of the seasonality, seasonal reversal, and long-term reversal factors. In the first column of Panel A, for example, we regress the monthly returns on the seasonality factor on the returns on the market, size, and value factors. These spanning regressions assess the extent to which the left-hand-side factor (here, the seasonality factor) contains information that is not present in the set of the right-hand-side factors (here, the market, size, and value factors).

The alphas from these spanning regressions have two interpretations. The first interpretation pertains to the investment problem. A statistically significant alpha indicates that an investor, who currently trades the right-hand-side factors, can increase his portfolio's Sharpe ratio significantly by also trading the left-hand-side factor. The second interpretation relates to comparing different asset pricing models. A statistically significant alpha indicates that an asset pricing model that adds the left-hand-side factor to the set of right-hand-side factors is statistically superior to a model that contains only the right-hand-side factors (Barillas and Shanken, 2017).

In the regression for Column 1, which explains seasonalities with the three-factor model, the alpha is 64 basis points per month with a t-value of 8.79. As suggested by the Table 1 Fama-MacBeth regressions, the seasonality factor thus contains a substantial amount of information about expected returns. In the regressions for Columns 2

Table 8
Incremental information of the seasonality, seasonal reversal, and long-term reversal factors.
This table reports estimates from spanning regressions. In Panel A, the left-hand side variable is the monthly return on the seasonality factor or the seasonal reversal factor; in Panel B, the return on the annual minus non-annual factor or the long-term reversal factor. The explanatory variables are the four factors of the Carhart (1997) four-factor model (market, size, value, and momentum) and, depending on the specification, the seasonality or the seasonal reversal factor. t-values of the alpha are reported in parentheses. The factor return data are from January 1963 to December 2016.

Panel A: Seasonality factor and seasonal factor

	Dependent variable					
Explanatory	Seasonality factor			Seasonal reversal factor		
variable	(1)	(2)	(3)	(4)	(5)	(6)
Monthly percentage alpha						
$\hat{\alpha}$	0.64	0.67	0.62	0.35	0.30	0.25
	(8.79)	(8.10)	(7.32)	(6.17)	(5.33)	(3.91)
Factor loadings						
MKTRF	0.06	0.05	0.08	−0.18	−0.17	−0.18
SMB	−0.03	−0.03	−0.01	−0.02	−0.08	−0.08
HML	−0.14	−0.14	−0.21	0.52	0.43	0.44
UMD		−0.03	−0.04		0.04	0.04
LTREV		−0.02	−0.05		0.23	0.23
ANN						0.09
NANN			0.15			
Factor loadings, t-values						
MKTRF	2.34	2.08	2.94	−11.13	−11.61	−11.94
SMB	−0.99	−0.86	−0.44	−0.81	−4.10	−4.02
HML	−3.35	−3.08	−3.99	18.23	14.87	15.15
UMD		−0.90	−1.06		2.27	2.43
LTREV		−0.37	−1.18		7.94	8.07
ANN						2.62
NANN			2.51			

Panel B: Annual minus non-annual factor and long-term reversal factor

	Dependent variable			
Explanatory	Annual minus non-annual factor		Long-term reversal factor	
variable	(1)	(2)	(3)	(4)
Monthly percentage alpha				
$\hat{\alpha}$	0.66	0.67	0.04	−0.08
	(9.70)	(8.53)	(0.50)	(−0.96)
Factor loadings				
MKTRF	0.01	0.00	0.01	0.10
SMB	−0.04	−0.05	0.29	0.30
HML	0.03	0.01	0.46	0.20
UMD		−0.01		0.00
LTREV		0.03		
ANN				−0.06
NANN				0.49
Factor loadings, t-values				
MKTRF	0.29	0.21	0.22	3.54
SMB	−1.27	−1.58	7.18	8.10
HML	0.72	0.24	10.39	3.88
UMD		−0.21		−0.14
LTREV		0.76		
ANN				−1.19
NANN				7.74

and 3, we add to the right-hand-side the long-term reversal and seasonal reversal factors. The addition of these factors does not materially lower the alpha. In the Column 3 model with both of these additional factors, the alpha is 62 basis points per month with a t-value of 7.32.

Columns 2–6 of Panel A show that the seasonal reversal factor also contains information that is not present in the Carhart (1997) four-factor model, the long-term reversal factor, and the seasonality factor. The seasonal reversal factor's monthly three-factor model alpha is 35 basis points (t-value = 6.17). This alpha falls only slightly to 30 basis points (t-value = 5.33) when we also control for long-term reversals. Thus, long-term reversals are largely unrelated to seasonal reversals, and the seasonal reversal factor contains information that is not present in the long-term reversal factor. The regression for Column 6 includes the seasonality factor on the right-hand side. The alpha is now 25 basis points per month with a t-value of 3.91. This estimate suggests that seasonalities and seasonal reversals contain independent information about the cross section of returns. Both variables appear to provide noisy (but independent) estimates of the seasonal component of expected returns.

Panel B of Table 8 reports the coefficients from spanning regressions with either the annual minus non-annual factor or the long-term reversal factor as the dependent variable. The annual minus non-annual factor, which estimates expected returns from both the average same- and other-month returns, earns a monthly three-factor model alpha of 66 basis points (t-value = 9.70). Because this factor does not correlate substantially with momentum or long-term reversals, its alpha in a full model including these factors remains at 67 basis points (t-value = 8.53).

The long-term reversal factor does not contain information beyond that contained in the three-factor model. This factor's three-factor model alpha is 4 basis points (t-value = 0.50). The size and value factors are the culprits for this loss of significance. Long-term losers are, on average, small value stocks, so the long-term reversal factor loads positively on both the size and the value factors (Fama and French, 1996). The insignificant alpha indicates that an investor who already trades the market, size, and value factors would not benefit from also trading the long-term reversal factor. The finding that the seasonal reversal factor earns a statistically significant alpha in a regression that includes, among other things, size, value, and the long-term reversal factors is therefore important. The significance of this alpha indicates that the seasonal reversal factor is not just a different measure of long-term reversals, but a distinct effect.

Panel A of Table 8 shows that the correlation between the seasonality and seasonal reversal factors becomes positive once we control for the differences in their market betas. Column 3, for example, shows that the seasonality factor's slope against the seasonal reversal factor is 0.15 (t-value = 2.51). In Appendix Section A.7, we discuss simulations that assess the extent to which this correlation is consistent with both factors being based on noisy signals of the seasonalities in expected returns, $E_t(\tilde{r}_{i,t+1})$. We show that in the simulated data, just like in the actual data, the correlation between two highly statistically significant factors can be low even though the factors are noisy versions of the same signal.

5.3. Maximum Sharpe ratios

The spanning regressions of Table 8 suggest that investors would have earned higher Sharpe ratios by trading the seasonalities in expected returns in addition to the other factors. We next quantify these results by constructing ex post maximum Sharpe ratio portfolios from various combinations of factors listed in Table 7. We use factors' monthly returns from January 1963 to December 2016 to find the tangency portfolio for different combinations of factors and then compute and report the Sharpe ratio associated with this portfolio in Table 9.

An investor would have earned an annualized Sharpe ratio of 0.41 by investing in the market portfolio (no. 1) between January 1963 and December 2016. In Portfolio 2, the size, value, and momentum factors are very valuable. The optimal combination of these factors would have earned a Sharpe ratio of 1.08, and the investor would have obtained this Sharpe ratio by investing two-thirds of his portfolio in the value and momentum factors.

In Portfolios 3–7, we add to this set of factors different combinations of the long-term reversal, seasonality, and seasonal reversal factors. Consistent with its insignificant alpha in the regressions for Table 8, the addition of the long-term reversal factor does not change the maximum Sharpe ratio. Both the seasonality and seasonal reversal factors, however, substantially increase the Sharpe ratio. The seasonal reversal factor alone increases it from 1.08 to 1.34; the seasonality factor alone increases it to 1.69; and the addition of both of them increases it to 1.81. This last increase is consistent with both factors containing information about expected returns that is not present in the other. Portfolio 6, with both the seasonality and seasonal reversal factors, shows that the optimal portfolio is now two-thirds invested in these two factors alone. The remaining one-third is approximately evenly divided between the market, size, value, and momentum factors. In Portfolio 7, the maximum Sharpe ratio is nearly as high, 1.73, when we replace the two factors with the annual minus non-annual factor.

The shifts in portfolio weights in Table 9 that result from adding the seasonal factors are substantial. The shifts indicate that the seasonal effects (seasonalities and seasonal reversals) are very large in magnitude relative to the non-seasonal return predictors such as value and momentum.

5.4. Persistence

Seasonalities and seasonal reversals are persistent throughout an extended sample period from 1946 through 2016.[19] Fig. 6 illustrates this persistence by reporting the t-values associated with the average factor returns for ten-year rolling windows. The first data points in the figure, for example, correspond to a ten-year window from January 1946 to December 1956. The seasonality, seasonal reversal, and annual minus non-annual factors earn average returns of 74 basis points (t-value = 5.71), 24 basis points (t-value = 1.79), and 75 basis points (t-value = 6.09) over this ten-year period, and it is these t-values we report in Fig. 6. The period to the left from the vertical line in Fig. 6 is out-of-sample relative to the other results in this paper.

We find that none of the factors averages a negative return over any of the ten-year periods. The t-value for

[19] The factor data begin in 1946 because the portfolios are formed using returns over the prior 20-year period, and the CRSP data begin in January 1926.

Table 9
Maximum Sharpe ratios.
This table reports weights and ex-post maximum Sharpe ratios for strategies formed from market, value, size, momentum, seasonality, seasonal reversal, and long-term reversal factors. The seasonality, seasonal reversal, and long-term reversal factors are high-minus-low (HML)–like factors that first sort stocks into six portfolios by market capitalization and the sorting variable. The long-term reversal factor sorts stocks by their five-year return skipping a year; the seasonality factor sorts stocks by their average same-month return; the seasonal reversal factor sorts stocks by their average other-month return; and the annual minus non-annual factor sorts stocks by the difference between the average same-month and other-month return. The data are demeaned in each cross section before computing the average same- and other-month returns. Both averages use up to 20 years of historical data. The factor return data are from January 1963 through December 2016.

	Portfolio weight (percent)								
						Seasonality factor			Sharpe
#	MKTRF	SMB	HML	UMD	LTREV	ANN	NANN	AMN	ratio
1	100								0.41
2	21	11	42	27					1.07
3	21	10	40	26	3				1.08
4	7	6	25	14	2	47			1.69
5	24	11	3	16	−11		57		1.34
6	11	7	10	11	−4	38	28		1.81
7	10	7	18	13	0			53	1.73

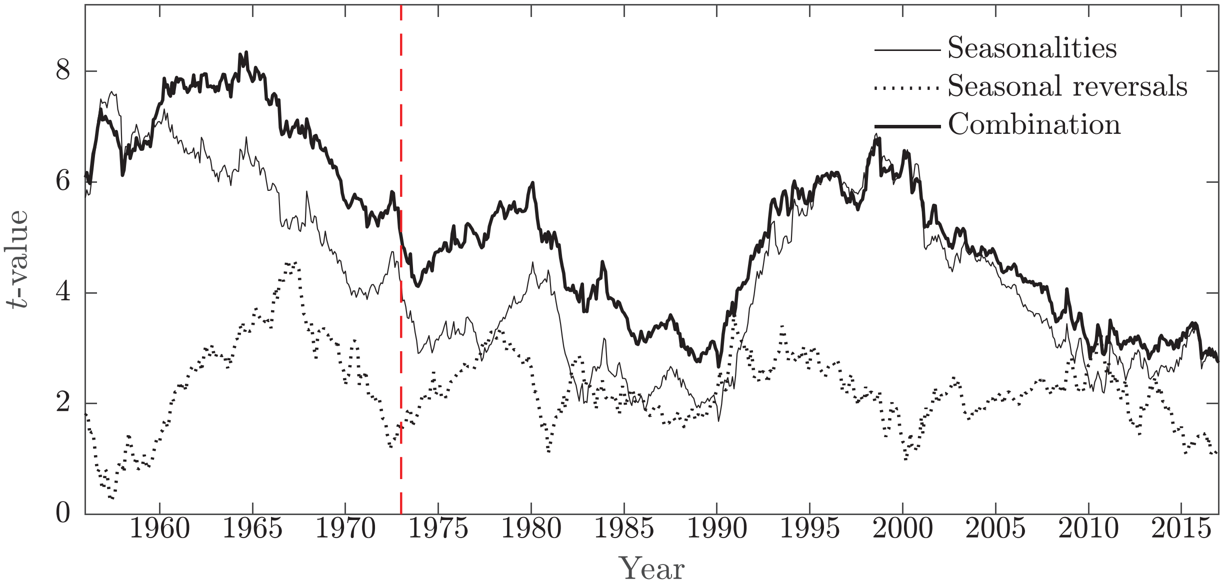


Fig. 6. Seasonality, seasonal reversal, and their combination factors, 1946–2016. This figure plots the *t*-values associated with the average premiums on three factors. The seasonality factor sorts stocks by their average same-month return; the seasonal reversal factor sorts stocks by their average other-month return; and the annual minus non-annual factor, labeled "Combination," sorts stocks by the difference between the average same-month and other-month return. We compute average factor returns using rolling ten-year windows and report the *t*-values for these rolling windows. The 1960 data points, for example, are the *t*-values of the factors for the ten-year period up to 1960. The period before the dashed vertical line at 1973 is an out-of-sample test period.

the combination factor, which sorts stocks into portfolios by the difference between the average same- and other-month returns, is above 2.5 for every ten-year period between January 1946 and December 2016. The existence of significant return seasonalities also in the early part of the sample is perhaps not surprising. In an early attempt at constructing a stock price index, Cover et al. (1930) deem it preferable to adjust stock prices for such seasonalities (Hartzmark and Solomon, 2018).[20]

The persistence of these factor premiums is important. Heston and Sadka (2008) show the seasonality effect in stock returns using data up December 2002. The data for the last few years of Fig. 6 are therefore entirely out of sample for the seasonality factor. This out-of-sample persistence of the effect alone suggests that return seasonalities and seasonal reversals are unlikely to be an artifact of data mining (McLean and Pontiff, 2016; Linnainmaa and Roberts, 2018). At the same time, this degree of persistence is puzzling, casting doubt on the risk-based explanations for the cross-sectional variation in expected returns.

[20] Cover et al. (1930, pp. 181–182) examine cross-sectional variation in the seasonal patterns in the returns of 12 companies. They classify the companies by the months in which they typically earned their highest and lowest returns, as well as by the amount of seasonal variation in returns: "The series with least month-to-month variation were Elgin National Watch and Pullman, Inc. Those with the greatest seasonal amplitude were Standard Oil of Indiana, Corn Products, and Stewart-Warner."

If the seasonality and seasonal reversal factors earn premiums to compensate for some combination of risks, these risks appear not to have materialized over the past 50 years. An alternative explanation, which would reconcile both the persistence of the seasonalities and seasonal reversals, and the fact that seasonal reversals seem to balance out seasonalities, is that seasonalities are due to temporary mispricing, not risk-based mechanisms. Seasonalities could be induced by investors consistently trading in the same direction in the same periods. These effects can persist in the data if the resulting seasonalities are not large enough to be exploited as stand-alone anomalies due to their high turnover. At the same time, even if the associated round-trip transaction costs are prohibitively large, investors could benefit from these seasonalities by using them to time their trades (Heston et al., 2010; Novy-Marx and Velikov, 2016).

6. Conclusion

What explains return seasonalities? In this paper, we consider two potential explanations for them: risk and mispricing. Consistent with the mispricing explanation, we show that seasonalities are balanced out by seasonal reversals. A stock's high seasonal return in one month is offset by low seasonal returns in the other months. We cannot reject the null hypothesis that the seasonalities in US equities reverse perfectly. These results speak against risk-based explanations, which make no predictions on seasonal reversals. If an asset earns an above-average return in one month, no economic reason exists that it should earn a below-average return in the other months.

If seasonalities are due to mispricing, we would expect to find seasonal reversals wherever seasonalities are found. Consistent with this prediction, we find evidence of seasonal reversals not only in monthly US equity returns but also in daily stock returns, international stock returns, country equity indices, and commodity returns. We find that return seasonalities across asset classes and frequencies are always offset or moderated by seasonal reversals.

Our insights on seasonal reversals improve the predictive power of seasonal trading strategies. We find that both same- and other-month returns contain independent information about future expected returns. Neither seasonalities nor seasonal reversals subsume each other, which is consistent with them containing independent information about expected returns.

Our results offer insight into the nature of return seasonalities and new empirical predictions. If seasonalities are due to mispricing, they could be driven by investors consistently trading in the same direction and the rest of the market not fully absorbing these trades. Chan et al. (2004), Heston et al. (2010), Bogousslavsky (2016), and Murphy and Thirumalai (2017) provide evidence consistent with this conjecture. Testing whether this conjecture holds for periods longer than daily trades and asset classes other than stocks is left for future work.

References

Ang, A., Hodrick, R., Xing, Y., Zhang, X., 2006. The cross section of volatility and expected returns. J. Financ. 61 (1), 259–299.

Asness, C., Moskowitz, T.J., Pedersen, L.H., 2013. Value and momentum everywhere. J. Financ. 68 (3), 929–985.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. J. Financ. 61 (4), 1645–1680.

Barber, B.M., Odean, T., Zhu, N., 2009a. Do retail trades move markets? Rev. Financ. Stud. 22 (1), 151–186.

Barber, B.M., Odean, T., Zhu, N., 2009b. Systematic noise. J. Financ. Mark. 12 (4), 547–569.

Barillas, F., Shanken, J., 2017. Which alpha? Rev. Financ. Stud. 30 (4), 1316–1338.

Berk, J.B., Green, R.C., Naik, V., 1999. Optimal investment, growth options, and security returns. J. Financ. 54 (5), 1553–1607.

Birru, J., 2018. Day of the week and the cross-section of returns. J. Financ. Econ. 130 (1), 182–214.

Bogousslavsky, V., 2016. Infrequent rebalancing, return autocorrelation, and seasonality. J. Financ. 71 (6), 2967–3006.

Carhart, M.M., 1997. On persistence in mutual fund performance. J. Financ. 52 (1), 57–82.

Chan, S.H., Leung, W.-K., Wang, K., 2004. The impact of institutional investors on the Monday seasonal. J. Bus. 77 (4), 967–986.

Cohen, R. B., Polk, C., 1996. An investigation of the impact of industry factors in asset-pricing tests. Unpublished working paper, London School of Economics and Political Science, London, UK.

Conrad, J., Kaul, G., 1998. An anatomy of trading strategies. Rev. Financ. Stud. 11, 489–519.

Cover, J.H., Revzan, D.A., Helms, W.M., Cohenour, V.J., 1930. A barometer of Chicago stock prices. J. Bus. Univ. Chicago 3 (2), 171–191.

Davis, J.L., Fama, E.F., French, K.R., 2000. Characteristics, covariances, and average returns: 1929 to 1997. J. Financ. 55 (1), 389–406.

De Bondt, W.F.M., Thaler, R.H., 1985. Does the stock market overreact? J. Financ. 40 (3), 379–395.

Dorn, D., Huberman, G., Sengmueller, P., 2008. Correlated trading and returns. J. Financ. 63 (2), 858–920.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns of stocks and bonds. J. Financ. Econ. 33 (1), 3–56.

Fama, E.F., French, K.R., 1996. Multifactor explanations of asset pricing anomalies. J. Financ. 51 (1), 55–84.

Fama, E.F., French, K.R., 2018. Choosing factors. J. Financ. Econ. 128 (2), 234–252.

Fama, E.F., MacBeth, J.D., 1973. Risk, return and equilibrium: Empirical tests. J. Polit. Econ. 71 (3), 607–636.

Feng, L., Seasholes, M., 2004. Correlated trading and location. J. Financ. 59, 2117–2144.

French, K.R., 1980. Stock returns and the weekend effect. J. Financ. Econ. 8 (1), 55–69.

Grinblatt, M., Keloharju, M., Linnainmaa, J., 2012. IQ, trading behavior, and performance. J. Financ. Econ. 104 (2), 339–362.

Hartzmark, S. M., Solomon, D. H., 2018. Reconsidering returns. Unpublished working paper, University of Chicago.

Heston, S.L., Korajczyk, R.A., Sadka, R., 2010. Intraday patterns in the cross-section of stock returns. J. Financ. 65 (4), 1369–1407.

Heston, S.L., Sadka, R., 2008. Seasonality in the cross-section of stock returns. J. Financ. Econ. 87 (2), 418–445.

Heston, S.L., Sadka, R., 2010. Seasonality in the cross-section of stock returns: The international evidence. J. Financ. Quant. Anal. 45 (5), 1133–1160.

Hirshleifer, D., Jiang, D., Meng, Y., 2020. Mood betas and seasonalities in stock returns. J. Financ. Econ. 137 (1), 272–295.

Hirshleifer, D., Teoh, S.H., Yu, J.J., 2011. Short arbitrage, return asymmetry, and the accrual anomaly. Rev. Financ. Stud. 24 (7), 2429–2461.

Hou, K., Xue, C., Zhang, L., 2020. Replicating anomalies. Rev. Financ. Stud. 33 (5), 2019–2133.

Kamstra, M.J., Kramer, L.A., Levi, M.D., 2003. Winter blues: A SAD stock market cycle. Am. Econ. Rev. 93 (1), 324–343.

Keloharju, M., Linnainmaa, J.T., Nyberg, P., 2016. Return seasonalities. J. Financ. 71 (4), 1557–1590.

Kumar, A., Lee, C., 2006. Retail investor sentiment and return comovements. J. Financ. 61 (5), 2451–2486.

Lakonishok, J., Maberly, E., 1990. The weekend effect: Trading patterns of individual and institutional investors. J. Financ. 45 (1), 231–243.

Lakonishok, J., Shleifer, A., Vishny, R.W., 1992. The impact of institutional trading on stock prices. J. Financ. Econ. 32, 23–43.

Lewellen, J., Lewellen, K., 2018. Institutional investors and corporate governance: The incentive to be engaged. Unpublished working paper, Dartmouth College.

Linnainmaa, J.T., Roberts, M.R., 2018. The history of the cross section of stock returns. Rev. Financ. Stud. 31 (7), 2606–2649.

Lo, A.W., MacKinlay, A.C., 1990. When are contrarian profits due to stock market overreaction? Rev. Financ. Stud. 3 (2), 175–205.

McLean, R.D., Pontiff, J., 2016. Does academic research destroy stock return predictability? J. Financ. 71 (1), 5–32.
Moskowitz, T.J., Grinblatt, M., 1999. Do industries explain momentum? J. Financ. 54 (4), 1249–1290.
Murphy, D., Thirumalai, R., 2017. Short-term predictability and repetitive institutional net order activity. J. Financ. Research 40 (4), 455–477.
Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. J. Financ. Econ. 78 (2), 277–309.
Novy-Marx, R., 2013. The other side of value: The gross profitability premium. J. Financ. Econ. 108 (1), 1–28.
Novy-Marx, R., Velikov, M., 2016. A taxonomy of anomalies and their trading costs. Rev. Financ. Stud. 29 (1), 104–147.
Ritter, J.R., 1988. The buying and selling behavior of individual investors at the turn of the year. J. Financ. 43 (3), 701–717.
Shumway, T., 1997. The delisting bias in CRSP data. J. Financ. 52 (1), 327–340.
Stambaugh, R.F., Yu, J., Yuan, Y., 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. J. Financ. 70 (5), 1903–1948.
Wermers, R., 1999. Mutual fund herding and the impact on stock prices. J. Financ. 54 (2), 581–622.

Journal of Financial Economics 137 (2020) 272–295

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Mood beta and seasonalities in stock returns[☆]

David Hirshleifer[a], Danling Jiang[b,*], Yuting Meng DiGiovanni[c]

[a] The Paul Merage School of Business, University of California at Irvine and NBER, United States
[b] College of Business, Stony Brook University, United States; School of Economics and Management, Southwest Jiaotong University, China. Mailing address: 306 Harriman Hall, Stony Brook, NY 11794 United States
[c] Muma College of Business, University of South Florida, United States

A R T I C L E   I N F O

Article history:
Received 20 March 2018
Revised 19 June 2019
Accepted 7 August 2019
Available online 7 February 2020

JEL classification:
G11
G12
G14
G41

Keywords:
Return seasonality
Investor mood
Mood beta
Market efficiency
Anomalies

A B S T R A C T

Existing research has found cross-sectional seasonality of stock returns—the periodic outperformance of certain stocks during the same calendar months or weekdays. We hypothesize that assets' different sensitivities to investor mood explain these effects and imply other seasonalities. Consistent with our hypotheses, relative performance across individual stocks or portfolios during past high or low mood months and weekdays tends to recur in periods with congruent mood and reverse in periods with noncongruent mood. Furthermore, assets with higher sensitivities to aggregate mood—higher mood betas—subsequently earn higher returns during ascending mood periods and earn lower returns during descending mood periods.

Published by Elsevier B.V.

1. Introduction

Extensive research has found several aggregate market return seasonalities—periodic variation in the mean returns of market index portfolios.[1] Recent studies have also identified seasonality in the cross-section of security returns—the periodic outperformance of certain securities relative to others in the same calendar month (Heston and Sadka, 2008, 2010), on the same day of the week (Keloharju et al., 2016), during certain weekdays (Birru, 2018), or during the pre-holiday period (Hirshleifer et al., 2016).[2]

We propose an integrated explanation for these effects based upon investor mood. This explanation covers seasonalities at both the aggregate and cross-sectional levels, and at both monthly and daily return frequencies. We term this explanation the *mood seasonality hypothesis*. Based on this, we test and show an extensive set of new empirical implications for return seasonalities. Consistent with the mood seasonality hypothesis, we find that historical seasonal

☆ We thank Alexander Barinov, Ling Cen, Kent Daniel, Iftekhar Hasan, Juhani Linnainmaa, Antti Llmanen, Lin Peng, Sang Byung Seo, Alain Wouassom, Chengxi Yin, and Bohui Zhang, and seminar participants at Auburn University, Chongqing University, CUNY-Baruch, Florida State University, Fordham University, McMaster University, Southwest Jiaotong University, Stony Brook University, University of South Florida, South University of Missouri at Columbia, the 2017 ASSA-AFE, the 2017 CICF, and the 2018 MFA. Danling Jiang acknowledges financial support from the National Natural Science Foundation of China (NSFC) (Grant #91746109). All errors remain ours.

* Corresponding author.
E-mail address: danling.jiang@stonybrook.edu (D. Jiang).

[1] See, e.g., Keim (1983), Lakonishok and Smidt (1988), and Kamstra et al. (2003).

[2] See Hartzmark and Solomon (2018) for a review of a general line of literature on market impounding of the information in recurring firm events.

https://doi.org/10.1016/j.jfineco.2020.02.003
0304-405X/Published by Elsevier B.V.

return of a security relative to other securities is a positive or negative forecaster of the security's relative future seasonal returns, with direction that depends on whether the investor moods during the historical and forecasting periods are congruent or noncongruent. Also motivated by the mood seasonality hypothesis, we propose a new measure of a stock's return sensitivity to mood variations, which we call mood beta, and show that mood beta is a strong predictor of seasonal asset returns in future periods in which mood is expected to rise or fall.

Mood here refers to emotion-induced variation in investor preferences, beliefs, or risk tolerance. While mood affects economic decision-making through a variety of pathways (Lerner et al., 2015), our focus is on the effects of affective valence—whether the mood is positive or negative (and in changes, improving or deteriorating). Mood can be viewed as a special case of investor sentiment (e.g., Baker and Wurgler, 2006, 2007), which is a general term that includes shifts in beliefs or preferences that can derive from either affective sources (i.e., feelings) or from non-affective sources such as shifts in investor attention (e.g., Hirshleifer and Teoh, 2003; Peng and Xiong, 2006), shifts in confidence (e.g., Daniel et al., 1998, 2001), or extrapolative expectation (e.g., Barberis et al., 2015, 2018).

The key premise of our tests is that there are predictable seasonal variations in investor mood (as motivated by our review of related literature in Section 2). We hypothesize that such seasonal mood shifts cause periodic investor optimism or pessimism in evaluating common factors in returns, which results in seasonal variations in factor mispricing.[3] Accordingly, factor mispricing leads to return seasonality and predictability in the cross-section owing to cross-sectional dispersion in factor loadings. During periods in which mood is improving, both the aggregate market and assets with higher sensitivities to ascending mood earn higher average returns. The reverse holds when mood is deteriorating. High mood sensitivity of an asset results from high loadings on mood-mispriced factors.

We suggest that an asset's sensitivity to seasonal mood shifts can be captured by its average historical returns during different seasonal periods when mood is ascending or descending, or by its return sensitivity to aggregate returns during those periods—the mood beta. We hypothesize that these measures of mood sensitivity predict asset returns in future periods when ascending or descending mood is anticipated.

We test these ideas by selecting calendar months and weekdays that are associated with ascending or descending moods based on various experimental, survey, and empirical research. We hereafter refer to these calendar months or weekdays more briefly as hypothesized high or low mood periods. Specifically, we hypothesize January, March, and Friday to be high mood periods, as these are associated with the highest equal-weighted market returns among all calendar months or weekdays during our sample period 1963–2016. Similarly, we hypothesize September, October, and Monday to be low mood periods, as these are associated with the lowest full sample equal-weighted market returns for the given frequency.

These average return patterns also confirm the seasonal psychology suggested by previous literature. Early January is associated with the uplifted mood of the New Year period (Thaler, 1987; Bergsma and Jiang, 2016), March is associated with the highest recovery from seasonal affective disorder (SAD) (Kamstra et al., 2017), and Friday induces an upbeat mood in anticipation of the weekend break (e.g., Birru,2018).[4] In contrast, September to October are associated with the highest onset of the SAD effect (Kamstra et al., 2017), and Monday induces downbeat mood at the start of the week (e.g., Birru, 2018).

The mood seasonality hypothesis first predicts that relative asset performance in the cross-section will recur during congruent-mood periods and reverse during noncongruent-mood periods. For example, if Asset A outperforms Asset B, on average, in January and March, then we expect A to underperform Asset B next September and October (reversal) but to outperform Asset B next January and March (recurrence), and we expect such patterns to repeat for years after the conditioning date. Similarly, if A outperforms B on Friday, we expect this average relative performance to alternate between Mondays and Fridays for months after the conditioning date.

To test for such effects, we use the historical seasonal returns in the past high or low mood periods to forecast returns in future hypothesized high (January, March, and Friday) and low (September, October, and Monday) mood periods in Fama-MacBeth regressions. The seasonal historical return is first proxied by the average asset return in these prespecified high or low mood periods in recent years or weeks.[5] Higher realized returns of the broad market in any month could in part reflect more optimistic mood shifts and lower returns less optimistic mood shifts. Thus we use another proxy for an asset's historical seasonal return: its average return earned during the recent months or weekdays with the highest or lowest equal-weighted market excess return in a given year or week.

We use three sets of test assets over the sample period of 1963–2016: the full cross-section of individuals stocks, the 94 Baker and Wurgler (2006, henceforth BW) portfolios and 79 Keloharju et al. (2016; henceforth KLN) portfolios, with the portfolios formed by sorting individual stocks on various firm characteristics. Consistent with our hypothesis, we show the novel finding that the relative performance across assets during a mood period indeed tends

[3] Shifts in mood may also cause seasonal changes in risk tolerance (e.g., Kamstra et al., 2003), which may also induce factor-wide price changes.

[4] DellaVigna and Pollet (2009) hypothesize that Fridays are associated with more investor inattention. This attention-based hypothesis predicts weaker market reactions to both positive and negative news announced on Fridays but does not predict an average misreaction. The mood seasonality hypothesis predicts more favorable market reactions to all news announced on Fridays, implying a positive average misreaction. It is, of course, possible that both attention and mood effects are present.

[5] These return predictors are similar to the historical same-month (Heston and Sadka, 2008) or same-weekday (Keloharju et al., 2016) return variables but differ in that these historical seasonal returns are not confined to the same calendar months; instead they are averaged across different calendar months with congruent mood (such as January and March or September and October).

to recur in future periods of hypothesized congruent mood and to reverse in future periods of hypothesized noncongruent mood. We call the former the congruent-mood recurrence effect and the latter the noncongruent-mood reversal effect. A long-short portfolio that exploits the mood recurrence or reversal effect generates a monthly risk-adjusted return of 0.33% to 1.80%, or a daily risk-adjusted return of 2 to 10 basis points.

In robustness checks, we also identify the hypothesized high or low mood months and weekdays in the future, not by prestipulated periods but by using the historical average equal-weighted market return up to the preceding year or month or the full sample average equal-weighted returns in only odd or even years. These alternative methods identify future hypothesized high and low mood periods out of sample for testing return predictability and therefore help minimize the concern of an in-sample bias. We observe similar cross-sectional return patterns in which relative asset performances flip across high and low mood periods.

A second innovation of this paper is to introduce the concept of mood beta, an asset's return sensitivity to investor mood variations. We estimate an asset's mood sensitivity as the slope coefficient in the regression of the asset's returns earned during the recent high and low mood periods upon the corresponding equal-weighted market returns.[6] We call the slope coefficient the mood beta. Alternatively, the sensitivity of return to mood can be estimated as the difference between mean historical returns in ascending versus descending mood periods normalized by the mean difference in market returns during those periods. For each stock or portfolio, we estimate its mood betas by using its monthly or weekday returns during the high and low mood periods in the latest ten years or six months. We also construct a composite mood beta as the first principal component of the two mood betas.

Our mood sensitivity hypothesis predicts that assets with high mood betas will, on average, outperform during periods with ascending moods and underperform when moods are descending. Consistent with this, we find strong evidence that assets with high mood betas earn higher average future returns during January and March and on Friday and earn lower average future returns during September and October and on Monday. Furthermore, mood betas vary with firm characteristics and industries in an intuitive pattern. Hard-to-value stocks and industries, and those sensitive to high sentiment (in the sense of Baker and Wurgler, 2006), have high mood betas, while easy-to-value assets and those less subject to sentiment have low mood betas.

To assess the economic magnitude of the mood beta effect, we form a hedge portfolio. The hedge portfolio goes long the highest mood beta decile and short the lowest decile during periods when mood is expected to rise and flips the long and short sides during periods when mood is expected to fall. This captures the expected positive return spread in both favorable and unfavorable mood periods. This hedge portfolio produces a significant Fama-French five-factor alpha of 1.5% or more per month and 12 basis points or more per day. As predicted, after controlling for mood beta, historical seasonal returns tend to have substantially reduced ability, sometimes with a reversed sign, to forecast returns in future high or low mood periods. These findings suggest that mood beta offers a unique and integrated explanation for a wide and varied set of seasonal return recurrence and reversal effects.

The effect of mood beta is robust to controls for market beta and the sentiment beta of Baker and Wurgler (2007). We perform a horse race between mood beta, market beta, and sentiment beta in Fama-MacBeth regressions. This is important since it would not be surprising if the different betas were correlated across assets. However, this association is far from perfect. We find that mood beta continues to be a strong forecaster of future returns after controlling for these other betas and a set of firm characteristics.

In contrast, in the same regressions, market beta is a negative predictor of future returns, consistent with the documented low-risk anomaly (Baker et al., 2011; Frazzini and Pedersen, 2014). The sentiment beta is a positive return predictor. We also find no evidence that the market beta or the sentiment beta consistently forecasts stock returns in future hypothesized high and low mood periods with opposite signs as predicted by our hypothesis for mood betas. Thus these other betas are not the source of the mood beta effects we identify.

Regardless of whether the effects shown in this paper derive from investor mood, as we hypothesize, they constitute a rich set of newly identified conditional return seasonalities that deserve attention. Mood beta provides a possible integrated explanation for this wide range of effects, and it is otherwise far from obvious how to explain them.

2. Background and hypotheses

Our mood seasonality hypothesis is motivated by the psychology of mood and by a line of literature on stock return seasonality. The basic month-of-the-year effect refers to the finding that aggregate stock markets tend to do better in certain calendar months (e.g., January) and do worse in other calendar months such as September and October (Lakonishok and Smidt, 1988; Bouman and Jacobsen, 2002). Cross-sectionally, Heston and Sadka (2008, 2010) find that relative performance across stocks tends to persist for years in the same calendar month. They rule out various possible explanations based upon volume, volatility, industry, earnings, and dividends but do not propose an explanation for this cross-sectional return seasonality.

We hypothesize that both the aggregate and cross-sectional seasonalities are induced by seasonal variations in investor mood. The strong early January performance of stock markets, especially among small firms (Keim, 1983), may derive from investor optimism at the turn of the year (e.g., Ritter, 1988; Doran et al., 2012;

[6] This is based on the premise that extreme mood periods also have relatively variable mood, improving the signal-to-noise ratio in estimation of mood beta. At the monthly level, these historical mood months include January, March, September, and October as well as the two highest and two lowest months in terms of realized equal-weighted market excess returns in a given year. At the weekday level, these historical mood weekdays include Monday and Friday as well as the highest and the lowest weekdays in terms of realized equal-weighted market excess returns in a given week.

Bergsma and Jiang, 2016; Kaustia and Rantapuska, 2016). The weak September and October performance may derive from the declining number of hours of daytime sunlight starting in early autumn, which is known to induce the SAD effect (Kamstra et al., 2003). Among all months, Kamstra et al. (2017) show that September to October are associated with the largest net increase in the proportion of SAD-affected individuals and the biggest fund flow from risky to safe assets. They also show that around March, the opposite is observed as the daylight hours start to increase.

The above evidence is consistent with the possibility that investor mood is improving in early January and around March (and quite possibly in February as well) while deteriorating in September to October. Also potentially consistent with this, during our sample period of 1963–2016, the average stock excess return (CRSP equal-weighted index return minus the riskfree rate) is highest in January, followed by March, and lowest in October, followed by September. We therefore use September and October as proxies for investors being in descending moods. For symmetry, we use two months for ascending mood—January and March as suggested by past literature.[7]

Previous literature has also shown the day-of-the-week effect, the finding that aggregate stock markets tend to do better at the end of the week (Friday) and worse at the beginning of the week (Monday) (French, 1980; Lakonishok and Smidt, 1988). Survey evidence suggests a downbeat mood on Mondays and an upbeat mood on Fridays among both the general and the investing populations (e.g., Rossi and Rossi, 1977; McFarlane et al., 1988; Stone et al., 2012; Helliwell and Wang, 2014).

In the cross-section, KLN find that stocks' relative performance on a given weekday persists for subsequent weeks on the same weekday. They hypothesize that cross-sectional return seasonality is a manifestation of seasonal factor premia but do not explore the economic or psychological sources of it. We instead point out the affective psychology behind the seasonal factor premia. Birru (2018) finds that the performance of many major anomaly strategies exhibit opposite return patterns on Monday versus Friday based on whether the short leg is betting on speculative or safe stocks and links these patterns to investor mood. As in Birru's paper, we use Friday to represent improving mood and Monday for deteriorating mood. However, we also study at the monthly frequency and individual stocks as well as mood betas of various test assets.

We hypothesize that these seasonal mood variations cause seasonal factor mispricing, resulting from mood-induced overoptimism or overpessimism about future expected factor payoffs. The factor-wide mispricing is inherited by the cross-section of stocks and portfolios according to their factor loadings. The general idea that factor-level mispricing predicts the cross section is modeled explicitly by Daniel et al. (2001) in the context of investor overconfidence. Differences in cross-sectional mispricing can also be induced by cross-stock differences in cost of arbitrage.

Baker and Wurgler (2006, 2007) provide evidence that shifts in general market sentiment influence stocks with different characteristics such as valuation uncertainty and cost of arbitrage. In our context, assets with higher loadings on the mispriced factor are more sensitive to mood shocks and will inherit greater factor mispricing. Thus, when a factor gets overpriced or underpriced upon seasonal mood shifts, assets with greater mood sensitivity will earn higher or lower returns accordingly.[8]

This effect manifests in the following two mood seasonality hypotheses that we test:

Hypothesis 1. (The mood recurrence and reversal effects): In the cross-section, a security's historical seasonal returns are positively correlated with its future seasonal returns under a congruent-mood period and negatively related to its future seasonal returns under a noncongruent-mood period.

Hypothesis 2. (The mood beta effect): Mood beta, which measures an asset's return sensitivity to mood, positively predicts the cross-section of security returns during ascending mood periods and negatively predicts the cross-section of security returns during descending mood periods.

Broadly, our study adds to research that explores how investor mood affects financial decision-making and asset prices. The effects of emotion are relatively neglected compared to the large body of research in behavioral finance on cognitive biases and nonstandard preferences such as prospect theory. There has been some past empirical research on feelings and financial decisions. Previous research reports that people in a more positive mood tend to be more risk-tolerant and exhibit a higher demand for risky assets (Bassi et al., 2013; Kaplanski et al., 2015; Breaban and Noussair, 2017). Weather conditions, sports outcomes, and aviation disasters are associated with aggregate stock market returns (Saunders 1993; Hirshleifer and Shumway, 2003; Edmans et al., 2007; Kaplanski and Levy, 2010), returns of individual stocks, perceived stock overpricing by institutional investors (Goetzmann et al., 2015), market reactions to earnings announcements (deHaan et al., 2017; Jiang et al., 2019), individuals' sentiment about the economy and life satisfaction (Makridis 2018), and firm hiring and investment decisions as well as hiring and creating new businesses (Chhaochharia et al., 2019).

3. Mood recurrence and reversal effects

Our US sample includes common stocks traded on the NYSE, Amex, and Nasdaq from January 1963 to December 2016. Daily and monthly stock and market portfolio returns, as well as other trading information, are obtained from the Center for Research in Security Prices (CRSP). Accounting data are obtained from Compustat.

[7] Untabulated robustness checks show that results are very similar if we instead use January and February for ascending mood. Also, we find qualitatively similar effects if we use only January to proxy for a high mood state.

[8] Mood can also induce shifts in risk tolerance. This has similar implications to the shifts in optimism and pessimism that we focus on since the returns of riskier stocks will be more sensitive to shifts in mood than less risky stocks. However, it is not clear that mood will affect risk tolerance at the high frequency of our daily tests.

We use three sets of test assets: the full cross section of individual stocks, the 94 BW portfolios and 79 KLN portfolios. The BW portfolios are formed monthly based on ten firm characteristics: firm age (AGE), book-to-market equity (B/M), dividends to equity (D/BE), external financing (EF/A), market equity (ME), sales growth (SG), tangible assets (PPE/A), research & development (R&D/A), return on equity (ROE), and return volatility (SIGMA). As in BW, we use the NYSE breakpoints for each characteristic to form portfolio deciles and to calculate equal-weighted portfolio returns. Nonpositive earnings, dividends, PPE, or R&D firms are included in a portfolio separately from the deciles sorted based on positive values of that characteristic.

The KLN portfolios are formed monthly based on six firm characteristics: B/M, ME, price momentum based on cumulative returns from month $t-12$ to $t-2$ (MOM), gross profitability (GP), dividend yield (D/P), and earnings-to-price (E/P). Further added to the KLN portfolios are the Fama-French 17 industry portfolios. As in KLN, we use breakpoints based on all firms to form the deciles, but we calculate equal-weighted, as opposed to value-weighted, portfolio returns because we believe that mood should have a stronger impact on small firms than on large firms. Firms with nonpositive earnings or dividend firms are included in separate portfolios from the decile portfolios. All definitions of the seasonal returns, firm characteristics, and portfolio formation are defined in the Data Appendix. Table 1 reports the seasonal returns summary statistics.

3.1. Month-level mood effects

During our sample period of 1963–2016, consistent with the psychology accounts reviewed in Section 2, the average stock excess return (CRSP equal-weighted index return minus the riskfree rate) is highest in January (5.06%), second highest in March (1.26%), lowest in October (−0.84%), and second lowest in September (−0.29%). Thus, in our main tests we use January and March as proxies for the hypothesized high mood months and September and October for the hypothesized low mood months. Later we explore the robustness of our findings to alternative definitions of hypothesized high and low mood months.

Using these four months, we first test for the return recurrence and reversal effect across congruent and noncongruent-mood month. The return recurrence test is similar to tests of the same-calendar-month effect found by Heston and Sadka (2008), but we do not differentiate January from March or September from October, as they proxy for the past high versus the low mood state, respectively.

3.1.1. The mood recurrence effect: returns during prespecified mood months as predictors

Specifically, we estimate the following Fama-MacBeth (FMB) regressions of the hypothesized high (January and March) or low (September and October) mood month returns across assets on their historical seasonal returns earned during these prespecified congruent-mood months at three sets of annual lags:

$$RET_{\text{high(Low)},\ t} = \eta_{k,t} + \gamma_{k,t} RET_{\text{high(Low)},\ t-k} + \varepsilon_t, \tag{1}$$

where $k = 1$, 2–5, and 6–10, and $RET_{\text{high(Low)},t}$ is the current mood month (high or low) asset return in year t, and $RET_{\text{high(Low)},\ t-k}$ is the historical average congruent (high or low) mood month return in year $t-k$ for the same asset. For example, for annual lag $k = 1$, the independent variable is the average January and March return of an asset of the prior year when forecasting January or March returns of the current year, and it is the average September and October return of the prior year when predicting current September or October returns. For multiple year lags (e.g., 2–5 or 6–10), the annual independent variables are averaged across the designated annual lags before being used as an independent variable in the regression.

We run cross-sectional regressions as in regression (1) for each mood month, and the estimates of $\gamma_{k,\ t}$ are averaged across the full sample period to yield the estimate for γ_k, reported as the FMB regression coefficient. Such regressions help to assess whether certain stocks tend to repeatedly outperform other stocks during the congruent-mood months year after year. We follow Heston and Sadka (2008) and call the slope coefficient estimate γ_k the "return response."

Our regression estimates for individual stocks are reported in Table 2, Panel A, Column (1). There is an insignificant coefficient for the first lag and positive and significant return responses for annual lags 2–5 (coefficient = 1.82%, $t = 2.65$) and lags 6–10 (coefficient = 4.37%, $t = 4.88$). The return responses represent significant economic impacts. For example, for annual lags 2–5, the return response suggests a one standard deviation (7.86%) increase in the prior mood month return leads to a 14 basis points (7.86%×1.82%) increase in the current congruent-mood return, or an 8.7% increase relative to the mean mood month return (1.64%) in each congruent-mood month during the next two to five years.

Moving to Panels B and C for the BW and KLN portfolios, the return responses are all positive, ranging from 19.20% to 48.75%, and significant at all three sets of lags with t statistics ranging from 4.23 to 7.09. The implied economic effect is larger; a one standard deviation increase in the historical return measure implies 60%−86% higher returns relative to the mean in each subsequent congruent-mood month up to ten years. Thus, our evidence confirms that asset returns exhibit recurrence across congruent-mood months for at least ten years after the conditioning date.

3.1.2. The mood recurrence effect: returns during realized mood months as predictors

Next, we expand the mood recurrence effect predicted based upon prespecified mood months by considering historical seasonal returns during realized mood months. We measure realized positive and negative mood periods using the top two and bottom two months, ranked by the equal-weighted CRSP excess returns realized in a given year.[9] The

[9] Our results hold if we focus on only the highest and lowest realized market return months. Further, we believe that the equal-weighted market index can more accurately reflect the collective mood effect for individual stocks than the value-weighted index, as individual investors are more prone to the mood influence and prefer trading small stocks.

Table 1
Summary Statistics.
This table reports the summary statistics of the main variables. The analyses include common stocks traded on the NYSE, Amex, or Nasdaq. All variables are defined in the Data Appendix. The sample period is from January 1963 to December 2016.

Panel A: Returns and betas of test assets

Variables	Mean	Median	Standard deviation	10% percentile	25% percentile	75% percentile	90% percentile
Individual stocks							
Month-level							
RET_{High}	3.65	1.11	20.67	−13.46	−5.36	9.38	21.69
RET_{Low}	−0.38	−0.22	17.38	−17.95	−8.08	6.29	16.00
$RET_{High/Low}$	1.64	0.00	19.20	−15.75	−6.67	7.83	18.85
RET_{RHigh}	8.59	5.17	22.24	−8.51	−1.10	14.38	27.78
RET_{RLow}	−6.37	−5.30	15.50	−23.53	−13.34	0.49	7.53
$RET_{RHigh/RLow}$	1.09	0.00	20.57	−17.96	−8.19	8.00	19.74
Weekday-level							
RET_{High}	0.22	0.00	4.45	−3.17	−1.1	1.23	3.64
RET_{Low}	−0.09	0.00	4.58	−3.77	−1.43	1.04	3.37
$RET_{High/Low}$	0.07	0.00	4.51	−3.45	−1.26	1.14	3.51
RET_{RHigh}	0.83	0.00	4.69	−2.50	−0.43	1.98	4.60
RET_{RLow}	−0.71	−0.08	4.46	−4.49	−2.08	0.26	2.46
$RET_{Rhigh/Rlow}$	0.06	0.00	4.64	−3.66	−1.36	1.23	3.70
$\beta^{Mood}{}_{Month}$	1.02	0.93	0.69	0.30	0.58	1.34	1.81
$\beta^{Mood}{}_{Weekday}$	1.06	0.96	1.08	0.01	0.41	1.58	2.28
β^{Mood}	0.00	−0.12	1.00	−1.13	−0.68	0.56	1.30
$\beta^{MKT}{}_{Month}$	1.09	1.04	0.65	0.35	0.65	1.45	1.89
$\beta^{MKT}{}_{Weekday}$	0.74	0.66	3.95	−0.06	0.23	1.17	1.70
β^{SENT}	0.24	0.20	2.91	−2.55	−0.99	1.44	3.11
Baker and Wurgler (BW) portfolios							
Month-level							
RET_{High}	3.25	2.44	6.41	−3.33	−0.50	6.51	10.39
RET_{Low}	−0.02	0.77	6.52	−7.36	−3.03	3.61	7.02
$RET_{High/Low}$	1.61	1.61	6.67	−5.35	−1.72	4.92	8.95
RET_{RHigh}	8.01	7.15	4.90	3.20	4.99	9.93	13.24
RET_{RLow}	−5.75	−4.52	5.05	−11.76	−7.66	−2.47	−0.93
$RET_{RHigh/RLow}$	1.13	0.94	8.49	−8.68	−4.52	7.15	10.79
Weekday-level							
RET_{High}	−0.07	0.00	1.07	−1.15	−0.48	0.42	0.92
RET_{Low}	0.18	0.21	0.85	−0.72	−0.19	0.59	1.03
$RET_{High/Low}$	0.06	0.11	0.97	−0.95	−0.34	0.52	0.98
RET_{RHigh}	0.85	0.68	0.86	0.09	0.35	1.13	1.75
RET_{RLow}	−0.73	−0.51	0.98	−1.82	−1.08	−0.13	0.15
$RET_{RHigh/RLow}$	0.06	0.12	1.21	−1.27	−0.53	0.69	1.28
$\beta^{Mood}{}_{Month}$	0.95	0.95	0.22	0.68	0.83	1.08	1.20
$\beta^{Mood}{}_{Weekday}$	1.05	1.06	0.21	0.78	0.93	1.18	1.30
β^{Mood}	0.00	−0.04	1.00	−1.18	−0.58	0.62	1.26
$\beta^{MKT}{}_{Month}$	1.09	1.10	0.22	0.82	0.96	1.23	1.36
$\beta^{MKT}{}_{Weekday}$	0.82	0.81	0.25	0.52	0.63	1.00	1.13
β^{SENT}	0.20	0.18	0.38	−0.25	−0.04	0.42	0.69
Keloharju, Linnainmaa, and Nyberg (KLN) portfolios							
Month-level							
RET_{High}	3.45	2.60	6.86	−3.54	−0.54	6.75	10.92
RET_{Low}	−0.14	0.58	6.76	−7.72	−3.23	3.58	7.08
$RET_{High/Low}$	1.65	1.59	7.04	−5.72	−1.84	5.16	9.27
RET_{RHigh}	8.15	7.20	5.47	2.96	4.88	10.19	13.96
RET_{RLow}	−5.91	−4.71	5.30	−12.31	−8.03	−2.44	−0.81
$RET_{RHigh/RLow}$	1.12	0.80	8.85	−9.05	−4.72	7.21	11.13
Weekday-level							
RET_{High}	0.19	0.22	0.88	−0.71	−0.19	0.61	1.06
RET_{Low}	−0.08	−0.01	1.1	−1.16	−0.49	0.42	0.93
$RET_{High/Low}$	0.06	0.11	1.00	−0.96	−0.34	0.53	1.00
RET_{RHigh}	0.84	0.66	0.90	0.06	0.33	1.12	1.77
RET_{RLow}	−0.71	−0.50	1.02	−1.85	−1.08	−0.11	0.18
$RET_{RHigh/RLow}$	0.06	0.12	1.24	−1.27	−0.52	0.68	1.28
$\beta^{Mood}{}_{Month}$	0.98	0.99	0.23	0.70	0.86	1.10	1.21
$\beta^{Mood}{}_{Weekday}$	1.04	1.05	0.25	0.74	0.89	1.19	1.33
β^{Mood}	0.00	0.00	0.99	−1.15	−0.63	0.61	1.21
$\beta^{MKT}{}_{Month}$	1.09	1.10	0.24	0.79	0.95	1.25	1.38
$\beta^{MKT}{}_{Weekday}$	0.80	0.78	0.29	0.45	0.58	1.00	1.18
β^{SENT}	0.23	0.20	0.45	−0.30	−0.06	0.49	0.81

(*continued on next page*)

Table 1 (*continued*)

Panel B: Firm characteristics

Variables	Mean	Median	Standard deviation	10% percentile	25% percentile	75% percentile	90% percentile
AGE	155	97	168	15	39	208	384
B/M	0.91	0.67	0.92	0.19	0.36	1.13	1.81
D/BE	0.02	0.00	0.04	0.00	0.00	0.04	0.06
D/P	0.02	0.00	0.02	0.00	0.00	0.02	0.05
ROE	0.10	0.09	0.09	0.00	0.00	0.15	0.21
EF/A	0.09	0.05	0.24	−0.08	−0.01	0.15	0.32
E/P	−0.03	0.05	0.44	−0.21	−0.01	0.09	0.15
GP	0.32	0.29	0.30	0.03	0.12	0.48	0.69
SG	0.21	0.10	0.78	−0.14	−0.01	0.23	0.48
ME	1.42	0.08	9.84	0.01	0.02	0.40	1.76
MOM	0.13	0.05	0.60	−0.46	−0.21	0.33	0.72
PPE/A	0.53	0.45	0.39	0.09	0.22	0.77	1.08
R&D/A	0.04	0.00	0.10	0.00	0.00	0.03	0.11
SIGMA	0.14	0.11	0.10	0.05	0.08	0.17	0.24

Table 2

Mood month return recurrence and reversal effects.

This table reports the estimates of Fama-MacBeth regressions to test for return recurrence and reversal effects across mood months in the cross-section. For the congruent-mood recurrence effect, we regress high (low) mood month returns across assets on their own past high (low) mood month returns or their own past returns during the two realized high (low) mood months. $\text{RET}_{\text{High(Low)}}$ refers to the high (or low) mood months identified using the full sample equal-weighted market excess returns: January and March (September and October). $\text{RET}_{\text{RHigh(RLow)}}$ refers to the high (or low) mood months identified using the realized equal-weighted excess market returns in a given year. For the noncongruent-mood reversal effect, the independent variables are flipped to forecast the future high (low) mood month returns. The reported coefficient is the time-series average of the return responses, reported in percentages for annual lags up to ten. For regressions with year lags 2–5 or 6–10, the annual independent variables are averaged across the designated lags before used in the regression. The reported Fama-MacBeth *t*-statistics are in parentheses and are corrected for heteroskedasticity and autocorrelation using Newey and West (1987). The symbols *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, based on two-tailed tests. All variables are defined in the Data Appendix. The sample period is from January 1963 to December 2016.

Dep. var.	*Congruent-mood recurrence* $\text{RET}_{\text{High(Low)}}$		*Noncongruent-mood reversal* $\text{RET}_{\text{High(Low)}}$	
Indep. var. (Lagged)	$\text{RET}_{\text{High(Low)}}$	$\text{RET}_{\text{RHigh(RLow)}}$	$\text{RET}_{\text{Low(High)}}$	$\text{RET}_{\text{RLow(RHigh)}}$
Year lag (k)	(1)	(2)	(3)	(4)
Panel A: Individual stocks				
1	1.05	1.37*	−3.00***	−3.99***
	(1.54)	(1.76)	(−3.01)	(−3.38)
2~5	1.82***	3.20***	−5.63***	−8.65***
	(2.65)	(2.64)	(−5.77)	(−5.95)
6~10	4.37***	5.44***	−2.65***	−6.38***
	(4.88)	(3.65)	(−3.58)	(−4.55)
Panel B: Baker and Wurgler (BW) portfolios				
1	20.63***	20.39***	−11.30*	−16.70***
	(4.23)	(4.41)	(−1.76)	(−3.00)
2~5	43.03***	29.35***	−30.0***	−26.2***
	(4.74)	(4.70)	(−3.53)	(−4.27)
6~10	48.75***	35.89***	2.04	−28.7***
	(6.38)	(5.29)	(0.18)	(−3.86)
Panel C: Keloharju, Linnainmaa, and Nyberg (KLN) portfolios				
1	19.20***	17.96***	−5.55	−11.50**
	(4.52)	(4.45)	(−1.12)	(−2.24)
2~5	32.40***	26.39***	−22.0***	−27.00***
	(4.36)	(4.32)	(−3.09)	(−4.30)
6~10	47.08***	33.23***	−3.11	−25.10***
	(7.09)	(5.40)	(−0.42)	(−3.91)

rationale, as discussed previously, relies on the assumption that extreme realized average returns are likely to reflect extreme mood swings.

Using FMB regressions, we employ the relative performance across assets in these recent realized high and low mood months to forecast the cross-section of returns in subsequent, hypothesized congruent-mood months:

$$RET_{\text{High(Low)},\ t} = \eta_{k,t} + \gamma_{k,t} RET_{\text{RHigh(Rlow)},\ t-k} + \varepsilon_t, \qquad (2)$$

where $RET_{\text{RHigh(Rlow)},\ t-k}$ is the historical return during the two highest (lowest) market return months realized in year $t-k$. The return responses are reported in Column (2) of Table 2. For individual stocks, we obtain positive and significant return responses for all three sets of annual lags, significant at the 10%, 1% and 1% levels, respectively. The average return response for lags 2–5 is 3.20%, implying that a one standard deviation (3.05%) return increase in the historical realized extreme mood month leads to 10 basis points, or a 6%, higher returns relative to the mean in each of the future hypothesized congruent-mood months of the subsequent five years.

For the BW and KLN portfolios, the return responses are all positive, ranging from 18% to 36%, and significant at the 1% level. The implied economic impact is considerably larger; a one standard deviation change in the historical return measure leads to 101% to 227% higher returns relative to the mean in each of future mood months. This evidence supports our conjecture that cross-sectional returns recur across the congruent-mood months even when we identify mood swings in the past using realized average stock returns.

3.1.3. The mood reversal effect: returns during prespecified mood months as predictors

Next, we test for the cross-sectional reversal effect across prespecified, noncongruent, recurrent mood periods, again proxied by January and March for high moods and September and October for low moods. In such regressions, we simply switch the independent variables in regression (1) when forecasting future high or low mood month returns, through which we test whether the historical high mood month returns reverse during future low mood months and vice versa.

In Column (3) of Table 2, we report the regression estimates. For individual stocks, the return responses are all negative and significant at the 1% for the three sets of lags. The coefficient for annual lags 2–5 is −5.63% ($t = -5.77$), suggesting that a one standard deviation increase in the most recent noncongruent-month return leads to a 27% lower return relative to the mean in each of the noncongruent-mood months in the subsequent five years.

The return response is negative and significant for annual lags up to five for the BW portfolios and only for lags 2–5 for the KLN portfolios. In both cases, the economic effect represents a 41% to 53% return reduction resulting from a one standard deviation increase in the historical return. Most interestingly, for $k = 1$, reversal is observed for individual stocks and the BW portfolios despite the fact that monthly returns in the prior year typically exhibit a momentum effect (Jegadeesh and Titman 1993). The evidence thus shows that a cross-sectional reversal effect takes place across prespecified, noncongruent-mood states, at least for a few subsequent years.

3.1.4. The mood reversal effect: returns during realized mood months as predictors

The reversal effect can also be identified by using recent realized mood periods identified by extreme historical equal-weighted CRSP excess returns. The regressions are done by switching the independent variables in regression (3.2). In Column (4) of Table 2, we report the estimates from regressions of the current hypothesized high or low mood month returns across stocks on their own historical returns in prior years during the recent realized low or high mood months, respectively.

We obtain significant negative return responses across all lags for all three sets of test assets. For lags 2–5, the return response is −8.65% ($t = -5.95$) for individual stocks, −26.2% ($t = -4.27$) for the BW portfolios, and −27.0% ($t = -4.30$) for the KLN portfolios. These return responses represent a 16% to 108% lower monthly return relative to the mean for a one standard deviation increase in the historical realized noncongruent-mood month return. This is again a remarkably strong return reversal effect at a time when investor mood is expected to reverse.

Taken together, our results in Table 2 suggest the existence of strong congruent-mood recurrence effects and noncongruent-mood reversal effects at the monthly frequency, regardless of whether we identify historical mood months using average or realized market performances. The estimated economic effect is stronger for portfolios than for individual stocks. These effects seemingly connect independent cross-sectional seasonalities across different calendar months with the congruent or noncongruent mood.

3.2. Weekday-level mood effects

At a higher frequency, we explore whether the cross-sectional recurrence and reversal effects are present across days of the week with hypothesized moods: Mondays and Fridays. In untabulated tests, we first verify the findings from previous studies that stocks as a whole, measured by the equal-weighted market portfolio, earn higher returns on Fridays (19 basis points) and lower returns on Monday (−10 basis points) during our sample period 1963–2016. We then go beyond previous findings to examine weekday congruent-mood recurrence and noncongruent-mood reversal effects.

3.2.1. The mood recurrence effect: returns during prespecified mood weekdays as predictors

We examine the congruent-mood recurrence effect at the weekday frequency using FMB regressions, similar to KLN but using only Monday and Friday stock returns. We rerun regression as in regression (1) for Mondays, proxying hypothesized low moods, and Fridays, proxying hypothesized high moods.

For individual stocks, Column (1) in Table 3 shows that historical Monday/Friday weekday returns across stocks are strong positive predictors of their subsequent congruent-mood-weekday returns beyond the first lag, which has an insignificant return response. The return responses for week lags 2–10 and 11–20 are 1.96% ($t = 9.90$) and 2.53% ($t = 13.22$), statistically significant at the 1% level, implying a 52% to 62% higher future Monday/Friday return for a one standard deviation increase in the historical congruent-mood-weekday return.[10] The insignificance at the first lag is also observed by KLN, owing to the short-term reversal effect of one-month returns (Jegadeesh, 1990) that appears to be unusually strong during the first week.[11]

[10] Untabulated tests show that the predictive power of the same-weekday return persists for at least 50 weeks at the individual stock level.

[11] Keloharju et al. (2016) show that past daily returns are in general negatively related to future daily returns in the subsequent four weeks, except for the same-weekday returns, which are much less negative or slightly positive.

Table 3
Mood weekday return recurrence and reversal effects.
This table reports the estimates of Fama-MacBeth regressions to test for return recurrence and reversal effects across mood weekdays in the cross-section. The dependent variable is the asset return on Friday or Monday. For the congruent-mood recurrence effect, we regress high (Friday) or low (Monday) mood weekday returns across assets on their own past average Friday or Monday returns or their own past returns during the realized high or low mood weekdays. $RET_{High(Low)}$ refers to the high (low) mood weekdays identified using the full sample equal-weighted market excess returns: Friday (Monday). $RET_{RHigh(RLow)}$ refers to the high and low mood weekdays identified using the realized equal-weighted market excess returns in a given week. For the noncongruent-mood reversal effect, the independent variables are switched to forecast the future high (low) mood weekday returns. The reported coefficient is the time-series average of the return responses, reported in basis points for weekly lags up to 20. For regressions with week lags 2–10 or 11–20, the weekly independent variables are averaged across the lags before used in the regression. The reported Fama-MacBeth t-statistics are in parentheses and are corrected for heteroskedasticity and autocorrelation using Newey and West (1987). The symbols *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, based on two-tailed tests. All variables are defined in the Data Appendix. The sample period is from January 1963 to December 2016.

Dep. var.	*Congruent-mood recurrence* $RET_{High(Low)}$		*Noncongruent-mood reversal* $RET_{High(Low)}$	
Indep. var. (Lagged)	$RET_{High(Low)}$	$RET_{RHigh(RLow)}$	$RET_{Low(High)}$	$RET_{RLow(RHigh)}$
Week lag (k)	(1)	(2)	(3)	(4)
Panel A: Individual stocks				
1	0.01	−0.62***	−4.87***	−1.90***
	(0.07)	(−6.67)	(−31.6)	(−17.2)
2~10	1.96***	1.43***	−1.80***	−1.53***
	(9.90)	(5.39)	(−9.15)	(−5.59)
11~20	2.53***	2.28***	−0.92***	−1.34***
	(13.22)	(8.67)	(−4.65)	(−5.14)
Panel B: Baker and Wurgler (BW) portfolios				
1	7.00***	6.61***	2.38***	0.84*
	(13.94)	(13.93)	(4.44)	(1.77)
2~10	22.36***	14.30***	−5.80***	−4.63***
	(17.30)	(12.93)	(−4.30)	(−4.05)
11~20	17.13***	12.09***	−9.14***	−7.69***
	(11.94)	(10.61)	(−6.77)	(−6.79)
Panel C: Keloharju, Linnainmaa, and Nyberg (KLN) portfolios				
1	8.40***	8.20***	4.91***	2.37***
	(15.31)	(16.17)	(8.52)	(4.67)
2~10	24.63***	14.77***	−1.17	−1.46
	(18.37)	(13.05)	(−0.85)	(−1.28)
11~20	19.51***	11.65***	−5.57***	−4.68***
	(13.41)	(10.23)	(−4.00)	(−4.13)

For the BW and KLN portfolios, the return responses are all positive and significant at the 1% level across the three sets of lags. The size of the return response implies a 101% to 160% higher future Monday/Friday portfolio return for a one standard deviation increase in the historical congruent weekday return.[12] Thus, our evidence confirms recurrent relative performances across stocks or portfolios across prespecified congruent-mood weekdays: Monday and Friday.

3.2.2. The mood recurrence effect: returns during realized mood weekdays as predictors

We extend the mood-weekday recurrence effect to identifying realized mood weekdays by using the two days with the highest or lowest CRSP equal-weighted excess return realized in a given week. Then we test whether cross-sectional performance in prior realized extreme mood periods recurs on subsequent weekdays with hypothesized congruent moods (Friday and Monday), similar to regression (3.2).

Column (2) of Table 3 reports the estimates. Across the three panels, the return responses are all significantly positive across assets and week lags except for the first lag of individual stocks, again likely owing to the short-term return reversal effects at the individual stock level. For week lags 2–10, the return responses are 1.43% (t = 5.39), 14.30% (t = 12.93), 14.77% (t = 13.05), for individual stocks, the BW portfolios and the KLN portfolios, respectively. These return responses represent 44% to 243% higher returns for a one standard deviation increase in the predictor for each Monday or Friday during the next two to ten weeks.

3.2.3. The mood reversal effect: returns during prespecified mood weekdays as predictors

For the reversal effect across noncongruent weekdays, we regress Friday or Monday returns across stocks on their noncongruent-mood-weekday returns (Monday or Friday, respectively) in prior weeks. That is, we switch the independent variables in regression (1) when forecasting returns on the hypothesized future high and low weekdays.

As reported in Column (3) of Table 3, Panel A, we observe a significant negative return response for all three sets of lags for individual stocks. For lags 2–10, the return response is −1.80% (t = −9.15), suggesting a 48% lower return relative to the mean is expected during Mondays (Fridays) of the next two to ten weeks for a one standard deviation increase in the average Friday (Monday) return over

[12] Untabulated tests show that the predictive power of the congruent-mood-weekday return persists for at least 50 weeks at the portfolio level.

the past two to ten weeks. In Panels B and C, the significant negative return response is present for lags 2–10 and 11–20 for the BW portfolios and only for lags 11–20 for the KLN portfolios, suggesting a weaker return reversal effect across prespecified noncongruent-mood weekdays at the portfolio level.

3.2.4. The mood reversal effect: returns during realized mood weekdays as predictors

Analogous to the monthly returns, a stronger reversal effect is also observed across noncongruent-mood weekdays when the prior mood is identified using historical, realized extreme equal-weighted CRSP excess returns. We regress hypothesized high or low mood weekday (i.e., Friday or Monday) returns across assets on their historical returns realized on the weekday with the lowest or highest market return of the prior weeks for three sets of week lags, k = 1, 2–10, 6–20, when mood is presumably noncongruent.

For individual stocks, the return responses reported in Column (4) of Table 3, Panel A are all negative and significant at the 5% level or better. The economic impact is large; a one standard deviation increase in past average noncongruent-mood-weekday return corresponds to a 146%, 56%, and 27% lower return relative to the mean, respectively, for each of the next 1, 10, and 20 Monday and Fridays.

When we move to Panels B and C, however, for the BW and KLN portfolios, the return response is positive for the first lag. It turns negative and significant when we move to longer lags, suggesting the reversal effects take place only after the first few weeks. Overall, this evidence indicates that when investor mood switches between noncongruent states in a predictable way, cross-sectional return reversals occur strongly at the individual stock level and to some extent at the portfolio level.

4. Mood beta effect

The evidence in Tables 2 and 3 provides support for our mood seasonality hypothesis: relative stock performance tends to recur between congruent-mood periods and to reverse between noncongruent-mood periods across the cycle of calendar months and weekdays. We next employ mood beta to integrate the various seasonality effects.

We measure mood beta by an asset's return sensitivity to the equal-weighted market excess returns during the past high and low mood periods. This approach starts from the idea that shifts in mood are manifested in both returns on the equal-weighted market portfolio as well as on individual assets. Unfortunately, regressing asset returns on aggregate returns reflects both variation in mood and in fundamentals. However, it is likely that extreme mood periods also have relatively variable mood. Thus restricting the regression to extreme mood periods can potentially improve the signal-to-noise ratio in estimation of mood beta. As a robustness check, we alternatively estimate mood sensitivity using mean return differences between high and low mood periods.

4.1. Monthly mood beta

The first mood beta is estimated by using monthly returns during the past high and low mood periods. Specifically, using a 10-year rolling window by requiring a minimum of 40 observations, we estimate mood beta for each asset from time-series regressions of the asset's historical excess returns earned during prespecified and realized high and low mood months ($XRET_{i,\text{ MoodMonth}}$) on the contemporaneous equal-weighted CRSP excess returns ($XRET_{A,\text{ MoodMonth}}$).

$$XRET_{i,\text{MoodMonth}} = \alpha_i + \beta^{\text{Mood}}_{i,\text{ month}} XRET_{A,\text{MoodMonth}} + \varepsilon_i. \quad (3)$$

The regression thus includes eight months in a year: four prespecified (January, March, September, and October) and four realized high and low mood months (the top two and bottom two months with the highest and lowest realized equal-weighted market returns).[13] The estimated $\beta^{\text{Mood}}_{i,\text{ month}}$ is called the monthly mood beta. It measures the average return change of an asset in response to a 1% aggregate return change in the identified historical mood months.

In unreported tests, we obtain similar results if we estimate mood beta using only four months a year based on either prespecified or realized moods. We also use an alternative mood beta measure, defined as the ratio $(\overline{XRET}_{i,\text{HighRHigh}} - \overline{XRET}_{i,\text{LowRLow}})/(\overline{XRET}_{A,\text{HighRHigh}} - \overline{XRET}_{A,\text{LowRLow}})$, where each variable indicates average excess returns across the high or low mood months. The ratio-based mood beta also captures the average return change for an asset when the average aggregate return increases by one percentage point from periods with declining moods to periods with improving moods. The results using this ratio are qualitatively similar to those using the regression-based mood beta.

Moving to the second stage, we run FMB regressions of asset returns in the current hypothesized high (January and March) or low mood month (September and October) on their mood beta, estimated using prior return information ending in year $t-5$ to $t-2$, the lags for which we observe robust mood recurrence and reversal effects in Table 2. Mood betas are averaged across multiple annual lags for use as a regressor. Estimates for lags of t = 1 or 6–10 are unreported and are similar to the baseline regressions.

Our mood seasonality Hypothesis 2 predicts that high mood beta stocks will do better in subsequent high mood months and will do worse in subsequent low mood months. Thus, our cross-sectional regressions flip the sign of the mood beta (equivalent to flipping the sign of estimated slope coefficient) when forecasting low mood month returns. In consequence, the estimated coefficient ($\lambda_{k,\ t}$ in regression (4) below) in the cross-sectional regressions is expected to be positive. As phase two of the FMB regressions, we average $\lambda_{k,\ t}$ across time to yield λ_k, which

[13] If a month appears twice in a year based on the two identification criteria, then it counts as two observations in the regressions. This of course induces residual correlation in the regression, but this is not a concern for our purposes. We use these regressions to estimate the mood beta for use in other tests, not to evaluate its statistical significance.

Table 4
Mood beta to predict cross section of returns.

This table examines the predictive power of mood beta to forecast future mood month or weekday returns in Fama-MacBeth regressions. The key independent variable (β^{Mood}) refers to the monthly mood beta ($\beta^{Mood}{}_{Month}$) when we forecast mood month returns and the weekday mood beta ($\beta^{Mood}{}_{Weekday}$) when we forecast mood weekday returns. When forecasting future returns during a high mood state, the independent variable is the stock's historical β^{Mood}, and when forecasting future returns during a low mood state, it is $-\beta^{Mood}$. The other independent variable is the residual return earned during congruent or noncongruent-mood months or weekdays in the past, which is orthogonalized to β^{Mood}. Estimates for regressions with year lags 2–5 and week lags 2–10 are reported. Mood betas and residual returns are averaged across the designated year or week lags before used as a regressor. Regression estimates are reported percentages. The reported Fama-MacBeth *t*-statistics are in parentheses and are corrected for heteroskedasticity and autocorrelation using Newey and West (1987). The symbols *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, based on two-tailed tests. All variables are defined in the Data Appendix. The sample period is from January 1968 to December 2016.

Dep. var.	*Congruent-mood recurrence* $RET_{High(Low)}$				*Noncongruent-mood reversal* $RET_{High(Low)}$			
Indep. var.	(1)		(2)		(3)		(4)	
	$\pm\beta^{Mood}$	$RET^{\perp}_{High(Low)}$	$\pm\beta^{Mood}$	$RET^{\perp}_{RHigh(RLow)}$	$\pm\beta^{Mood}$	$RET^{\perp}_{Low(High)}$	$\pm\beta^{Mood}$	$RET^{\perp}_{RLow(RHigh)}$
Panel A: Individual stocks								
Year lag	1.47***	0.76	1.48***	−4.91***	1.47***	−3.87***	1.48***	−0.31
2~5	(4.83)	(1.06)	(4.84)	(−4.67)	(4.83)	(−4.01)	(4.84)	(−0.35)
Week lag	0.05***	1.66***	0.05***	0.27	0.05***	−1.64***	0.05***	−0.69***
2~10	(7.59)	(8.72)	(7.61)	(1.32)	(7.59)	(−8.94)	(7.56)	(−3.30)
Panel B: Baker and Wurgler (BW) portfolios								
Year lag	2.73***	30.47***	2.73***	9.52**	2.73***	−10.30*	2.73***	10.91**
2~5	(5.26)	(6.00)	(5.26)	(2.00)	(5.26)	(−1.95)	(5.26)	(2.06)
Week lag	0.12***	19.30***	0.12***	6.07***	0.12***	−1.68	0.12***	9.95***
2~10	(11.11)	(18.53)	(11.13)	(5.67)	(11.11)	(−1.54)	(11.10)	(9.27)
Panel C: Keloharju, Linnainmaa, and Nyberg (KLN) portfolios								
Year lag	2.95***	24.94***	2.95***	8.35	2.95***	−12.4*	2.95***	2.06
2~5	(6.03)	(4.45)	(6.03)	(1.61)	(6.03)	(−1.95)	(6.03)	(0.36)
Week lag	0.10***	24.76***	0.10***	10.61***	0.10***	0.90	0.10***	10.50***
2~10	(9.43)	(21.23)	(9.42)	(9.33)	(9.43)	(0.73)	(9.39)	(8.95)

we call the mood premium and captures the average size of the positive return spread between the high and low mood beta assets in high mood periods and that of the negative return spread in low mood periods.

Furthermore, to explore the extent to which the congruent-mood recurrence effects are explained by mood beta, we orthogonalize the historical seasonal returns on mood beta. The orthogonalized historical seasonal returns, denoted as $RET^{\perp}_{High(Low)}$, proxy for firm-specific mood sensitivity or a component that is totally unrelated to mood.

$$RET_{High,t} = \eta_{k,t} + \lambda_{k,t}\beta^{Mood}_{i,Month,t-k} + \gamma_{k,t}RET^{\perp}_{High,t-k} + \varepsilon_t, \text{ and}$$
$$RET_{Low,t} = \eta_{k,t} - \lambda_{k,t}\beta^{Mood}_{i,Month,t-k} + \gamma_{k,t}RET^{\perp}_{Low,t-k} + \varepsilon_t. \quad (4)$$

As reported in Column (1) of Table 4, Panel A, for individual stocks (year lag 2–5), the estimated mood premium (γ_k) is 1.47% ($t = 4.83$), implying that a one standard deviation increase in mood beta (0.69) leads to an average 101 basis points (= 1.47%*0.69) return increase (decrease) in each of the next ten Januaries and Marches (Septembers and Octobers).

After accounting for the correlation with mood beta, the coefficient of $RET^{\perp}_{High(Low)}$ becomes insignificant. The visible reduction in the predictive power of the historical seasonal return relative to that of the baseline seasonal return predictive regression (Column (1) of Table 2) suggests that mood beta captures a major and stable component of the historical seasonal returns.

For the BW and KLN portfolios, the mood premium estimates reported in Panels B and C nearly double, 2.73% and 2.95% per month, significant at the 1% level. $RET^{\perp}_{High(Low)}$, however, continues to carry a significant positive coefficient for the portfolios. Replacing $RET^{\perp}_{High(Low)}$ with $RET^{\perp}_{RHigh(RLow)}$ in Column (2) of Table 4 has no effect on the forecasting power of mood beta but tends to diminish and sometimes flip the sign of the coefficient on $RET^{\perp}_{RHigh(RLow)}$.

To test whether mood beta explains the noncongruent-mood reversal effect, we add both mood beta and the orthogonalized historical seasonal returns earned during noncongruent-mood month ($RET^{\perp}_{Low(high)}$) to the FMB regressions as below:

$$RET_{High,t} = \eta_{k,t} + \lambda_{k,t}\beta^{Mood}_{i,Month,t-k} + \gamma_{k,t}RET^{\perp}_{Low,t-k} + \varepsilon_t, \text{ and}$$
$$RET_{Low,t} = \eta_{k,t} - \lambda_{k,t}\beta^{Mood}_{i,\ Month,t-k} + \gamma_{k,t}RET^{\perp}_{High,t-k} + \varepsilon_t. \quad (5)$$

Shown in Column (3) of Table 4, $RET^{\perp}_{Low(high)}$ tends to exhibit considerably diminished predictive power; it is statistically significant for individual stocks and marginally significant for two sets of portfolios. In contrast, the mood premium ($\lambda_{k,\ t}$) estimates remain positive and significant at the 1% level for all three cases.

Next, we replace $RET^{\perp}_{Low(high)}$ in regressions (4.3) with the orthogonalized historical returns earned during the realized high or low mood months ($RET^{\perp}_{RLow(RHigh)}$). The estimates are reported under Columns (4) in Table 4. These orthogonalized historical return measures lose their predictive power for two of the three test assets. In contrast, the mood premium estimates remain virtually unchanged. The findings overall suggest that mood beta accounts for a majority, if not all, of the month-level return recurrence and reversal effects.

4.2. Weekday mood beta

Moving to weekday returns, we estimate mood beta for each asset from time-series regressions of a stock's excess return during the prespecified and realized high and low mood days of the week on the corresponding equal-weighted market excess returns using a six-month rolling window (by requiring a minimum of 50 observations):

$$XRET_{i,\text{MoodWeekday}} = \alpha_i + \beta_{i,\ \text{weekday}}^{\text{Mood}} XRET_{A,\text{MoodWeekday}} + \varepsilon_i. \tag{6}$$

The regression thus includes Fridays, Mondays, and the weekdays with the highest and lowest equal-weighted market excess returns realized in a week. The estimated coefficient on the market excess return is called the weekday mood beta. We obtain qualitatively similar results if we define mood beta as a ratio: $(\overline{XRET}_{i,\text{HighRHigh}} - \overline{XRET}_{i,\text{LowRLow}})/(\overline{XRET}_{A,\text{HighRHigh}} - \overline{XRET}_{A,\text{LowRLow}})$.

We next use the estimated $\beta_{i,\ \text{Weekday}}^{\text{Mood}}$ to forecast future returns on hypothesized high and low mood weekdays (Fridays and Mondays) by controlling for the orthogonalized historical seasonal returns earned during congruent-mood weekday ($RET^{\perp}_{\text{High(Low)}}$).

$$RET_{\text{High},t} = \eta_{k,t} + \lambda_{k,t}\beta^{\text{Mood}}_{i,\text{Weekday},t-k} + \gamma_{k,t}RET^{\perp}_{\text{high},t-k} + \varepsilon_t, \text{ and}$$
$$RET_{\text{Low},t} = \eta_{k,t} - \lambda_{k,t}\beta^{\text{Mood}}_{i,\text{Weekday},t-k} + \gamma_{k,t}RET^{\perp}_{\text{Low},t-k} + \varepsilon_t. \tag{7}$$

We focus on the mood betas estimated using prior six-month weekday returns ending in week $t-10$ to $t-2$. Mood betas across the multiple lags are averaged to generate the mood beta regressor. Estimates for lags of $t = 1$ or 11–20 are unreported and are similar to the baseline regressions. As reported in Column (1) of Table 4 (week lag 2–10), the estimated mood beta premium is positive and significant at the 1% level for all three sets of test assets, with the size of the daily premium at 5 basis points for individual stocks, 12 basis points for the BW portfolios, and 10 basis points for the KLN portfolios. The estimated return response on $RET^{\perp}_{\text{High(Low)}}$ remains positive and significant for all three sets of test assets.

In Column (2) of Table 4, we report the estimates for a similar specification as in regression (4.5) in which we replace $RET^{\perp}_{\text{High(Low)}}$ with $RET^{\perp}_{\text{RHigh(RLow)}}$, identified from realized weekday returns of the equal-weighted market portfolio. Again, all mood beta premia are positive and significant at the 1% level, while only one out of three coefficients of $RET^{\perp}_{\text{RHigh(RLow)}}$ is significant.

Next we replace $RET^{\perp}_{\text{High(Low)}}$ in regression (4.5) by those earned during the noncongruent-mood weekdays, either using $RET^{\perp}_{\text{Low(High)}}$ or $RET^{\perp}_{\text{RLow(RHigh)}}$, to assess how the noncongruent-mood reversal effect is related to mood beta. The estimates are reported under Columns (3) and (4) of Table 4. Mood beta premia remain positive and significant in all cases. But only two out of six coefficients on $RET^{\perp}_{\text{Low(High)}}$ and $RET^{\perp}_{\text{RLow(RHigh)}}$ remain negative and significant. After accounting for the mood beta, the noncongruent-mood reversal effects tend to weaken or disappear and even turn into a return recurrence effect.

This evidence suggests that mood beta explains a considerable portion of the noncongruent reversal effects at the weekday level. Since betas are estimated with error, it is possible that the true mood beta is the entire source of these effects.

4.3. Mood beta, market beta and sentiment beta

We next consider whether the mood beta effects that we find derive from traditional market beta. Under rational risk-based asset pricing theory, the market premium should be positive, which implies that market beta should be positively related to expected returns in predesignated months or days. This prediction, however, is contradicted by our estimates of negative premia on mood beta during Septembers, Octobers, and Mondays (see Panels A and B of Fig. 2, which will be discussed later in Section 5.3).

Nevertheless, to further address the possibility that the mood beta effects may derive in part from market beta, we perform tests that control for market beta in our regressions with mood beta, where market beta is estimated in a fashion analogous to the corresponding mood beta (using monthly or daily returns) except that all month or weekday asset returns and value-weighted market returns are used in the estimation.

Another possible concern is that mood beta may be a proxy for the sentiment beta (Baker and Wurgler 2007). The purpose of studying mood beta is to capture stock sensitivity to shifts in investor emotions—the affective aspect of investor sentiment. Investor emotions can potentially behave differently from other aspects of sentiment (such as, for example, swings in investor attention between different aspects of the economic environment). Thus we do not expect mood beta to be perfectly correlated with sentiment beta over time or across securities.

To verify whether mood beta has incremental explanatory power, we include sentiment beta in the regression, where sentiment beta is estimated using the most recent 60 (at least 36) monthly returns regressed on the monthly changes in the BW sentiment index (orthogonalized to macroeconomic variables) together with the CRSP value-weighted index returns. An alternative measure of the sentiment index that is the principal component of the monthly changes in five sentiment index components yields qualitatively similar results.

Our regressions are designed to forecast future asset returns during the hypothesized high and low mood periods (months or weekdays) using mood beta with controls for market beta or sentiment beta. Again, we focus on year lags 2–5 for month-level return regressions and week lags 2–10 for weekday return regressions. The estimates reported in Table 5 indicate that, across all three test assets and specifications, the mood premium remains significant in the presence of market beta or sentiment beta.

In contrast, during the hypothesized high and low mood periods, market beta tends to carry a significantly negative premium, contrary to the prediction of the rational risk-based theory. This phenomenon is referred to as the low-risk anomaly (Baker et al., 2011; Frazzini and Pedersen, 2014). We also find that sentiment beta tends to carry a positive coefficient, suggesting that high-sentiment

Table 5
Mood beta, market beta, and sentiment beta.

This table examines the predictive power of mood beta, market beta, and sentiment beta to forecast future mood month or weekday returns in Fama-MacBeth regressions. The mood betas ($\beta^{\text{Mood}}{}_{\text{Month}}$, $\beta^{\text{Mood}}{}_{\text{Weekday}}$) are estimated using monthly or weekday returns during the historical mood states, as defined as in Table 4. The market betas ($\beta^{\text{Mkt}}{}_{\text{Month}}$, $\beta^{\text{Mkt}}{}_{\text{Weekday}}$) are estimated by regressing all historical monthly or weekday asset returns on the corresponding returns on the value-weighted CRSP index during a rolling window. The sentiment beta (β^{Sent}) is estimated by regressing the monthly asset returns on the monthly changes in the Baker and Wurgler (2006) sentiment index together with the value-weighted CRSP index returns over a rolling 60-month window. When forecasting future returns during a high mood month or weekday, the independent variable is the asset's historical β^{Mood}. When forecasting future returns during a low mood month or weekday, it is $-\beta^{\text{Mood}}$. Regression estimates are reported in percentages. The reported Fama-MacBeth *t*-statistics are in parentheses and are corrected for heteroskedasticity and autocorrelation using Newey and West (1987). The symbols *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, based on two-tailed tests. All variables are defined in the Data Appendix.

Dep. var.	*Mood month return* RET$_{\text{High(Low)}}$				*Dep. var.*	*Mood weekday return* RET$_{\text{High(Low)}}$			
Indep. var.	(1)		(2)		*Indep.* var.	(3)		(4)	
Year lag (k)	$\pm\,\beta^{\text{Mood}}{}_{\text{Month}}$	$\beta^{\text{Mkt}}{}_{\text{Month}}$	$\pm\,\beta^{\text{Mood}}{}_{\text{Month}}$	β^{Sent}	*Week lag (k)*	$\pm\,\beta^{\text{Mood}}{}_{\text{Weekday}}$	$\beta^{\text{Mkt}}{}_{\text{Weekday}}$	$\pm\,\beta^{\text{Mood}}{}_{\text{Weekday}}$	β^{Sent}
Panel A: Individual stocks									
2~5	1.98***	−0.87***	1.52***	−0.01	*2~10*	0.07***	−0.09***	0.05***	0.005***
	(5.78)	(−3.99)	(4.86)	(−0.65)		(8.97)	(−7.69)	(6.08)	(4.83)
Panel B: Baker and Wurgler (BW) portfolios									
2~5	5.49***	−3.56***	2.61***	0.91***	*2~10*	0.42***	−0.39***	0.11***	0.05***
	(6.62)	(−5.06)	(5.09)	(4.41)		(13.14)	(−8.09)	(9.34)	(7.38)
Panel C: Keloharju, Linnainmaa, and Nyberg (KLN) portfolios									
2~5	6.00***	−3.28***	3.10***	0.55**	*2~10*	0.44***	−0.41***	0.11***	0.05***
	(7.65)	(−4.44)	(5.69)	(2.55)		(12.65)	(−7.40)	(9.04)	(7.55)

beta stocks tend to earn higher average returns in these periods. In conclusion, neither market beta nor sentiment beta subsumes the ability of mood beta to predict returns across high and low mood periods in the very specific way in which it does—with opposite signs.

4.4. Composite mood beta

So far, for each asset, we have two mood betas, estimated from monthly and weekday returns during the historical high and low mood periods. To further reduce noise, we form a composite mood beta (β_i^{Mood}) as the first principal component of the monthly mood beta ($\beta_{i,\ \text{month}}^{\text{Mood}}$) and the weekday mood beta ($\beta_{i,\ \text{Weekday}}^{\text{Mood}}$), extracted month by month in the cross-section of individual stocks or portfolios. The monthly mood beta $\beta_{i,\ \text{month}}^{\text{Mood}}$ is updated annually, so the within-year variation for a given stock comes solely from the variation in $\beta_{i,\ \text{Weekday}}^{\text{Mood}}$.

The composite mood beta has an average eigenvalue of 1.34, 1.60, and 1.47, respectively, for the three sets of test assets and, by construction, zero mean and unit standard deviation. The average weight is roughly equal across the two mood beta components in the composite mood beta. The evidence suggests that there is important commonality among the two mood betas that is picked up by the composite mood beta.

In Table 1 we report the summary statistics of the composite mood beta, and in Fig. 1 we plot the time-series average of the composite beta for each of the BW and KLN portfolios. Fig. 1 reveals that mood beta tends to be higher for younger firms than older firms, growth firms than value firms, nondividend payers than payers, small firms than larger firms, high R&D firms than low R&D firms, high volatility firms than low volatility firms, low dividend-yield firms than high dividend-yield firms, and low earnings-to-price firms than high earnings-to-price firms.

Some other attributes exhibit a V-shaped or inverse V-shaped relation with mood beta. For example, mood beta is higher for both extreme winners and extreme losers, firms with extremely high or extremely low return on equity, with the highest or lowest external financing, and with the fastest or the slowest sales growth. Mood beta is lower for firms with zero or extremely high tangible assets. Across industries, the highest mood beta is observed for the machinery and business equipment industry, and by far the lowest mood beta is seen for the utilities industry.

Many of these patterns for mood beta are similar to those associated with the BW sentiment beta. These patterns support the notion that hard-to-value firms and attention-grabbing firms are more heavily influenced by investor mood swings than easy-to-value or easy-to-neglect firms. In subsequent tests, we use the composite mood beta to assess the profitability of trading strategies as well as to conduct multivariate tests.

4.5. Long-short portfolios based on historical seasonal returns and mood beta

Our FMB regression results presented in the previous sections suggest that historical congruent-mood returns, noncongruent-mood returns, and mood beta all have the ability to forecast cross-sectional returns in future hypothesized high (January, March, and Friday) and low (September, October, and Monday) mood periods. To quantify the economic magnitude of these effects, we next examine the profitability of various trading strategies derived from these findings.

We form a long-short portfolio for each predictor based on portfolio deciles sorted each month according to the predictor. If the historical seasonal return is used as the predictor, the hedge portfolio always goes long the highest and short the lowest decile. If mood beta is used

Fig. 1. Mood betas of characteristics-sorted portfolios.
This figure reports the average composite mood beta (β^{Mood}) for portfolios sorted based on firm characteristics used by Baker and Wurgler (2006) (Panels A through J) and Keloharju et al. (2016) (Panels K through O excluding size and book-to-market portfolios, which are reported in Panels B and G). All variables are defined in the Data Appendix.

Table 6
Long-short portfolios based on historical seasonal returns and mood beta: month-level tests.

This table reports the mean and abnormal returns on the long-short portfolios sorted based on historical congruent, noncongruent-mood month returns. or mood betas. Each month, we sort stocks into deciles based on the average historical congruent ($RET_{High(Low)}$, $RET_{Bhigh(RLow)}$), noncongruent ($RET_{Low(High)}$, $RET_{RLow(Rhigh)}$) mood month returns, or mood betas ($\beta^{Mood}{}_{Month}$, $\beta^{Mood}{}_{Weekday}$, and β^{Mood}) during years $t-2$ through $t-5$ and calculate equal-weighted portfolio returns. The long-short portfolios based on historical returns go long the highest decile and short the lowest decile. The long-short portfolios based on mood beta go long the highest decile and short the lowest mood beta decile during the high mood months (January and March) and flip the long and short lags during the low mood months (September and October). The abnormal returns are estimated using the four-factor model (Fama-French-Carhart 1997) and the Fama-French (2015) five-factor model. Estimates are reported in percentage points. The Newey-West t-statistics are reported in parentheses. All variables are defined in the Data Appendix.

Sorting variables	$RET_{High(Low)}$ (1)	$RET_{RHigh(RLow)}$ (2)	$RET_{Low(High)}$ (3)	$RET_{RLow(RHigh)}$ (4)	$\beta^{Mood}{}_{Month}$ (5)	$\beta^{Mood}{}_{Weekday}$ (6)	β^{Mood} (7)
Panel A: Individiual stocks							
Mean return	0.26*	0.24	−1.48***	−1.74***	2.74***	1.15**	2.23***
	(1.78)	(0.78)	(−6.90)	(−5.36)	(5.08)	(2.58)	(3.95)
4-factor-adjusted	0.33*	0.46	−1.40***	−1.65**	2.69***	1.03**	2.15***
	(1.85)	(1.34)	(−5.75)	(−4.27)	(5.04)	(2.12)	(3.75)
5-factor-adjusted	0.33*	0.54	−1.40***	−1.74**	2.85***	1.22**	2.37***
	(1.76)	(1.55)	(−5.63)	(−4.44)	(5.23)	(2.51)	(4.07)
Panel B: Baker and Wurgler (BW) portfolios							
Mean Return	1.43***	1.50***	−0.77***	−1.39***	1.74***	1.00***	1.58***
	(5.68)	(4.78)	(−3.52)	(−4.49)	(4.92)	(3.21)	(4.38)
4-factor-adjusted	1.42***	1.57**	−0.74***	−1.53**	1.72***	0.99***	1.56***
	(5.23)	(4.34)	(−3.03)	(−4.29)	(4.62)	(3.09)	(4.13)
5-factor-adjusted	1.47***	1.65**	−0.77***	−1.62**	1.82***	1.08***	1.67***
	(5.06)	(4.47)	(−3.10)	(−4.44)	(4.79)	(3.30)	(4.32)
Panel C: Keloharju, Linnainmaa, and Nyberg (KLN) portfolios							
Mean return	1.67***	1.45***	−0.69***	−1.34***	1.87***	0.46	1.45***
	(6.27)	(4.36)	(−2.71)	(−4.11)	(5.44)	(1.54)	(4.12)
4-factor-adjusted	1.69***	1.57**	−0.80***	−1.41**	1.90***	0.52	1.49***
	(5.51)	(4.31)	(−3.48)	(−3.87)	(4.99)	(1.65)	(3.94)
5-factor-adjusted	1.80***	1.66**	−0.82***	−1.47**	1.99***	0.61*	1.59***
	(5.81)	(4.45)	(−3.49)	(−3.93)	(5.14)	(1.90)	(4.13)

instead, the hedge portfolio goes long the highest and short the lowest decile during the hypothesized high mood periods and flips the long and short legs when low moods are anticipated.

Table 6 reports results at the month level. We employ four strategies based on historical seasonal returns earned during the prespecified high and low and the realized high and low mood months and three mood beta strategies based on the monthly, the weekday, and the composite mood betas. The strategies are implemented by using the signals measured with annual lags 2–5. In addition to reporting mean returns, we report the estimated risk-adjusted returns (i.e., alphas) for these long-short portfolios based on the Fama-French-Carhart four-factor model (Carhart 1997) and the Fama and French (2015) five-factor model.

Across the three sets of test assets, Table 6 shows that the trading strategies that capture the congruent-mood recurrence effects (Columns (1) and (2)) work better for the BW and KLN portfolios, with a five-factor alpha ranging from 1.47% ($t=5.06$) to 1.80% ($t=5.81$) per month. Those based on the noncongruent-mood reversal effects (Columns (3) and (4)) work better for individual stocks, with a five-factor alpha of −1.40% ($t=-5.63$) and −1.74% ($t=-4.44$) per month.

The trading strategies based on three mood betas (Columns (5)-(7)) tend to be more profitable for individual stocks. The monthly five-factor alphas for the mood beta strategies across three sets of assets range from 0.61% ($t=1.90$) to 2.85% ($t=5.23$). The composite-mood-beta-based strategies (Column (7)) work well for all three sets of test assets, generating a monthly five-factor alpha ranging from 1.59% to 2.37%, all significant at the 5% level or better. It is particularly profitable for individuals stocks. This contrasts with the relatively weaker results based on the return recurrence effects in (1) and (2).

Next, in Table 7 we apply the trading strategies to forecasting future mood weekday returns (Mondays and Fridays) using predictors with weekly lags 2–10. Here we observe all positive alphas for strategies based upon the congruent-mood recurrence effect (Columns (1) and (2)); the alphas range from 3 to 12 basis points per day across the three assets, nearly all significant at the 5% level or better. The strategies based on the noncongruent-mood reversal effect (Columns (3) and (4)) are profitable for individual stocks (8 to 10 basis points per day) and for the BW portfolios (2 to 3 basis points per day), but are unprofitable for the KLN portfolios.

In contrast, mood-beta-based strategies (Columns (5)-(7)) are highly profitable throughout all measures, with alphas ranging from 8 to 17 basis points a day, all significant at the 5% level or better. Overall, mood beta implies more stable and profitable trading strategies across all three assets over the full sample period.[14]

[14] Fig. A1 in the Online Appendix indicates that the long-short profits exhibit diminishing returns since 2001, which may derive from the rising influence of sophisticated investors such as hedge funds.

Table 7
Long-short portfolios based on historical seasonal returns and mood beta: weekday-level tests.

This table reports the mean and abnormal returns on the long-short portfolios sorted based on historical congruent ($RET_{High(Low)}$, $RET_{Bhigh(RLow)}$), noncongruent ($RET_{Low(High)}$, $RET_{RLow(Rhigh)}$) mood weekday returns, or mood betas ($\beta^{Mood}{}_{Month}$, $\beta^{Mood}{}_{Weekday}$, and β^{Mood}). Each day of the week, we sort stocks into deciles based on the average historical congruent or noncongruent-mood weekday returns, or mood beta during weeks $t-2$ through t – 10, and calculate equal-weighted portfolio returns. The long-short portfolios based on historical returns go long the highest decile and short the lowest decile. The long-short portfolios based on mood beta are long the highest decile and short the lowest mood beta decile during the high mood weekday (Friday) and reverses the long and short lags during the low mood weekday (Monday). The abnormal returns are estimated using the four-factor model (Fama-French-Carhart 1997) and the Fama-French (2015) five-factor model. Estimates are reported in percentage points. The Newey-West t-statistics are reported in parentheses. All variables are defined in the Data Appendix.

Sorting variables	$RET_{High(Low)}$ (1)	$RET_{RHigh(RLow)}$ (2)	$RET_{Low(High)}$ (3)	$RET_{RLow(RHigh)}$ (4)	$\beta^{Mood}{}_{Month}$ (5)	$\beta^{Mood}{}_{Weekday}$ (6)	β^{Mood} (7)
Panel A: Individual stocks							
Mean return	0.07***	0.03**	−0.10***	−0.07***	0.17***	0.13***	0.17***
	(7.55)	(2.03)	(−11.32)	(−5.31)	(12.77)	(7.25)	(9.40)
4-factor-adjusted	0.06**	0.03*	−0.10***	−0.08***	0.17***	0.13**	0.17***
	(7.12)	(1.70)	(−13.06)	(−5.37)	(11.28)	(6.86)	(8.73)
5-factor-adjusted	0.07***	0.03**	−0.10***	−0.08***	0.17***	0.13***	0.17***
	(7.84)	(1.97)	(−12.11)	(−5.17)	(11.27)	(6.86)	(8.70)
Panel B: Baker and Wurgler (BW) portfolios							
Mean Return	0.09***	0.08***	−0.02***	−0.02***	0.13***	0.08***	0.13***
	(15.06)	(10.83)	(−3.04)	(−3.35)	(19.22)	(11.23)	(16.98)
4-Factor-Adjusted	0.09***	0.08***	−0.03***	−0.03***	0.13***	0.08***	0.12***
	(14.54)	(10.12)	(−4.68)	(−3.43)	(16.66)	(10.38)	(14.82)
5-Factor-Adjusted	0.09***	0.08***	−0.03***	−0.02***	0.13***	0.08***	0.12***
	(14.93)	(10.35)	(−4.23)	(−3.23)	(16.60)	(10.29)	(14.85)
Panel C: Keloharju, Linnainmaa, and Nyberg (KLN) portfolios							
Mean return	0.12***	0.09***	−0.00	−0.01	0.12***	0.08***	0.12***
	(15.15)	(9.71)	(−0.47)	(−1.10)	(15.78)	(8.43)	(13.78)
4-factor-adjusted	0.12**	0.09***	−0.01	−0.01	0.12***	0.08**	0.12***
	(15.87)	(9.03)	(−1.53)	(−1.32)	(13.36)	(7.73)	(12.03)
5-factor-adjusted	0.12**	0.09***	−0.01	−0.01	0.12***	0.08**	0.12***
	(16.33)	(9.28)	(−0.89)	(−1.18)	(13.37)	(7.73)	(12.16)

4.6. Multivariate Fama-MacBeth regressions

Last, we conduct multivariate tests using firm-level FMB regressions to assess the incremental predictive power of mood beta relative to a set of firm characteristics and return predictors—after accounting for the possible ability of these return predictors to forecast returns with opposing signs across the high and low mood periods. In Table 8, we report the estimates separately for the mood month and mood day regressions, in which lagged mood beta is used to forecast stock returns in future hypothesized high and low mood months (January, March, September, and October) or mood days (Monday and Friday).

We include regressions with only the composite mood beta (regressions 1 and 4), with market beta and sentiment beta (regressions 2 and 5), and additionally controlling for a set of firm characteristics used to form the BW and KLN portfolios (regressions 3 and 6). The characteristics include *ME, B/M, MOM, EF/A, GP*, PPE/A, *SG*, and *SIGMA*. We exclude several potential controls that are too closely related to the preceding ones: firm age, dividend-to-price, earnings-to-price, dividend-to-book equity, ROE, and R&D. Like the included controls, these measure or correlate with firm size, fundamental-to-price ratio, profitability, or asset tangibility.

To account for the opposite relation between mood beta and future returns in high versus low mood periods, we add a negative sign to the dependent variable realized in low mood months or weekdays so that mood beta has an expected positive coefficient for both mood states. This sign adjustment is equivalent to flipping signs of all independent variables, from market beta and sentiment beta to firm characteristics, between the anticipated high and low mood periods.

The regression estimates reported in Table 8 show that mood beta has a positive and significant coefficient throughout all regressions; the coefficient ranges from 0.59% to 1.80% per month and from 2.37 basis points to 4.65 basis points per day, all significant at the 5% level or better. Including all other forecasters (Regressions 3 and 6) roughly halves the size of the coefficient on mood beta.

In contrast, market beta and sentiment beta have mixed or insignificant coefficients, exhibiting no clear pattern. Among other firm characteristics, only size, momentum, and gross profitability exhibit a consistent and significant relation with future mood returns, which suggests that their relationship with future returns also tend to flip between high and low mood periods.

Overall, the evidence in the multivariate FMB regressions supports our prior finding that mood beta positively predicts stock returns when investors experience ascending moods and negatively do so when descending moods occur. The forecasting power is robust to controls for market beta, sentiment beta, and a host of firm characteristics.

5. Additional tests and robustness checks

This section presents additional tests and robustness checks to address several possible concerns about our main

Table 8
Fama-MacBeth regression at the firm level.

This table reports Fama-MacBeth regression estimates on lagged mood beta, market beta, sentiment beta, and a host of firm characteristics in forecasting future returns during the high and low mood month or weekdays. The test assets are the full cross-section of individual stocks. The dependent variable is the monthly return in percentagess or the weekday return in basis points. Columns (1), (2), and (3) report the estimates for the regressions that forecast the positive January and March (high mood months) returns and the negative September and October (low mood months) returns. Columns (4), (5), and (6) report the estimates for the regressions that forecast the positive high mood day (Friday) returns and the negative low mood day (Monday) returns. The Newey-West *t*-statistics are reported in parentheses. All independent variables are lagged. They are also standardized to have zero mean and unit variance and are defined in the Data Appendix.

Dep. var.	$+\ \mathrm{RET}_{\mathrm{High}}$ and $-\ \mathrm{RET}_{\mathrm{Low}}$					
	Mood month			Mood day		
Indep. var.	(1)	(2)	(3)	(4)	(5)	(6)
β^{Mood}	1.15***	1.80***	0.59**	4.65***	4.14***	2.37***
	(3.70)	(3.83)	(2.19)	(9.39)	(12.69)	(7.38)
$\beta^{Mkt}{}_{Weekday}$		−0.74*	−0.56		0.61	1.38**
		(−1.85)	(−0.80)		(0.98)	(2.39)
β^{Sent}		−0.03	0.13		0.58***	0.38**
		(−0.16)	(0.57)		(3.07)	(1.99)
Log(ME)			−1.01***			−3.69***
			(−5.66)			(−9.49)
Log(B/M)			0.05			−1.88***
			(0.21)			(−9.06)
MOM			−1.63***			−0.70**
			(−2.77)			(−2.44)
EF/A			0.03			0.25
			(0.15)			(1.24)
GP			−0.47**			−1.59***
			(−2.04)			(−8.53)
PPE/A			−0.08			0.42**
			(−0.61)			(2.13)
SG			−0.24			0.07
			(−0.81)			(0.37)
SIGMA			0.30			2.43***
			(0.87)			(8.02)
# of mons/days	196	188	188	4816	4621	4621
Avg.# of stocks	2577	2814	2577	2639	2780	2555
Adj. R2	0.25%	0.36%	1.23%	0.78%	1.62%	2.61%

findings of the mood recurrence, reversal, and mood beta effects.

5.1. Alternative definitions of hypothesized mood periods

A possible concern is that our results may be driven by a look-ahead or in-sample bias. Our hypothesized high and low mood periods are identified by the average highest and lowest aggregate returns in the full sample period. Thus market beta will be strongly positively related to returns in the high market return periods and will be negatively related to returns during the periods when realized market returns are, on average, negative. If mood beta is correlated with market beta, then such patterns may apply to mood beta as well.

The evidence opposes this explanation. We show in Table 8 (discussed in Section 4.6) and Fig. 2 (further discussed in Section 5.3) that market beta does not predict returns with opposite signs in the hypothesized high and low mood periods. In contrast, the mood beta does. Thus market beta does not capture this key mood beta effect.[15]

We also raise a conceptual objection to this explanation and present a set of robustness checks. Conceptually, rational factor pricing models imply that the ex ante market risk premium is positive on all months or days. In long enough samples, systematic patterns of negative daily or monthly premia for market beta would be ruled out. Suppose that our mood beta is a proxy for the true market beta, then there should typically be a positive premium for mood beta. This is dissonant with our finding that in some seasonal periods there is a negative premium associated with mood beta. At best, such a finding would be limited to specific subsamples (especially when the high and low mood periods are selected by also using information from the periods being predicted). This contrasts with our mood seasonality hypothesis that predicts such effects will systematically occur in sample as mood varies predictably across months and weekdays.

We also conduct a set of robustness checks in Tables 9 and 10 by changing the predictable future mood periods from the originally predesignated January, March, September, October, Monday, and Friday to others based on alternative identifications. The purpose of these tests is to ascertain whether our findings are limited to specific periods.

We consider four alternative identifications of mood periods based on the extreme, realized monthly or weekday equal-weighted market excess returns observed in four windows, which include (1) from 1927 to the preceding year before the forecast is made (*Expanding window*); (2) over the rolling 50-year (10-year) window for monthly (weekday) returns ending in the prior year (month) (*Rolling window*); (3) during even years when we forecast returns in odd years (*Even years to forecast odd years*); and (4) during odd years when we forecast returns in even years (*Odd years to forecast even years*). The commonality in the four tests is that the future mood months or weekdays are identified using only historical or split-sample data that exclude the return information of the mood months or weekday being forecasted.

At the month level, Table 9 reports the long-short portfolio returns based on historical mood month returns. The results indicate strong congruent-mood recurrence effects for portfolios and strong noncongruent-mood reversal effects for individual stocks, a pattern generally similar to our baseline results in Table 7. Moving to weekday-level tests in Table 10, the congruent-mood recurrence effects generally remain strong across three sets of tests assets. The noncongruent-mood reversal effects are strong for individual stocks and the BW portfolios but weak for the KLN portfolios, again similar to the baseline results in Table 7.

More importantly, the long-short returns of the composite mood beta portfolios, reported in Tables 9 and 10, Column (5), exhibit similar patterns to our baseline tests.

[15] So this mood beta effect could only be driven by market beta if one believes that mood beta is a more accurate proxy for the true market beta than market beta directly estimated from recent monthly or daily returns. It is not clear why this would be the case.

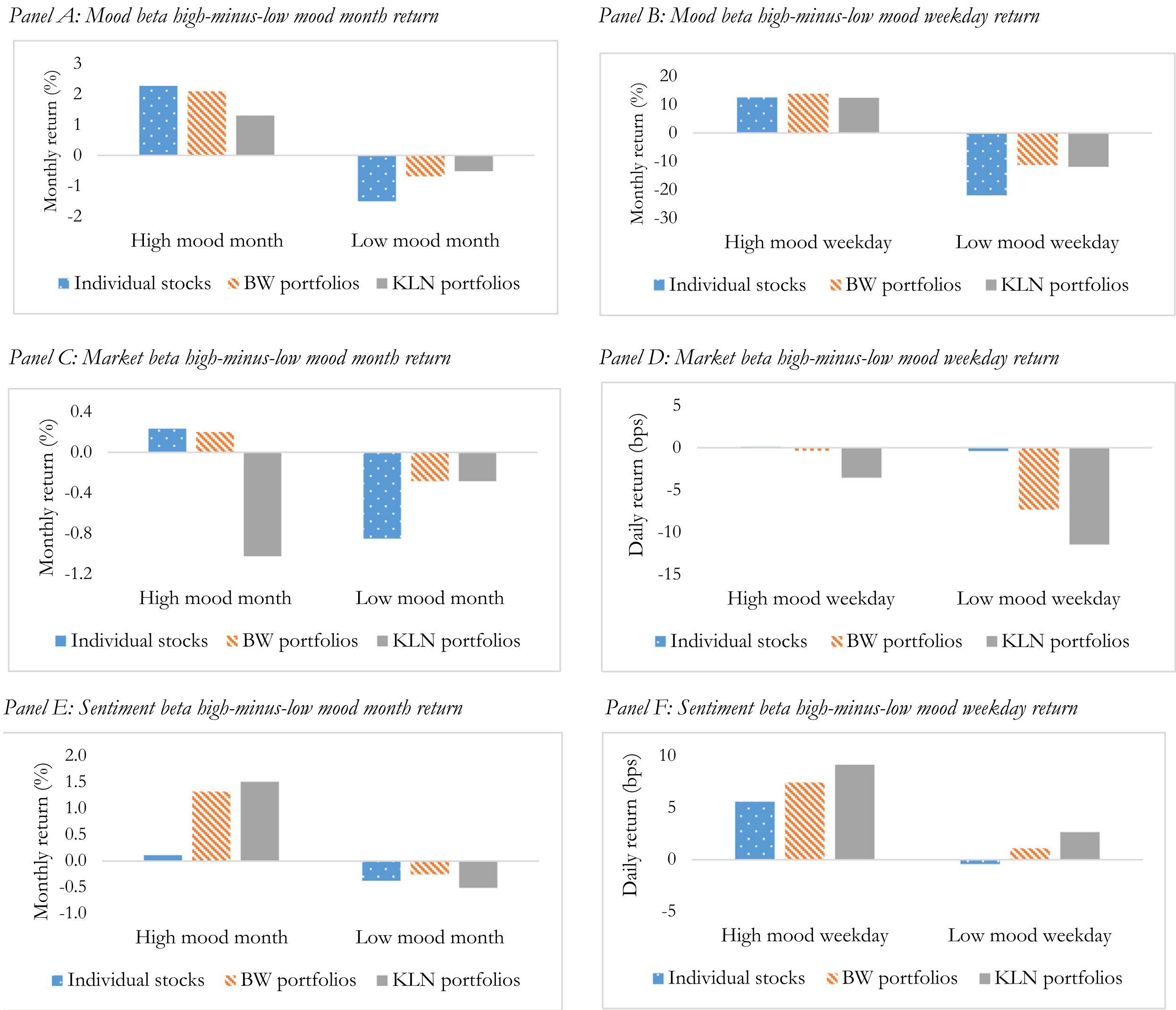


Fig. 2. High-minus-low portfolios across high and low mood periods: mood beta versus market beta and sentiment beta.
This figure reports the average high-minus-low portfolio returns sorted based on the composite mood beta (β^{Mood}) across high and low mood states. The high-minus-low portfolio is long the top decile of assets with the highest mood beta and short the bottom decile of assets with the lowest mood beta. High mood months include January and March and high mood weekday includes Friday. Low mood months include September and October and low mood weekday includes Monday. The three sets of test assets include the full cross-section of individual stocks, the 94 BW portfolios and the 79 KLN portfolios. The average returns of the long-short portfolios are expressed in percentage points in left-hand-side panels for month-level tests and in basis points in right-hand-side panels for weekday-level tests. The composite mood beta is defined in the Data Appendix.

Across all approaches of identifying future mood periods and for all three sets of test assets, mood-beta-based long-short portfolios earn positive and significant average returns in 8 out of 12 cases, and marginally significant average returns in 2 out of 12 cases, leaving only two cases insignificant.

Overall, the long-short strategies produce an average return of about 1% per month or above 10 basis points per day, a magnitude that is slightly smaller at the monthly level, but similar at the weekday level, to the baseline strategies based on predesignated mood months or weekdays. Thus, using the alternative definitions of future predictable mood periods leads to qualitatively similar, and quantitatively comparable results. Therefore our findings of the congruent-mood recurrence and noncongruent-mood reversal effects and the predictive power of mood beta are unlikely to be driven by a look-ahead or in-sample bias.

5.2. Is January atypical?

Heston and Sadka (2008) show that January is associated with a stronger same-month return persistence effect. We extend their analysis by studying the effects of mood recurrence, reversal, and mood beta separately for January and non-January months.

At the monthly level, Table A1 of the Online Appendix reports the FMB regression coefficients. The results show that, indeed, both month-level mood effects are visibly stronger during January months. Although some of the

Table 9
Long-short portfolio returns: alternative definition of mood months.

This table reports the mean returns on the long-short portfolios earned in future mood months defined using various alternative methods, and the long-short portfolios are sorted on the average historical congruent ($RET_{High(Low)}$, $RET_{Bhigh(RLow)}$), noncongruent ($RET_{Low(High)}$, $RET_{RLow(Rhigh)}$) mood month returns, and the composite mood betas (β^{Mood}) during years $t-2$ through $t-5$. The future high (low) mood months are identified using the highest (lowest) mean equal-weighted market excess monthly returns earned (1) from 1927 to the most recent year (Expanding window), (2) over the most recent 50 years (Rolling window), (3) during even years to forecast only odd years from 1963 to 2016 (Even years to forecast odd years); and (4) during odd years to forecast only even years from 1963 to 2016 (Odd years to forecast even years). Regression estimates are reported in percentages. The Newey-West t-statistics are reported in parentheses. All variables are defined in the Data Appendix.

	Dependent variable: month-level $RET_{High(Low)}$				
Sorting variable	$RET_{High(Low)}$ (1)	$RET_{RHigh(RLow)}$ (2)	$RET_{Low(High)}$ (3)	$RET_{RLow(RHigh)}$ (4)	β^{Mood} (5)
Panel A: Individual stocks					
Expanding window	0.32**	0.21	−1.37***	−1.43***	1.82***
	(2.30)	(0.71)	(−6.92)	(−4.62)	(3.33)
Rolling window	−0.29	−0.00	−1.32***	−1.58***	1.20**
	(−1.63)	(−0.01)	(−6.67)	(−4.84)	(2.26)
Even years to forecast odd years	−0.08	−0.55	−1.50***	−1.40***	0.77
	(−0.37)	(−1.06)	(−4.66)	(−2.63)	(1.01)
Odd years to forecast even years	0.29	0.12	−1.33***	−1.34***	1.84**
	(1.54)	(0.28)	(−4.83)	(−2.95)	(2.28)
Panel B: Baker and Wurgler (BW) portfolios					
Expanding window	1.27***	1.31***	−0.67***	−1.13***	1.32***
	(5.14)	(4.27)	(−3.30)	(−3.75)	(3.79)
Rolling window	0.90***	0.86**	−0.13	−0.62*	0.90**
	(3.94)	(2.58)	(−0.53)	(−1.82)	(2.36)
Even years to forecast odd years	1.11***	0.94**	−0.20	−0.63	0.88*
	(3.23)	(2.19)	(−0.61)	(−1.56)	(1.82)
Odd years to forecast even years	1.21***	1.37***	−0.87***	−1.32***	1.26***
	(3.23)	(3.53)	(−2.64)	(−3.12)	(2.71)
Panel C: Keloharju, Linnainmaa, and Nyberg (KLN) Portfolios					
Expanding window	1.49***	1.23***	−0.64**	−1.06***	1.19***
	(5.30)	(3.74)	(−2.58)	(−3.19)	(3.37)
Rolling window	1.38***	0.78**	0.11	−0.63*	0.72*
	(5.45)	(2.35)	(0.43)	(−1.77)	(1.81)
Even years to forecast odd years	1.62***	0.84*	−0.05	−0.73*	0.78
	(4.40)	(1.76)	(−0.12)	(−1.66)	(1.53)
Odd years to forecast even years	1.12***	1.31***	−0.89***	−1.33***	1.17***
	(2.63)	(3.18)	(−2.74)	(−3.13)	(2.72)

historical seasonal returns lose predictive power in non-January mood months, mood beta retains its significant power in the forecasts for all three test assets.

Moving to Table A2 for the weekday-level tests, however, we find that congruent-mood recurrence and noncongruent-mood reversal effects are slightly stronger in non-January months and so is the forecast power of mood beta; the coefficient of mood beta in non-January months is several times that in January. Taken together, the mood effects are especially strong in January for monthly-level tests but not so for weekday-level tests.

5.3. Mood beta in high and low mood periods

Does mood beta predict future returns in the hypothesized high or low mood periods? We find that the predictability comes from both periods. As plotted in Panel A of Fig. 2, the three high-minus-low portfolios based on the composite mood beta yield positive average returns during high mood months (January and March) and weekdays (Friday) and negative average returns during low mood months (September and October) and weekdays (Monday). This is in contrast to the prediction of rational risk theory that higher loadings on fundamental risk factors should consistently receive risk premia of the same sign.

As a comparison, we plot in Panels B and C of Fig. 2 the high-minus-low portfolio returns based on the market beta and sentiment beta, again, separately for the hypothesized high and low mood months and weekdays. Panel B shows that high market beta assets tend to underperform during low mood periods but do not consistently overperform during high mood periods. Panel C, however, shows that while high sentiment beta assets do overperform during high mood periods, they do not consistently underperform during low mood periods. Thus, the effects of mood beta are far from fully captured by either market beta or the sentiment beta. Mood beta captures the distinctive opposite-sign relation with average returns across high and low mood periods.

5.4. Mood beta effect across time periods

Is the mood beta effect persistent across time? If there is mood-induced mispricing, mood-insensitive investors have an incentive to arbitrage it away, which implies that the effect should have weakened over time with the rising assets under management of hedge funds and active institutional investors in general.

In Fig. A1 of the Online Appendix, we plot the annual return of the mood beta strategies (that flips the long

Table 10
Long-short portfolio returns: alternative definition of mood weekdays.

This table reports the mean returns on the long-short portfolios earned in future mood weekdays defined using various alternative methods, and the long-short portfolios are sorted on the average historical congruent ($RET_{High(Low)}$, $RET_{Bhigh(RLow)}$), noncongruent ($RET_{Low(High)}$, $RET_{RLow(Rhigh)}$) mood weekday returns, and the composite mood betas (β^{Mood}) during weeks $t-2$ through $t-10$. The high (low) mood weekdays are identified using the highest (lowest) mean equal-weighted market weekday excess returns (1) from 1927 to the most recent year (Expanding window), (2) over the most recent ten years (Rolling window), (3) during even years to forecast only odd years from 1963 to 2016 (Even years to forecast odd years); and (4) during odd years to forecast only even years from 1963 to 2016 (Odd years to forecast even years). Regression estimates are reported in percentages. . The Newey-West t-statistics are reported in parentheses. All variables are defined in the Data Appendix.

	Dependent variable: weekday-level $RET_{High(Low)}$				
Sorting variable	$RET_{High(Low)}$ (1)	$RET_{RHigh(RLow)}$ (2)	$RET_{Low(High)}$ (3)	$RET_{RLow(RHigh)}$ (4)	β^{Mood} (5)
Panel A: Individual stocks					
Expanding window	0.06***	0.04***	−0.09***	−0.07***	0.16***
	(7.10)	(2.84)	(−10.29)	(−5.16)	(8.64)
Rolling window	0.06***	0.04***	−0.08***	−0.08***	0.16***
	(6.79)	(2.83)	(−9.11)	(−5.35)	(8.23)
Even years to forecast odd years	0.08***	0.01	−0.06***	−0.13***	0.16***
	(7.11)	(0.51)	(−5.94)	(−6.41)	(6.33)
Odd years to forecast even years	0.06***	0.00	−0.12***	−0.17***	0.20***
	(4.46)	(0.04)	(−8.92)	(−7.72)	(6.99)
Panel B: Baker and Wurgler (BW) portfolios					
Expanding window	0.08***	0.07***	−0.02***	−0.02***	0.12***
	(13.39)	(10.01)	(−2.70)	(−3.21)	(15.22)
Rolling window	0.08***	0.07***	−0.01	−0.02***	0.11***
	(13.02)	(9.93)	(−0.99)	(−2.85)	(14.08)
Even years to forecast odd years	0.10***	0.08***	−0.00	−0.02**	0.12***
	(13.19)	(8.50)	(−0.18)	(−2.54)	(12.35)
Odd years to forecast even years	0.09***	0.08***	−0.02**	0.08***	0.13***
	(8.81)	(7.59)	(−2.36)	(7.59)	(11.61)
Panel C: Keloharju, Linnainmaa, and Nyberg (KLN) portfolios					
Expanding window	0.12***	0.09***	0.00	−0.01	0.11***
	(14.15)	(9.45)	(0.05)	(−0.86)	(12.72)
Rolling window	0.11***	0.09***	0.01*	−0.00	0.10***
	(13.81)	(9.34)	(1.70)	(−0.37)	(11.60)
Even years to forecast odd years	0.13***	0.09***	0.02**	0.00	0.10***
	(13.50)	(6.81)	(2.14)	(0.35)	(8.88)
Odd years to forecast even years	0.11***	0.11***	−0.01	−0.02	0.14***
	(8.73)	(6.92)	(−1.00)	(−1.25)	(10.12)

and short legs across high versus low mood periods) from 1967 to 2016, separately for the three sets of test assets in three panels. The plots show strong performance of these strategies since the early 1970s through early 2000s, which peaked in 2000 and became less positive and more volatile since then.

We also separately report the mean long-short portfolio returns across three time periods (before 1980, from 1981 to 2000, and after 2000) in Fig. A1. The evidence suggests a diminishing profitability associated with the strategies since 2001. This is not surprising given the corresponding rise of sophisticated, active investors, who may have eliminated much mispricing related to calendar time seasonality effects. These strategies continue to yield an average positive return in recent years, though the small magnitude may be unattractive after accounting for transaction costs.

It is worth noting that these strategies are implemented only in four months a year or two days a week, so they effectively achieve zero gains during other periods. Thus in a full year of trading, performance is heavily diluted by the inactive periods. By the same token, the strategy has a hidden virtue: during inactive periods, the strategy does not exhaust the investor's capital constraints/risk-bearing capacity, so the investor can potentially redeploy capital to earn alpha on other strategies. This hidden benefit should be kept in mind in considering the strategy's annual return performance.

5.5. Mood beta effect after portfolio formation

How much mispricing does mood generate, and how long does it take to correct the mispricing? We consider either a monthly or daily frequency. To answer this question, we track the total seasonal mood-induced mispricing by calculating the cumulative returns on a long-short portfolio formed based upon mood beta deciles during the mood month over the next 12 months or the weekday over the next 5 trading days, during which we do not rebalance the long-short portfolios. Thus a cumulative return of zero at some point indicates a complete correction of mispricing generated starting at portfolio formation during the pre-specified high or low mood periods.

Fig. A2 of the Online Appendix plots these cumulative long-short returns for the three sets of test assets (individual stocks, the BW portfolios, and the KLN portfolios) and separately for the high versus low mood conditioning periods. We focus on the magnitude of the initial mispricing and the number of months or days it takes the mispricing to correct to zero during a 12-month or 5-trading-day period.

We cannot be sure that pricing is correct at any given point in the season, so our focus is on the cumulative increment mispricing after a given starting point. In the left figure of Panel A (plotting the month-level patterns for individual stocks), we find that the average initial long-short return in the high mood months (January and March) is 4.83%. If we interpret the starting point as correct pricing, this suggests that during January and March, high mood beta stocks, on average, become incrementally overpriced by nearly 5% relative to low mood beta stocks. This overpricing takes about nine months to correct. (The long-short return next turns slightly negative when low mood months ensue and then gradually rises until the end of the 12-month period.) As for the low mood months (September and October), the long-short portfolio earns an average of −1.06% initially. If we interpret the starting point as correct pricing, this indicates that the low-mood months induce incremental underpricing of about 1%. This takes about three months to correct to zero. (The long-short return then moves up to a positive level about 5% after high mood months arrive).

The right figure of Panel A plots the cumulative long-short returns across the five-trading-day cycle for individual stocks. Here the initial average positive long-short return (5.67 basis points) earned on Fridays quickly turns to negative during the next trading day (usually Monday) and stays negative for four days before the next cycle starts. The initial average negative return earned on Mondays (−21.34 basis points) worsens the next day before reversing gradually in the next three days. Neither long-short return recovers back to zero during the five trading days. Our untabulated tests show that these cumulative negative returns do not recover for an extended period of a few weeks, suggesting a negative premium attached to high mood beta, at least at the weekday level for individual stocks.

We observe somewhat similar patterns when moving to the two sets of portfolios in Panels B and C. The 12-month or 5-trading-day period may end in positive or negative zones, depending on the frequency or assets we examine. The overall picture is that mood-induced mispricing created during the high or low mood period is corrected over the subsequent several months or days, and correction to the initial mispricing is concentrated in the periods when noncongruent moods arrive. Although the effects for January and Monday are large, several of the effects that we show would be hard for investors to exploit at a large scale given transaction costs and the required rebalancing frequency, which may help explain why they have persisted.

6. Conclusion

We propose and test a mood seasonality hypothesis, which asserts that seasonal variations in investor mood are in part responsible for both aggregate and cross-sectional return seasonalities. Consistent with this hypothesis, we find a variety of strong, novel cross-sectional mood recurrence and reversal effects across calendar months and weekdays. Assets that outperform in the past periods when investors are in ascending moods tend to outperform in future periods when an ascending mood is expected and to underperform in future periods when a descending mood is expected.

Our empirical results also highlight the role of mood beta, which measures a security's return sensitivity to market-wide mood-induced mispricing, in integrating various mood recurrence and reversal effects. Across the board, we observe that high mood beta stocks outperform during future ascending mood periods and underperform during future descending mood period. The predictive power of mood beta is incremental to market beta, sentiment beta, and a host of firm characteristics.

It is unclear how to reconcile our findings with a rational risk-based story in which predictable, seasonal cross-sectional return reversals require either seasonal, negative risk premiums or seasonal reversals in the cross-section of market betas or factor loadings. This does not seem very plausible, especially at the daily frequency. The evidence in KLN provides the insight that both aggregate and cross-sectional return seasonalities are manifestations of seasonal shifts in factor premia—though not necessarily rational risk premia. Our evidence points to one source of such seasonal factor return predictability: that seasonal factor mispricing is induced by seasonal variations in mood.

Data Appendix: Variable definition

A.1. Returns and betas of test assets

Variables	Definitions
RET_{High}	Monthly or weekday return during the high mood months or weekdays identified by high full sample average equal-weighted market excess returns. The high mood state refers to January and March at the month level and Friday at the weekday level. When used as an independent variable or sorting variable at the month level tests, it is the average return during January and March in a given year. Reported in percentages.
RET_{Low}	Monthly or weekday return during the low mood months or weekdays identified by low full sample average equal-weighted market excess returns. The low mood state refers to September and October at the month level and Monday at the weekday level. When used as an independent or sorting variable at the month level tests, it is the average return during September and October in a given year. Reported in percentages.
RET_{RHigh}	Monthly or weekday return during the high mood months or weekdays identified by realized high equal-weighted market excess returns. The realized high mood state refers to the two months in a year or the one day in a week with the highest equal-weighted CRSP excess return. When used as an independent or sorting variable at the month level tests, it is the average return during the two realized high mood months in a given year. Reported in percentages.
RET_{RLow}	Monthly or weekday return during the low mood months or weekdays identified by realized low equal-weighted market excess returns. The realized low mood state refers to the two months in a year or the one day in a week with the lowest equal-weighted CRSP excess return. When used as an independent or sorting variable at the month level tests, it is the average return during the two realized low mood months in a given year. Reported in percentages.
${\beta^{Mood}}_{Month}$	Monthly-return-estimated mood beta, or monthly mood beta, estimated by regressing an asset's excess returns during the four high and four low mood months (as defined for the mood state returns) on the equal-weighted CRSP excess return over a ten-year rolling window ending in the prior year, updated annually. A minimum of 40 observations are required.
${\beta^{Mood}}_{Weekday}$	Weekday-return-estimated mood beta, or weekday mood beta, estimated by regressing an asset's excess returns during the two high and two low mood weekdays (as defined for the mood state returns) on the equal-weighted CRSP excess return over a six-month rolling window ending in the prior month, updated monthly. A minimum of 50 observations are required.
β^{Mood}	Composite mood beta, defined as the principal component of ${\beta^{Mood}}_{Month}$ and ${\beta^{Mood}}_{Weekday}$, extracted monthly. It is normalized to have zero mean and unit standard deviation.
${\beta^{MKT}}_{Month}$	Monthly-return-estimated market beta, estimated from a market model using monthly returns over a ten-year rolling window ending in the prior year, updated annually. The market portfolio is proxied by the value-weighted CRSP index.
${\beta^{MKT}}_{Weekday}$	Weekday-return-estimated market beta, estimated from a market model using daily returns over a six-month rolling window ending in the prior month, updated monthly. The market portfolio is proxied by the value-weighted CRSP index.
β^{SENT}	Sentiment beta, estimated from regressions of monthly returns (in percentages) over a 60-month rolling window (requiring at least 36 monthly observations) on the monthly changes in the Baker and Wurgler (2006) orthogonalized sentiment index, controlling for the CRSP value-weighted returns, updated monthly.

A.2. Firm characteristics

A.2.1. Baker and Wurgler (BW) Portfolios

Variables	Definitions
AGE	Firm age as measured by the number of months since the firm's first appearance on CRSP, measured as of the most recent month.
B/M	Book-to-market equity. We define book equity (BE) as stockholders' equity, plus balance sheet deferred taxes (TXDB) and investment tax credit (ITCB), plus postretirement benefit liabilities (PRBA), minus the book value of preference stocks. Set TXDB, ITCB, or PRBA to zero if unavailable. Depending on availability, in order of preference, we use redemption (PSTKRV), liquidation (PSTKL), carrying value (PSTK), or zero if none is available. Stockholders' equity is measured as the book value of shareholder equity (SEQ). If SEQ is missing, we use the book value of common equity (CEQ) plus the book value of preferred stock. If CEQ is not available, we use the book value of assets (AT) minus total liabilities (LT). To compute B/M, we match BE for the fiscal year ending in calendar year $t-1$ with the firm's market equity at the end of December of year $t-1$ and then match this B/M to returns from July of year t through June of year $t+1$. This variable is winsorized annually at the 0.5% and 99.5% levels.
D/BE	Dividends to equity, defined as dividends per share at the ex date (DVPSX_F) of fiscal year end times Compustat shares outstanding (CSHO) dividend by book equity. Zero dividend firms are included in a separate portfolio from the deciles. We match D/BE for the fiscal year ending in calendar year $t-1$ to returns from July of year t through June of year $t+1$. This variable is winsorized annually at the 0.5% and 99.5% levels.
EF/A	External finance, defined as the change in total assets (AT) minus the change in retained earnings (RE) divided by assets (AT). If retained earnings is missing, it is replaced by net income (NI) minus common stock dividends (DVC). We match EF/A in June of year t to returns from July of year t through June of year $t+1$.This variable is winsorized annually at 0.5% and 99.5% levels.
ME	Market equity, measured by price (PRC) times shares outstanding (SHROUT) from the end of the latest June. We match ME in June of year t to returns from July of year t through June of year $t+1$.

(continued on next page)

(*continued*)

Variables	Definitions
SG	Sales growth, defined as the change in net sales (SALE) divided by prior-year net sales. We match SG for the fiscal year ending in calendar year $t-1$ to returns from July of year t through June of year $t+1$. This variable is winsorized annually at the 0.5% and 99.5% levels.
PPE/A	Tangible assets, defined as property, plant, and equipment (PPEGT) over assets (AT). Zero PPEGT firms are included in a separate portfolio from the deciles. We match PPE/A for the fiscal year ending in calendar year $t-1$ to returns from July of year t through June of year $t+1$. This variable is winsorized annually at the 0.5% and 99.5% levels.
R&D/A	Research and development expense (XRD) over assets (AT). We do not consider this variable prior to 1972, following Baker and Wurgler (2006). Zero XRD firms are included in a separate portfolio from the deciles. We match R&D/A for the fiscal year ending in calendar year $t-1$ to returns from July of year t through June of year $t+1$. This variable is winsorized annually at the 0.5% and 99.5% levels.
ROE	Return on equity, defined as earnings dividend by book equity. Earning is income before extraordinary items (IB) plus income statement deferred taxes (TXDB) minus preferred dividends (DVP). Book equity (BE) is as defined as for B/M. ROE is set to zero if earning is negative. Zero ROE firms are included in a separate portfolio from the deciles. We match D/BE for the fiscal year ending in calendar year $t-1$ to returns from July of year t through June of year $t+1$. This variable is winsorized annually at the 0.5% and 99.5% levels.
SIGMA	Return volatility, measured by the standard deviation of monthly returns over the 12 months ending in June. We match SIGMA measured as of June of year t to monthly returns from July of year t through June of year $t+1$.

A.2.2. Keloharju, Linnainmaa, and Nyberg (KLN) Portfolios

Variables	Definitions
ME	As defined in Appendix 2.1.
B/M	As defined in Appendix 2.1.
MOM	Price momentum measured by the cumulative return from month $t-12$ through $t-2$, matched to return in month t.
GP	Gross profitability, defined as annual revenues (REVT) minus cost of goods sold (COGS), divided by book equity (BE) for the last fiscal year end in $t-1$, where BE is as defined in Appendix 2.1 for B/M. We match GP for the fiscal year ending in calendar year $t-1$ to returns from July of year t through June of year $t+1$. This variable is winsorized annually at the 0.5% and 99.5% levels.
D/P	Dividend yield, defined as ex-date dividends per share (DVPSX_F) scaled by ex-date price per share (PRCC_F) at the fiscal year end. Zero dividend firms are included in a separate portfolio from the deciles. We match D/P for the fiscal year ending in calendar year $t-1$ to returns from July of year t through June of year $t+1$. This variable is winsorized annually at the 0.5% and 99.5% levels.
E/P	Earnings yield, defined as earnings per share including extraordinary items (EPSFI) scaled by price per share (PRCC_F) at the fiscal year end. Zero earnings firms are included in a separate portfolio from the deciles. We match E/P for the fiscal year ending in calendar year t – 1 to returns from July of year t through June of year $t+1$. This variable is winsorized annually at the 0.5% and 99.5% levels.
Industries	We use the Fama-French 17 industry portfolios formed at the end of June each year based on its four-digit SIC code at that time. The industries include food, mines (mining and minerals), oil (oil and petroleum products), clothes (textiles, apparel and footware), consumer durables, chemicals, consumer goods (drugs, soap, perfumes, and tobacco), construction (construction and construction materials), steel (steel works etc.), fabricated products, machine (machinery and business equipment), cars (automobiles), transportation, utilities, retail stores, financial (banks, insurance companies, and other financials), and other.

References

Baker, M., Bradley, B., Wurgler, J., 2011. Benchmarks as limits to arbitrage: understanding the low-volatility anomaly. Financial Anal. J. 67, 1–15.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. J. Finance 61, 1645–1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. J. Econ. Perspect. 21, 129–152.

Barberis, N., Greenwood, R., Jin, L., Shleifer, A., 2015. X-CAPM: an extrapolative capital asset pricing model. J. Financ. Econ. 115, 1–24.

Barberis, N., Greenwood, R., Jin, L., Shleifer, A., 2018. Extrapolation and bubbles. J financ econ 129, 203–227.

Bassi, A., Colacito, R., Fulghieri, P., 2013. 'O sole mio: an experimental analysis of weather and risk attitudes in financial decisions. Rev. Financ. Stud. 26, 1824–1852.

Bergsma, K., Jiang, D., 2016. Cultural new year holidays and stock returns around the world. Financ. Manage 45, 3–35.

Birru, J., 2018. Day of the week and the cross-section of returns. J. Financ. Econ. 130, 182–214.

Bouman, S., Jacobsen, B., 2002. The halloween indicator, "sell in may and go away"; another puzzle. Am. Econ. Rev. 92, 1618–1635.

Breaban, A., Noussair, C., 2017. Emotional state and market behavior. Rev. Financ. 22, 279–309.

Carhart, M.M., 1997. On persistence in mutual fund performance. J. Finance 52, 57–82.

Chhaochharia, V., Kim, D., Korniotis, G., Kumar, A., 2019. Mood, firm behavior, and aggregate economic outcomes. J. Financ. Econ. 132, 427–450.

Daniel, K.D., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under and over-reactions. J. Finance 53, 1839–1886.

Daniel, K.D., Hirshleifer, D., Subrahmanyam, A., 2001. Overconfidence, arbitrage, and equilibrium asset pricing. J. Finance 56, 921–965.

deHaan, E., Madsen, J., Piotroski, J., 2017. Do weather-induced moods affect the processing of earnings news? Journal of Accounting Research 55, 509–550.

DellaVigna, S., Pollet, J., 2009. Investor inattention and Friday earnings announcements. J. Finance 64, 709–749.

Doran, J.S., Jiang, D., Peterson, D.R., 2012. Gambling preference and the new year effect of assets with lottery features. Rev. Financ. 16, 685–731.

Edmans, A., García, D., Norli, Ø, 2007. Sports sentiment and stock returns. J. Finance 62, 1967–1998.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116, 1–22.

Frazzini, A., Pedersen, L.H., 2014. Betting against beta. J. Financ. Econ. 111, 1–25.

French, K., 1980. Stock returns and the weekend effect. J. Financ. Econ. 8, 55–69.

Goetzmann, W.N., Kim, D., Kumar, A., Wang, Q., 2015. Weather-induced mood, institutional investors, and stock returns. Rev. Financ. Stud. 28, 73–111.

Hartzmark, S.M., Solomon, D.H., 2018. Recurring firm events and predictable returns: the within-firm time series. Ann. Rev. Financ. Econ. 10, 499–517.

Helliwell, J.F., Wang, S., 2014. Weekends and subjective well-being. Soc. Indic Res. 116, 389–407.

Heston, S.L., Sadka, R., 2008. Seasonality in the cross-section of stock returns. J. Financ. Econ. 87, 418–445.

Heston, S.L., Sadka, R., 2010. Seasonality in the cross-section of stock returns: the international evidence. J. Financ. Quant. Anal. 45, 1133–1160.

Hirshleifer, D., Jiang, D., Meng, Y., Peterson, D.R., 2016. 'Tis the season! Pre-holiday cross-Sectional Return Seasonalities. University of California, Irvine Unpublished working paper.
Hirshleifer, D., Shumway, T., 2003. Good day sunshine: stock returns and the weather. J. Finance 58, 1009–1032.
Hirshleifer, D., Teoh, S., 2003. Limited attention, information disclosure, and financial reporting. J. Account. Econ. 36, 337–386.
Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. J. Finance 45, 881–898.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. J. Finance 48, 65–91.
Jiang, D., Norris, D., Sun, L., 2019. Weather, Institutional Investors, and Earnings News. Stony Brook University Unpublished working paper.
Kamstra, M.J., Kramer, L.A., Levi, M.D., 2003. Winter blues: a sad stock market cycle. Am. Econ. Rev. 93, 324–343.
Kamstra, M.J., Kramer, L.A., Levi, M.D., Wermers, R., 2017. Seasonal asset allocation: evidence from mutual fund flows. J. Financ. Quant. Anal. 52, 71–109.
Kaplanski, G., Levy, H., 2010. Sentiment and stock prices: the case of aviation disasters. J. Financ. Econ. 95, 174–201.
Kaplanski, G., Levy, H., Veld, C., Veld-Merkoulova, Y., 2015. Do happy people make optimistic investors? J. Financ. Quant. Anal. 50, 145–168.
Kaustia, M., Rantapuska, E., 2016. Does mood affect trading behavior? J. Financ. Markets 29, 1–26.
Keim, D.B., 1983. Size-related anomalies and stock return seasonality: further empirical evidence. J. Financ. Econ. 12, 13–32.
Keloharju, M., Linnainmaa, J.T., Nyberg, P., 2016. Return seasonalities. J. Finance 71, 1557–1590.
Lakonishok, J., Smidt, S., 1988. Are seasonal anomalies real? a ninety-year perspective. Rev. Financ. Stud. 1, 403–425.
Lerner, J.S., Li, Y., Valdesolo, P., Kassam, K.S., 2015. Emotion and decision making. Annu. Rev. Psychol. 66, 799–823.
Makridis, C., 2018. Can You Feel the heat? Extreme temperatures, Stock returns, and Economic Sentiment. MIT Unpublished working paper.
McFarlane, J., Martin, C.L., Williams, T.M., 1988. Mood fluctuations: women vesus men and menstrual versus other cycles. Psychol. Women Q 12, 201–223.
Newey, W.K., West, K.D., 1987. Hypothesis testing with efficient method of moments estimation. Int. Econ. Rev. (Philadelphia) 28, 777–787.
Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. J. Financ. Econ. 80, 563–602.
Ritter, J.R., 1988. The buying and selling behavior of individual investors at the turn of the year. J. Finance 43, 701–717.
Rossi, A.S., Rossi, P.E., 1977. Body time and social time: mood patterns by menstrual cycle phase and day of the week. Soc. Sci. Res. 6, 273–308.
Saunders, E.M., 1993. Stock prices and wall street weather. Am. Econ. Rev. 83, 1337–1345.
Stone, A.A., Schneider, S., Harter, J.K., 2012. Day-of-week mood patterns in the united states: on the existence of 'Blue monday', 'Thank god it's friday' and weekend effects. J. Pos. Psychol. 7, 306–314.
Thaler, R.H., 1987. Anomalies: the January effect. J. Econ. Perspect. 1, 107–201.

Journal of Financial Economics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Do investors care about carbon risk?☆

Patrick Bolton[a,b], Marcin Kacperczyk[a,b,*]

[a] Columbia University, Imperial College London, CEPR, and NBER, United States
[b] Imperial College London and CEPR, United Kingdom

ARTICLE INFO

Article history:
Received 21 October 2019
Revised 24 September 2020
Accepted 23 October 2020
Available online xxx

JEL classification:
G12
G23
G30
D62

Keywords:
Carbon emissions
Climate change
Stock returns
Institutional investors

ABSTRACT

We study whether carbon emissions affect the cross-section of US stock returns. We find that stocks of firms with higher total carbon dioxide emissions (and changes in emissions) earn higher returns, controlling for size, book-to-market, and other return predictors. We cannot explain this carbon premium through differences in unexpected profitability or other known risk factors. We also find that institutional investors implement exclusionary screening based on direct emission intensity (the ratio of total emissions to sales) in a few salient industries. Overall, our results are consistent with an interpretation that investors are already demanding compensation for their exposure to carbon emission risk.

1. Introduction

Many studies seek to explain the cross-sectional pattern of stock returns based on exposures to aggregate risk factors such as size and book-to-market ratios, or firm-specific risk linked to observable firm characteristics. One variable that has so far been missing from the analysis is corporate carbon emissions. This omission may be for historical reasons, as concerns over global warming linked to carbon dioxide (CO_2) emissions from human activity have only recently become salient. But, both the evidence of rising temperatures and the renewed policy efforts to curb CO_2 emissions raise the question of whether carbon emissions represent a material risk for investors that is reflected in the cross-section of stock returns and portfolio holdings.

Two major developments, in particular, suggest that this may be the case. First, the Paris COP 21 climate agreement of December 2015, with 195 signatories committing to limit global warming to well below 2 °C above pre-industrial levels. Second, the rising engagement of the finance industry with climate change, largely as a result of the call to non-governmental actors to join the fight against climate change at the COP 21. Institutional investors are increasingly tracking the greenhouse gas emissions of listed firms and forming coalitions such as Climate Action 100+ to engage with companies to reduce their

☆ We thank Jawad Addoum, Franklin Allen, Eric Bouyé, Kent Daniel, Charles Donovan, Elyse Douglas, Gerry Garvey, Lukasz Pomorski, Ailsa Roell, Zacharias Sautner, Gireesh Shrimali, Michela Verardo, Jeff Wurgler, and the referee for helpful suggestions. We also appreciate comments from seminar participants at Blackrock, EFA Conference, IESE, Newcastle University, the New Frontiers in Investment Research conference, NHH Bergen, NYU Law Roundtable on Climate Disclosure, UBS, University of Cardiff, and University College Dublin. We are grateful to Trucost for giving us access to their corporate carbon emissions data, and to Jingyu Zhang for very helpful research assistance. This project has received funding from the European Research Council (ERC) under the ERC Advanced Grant programme (grant agreement No. 885552 Investors and Climate Change).

* Corresponding author.
E-mail addresses: pb2208@columbia.edu (P. Bolton), mkacperc@ic.ac.uk (M. Kacperczyk).

https://doi.org/10.1016/j.jfineco.2021.05.008

carbon emissions.[1] More and more asset owners are following the lead of the Church of England Pension Fund, whose stated goal is "to demonstrate transparently that it has delivered on its commitment to be aligned to the Paris Agreement."[2]

Even if the US has pulled out of the Paris Agreement under the Trump administration, and even if the commitments of the other remaining signatories are only partially credible, major curbs in CO_2 emissions are likely to be introduced over the next decade. Primarily affected by these curbs are the companies with operations generating high CO_2 emissions, or with activities linked to companies in the value chain that have high CO_2 emissions. In light of these developments, one would expect to see the risk with respect to carbon emissions to be reflected in the cross-section of stock returns. Yet, considerable skepticism remains, not least in the US where the Trump administration had worked to upend regulations that limit CO_2 emissions. For example, Darren Woods, ExxonMobil's CEO, recently declared that "Individual companies setting targets and then selling assets to another company so that their portfolio has a different carbon intensity has not solved the problem for the world." And that ExxonMobil was "taking steps to solve the problem for society as a whole and not try and get into a beauty competition."[3]

The lack of consensus among institutional investors around climate change naturally raises the possibility that carbon risk may not yet be reflected in asset prices. To find out, in this paper we systematically explore whether investors demand a carbon risk premium by looking at how stock returns vary with CO_2 emissions across firms and industries. We undertake a standard cross-sectional analysis, asking whether carbon emissions affect cross-sectional US stock returns.

There are several ways in which one might expect CO_2 emissions to affect stock returns. First, since CO_2 emissions are tied to fossil-fuel energy use, returns are affected by fossil-fuel energy prices and commodity price risk. Relatedly, firms with disproportionately high CO_2 emissions may be exposed to carbon pricing risk and other regulatory interventions to limit emissions. The firms that are most reliant on fossil energy are also more exposed to technology risk from lower-cost renewable energy. Forward-looking investors may seek compensation for holding the stocks of disproportionately high CO_2 emitters and the associated higher carbon risk they expose themselves to, giving rise to a positive relation in the cross-section between a firm's own CO_2 emissions and its stock returns. We refer to this as the carbon risk premium hypothesis.

An interesting question is whether carbon emissions are perceived to be a systematic risk factor and whether the carbon risk premium is tied to loadings on this risk factor. Carbon emissions could be a systematic risk factor if expected regulatory interventions to curb emissions apply uniformly to all emissions. For example, if a large federal carbon tax were to be introduced, this would be a systematic shock affecting all companies with significant emissions. Alternatively, most regulatory interventions could be introduced in a piecemeal way at the state, industry, and municipal level. Similarly, technological improvements in the use of renewable energy could be mostly targeted to particular operations or sectors. In this case, one would not expect carbon emissions to be a systematic risk factor.

A second hypothesis is that financial markets are pricing carbon risk inefficiently and the risk associated with carbon emissions is underpriced. Carbon risk may not be fully integrated by most investors, who by force or habit look at future cash-flow projections through local thinking à la Gennaioli and Shleifer (2010), ignoring unrepresentative information about global warming and its attendant risks. To be sure, the cash-flow scenarios commonly used by financial analysts do not directly refer to carbon emissions and their possible future repricing. A recent study by In et al. (2019) on a different sample than ours finds that a portfolio that is long stocks of companies with low carbon emissions and short stocks of companies with high emissions generates positive abnormal returns. We refer to this hypothesis as the market inefficiency, or carbon alpha, hypothesis. An important question we explore is whether financial markets underprice carbon risk (after controlling for other known risk factors, industry, and firm characteristics) to the point that responsible investors, who care about carbon emissions and climate change, could be "doing well by doing good."

A third hypothesis is that the stocks of firms with high emissions are like other "sin stocks"; they are shunned by socially responsible, or ethical, investors to such an extent that the spurned firms present higher stock returns. A key question in this respect is how investors identify the firms to be divested from. Do they look at carbon emissions at the firm level, or do they pigeonhole firms into broader categories such as the industry they operate in? Even socially responsible investors that care about climate change may use sparse models (à la Gabaix, 2014) and not look much beyond industry categorizations, such as the energy and electric utility sectors, which produce a disproportionate share of CO_2 emissions. Prominent divestors like the Rockefeller Brothers Fund, for example, who have pledged to divest from fossil fuel companies, largely focus on energy companies that extract coal and oil from tar sands.[4] We refer to this as the divestment hypothesis.

A pioneer in producing company-level CO_2 emissions data is the Carbon Disclosure Project (CDP).[5] It has been joined by other leading providers of carbon data, including MSCI ESG Research and Trucost, among others.[6] While more and more institutional investors make use of the data, it is not known how much individual companies' stock returns are actually affected by the availability of

[1] See http://www.climateaction100.org/.

[2] Statement made by Adam Matthews, the fund's director of ethics and engagement. The Church of England Pension Fund is co-chairing the IIGCC initiative.

[3] Quoted in Exxon CEO Calls Rivals' Climate Targets a 'Beauty Competition' by Kevin Crowley, *Bloomberg News*, March 5, 2020, https://www.bnnbloomberg.ca/exxon-ceo-calls-rivals-climate-targets-a-beauty-competition-1.1400957.

[4] See https://www.rbf.org/mission-aligned-investing/divestment.

[5] See http://www.cdp.net/en-US/Pages/About-Us.aspx.

[6] See https://www.msci.com/climate-change-solutions and https://www.trucost.com/policy-academic-research.

these more granular CO_2 emissions data to financial analysts. Our study relies on the Trucost EDX data, which cover around 1000 listed companies since fiscal year 2005, and over 2900 listed companies in the US since fiscal year 2016. We match these data with the FactSet returns and balance sheet data for all US-listed companies from 2005 to 2017.

Carbon emissions from a company's operations and economic activity are typically grouped into three different categories: direct emissions from production (scope 1), indirect emissions from consumption of purchased electricity, heat, or steam (scope 2), and other indirect emissions from the production of purchased materials, product use, waste disposal, outsourced activities, etc. (scope 3). The scope 3 category in turn is separated into upstream and downstream indirect emissions. The data on scope 1 and scope 2 emissions are widely reported. Scope 3 emissions on the other hand are estimated using an input-output matrix. Although scope 3 emissions are the most important component of companies' emissions in a number of industries (e.g., automobile manufacturing), they have not been reported by companies until recently.

Our main broad finding is that carbon emissions significantly affect stock returns. For all three categories of emissions, we find a positive and statistically significant effect on firms' stock returns. We designate the higher returns associated with higher emissions as a carbon premium. We estimate how this carbon premium is related to three different measures of corporate emissions: 1) the total level of emissions; 2) the year-by-year change in emissions; and 3) emission intensity, which measures carbon emissions per unit of sales. A striking result is that the carbon premium is related to the level of (and to changes in) emissions, but not to emission intensity. One reason why the premium is tied to total emissions is that regulations limiting emissions are more likely to target activities where the level of emissions is highest. For example, in its planned climate stress test, the Bank of England focuses only on large firms and measures risk in terms of required reductions in the level of emissions (see the 2021 biennial exploratory scenario on the financial risks from climate change (Bank of England discussion paper, 2019)). Similarly, since technological change generally involves a fixed cost, renewable energy is more likely to displace fossil fuels in firms where returns to scale are highest. Another consideration is that since emission intensity is a ratio, it is likely to be a noisier metric of carbon risk exposure. Two firms with identical emission intensities may vary substantially in their levels of emissions. Indeed, this is what we find: the correlation coefficient between the level of scope 1 emissions and emission intensity is 0.6, and significantly less for scope 2 and scope 3. Nevertheless, it is somewhat surprising that we find no premium associated with emission intensity since emission-intensive firms might well be the first to become unprofitable should the carbon price rise. Investors would then demand a premium for holding these firms.

Interestingly, there is also a significant carbon premium associated with the year-to-year growth in emissions. As one would expect, we find that the level of emissions is highly persistent. Hence, emission levels reflect a long-run risk exposure with respect to carbon emissions. Changes in emissions, in turn, reflect short-run effects; how much worse, or better, carbon risk gets. Of course, changes in emissions could also indicate changes in earnings, but we control for this effect by adding the company's return on equity, sales growth, and earnings growth, among our independent variables.

The carbon premium is economically significant: A one-standard-deviation increase in respectively the level and change of scope 1 emissions leads to a 15-bps and 26-bps increase in stock returns, or respectively a 1.8% and 3.1% annualized increase. In addition, a one-standard-deviation increase in the level and change of scope 2 emissions leads to respectively a 24-bps and 18-bps increase in stock returns, or a 2.9% and 2.2% annualized increase. Finally, a corresponding one-standard-deviation increase in the level and change of scope 3 emissions increases stock returns by 33 bps and 31 bps per month, or 4.0% and 3.8% on an annual basis. Importantly, firms with higher emissions generate higher returns, after controlling for size, book-to-market, momentum, other well-recognized variables that predict returns, and firm characteristics, such as the value of property, plant & equipment (PPE), and investment over assets.

Other things equal, a carbon premium is the reflection of a lower investor demand for stocks of companies associated with high emissions. In equilibrium, this lower demand translates into a lower stock price, and possibly also lower holdings of high-emission stocks by some categories of investors. Following Hong and Kacperczyk (2009), we explore to what extent companies with high carbon emissions are treated like "sin stocks" by institutional investors. We find that, in aggregate, institutional investors do hold a significantly smaller fraction of companies with high scope 1 emission intensity, but they are not underweight companies with high levels of emissions. When we disaggregate by investor categories (mutual funds, insurance companies, banks, pension funds, and hedge funds), we find that insurance companies, pension funds, and mutual funds are underweight scope 1 emission intensity. The negative ownership effect of moving from high to low scope 1 emission-intensity firms is economically large and accounts for about 15%–20% of the cross-sectional variation in the ownership variable. This finding is in line with the rise in the sustainable investment movement and the popular negative exclusionary screening investment strategy followed by funds with an environmental, social, and governance (ESG) tilt.[7]

We find that divestment is only based on scope 1 emission intensity. This is true both in aggregate and for each institutional investor category. Essentially, institutional investors have been applying exclusionary screens (or not) solely on the basis of scope 1 emission intensity. Even more remarkable, we find that when we exclude the in-

[7] See Krueger, Sautner, and Starks (2020). Also, according to the Global Sustainable Investment Review 2018, negative/exclusionary screening is the largest sustainable investment strategy globally, representing $19.8 trillion of assets under management. http://www.gsi-alliance.org/wp-content/uploads/2019/03/GSIR_Review2018.3.28.pdf.

dustries with the highest CO_2 emissions (oil & gas, utilities, and motor industries), there is no significant exclusionary screening at all by institutional investors. In other words, the exclusionary screening is done entirely in these salient industries; in all other industries, there is no significant divestment. Overall, these findings lead us to reject the divestment hypothesis. First, although there is significant divestment by institutional investors, it is not directly linked to an effect on stock returns. Institutional investor portfolios are significantly underweight firms with high scope 1 emission intensity, but stock returns are not affected significantly by emission intensity.

Our finding that stock returns are positively related to the level (and changes) of carbon emissions is largely consistent with the view that investors are pricing in a carbon risk premium at the firm level. This result contradicts the carbon alpha hypothesis, whereby investors holding a portfolio long stocks of companies with low carbon emissions and short stocks of companies with high emissions generates positive abnormal returns. Garvey et al. (2018) and In et al. (2019) suggest that portfolios that sort stocks by emission intensity (going long stocks with low intensity and short stocks with high intensity) generate a positive alpha. In contrast, we find that there is no significant effect of carbon intensity on stock returns. Our study differs in two important respects from theirs. First, we cover a different time period and sample of firms. Second, we control for industry, firm characteristics, and known risk factors, while neither of these studies includes all of these controls. Controlling for industry significantly affects the results. Also, in contrast to In et al. (2019), we analyze the effects of carbon emissions for each scope category separately, thereby avoiding double counting.

Another important finding is that the carbon premium has only materialized recently. We show that if we look back to the 1990s by imputing the 2005 cross-sectional distribution of total emissions to the 1990s, there is no significant carbon premium, consistent with the view that investors at that time likely did not pay as much attention to carbon emissions. However, if we apply the same analysis to our sample period, by imputing the 2017 cross-sectional distribution of emissions back throughout our sample period, we find that there is a highly significant carbon premium.

To summarize, investors seem to take a somewhat schizophrenic attitude to carbon emissions. On the one hand, institutional investors clearly want to take a proactive approach by divesting from industries with high CO_2 emissions. On the other hand, this categorical exclusionary screening approach only partially addresses the carbon risk issue. Indeed, investors price in a carbon emission risk premium at the firm level in all industries even though divestment is concentrated in the industries with the highest CO_2 emissions (oil & gas, utilities, and transportation industries). If there is one general lesson that emerges from our analysis it is that carbon risk cannot just be reduced to a fossil fuel *supply* problem. It is also a *demand* problem. Once one factors in both the supply and demand aspects, all companies in all sectors are exposed to various degrees to carbon emissions risk. A coarse exclusionary approach focusing only on the energy and utility sectors misses the full extent of the problem investors face. Accounting for carbon risk is also required on the demand side, which inevitably involves the careful tracking of emissions at the firm level in all sectors.

Our study is related to a rapidly growing literature on climate change and financial markets. An early study by Matsumura, Prakash, and Vera-Munoz (2014) of S&P 500 firms between 2006 and 2008 looks at the effects of direct carbon emissions on firm value, and the effects of voluntary public disclosure of emissions (through CDP) on firm value. They find that higher emissions are associated with lower firm values, but that voluntary disclosure mitigates the negative valuation effect of emissions. Relatedly, Chava (2014) looks at the effects of environmental concerns, as reflected in KLD ratings, on firms' cost of capital. He finds that firms that derive substantial revenues from the sale of coal or oil, as reflected in a KLD rating, are associated with a higher implied cost of capital. In an extensive survey of institutional investors, Krueger et al. (2020) also find that institutional investors believe that carbon emissions represent a material risk. Among their responses, institutional investors also say that they do not believe that there is substantial underpricing of carbon risk. Andersson et al. (2016) propose a carbon risk hedging strategy for passive investors based on low carbon indexes.

More recently, Ilhan et al. (2020) find that carbon emissions increase downside risk as reflected in out-of-the-money put option prices. Hsu et al. (2019) look at the effects of environmental pollution on the cross-section of stock returns. They find that highly polluting firms are more exposed to environmental regulation risk and command higher average returns. Engle et al. (2020) construct an index of climate news through textual analysis of *The Wall Street Journal* and other media and show how a dynamic portfolio strategy can be implemented that hedges risk with respect to climate change news. Görgen et al. (2019) construct a carbon-risk factor and estimate a carbon beta for firms. Monasterolo and De Angelis (2019) explore whether investors demand higher risk premia for carbon-intensive assets following the COP 21 agreement.

Other related studies explore the asset pricing consequences of greater material risks linked to climate events and global warming. Hong et al. (2019) find that the rising drought risk caused by climate change is not efficiently priced by stock markets. Several studies examine climate change and real estate prices. Baldauf et al. (2020) find little evidence of declining prices as a result of greater flood risk due to sea level rise. Bakkensen and Barrage (2017) find that climate risk beliefs in coastal areas are highly heterogeneous and that rising flood risk due to climate change is not fully reflected in coastal house prices. Bernstein et al. (2019) find that coastal homes vulnerable to sea-level rise are priced at a 6.6% discount relative to similar homes at higher elevations. However, in a related study, Murfin and Spiegel (2020) find no evidence that sea level rise risk is reflected in residential real estate prices. Finally, Giglio et al. (2018) use real estate pricing data to infer long-run discount rates for valuing investments in climate change abatement.

The remainder of the paper is organized as follows. In Section 2, we describe the data and provide summary statistics. We discuss the results in Section 3. Concluding remarks are in Section 4.

2. Data and sample

Our primary database covers the 2005–2017 period and is largely a result of matching two data sets by Trucost and FactSet in the US. Trucost provides information on corporate carbon and other greenhouse gas emissions. FactSet provides data on stock returns, corporate fundamentals, and institutional ownership. We performed the matching using ISIN as a main identifier. In some instances, in which ISIN was not available to create a perfect match, we relied on matching based on company names (after standardizing the company names in FactSet and Trucost we match the names with a similarity score of one). Finally, when there are multiple subsidiaries of a given company, we used the primary location as a matching entity. The ultimate matching produced 3421 unique companies out of 3481 companies available in Trucost. Among the 60 companies we were not able to match, more than half are not exchange listed and the remaining ones are small. Hence, we believe our data cover almost the entire universe of companies with available emission data.

2.1. Data on corporate carbon emissions

Firm-level carbon emissions data are assembled by seven main providers: CDP, Trucost, MSCI, Sustainalytics, Thomson Reuters, Bloomberg, and ISS. All these providers follow the Greenhouse Gas Protocol that sets the standards for measuring corporate emissions.[8] More and more companies disclose their greenhouse gas emissions, and most large corporations report their emissions to CDP. Other providers rely on the CDP data and supplement it with other sources. Emissions can be measured directly at the source or more commonly by applying conversion factors to energy use. The Greenhouse Gas Protocol distinguishes between three different sources of emissions: scope 1 emissions, which cover direct emissions over one year from establishments that are owned or controlled by the company; these include all emissions from fossil fuel used in production. Scope 2 emissions come from the generation of purchased heat, steam, and electricity consumed by the company. Scope 3 emissions are caused by the operations and products of the company but occur from sources not owned or controlled by the company. These include emissions from the production of purchased materials, product use, waste disposal, and outsourced activities.

In some sectors, like automobile manufacturing, by far the most important component of their emissions is the aggregation of all their scope 3 emissions. The Greenhouse Gas Protocol distinguishes between 15 different categories of scope 3 emissions, including purchased goods and services, capital goods, upstream & downstream transportation and distribution, waste generated in operations, business travel, employee commuting, processing & use of sold products, and end-of-life treatment of sold products.[9] According to CDP's 2016 Climate Change Report, most scope 3 emissions are concentrated in two categories: purchased goods and services (around 44%) and use of sold products (around 48%).[10] The Greenhouse Gas Protocol provides detailed guidance on how to identify a company's most important sources of scope 3 emissions and how to calculate them. For purchased goods and services, this basically involves measuring inputs, or "activity data," and applying emission factors to these purchased inputs that convert activity data into emissions data. The upstream scope 3 data from Trucost that we use is constructed using an input-output model that provides the fraction of expenditures from one sector across all other sectors of the economy. This model is extended to include sector-level emission factors, so that an upstream scope 3 emissions estimate can be determined from each firm's expenditures across all sectors from which it obtains its inputs (see Trucost, 2019). Downstream scope 3 emissions caused using sold products can also be estimated and are increasingly reported by companies. Trucost has recently started assembling this data, but we were not able to include it our study.

Because they are easier to measure, and because disclosure requirements are stricter, data on scope 1 and scope 2 have been more systematically reported and accurately estimated. As Busch et al. (2018) show, there is very little variation in the reported scope 1 and 2 emissions data across the data providers. Correlations in the reported scope 1 (scope 2) data average 0.99 (0.98), across the five providers CDP, Trucost, MSCI, Sustainalytics, and Thomson Reuters.[11] However, when it comes to estimated scope 1 and scope 2 emissions (when reported data are missing), the correlations drop to 0.79 and 0.63, respectively for the three providers, Trucost, MSCI, and Sustainalytics, that offer these estimates. Finally, only two data providers, Trucost and ISS ESG, provide estimates of scope 3 emissions. The Trucost EDX database we use in our main analysis reports all three scopes of carbon emissions in units of tons of CO_2 emitted in a year. We report the summary statistics of the emissions variables in Panel A of Table 1.

The average firm in our sample produces 1.97 million tons of scope 1 emissions, and is tied to 1.72 million tons of scope 3 emissions. The quantity of scope 2 emissions is relatively smaller, at 342,000 tons of CO_2 equivalent. Notably, the median number is the largest for scope 3 emissions, as almost all companies in our sample are tied to a significant quantity of such emissions. The scope 1, 2, and 3 measures are in units of tons of CO_2 and normalized using the natural log scale. We also report annual growth rates in each emission measure. To mitigate the impact of outliers, we winsorize all growth measures at the 2.5% level. The carbon intensity of a company is expressed as tons of CO_2 equivalent divided by the com-

[8] See https://ghgprotocol.org.

[9] See http://ghgprotocol.org/standards/scope-3-standard.

[10] See CDP 2016 Climate Change Report "Tracking Progress on Corporate Climate Action."

[11] More than 6,300 companies worldwide answered CDP's climate change questionnaire in 2018. Of these, 76% disclosed scope 1 emissions, 68% scope 2 emissions, and 38% scope 3 emissions (see https://www.cdp.net).

Table 1
Summary statistics.

This tables reports summary statistics (averages, medians, and standard deviations) for the variables used for the six sets of regressions. The sample period is 2005–2017. **Panel A** reports the emission variables. **Panel B** reports the cross-sectional return variables. *RET* is the monthly stock return; LOG*SIZE* is the natural logarithm of market capitalization (in $ million); *B/M* is the book value of equity divided by market value of equity; *ROE* is the return on equity; *LEVERAGE* is the book value of leverage defined as the book value of debt divided by the book value of assets; *MOM* is the cumulative stock return over the one-year period; *INVEST/A* is the CAPEX divided by book value of assets; *HHI* is the Herfindahl index of the business segments of a company with weights proportional to revenues; LOG*PPE* is the natural logarithm of plant, property & equipment (in $ million); *BETA* is the CAPM beta calculated over the one year period; *VOLAT* is the monthly stock return volatility calculated over the one year period. **Panel C** reports the time-series variables. *MKTRF* is the monthly return on the value-weighted stock market net of the risk free rate; *HML* is the monthly return on the portfolio long value stocks and short growth stocks; *SMB* is the monthly return on the portfolio long small-cap stocks and short large-cap stocks; *MOM* is the monthly return on the portfolio long 12-month stock winners and short 12-month past losers; *CMA* is the monthly return of a portfolio that is long on conservative investment stocks and short on aggressive investment stocks; *BAB* is the monthly return of a portfolio that is long on low-beta stocks and short on high-beta stocks; *LIQ* is the liquidity factor of Pastor and Stambaugh; *NET ISSUANCE* is the monthly return of a portfolio that is long on high-net-issuance stocks and short on low-net-issuance stocks. Net issuance for year *t* is the change in the natural log of split-adjusted shares outstanding from the fiscal yearend in *t*-2 to the fiscal yearend in *t*-1; *IDIO VOL* is the monthly return of a portfolio that is long on low idiosyncratic volatility stocks and short on high idiosyncratic volatility stocks. **Panel D** reports the ownership variables. $IO_{i,t}$ is the fraction of the shares of company *i* held by institutions in the FactSet database at the end of year *t. IO* is calculated by aggregating the shares held by all types of institutions at the end of the year, and then dividing this amount by shares outstanding at the end of the year. *IO_BANKS* is the ownership by banks; *IO_INSURANCE* is the ownership by insurance companies; *IO_INVESTCOS* is the ownership by investment companies (e.g., mutual funds); *IO_ADVISERS* is the ownership by independent investment advisers; *IO_PENSIONS* is the ownership by pension funds; *IO_HFS* is the ownership by hedge funds. $PRINV_{i,t}$ is the inverse of firm *i*'s share price at the end of year *t; TOT* $VOLAT_{i,t}$ is the standard deviation of daily stock returns for company *i* over the one-year period; $VOLUME_{i,t}$ is the average daily trading volume (in $million) of stock *i* over the calendar year *t;* $NASDAQ_{i,t}$ is an indicator variable equal to one if a stock *i* is listed on NASDAQ in year *t*, and zero otherwise; $SP500_{i,t}$ is an indicator variable equal to one if a stock *i* is part of the S&P 500 Index in year *t*, and zero otherwise.

Variable	Mean	Median	Std. Dev.
Panel A: Emission variables			
Log (Carbon Emissions Scope 1 (tons CO_2e))	10.55	10.47	2.95
Log (Carbon Emissions Scope 2 (tons CO_2e))	10.52	10.66	2.36
Log (Carbon Emissions Scope 3 (tons CO_2e))	12.31	12.46	2.25
Growth Rate in Carbon Emissions Scope 1 (winsorized at 2.5%)	0.08	0.03	0.36
Growth Rate in Carbon Emissions Scope 2 (winsorized at 2.5%)	0.14	0.05	0.45
Growth Rate in Carbon Emissions Scope 3 (winsorized at 2.5%)	0.09	0.06	0.24
Carbon Intensity Scope 1 (tons CO_2e/USD m.)/100 (winsorized at 2.5%)	1.92	0.15	5.88
Carbon Intensity Scope 2 (tons CO_2e/USD m.)/100 (winsorized at 2.5%)	0.34	0.18	0.46
Carbon Intensity Scope 3 (tons CO_2e/USD m.) /100 (winsorized at 2.5%)	1.58	0.98	1.59
Carbon Intensity Direct (winsorized at 2.5%)/100	2.12	0.16	6.45
Carbon Intensity Indirect (winsorized at 2.5%)/100	1.04	0.58	1.31
GHG Direct Impact Ratio (winsorized at 2.5%)	0.75	0.06	2.29
GHG Indirect Impact Ratio (winsorized at 2.5%)	0.71	0.47	0.68
Panel B: Cross-sectional return variables			
RET (%)	1.14	1.08	10.84
LOGSIZE	8.25	8.25	1.57
B/M (winsorized at 2.5%)	0.50	0.39	0.41
LEVERAGE (winsorized at 2.5%)	0.24	0.22	0.18
MOM (winsorized at 0.5%)	0.15	0.11	0.45
INVEST/A (winsorized at 2.5%)	0.05	0.03	0.05
ROE (winsorized at 2.5%, in%)	9.76	11.32	21.23
HHI	0.82	1.00	0.24
LOGPPE	6.22	6.34	2.26
BETA	1.10	1.05	0.44
VOLAT (winsorized at 0.5%)	0.10	0.08	0.06
SALESGR (winsorized at 0.5%)	0.02	0.03	0.30
EPSGR (winsorized at 0.5%)	0.01	0.00	0.43
Panel C: Time-series variables			
MKTRF (in%)	0.70	1.06	4.08
HML (in%)	0.00	−0.22	2.57
SMB (in%)	0.07	0.04	2.26
MOM (in%)	0.07	0.36	4.53
CMA (in%)	0.02	−0.06	1.39
BAB (in%)	0.49	0.74	2.66
LIQ (in%)	0.15	0.38	3.59
NET ISSUANCE (in%)	0.51	0.55	1.65
IDIO VOL (in%)	−0.18	0.03	5.27
Panel D: Ownership variables			
IO (in%)	76.84	82.93	22.22
IO_BANKS (in%)	0.10	0.07	0.16
IO_INSURANCE (in%)	0.35	0.13	3.11
IO_INVESTCOS. (in%)	18.19	18.37	8.64
IO_ADVISERS (in%)	43.94	46.11	15.39
IO_PENSIONS (in%)	3.40	3.51	2.31
IO_HFS (in%)	10.87	7.73	10.04
PRINV (winsorized at 0.5%)	0.05	0.03	0.11
VOLAT (winsorized at 0.5%)	0.10	0.08	0.06
VOLUME (in $million) (winsorized at 2.5%)	0.44	0.21	0.56
NASDAQ	0.30	0.00	0.46
SP500	0.37	0.00	0.48

Table 2
Stock characteristics by emission calculation.
The table reports the sample means of the main variables over the 2005–2017 period. All variables are defined in Table 1. Imputed includes all firms for which Trucost estimates the levels of emissions. Direct includes all firms for which data is directly available.

Calculation Method	Imputed	Direct
SCOPE 1 TOT	1,366,013	5,954,876
SCOPE 2 TOT	264,203	957,827
SCOPE 3 TOT	1,433,741	4,057,516
SCOPE 1 INT	211.76	588.91
SCOPE 2 INT	35.89	68.26
SCOPE 3 INT	158.11	197.92
RET (%)	1.00	1.09
LOGSIZE	8.22	9.64
B/M	0.50	0.48
LEVERAGE	0.24	0.27
MOM	0.15	0.13
INVEST/A	0.05	0.05
ROE	9.88	14.89
HHI	0.84	0.72
LOGPPE	6.19	8.03
BETA	1.13	1.04
VOLAT	0.10	0.07
SALESGR (%)	1.67	−0.16
EPSGR (%)	1.53	0.25

pany's revenues in million US dollar units, also winsorized at the 2.5% level. The average (unwinsorized) scope 1 intensity in our sample equals 265.26 tons/million, while the intensities for scope 2 and scope 3 are 39.64 tons/million and 164.22 tons/million, respectively. The EDX database also provides information on whether the emissions data was reported or estimated, which allows us to do a sensitivity analysis and determine how the results are affected by the exclusion of the estimated data. We describe how the data breaks down into firms with reported and estimated emissions data in Table 2. As Matsumura, Prakash, and Vera-Munoz (2014) note, firms that do not report their emissions are typically smaller (and therefore have smaller emissions) and are less profitable. But in other respects, firms that report their emissions have similar characteristics to those that do not. In particular, their stock returns, volatility, leverage, book-to-market ratios, capital expenditures, and betas are very similar.

We also report alternative measures Trucost provides, in particular: i) *CARBON DIRECT*, which adds three additional greenhouse gas to the GHG Protocol scope 1 measures; ii) *CARBON INDIRECT*, which covers a slightly broader set of emissions by the direct suppliers to a company than scope 2; iii) *GHG DIRECT*, measured in US dollars, which covers all direct external environmental impacts of a company. Trucost applies a monetary value to GHG emissions quantities, which represents the global average damage of each environmental impact; and iv) *GHG INDIRECT*, which covers indirect supply chain environmental impacts. These are estimated impacts based on Trucost's environmental impact models. Again, these are reported in US dollars and represent the global average damages of each environmental impact.

How correlated are these different emission variables? We report the cross-correlations in Panel A of Table 3. As one would expect, the levels of all three categories of emissions are positively correlated. Yet, the coefficients are relatively small. Similarly, the level of scope 1 emissions is obviously positively correlated with scope 1 emission intensity, but the size of the coefficient is only 0.6, reflecting the fact that two firms with the same scope 1 intensity may have very different levels of emissions. A large firm, with high emissions, can have the same emission intensity as a small firm. The low correlation between levels and intensity is even more pronounced for scope 2 (0.24) and scope 3 (0.27). In Panel B, we also report the autocorrelation coefficients for the different measures of emissions. Emission levels for all three categories are highly persistent, with an autocorrelation coefficient of 0.977 for scope 1, 0.955 for scope 2, and 0.967 for scope 3. Interestingly, the year-to-year growth in emissions also has some persistence, especially for scope 3 emissions. As for the emission intensity variables they are, not surprisingly, also highly persistent as sales are highly persistent.

We also analyze the average values of all three emission sources over time. Fig. 1 and Table 4 present the results. As one might expect, there is a steady decline in scope 1 and scope 3 emissions at the firm level over time as a result of energy efficiency improvements, technological innovations, and an increase in the reliance on renewable energy sources. There is a sharp decline in scope 1 emissions from 2015 to 2016. However, this mainly reflects the addition by Trucost of many smaller firms to the sample in 2015, as can be seen in Fig. 2. The addition of all these firms to the sample also explains why total scope 3 emissions sharply increase from 2015 to 2016, and why total scope 1 emissions remain flat even though per-firm emissions decline. All these results are further confirmed by the numbers in Panels A and B of Table 4; averages for all firms in our sample are in Panel A while those conditioned on the presence in the sample prior to 2015 are in Panel B. We can see that when we drop the new firms added in 2016 from the sample, the averages for 2016 and 2017 are very close to the numbers in 2015. While we still observe some decline in scope 1 emissions, there is no such decline in scope 2 and scope 3 emissions. If anything, the numbers for scope 3 emissions go up, although not by much.

We note that firms with significant emissions are represented in a wide range of industries. In Table 5, we present the distribution of firms in our sample with respect to the six-digit Global Industry Classification (GIC 6). Banks, biotech, and oil & gas are the most represented industries, with each one having more than 150 firms.[12] In Table 6, we provide a list of industries with the highest and the lowest intensity of emissions. Power, electric, and multi-utility industries produce the most scope 1 emissions, while consumer finance, thrifts and mortgages, and capital markets are the cleanest. The ranking is somewhat different when we classify industries with respect to their scope 2 and scope 3 emissions. Metals and mining, electric utilities, and construction materials are the three most scope 2 emission-intensive industries (the cleanest industries mimic those based on scope 1 classification). In turn,

[12] Some firms in this table are classified into multiple industries; hence, the total number of firms in the table (3917) exceeds the number of unique firms in our sample (3421).

Table 3
Carbon emissions: correlations.

The sample period is 2005–2017. Panel A presents the cross-correlations among emission variables. Panel B presents the coefficients from estimating the AR(1) model for various measures of emissions. All regressions include year-month fixed effects. We cluster standard errors at firm and year dimensions. The emission variables are defined in Table 1. ***1% significance; **5% significance; *10% significance.

Panel A: Cross-correlations

	SCOPE 1 TOT	SCOPE 2 TOT	SCOPE 3 TOT	SCOPE 1 INT	SCOPE 2 INT	SCOPE 3 INT
SCOPE 1 TOT	1.00					
SCOPE 2 TOT	0.39	1.00				
SCOPE 3 TOT	0.51	0.75	1.00			
SCOPE 1 INT	0.60	0.03	0.03	1.00		
SCOPE 2 INT	0.05	0.24	0.02	0.10	1.00	
SCOPE 3 INT	0.21	0.09	0.27	0.25	0.10	1.00

Panel B: Autocorrelations

VARIABLES	(1) LOG (SCOPE 1)	(2) LOG (SCOPE 2)	(3) LOG (SCOPE 3)	(4) ΔSCOPE 1	(5) ΔSCOPE 2	(6) ΔSCOPE 3	(7) SCOPE 1 INT	(8) SCOPE 2 INT	(9) SCOPE 3 INT
LOG (SCOPE 1 TOT)$_{t-1}$	0.977***								
	(0.003)								
LOG (SCOPE 2 TOT)$_{t-1}$		0.955***							
		(0.005)							
LOG (SCOPE 3 TOT)$_{t-1}$			0.967***						
			(0.004)						
ΔSCOPE 1$_{t-1}$				0.045*					
				(0.021)					
ΔSCOPE 2$_{t-1}$					0.025				
					(0.015)				
ΔSCOPE 3$_{t-1}$						0.190***			
						(0.047)			
SCOPE 1 INT$_{t-1}$							0.945***		
							(0.005)		
SCOPE 2 INT$_{t-1}$								0.946***	
								(0.012)	
SCOPE 3 INT$_{t-1}$									0.969***
									(0.021)
Constant	0.281***	0.573***	0.475***	0.057***	0.106***	0.053***	0.065***	0.026***	0.031
	(0.033)	(0.052)	(0.046)	(0.001)	(0.002)	(0.003)	(0.013)	(0.004)	(0.033)
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	156,446	156,374	156,578	122,686	122,602	122,794	156,578	156,578	156,578
R-squared	0.972	0.945	0.975	0.014	0.020	0.085	0.962	0.850	0.964

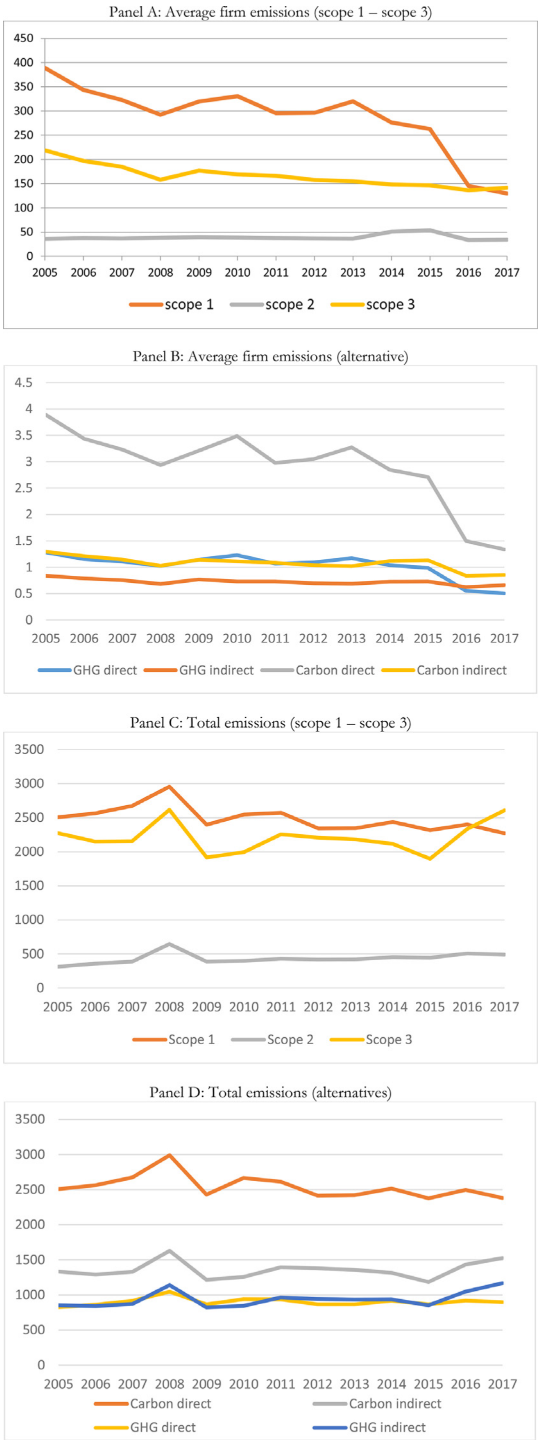


Fig. 1. Carbon emissions: time series summary.
The data source is Trucost and the data sample period is 2005–2017. Panels A and B present average firm emissions (in tons of CO_2 equivalent to revenues in \$ million). The emissions are broken down into scope 1, scope 2, and scope 3 emissions. In Panel B, *GHG Direct* and *GHG Indirect* are impact ratios expressed as a percentage of costs in revenues (in \$ million). *Carbon direct* and *Carbon indirect* are intensities expressed in tons of CO_2 equivalent to revenues in \$ million. Panels C and D present the total emissions (across all firms) per year.

food products, metals and mining, and construction materials are the three most scope 3 emission-intensive industries. Internet software and services, health care technologies, and software are the three least emission-intensive industries. The Trucost industry classification is finer than the GIC six-digit classification. Given that we control for industry a natural question is how sensitive the results are to the classification itself. The classification in theory could be so fine that it includes only one firm in each industry or so coarse that it includes all firms in one industry. Adding industry fixed effects would be meaningless under these polar classification systems. As a robustness check, we also perform our analysis under the GIC classification and report the results in Table A.4 in the Appendix.

Finally, we observe not only substantial variation in the growth rates of emissions across different industries, but also significant variation in the rates of all three categories of emissions across firms within the same industry. Fig. 3 displays the time series plots of the average cross-sectional standard deviations of emission growth rates across all firms (Panel A) and across all firms within a given GIC 6 industry (Panel B). Even though the scale of the variation in Panel A is larger than that in Panel B, there is still a significant dispersion in emissions in Panel B. Moreover, the standard deviation in carbon emission growth rates is very stable over time. In particular, the standard deviation did not significantly change following the addition of new firms to the sample in 2015.

2.2. Variables in cross-sectional return regressions

Our empirical analysis of stock returns employs a monthly measure of returns as a dependent variable. In our cross-sectional return regressions, the dependent variable $RET_{i,t}$ is the monthly return of an individual stock i in month t. Our return data primarily comes from FactSet, but for a small subset of delisted firms, we replace the return data with delisting-adjusted values from Compustat. Finally, we remove observations with returns greater than 100% to mitigate the impact of outliers. The number of excluded firm/month observations is 109 and its exclusion does not materially affect our results. However, using unrestricted returns data would be problematic as the data, for example, include four observations with monthly returns greater than 10,000%.

Our control variables are defined as follows: $LOGSIZE_{i,t}$ is the natural logarithm of firm i's market capitalization (price times shares outstanding) at the end of year t; $B/M_{i,t}$ is firm i's book value divided by its market capitalization at the end of year t; *LEVERAGE* is the book leverage of the company; $ROE_{i,t}$ is the firm's earnings performance, given by the ratio of firm i's net yearly income divided by the value of its equity; $MOM_{i,t}$ is the average of the most recent 12 months' returns on stock i, leading up to and including month t-1; *INVEST/A* represents the firm's capital expenditures divided by the book value of its assets; *HHI* is the Herfindahl concentration index of firms with respect to different business segments, based on each segment's revenues; LOG*PPE* is the natural logarithm, of the firm's property, plant, and equipment; $BETA_{i,t}$ is the market beta of firm i in year t, calculated over the one year period

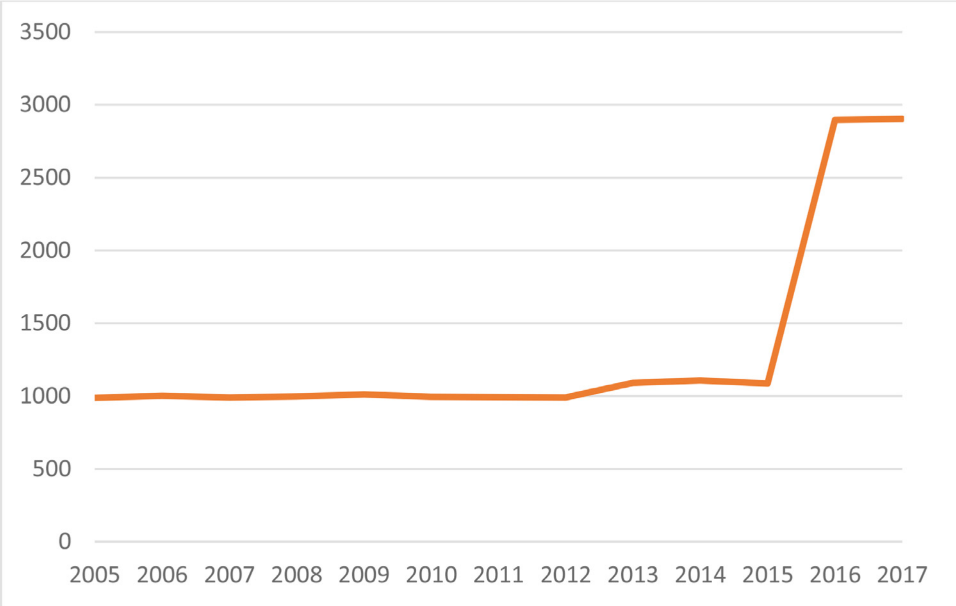


Fig. 2. Carbon emissions: sample selection.
The data source is Trucost. The figure presents the number of firms with valid emission data over the 2005–2017 period.

Table 4
Carbon emissions over time.

The table reports the cross-sectional averages of scope 1, scope 2, and scope 3 levels and intensity variables over the 2005–2017 period. Panel A considers a full sample of firms. Panel B is restricted to a sample of firms that existed prior to 2016. The emissions variables are defined in Table 1.

Panel A: Full sample

Year	SCOPE 1 TOT	SCOPE 2 TOT	SCOPE 3 TOT	SCOPE 1 INT	SCOPE 2 INT	SCOPE 3 INT
2005	2,697,225	335,402	2,414,925	411.16	37.55	229.79
2006	2,775,999	379,869	2,229,797	373.64	39.17	205.90
2007	2,893,335	410,656	2,281,158	341.57	37.38	193.13
2008	3,147,450	683,294	2,750,231	308.70	39.75	164.33
2009	2,482,940	385,670	1,907,531	334.35	41.41	184.06
2010	2,655,585	400,848	1,987,772	339.68	40.47	173.56
2011	2,639,823	440,716	2,217,712	305.06	40.20	169.39
2012	2,417,298	431,992	2,222,692	308.23	39.57	160.65
2013	2,223,849	398,491	2,046,741	335.82	39.22	159.69
2014	2,255,386	425,080	1,979,578	281.89	54.37	152.26
2015	2,161,598	419,362	1,783,537	273.32	56.79	150.77
2016	883,498	184,335	858,982	154.25	33.66	139.00
2017	809,277	176,805	935,203	139.29	33.88	145.53

Panel B: Legacy sample

Year	SCOPE 1 TOT	SCOPE 2 TOT	SCOPE 3 TOT	SCOPE 1 INT	SCOPE 2 INT	SCOPE 3 INT
2005	2,697,225	335,402	2,414,925	411.16	37.55	229.79
2006	2,775,999	379,869	2,229,797	373.64	39.17	205.90
2007	2,893,335	410,656	2,281,158	341.57	37.38	193.13
2008	3,147,450	683,294	2,750,231	308.70	39.75	164.33
2009	2,482,940	385,670	1,907,531	334.35	41.41	184.06
2010	2,655,585	400,848	1,987,772	339.68	40.47	173.56
2011	2,639,823	440,716	2,217,712	305.06	40.20	169.39
2012	2,417,298	431,992	2,222,692	308.23	39.57	160.65
2013	2,223,849	398,491	2,046,741	335.82	39.22	159.69
2014	2,255,386	425,080	1,979,578	281.89	54.37	152.26
2015	2,161,598	419,362	1,783,537	273.32	56.79	150.77
2016	1,993,060	404,850	1,874,254	269.09	45.78	167.35
2017	1,922,550	404,904	2,149,459	243.38	44.95	176.12

Table 5
Industry representation by number of firms.

The table reports the distribution of unique firms in our sample with regard to GIC 6 industry classification. *Total* represents the total number of firms in our sample. The sample period is 2005–2017.

GIC 6	Industry Name	# of Firms
1	Energy Equipment & Services	75
2	Oil, Gas & Consumable Fuels	164
3	Chemicals	81
4	Construction Materials	17
5	Containers & Packaging	21
6	Metals & Mining	47
7	Paper & Forest Products	12
8	Aerospace & defense	46
9	Building Products	32
10	Construction & Engineering	36
11	Electrical Equipment	54
12	Industrial Conglomerates	16
13	Machinery	118
14	Trading Companies & Distributors	40
15	Commercial Services & Supplies	69
16	Professional Services	42
17	Air Freight & Logistics	15
18	Airlines	13
19	Marine	27
20	Road & Rail	31
21	Transportation Infrastructure	5
22	Auto Components	43
23	Automobiles	8
24	Household Durables	64
25	Leisure Products	21
26	Textiles, Apparel & Luxury Goods	41
27	Hotels, Restaurants & Leisure	95
28	Diversified Consumer Services	38
29	Media	83
30	Distributors	8
31	Internet & Direct Marketing Retail	45
32	Multiline Retail	17
33	Specialty Retail	110
34	Food & Staples Retailing	27
35	Beverages	17
36	Food Products	57
37	Tobacco	9
38	Household Products	12
39	Personal Products	15
40	Health Care Equipment & Supplies	109
41	Health Care Providers & Services	77
42	Health Care Technology	20
43	Biotechnology	203
44	Pharmaceuticals	87
45	Life Sciences Tools & Services	34
46	Banks	260
47	Thrifts & Mortgage Finance	61
48	Diversified Financial Services	28
49	Consumer Finance	37
50	Capital Markets	92
51	Mortgage Real Estate Investment Trusts (REITs)	22
52	Insurance	111
53	Internet Software & Services	100
54	IT Services	102
55	Software	150
56	Communications Equipment	47
57	Technology Hardware, Storage & Peripherals	34
58	Electronic Equipment, Instruments & Components	82
59	Semiconductors & Semiconductor Equipment	103
60	Diversified Telecommunication Services	34
61	Wireless Telecommunication Services	15
62	Media	49
63	Entertainment	22
64	Interactive Media & Services	29
65	Electric Utilities	42
66	Gas Utilities	17
67	Multi-Utilities	30
68	Water Utilities	13
69	Independent Power and Renewable Electricity Producers	17
70	Equity Real Estate Investment Trusts (REITs)	184
71	Real Estate Management & Development	35
Total		3917

Table 6
Carbon emission production by industry.
Panel A reports the top 10 of GIC 6 industries in terms of average emission production (scope 1, scope 2, scope 3). Panel B reports the bottom 10 of GIC 6 industries in terms of average emission production (scope 1, scope 2, scope 3). The sample period is 2005–2017. The emission variables are expressed in tons of CO_2e.

Panel A: Largest emissions (avg.)					
GIC 6	Scope 1	GIC 6	Scope 2	GIC 6	Scope 3
69	33,300,000	34	2,163,081	23	18,700,000
65	30,700,000	23	2,094,174	36	11,800,000
18	17,600,000	6	1,749,360	37	6,847,386
67	17,200,000	3	1,475,783	12	6,575,213
6	6,343,545	7	1,375,637	35	6,106,099
2	6,302,663	60	1,219,956	2	6,049,237
17	4,316,221	12	1,014,037	34	5,882,429
4	3,827,648	38	994,783	38	4,313,762
7	3,286,922	32	825,501	6	3,580,245
3	3,280,770	2	820,777	22	3,285,134
Panel B: Smallest emissions (avg.)					
GIC 6	Scope 1	GIC 6	Scope 2	GIC 6	Scope 3
47	601	47	1756	47	15,193
50	6767	42	11,824	51	27,069
46	6965	19	21,798	68	41,182
49	7469	16	22,653	42	64,097
64	7649	43	24,606	71	84,764
51	8770	50	35,404	70	102,300
53	8898	51	36,013	16	114,132
55	9132	66	39,177	46	116,073
42	11,657	45	44,082	28	145,311
16	17,895	46	45,627	43	151,772

using daily data; $VOLAT_{i,t}$ is the standard deviation of returns based on the past 12 months of monthly returns; $SALESGR_{i,t}$ is the dollar change in annual firm revenues normalized by last month's market capitalization; $EPSGR_{i,t}$ is the dollar change in annual earnings per share, normalized by the firm's equity price. To eliminate the impact of outliers, we winsorize *B/M, LEVERAGE,* and *INVEST/A* at the 2.5% level, and *MOM, VOLAT, SALESGR*, and *EPSGR* at the 0.5% level. We report the summary statistics of these variables in Panel B of Table 1.

The average firm's monthly stock return is 1.14%, with a standard deviation of 10.84%. The average firm has a market capitalization of $13 billion, with a median value of $3.8 billion. The average book-to-market ratio is 0.50, while the average book leverage is 24%. The average market beta is 1.10, slightly more than that of the market.

2.3. Variables in the time series return regressions

The variables for our time series regressions are defined as follows: $MKTRF_t$ is the monthly return of the CRSP value-weighted portfolio in month *t*, net of the risk-free rate; SMB_t, HML_t, MOM_t, and CMA_t are well-known portfolio return series downloaded from Ken French's website: *SMB* is the monthly return of a portfolio that is long on small stocks and short on large stocks; *HML* is the monthly return of a portfolio that is long on high book-to-market stocks and short on low book-to-market stocks; *MOM* is the monthly return of a portfolio that is long on past one-year return winners and short on past one-year return losers; *CMA* is the monthly return of a portfolio that is long on conservative investment stocks and short on aggressive investment stocks. *BAB* is the monthly return of a portfolio that is long on low-beta stocks and short on high-beta stocks; *LIQ* is the liquidity factor of Pastor and Stambaugh; *NET ISSUANCE* is the monthly return of a portfolio that is long on high-net-issuance stocks and short on low-net-issuance stocks. Net issuance for year *t* is the change in the natural log of split-adjusted shares outstanding from the fiscal yearend in *t*-2 to the fiscal yearend in *t*-1; *IDIO VOL* is the monthly return of a portfolio that is long on low idiosyncratic volatility stocks and short on high idiosyncratic volatility stocks. We present the summary statistics for the various portfolio returns in Panel C of Table 1.

The average market risk premium in our sample is 0.7% per month. Other factors with relatively high risk premia are net issuance and *BAB*. Somewhat atypically, the value factor return in our sample is equal to 0%. Similarly, the momentum factor generates a mere 0.07% per month, and the volatility factor has a negative return of −0.18% per month.

2.4. Variables in divestment regressions

Our institutional ownership regression variables are: $IO_{i,t}$, which is the fraction of the shares of company *i* held by institutions in the FactSet database at the end of year *t. IO* is calculated by aggregating the shares held by all types of institutions at the end of the year, and then dividing this value by the number of shares outstanding at the end of the year. We further decompose the institutional ownership with respect to subgroups of owners. *IO_BANKS* is the ownership by banks; *IO_INSURANCE* is the ownership by insurance companies; *IO_INVESTCOS*

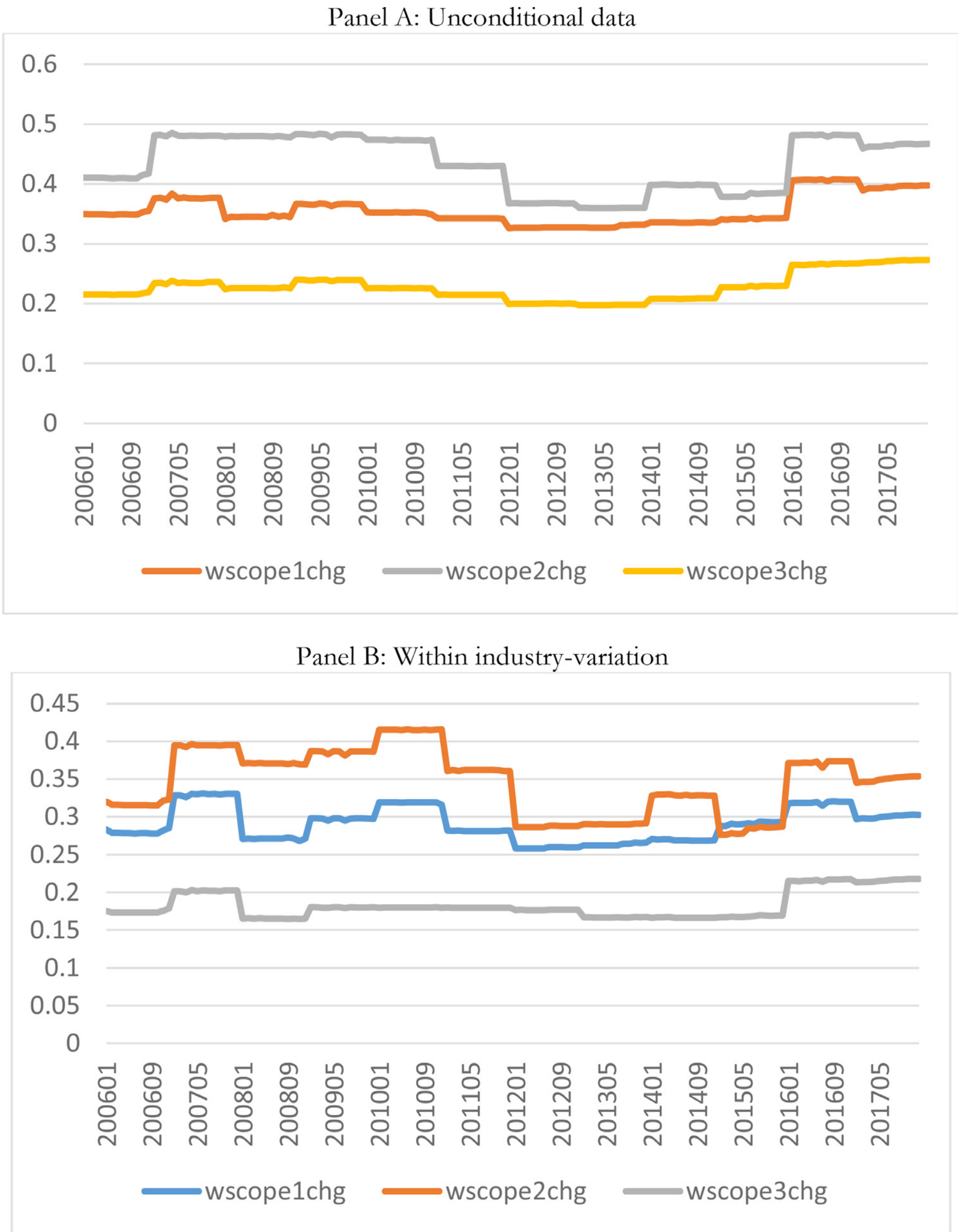


Fig. 3. Standard deviation of carbon emission growth rates.
The data source is Trucost. Panel A presents the cross-sectional standard deviation of firm-level emissions. Panel B presents the cross-sectional standard deviation of firm level emissions within GIC-6 industries, all averaged across all the industries in a given year. All emissions are broken down into scope 1, scope 2, and scope 3, over the 2005–2017 period. The emission levels are measured in millions of tons of CO_2 equivalent and are winsorized at the 1% level.

is the ownership by investment companies (e.g., mutual funds); *IO_ADVISERS* is the ownership by independent investment advisers; *IO_PENSIONS* is the ownership by pension funds and *IO_HFS* is the ownership by hedge funds. Even though the total institutional ownership captures the intensive margin only, the range of disaggregated ownership variables varies from 0% to 100% (as long as the total institutional ownership in the data has a positive value).

The control variables in the ownership regressions include $PRINV_{i,t}$, which is the inverse of firm *i*'s share price at the end of year *t*; $VOLAT_{i,t}$ is the standard deviation of monthly stock returns for firm *i* over the one-year period; $VOLUME_{i,t}$ is the average daily trading volume (in $million) of stock *i* over the calendar year *t*. $NASDAQ_{i,t}$ is an indicator variable equal to one if a stock *i* is listed on NASDAQ in year *t*, and zero otherwise; $SP500_{i,t}$ is an indicator variable equal to one if a stock *i* is part of the S&P 500 Index in year t, and zero otherwise. We report the summary statistics for these variables in Panel D of Table 1.

The average *IO* is 0.77, and the cross-sectional standard deviation of *IO* is 0.22. In other words, in a typical year, a typical firm has about 77% of its shares held by institutional investors, and the standard deviation of institutional ownership in a typical cross-section is 22%. Among the different institutional owners, independent advisers are the biggest holders, with an average stock's ownership equal to 43.9%, followed by investment companies with an average 18.2% ownership. Banks and insurance companies, in turn, are the smallest institutional owners. The average daily stock return volatility in our sample is 10% or annualized 158.7%. The average daily stock volume is $440,000. Finally, about 30% of stock-month observations are companies listed on NASDAQ, and 37% observations are firms from the S&P 500 Index.

3. Results

We begin our analysis by investigating the determinants of scope 1, scope 2, and scope 3 emissions. We then turn to the evaluation of the carbon return premium in the cross-section of stocks. We next explore the time-series properties of the cross-sectional carbon premium with respect to well-known risk factors. Finally, we consider the divestment hypothesis by looking at institutional ownership patterns.

3.1. Determinants of carbon emissions

Since emissions are not reported by all companies, one basic issue to explore first is how companies that do report their emissions compare with non-reporting companies. To assess the quantitative differences on the extensive margin, we compare various firm-level characteristics for the reporting and non-reporting firms. We describe basic summary statistics of the two categories of firms in Table A.1 of the Appendix. As one might expect, we find that larger firms are more likely to report their emissions. Also, firms with lower book-to-market ratios and higher book leverage are more likely to report emissions. At the same time, the two groups of firms do not differ significantly in terms of their stock returns or investment levels.

Next, we assess the differences in emission levels, year-by-year changes, and emission intensities across firms using a regression framework. Our dependent variables are levels, changes, and intensities of scope 1, scope 2, and scope 3. Since there is little theory that can guide us on what determines the level of carbon emissions, especially with regard to their different sources, we include a host of firm-level variables, comprising LOG*SIZE*, *B/M*, *ROE*, *LEVERAGE*, *INVEST/A*, *HHI*, LOG*PPE*, *SALESGR*, and *EPSGR*. To reflect the possibility that firm-level emissions could concentrate across firms and over time, we cluster standard errors at the firm and year levels. Standard errors in all panel regressions become significantly smaller in alternative specifications that cluster at the firm, industry, time, or industry and time levels. We present the results in Table 7.

Not surprisingly, all three categories of emission levels, and changes in emissions, are significantly positively related to LOG*SIZE*. However, scope 1 and scope 3 emission intensities are weakly negatively related to LOG*SIZE*. The level of emissions is also significantly associated with high book-to-market ratios, high tangible capital (PPE), highly levered firms, and firms with high growth in sales and earnings. On the other hand, the level of emissions is lower for firms with high capital expenditures, although these growth firms are associated with high increases in emissions. Interestingly, only diversification (HHI) and tangible capital significantly affect emission intensity.

3.2. Evidence on cross-sectional returns

For all three categories of emissions, we relate in turn the level of companies' emissions, the year-to-year growth in emissions, and the companies' emission intensity to their corresponding stock returns in the cross-section. We first estimate the following cross-sectional regression model using pooled OLS:

$$
\begin{aligned}
RET_{i,t} &= a_0 + a_1 LOG\ (TOT\ Emissions)_{i,t} \\
&\quad + a_2 Controls_{i,t-1} + \mu_t + \varepsilon_{i,t},
\end{aligned}
\tag{1}
$$

where $RET_{i,t}$ measures the stock return of company i in month t and *Emissions* is a generic term alternately standing for *SCOPE 1*, *SCOPE 2*, and *SCOPE 3* emissions. The vector of controls includes a host of firm-specific variables known to predict returns, such as LOG*SIZE*, *B/M*, *ROE*, *LEVERAGE*, *MOM*, *INVEST/A*, *HHI*, LOG*PPE*, *BETA*, *VOLAT*, *SALESGR*, and *EPSGR*.[13] We also include year/month fixed effects. We cluster standard errors at the firm and year levels. Our coefficient of interest is a_1.

We report the results in Table 8, Panel A. Column 1 shows the results for *SCOPE 1*; column 2 for *SCOPE 2*, and column 3 for *SCOPE 3*. For all three categories of emissions, we find a positive and statistically significant effect on firms' stock returns. The effect is also economically significant: a one-standard-deviation increase in *SCOPE 1* leads to a 13-bps increase in stock returns, or 1.5% annualized, and a one-standard-deviation increase in *SCOPE 2* leads to a 23-bps increase in stock returns, or 2.8% annualized. Finally, a one-standard-deviation increase in *SCOPE 3* increases stock returns by 30 bps per month, or 3.6% annualized.

Since emissions tend to cluster significantly within specific industries, a question of interest is whether the firm-specific differences can be attributed to industry-specific effects. To examine this possibility, we additionally include industry-fixed effects using the Trucost industry classification. The results presented in columns 4 to 6 are quite striking. Including industry effects significantly strengthens the cross-sectional dispersion of returns due to carbon emissions. In fact, the economic significance increases by anywhere between 70% and 280% relative to the model without industry effects.

We also plot the time series of the cumulative values of the unadjusted and industry-adjusted carbon premia in

[13] *HHI*, *SALESGR*, and *EPSGR* are measured as of time t to reflect the fact that all three may have a nontrivial contemporaneous effect on the level of emissions at time t.

Table 7
Determinants of carbon emissions.
The sample period is 2005–2017. The dependent variables are natural logarithm of total emissions, percentage change in total emissions, and carbon intensity. All variables are defined in Table 1. We report the results of the pooled regression with standard errors clustered at the firm level and year (in parentheses). All regressions include year-month fixed effects and industry-fixed effects. ***1% significance; **5% significance; *10% significance.

Variables	(1) LOG (SCOPE 1)	(2) LOG (SCOPE 2)	(3) LOG (SCOPE 3)	(4) ΔSCOPE 1	(5) ΔSCOPE 2	(6) ΔSCOPE 3	(7) SCOPE 1 INT	(8) SCOPE 2 INT	(9) SCOPE 3 INT
LOGSIZE	0.438***	0.571***	0.572***	0.026***	0.026***	0.027***	−0.118*	0.002	−0.021**
	(0.036)	(0.032)	(0.022)	(0.008)	(0.008)	(0.006)	(0.063)	(0.006)	(0.009)
B/M	0.464***	0.555***	0.562***	−0.033**	−0.038	−0.041**	−0.003	0.003	0.000
	(0.060)	(0.059)	(0.054)	(0.015)	(0.021)	(0.017)	(0.107)	(0.010)	(0.013)
ROE	0.006***	0.006***	0.007***	−0.002***	−0.002***	−0.001***	−0.002	−0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)	(0.000)	(0.000)
LEVERAGE	0.531**	0.625***	0.574***	0.026	0.010	0.019	0.364	0.002	−0.056*
	(0.196)	(0.188)	(0.162)	(0.020)	(0.030)	(0.023)	(0.230)	(0.030)	(0.030)
INVEST/A	−2.026***	−1.950***	−2.457***	0.676***	0.706***	0.530***	−0.586	−0.067	−0.446**
	(0.489)	(0.460)	(0.432)	(0.145)	(0.132)	(0.117)	(1.161)	(0.153)	(0.201)
HHI	−1.044***	−0.569***	−0.499***	0.014	−0.024	0.023**	−2.185***	0.009	−0.260***
	(0.119)	(0.081)	(0.063)	(0.021)	(0.024)	(0.008)	(0.497)	(0.030)	(0.062)
LOGPPE	0.376***	0.372***	0.317***	−0.033***	−0.034***	−0.030***	0.127***	0.025***	0.026***
	(0.036)	(0.037)	(0.023)	(0.005)	(0.006)	(0.006)	(0.042)	(0.007)	(0.007)
SALESGR	0.237***	0.190**	0.231**	0.311***	0.343***	0.320***	−0.085	−0.019**	0.010
	(0.059)	(0.062)	(0.077)	(0.042)	(0.041)	(0.030)	(0.070)	(0.007)	(0.024)
EPSGR	0.137**	0.146**	0.144**	−0.005	−0.011	0.001	0.009	0.006**	−0.002
	(0.049)	(0.049)	(0.050)	(0.008)	(0.012)	(0.006)	(0.038)	(0.003)	(0.006)
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	189,187	189,115	189,283	156,506	156,410	156,578	189,283	189,283	189,283
R-squared	0.899	0.849	0.905	0.150	0.136	0.320	0.786	0.650	0.935

Table 8
Carbon emissions and stock returns.
The sample period is 2005–2017. The dependent variable is *RET*. All variables are defined in Table 1. We report the results of the pooled regression with standard errors clustered at the firm and year level (in parentheses). All regressions include year-month fixed effects. In the regressions for columns 4 through 6, we additionally include industry-fixed effects. Panel A reports the results for the natural logarithm of total firm-level emissions; Panel B reports the results for the percentage change in carbon total emissions; Panel C reports the results for carbon emission intensity. ***1% significance; **5% significance; *10% significance.

Panel A: Total emissions

Variables	(1)	(2)	(3)	(4)	(5)	(6)
LOG (SCOPE 1 TOT)	0.043**			0.164***		
	(0.023)			(0.036)		
LOG (SCOPE 2 TOT)		0.098**			0.167***	
		(0.042)			(0.048)	
LOG (SCOPE 3 TOT)			0.135**			0.312***
			(0.046)			(0.071)
LOGSIZE	−0.140	−0.184	−0.193	−0.302*	−0.327*	−0.410**
	(0.163)	(0.167)	(0.165)	(0.148)	(0.154)	(0.163)
B/M	0.460	0.469	0.444	0.656**	0.642**	0.562**
	(0.260)	(0.266)	(0.258)	(0.234)	(0.229)	(0.224)
LEVERAGE	−0.559*	−0.579*	−0.498*	−0.699***	−0.712***	−0.790***
	(0.272)	(0.280)	(0.274)	(0.177)	(0.171)	(0.167)
MOM	0.321	0.348	0.338	0.284	0.294	0.301
	(0.276)	(0.272)	(0.274)	(0.291)	(0.290)	(0.290)
INVEST/A	−2.218	−1.914	−1.587	0.277	0.267	0.699
	(1.740)	(1.794)	(1.838)	(2.111)	(2.126)	(2.082)
ROE	0.010*	0.009	0.008	0.009*	0.009*	0.007*
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
HHI	0.032	−0.026	0.137	0.130*	0.052	0.111
	(0.110)	(0.112)	(0.101)	(0.072)	(0.073)	(0.071)
LOGPPE	−0.015	−0.027	−0.045	0.020	0.019	−0.017
	(0.100)	(0.088)	(0.090)	(0.058)	(0.058)	(0.057)
BETA	0.059	0.023	0.047	0.045	0.040	0.063
	(0.131)	(0.131)	(0.130)	(0.148)	(0.147)	(0.146)
VOLAT	0.978	0.674	0.749	0.622	0.501	0.549
	(3.571)	(3.415)	(3.506)	(3.290)	(3.285)	(3.269)
SALESGR	0.692	0.688	0.672	0.679	0.686	0.648
	(0.429)	(0.430)	(0.420)	(0.412)	(0.412)	(0.407)
EPSGR	0.592**	0.589**	0.575**	0.637**	0.636**	0.615**
	(0.234)	(0.231)	(0.232)	(0.231)	(0.233)	(0.227)
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	184,288	184,216	184,384	184,288	184,216	184,384
R-squared	0.203	0.204	0.204	0.206	0.206	0.206

Panel B: Growth rate in total emissions

Variables	(1)	(2)	(3)	(4)	(5)	(6)
ΔSCOPE 1	0.641***			0.627***		
	(0.153)			(0.144)		
ΔSCOPE 2		0.345**			0.321**	
		(0.125)			(0.120)	
ΔSCOPE 3			1.203***			1.186***
			(0.318)			(0.314)
LOGSIZE	−0.023	−0.013	−0.037	−0.107	−0.099	−0.121
	(0.110)	(0.112)	(0.111)	(0.114)	(0.115)	(0.117)
B/M	0.391	0.388	0.410*	0.771**	0.764**	0.789***
	(0.232)	(0.233)	(0.226)	(0.257)	(0.257)	(0.246)
LEVERAGE	−0.433*	−0.414*	−0.441*	−0.794***	−0.785***	−0.799***
	(0.217)	(0.216)	(0.213)	(0.213)	(0.217)	(0.214)
MOM	0.204	0.217	0.166	0.160	0.175	0.124
	(0.265)	(0.268)	(0.267)	(0.264)	(0.266)	(0.264)
INVEST/A	−2.508	−2.244	−2.638	−0.620	−0.463	−0.807
	(1.820)	(1.848)	(1.867)	(2.326)	(2.291)	(2.341)
ROE	0.009**	0.009**	0.009**	0.008**	0.008**	0.009**
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
HHI	−0.143	−0.112	−0.162	−0.072	−0.056	−0.089
	(0.154)	(0.153)	(0.151)	(0.098)	(0.097)	(0.102)
LOGPPE	−0.006	−0.015	0.006	0.053	0.045	0.066
	(0.058)	(0.057)	(0.060)	(0.041)	(0.041)	(0.044)
BETA	0.109	0.119	0.106	0.155	0.166	0.145

(Continued on next page)

Table 8
(*Continued*)

Panel B: Growth rate in total emissions						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
	(0.165)	(0.165)	(0.168)	(0.158)	(0.157)	(0.162)
VOLAT	1.853	2.004	1.800	1.373	1.504	1.341
	(4.240)	(4.226)	(4.274)	(4.072)	(4.075)	(4.107)
SALESGR	0.459	0.544	0.280	0.463	0.549	0.284
	(0.447)	(0.454)	(0.430)	(0.429)	(0.434)	(0.402)
EPSGR	0.573**	0.573**	0.568**	0.641**	0.641**	0.636**
	(0.247)	(0.246)	(0.250)	(0.263)	(0.263)	(0.266)
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	153,051	152,955	153,123	153,051	152,955	153,123
R-squared	0.218	0.218	0.218	0.221	0.221	0.222

Panel C: Emission intensity						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
SCOPE 1 INT	−0.010			0.005		
	(0.012)			(0.006)		
SCOPE 2 INT		0.145			0.081	
		(0.121)			(0.074)	
SCOPE 3 INT			0.055			0.048
			(0.033)			(0.075)
LOGSIZE	−0.154	−0.133	−0.124	−0.229	−0.230	−0.229
	(0.169)	(0.159)	(0.164)	(0.142)	(0.141)	(0.142)
B/M	0.456	0.470	0.479*	0.732**	0.732**	0.732**
	(0.264)	(0.269)	(0.258)	(0.244)	(0.243)	(0.244)
LEVERAGE	−0.545*	−0.558*	−0.532*	−0.608***	−0.606***	−0.603***
	(0.264)	(0.269)	(0.263)	(0.195)	(0.195)	(0.196)
MOM	0.332	0.321	0.317	0.282	0.282	0.281
	(0.277)	(0.279)	(0.279)	(0.292)	(0.292)	(0.291)
INVEST/A	−1.953	−2.047	−1.916	−0.041	−0.037	−0.022
	(1.815)	(1.823)	(1.867)	(2.123)	(2.127)	(2.134)
ROE	0.010*	0.010*	0.010*	0.010**	0.010**	0.010**
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
HHI	−0.139	−0.069	0.028	−0.032	−0.043	−0.030
	(0.137)	(0.113)	(0.082)	(0.074)	(0.072)	(0.067)
LOGPPE	0.034	0.010	0.006	0.081	0.079	0.080
	(0.099)	(0.087)	(0.093)	(0.065)	(0.064)	(0.066)
BETA	0.047	0.045	0.051	0.035	0.034	0.036
	(0.131)	(0.131)	(0.131)	(0.148)	(0.148)	(0.148)
VOLAT	1.027	0.978	1.028	0.577	0.558	0.572
	(3.512)	(3.527)	(3.563)	(3.296)	(3.297)	(3.300)
SALESGR	0.709	0.714	0.712	0.718	0.719	0.717
	(0.435)	(0.432)	(0.427)	(0.414)	(0.413)	(0.413)
EPSGR	0.600**	0.600**	0.600**	0.660**	0.660**	0.661**
	(0.234)	(0.232)	(0.232)	(0.235)	(0.236)	(0.235)
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	184,384	184,384	184,384	184,384	184,384	184,384
R-squared	0.203	0.203	0.203	0.206	0.206	0.206

Fig. 4. Because different emission variables have different supports, we express the magnitudes in terms of unit standard deviation of each variable at each cross-section in time, so that all plots of the cumulative effect show comparable numbers in terms of economic significance. As can be seen in the figure, there are large positive cumulative returns for all measures of total emissions. The economic magnitudes of the effect become even larger once we factor in differences in industry exposures.

We next estimate the same cross-sectional regression model (1) replacing the level of emissions (*LOG (Emissions TOT)*) with the year-to-year growth in emissions (*Δ(Emissions)*). The results are reported in Table 8, Panel B. We find again a positive and statistically significant effect of the growth in emissions on stock returns. Interestingly, controlling for industry makes almost no difference when it comes to the effect of the growth in emissions. To allay any concern that our results may be driven by the correlation between emissions and size, we provide additional robustness tests in which we estimate univariate regression models with respective emission variables only, and regressions with emissions and size only. The results, reported in Table A.2 of the Online Appendix indicate that size is an important control when one considers the level of total emissions as a regressor but it is not as important in the model with the growth rate of emissions. Note

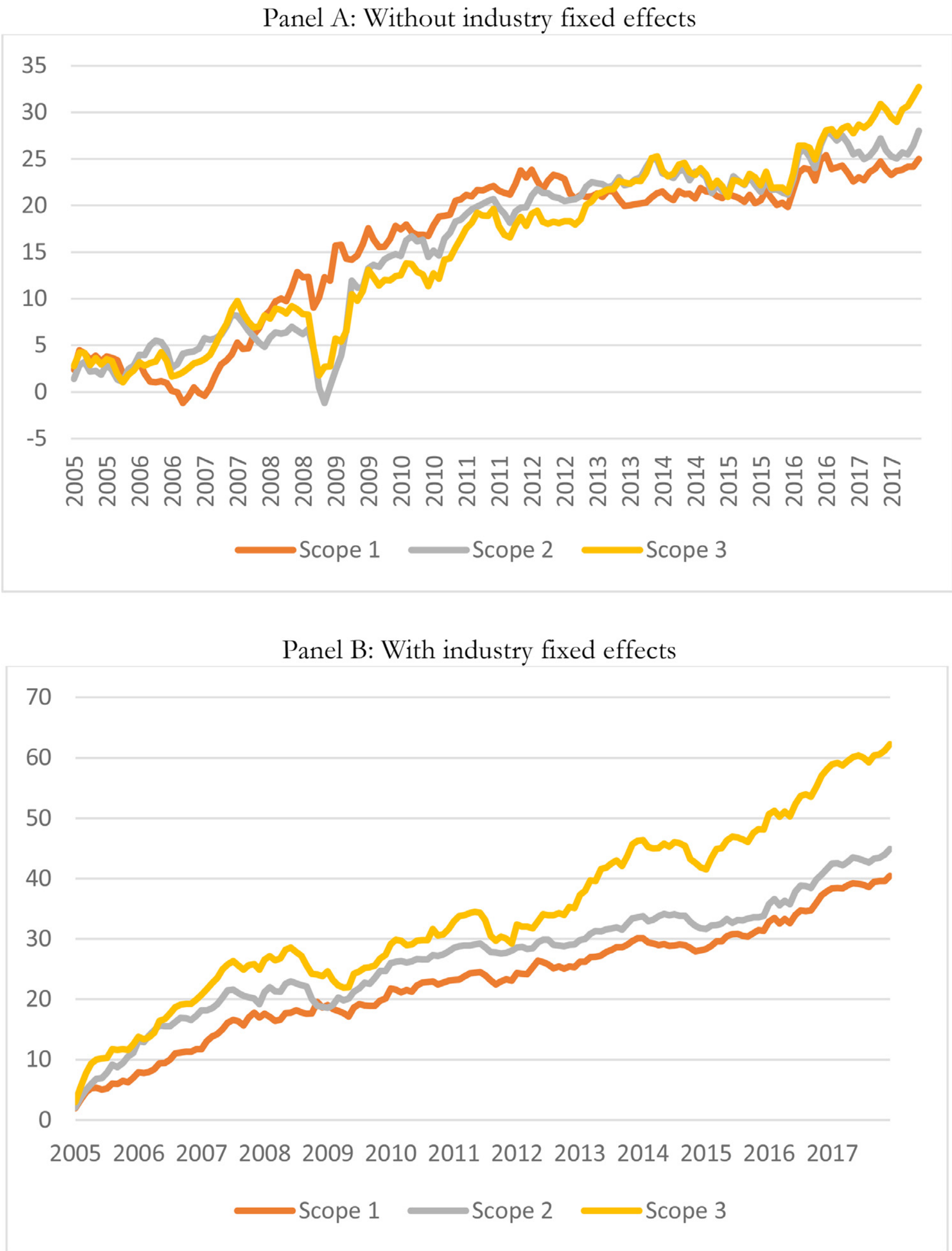


Fig. 4. Carbon cumulative return premia: level effect.
Figures show the cumulative values of carbon premia estimated from the cross-sectional regressions of monthly returns on the natural logarithm of the level of scope 1, scope 2, and scope 3 emissions. The regressions include the same set of controls as in Table 7. Panel A shows the plots for the model without industry fixed effects, while Panel B shows the results with industry-fixed effects as additional control. The data source is Trucost and the sample period is 2005–2017.

also that *ROE* has a significant positive effect on stock returns under this specification (it is insignificant in the specification with emission levels). We attribute this to the fact that firms with high emission growth likely also have higher earnings, which could result in higher stock returns (to the extent that the higher earnings outcome is unanticipated).

We also plot the time series of the cumulative values of the unadjusted and industry-adjusted carbon premia related to the growth in emissions in Fig. 5. All measures of emissions exhibit a steady rate of increase in the carbon premium over time.

Finally, we estimate the cross-sectional regression model in (1) for emission intensities. We report the results in Table 8, Panel C. There is no significant effect of emission intensity on returns for any of the three categories of emissions, whether we control for industry or not. The cumulative effect of emission intensity on the carbon pre-

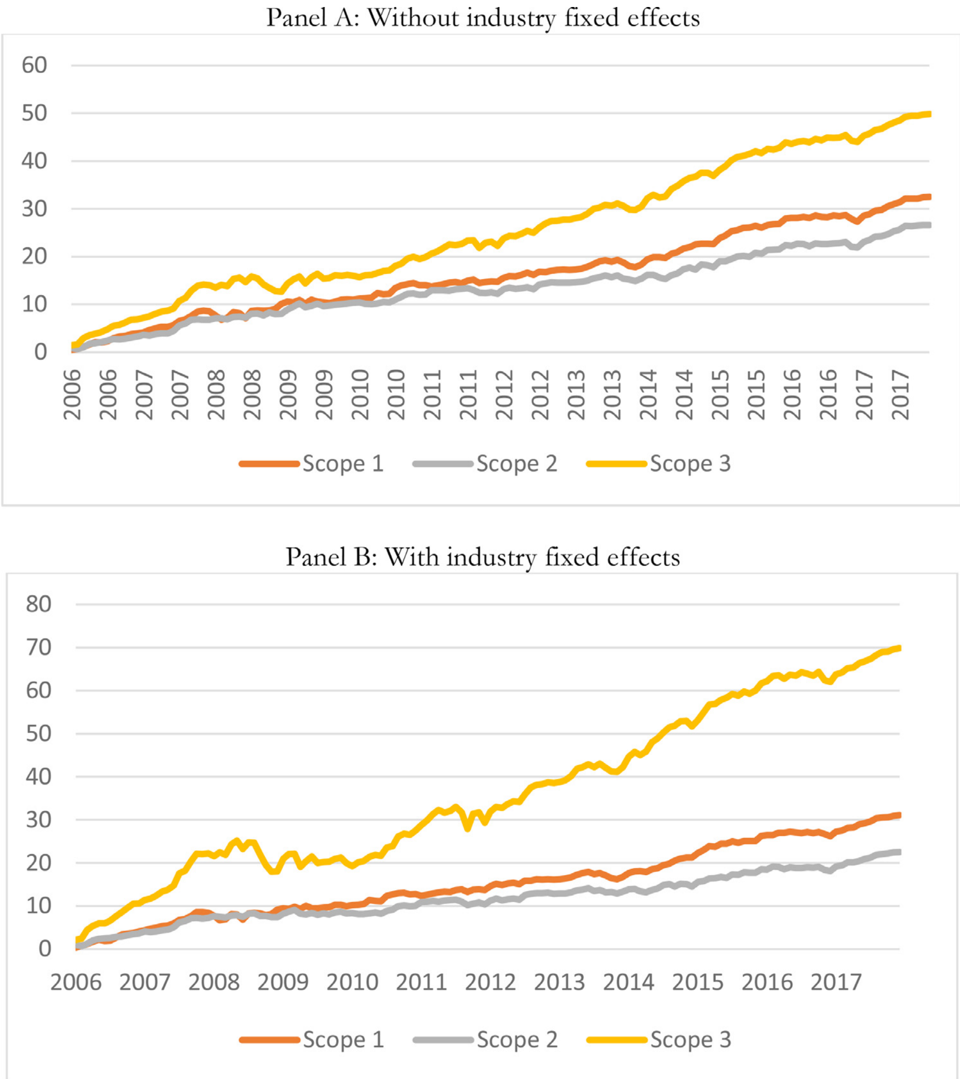


Fig. 5. Carbon cumulative return premia: change effect.
Figures show the cumulative values of carbon premia estimated from the cross-sectional regressions of monthly returns on the percentage changes (year over year) of scope 1, scope 2, and scope 3 emission levels. The regressions include the same set of controls as in Table 7. Panel A shows the plots for the model without industry fixed effects, while Panel B shows the results with industry-fixed effects as additional control. The data source is Trucost and the sample period is 2005–2017.

mium, presented in Fig. 6, is also quite weak, with the exception of scope 2 for which we observe a slightly positive trend. Overall, these results reveal that there is a significant carbon premium with respect to the level of emissions, reflecting firms' long-run risk exposure to carbon emissions, and a premium with respect to the growth in emissions, which capture the more short-term evolution of firms' risk exposure to future emissions.

One open question with our analysis above is that we use carbon emission data in year t to explain monthly returns over the same year t. This could conceivably introduce a look-ahead bias. That is, under this specification we might unwittingly relate stock returns for some months in year t to emission data that might not yet have been available to investors. To address this question, we undertake the following robustness check. We relate monthly stock returns with a lag of respectively 0 to 12 months between the time when emissions are reported and the month when returns are realized.

Another interpretation of the results with lagged returns is that investors have limited attention and do not immediately absorb new information about carbon emissions at the firm level (Kacperczyk et al., 2016). In that case, carbon emissions for year t will be gradually reflected in returns over the year. An additional consideration is that investors obtain information about carbon emissions from multiple sources that are not all available at the same time. For example, a lot of firms disclose their emissions first to the CDP, data which then is merged into and combined with other sources by Trucost. Different information that is likely to be highly correlated with other information (given that all providers use the same data collection protocols)

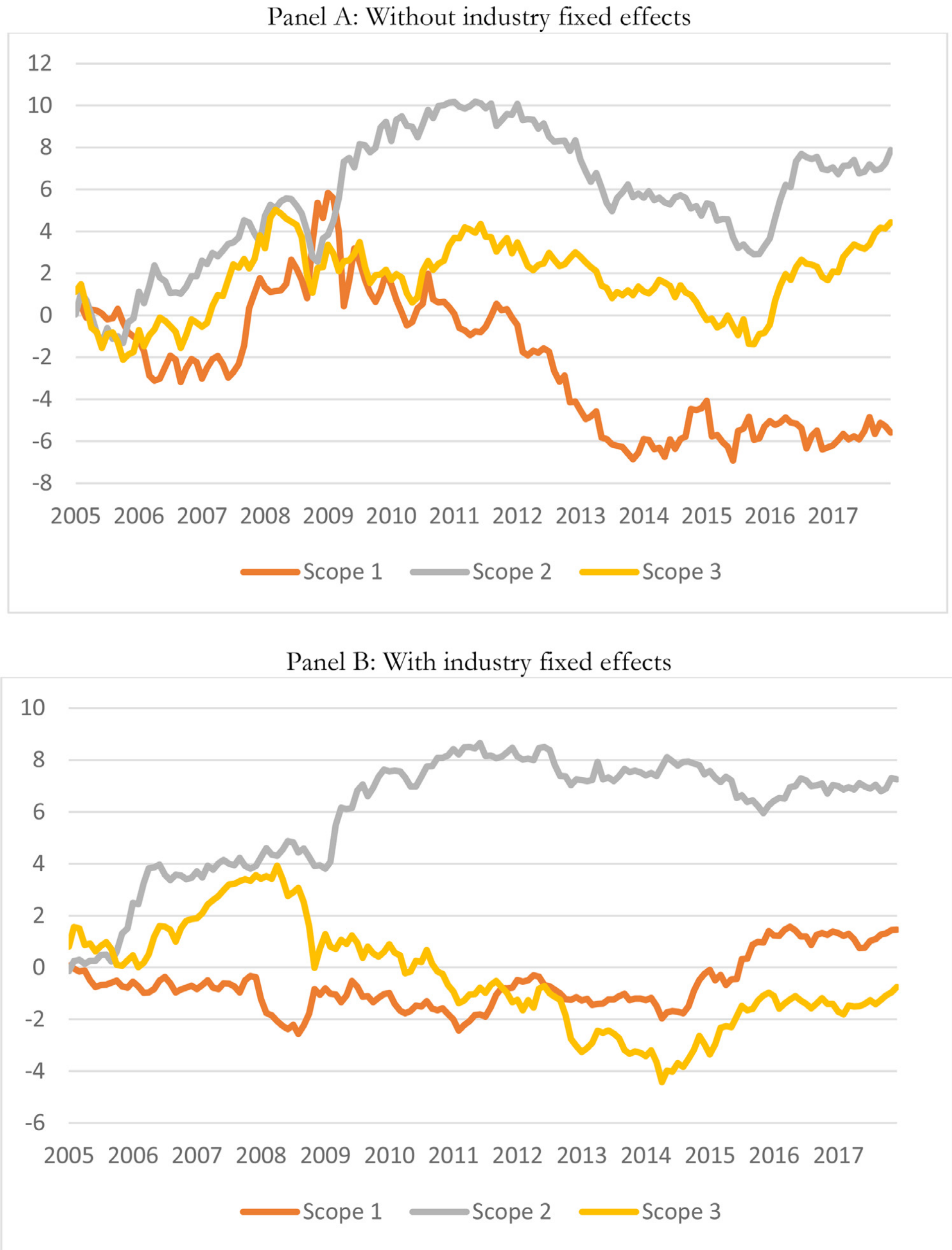


Fig. 6. Carbon cumulative return premia: intensity effect.
Figures show the cumulative values of carbon premia estimated from the cross-sectional regressions of monthly returns on the carbon intensity of scope 1, scope 2, and scope 3 emissions. The regressions include the same set of controls as in Table 7. Panel A shows the plots for the model without industry fixed effects, while Panel B shows the results with industry-fixed effects as additional control. The data source is Trucost and the sample period is 2005–2017.

becomes available at different times. This is another reason why carbon emissions are only gradually reflected in stock returns.

We report the results in Table A.3. A remarkable pattern emerges from this analysis. Panel A1 reports the results for LOG *(SCOPE 1 TOT)*. The coefficient is statistically significant for the first month (without industry fixed effects), remains significant at the 5% level until month 6 (with industry fixed effects), and is insignificant thereafter. Not surprisingly, it takes time for information about emissions to be reflected in stock prices, but eventually (after six months or so) this information appears to be fully absorbed. Essentially the same pattern is observed for the level of scope 2 and scope 3 emissions (with a somewhat faster (slower) integration of scope 2 (scope 3) emission information into stock prices), as the results in Panels A2 and A3 show. The same pattern is present for the growth in total emissions, as can be seen in panels B1, B2, and B3. However, emission intensity is nearly always insignificant, as we report in Panels C1, C2, and C3. The only visible ex-

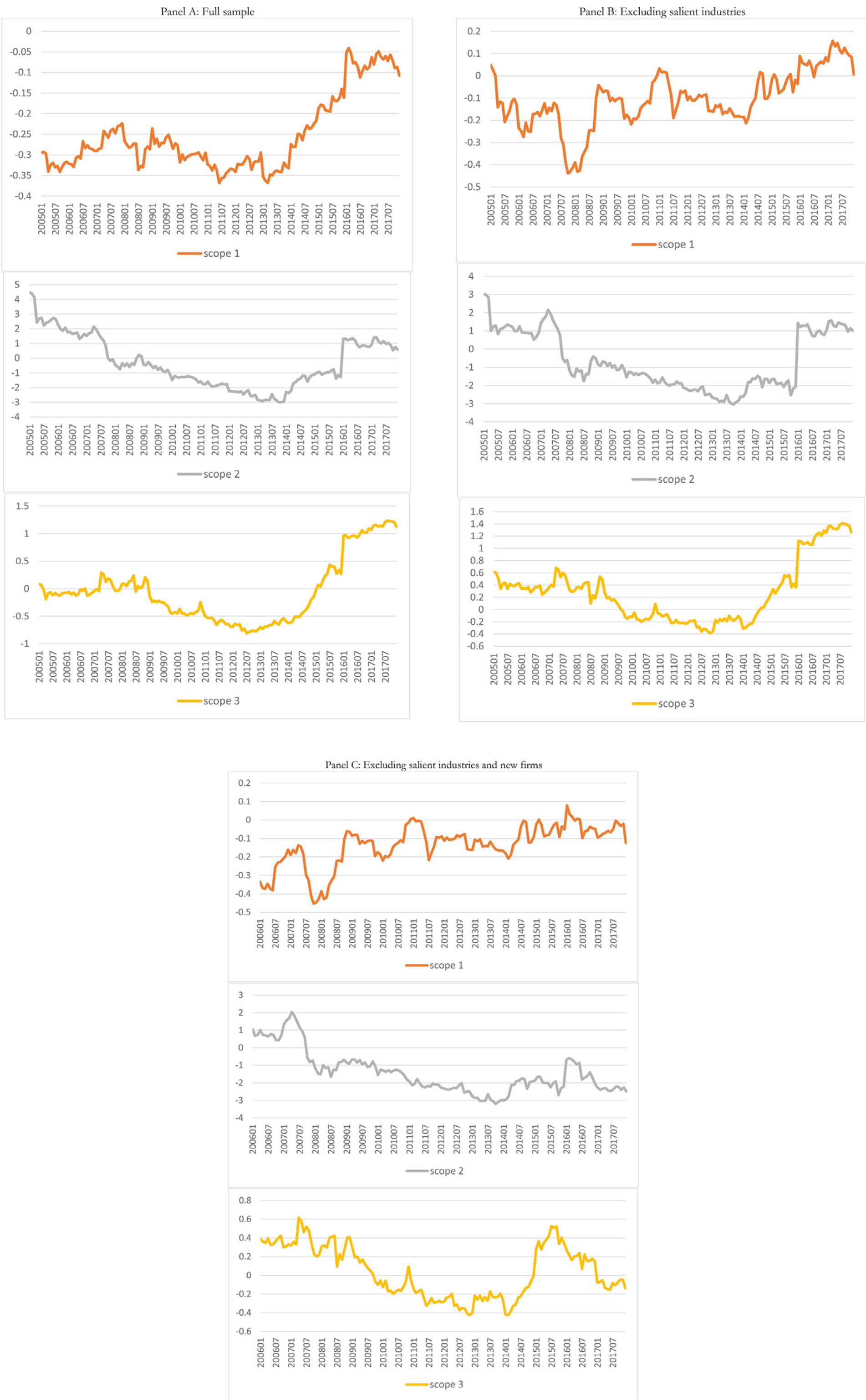


Fig. 7. Carbon intensity and institutional ownership: cumulative effect.
Figures show the cumulative values of the coefficient of emission intensity estimated from the cross-sectional regressions of monthly firm-level institutional ownership on scope 1, scope 2, and scope 3 emissions intensity. The regressions include the same set of controls as Table 11. Panel A shows the plots for the full sample, Panel B shows the results for the sample of firms excluding salient industries (GIC 19, 20, 23), Panel C shows the results for the sample of firms excluding the same salient industries and also firms that are added to the sample post 2015. The data source is Trucost and the sample period is 2005–2017.

Table 9
Carbon emissions and stock returns net of earnings returns.
The sample period is 2005–2017. The dependent variable is *RET* net of daily return realized on the earnings announcement day. All variables are defined in Table 1. We report the results of the pooled regression with standard errors clustered at the firm and year level (in parentheses). All regressions include year-month fixed effects. In the regressions for columns 4 through 6, we additionally include industry-fixed effects. Panel A reports the results for the natural logarithm of total emissions; Panel B reports the results for the percentage change in carbon total emissions; Panel C reports the results for carbon emission intensity. ***1% significance; **5% significance; *10% significance.

Panel A: Total emissions						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
LOG (SCOPE 1 TOT)	0.044*			0.152***		
	(0.024)			(0.031)		
LOG (SCOPE 2 TOT)		0.088**			0.150***	
		(0.040)			(0.044)	
LOG (SCOPE 3 TOT)			0.121**			0.279***
			(0.047)			(0.067)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	184,288	184,216	184,384	184,288	184,216	184,384
R-squared	0.220	0.221	0.220	0.223	0.223	0.223
Panel B: Growth rate in total emissions						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
ΔSCOPE 1	0.552***			0.532***		
	(0.137)			(0.131)		
ΔSCOPE 2		0.288**			0.266**	
		(0.111)			(0.108)	
ΔSCOPE 3			0.896**			0.882**
			(0.313)			(0.316)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	153,051	152,955	153,123	153,051	152,955	153,123
R-squared	0.235	0.236	0.235	0.239	0.239	0.239
Panel C: Emission intensity						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
SCOPE 1 INT	−0.008			0.004		
	(0.011)			(0.007)		
SCOPE 2 INT		0.155			0.079	
		(0.124)			(0.068)	
SCOPE 3 INT			0.050			0.029
			(0.032)			(0.071)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	184,384	184,384	184,384	184,384	184,384	184,384
R-squared	0.220	0.220	0.220	0.223	0.223	0.223

ception is scope 1 emission intensity, which is significant at the 5% level in month 6 in the model with industry fixed effects. We conclude from this analysis that our results are not biased by a look-ahead effect.

Another possible explanation is that firms with higher emissions have also been exposed to unexpected positive value shocks. We explore this hypothesis by analyzing returns that strip out the effect of earnings surprises. Specifically, we subtract from the monthly stock returns the component that is realized on earnings announcement days and re-estimate the regression model in (1) with the adjusted returns. We report the results in Table 9 for the level of total emissions (Panel A), for the growth rate of emissions (Panel B), and for emission intensity (Panel C). We find no significant differential effect of earnings announcements on the carbon premium. Stocks with higher levels and growth rates of emissions still have higher returns, and emission intensity is still insignificant.

3.3. Carbon premium and risk factors

Is the carbon premium linked to traditional risk factors? To answer this question, we estimate the following time-series regression model using monthly data:

$$a_{1,t} = c_0 + \boldsymbol{cF_t} + \varepsilon_t, \tag{2}$$

where $a_{1,t}$ is the carbon return premium estimated from the cross-sectional Fama-MacBeth regression in Eq. (1); $\boldsymbol{F}$ is a set of factor-mimicking portfolios that includes *MKTRF, HML, SMB, MOM, CMA, BAB, LIQ, NET ISSUANCE*, and

IDIO VOL. These factors have been widely used in many studies of asset prices. There are also economic reasons to believe that they could be meaningfully related to our carbon factor. Specifically, the first five factors correspond to the classic framework of Fama and French. In light of our results reported above, firm-level emissions are related to firm size and to firms' growth opportunities; hence we include both the SMB and HML factors. The investment factor, *CMA*, controls for any differences in investments across firms. The market and momentum factors are standard controls in all time-series regressions. The BAB factor controls for the possibility that high carbon risk firms may be exposed to margin investments. The liquidity factor controls for possible differences in market liquidity among firms with different levels of carbon emissions, which could arise if some firms are not as actively traded as others due to ESG norm-based reasons. The net-issuance factor controls for any variation in capital structure and market timing by firm managers. Finally, the idiosyncratic volatility factor controls for the possibility that the measure of risk we capture may be idiosyncratic in nature. We calculate the standard errors of the coefficients using the Newey-West procedure with 12 lags to account for autocorrelation in error terms. Our coefficient of interest is c_0, which measures the residual carbon premium, controlling for other risk/style factors.

Panel A in Table 10 shows the results for the carbon premium related to total emissions. In the odd columns, we report the unconditional carbon premium as a benchmark. In the even columns, we report results from regressions that add various factors *MKTRF, HML, SMB, MOM, CMA, BAB, LIQ, NET ISSUANCE*, and *IDIO VOL*. Comparing the odd and even columns for the respective scope categories of emissions, we find that the carbon premium remains statistically and economically significant after we adjust for differential factor exposures. However, the economic size of the premium is about 10%–20% smaller in magnitude. Overall, the regression intercepts from the cross-sectional return regressions are both economically and statistically significant in the presence of various risk factors.

Panel B shows the results for the carbon premium related to the growth rate in total emissions. We find again that the set of standard risk factors cannot explain the average value of the carbon premium for any of the emissions categories. This time, however, the difference in magnitudes across specifications is much smaller. Panel C gives the results for emission intensity. Whether unconditionally or conditionally on the risk factors, we find no significant carbon premium.

Overall, our time-series regression results show that the carbon premium cannot be explained by known risk factors. This result reinforces the finding in Section 3.2 that the level of carbon emissions contains independent information about the cross-section of average returns.

3.4. The divestment hypothesis

An important possible explanation for the observed carbon premium could be under-diversification resulting from divestment and exclusionary screening of stocks with high carbon emissions by institutional investors implementing a sustainable investment policy. To the extent that some investors may shun companies with high carbon emissions, risk sharing would be limited, and idiosyncratic risk could be priced (e.g., Merton, 1987; Hong and Kacperczyk, 2009). If the extent of such divestment is high, one would expect to see significant pricing effects.

We test this possibility by looking at the portfolio holdings of institutional investors. Formally, we estimate the following pooled regression model:

$$IO_{i,t} = d_0 + d_1 Emission_{j,t} + d_2 Controls_{j,t} + \varepsilon_{i,t}. \quad (3)$$

We consider ownership effects based on carbon intensity, the measure that is most aligned with explicit mandates imposed by socially sensitive asset managers. In the Online Appendix, Table A.4, we also present the results for the less commonly used measures of total emissions and growth in emissions. As these results confirm, these variables have no significant impact on institutional investor portfolios. The vector of controls includes LOG*SIZE, PRINV, B/M, MOM, BETA, VOLAT, VOLUME, NASDAQ*, and *SP500*. All regressions include year/month fixed effects. Also, carbon emissions tend to vary geographically, due to resource-driven firm locations. It is thus possible that the geographic location may also interact with ownership incentives. We test this idea by including in the ownership regression state fixed effects determined by the firm headquarters' locations (in even columns). Our coefficient of interest is d_1, which measures the degree of avoidance of firms with greater carbon emissions. We cluster standard errors at the industry and year levels.

In Panel A of Table 11, we report the results for the aggregate institutional ownership measure. Columns 1 and 2 show the results for *SCOPE 1 INT*, respectively without and with state fixed effects. Both coefficients are negative and statistically significant at the 5% level. The economic effect of the divestment is relatively modest: a one-standard-deviation increase in *SCOPE 1* leads to approximately a 1.3-percentage-point decrease in aggregate institutional ownership, which is about 6.3% of the cross-sectional standard deviation in ownership. In contrast, the coefficients are statistically insignificant for *SCOPE 2 INT* and *SCOPE 3 INT*, indicating that the exclusionary screens institutional investors apply in constructing their portfolios are entirely based on *SCOPE 1 INT*.

The institutional investor world pools a number of different constituencies with possibly different investor pressures. We conjecture that certain institutions, such as insurance companies, investment advisers, or pension funds, are more likely to face investor pressure, and thus they avoid high-emission companies, as opposed to mutual funds and hedge funds who are natural arbitrageurs. We test this hypothesis formally by dividing the institutional investors' universe into six categories: banks, insurance companies, investment companies, independent advisers, pension funds, and hedge funds. For each category, we obtain their stock-level institutional ownership and estimate the regression model in (3) for each of them separately. Panel B reports the results broken down by investor category. We observe a strong cross-sectional variation in the ownership patterns. Insurance companies, investment advisers, and pension funds tend to hold less of the high scope 1 emission companies. At the same time, we observe positive, though weaker, ownership effects for banks, investment companies, and hedge funds, consistent with these groups being natural arbitrageurs. The divestment effects are economically large. A movement in *SCOPE 1 INT* from one standard deviation below the mean to one standard deviation above the mean, corresponding to a spread between low and high-emission firms leads to a reduction in ownership by 21%, 5%, and 4% of the

Table 10
Can the carbon premium be explained by risk factors?

The sample period is 2005–2017. The dependent variable is the monthly carbon premium estimated each period using a cross-sectional return regression. All variables are defined in Table 1. We report the results of the time-series regression with standard errors adjusted for autocorrelation with 12 lags using Newey-West test (in parentheses). Panel A reports the results for the natural logarithm of contemporaneous total emissions; Panel B reports the results for the percentage change in carbon emissions; Panel C reports the results for carbon emission intensity. ***1% significance; **5% significance; *10% significance.

Panel A: Total emissions

	LOG (SCOPE 1 TOT)		LOG (SCOPE 2 TOT)		LOG (SCOPE 3 TOT)	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
MKTRF		−1.176		3.298***		3.429**
		(0.714)		(1.084)		(1.357)
HML		−6.020***		−4.284**		−6.444**
		(1.598)		(1.759)		(2.537)
SMB		−0.331		1.184		1.539
		(0.887)		(2.858)		(1.840)
MOM		0.399		−3.853**		−3.580***
		(0.559)		(1.721)		(1.281)
CMA		0.086***		0.053		0.116***
		(0.028)		(0.036)		(0.036)
BAB		0.772		0.303		1.581
		(0.824)		(1.749)		(1.681)
LIQ		2.658***		0.816		3.094***
		(0.768)		(1.135)		(1.016)
NET ISSUANCE		1.250		−1.603		0.376
		(1.015)		(2.207)		(2.352)
IDIO VOL		1.566**		0.986		0.414
		(0.723)		(1.332)		(1.319)
Constant	0.058**	0.053**	0.085**	0.070***	0.103***	0.065**
	(0.026)	(0.023)	(0.037)	(0.027)	(0.035)	(0.027)
Industry adj.	No	No	No	No	No	No
Adj. R2	0.001	0.331	0.001	0.335	0.001	0.247
Observations	156	156	156	156	156	156

Panel B: Growth rate in total emissions

	ΔSCOPE 1		ΔSCOPE 2		ΔSCOPE 3	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
MKTRF		4.847		−2.463		8.303
		(5.605)		(2.516)		(8.965)
HML		−8.427**		−5.897*		−17.483**
		(3.853)		(3.362)		(7.113)
SMB		−15.284**		−9.960*		−23.109*
		(6.419)		(5.667)		(13.738)
MOM		3.223		3.703		9.171
		(4.704)		(2.727)		(8.912)
CMA		−0.159*		−0.153***		−0.468***
		(0.087)		(0.058)		(0.168)
BAB		−8.919***		2.396		11.861
		(3.255)		(2.036)		(8.199)
LIQ		0.808		−1.343		9.512*
		(2.495)		(2.342)		(4.847)
NET ISSUANCE		4.702		1.724		15.976
		(5.262)		(4.821)		(13.211)
IDIO VOL		3.851		6.477*		16.111
		(6.820)		(3.474)		(11.811)
Constant	0.640***	0.643***	0.435***	0.463***	1.559***	1.424***
	(0.089)	(0.120)	(0.065)	(0.063)	(0.237)	(0.250)
Industry adj.	No	No	No	No	No	No
Adj. R2	0.001	0.107	0.001	0.178	0.001	0.290
Observations	144	144	144	144	144	144

(*continued on next page*)

Table 10
(*continued*)

Panel C: Emission intensity						
	SCOPE 1 INT		SCOPE 2 INT		SCOPE 3 INT	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
MKTRF		−0.793***		1.790		0.820
		(0.177)		(2.810)		(0.880)
HML		−0.927***		−6.181		−4.063**
		(0.315)		(4.340)		(1.635)
SMB		−1.027**		−9.486		−0.722
		(0.519)		(6.371)		(1.214)
MOM		0.855***		−1.195		−0.449
		(0.214)		(2.970)		(0.597)
CMA		0.001		0.008		0.039
		(0.007)		(0.101)		(0.031)
BAB		0.302		−4.055		−0.645
		(0.391)		(3.961)		(0.915)
LIQ		0.229		0.372		2.608***
		(0.297)		(2.942)		(0.800)
NET ISSUANCE		0.445		−6.006		−0.139
		(0.304)		(5.742)		(1.159)
IDIO VOL		0.333		8.908***		0.424
		(0.293)		(3.069)		(0.723)
Constant	−0.006	−0.004	0.121	0.181*	0.018	0.012
	(0.008)	(0.007)	(0.102)	(0.097)	(0.027)	(0.028)
Industry adj.	No	No	No	No	No	No
Adj. R2	0.001	0.413	0.001	0.135	0.001	0.104
Observations	156	156	156	156	156	156

cross-sectional standard deviation of ownership for investment advisers, insurance companies, and pension funds, respectively. In particular, given its large aggregate shares of stock holdings, the effect through investment advisers could lead to significant pricing effects. In sharp contrast to the results for *SCOPE 1 INT*, we observe that (with the exception of banks loading up positively on *SCOPE 3 INT*) all coefficients for the different investor types are small and statistically insignificant, which suggests that institutional investors do not seem to discriminate between stocks with regard to their scope 2 and scope 3 emission intensities.

Overall, institutional investors do significantly divest from firms associated with high *SCOPE 1 INT*. They do not screen companies based on the level of their emissions (or growth in emissions), even though the carbon premium is associated with these variables. They prefer to screen firms based on how efficiently they use fossil fuel energy and do not seem to be concerned about reducing their exposure to the level of carbon emissions per se. We conclude from these findings that, unlike for "sin" stocks (as shown by Hong and Kacperczyk, 2009), limited risk sharing caused by divestment cannot alone explain why we observe a return premium for companies with higher levels (and growth) of emissions.

3.5. Coarse categorization

It is often pointed out that only a handful of industries produce the most significant fraction of carbon emissions.[14] The typical salient industries that are mentioned are oil & gas (GIC = 2), utilities (GIC = 65–69), and transportation (GIC = 19, 20, and 23). It is therefore natural to wonder whether our results are disproportionately driven by these sectors, and whether our cross-sectional carbon premium would become significantly smaller if we exclude these industries from our analysis.

In Table 12, we report the results for the subset of firms, excluding the sectors mentioned above. Panel A reports the results for total emissions, Panel B for the growth rate in emissions, and Panel C for emission intensity. Compared with the results in Table 8, we observe that, if anything, excluding these salient sectors strengthens the results on the firm-level carbon premium. These findings imply that there is a coarser categorization of companies by investors within the salient industries, where returns are less sensitive to differences in emissions across firms.

In Table 13, we report the results on institutional ownership when the salient high-CO_2 industries are excluded. Consistent with Gabaix (2014), we find that coarse industry-level categorization drives our divestment results. Indeed, there is no significant divestment in the other industries. This is true in the aggregate as well as for the different categories of investors. It is as if investors decided to reduce their exposure to certain industries by divesting from some firms but holding on to the best in class in terms of scope 1 emission intensity in those industries. In Table A.5 of the Online Appendix, we provide additional evidence on this result with respect to levels and changes in emissions. We do not observe any divestment based on levels of emissions, but some divestment based on the

[14] For instance, in a 2016 report, the International Energy Agency estimates that 39% of CO_2 emissions come from electricity and heat production, 30% from transport, and 11% from industrial production (see https://www.iea.org/media/statistics/Energy_and_CO2_Emissions_in_the_OECD.pdf).

Table 11
Carbon emissions and institutional ownership.
The sample period is 2005–2017. The dependent variable in Panel A is *IO*. The dependent variables in Panel B, Panel C, and Panel D are *IO_BANK, IO_INSURANCE, IO_INVESTCOS, IO_ADVISERS, IO_PENSIONS,* and *IO_HFS.* Panels A-D present the result for contemporaneous measures of emission intensity. Panel B presents the results for *SCOPE 1*, Panel C presents the results for *SCOPE 2*, and Panel D presents the results for *SCOPE 3*. All variables are defined in Table 1. We report the results of the pooled regression with standard errors clustered at the industry and year level (in parentheses). All regressions include year-month fixed effects. In Panel A, the regressions for columns 2, 4, and 6 additionally include state-fixed effects. All regressions in Panels B-D include state fixed effects. ***1% significance; **5% significance; *10% significance.

Panel A: Aggregate ownership (Emission intensity)

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	(0.085)	(0.083)				
SCOPE 2 INT			−0.383	−0.381		
			(1.621)	(1.610)		
SCOPE 3 INT					0.094	−0.130
					(0.550)	(0.581)
LOGSIZE	2.078	1.847	2.096	1.859	2.104	1.850
	(1.510)	(1.702)	(1.484)	(1.678)	(1.499)	(1.706)
PRINV	−29.353***	−37.098***	−29.333***	−37.161***	−29.308***	−37.200***
	(5.614)	(6.448)	(5.611)	(6.392)	(5.640)	(6.476)
MOM	−1.453	−1.792*	−1.542	−1.871**	−1.544	−1.858*
	(0.937)	(0.876)	(0.895)	(0.823)	(0.920)	(0.856)
B/M	−1.165	−0.890	−1.533	−1.205	−1.498	−1.216
	(1.423)	(1.602)	(1.366)	(1.541)	(1.339)	(1.549)
BETA	9.123***	9.470***	9.332***	9.705***	9.300***	9.695***
	(1.508)	(1.459)	(1.421)	(1.375)	(1.430)	(1.388)
VOLAT	−7.617	4.118	−6.867	4.770	−7.095	4.532
	(14.257)	(12.827)	(13.550)	(11.939)	(14.024)	(12.565)
VOLUME	−4.427***	−4.612**	−4.379***	−4.568**	−4.389***	−4.582**
	(1.400)	(1.636)	(1.422)	(1.650)	(1.378)	(1.626)
NASDAQ	−1.159	−1.529	−0.875	−1.255	−0.751	−1.292
	(1.467)	(1.700)	(1.431)	(1.638)	(1.303)	(1.505)
SP500	2.559	1.711	2.418	1.508	2.394	1.510
	(2.120)	(2.093)	(2.122)	(2.088)	(2.129)	(2.095)
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effect	No	Yes	No	Yes	No	Yes
Observations	170,701	160,406	170,701	160,406	170,701	160,406
R-squared	0.121	0.166	0.118	0.162	0.118	0.162

Panel B: Disaggregate ownership

Variables	(1) Banks	(2) Insurance	(3) Invest. Cos.	(4) Advisers	(5) Pensions	(6) Hedge Funds
SCOPE 1 INT	0.001**	−0.011*	0.026	−0.258***	−0.009*	0.033
	(0.000)	(0.005)	(0.022)	(0.056)	(0.004)	(0.028)
SCOPE 2 INT	0.009	−0.253	−0.139	−0.156	0.049	0.108
	(0.006)	(0.144)	(0.406)	(0.992)	(0.097)	(0.441)
SCOPE 3 INT	0.004*	−0.021	0.038	0.052	0.028	−0.230
	(0.002)	(0.071)	(0.115)	(0.409)	(0.030)	(0.151)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	160,406	160,406	160,406	160,406	160,406	160,406

growth of scope 2 and scope 3 emissions. This divestment is particularly strong for pension funds.

3.6. Investor awareness

The carbon premium in stock returns could also be affected by the changing awareness of investors about carbon risk. In particular, one would expect that periods with greater climate change awareness would have a higher carbon premium. We evaluate this hypothesis in two ways. First, we compare the estimated carbon premium before and after the Paris Agreement in 2015. Second, we impute carbon emissions in the 1990s based on their levels in 2005 and estimate the carbon premium over this decade and compare this premium to the one obtained over our sample period, when similarly imputing back carbon emissions based on the levels of emissions in 2017. Both tests offer complementary views on the role of changing investors' attention. The first test allows us to assess the short-term impact of changing attention, while the second test is more suited for the long-term changes in attention.

The Paris Agreement possibly raised both the awareness of risks tied to carbon emissions and the prospect of regulatory interventions to limit carbon emissions. One could therefore expect that the carbon risk premium would increase after 2015 following the Paris Agreement. We test

Table 12
Carbon emissions and stock returns: excluding salient industries.

The sample period is 2005–2017. The dependent variable is *RET*. All variables are defined in Table 1. We report the results of the pooled regression with standard errors clustered at the industry level (in parentheses). The sample excludes companies in the oil and gas (gic=2), utilities (gic=65–69), and transportation (gic=18, 19, 23) industries All regressions include year-month fixed effects. In the regressions for columns 4–6, we additionally include industry-fixed effects. Panel A reports the results for the natural logarithm of total emissions; Panel B reports the results for the percentage change in carbon emissions; Panel C reports the results for carbon emission intensity. ***1% significance; **5% significance; *10% significance.

Panel A: Total emissions						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
LOG (SCOPE 1 TOT)	0.072**			0.177***		
	(0.025)			(0.044)		
LOG (SCOPE 2 TOT)		0.097**			0.227***	
		(0.039)			(0.057)	
LOG (SCOPE 3 TOT)			0.117**			0.324***
			(0.048)			(0.074)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	164,094	164,166	164,190	164,094	164,166	164,190
R-squared	0.213	0.213	0.213	0.216	0.216	0.216

Panel B: Growth rate in total emissions						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
ΔSCOPE 1	0.657***			0.630***		
	(0.151)			(0.142)		
ΔSCOPE 2		0.463***			0.438***	
		(0.117)			(0.112)	
ΔSCOPE 3			1.480***			1.456***
			(0.321)			(0.322)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	135,522	135,570	135,594	135,522	135,570	135,594
R-squared	0.230	0.230	0.230	0.233	0.233	0.233

Panel C: Emission intensity						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
SCOPE 1 INT	0.004			−0.012		
	(0.016)			(0.016)		
SCOPE 2 INT		0.154			0.150	
		(0.102)			(0.112)	
SCOPE 3 INT			0.054			0.160*
			(0.035)			(0.078)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	164,190	164,190	164,190	164,190	164,190	164,190
R-squared	0.213	0.213	0.213	0.216	0.216	0.216

this hypothesis by estimating the regression model in (1) on the two sub-periods: 2005–2015, and 2016–2017.[15] We report the results in Table 14. We find that indeed the premium associated with all three categories of emissions is larger during the 2016–2017 subperiod, especially for scope 1 and scope 2. This could be seen as evidence that investors care more about carbon risk following the Paris Agreement. However, an important caveat is that our sample increases after 2015, so that the difference in returns pre and post Paris could be attributed to the new firms that were added to our sample. We explore this possibility in the Online Appendix and indeed find in Table A.6 that the increase in return premium is mostly due to the addition of the new firms. We find that when we exclude the new firms, the carbon premium becomes insignificant in the two years following the Paris Agreement. The insignificance of the carbon premium does not necessarily mean that carbon risk is no longer priced after the Paris Agreement in 2015; it could be due to a weak statistical power given how noisy returns tend to be.

We further explore the cross-sectional variation of the effect of the Paris Agreement by examining whether the awareness that the Paris Agreement raised had a differential effect on the returns of firms with different exposures to carbon policy risk. We measure the exposures using our three measures of firm-level emissions. Our treatment

[15] To enhance the statistical robustness of our results, we now cluster standard errors at the firm and year-month levels.

Table 13
Carbon emissions and institutional ownership: excluding salient industries.
The sample excludes companies in the oil & gas (gic=2), utilities (gic=65–69), and transportation (gic=18, 19, 23) industries. The sample period is 2005–2017. Panel A presents the results for aggregate ownership for contemporaneous carbon intensity measures, Panel B for disaggregated ownership. All variables are defined in Table 1. We report the results of the pooled regression with standard errors clustered at the industry and year level (in parentheses). All regressions include year-month fixed effects. In Panel A, the regressions for columns 2, 4, and 6, the regressions additionally include state-fixed effects. All regressions for Panel B results include state fixed effects. ***1%; **5%; *10% significance.

Panel A: Aggregate ownership

Variables	(1)	(2)	(3)	(4)	(5)	(6)
SCOPE 1 INT	−0.015	−0.007				
	(0.094)	(0.104)				
SCOPE 2 INT			−0.565	−0.525		
			(1.968)	(2.024)		
SCOPE 3 INT					0.421	0.246
					(0.538)	(0.568)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effect	No	Yes	No	Yes	No	Yes
Observations	152,799	143,337	152,799	143,337	152,799	143,337
R-squared	0.126	0.169	0.126	0.169	0.127	0.170

Panel B: Disaggregate ownership

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Banks	Insurance	Invest. Cos.	Advisers	Pensions	Hedge Funds
SCOPE 1 INT	0.001*	−0.013	−0.059	−0.060	0.009	0.114
	(0.000)	(0.012)	(0.041)	(0.078)	(0.010)	(0.068)
SCOPE 2 INT	0.006	−0.298*	−0.320	−0.224	0.051	0.261
	(0.006)	(0.164)	(0.487)	(1.252)	(0.124)	(0.523)
SCOPE 3 INT	0.004*	−0.015	0.063	0.436	0.041	−0.282
	(0.002)	(0.077)	(0.125)	(0.376)	(0.031)	(0.170)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	143,337	143,337	143,337	143,337	143,337	143,337

sample is the subset of firms in the largest quartile of the distribution of firms sorted by the size of their carbon emission as of the end of 2014. We match these firms with a control group of firms with similar characteristics identified by two different techniques: the nearest neighbor and the Mahalanobis distance. The matching characteristics we use are the same as those we include in our return regressions. We report the results based on the nearest neighbor matching in Table 15. The results based on Mahalanobis matching are qualitatively similar.

To validate the quality of our matching, in Table A.7, we show, as an example, the balance test for the matched samples of treatment and control firms based on the scope 1 emission levels. We find that the two samples are not very different from each other along many firm-level dimensions. Notable exceptions are market capitalization, book-to-market ratio, return on equity, and property plant and equipment for which the differences are statistically significant, though not economically large. Importantly, the differences in market capitalization and PPE are expected given that the treatment sample is based on the size of firm emissions, which are strongly correlated with both characteristics. Next, we compare the returns of firms in the treatment and control groups in the one-year period around the Paris Agreement of December 2015. Formally, we estimate the following difference-in-differences regression model:

$$RET_{i,t} = e_0 + e_1 TREAT * AFTER_{j,t} + e_2 Controls_{i,t} + e_3\mu_i + e_4\mu_t + \varepsilon_{i,t}, \tag{4}$$

where *TREAT* is a generic indicator variable taking the value one for firms in the treatment sample and zero for firms in the control sample, and *AFTER* is an indicator variable equal to zero for the period 2015/05–2015/11 and equal to one for the period 2015/12–2016/05. We also include firm and year-month fixed effects in the regression. We estimate this model separately for each scope and emission measure. In the regressions, the sorts correspond to each scope measure, which then separately identify each individual treatment variable. Our coefficient of interest is e_1, which measures the differential effect of the change on firms with high emissions and firms with low emissions.

In Panel A of Table A.7, we present the results for the level of total emissions. We find a strong and positive effect on returns based on scope 1 emissions, but no significant effects for the other two scopes of emissions. The effect is economically large, implying that the Paris Agreement resulted in an average increase in returns of more than 10.6% over the six-month period. In Panel B, we show the results based on changes in emissions. While the magnitudes of the results for scope 1 and scope 3 based on the model with industry fixed effects are fairly large

Table 14
Carbon emissions and stock returns: sub-periods.

The sample period is 2005–2017. The dependent variable is *RET*. All variables are defined in Table 1. We report the results of the pooled regression with standard errors clustered at the firm and year/month level (in parentheses). All regressions include year-month fixed effects and industry-fixed effects. We report the results for the natural logarithm of contemporaneous total emissions in Panel A; the results for the growth rate in firm emissions in Panel B; and the results for emission intensity in Panel C. ***1% significance; **5% significance; *10% significance.

Panel A: Total emissions

	2005–2015			2016–2017		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
LOG (SCOPE 1 TOT)	0.127***			0.205**		
	(0.037)			(0.075)		
LOG (SCOPE 2 TOT)		0.127***			0.233**	
		(0.042)			(0.087)	
LOG (SCOPE 3 TOT)			0.265***			0.340***
			(0.086)			(0.107)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	121,694	121,622	121,778	62,594	62,594	62,606
R-squared	0.268	0.269	0.269	0.115	0.115	0.115

Panel B: Growth rate in total emissions

	2005–2015			2016–2017		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
ΔSCOPE 1	0.610***			0.629**		
	(0.161)			(0.249)		
ΔSCOPE 2		0.265***			0.459**	
		(0.097)			(0.193)	
ΔSCOPE 3			1.259***			1.032**
			(0.355)			(0.436)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	108,888	108,804	108,948	44,163	44,151	44,175
R-squared	0.278	0.279	0.279	0.089	0.089	0.089

Panel C: Emission intensity

	2005–2015			2016–2017		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
SCOPE 1 INT	0.005			0.010		
	(0.007)			(0.019)		
SCOPE 2 INT		0.091			0.117	
		(0.094)			(0.125)	
SCOPE 3 INT			0.030			0.040
			(0.091)			(0.087)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	121,778	121,778	121,778	62,606	62,606	62,606
R-squared	0.268	0.268	0.268	0.114	0.114	0.114

(between 3.7% and 4.5%), they are statistically insignificant. In Panel C, we present the results based on carbon intensity. Surprisingly, we find a strong negative coefficient for scope 3 emissions. The effects for the other two scopes are small and insignificant. Overall, these results on the differential cross-sectional effects of the Paris Agreement are broadly consistent with our other results but their statistical significance is relatively small. Again, one of the reasons could be the relatively small statistical power of the tests, as returns are generally quite noisy. Another reason could be that the salient effects, such as Paris Agreement, take a longer time to materialize in investors' beliefs.

To offer a longer-term perspective on the changing investors' beliefs, we exploit the fact that climate change and carbon emissions were not yet salient issues in the 1990s. It is only in the last two decades, with the accumulation of CO_2 in the atmosphere and the repeated record-breaking temperatures, that climate change has turned into a widespread concern. This naturally raises the question of whether stock returns were already affected by corporate carbon emissions in the 1990s. If information about firm-

Table 15
Paris Agreement and stock returns: difference-in-differences estimation.
The dependent variable is *RET*. Our treatment sample is the subset of firms in the largest quartile of the distribution of firms sorted by the size of their carbon emission as of the end of 2014. We match these firms with a control group of firms with similar characteristics identified by the nearest neighbor method. The matching characteristics we use are the same as those in our return regressions. *TREAT* is a generic indicator variable taking the value one for firms in the treatment sample and zero for firms in the control sample, and *AFTER* is an indicator variable equal to zero for the period 2015/05–2015/11 and equal to one for the period 2015/12–2016/05. We estimate this model separately for each scope and emission measure. In the regressions, the sorts correspond to each scope measure which then separately identify each individual treatment variable. We also include firm and year-month fixed effects in the regression. All variables are defined in Table 1. Standard errors (in parentheses) are clustered at the firm and year/month level. All regressions for columns 4–6 include industry-fixed effects. We report the results for the natural logarithm of contemporaneous total emissions in Panel A; the results for the growth rate in firm emissions in Panel B; and the results for emission intensity in Panel C. ***1% significance; **5% significance; *10% significance.

Panel A: Total emissions						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
TREAT1*AFTER	10.615***			10.705***		
	(1.175)			(1.200)		
TREAT2*AFTER		−1.783			−1.681	
		(5.861)			(5.821)	
TREAT3*AFTER			−8.917			−8.782
			(6.081)			(6.127)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	5452	6604	6604	5452	6604	6324
Panel B: Growth rate in total emissions						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
TREAT1*AFTER	0.438			4.425		
	(4.426)			(3.373)		
TREAT2*AFTER		−3.712			0.361	
		(3.541)			(2.592)	
TREAT3*AFTER			0.396			3.671
			(4.338)			(3.927)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	5764	5706	5901	5764	5706	5901
Panel C: Emission intensity						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
TREAT1*AFTER	2.825			2.855		
	(5.876)			(5.994)		
TREAT2*AFTER		−0.016			0.021	
		(5.344)			(5.417)	
TREAT3*AFTER			−7.614***			−7.749***
			(2.070)			(2.128)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	4540	4853	4736	4540	4853	4736

level emissions was scarce and/or investors did not pay attention to carbon risk, one would expect that the pricing effects we identify between 2005 and 2017 would be much smaller back then. Given that our carbon emissions data begins in 2005, we cannot evaluate this hypothesis directly. However, we can impute back the unobserved emissions data for each firm in the 1990s from the values we observe later on. In other words, since emission levels are highly autocorrelated and the cross-sectional variation in emissions is stable over time (see Fig. 3), it seems reasonable, as a first pass, to assume that the cross-sectional variation of emissions in the 1990s tracks closely that observed in our data.

Specifically, we make the assumption that each firm with stocks trading during the 1990s has an emission intensity equal to the first officially reported value in the 2005–2017 period. We then collect the time-series information on each company's revenues for the 1990–1999 period and impute the total value of emissions for each firm by taking the product of the emission intensity coefficient and the firm's time-varying sales. We thus obtain a panel of imputed total corporate emissions for 1990–1999. We

Table 16
Carbon emissions and stock returns (imputed emissions).

The sample period is 1990–1999. The dependent variable is *RET*. All variables are defined in Table 1. We report the results of the pooled regression with standard errors clustered at the firm and year level (in parentheses). All regressions include year-month fixed effects. In the regressions for columns 4 through 6, we additionally include industry-fixed effects. The total level of emissions is imputed using the earliest observed level of emission intensity for each firm for the period 2005–2017 (in Panel A) and for 1990–1999 (in Panel B) and scaling it by respective revenue values. ***1% significance; **5% significance; *10% significance.

Panel A: (2005–2017)						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
LOG (SCOPE 1 TOT)	0.097***			0.291***		
	(0.024)			(0.046)		
LOG (SCOPE 2 TOT)		0.186***			0.336***	
		(0.043)			(0.065)	
LOG (SCOPE 3 TOT)			0.245***			0.585***
			(0.043)			(0.127)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	161,122	161,062	161,313	161,122	161,062	161,313
R-squared	0.199	0.200	0.200	0.203	0.203	0.204
Panel B: (1990–1999)						
Variables	(1)	(2)	(3)	(4)	(5)	(6)
LOG (SCOPE 1 TOT)	−0.037			0.082		
	(0.034)			(0.078)		
LOG (SCOPE 2 TOT)		0.033			0.236	
		(0.045)			(0.134)	
LOG (SCOPE 3 TOT)			0.005			0.318*
			(0.059)			(0.162)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/month F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	No	No	Yes	Yes	Yes
Observations	59,878	59,878	59,878	59,878	59,878	59,878
R-squared	0.149	0.149	0.149	0.156	0.156	0.156

do exactly the same for emissions over our sample period. That is, we take the emission intensity coefficient for 2017 and impute back total emissions over the 2017–2005 sample period by multiplying this coefficient with the firm's sales year by year. This latter imputation has the additional benefit of adding imputed emissions to our sample for all the new firms added to our sample in 2016 and provides another robustness check of our findings.

Next, we estimate the regression model in (1) using the imputed emission values for both time periods and report the results in Table 16. The process of imputation is not suitable to obtain the variation in emission growth rates since changes in emissions would vary one to one with changes in revenues. We have therefore considered an alternative model in which we have fixed the growth rates at the first available reported value and used it for all dates in the 1990–1999 period. The results from this estimation, available upon request, indicate that again the carbon premium is insignificant. The results in Panel A for the period of our sample indicate that this imputation works and that there is a significant carbon premium associated with the imputed level of emissions for all three scope categories. Notably, the magnitude of the results is even stronger than for the reported emission data. In contrast, the results in Panel B for the 1990s indicate that there was no significant carbon premium over this period. This finding is consistent with the quite plausible view that investors did not yet internalize carbon risk over this time period, but began to do so in the last two decades, as reporting on climate change, the effects of global warming, technological progress in renewable energy, and political action to curb carbon emissions intensified.

3.7. Robustness

We have explored a number of alternatives that provide insight on the effects we document. We report the specific tables in the Online Appendix. Below, we briefly summarize the main findings in these tables.

First, we estimate the carbon premium excluding the period of the financial crisis, which we define as the period from August 2007 to July 2009. The reason for excluding the financial crisis is that during this period the level of emissions is artificially low because of the crisis and stock returns are highly volatile. As a result, the relation between stock returns and carbon emissions may be distorted by the observations from the crisis period. Broadly, we find in Table A.8 that excluding the crisis period does not affect our results in a major way.

Second, we explore the robustness of our results to the alternative GIC 6-digit industry classification. How much does this alternative classification affect changes in the estimates when industry fixed effects are included? Again, the results, reported in Table A.9, are broadly similar to those obtained under the finer Trucost industry classification. Third, we exclude the salient industries from our analysis of the carbon premium pre and post Paris Agreement. The results are reported in Table A.10. If anything, the increase in the size of the premium is more pronounced in the non-salient industries (with the exception, possibly, of scope 3 emissions).

Fourth, we split the sample into two categories of firms, those that report their emissions and those for which emissions are estimated, and contrast how the carbon premium varies across the two categories. The results are reported in Table A.11. The coefficient for the level of scope 1 emissions is slightly smaller and slightly less significant for firms that disclose their emissions than for firms that do not. This is not entirely surprising given that, other things equal, firms are more likely to disclose their emissions if their performance on that dimension is better. Alternatively, firms that go out of their way to disclose may also have taken steps to reduce their emissions.[16] Overall, the carbon premium is larger and more significant for the firms that do not disclose their emissions for all categories of emissions and for both emission levels and the growth in emissions (with one exception for scope 3 emission levels).

Fifth, we estimate the premium associated with the level and intensity of all three categories of emissions added up together. This is to facilitate comparison with the results in Garvey et al. (2018) and In et al. (2019). As one might expect based on our results for the disaggregated emissions, there is a highly significant premium associated with the level of emissions, but not with emission intensity. The results are presented in Table A.12. Sixth, we also report how institutional investor portfolios are not underweight companies with high levels of emissions (or high growth rates) in Table A.4. If anything, institutional investors load up on scope 2 and scope 3 emission levels. This could be a mechanical effect of their exclusionary screening policies based on scope 1 emission intensity.

Seventh, we further report how institutional investor portfolios are affected by the level of emissions in the companies they hold outside the salient industries tied to fossil fuels. We report the results in Table A.5. Interestingly, institutional investor portfolios load up on all three scope emission levels in the non-salient industries. Again, this is likely the consequence of institutional investors' exclusionary screening in the salient industries. If they hold less in these industries, they must hold more in other industries. Table A.13 also reports the exposure to emission levels of institutional investors' portfolios in the salient industries. Here we observe that their portfolios do not exhibit a significant tilt away from or into firms with high emission levels (with the exception of scope 3 emissions, where they are significantly underweight).

[16] The magnitudes of the coefficients of the estimated emissions could also be affected by measurement error. In general, such measurement error leads to attenuation bias; irrespective of the direction of the bias our comparisons should be treated with caution in the absence of a systematic adjustment for such an error.

Eighth, we explore how sensitive the carbon premium is to the addition of other firm characteristics besides size. Table A.2 reports the results. It turns out that, controlling for other firm characteristics such as B/M, PPE, leverage, etc. matters. Without these controls, there is no significant premium associated with the level of emissions; however, the growth in emissions remains highly significant). Note also that when we add industry fixed effects, adding size as a control or not affects results, with a significant premium associated with the level of scope 1 emissions appearing only when we control for size.

Ninth, we check the robustness of our ownership regressions with respect to outliers using the natural logarithm transformation. The results, in Table A.14, indicate that there is no significant difference compared to our baseline results in Table 8. Tenth, we estimate the carbon premium on only the subset of firms for which we have carbon emission data before 2016. The results are reported in Table A.15. Although the size of the premium is a little smaller, it is broadly in line with the one estimated on the full sample.

4. Conclusion

How is climate change affecting stock returns? This is a fundamental question for the burgeoning field of climate change and finance. It is also a fundamental question for policy makers who are seeking to enlist investors in the fight against climate change. We address this question by undertaking a cross-sectional stock returns analysis, with carbon emissions as a firm characteristic, and find robust evidence that carbon emissions significantly and positively affect stock returns. There is a straightforward link between climate change mitigation and the reduction in carbon emissions. Whether through the production of their goods and services, or through the use of their products, firms are differentially affected by policies to curb carbon emissions and by renewable-energy technology shocks. Our evidence suggests that investors are discerning these cross-sectional differences and are pricing in carbon risk. We also find that the carbon premium cannot be explained through a sin stock divestment effect. Divestment takes place in a coarse way in a few industries such as oil and gas, utilities, and automobiles, and is entirely based on scope 1 emission intensity screens. Notably, we find no carbon premium associated with emission intensity. Moreover, outside the salient industries where all the divestment takes place, we find a robust, persistent, and significant carbon premium at the firm level for all three categories of emission levels and growth rates.

References

Andersson, M., Bolton, P., Samama, F., 2016. Hedging climate risk. Financ. Anal. J. 72 (3), 13–32.
Baldauf, M., Garlappi, L., Yannelis, C., 2020. Does climate change affect real estate prices? Only if you believe in it. Rev. Financ. Stud. 33, 1256–1295.
Bernstein, A., Gustafson, M., Lewis, R., 2019. Disaster on the horizon: the price effect of sea level rise. J. Financ. Econ 134 (2), 253–272.
Busch, T., Johnson, M., Pioch, T., 2018. Consistency of Corporate Carbon Emission Data. University of Hamburg Report WWF Deutschland, Hamburg.

Chava, S., 2014. Environmental externalities and cost of capital. Manage. Sci. 60 (9), 2223–2247.
Engle, R., Giglio, S., Lee, H., Kelly, B., Stroebel, J., 2020. Hedging climate change news. Rev. Financ. Stud. 33, 1184–1216.
Giglio, S., Maggiori, M., Rao, K., Stroebel, J., Weber, A., 2018. Climate Change and Long-Run Discount rates: Evidence from Real Estate Chicago Booth Research Paper No. 17-22.
Gabaix, X., 2014. A sparsity-based model of bounded rationality. Q. J. Econ. 129 (4), 1661–1710.
Gennaioli, N., Shleifer, A., 2010. What comes to mind. Q. J. Econ. 125 (4), 1399–1433.
Garvey, G.T., Iyer, M., Nash, J., 2018. Carbon footprint and productivity: does the “E” in ESG capture efficiency as well as environment? J. Invest. Manage. 16 (1), 59–69.
Görgen, M., Jacob, A., Nerlinger, M., Riordan, R., Rohleder, M., Wilkens, M., 2019. Carbon risk. Unpublished working Paper. University of Augsburg.
Hong, H., Kacperczyk, M., 2009. The price of sin: the effects of social norms on markets. J. Financ. Econ. 93 (1), 15–36.
Hong, H., Li, F.W., Xu, J., 2019. Climate risks and market efficiency. J. Econom. 208 (1), 265–281.
Hsu, P.-.H., Li, K., Tsou, C.Y., 2019. The Pollution Premium Unpublished working paper. HKUST.
Ilhan, E., Sautner, Z., Vilkov, G., 2020. Carbon tail risk. Rev. Financ. Stud. forthcoming.
In, S.Y., Park, K.Y., Monk, A., 2019. Is 'being green' Rewarded in the market? An empirical Investigation of Decarbonization and Stock Returns Unpublished working paper. Stanford Global Project Center https://ssrn.com/abstract=3020304 .
Kacperczyk, M., Van Nieuwerburgh, S., Veldkamp, L., 2016. A rational theory of mutual funds' attention allocation. Econometrica 84 (2), 571–626.
Krueger, P., Sautner, Z., Starks, L., 2020. The importance of climate risks for institutional investors. Rev. Financ. Stud. 33, 1067–1111.
Merton, R.C., 1987. A simple model of capital market equilibrium with incomplete information. J. Finance 42, 483–510.
Monasterolo, I., De Angelis, L., 2019. Blind to Carbon risk? An analysis of Stock Market's Reaction to the Paris Agreement. University of Bologna Unpublished working paper.
Murfin, J., Spiegel, M., 2020. Is the risk of sea level rise capitalized in residential real estate? Rev. Financ. Stud. 33, 1217–1255.

Financial Analysts Journal

ISSN: 0015-198X (Print) 1938-3312 (Online) Journal homepage: https://www.tandfonline.com/loi/ufaj20

Hedging Climate Risk

Mats Andersson, Patrick Bolton & Frédéric Samama

Financial Analysts Journal
Volume 72 • Number 3
©2016 CFA Institute

PERSPECTIVES

Hedging Climate Risk

Mats Andersson, Patrick Bolton, and Frédéric Samama

We present a simple dynamic investment strategy that allows long-term passive investors to hedge climate risk without sacrificing financial returns. We illustrate how the tracking error can be virtually eliminated even for a low-carbon index with 50% less carbon footprint than its benchmark. By investing in such a decarbonized index, investors in effect are holding a "free option on carbon." As long as climate change mitigation actions are pending, the low-carbon index obtains the same return as the benchmark index; but once carbon dioxide emissions are priced, or expected to be priced, the low-carbon index should start to outperform the benchmark.

Whether or not one agrees with the scientific consensus on climate change, both climate risk and climate change mitigation policy risk are worth hedging. The evidence on rising global average temperatures has been the subject of recent debates, especially in light of the apparent slowdown in global warming over 1998–2014.[1] The perceived slowdown has confirmed the beliefs of climate change doubters and fueled a debate on climate science widely covered by the media. This ongoing debate is stimulated by three important considerations.

The first and most obvious consideration is that not all countries and industries are equally affected by climate change. As in other policy areas, the introduction of a new regulation naturally gives rise to policy debates between the losers, who exaggerate the costs, and the winners, who emphasize the urgency of the new policy. The second consideration is that climate mitigation has typically not been a "front burner" political issue. Politicians often tend to "kick the can down the road" rather than introduce policies that are costly in the short run and risk alienating their constituencies—all the more so if there is a perception that the climate change debate is not yet fully settled and that climate change mitigation may not require urgent attention. The third consideration is that although the scientific evidence on the link between carbon dioxide (CO_2) emissions and the greenhouse effect is overwhelming, there is considerable uncertainty regarding the rate of increase in average temperatures over the next 20 or 30 years and the effects on climate change. There is also considerable uncertainty regarding the "tipping point" beyond which catastrophic climate dynamics are set in motion.[2] As with financial crises, the observation of growing imbalances can alert analysts to the inevitability of a crash but still leave them in the dark as to when the crisis is likely to occur.

This uncertainty should be understood as an increasingly important risk factor for investors, particularly long-term investors. At a minimum, the climate science consensus tells us that the risks of a climate disaster are substantial and rising. Moreover, as further evidence of climate events linked to human-caused emissions of CO_2 accumulates and global temperatures keep rising, there is an increased likelihood of policy intervention to limit these emissions.[3] The prospect of such interventions has increased significantly following the Paris Climate Change Conference and the unanimous adoption of a new universal agreement on climate change.[4] Of course, other plausible scenarios can be envisioned whereby the Paris agreement is not followed by meaningful policies. From an investor's perspective, there is therefore a risk with respect to both climate change and the timing of climate mitigation policies. Still, overall, investors should—and some are beginning to—factor carbon risk into their investment policies. It is fair to say, however, that there is still little awareness of this risk factor among (institutional)

Mats Andersson is CEO of AP4, Stockholm. Patrick Bolton is the Barbara and David Zalaznick Professor of Business at Columbia University, New York City. Frédéric Samama is deputy global head of institutional clients at Amundi Asset Management, Paris.

Editor's note: The views expressed in this article are those of the authors and do not necessarily reflect the views of the Amundi Group, AP4, or MSCI.

Editor's note: This article was reviewed and accepted by Robert Litterman, executive editor at the time the article was submitted.

investors.[5] Few investors are aware of the carbon footprint and climate impact of the companies in their portfolios. Among investors holding oil and gas stocks, few are aware of the risks they face with respect to those companies' stranded assets.[6]

In this article, we revisit and analyze a simple, dynamic investment strategy that allows long-term passive investors—a huge institutional investor clientele that includes pension funds, insurance and re-insurance companies, central banks, and sovereign wealth funds—to significantly hedge climate risk while essentially sacrificing no financial returns. One of the main challenges for long-term investors is the uncertainty with respect to the timing of climate mitigation policies. To use another helpful analogy with financial crises, it is extremely risky for a fund manager to exit (or short) an asset class that is perceived to be overvalued and subject to a speculative bubble because the fund could be forced to close as a result of massive redemptions before the bubble has burst. Similarly, an asset manager looking to hedge climate risk by divesting from stocks with high carbon footprints bears the risk of underperforming his benchmark for as long as climate mitigation policies are postponed and market expectations about their introduction are low. Such a fund manager may well be wiped out long before serious limits on CO_2 emissions are introduced.

A number of "green" financial indexes have existed for many years. These indexes fall into two broad groups: (1) pure-play indexes that focus on renewable energy, clean technology, and/or environmental services and (2) "decarbonized" indexes (or "green beta" indexes), whose basic construction principle is to take a standard benchmark, such as the S&P 500 or NASDAQ 100, and remove or underweight the companies with relatively high carbon footprints.[7] The "first family" of green indexes offers no protection against the timing risk of climate change mitigation policies. But the "second family" of decarbonized indexes does: An investor holding such a decarbonized index is hedged against the timing risk of climate mitigation policies (which are expected to disproportionately hit high-carbon-footprint companies) because the decarbonized indexes are structured to maintain a low tracking error with respect to the benchmark indexes.

Thus far, the success of pure-play indexes has been limited. One important reason, highlighted in **Table 1**, is that since the onset of the financial crisis in 2007–2008, these index funds have significantly underperformed market benchmarks.

Besides the fact that clean tech has been overhyped,[8] one of the reasons why these indexes have underperformed is that some of the climate mitigation policies in place before the financial crisis have been scaled back (e.g., in Spain). In addition, financial markets may have rationally anticipated that one of the consequences of the financial crisis would be the likely postponement of the introduction of limits on CO_2 emissions. These changed expectations benefited the carbon-intensive utilities and energy companies more than other companies and may explain the relative underperformance of the green pure-play indexes. More importantly, the reach of the pure-play green funds is very limited because they concentrate investments in a couple of subsectors and, in any case, cannot serve as a basis for building a core equity portfolio for institutional investors.

The basic point underlying a climate risk–hedging strategy that uses decarbonized indexes is to go beyond a simple divestment policy or investments in only pure-play indexes and instead keep an aggregate risk exposure similar to that of standard market benchmarks. Indeed, divestment of high-carbon-footprint stocks is just the first step. The second key step is to optimize the composition and weighting of the decarbonized index in order to minimize the tracking error (TE) with the reference benchmark index. It turns out that TE can be virtually eliminated, with the overall carbon footprint of the decarbonized index remaining substantially lower than that of the reference index (close to 50% in terms of both carbon intensities and absolute carbon emissions). Decarbonized indexes have thus far essentially matched or even outperformed the benchmark index.[9] In other words, investors holding a decarbonized index have been able to significantly

Table 1. Pure-Play Clean Energy Indexes vs. Global Indexes

	S&P 500	NASDAQ 100	PP 1	PP 2	PP 3	PP 4	PP 5
Annualized return	4.79%	11.40%	5.02%	–8.72%	2.26%	–8.03%	–1.89%
Annualized volatility	22.3	23.6	24.1	39.3	30.2	33.8	37.3

Notes: Table 1 gives the financial returns of several ETFs that track leading clean energy pure-play indexes. Pure Play 1 refers to Market Vectors Environmental Services ETF, Pure Play 2 to Market Vectors Global Alternative Energy ETF, Pure Play 3 to PowerShares Cleantech Portfolio, Pure Play 4 to PowerShares Global Clean Energy Portfolio, and Pure Play 5 to First Trust NASDAQ Clean Edge Green Energy Index Fund. Annualized return and volatility were calculated using daily data from 5 January 2007 to the liquidation of Pure Play 1 on 12 November 2014.
Sources: Amundi and Bloomberg (1 September 2015).

reduce their carbon footprint exposure without sacrificing any financial returns. In effect, these investors are holding a "free option on carbon": So long as the introduction of significant limits on CO_2 emissions is postponed, they can obtain the same returns as on a benchmark index. But from the day CO_2 emissions are priced meaningfully and consistently and limits on CO_2 emissions are introduced, the decarbonized index should outperform the benchmark.[10] A climate risk–hedging policy around decarbonized indexes is essentially an unlevered minimum risk arbitrage policy that takes advantage of a currently mispriced risk factor (carbon risk) in financial markets. Although larger arbitrage gains are obtainable by taking larger risks (and this climate risk–hedging strategy errs on the side of caution), the strategy is particularly well suited for long-term passive investors who seek to maximize long-term returns while limiting active stock trading over time.

A Green Index without Relative Market Risk: The Basic Concept

Investor perceptions of lower financial returns from green index funds could explain why green indexes have thus far remained a niche market. Another reason might be the design of most green indexes, which lend themselves more to a bet on clean energy than a hedge against carbon risk. In contrast, the design we support allows passive long-term investors to hedge carbon risk. Thus, the goal is not just to minimize exposure to carbon risk by completely divesting from any company with a carbon footprint exceeding a given threshold, but also to minimize the tracking error of the decarbonized index with the benchmark index. We support this design because it implements a true dynamic hedging strategy for passive investors and can easily be scaled to significantly affect not only portfolios' footprints but also (eventually) the real economy.[11]

The basic idea behind index decarbonization is to construct a portfolio with fewer composite stocks than the benchmark index but with similar aggregate risk exposure to all priced risk factors. This approach is possible because, as Koch and Bassen (2013) showed, carbon risk is asymmetrically concentrated in a few firms.[12] Ideally, the only major difference in aggregate risk exposure between the two indexes would be with respect to the carbon risk factor, which would be significantly lower for the decarbonized index. So long as carbon risk remains unpriced by the market, the two indexes will generate similar returns (i.e., offer the same compensation for risk demanded by the representative investor), thus achieving no or minimal TE. But once carbon risk is priced or is expected to be priced by the market, the decarbonized index should start outperforming the benchmark.

The central underlying premise of this strategy is that financial markets currently underprice carbon risk. Moreover, our fundamental belief is that eventually, if not in the near future, financial markets will begin to price carbon risk. Our premise leads inevitably to the conclusion that a decarbonized index is bound to provide higher financial returns than the benchmark index. We believe that the evidence in support of our premise is overwhelming. Currently, virtually all financial analysts overlook carbon risk. Only in 2014 did a discussion about stranded assets make it into a report from a leading oil company for the first time, and the report mostly denied any concern that a fraction of proven reserves might ever become stranded assets.[13] Only a few specialized financial analysts[14] factor stranded assets into their valuation models of oil company stocks. Nor, apart from a few exceptions,[15] do financial analysts ever evoke carbon-pricing risk in their reports to investors. In sum, current analysts' forecasts assume by default that there is no carbon risk. Under these circumstances, it takes a stretch of the imagination to explain that financial markets somehow currently price carbon risk correctly. Even more implausible is the notion that financial markets currently price carbon risk excessively. Only in this latter scenario would investors in a decarbonized index face lower financial returns than in the benchmark index.

Some might object that our fundamental belief that financial markets will price carbon risk in the future is not particularly plausible. After all, the evidence from many climate talks' failures following Kyoto suggests, if anything, that global carbon pricing in the near future is extremely unlikely. If that should be the case, our investor in the decarbonized index would simply match the returns of the benchmark index—a worst-case scenario. Any concrete progress in international negotiations—and the implementation of nationally determined independent contributions agreed to in Paris—will change financial market expectations about carbon risk and likely result in higher financial returns on the low-TE index relative to the benchmark index.

The Decarbonized Index Optimization Problem. Given our basic premise and fundamental belief, the next question is how to go about constructing the green index. There are several possible formulations of the problem in practice. One formulation is to eliminate high-carbon-footprint composite stocks, with the objective of meeting a target carbon footprint reduction for the green index, and then to reweight the remaining stocks in order to minimize tracking error with the benchmark index. The dual formulation is to begin by imposing a constraint on maximum allowable tracking error with the benchmark index and then, subject to this constraint, exclude and reweight composite stocks in the benchmark index to maximize the green index's carbon footprint reduction. Although there is no compelling reason to choose one formulation over the other, we favor the second formulation, which seeks to minimize tracking error subject to meeting a carbon footprint reduction target.

Another relevant variation in the design of the constrained optimization problem is whether to (1) require at the outset the complete exclusion of composite stocks of the worst performers in terms of carbon footprint or (2) allow the green index to simply underweight high-carbon-footprint stocks without completely excluding them. Although the latter formulation is more flexible, it has drawbacks, which we discuss later in the article.

We confine our analysis to essentially two alternatives among the many possible formulations of the constrained optimization problem for the construction of a decarbonized index that trades off exposure to carbon, tracking error, and expected returns. We describe both formulations formally, under the simplifying assumption that only one sector is represented in the benchmark index.

The two portfolio optimization problems can be simply and easily represented. Suppose that there are N constituent stocks in the benchmark index and that the weight of each stock in the index is given by $w_i^b = \left[\frac{\text{Mkt cap}(i)}{\text{Total mkt cap}}\right]$. Suppose next that each constituent company is ranked in decreasing order of carbon intensity, q_l^i, with company $l = 1$ having the highest carbon intensity and company $l = N$ the lowest (each company is thus identified by two numbers $[i,l]$, with the first number referring to the company's identity and the second to its ranking in carbon intensity).

In the first problem, the green portfolio can be constructed by choosing new weights, w_i^g, for the constituent stocks to solve the following minimization problem:

$$\text{Min TE} = sd\left(R^g - R^b\right),$$

where

$$w_j^g = 0 \text{ for all } j = 1,\ldots,k$$

$$0 \le w_i^g \text{ for all } i = k+1,\ldots,N$$

sd = standard deviation

That is, the decarbonized index is constructed by first excluding the k worst performers in terms of carbon intensity and reweighting the remaining stocks in the green portfolio to minimize TE.[16] This decarbonization method follows transparent rules of exclusion, whatever the threshold k.

In the second problem formulation, the first set of constraints $(w_j^g = 0 \text{ for all } j = 1,\ldots,k)$ is replaced by the constraint that the green portfolio's carbon intensity must be smaller than a given threshold: $\sum_{l=1\ldots N} q_l w_l^g \le Q$. In other words, the second problem is a design, which potentially does not exclude any constituent stocks from the benchmark index and seeks only to reduce the carbon intensity of the index by reweighting the stocks in the green portfolio. Although the second problem formulation (pure optimization) dominates the first (transparent rules) for the same target aggregate carbon intensity, Q, because it has fewer constraints, it has a significant drawback in terms of the methodology's opacity and the lack of a clear signal for which constituent stocks to exclude on the basis of their relatively high carbon intensity.

Optimization Procedure. For both problem formulations, the *ex ante* TE—given by the estimated standard deviation of returns of the decarbonized portfolio from the benchmark—is estimated by using a multifactor model of aggregate risk (see Appendix D for more detailed information). This multifactor model significantly reduces computations, and the decomposition of individual stock returns into a weighted sum of common factor returns and specific returns provides a good approximation of individual stocks' expected returns. More formally, under the multifactor model the TE minimization problem has the following structure:

$$\text{Min}\left[\sqrt{\left(W^p - W^b\right)'\left(\boldsymbol{\beta}\boldsymbol{\Omega}_f\boldsymbol{\beta}' + \boldsymbol{\Delta}^{AR}\right)\left(W^p - W^b\right)}\right],$$

where

$$w_l^g = 0 \text{ for all } l = 1,\ldots,k$$

$$0 \le w_l^g \text{ for all } l = k+1,\ldots,N$$

$\left(W^p - W^b\right)$ = the vector of the difference in portfolio weights of the decarbonized portfolio and the benchmark

$\boldsymbol{\Omega}_f$ = the variance–covariance matrix of factors

$\boldsymbol{\beta}$ = the matrix of factor exposures

$\boldsymbol{\Delta}^{AR}$ = the diagonal matrix of specific risk variances

Risk Mitigation Benefits of Low Tracking Error. To explore more systematically the potential benefits of achieving a bounded tracking error, we ran a number of simulations with the pure optimization methodology and determined a

TE–carbon efficiency frontier for a decarbonized index constructed from the MSCI Europe Index. As illustrated in **Figure 1**, achieving a nearly 100% reduction in the MSCI Europe carbon footprint would come at the price of a huge tracking error of more than 3.5%.[17]

Such a large TE would expose investors in the decarbonized index to significant financial risk relative to the benchmark—even in a good scenario whereby the decarbonized index is expected to outperform the benchmark as a result of climate mitigation policies. **Figure 2** depicts the risk that a large TE might expose investors to and how that risk can be mitigated by lowering the TE. We first posit a scenario whereby the expected yearly return of the green index is 2.5% higher than that of the benchmark[18] and show (with a confidence interval of two standard deviations) that a 3.5% TE could expose investors to losses relative to the benchmark in the negative scenario.

As Figure 2 illustrates, if we lower the TE of the decarbonized index from 3.5% to 1.2%, the decarbonized index generates returns at least as high as those of the benchmark *even in the worst-case scenario*.

Illustrative Example. A simple example can illustrate in greater detail how a low-carbon, low-TE index might be constructed and how its financial returns—relative to the benchmark—would vary with (expectations of) the introduction of carbon taxes. Let us consider a portfolio of four stocks (A, B, C, D), each priced at 100. The first two stocks (A, B) are, say, oil company stocks; stock C is outside the oil industry, but its price is perfectly correlated with the oil industry stock price; and stock D is a company whose stock price is uncorrelated with the oil industry. The pre–carbon taxation returns on these stocks are 20%, 20%, 20%, and 30%, respectively. On the one hand, we assume that stocks A and B have a relatively high carbon footprint, which would expose them to relatively high implied carbon taxation—40% and 10% of earnings, respectively. On the other hand, we assume that stocks C and D have no carbon tax exposure. We then construct the low-carbon, low-TE index as follows: (1) We filter out entirely stocks A and B, (2) we treble the weighting of stock C to maintain the same overall exposure to the oil sector as the benchmark portfolio, and (3) we leave the weighting of stock D unchanged. If the introduction of carbon taxes is expected, the price of stock A will drop to 72 and the price of stock B will increase to 108, whereas the price of stock C will increase to 120 and the price of stock D will rise to 130. What are the implications for returns on the low-carbon, low-TE index relative to the benchmark? In this scenario, the low-TE index would outperform the benchmark by 14%.

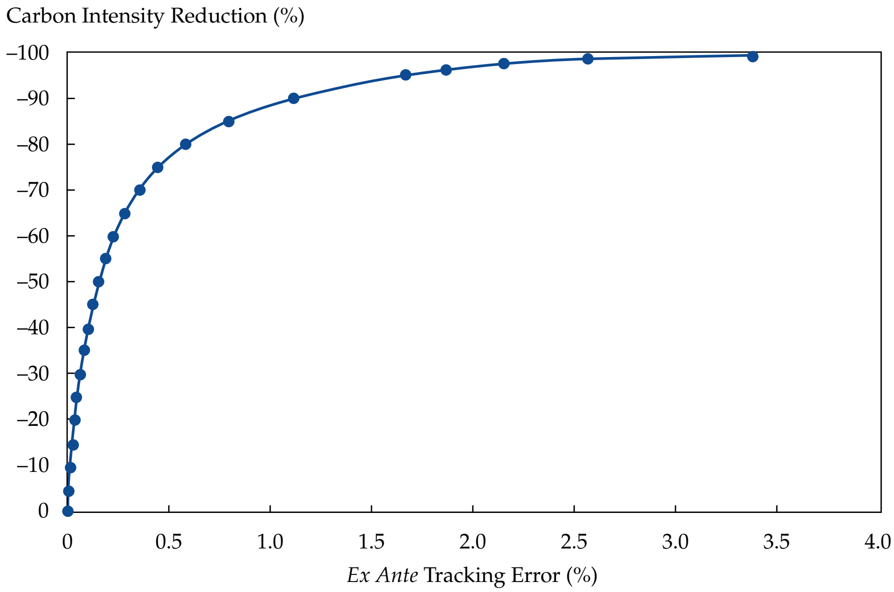
Tracking Error Management and Carbon Risk Repricing. Index managers seek to limit *ex ante* TE. However, some enhanced indexes, such as decarbonized indexes, also seek to increase returns relative to the benchmark. Although the two goals may seem in conflict, we note that the optimization procedure focuses on *ex ante* TE and excess returns are necessarily measured *ex post*. Therefore, if the risk model used to limit *ex ante* TE does not take into account carbon risk (or any factor responsible for a divergence of returns), a small *ex ante* TE can be compatible with active returns *ex post*. Two polar carbon-repricing scenarios can be considered: (1) a smooth repricing with moderate regulatory and technological changes that progressively impair the profitability of carbon-intensive companies and (2) a sharp repricing caused by unanticipated disruptive technologies or regulations. In the first scenario, investors could experience active positive returns with *ex post* TE in line with *ex ante* TE. In the second scenario, investors in a decarbonized index could experience a peak in *ex post* TE with active positive returns.

Beyond Optimization: Methodological Considerations and Caveats

In this section, we consider other issues besides portfolio optimization, including the benefits of clear signaling via transparent rules, trade-offs involved in different designs of decarbonized indexes and different normalizations of carbon footprints, how to deal with anticipated changes in companies' carbon footprints, and a few caveats.

Benefits of Clear Signaling through Transparent Rules. As all issuers well understand, inclusion in or exclusion from an index matters and is a newsworthy event. We believe that inclusion in a decarbonized index ought to have similar value. Clearly communicating which constituent stocks are in the decarbonized index would not only reward the included companies for their efforts in reducing their carbon footprint but also help discipline the excluded companies. This pressure might induce excluded companies to take steps to reduce their carbon footprint and to reward their CEOs for any carbon footprint reductions.[19] Because companies' exclusion from the index would be reevaluated yearly, it would also induce healthy competition to perform well with respect to carbon footprints, with the goal of rejoining the index.[20] Finally, clear communications concerning exclusion criteria based on carbon footprints would inspire a debate on whether greenhouse gas (GHG) emissions are properly measured and would lead to improvements in the

Figure 1. Carbon Frontier on the MSCI Europe Index



Source: Amundi (30 June 2015).

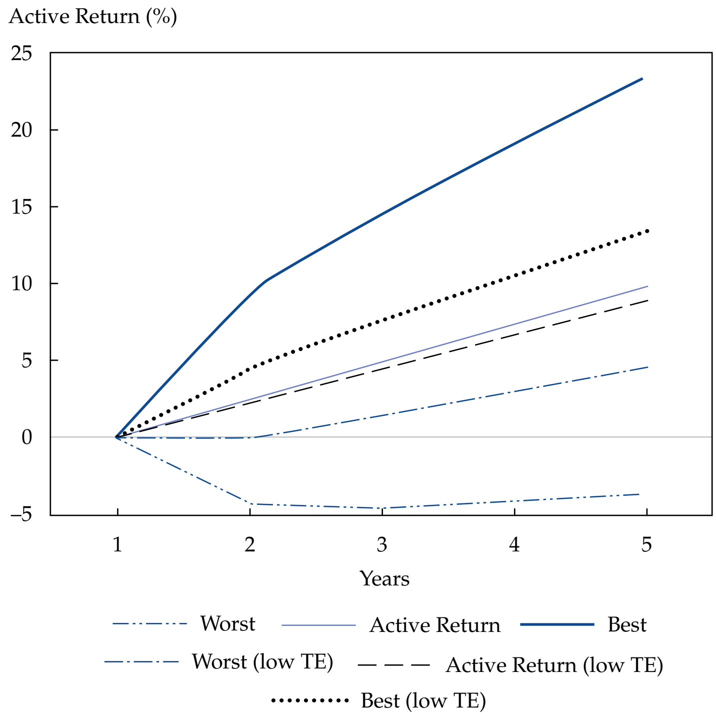
methodology for determining companies' carbon footprints.

Design Trade-Offs. A number of trade-offs are involved in the design of a decarbonized index. For example, an obvious question about balancing concerns the sector composition of the benchmark index. To what extent should the decarbonized index seek to preserve the sector balance of the benchmark? While seeking to preserve sector composition, should the filtering out of high-carbon-footprint stocks be performed sector by sector or across the entire benchmark index portfolio? Some believe that a sector-blind filtering out of companies by the size of their carbon footprint would result in an unbalanced decarbonized index that essentially excludes most of the fossil energy sector, electric utilities, and mining and materials companies. Obviously, such an unbalanced decarbonized index would have a very high tracking error and would be undesirable. Interestingly, however, a study of the world's 100 largest companies has shown that more than 90% of the world's GHG emissions are attributable to sectors other than oil and gas (see Climate Counts 2013). Hence, a sector-by-sector filtering approach could result in a significantly reduced carbon footprint while still maintaining a sector composition roughly similar to that of the benchmark. Later in the article, we show more concretely how much carbon footprint reduction can be achieved by decarbonizing the S&P 500 and MSCI Europe indexes.

One simple way to address this issue is to look at the decarbonized portfolio's TE for the different optimization problems and pick the procedure that yields the decarbonized index with the lowest TE. But there may be other relevant considerations besides TE minimization. For example, one advantage of a sector-by-sector filtering approach with transparent rules (subject to the constraint of maintaining roughly the same sector balance as that of the benchmark index) is that excluded companies can more easily determine their carbon footprint ranking in their industry and how much carbon footprint reduction it would take for their stock to be included in the decarbonized index. In other words, a sector-by-sector filtering approach would foster greater competition within each sector for companies to lower their carbon footprint. Another related benefit is that the exclusion of the worst GHG performers in the sector would also reduce exposure to companies that fare poorly on other material sustainability factors (given that carbon footprint reduction is a good proxy for investments in other material sustainability factors).[21]

Normalization of the Carbon Footprint. Because the largest companies in the benchmark index are likely to be the companies with the highest GHG emission levels, a filtering rule that excludes the stocks of companies with the highest absolute emission levels will tend to be biased against the largest companies, which could result in a high TE for the decarbonized index. Accordingly, some normalization of companies' carbon footprints is appropriate. Another reason to normalize the absolute carbon footprint measure is that a filter based on a normalized measure would be better at selecting the least wasteful companies in terms of GHG emissions. That

Figure 2. Returns and Risk with Low Tracking Error



Source: Amundi.

is, a normalized carbon footprint measure would better select companies on the basis of their energy efficiency. A simple and comprehensive, if somewhat rudimentary, normalization would be to divide each company's carbon footprint by sales. Normalizations adapted to each sector are preferable and could take the form of dividing CO_2 emissions by (1) tons of output in the oil and gas sector, (2) revenue from transporting one tonne over a certain distance in the transport sector, (3) total GWh (gigawatt-hour) electricity production in the electric utility sector, (4) square footage of floor space in the housing sector, or (5) total sales in the retail sector.

Changes in Companies' Carbon Footprints. Ideally, the green filter should take into account expected future carbon footprint reductions resulting from current investments in energy efficiency and reduced reliance on fossil fuels. Similarly, the green filter should penalize oil and gas companies that invest heavily in exploration with the goal of increasing their proven reserves, which raises the risk of stranded assets for such companies. This "threat" would provide an immediate incentive to any company with an exceptionally high carbon footprint to make investments to reduce it and would boost the financial returns of the decarbonized index relative to the benchmark.

Caveats. Whenever an investment strategy that is expected to outperform a market benchmark is pitched, a natural reaction is to ask, what's the catch? As explained earlier, the outperformance of the decarbonized index is premised on the fact that financial markets currently do not price carbon risk. Thus, an obvious potential flaw in our proposed climate risk–hedging strategy is the possibility that financial markets currently *overprice* carbon risk. While this overpricing is being corrected, the decarbonized index would underperform the benchmark index. We strongly believe this argument to be implausible because the current level of awareness of carbon risk remains very low outside a few circles of asset owners, a handful of brokers, and asset managers. Another highly implausible scenario is that somehow today's high-carbon-footprint sectors and companies will be tomorrow's low-carbon-footprint sectors and companies. One story to back such a scenario could be that the high-GHG emitters have the most to gain from carbon sequestration and will thus be the first to invest in that technology. Under this scenario, the decarbonized index would underperform the benchmark precisely when carbon taxes are introduced. This scenario is not in itself a crushing objection because the green filter can easily take into account investments in carbon sequestration as a criterion for inclusion in the index. In the end, this scenario simply suggests a reason for the carbon filter to take into account measures of companies' predicted carbon footprints.

A more valid concern is whether companies' carbon footprints are correctly measured and whether the filtering based on carbon intensity fits its purpose. Is there a built-in bias in the way carbon footprints are measured, or is the measure so noisy that investors could be exposed to many carbon measurement risks? A number of organizations—Trucost, CDP (formerly Carbon Disclosure Project), South Pole Group, and MSCI ESG Research—provide carbon footprint measures of the largest publicly traded companies, measures that can sometimes differ from one organization to another.[22] For example, it has been observed that GHG emissions associated with hydraulic fracturing for shale gas are significantly underestimated because the high methane emissions involved in the hydraulic fracturing process are not counted. Thus, what would appear to be—according to current carbon footprint measurements—a welcome reduction in carbon footprints following the shift from coal to shale gas could be just an illusion. Consequently, a green filter that relies on this biased carbon footprint measure risks exposing investors to more rather than less carbon risk.

As described in greater detail in Appendix C, GHG emissions are divided into three scopes: Scope 1, which measures direct GHG emissions; Scope 2, which concerns indirect emissions resulting from the company's purchases of energy; and Scope 3, which covers third-party emissions (suppliers and consumers) tied to the company's sales. Although Scope 3 emissions may represent the largest fraction of GHG emissions for some companies (e.g., consumer electronics companies and car manufacturers),[23] there is currently no systematic, standardized reporting of these emissions. This lack is clearly a major limitation and reduces the effectiveness of all existing decarbonization methodologies. For example, excluding the most-polluting companies in the automobile industry and the auto components industry on the basis of current emission measures would lead mostly to the exclusion of auto components companies. Automobile manufacturers would largely be preserved because most of the carbon emissions for a car maker are Scope 3 emissions. As reliance on decarbonized indexes grows in scale, however, more resources will likely be devoted to improving the quality of Scope 3 and the other categories of GHG emissions. The inclusion of Scope 3 emissions would also better account for green product innovations by materials companies that bolster the transition toward a low-carbon economy. For instance, aluminum producers might be excluded under the current GHG measures owing to their high carbon intensity even though aluminum will fare better than other materials in the transition to renewable energy.

There are three evident responses to these existing measurement limitations. First, drawing an analogy with credit markets, we know that a biased or noisy measure of credit risk by credit-rating agencies has never been a decisive reason for abolishing credit ratings altogether. Credit ratings have provided an essential reinforcement of credit markets for decades despite important imprecisions in their measurements of credit risk, which have been pointed out by researchers of credit markets over time. Second, as with credit ratings, methodologies for measuring carbon footprints will be improved, especially when the stakes involved in measuring carbon footprints correctly increase because of the role of these measures in any green filtering process. Third, the design of the decarbonized index itself offers protection against carbon footprint measurement risk; if there is virtually no tracking error with the benchmark, investors in the decarbonized index are partly hedged against this risk.

Finally, a somewhat more technical worry is that the stocks excluded from the decarbonized index could also be the most volatile stocks in the benchmark index because these stocks are the most sensitive to speculation about climate change and climate policy. If that is the case, tracking error cannot be eliminated entirely, but that should not be a reason for deciding not to invest in the decarbonized index. On the contrary, the decarbonized index will then have a higher Sharpe ratio than the benchmark, commensurate with a higher TE.[24]

To summarize, our proposed strategy for hedging climate risk is especially suitable for passive long-term investors. Rather than a risky bet on clean energy (at least in the short run), we have described a decarbonized index with minimal tracking error that offers passive investors a significantly reduced exposure to carbon risk, allowing them to "buy time" and limit their exposure with respect to the timing of the implementation of climate policy and a carbon tax. Thus, a key difference between this approach and existing green indexes is switching the focus from the inevitable transition to renewable energy to the timing risk with respect to climate policy. As we show later in the article, carbon exposure can be reduced significantly—with maximum insurance against the timing of climate policy—by minimizing tracking error with the benchmark index. We believe that this approach is essentially a win-win strategy for all passive asset owners and managers. Moreover, should this strategy be adopted by a large fraction of passive index investors—a market representing close to $11 trillion in assets, according to a recent study[25] (Boston Consulting Group 2015)—companies will feel the pressure to improve their performance on GHG emissions and debates about carbon emissions will surely be featured prominently in the financial press.[26] It constitutes, therefore, an easy entry point for a wide clientele of investors and could trigger the mobilization of a much broader ecosystem dedicated to the analysis and understanding of climate-related transition risks.

Decarbonized Indexes in Practice: How Small Are Their Carbon Footprints?

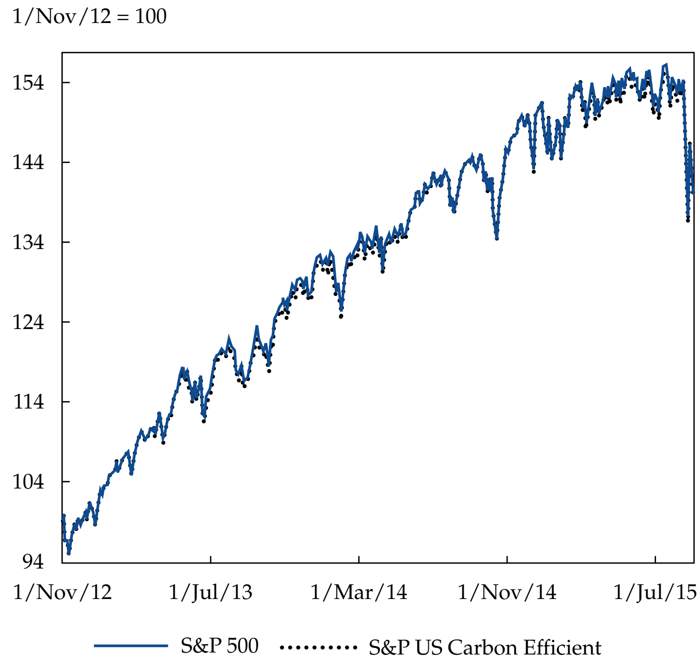
There are several examples of decarbonized indexes. AP4, the Fourth Swedish National Pension Fund (Fjärde AP-fonden), is, to our knowledge, the first institutional investor to adopt a systematic approach that uses some of these decarbonized indexes to significantly hedge the carbon exposure of its global equity portfolio. In 2012, AP4 decided to hedge the carbon exposure of its US equity holdings in the S&P 500 by switching to a decarbonized portfolio with a low TE relative to the S&P 500 through the replication of the S&P 500 Carbon Efficient Select Index. This index excludes the 20% worst performers in terms of carbon intensity (CO_2/Sales) as measured by Trucost, one of the leading companies specializing in the measurement of the environmental impacts of publicly traded companies. An initial design constraint on the decarbonized index is to ensure that stocks removed from the S&P 500 do not exceed a reduction in the Global Industry Classification Standard (GICS) sector weight of the S&P 500 by more than 50%. A second feature of the S&P 500 Carbon Efficient Select Index is the readjustment of the weighting of the remaining constituent stocks to minimize TE with the S&P 500. Remarkably, this decarbonized index reduces the overall carbon footprint of the S&P 500 by roughly 50%,[27] with a TE of no more than 0.5%. This first model of a decarbonized index strikingly illustrates that significant reductions in carbon exposure are possible without sacrificing much in the way of financial performance or TE. In fact, AP4's S&P 500 Carbon Efficient Select Index portfolio has outperformed the S&P 500 by about 24 bps annually since it first invested in the decarbonized index in November 2012, as **Figure 3** shows, which is in line with the 27 bp annual outperformance of the S&P 500 Carbon Efficient Select Index since January 2010.

AP4 has extended this approach to hedging climate risk to its equity holdings in emerging markets.[28] Relying on carbon footprint data from MSCI ESG Research, AP4 has sought to exclude from the MSCI EM Custom ESG Index not only the companies with the highest GHG emissions but also the worst companies in terms of stranded-asset risk. Turning to its Pacific-ex-Japan stock holdings, AP4 has applied a similar methodology in constructing its decarbonized portfolio, excluding the companies with the largest reserves and highest carbon emissions intensity while maintaining both sector and country weights in line with its initial index holdings in the region.

More recently, AP4, FRR (Fonds de réserve pour les retraites, or the French pensions reserve fund), and Amundi have worked with MSCI to develop another family of decarbonized indexes with a slightly different design. The result is the MSCI Global Low Carbon Leaders Index family—based on existing MSCI equity indexes (e.g., MSCI ACWI, MSCI World, and MSCI Europe)—which addresses two dimensions of carbon exposure. It excludes from the indexes the worst performers in terms of both carbon emissions intensity and fossil fuel reserves intensity while maintaining a maximum turnover constraint as well as minimum sector and country weights. The remaining constituent stocks are then rebalanced to minimize TE with the respective benchmarks.[29] **Table 2** compares the performance of the resulting decarbonized indexes, based on a backtest, with that of the MSCI Europe Index. As Table 2 shows, the Low Carbon Leaders Index delivers a remarkable 90 bp annualized outperformance over the MSCI Europe Index for November 2010–February 2016, with a similar volatility and a 0.7% tracking error.

At the end of January 2016, we conducted a performance attribution analysis, after the MSCI Europe Low Carbon Leaders Index was launched, for the period November 2014–January 2016,[30] when the outperformance was particularly strong (an overall 189 bps[31]). Our analysis shows how to distinguish which part of the performance is due to sector allocation (allocation effect[32]) and which part is due to stock selection within sectors (selection effect[33]). At the sector level (using the GICS[34] taxonomy), the allocation effect is responsible for 37 bps of outperformance, with the underweighting of the energy and materials sectors responsible for 40 bps and 20 bps, respectively. More importantly, the effect of screening out the worst GHG performers within a sector is greater than the allocation effect, with a 120 bp outperformance. Interestingly, the positive screening effect is concentrated in two sectors, Materials (127 bps) and Utilities (25 bps; see **Table E1** in Appendix E). The largest negative contributor, Consumer Staples, had an allocation effect of –37 bps and a selection effect of –8 bps.

Figure 3. S&P 500 and S&P US Carbon Efficient Indexes



Sources: Amundi and Bloomberg (31 August 2015).

We conducted a second-level analysis (industry level; see **Table E2** and **Table E3** in Appendix E) that focused on the largest contributor, the materials sector, and found that the index was strongly underweighted in the diversified metals and mining (DM&M) stocks, with a 68 bp allocation effect and a 36 bp selection effect. The reason behind this underweighting is that coal represents the major part of DM&M reserves. As for the utilities sector, the index was underweighted on multi-utilities because of their high emissions (an 11 bp selection effect and an 8 bp allocation effect). Stock performance for these two sectors was related to trends in the energy sector (mostly a fall in coal prices).

AP4, MSCI, FRR, and Amundi have further explored the robustness of these decarbonized indexes to other exclusion rules and to higher carbon footprint reductions. They found that there is not much to be gained by using more flexible criteria that permit less than 100% exclusion of high-carbon-footprint stocks. **Table 3** compares the performances of a fully "optimized" portfolio, with no strict exclusion of the worst performers, and a portfolio based on the "transparent exclusion rules" outlined earlier. Whether in terms of reduced exposure to carbon or overall tracking error, the two portfolios deliver similar results.

Interestingly, however, the two methods for constructing the decarbonized index yield substantial sector differences in TE contribution, which is concentrated in two sectors (materials and energy) for the fully optimized index. In contrast, the limit put on total sector exclusion in the Low Carbon Leaders Index (with transparent rules) spreads the effort across several sectors (see **Figure F1** in Appendix F for a detailed breakdown of the contributions to specific risks).

Conclusion

Our decarbonized index investment strategy stands on its own as a simple and effective climate risk–hedging strategy for passive long-term institutional investors, but it is also an important complement to climate change mitigation policies. Governments have thus far focused mostly on introducing policies to control or tax GHG emissions and to build broad international agreements for the global implementation of such policies (for a discussion of the pros and cons of cap-and-trade mechanisms versus a GHG emissions tax, see Guesnerie and Stern 2012).[35] Governments have also provided subsidies to the solar and wind energy sectors, thereby boosting a small-business constituency that supports climate change mitigation policies. Similarly, index decarbonization can boost support for such policies from a large fraction of the investor community. In addition, as more and more funds are allocated to decarbonized indexes, stronger market incentives will materialize, inducing the

Table 2. Financial Performance of Transparent Rules on MSCI Europe

Key Metrics	MSCI Europe Index	MSCI Europe Low Carbon Leaders Index
Total return[a]	7.8%	8.7%
Total risk[a]	13.2%	13.2%
Return/risk	0.59	0.65
Sharpe ratio	0.57	0.63
Active return[a]	0%	0.9%
Tracking error[a]	0%	0.7%
Information ratio	NA	1.16
Historical beta	1.00	1.16
Turnover[b]	1.8%	9.9%
Securities excluded	NA	93
Market cap excluded	NA	21.4%
Reduction in carbon emissions intensity (tCO$_2$/US$ millions)	NA	52%
Reduction in carbon reserves intensity (tCO$_2$/US$ millions)	NA	66%

NA = not applicable.
Notes: The index of low-carbon leaders is reviewed and updated every six months (in May and November). This table was created after the November 2015 review of the list of index constituents.
[a]Gross returns were annualized in euros for 30 November 2010–29 February 2016.
[b]Annualized one-way index turnover for 30 November 2010–29 February 2016.

Table 3. Carbon and Financial Performances of Transparent Rules on MSCI Europe

	Optimized Index (low-carbon target)	Transparent Rules (low-carbon leaders)
Reduction in carbon emissions intensity (tCO$_2$/US$ millions)	82%	62%
Reduction in carbon reserves intensity (tCO$_2$/US$ millions)	90%	81%
Tracking error[a]	0.9%	0.72%

Note: Backtests were run over a four-year period, from 30 November 2010 to 30 June 2014.
[a]Gross returns were annualized in euros for 30 November 2010–31 July 2015.
Source: MSCI.

world's largest corporations—the publicly traded companies—to invest in reducing GHG emissions. Moreover, the encouragement of climate risk hedging can have real effects on reducing GHG emissions even before climate change mitigation policies are introduced. The mere expectation that such policies will be introduced will affect the stock prices of the highest-GHG emitters and reward those investors that have hedged climate risk by holding a decarbonized index. Finally, the anticipation of the introduction of climate change mitigation policies will create immediate incentives to initiate a transition to renewable energy.

A simple, costless policy in support of climate risk hedging that governments can adopt immediately is to mandate disclosure of the carbon footprint of their state-owned investment arms (public pension funds and sovereign wealth funds). Such a disclosure policy would have several benefits.

Given that climate change is a financial risk, disclosure provides investors (and citizens) with relevant information on the nature of the risks they are exposed to. Remarkably, some pension funds have already taken this step by disclosing their portfolios' carbon footprint—in particular, ERAFP and FRR in France; KPA Pension, the Church of Sweden, and the AP funds in Sweden; APG in the Netherlands; and the Government Employees Pension Fund (GEPF) in South Africa.

Given that citizens and pensioners will ultimately bear the costs of climate change mitigation, disclosure of their carbon exposure through their pension or sovereign wealth funds helps internalize the externalities of climate change. Indeed, investment by a public pension fund in polluting companies generates a cost borne by its government and trustees and thereby lowers the overall returns on investment. The China Investment Corporation

(CIC), China's sovereign wealth fund, has already made some statements in that direction.

Disclosure of the carbon footprint of a sovereign wealth fund's portfolio can be a way for sovereign wealth funds of oil- and gas-exporting countries to bolster risk diversification and hedging of commodity and carbon risk through their portfolio holdings. The basic concept underlying a sovereign wealth fund is to diversify the nature of the country's assets by extracting the oil and gas under the ground and thereby "transforming" these assets into "above-ground" diversifiable financial assets. Thus, it makes sense to follow up this policy by diversifying investments held by the sovereign wealth fund away from energy companies and other stock holdings that have a large carbon exposure. Interestingly, the French government recently approved a law on energy transition that requires French institutional investors to disclose their climate impact and carbon risk exposure.[36]

A more direct way to support investment in low-carbon, low-TE indexes is to push public asset owners and their managers to make such investments. Governments could thus play an important role as catalysts to accelerate the mainstream adoption of such investment policies. In this respect, it is worth mentioning the interesting precedent of the recent policy of the Shinzō Abe administration in Japan to support the development of the JPX-Nikkei Index 400. What is particularly noteworthy is that the Abe administration sees this index as an integral part of its "third arrow" plan to reform Japan's companies. GPIF—by far the largest Japanese public investor, with more than $1.4 trillion of assets under management—has adopted the new index. This example illustrates how the combination of a newly designed index with a policymaking objective and the adoption of that index by a public asset owner can be a catalyst for change.

In his book *Finance and the Good Society*, Robert J. Shiller (2012, p. 7) advances a welcome and refreshing perspective on financial economics:

> Finance is not about "making money" per se. It is a "functional" science in that it exists to support other goals—those of society. The better aligned society's financial institutions are with its goals and ideals, the stronger and more successful the society will be.

It is in this spirit that we have pursued our research on how investors can protect their savings from the momentous risks associated with GHG emissions and their long-term, potentially devastating effect on climate change. Climate change has mostly and appropriately been the bailiwick of scientists, climatologists, governments, and environmental activists. There has been relatively little engagement by finance with this important issue, but investors and financial markets cannot continue to ignore climate change. The effects of rising temperatures, the increasingly extreme weather events climate change generates, and the climate change mitigation policy responses it could provoke may have dramatic consequences for the economy and thus investment returns. Therefore, financial innovation should be explored so that the power of financial markets can be used to address one of the most challenging global threats faced by humankind.

Besides offering investors a hedging tool against the rising risks associated with climate change, a decarbonized index investment strategy can mobilize financial markets to support the common good. As a larger and larger fraction of the index-investing market is devoted to decarbonized indexes, a virtuous cycle will be activated and enhanced whereby the greater awareness of carbon footprints and GHG emissions will exert a disciplining pressure to reduce CO_2 emissions and will gradually build an investor constituency that supports climate change mitigation policies. Governments, businesses, technology innovators, and society will thus be encouraged to implement changes that accelerate the transition to a renewable energy economy.

Our basic premise/working assumption is that to foster the engagement of financial markets with climate change, it is advisable to appeal to investors' rationality and self-interest. Our argument is simply that even if some investors are climate change skeptics, the uncertainty surrounding climate change cannot be used to dismiss climate change and related mitigation policies as a zero probability risk. Any rational investor with a long-term perspective should be concerned about the absence of a market for carbon and the potential market failures that could result from this incompleteness. A dynamic decarbonized index investment strategy seeks to fill this void, offering an attractive hedging tool even for climate change skeptics.

Finally, the decarbonization approach we have described for equity indexes can also be applied to corporate debt indexes. Although the focus in fixed-income markets has been on green bonds, corporate debt indexes—decarbonized along the same lines as equity indexes (screening and exclusion based on carbon intensity and fossil fuel reserves while maintaining sector neutrality and a low TE)—could be a good complement to green bonds. Similarly, low-water-use indexes and other environmental leader indexes can be constructed in the same way as our decarbonized index.

We thank the Rockefeller Foundation for its support of this research project. For their helpful comments, we are grateful to Bertrand Badré, Lorenzo Bernasconi, *Pierre-Olivier Billard, Pascal Blanqué, Jean Boissinot, Remy Briand, Pierre Cailleteau, Yves Chevalier, Catherine Crozat, Bob Eccles, Christiana Figueres, Remco Fischer, Mark Fulton, Evan Greenfield, Mickaël Hellier, Harrison Hong, Maximilian Horster, Haizhou Huang, Mikael Johansson, Tegwen Le Berthe, Linda-Eling Lee, Mark Lewis, Bob Litterman, Jun Ma, Justin Mundy, Carl Page, Fredrik Regland, Ailsa Roell, Olivier Rousseau, Alessandro Russo, Michael Sheren, Martin Skancke, Lord Nicholas Stern, Laurent Trottier, Eric Usher, and Gernot Wagner. We thank Timothée Jaulin for excellent research assistance.*

Appendix A. Current Context of Climate Legislation

The United Nations Framework Convention on Climate Change (UNFCCC) coordinates global policy efforts toward the stabilization of GHG concentrations in the atmosphere, with a widely accepted policy target for the upcoming decades of limiting GHG emissions to keep average temperatures from rising more than 2°C by 2050. However, no concrete policies limiting GHG emissions have yet been agreed to that make this target a realistic prospect. To give an idea of what this target entails, scientists estimate that an overall limit on the concentration of CO_2 in the atmosphere between 350 parts per million (ppm) and 450 ppm should not be exceeded if we are to have a reasonable prospect of keeping temperatures from rising by more than 2°C (IPCC 2014). Maintaining CO_2 concentrations under that limit would require keeping global CO_2 emissions below roughly 35 billion tons a year, which is more or less the current rate of emissions; it was 34.5 gigatons (Gt) in 2012, according to the European Commission.

Although the process led by the UNFCCC stalled during many years following the adoption of the Kyoto Protocol, a number of countries have taken unilateral steps to limit GHG emissions in their jurisdictions. Thus, a very wide array of local regulations, as well as legislation focused on carbon emission limits and clean energy, has been introduced in the past decade—for example, 490 new regulations were put in place in 2012 as opposed to only 151 in 2004 and 46 in 1998 (UNEP FI 2013). Moreover, after promising signs of greater urgency concerning climate policies in both the United States[37] and China, the "Paris agreement" negotiated during the climate conference in Paris in December 2015 marked "an unprecedented political recognition of the risks of climate change."[38]

The Paris agreement, however, does not detail a course for action and entails many nonbinding provisions with no penalties imposed on countries unwilling or unable to reach their targets. But if the prospect of a global market for CO_2 emission permits—or even a global carbon tax—also seems far off, the establishment of a national market for CO_2 emission permits in China in the next few years could be a game changer. Indeed, in the U.S.–China Joint Announcement on Climate Change and Clean Energy Cooperation, China has pledged to cap its CO_2 emissions around 2030 and to increase the non-fossil-fuel share of its energy consumption to around 20% by 2030.[39] Moreover, following the launch of seven pilot emissions-trading schemes (ETSs), which are currently in operation, China's National Development and Reform Commission (NDRC) stated that it aimed to establish a national ETS during its five-year plan (2016–2020).[40]

Yet, despite China's impressive stated climate policy goals and the Paris agreement, substantially more reductions in CO_2 emissions need to be implemented globally to have an impact on climate change. In particular, the global price of CO_2 emissions must be significantly higher to induce economic agents to reduce their reliance on fossil fuels or to make carbon capture and storage worthwhile (current estimates indicate that a minimum carbon price of $25–$30 per ton of carbon dioxide equivalent [CO_2e] is required to cover the cost of carbon capture).[41] Therefore, with the continued rise in global temperatures and the greater and greater urgency regarding strong climate mitigation policies in the coming years, policymakers may at last realize that they have little choice but to implement radical climate policies, resulting in a steep rise in the price of carbon. On top of national governments' mobilization and international agreements, major religious authorities have recently expressed their concerns about climate change, urging both governments and civil society to act.[42]

Appendix B. Risk of Stranded Assets

The notion of stranded assets was introduced by the Carbon Tracker Initiative (2011, 2013)[43] and the Generation Foundation (2013). It refers to the possibility that not all known oil and gas reserves will be exploitable should the planet reach the peak of sustainable concentrations in the atmosphere before all oil and gas reserves have been exhausted. A plausible back-of-the-envelope calculation goes as follows: According to the Carbon Tracker Initiative (2011), Earth's proven fossil fuel reserves amount to approximately 2,800 Gt of CO_2 emissions. But to maintain the objective of no warming greater than 2°C by 2050 (with at least a 50% chance), the maximum amount of allowable emissions is roughly half, or 1,400 Gt of CO_2. In other words, oil companies' usable proven reserves are only about half of reported reserves. Responding to a shareholder resolution, ExxonMobil published in 2014, for the first time ever, a report describing how it assesses the risk of stranded assets.[44] Much of the report is an exercise in minimizing shareholders' and analysts' concerns about stranded-asset risk by pointing to the International Energy Agency's projections on growing energy demand without competitive substitutes leading to higher fossil fuel prices. Nonetheless, it cannot be entirely ruled out that investors will see a growing fraction of proven reserves as unexploitable because they are simply too costly—whether because of the emergence of cheap, clean, and reliable substitutes in the form of competitive clean energy or because climate mitigation policies become an increasingly binding reality (or, most likely, both).

Appendix C. Carbon Data

In this appendix, we offer further details on the available carbon emissions and carbon reserves data as well as the main providers of the carbon data we used.

Nature of Carbon Emissions and Carbon Reserves Data

Carbon emissions and carbon reserves relate to a wide array of greenhouse gases (GHGs) and hydrocarbon reserves. The standard unit of measurement is the metric ton of carbon dioxide equivalent ($MtCO_2e$), usually shortened to tons of carbon. Regarding GHG emissions, the most widely used international carbon-accounting tool for governments and businesses is the GHG protocol. This protocol serves as the foundation for almost every GHG standard in the world—notably, the International Organization for Standardization (ISO) and the Climate Registry. Corporate users include BP, Shell, General Motors, GE, AEG, Johnson & Johnson, Lafarge, and Tata Group. Noncorporate users include trading schemes (EU ETS, UK ETS, Chicago Climate Exchange); nongovernmental organizations (CDP, WWF, Global Reporting Initiative); and government agencies in China, the United States, US states, Canada, Australia, Mexico, and other jurisdictions.

According to the protocol, GHG emissions are divided into three scopes. Scope 1 relates to direct GHG emissions—that is, emissions that occur from sources owned or controlled by the company (e.g., emissions from fossil fuels burned on site or in leased vehicles). Scope 2 emissions are indirect GHG emissions resulting from the purchase of electricity, heating, cooling, or steam generated off-site but purchased by the entity. Scope 3 emissions encompass indirect emissions from sources not owned or directly controlled by the entity but related to its activities (e.g., employee travel and commuting, vendor supply chain). Obviously, Scope 3 emissions represent the largest GHG impact for many companies, whether in upstream activities (e.g., consumer electronics) or downstream activities (e.g., automotive industry). Scope 3 emissions reporting still lacks standardization, however, and the reporting level remains low; only 180 of the Fortune 500 companies reported on some portion of their supply chain in 2013.[45]

The estimation of the CO_2 equivalent of carbon reserves is a three-step process that involves the classification and estimation of hydrocarbon reserves that are then translated into CO_2 emissions. Most of the time, the data used for estimation of fossil fuel reserves and stranded assets concern proven reserves (a 90% probability that at least the actual reserves will exceed the estimated proven reserves). Those data are publicly available and must be disclosed in company reports. Once the proven reserves are estimated in volume or mass, two steps remain. First, the calorific value of total fossil fuel reserves must be estimated. Second, that calorific value must be translated into carbon reserves by using a carbon intensity table.

Carbon Data Providers

At the two ends of the spectrum of carbon data providers, we found entities that simply aggregate data either provided directly by companies or publicly available and those that use only their internal models to estimate carbon emissions and reserves.

Corporations themselves are the primary providers of carbon data via two main channels: (1) CSR (corporate social responsibility) reports from 37% of the world's largest companies (with a market capitalization exceeding $2 billion) completely disclose their GHG emission information; (2) CDP provides the largest global carbon-related database, in partnership with Bloomberg, MSCI ESG, Trucost, and others. Companies respond to CDP's annual information request forms for the collection of climate change–related information; the number of respondents has increased from 235 in 2003 to 2,132 in 2011. Financial data vendors, such as Bloomberg, generally provide datasets sourced from CDP, CSR reports, and other relevant reports. The heterogeneity of sources explains the discrepancies that can sometimes be found in carbon footprint measurements.

Appendix D. TE Minimization with a Multifactor Risk Model

In this appendix, we describe the multifactor risk model that we used to determine the decarbonized portfolio with minimum tracking error. We reduce *ex ante* TE by first estimating factor returns, then estimating risk, and ultimately minimizing TE.

Ex Ante and *Ex Post* Tracking Error

Index managers usually seek a very low tracking error, but some may also seek higher returns by optimizing index replication (e.g., tax optimization, management of changes in index composition, management of takeover bids). For index managers, there is a trade-off between the goals of minimizing tracking error and maximizing return. Portfolio managers use two different measures of tracking error: (1) *Ex post* TE is the measure of the volatility of the realized active return deviations from the benchmark, and (2) *ex ante* TE is an estimation (or prediction) based on an estimated multifactor model.

Ex ante TE is a function of portfolio weights, benchmark weights, the volatility of stocks, and correlations across assets. Thus, to estimate portfolio risk once portfolio weights and benchmark weights are given, we need the covariance matrix of security returns. One can estimate such a covariance matrix by using historical data of security returns, but that method is burdensome and prone to estimation error (spurious correlations).

An alternative method is to use a multifactor model. We rely on the widely used Barra multiple-factor model (MFM),[46] which decomposes the return of an individual stock into the weighted sum of common factor returns and an idiosyncratic return as follows:

$$r_i = \beta_{country\ i} f_{country\ i} + \beta_{sector\ i} f_{sector\ i} + \beta_{size\ i} f_{size\ i} + \cdots + u_i$$

$$r_i = \sum_{j=1}^{j} \beta_{ji} \tilde{f}_j + u_i$$

$$\begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1k} \\ \vdots & \ddots & \vdots \\ \beta_{nk} & \cdots & \beta_{nn} \end{bmatrix} \begin{bmatrix} f_1 \\ \vdots \\ f_j \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

$$r = \beta f + u,$$

where

β_{ji} = the factor loading for security i on common factor j

f_j = the common factor return

u_i = the part of the return that cannot be explained by common factors

Estimating Factor Returns

Common factors used by Barra include industries, styles (size, value, momentum, and volatility), and currencies; 68 factors are used for the multiple-horizon US equity model.

Common factor returns are estimated using monthly stock returns. The time series of factor returns are then used to generate factor variances and covariances in the covariance matrix:

$$\begin{bmatrix} \mathrm{Var}(f_1) & \cdots & \mathrm{Cov}(f_1, f_k) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(f_k, f_1) & \cdots & \mathrm{Var}(f_k) \end{bmatrix}.$$

To capture variance and covariance dynamics and improve the predictive power of the model, Barra uses an exponential weighting scheme that gives more weight to recent data, and so, on average, the last two to three years of data represent 50% of the available information ("half life").

From Factor Returns to Risk Estimation

Similar to components of returns, components of risks can be divided into common factor sources and security-specific risks:

$$\mathrm{Var}(\text{total risk}) = \mathrm{Var}(\text{common factor risk}) + \mathrm{Var}(\text{active specific risk}),$$

and the multifactor equation becomes

$$\mathrm{Var}(r) = \mathrm{Var}(\boldsymbol{\beta} f + u)$$
$$\mathrm{Var}(r) = \boldsymbol{\beta\Omega}_f \boldsymbol{\beta}' + \boldsymbol{\Delta},$$

where

$\boldsymbol{\beta}$ = the matrix of factor exposures

$\boldsymbol{\beta}'$ = the transposed matrix

$\boldsymbol{\Omega}$ = the variance–covariance matrix for the k factors

$\boldsymbol{\Delta}$ = the diagonal matrix of specific risk variances

The volatility, σ_p, of any portfolio p, represented by a vector of portfolio weights $\mathbf{W}_p$, is thus

$$\sigma_p = \sqrt{\mathbf{W}_p \left(\boldsymbol{\beta\Omega}_f \boldsymbol{\beta}' + \boldsymbol{\Delta}\right) \mathbf{W}_p'}.$$

TE Minimization

In the case of tracking error minimization, the objective function is the *ex ante* tracking error; constraints can range from turnover limits to reweighting rules with or without active weight constraints, among others.

Let us consider an example of a low-carbon, low-TE, multi-utilities fund. First, we have a reference universe of 10 constituents: the multi-utilities industry group in the utilities sector in a large economic zone. We assign to each constituent an index weight equal to $\left[\text{Mkt cap}(i)/\text{Total mkt cap}\right]$ in order to obtain a market cap–weighted index, and we let $\left(w_1^b,\cdots,w_{10}^b\right)$ be the constituent stocks' weights. We rank the constituents according to their carbon intensity (e.g., CO_2e/GWh) and then adopt the following constraint (rule):

$$\begin{pmatrix} w_1^b \\ w_2^b \\ \vdots \\ w_{10}^b \end{pmatrix} => \begin{pmatrix} 0 \\ w_2 \\ \vdots \\ w_{10} \end{pmatrix}.$$

In other words, the optimal portfolio $(0,w_2,\cdots,w_{10})$ will be the result of the minimization of the following objective function:

$$\text{Min}\left[\sqrt{\left(W^p-W^b\right)'\left(\boldsymbol{\beta}\boldsymbol{\Omega}_f\boldsymbol{\beta}'+\boldsymbol{\Delta}\right)\left(W^p-W^b\right)}\right],$$

where

$$\forall i=1,\cdots,10;\ 0\le w_i$$
$$i=1;\ w_1=0,$$

and

$\left(W^p-W^b\right)$ = the active weights of the portfolio with regard to the benchmark

$\boldsymbol{\Omega}_f$ = the variance–covariance matrix of factors

$\boldsymbol{\beta}$ = the matrix of factor exposures

$\boldsymbol{\Delta}$ = the diagonal matrix of specific risk variances

Barra uses an optimization algorithm to minimize TE under the new constraint of excluding stock 1. It selects active weights depending on the factor loading of each security and the covariance between each factor in order to create a new portfolio that closely tracks the reference portfolio.

Appendix E. Performance Attribution in the MSCI Europe Low Carbon Leaders Index vs. the MSCI Europe Index

In this appendix, **Table E1**, **Table E2**, and **Table E3** give several measures of performance attribution for various sectors in the MSCI Europe Low Carbon Leaders Index versus the MSCI Europe Index.

Appendix F. Percentage Contributions to Specific Risks by Sector

In this appendix, **Figure F1** depicts the breakdown of the percentage contributions to specific risks by sector.

Figure F1. Percentage Contributions to Specific Risks by Sector

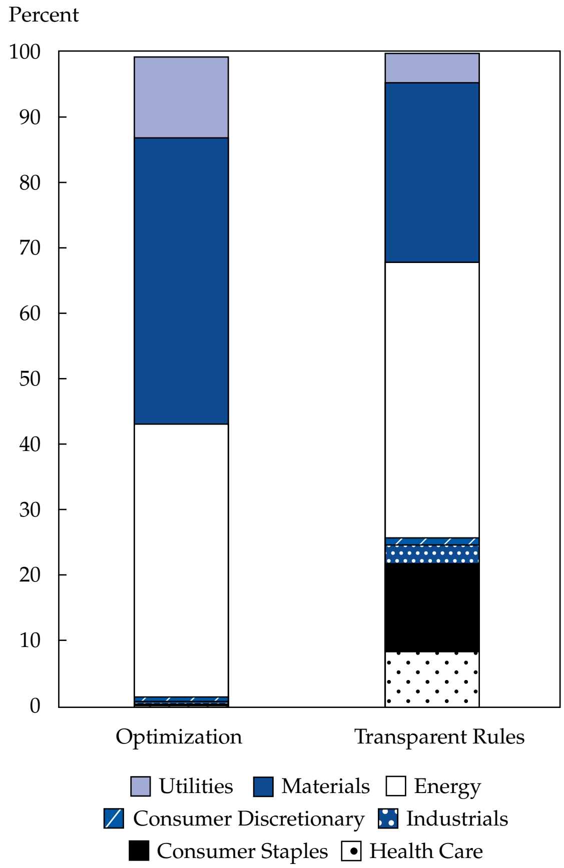


Table E1. MSCI Europe Low Carbon Leaders vs. MSCI Europe, 7 November 2014–31 January 2016

	MSCI Europe Low Carbon Leaders Index			MSCI Europe Index			Attribution Effect		
Sector	Weight	Total Return	Contribution to Return	Weight	Total Return	Contribution to Return	Allocation Effect	Selection Effect	Total Effect
Total	100.00	6.06	6.06	100.00	4.17	4.17	0.37	1.52	1.89
Materials	6.18	2.65	0.20	7.23	–17.72	–1.10	0.20	1.27	1.47
Utilities	3.87	7.55	0.30	4.00	0.83	0.04	0.02	0.25	0.27
Health care	13.48	11.16	1.29	13.84	9.28	1.12	0.00	0.21	0.21
Consumer discretionary	12.57	12.58	1.41	11.45	12.18	1.23	0.09	0.05	0.15
Industrials	12.93	7.74	0.98	11.04	7.11	0.74	0.06	0.07	0.14
Telecommunication services	5.61	17.44	0.89	4.95	16.58	0.70	0.08	0.05	0.13
Information technology	3.69	25.97	0.93	3.56	21.92	0.69	0.02	0.11	0.13
Financials	24.64	–4.18	–1.18	22.75	–4.55	–1.26	–0.15	0.11	–0.04
Energy	5.15	–26.05	–1.33	7.13	–16.82	–1.10	0.40	–0.52	–0.12
Consumer Staples	11.90	22.71	2.56	14.07	24.19	3.12	–0.37	–0.08	–0.45

Sources: Amundi; MSCI; FactSet.

Table E2. MSCI Europe Low Carbon Leaders vs. MSCI Europe—Materials Sector, 7 November 2014–31 January 2016

	MSCI Europe Low Carbon Leaders Index			MSCI Europe Index			Attribution Effect		
Sector	Weight	Total Return	Contribution to Return	Weight	Total Return	Contribution to Return	Allocation Effect	Selection Effect	Total Effect
Materials	6.18	2.65	0.20	7.23	–17.72	–1.10	0.20	1.27	1.47
Diversified metals and mining	0.75	–23.73	–0.36	1.84	–55.54	–1.15	0.68	0.36	1.04
Construction materials	0.47	28.56	0.10	0.75	–0.75	–0.01	0.01	0.11	0.12
Specialty chemicals	1.69	14.25	0.32	1.16	12.26	0.12	0.04	0.03	0.06
Steel	0.34	–23.61	–0.06	0.27	–43.40	–0.11	–0.04	0.09	0.06
Diversified chemicals	1.27	–7.61	–0.06	1.16	–9.39	–0.06	–0.02	0.02	0.00

Sources: Amundi; MSCI; FactSet.

Table E3. MSCI Europe Low Carbon Leaders vs. MSCI Europe—Utilities Sector, 7 November 2014–31 January 2016

	MSCI Europe Low Carbon Leaders Index			MSCI Europe Index			Attribution Effect		
Sector	Weight	Total Return	Contribution to Return	Weight	Total Return	Contribution to Return	Allocation Effect	Selection Effect	Total Effect
Utilities	3.87	7.55	0.30	4.00	0.83	0.04	0.02	0.25	0.27
Multi-utilities	1.43	–0.20	–0.01	1.82	–8.02	–0.13	0.08	0.11	0.19
Water utilities	0.38	21.29	0.09	0.21	21.24	0.04	0.03	0.00	0.03
Electric utilities	1.45	12.10	0.18	1.63	7.66	0.10	–0.03	0.05	0.03
Gas utilities	0.50	10.96	0.05	0.30	10.84	0.03	0.01	0.00	0.01
Renewable electricity	0.11	–3.12	0.00	0.04	–3.12	0.00	0.00	0.00	0.00

Sources: Amundi; MSCI; FactSet.

Notes

1. A recent study by a team from the National Oceanic and Atmospheric Administration found that this perceived slowdown was entirely the result of measurement errors in recorded ocean temperatures (Karl, Arguez, Huang, Lawrimore, McMahon, Menne, Peterson, Vose, and Zhang 2015).
2. For an analysis of the consequences of this deep uncertainty for the economics of carbon pricing, see Litterman (2012).
3. For a widely quoted speech on climate change and the "tragedy of horizon" and related "transition risks," see Carney (2015).
4. The United Nations Framework Convention on Climate Change (UNFCCC) coordinates global policy efforts toward the stabilization of greenhouse gas (GHG) concentrations in the atmosphere, with a widely accepted policy target for the coming decades of limiting GHG emissions to keep average temperatures from rising more than 2°C by 2050. However, no concrete policies limiting GHG emissions have yet been accepted that make this target a realistic prospect. Although the process led by the UNFCCC stalled following the adoption of the Kyoto Protocol, a number of countries have taken unilateral steps to limit GHG emissions in their own jurisdictions. The 21st Conference of the Parties to the UNFCCC, which was held in Paris in December 2015 (http://www.un.org/sustainabledevelopment/cop21/), is seen by many observers as a crucial milestone in the fight against climate change. For further details, see Appendix A.
5. A handful of organizations contribute to raising awareness of carbon risk among institutional investors. For example, the Portfolio Decarbonization Coalition (PDC)—co-founded by AP4, CDP, Amundi, and UNEP FI in September 2014—enables pioneers in the decarbonization of portfolios to share their knowledge and best practices. When it was founded, PDC set a target of $100 billion in institutional investment decarbonization to be reached by the time of the Paris conference in December 2015. It was able to significantly surpass this target, with its 25 members claiming $600 billion of decarbonized investments out of $3.2 trillion of assets under management. For more information, see http://unepfi.org/pdc/ and Top1000Funds (2015). Another example is the "Aiming for A" coalition—a group representing institutional investors—which engages carbon-intensive companies to "measure and manage their carbon emissions and move to a low-carbon economy."
6. For more information on stranded assets, see Appendix B.
7. The carbon footprint of a company refers to its annualized GHG emissions relative to a financial metric (e.g., revenue or sales) or a relevant activity metric (e.g., units produced). For further details, see the pertinent discussion later in the article as well as Appendix C.
8. See Gartner, Inc. (2016).
9. Later in the article, we report the performance results of the "decarbonized" S&P 500 and MSCI Europe indexes.
10. The mechanics that affect the relationship of carbon legislation, technological changes, and financial returns are obviously complex and not straightforward. But the purpose of decarbonized indexes is to circumvent these difficulties by focusing on an area with somewhat less uncertainty: the companies most exposed to carbon risk. Later in the article, we delve into further details.
11. To explore the links between portfolio decarbonization and the incentives it gives to companies to rechannel their investments and lower their carbon footprint, see http://unepfi.org/pdc/.
12. Koch and Bassen (2013) estimated an "equity value at risk from carbon" for European electric utilities, which is driven by their fossil fuel mix, and showed that a filter on companies with a high carbon-specific risk reduces the exposure to global carbon risk without otherwise affecting the risk–return performance of an equity portfolio.
13. See "Energy and Carbon—Managing the Risks," ExxonMobil report (March 2014).
14. These are mostly environmental, social, and governance (ESG) analysts, who until recently were largely segregated from mainstream equity analyst teams and whose audience consists predominantly of ethical investors.
15. HSBC is a notable exception, with its early integrated analysis of the materiality of carbon risk in the oil and gas as well as coal industries (HSBC 2008). Since then, the Carbon Tracker Initiative has been instrumental in raising awareness of stranded asset issues, and energy-focused analysts are increasingly and consistently integrating carbon-related risk into their analyses (see, e.g., HSBC 2012; Lewis 2014).
16. A multisector generalization of this optimization problem can break down the first set of constraints into companies that are excluded on the basis of their poor ranking in carbon intensity across all sectors, as well as companies that are excluded within each sector on the basis of either their poor carbon intensity score or high stranded assets relative to other companies in their sector.
17. Unless noted otherwise, tracking error is calculated *ex ante*.
18. This level of outperformance over such a time frame is hypothetical and for illustrative purposes only. Although we hope that a scenario of radical climate risk mitigation policy measures is possible in the near future, global climate policy implementation and its potential impact on equity valuation understandably remain a very speculative exercise.
19. In this respect, it is worth mentioning that Veolia and Danone now include carbon footprint improvement targets in their executive compensation contracts.
20. An interesting example of such a mechanism is the JPX-Nikkei Index 400, a new index based on both standard quantitative criteria (e.g., return on equity, operating profit, and market value) and more innovative qualitative criteria (e.g., a governance requirement of at least two independent outside directors). Launched with the support of the giant Japanese pension fund GPIF (Government Pension Investment Fund) to foster better corporate performance, the JPX-Nikkei 400 was quickly dubbed the "shame index." It is now carefully scrutinized by analysts, and companies are taking inclusion in the index more and more seriously.
21. For a discussion of the relationship between sustainability investments and shareholder value creation, see Khan, Serafeim, and Yoon (2015).
22. For an attempt at comparing different providers' results within a given universe, see http://www.iigcc.org/events/event/50-shades-of-green-carbon-foot-print-workshop. The differences that emerged came from different estimation models. But professionals agree that the measures are globally converging toward a much-improved harmonization.
23. For 60% of the companies in the MSCI World Index, at least 75% of emissions are from supply chains (Trucost 2013).
24. Moreover, most modern optimization techniques use factor exposures and correlations to reduce tracking error risk from such known systematic factors as volatility, small cap, and beta; they would therefore increase the weights on high-volatility/low-carbon stocks to replace high-volatility/high-carbon stocks.
25. Index and ETF investments represent a growing share of total investment products, amounting to almost 14% of total assets under management, with a year-over-year growth rate of 10% from 2013 to 2014.
26. Beyond the $11 trillion in index funds, asset owners that are members of CDP represent an asset base as high as $95 trillion (see CDP.net).
27. When AP4 started investing in 2012, a 48% reduction in carbon footprint was achieved.

28. For an early analysis of carbon-efficient indexes in emerging markets, see Banerjee (2010).
29. The criteria for excluding a stock from the index are straightforward: First, companies with the highest emissions intensity (as measured by GHG emissions/sales) are excluded, with a limit on cumulative sector weight exclusion of no more than 30%. Second, the largest owners of carbon reserves per dollar of market capitalization are excluded until the carbon reserves intensity of the index is reduced by at least 50%.
30. Our performance attribution analysis was for the MSCI Europe Low Carbon Leaders Index from 7 November 2014 to 29 January 2016.
31. During the same period, the MSCI North America Low Carbon Leaders Index outperformed the MSCI North America Index by 121 bps.
32. The allocation effect measures whether the choice of sector allocation led to a positive or negative contribution. All else being equal, overweighting outperforming sectors leads to a positive allocation effect.
33. The selection effect measures within each sector whether the portfolio manager selected the outperforming or underperforming stocks.
34. The Global Industry Classification Standard is an industry taxonomy consisting of 10 sectors, 24 industry groups, 67 industries, and 156 sub-industries.
35. Notable exceptions include the French government, which took a lead role ahead of the Paris conference in mobilizing the financial sector by requiring institutional investors to report on their climate risk exposure. A handful of central banks have also been instrumental in raising awareness of the possible hazards of climate change regulations and the potential mobilization of financial institutions. Significant contributions include the People's Bank of China and UNEP Inquiry (2015) report "Establishing China's Green Financial System" and the Bank of England's ongoing prudential review of climate-related risks to the financial sector.
36. See Article 173 of *Projet de loi relative à la transition énergétique pour la croissance verte*: "La prise en compte de l'exposition aux risques climatiques, notamment la mesure des émissions de gaz à effet de serre associées aux actifs détenus, ainsi que la contribution au respect de l'objectif international de limitation du réchauffement climatique et à l'atteinte des objectifs de la transition énergétique et écologique, figurent parmi les informations relevant de la prise en compte d'objectifs environnementaux." // "The information relative to the consideration of environmental objectives includes: the exposure to climate-related risks, including the GHG emissions associated with assets owned, and the contribution to the international goal of limiting global warming and to the achievement of the objectives of the energy and ecological transition."
37. Prominent voices in the business community have expressed their concern that the debate over climate policy has become too politicized. Also, in June 2014, the US Environmental Protection Agency unveiled an ambitious program calling for deep cuts in carbon emissions from existing power plants, with a 30% national target by 2030—which is equivalent to 730 million tons of carbon emission reductions, or about two-thirds of the nation's passenger vehicle annual emissions.
38. See "The Paris Agreement Marks an Unprecedented Political Recognition of the Risks of Climate Change," *Economist* (12 December 2015).
39. See https://www.whitehouse.gov/the-press-office/2014/11/11/fact-sheet-us-china-joint-announcement-climate-change-and-clean-energy-c.
40. The interregional ETS covering the Beijing, Tianjin, and Hebei Provinces was under discussion in February 2016, at the time of writing. In addition, the National Development and Reform Committee issued a paper in February 2016 that set up an agenda to ensure the establishment of a national ETS in 2017. We note that following China's lead, a movement is underway to move away from existing oil and gas subsidies. According to a recent IMF study by Coady, Parry, Sears, and Shang (2015), global subsidies for fossil fuels were estimated to be $333 billion in 2015.
41. The current price level is far below $30, with average carbon prices ranging from the lowest at RMB9.00/$t$$CO_2$e in Shanghai to the highest at RMB44.4/$t$$CO_2$e in Shenzhen, with others at RMB35 in Beijing, RMB23 in Tianjin, RMB22 in Hubei, RMB13 in Chongqing, and RMB14 in Guangdong (as of 4 March 2016); around EUR4.96/CO_2e (as of 7 March 2016) in Europe; and $7.5/$CO_2$e under the Regional Greenhouse Gas Initiative in the United States (as of 2 February 2016).
42. Pope Francis's *Laudato Si'* encyclical (published in May 2015), Muslim scholars' *Islamic Declaration on Global Climate Change* (published in August 2015), and US rabbis' *Rabbinic Letter on the Climate Crisis* (released in May 2015) show that climate change has become a shared concern among religious authorities.
43. For a recent study on the risk of stranded assets, see Lewis (2014).
44. See ExxonMobil (2014); Shell followed with its "Open Letter on Stranded-Asset Risk" in May 2014.
45. See https://www.greenbiz.com/blog/2013/08/12/hybrid-lcas-help-companies-size-scope-3-emissions.
46. For a thorough review of Barra equity risk modeling, see MSCI Barra (2007).

References

Banerjee, A. 2010. "Combating Global Warming in Emerging Markets with Carbon Efficient Indexes." *Journal of Environmental Investing*, vol. 1, no. 2: 29–38.

Boston Consulting Group. 2015. "Global Asset Management 2015: Sparking Growth with Go-to-Market Excellence." Boston Consulting Group (July).

Carbon Tracker Initiative. 2011. "Unburnable Carbon—Are the World's Financial Markets Carrying a Carbon Bubble?"

———. 2013. "Unburnable Carbon 2013: Wasted Capital and Stranded Assets."

Carney, Mark. 2015. "Breaking the Tragedy of the Horizon—Climate Change and Financial Stability." Speech given at Lloyd's of London (29 September).

Climate Counts. 2013. "Assessing Corporate Emissions Performance through the Lens of Climate Science" (18 December).

Coady, D., I. Parry, L. Sears, and B. Shang. 2015. "How Large Are Global Energy Subsidies?" IMF Working Paper No. 15/105 (15 May).

ExxonMobil. 2014. "Energy and Carbon—Managing the Risks." ExxonMobil Report (March).

Gartner, Inc. 2016. "Interpreting Technology Hype" (http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp).

Generation Foundation. 2013. "Stranded Carbon Assets: Why and How Carbon Should Be Incorporated in Investment Analysis" (30 October).

Guesnerie, R., and N. Stern. 2012. *Deux économistes face aux enjeux climatiques*. Paris: Le Pommier.

HSBC. 2008. "Oil and Carbon: Counting the Cost." HSBC Global Research (September).

———. 2012. "Coal and Carbon. Stranded Assets: Assessing the Risk." HSBC Global Research.

IPCC. 2014. "Climate Change 2014: Synthesis Report Summary for Policymakers." In *Contribution of Working Groups I, II, and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva: Intergovernmental Panel on Climate Change.

Karl, Thomas R., Anthony Arguez, Boyin Huang, Jay H. Lawrimore, James R. McMahon, Matthew J. Menne, Thomas C. Peterson, Russell S. Vose, and Huai-Min Zhang. 2015. "Possible Artifacts of Data Biases in the Recent Global Surface Warming Hiatus." *Science*, vol. 348, no. 6242 (June): 1469–1472.

Khan, M., G. Serafeim, and A. Yoon. 2015. "Corporate Sustainability: First Evidence on Materiality." Harvard Business School Working Paper No. 15-073 (March).

Koch, N., and A. Bassen. 2013. "Valuing the Carbon Exposure of European Utilities: The Role of Fuel Mix, Permit Allocation, and Replacement Investments." *Energy Economics*, vol. 36 (March): 431–443.

Lewis, Mark C. 2014. "Stranded Assets, Fossilised Revenues." ESG sustainability research report, Kepler Cheuvreux (24 April).

Litterman, R. 2012. "Tail Risk and the Price of Carbon Emissions." Working paper (5 December).

MSCI Barra. 2007. *Barra Risk Model Handbook*.

"The Paris Agreement Marks an Unprecedented Political Recognition of the Risks of Climate Change." 2015. *Economist* (12 December).

The People's Bank of China and UNEP Inquiry. 2015. "Establishing China's Green Financial System." Final Report of the Green Finance Task Force.

Shiller, R. 2012. *Finance and the Good Society*. Princeton, NJ: Princeton University Press.

Top1000Funds. 2015. "Institutional Investors Get Serious" (9 December).

Trucost. 2013. "Supply Chain Carbon Briefing: GHG Protocol Scope 3 Standard."

UNEP FI. 2013. "Portfolio Carbon: Measuring, Disclosing and Managing the Carbon Intensity of Investments and Investment Portfolios." UNEP Finance Initiative Investor Briefing (July).

Hedging Climate Change News

Robert F. Engle
NYU Stern and NBER

Stefano Giglio
Yale University, NBER, and CEPR

Bryan Kelly
Yale University, AQR Capital Management, and NBER

Heebum Lee
NYU Stern

Johannes Stroebel
NYU Stern, NBER, and CEPR

We propose and implement a procedure to dynamically hedge climate change risk. We extract innovations from climate news series that we construct through textual analysis of newspapers. We then use a mimicking portfolio approach to build climate change hedge portfolios. We discipline the exercise by using third-party ESG scores of firms to model their climate risk exposures. We show that this approach yields parsimonious and industry-balanced portfolios that perform well in hedging innovations in climate news both in sample and out of sample. We discuss multiple directions for future research on financial approaches to managing climate risk. (*JEL* G11, G18, Q54)

Received December 7, 2017; editorial decision December 12, 2018 by Editor Andrew Karolyi. Authors have furnished an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

Earth's climate is changing, but uncertainty around the trajectory and the economic consequences of climate change is substantial. As a result, investors around the world desire products that allow them to hedge against the

---

This version: May 23 2019. We thank Harrison Hong, Andrew Karolyi, and Ross Valkanov; participants at the Climate Finance Workshop at Columbia University, the Climate Finance Conference at Imperial College, and the Volatility Institute Conference at NYU Stern; and a number of anonymous referees for helpful comments. The Norwegian Finance Initiative and the Global Risk Institute provided generous grant support. We thank Konhee Chang for outstanding research assistance. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to Robert Engle, NYU Stern, 44 West 4th Street, New York, NY 10012; telephone: 212-998-0710. E-mail: rengle@stern.nyu.edu. Disclaimer: The views and opinions expressed are those of the authors and do not necessarily reflect the views of AQR Capital Management, its affiliates, or its employees; do not constitute an offer, solicitation of an offer, or any advice or recommendation, to purchase any securities or other financial instruments, and may not be construed as such.

doi:10.1093/rfs/hhz072

realizations of climate risk. Because of the long run and nondiversifiable nature of climate risk, standard futures or insurance contracts in which one party promises to pay the other in the event of a climate disaster are difficult to implement. Indeed, no counterparty could credibly guarantee to pay claims during a climate disaster event that might materialize in many decades, in part because a bad outcome would mandate all contracts to be paid at the same time. Individual investors are therefore largely constrained to self-insure against climate risk.

In this paper, we propose an approach for constructing climate risk hedge portfolios using publicly traded assets. We follow a dynamic hedging approach similar to Black and Scholes (1973) and Merton (1973). In this approach, rather than buying a security that directly pays off in the event of a future climate disaster, we construct portfolios whose short-term returns hedge *news* about climate change over the holding period. By hedging, period by period, the innovations in news about long-run climate change, an investor can ultimately hedge her long-run exposure to climate risk. In the short run, such a portfolio differs from the Markowitz mean-variance efficient portfolio and will thus exhibit a lower Sharpe ratio; but, in the long run, the dynamic hedging approach will compensate investors for losses that arise from the realization of climate risk.

The primary objective of this paper is to provide a rigorous methodology for constructing portfolios that use relatively easy-to-trade assets (equities) to hedge against risks that are otherwise difficult to insure. We show that our approach, which uses tools from standard asset pricing theory, does indeed allow us to construct portfolios that can successfully hedge climate news out of sample. Having said that, we do not view our resultant hedge portfolios as the definitive best hedges against climate change risk, but instead as a starting point for further exploration. Along these lines, we will discuss many valuable directions for future research on using financial markets to hedge climate risk.

The first challenge to implementing a dynamic hedging strategy for climate risk is to construct a time series that captures news about long-run climate risk, and which can therefore help us to construct an appropriate hedge target. We start from the observation that when there are events that plausibly contain such information about changes in climate risk, this will likely lead to newspaper coverage of these events; indeed, newspapers may even be the direct source that investors use to update their subjective probabilities of climate risks. Our approach in this paper therefore is to extract a climate news series from textual analysis of news sources. A wide range of events covered in newspapers can carry potentially relevant information. Indeed, the list of topics that are often covered by newspapers in relation to discussions about climate risk includes extreme weather events (e.g., floods, hurricanes, droughts, wildfires, extreme temperatures), physical changes to the planet (e.g., sea level changes, glacial melting, ocean temperatures), regulatory discussions, technical progress in alternative fuel delivery, and the price of fossil fuels.

We construct two complementary indices that measure the extent to which climate change is discussed in the news media. The first index is calculated as the correlation between the text content of *The Wall Street Journal* (WSJ) each month and a fixed climate change vocabulary, which we construct from a list of authoritative texts published by various governmental and research organizations. The WSJ is among the most salient media outlets for market participants, and thus our index captures the intensity of climate change discourse that is accessible to the investment community at very low cost.

Our WSJ Climate Change News Index associates increased climate change reporting with news about elevated climate risk, based on the idea that climate change primarily rises to the media's attention when there is a cause for concern. An alternative approach is to directly differentiate between positive and negative news in our index construction. To this end, we construct a second news-based climate index that is designed to focus specifically on bad news about climate change. This index applies sentiment analysis to climate-related articles to measure the intensity of *negative* climate news in a given month.

In this paper, we do not try to distinguish between different *types* of climate change news. In particular, we do not distinguish between news about physical damages from climate change and news about regulatory risks that are related to climate change. These two risk measures might move independent from each other. For example, the Paris accord, which led to a pledge to reduce carbon emissions, might have represented an increase in regulatory risk and a decrease in physical risk. Separately measuring news series about physical and regulatory climate risk represents an interesting avenue for future research. Also, our focus in this paper is on *global* climate change news. Our indices ignore news about local climate events, which are not covered in the WSJ or in a large cross-section of newspapers.

The second step in implementing our dynamic hedging strategy is to construct portfolios that allow us to hedge innovations in these two news series. In particular, we seek to systematically explore which stocks rise in value and which stocks fall in value when (negative) news about climate change materializes. Then, by constructing a portfolio that overweights stocks that perform well on the arrival of such negative news, an investor will have a portfolio that is well-positioned to profit the next time when such news about climate change materializes. Continued updating of this portfolio based on new information about the relationship between climate news and stock returns will ultimately lead to a portfolio which is long the winners from climate change and short the losers.

Our econometric approach to forming such hedge portfolios follows standard methods in the asset pricing literature. If climate risk represents a risk factor for asset markets (i.e., if it is a factor that drives the comovement of different assets), it is possible to construct a well-diversified portfolio the return of which isolates the exposure to that risk factor. Investors can then hedge their climate risk exposure by trading this portfolio without changing their exposures to the other risk factors in their portfolios. Various approaches to construct such hedge portfolios have been proposed in the literature. The two main ones are cross-sectional regressions like Fama-MacBeth (in which the hedging portfolio is obtained through period-by-period cross-sectional regressions of asset returns onto exposures to the risk factors), and direct projections of the risk factors onto a set of asset returns (the so-called "mimicking portfolio approach").[1] Among the many prominent papers in this literature are Fama and MacBeth (1973), Chen, Roll, and Ross (1986), Huberman, Kandel, and Stambaugh (1987), Breeden, Gibbons, and Litzenberger (1989), Lamont (2001), Balduzzi and Robotti (2008), Lönn and Schotman (2017), and Roll and Srivastava (2018). Giglio and Xiu (2018) study the asymptotic properties of the different estimators in large cross-sections, and investigate their robustness to model specification errors. In this paper, we will apply the mimicking portfolio approach, as advocated by Lamont (2001).

The challenge with implementing this mimicking portfolio approach is that we only observe a limited number of months of climate news realizations, but have a large set of assets that we could use to form hedge portfolios. This leads to concerns about data mining, where we might end up constructing hedge portfolios that perform very well in sample but that are not stable going forward. To address this concern, we use characteristics that proxy for a firm's exposure to climate risk to parsimoniously parameterize the weights of the hedge portfolios. For example, one such characteristic might be the carbon footprint of each firm. In particular, it might be that when there is news about increasing climate risk, individuals will buy low-carbon-footprint stocks and sell high-carbon-footprint stocks. If this were the case, one could construct a portfolio that increases in value when there is (negative) news about climate risk using thousands of long and short positions based on just one parameter, the firms' carbon footprints.

We implement this characteristics-based approach by using firm-level environmental performance scores constructed by the ESG ("**E**nvironmental, **S**ocial, and **G**overnance") data providers MSCI and Sustainalytics to proxy for firms' climate risk exposure.[2] In particular, we use these scores as characteristics on which to sort individual stocks to form portfolios. We then construct the final hedge portfolios by projecting innovations in our climate news indices onto these ESG-characteristic-sorted portfolios, together with standard Fama-French factor-sorted portfolios (market, size, and value).

[1] The literature on cross-sectional regressions, like Fama-MacBeth, typically focuses on estimating the risk premiums of the factor, but risk premiums are simply the average excess returns of the corresponding hedge portfolios.

[2] Again, there is a question of what *type* of climate change risk exposure these measures capture. Specifically, they may more closely capture regulatory risks than physical risks, and other characteristics could be added to the analysis to capture different types of climate change exposures. For example, one could perhaps proxy for firms' physical climate risk by the distance of firms' headquarters or production facilities from the sea. Exploring different firm-level measures of climate risk exposure (both physical and regulatory) constitutes an interesting avenue for future research.

When we compare our hedge portfolios to alternative hedge portfolios that add simple industry bets (such as positions in the energy exchange-traded fund XLE) to the standard Fama-French factors, we find that our ESG-characteristic-based mimicking portfolios procedure produces hedge portfolios that perform better than the alternatives in hedging innovations in climate risk. In particular, our portfolios deliver higher in-sample and out-of-sample correlations with those innovations. For example, the return of the hedge portfolio based on the Sustainalytics E-Scores achieves out-of-sample correlations with the WSJ index innovations as high as 30%. Our hedge portfolios also do not resemble industry bets; rather, they identify, both within and across industries, those firms with the largest exposures to climate change risk, yielding a climate hedge portfolio that is relatively industry-balanced.

Our work contributes to a burgeoning literature that studies how climate change affects asset markets, and how asset markets in turn may affect the dynamics of climate change. Andersson, Bolton, and Samama (2016) propose a passive investment strategy tilted to low-carbon stock as a hedge against climate risk, while Choi, Gao, and Jiang (2018) explore how investors update their information about climate risk. Hong, Li, and Xu (2019) investigate whether international stock markets efficiently price drought risk, and Kumar, Shashwat, and Wermers (2018) explore whether fund managers misestimate the risk of climate disasters. Baldauf, Garlappi, and Yannelis (2018), Bakkensen and Barrage (2018), Bernstein, Gustafson, and Lewis (2019), Giglio et al. (2018), and Murfin and Spiegel (2018) explore the pricing of climate risk in real estate markets, while Giglio, Maggiori, and Stroebel (2015), Giglio et al. (2018) use real estate pricing data to back out very long-run discount rates that are appropriate for valuing projects aimed at mitigating climate change. Daniel, Litterman, and Wagner (2015) apply standard asset pricing theory to calibrate the social cost of carbon.

1. Construction of the Hedge Portfolios: Theory

This section discusses our methodology to construct portfolios that hedge news about climate change. We denote by r_t an $n\times 1$ vector of excess returns over the risk-free rate of n assets at time t. We assume that these returns follow a linear factor model, in which asset returns are driven by innovations in climate news, which we denote by CC_t, as well as by p other (tradable or nontradable) risk factors v_t:

$$\underbrace{r_t}_{n\times 1}=(\underbrace{\beta_{CC}}_{n\times 1}\underbrace{\gamma_{CC}}_{1\times 1}+\underbrace{\beta_{CC}}_{n\times 1}\underbrace{(CC_t-E[CC_t])}_{1\times 1})+(\underbrace{\beta}_{n\times p}\underbrace{\gamma}_{p\times 1}+\underbrace{\beta}_{n\times p}\underbrace{v_t}_{p\times 1})+\underbrace{u_t}_{n\times 1}. \quad (1)$$

The vectors β_{CC} and β are risk exposures of the n assets to the climate news factor and the other p factors, respectively. Similarly, γ_{CC} and γ are the corresponding risk premiums for the climate news factor and the other risk factors. Finally, u_t is an idiosyncratic error term. In this basic setup, the risk exposures are constant; we relax this assumption below.

Our objective is to construct a hedge portfolio for CC_t. This is defined as a portfolio that has unit exposure (beta) to climate risk shocks CC_t, but no exposure to any of the other p factors v_t. This ensures that investors can change their exposure to climate risk by trading in this portfolio, without modifying their exposure to the other risk factors. The asset pricing literature has followed two main approaches to construct hedge portfolios: the Fama-MacBeth cross-sectional regression approach and the mimicking portfolio approach. Giglio and Xiu (2018) derive theoretical properties of the two estimators in large-dimensional settings.

In this paper, we follow the mimicking portfolio approach; for completeness, Appendix A.1 provides a review of the Fama-MacBeth procedure in our setting. In the mimicking portfolio approach, the climate risk factor CC_t is directly projected onto a set of excess returns of a set of portfolios, $\tilde{r}_t$:

$$CC_t = \xi + w'\tilde{r}_t + e_t. \tag{2}$$

The hedge portfolio for CC_t is constructed using the weights $\hat{w}$ estimated from this regression; its excess return is $h_t^{CC} = \hat{w}'\tilde{r}_t$. The vector e_t captures the measurement error in CC_t, so that this approach explicitly accounts for potential measurement error in the climate risk factor CC_t. A sufficient condition for this procedure to recover the desired hedge portfolio for climate news is that the returns of the portfolios used in the projection, $\tilde{r}$, span the same space as the true factors, (CC_t, v_t).[3]

1.1 Implementation and construction of the hedge portfolios

To build hedge portfolios using the mimicking portfolio approach, we choose a set of projection portfolios which are well diversified, so that idiosyncratic error is approximately eliminated, and which at the same time capture different dimensions of risk, so that their returns $\tilde{r}_t$ span the factor space. The portfolios used in the projection need to satisfy one further requirement. In particular, the setup described in Equation 1 includes the assumption that the risk exposures of the assets used in the estimation are constant over time. We therefore need to construct the portfolios $\tilde{r}$ in such a way that their exposures to the underlying risk factors are constant. A standard approach to achieve this is to form portfolios by sorting assets on characteristics. Indeed, to the extent that risk exposures of individual assets directly depend on these characteristics, sorting the assets by characteristics will ensure that the resultant portfolios have constant risk exposures. We follow this approach and choose a matrix

[3] Formally, write the model in the following compact form by calling f the vector of all factors: $f_t \equiv (CC_t, v_t)$, with covariance matrix Σ_f and β_f the matrix of betas: $\beta_f = (\hat{\beta}_{CC}, \hat{\beta})$. Call η the $(p+1)\times 1$ vector with 1 as the first element and 0 everywhere else, so that $CC_t = \eta' f_t$. The population vector of weights w is $Var(\tilde{r}_t)^{-1} Cov(\tilde{r}_t, CC_t)$. If returns $\tilde{r}_t$ span the same space as the true factors, this means there exists an invertible matrix H such that $\tilde{r}_t = Hf_t$. We can then write $w = (H\Sigma_f H')^{-1} H\Sigma_f \eta = H'^{-1}\eta$. The return of this portfolio is $h_t^{CC} = w'\tilde{r}_t = w'Hf_t = \eta' H^{-1} H f_t = \eta' f_t = CC_t$.

of firm-level characteristics Z_t, appropriately cross-sectionally normalized, to construct the portfolio returns as

$$\tilde{r}_t = Z'_{t-1} r_t,$$

where r_t are excess returns of individual stocks, and portfolio weights are equal to the normalized characteristics.[4] Substituting this expression into Equation 2, we write

$$CC_t = \xi + w' Z'_{t-1} r_t + e_t. \quad (3)$$

Equation 3 can be interpreted in two ways. It can either be thought of as a projection of the hedge target CC_t onto characteristic-sorted portfolios $Z'_{t-1} r_t$ that are assumed to have constant risk exposure and that span the entire factor space. Alternatively, it can be thought of as a constrained projection of CC_t on *all* individual asset returns r_t, but with time-varying weights $w' Z'_{t-1}$; the weights are modeled as a linear function of characteristics, so that any individual firm's weight depends on its risk exposure to the different factors. Equation 3 therefore performs a one-step dimension reduction that estimates the hedge portfolio, while modeling the time variation in risk exposures.

2. Hedging Climate Change News

In this section, we implement the mimicking portfolio approach to hedging climate risk that we described above. As we have highlighted in the Introduction, the relevant performance measure for the resultant hedge portfolios is how well they hedge innovations to climate news out of sample. However, given the relatively short time period for which we observe measures of both climate news and firm-level climate risk exposures, there are a limited number of out-of-sample test periods on which to evaluate the climate hedge portfolios.[5] As will become apparent below, there are many degrees of freedom in how to construct these hedge portfolios, including decisions about how to construct measures of firm-level climate risk exposures and about what other portfolios to include in regression 2. As a result, there is the danger of optimizing over these degrees of freedom to construct portfolios that provide optimal out-of-sample hedges to climate news over the short period we observe, but that may not be effective at hedging this news going forward.

To avoid such data mining concerns, we will clearly describe the various choices we encountered in the construction of the climate hedge portfolios.

[4] Note that we are exclusively working with excess returns, so there are no theoretical constraints on portfolio weights.

[5] In addition, even if we could easily extend our time series further into the past, it is unclear whether the additional sample periods would help us with constructing climate hedge portfolios today. In particular, it is plausible that climate risk has only started to be priced in stocks in recent years as investors' attention to this risk has increased. Indeed, some indirect evidence for such a suggestion comes from the fact that demand for ESG measures has substantially increased over the past few years. As a result, it is unclear whether firms with different climate risk exposures have had different excess returns in response to climate news that materialized in, say, the 1990s.

However, instead of optimizing over these degrees of freedom to find a portfolio that optimally hedges climate news over our short test sample, we make choices that appear reasonable to us, and that will hopefully lead to stable approaches to hedging climate news that is yet to occur. This discussion will highlight a number of important directions in which to further develop these climate hedge portfolios, and longer time series of measures of climate news and climate risk exposures will allow for more systematic ways of testing the true out-of-sample performance of different climate hedge portfolios.

2.1 Measuring climate change news

The first step in our analysis is to construct an index that measures innovations in news about climate risk. A variety of choices must be made when constructing this hedge target. How should we identify the news sources that reflect the information investors use in their climate risk-based investment decisions? Once we identify the appropriate news, how do we measure its relative intensity over time? How do we quantify the extent of good news versus bad news? And should one differentiate among subtypes of climate news (such as news about physical climate risks versus news about regulatory risks)?

Below, we follow two alternative approaches to building a climate news index. We believe they have the virtues of breadth and simplicity and offer scope for comparing trade-offs in some of our construction choices. At the same time, our indices have obvious imperfections and leave much room for other researchers to propose adjustments. Indeed, different investors might want to make different choices to ours in order to optimally align their hedge targets with the overall climate exposures of the rest of their portfolios. For example, investors with a strong coastal real estate portfolio might want to focus more on news about physical climate risk (because such real estate is strongly exposed to rising sea levels), while investors with a strong exposure to the coal industry might want to focus more on news about regulatory interventions in response to climate risk.[6]

2.1.1 *Wall Street Journal* climate change news index. The first index that we construct is based on climate news coverage in *The Wall Street Journal* (WSJ). Two considerations support our use of the WSJ. One is a desire to measure news that is relevant to and salient for investors concerned about climate risks, and the WSJ is among the most important media sources consumed by financial market participants. The second advantage is that we have access to the full text of WSJ articles since the early 1980s, which provides us with complete flexibility in choosing how to build the climate news index from raw news content.

[6] In addition, some researchers and investors may want to expand the list of publications they consider beyond our newspaper-based approach. Additional publications of interest could include coverage in scientific journals or social media posts.

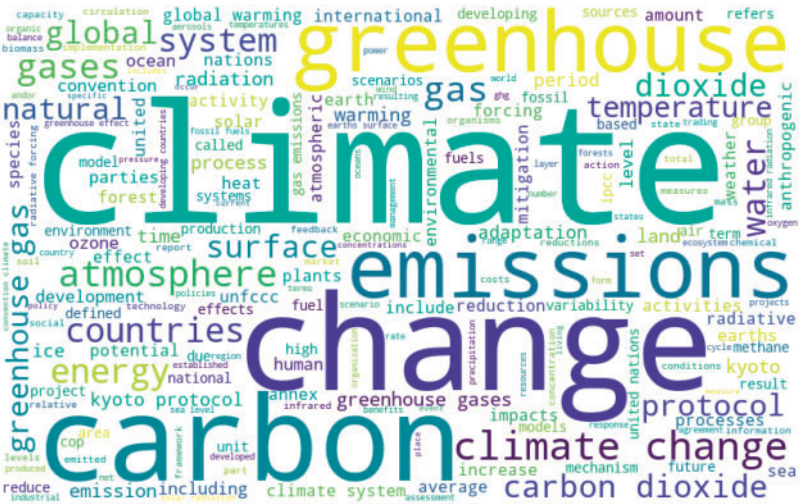


Figure 1
Climate change vocabulary
Word cloud summary of climate change vocabulary from a corpus of seventy-four authoritative climate change texts. Term sizes are proportional to their frequency in the corpus.

To quantify the intensity of climate news coverage in the WSJ, we compare the news content to a corpus of authoritative texts on the subject of climate change. In particular, we collect 19 climate change white papers from sources such as the Intergovernmental Panel on Climate Change (IPCC), the Environmental Protection Agency (EPA), and the U.S. Global Change Research Program. We complement these white papers with 55 climate change glossaries from sources such as the United Nations, NASA, the IPCC, the EPA, and others. Appendix A.2 presents the full list of these authoritative texts. We aggregate the seventy-four text documents into a “Climate Change Vocabulary (CCV),” which amounts to the list of unique terms (stemmed unigrams and bigrams) and the associated frequency with which each term appears in the aggregated corpus. Figure 1 provides an illustration of the CCV in the form of a word cloud, with term sizes proportional to their frequency. We form an analogous list of term counts for the WSJ. Each (daily) edition of WSJ is treated as a “document,” and term counts are tallied separately for each document. Next, we convert WSJ term counts into “term frequency–inverse document frequency,” or *tf-idf*, scores. Common terms that appear in most documents earn low scores because they are less informative about any individual document’s content (they have low *idf*), as do terms that are rare in a given article (they have low *tf*). The *tf-idf* transformation defines the most representative terms in a given document to be those that appear infrequently overall, but frequently in that specific document (see Gentzkow, Kelly, and Taddy 2018).

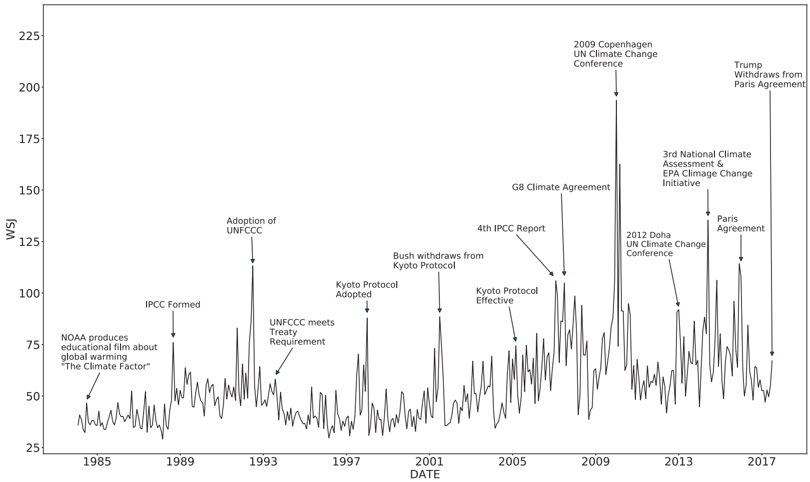


Figure 2
WSJ Climate Change News Index
This figure shows the WSJ Climate Change News Index from 1984 to 2017, annotated with climate-relevant news announcements.

The main choice going into our index construction is to treat the CCV as our definition of phraseology associated with climate change discourse. That is, our CCV takes a stand on the specific terms, and their relative usage intensity, to identify news about the topic of climate change. Like with the WSJ, we convert Climate Change Vocabulary term counts into *tf-idf*. We treat the aggregated CCV as a single document when calculating term frequencies, and apply the inverse document frequency calculation from the WSJ corpus.[7]

Finally, we construct our daily climate change index as the "cosine similarity" between the *tf-idf* scores for the CCV and each daily WSJ edition. Days in which the WSJ uses the same terms in the same proportion as the CCV earn an index value of one, while days in which the WSJ uses no words from the CCV earn an index value of zero. Approximately speaking, our raw WSJ Climate Change News Index describes the fraction of the WSJ dedicated to the topic of climate change each day, as defined by the texts that underlie the CCV. We scale this index by a factor of 10,000 to allow interpretation of the magnitudes of innovations in the index, which will represent our eventual hedge targets.

Figure 2 shows a time series of the WSJ Climate Change News Index since 1984. The figure shows that the intensity of climate news coverage has steadily increased since about the year 2000. In addition, the climate risk index spikes during salient climate events, such as the adoption of global climate treaties

[7] The choice to use the same *idf* for WSJ and CCV counts ensures that the document-frequency weights of CCV terms match the weights of WSJ terms. If we were to instead calculate *idf* based on the corpus of authoritative climate texts, we would down weight the most informative climate change terms and unduly distort the measurement of climate change discourse in the WSJ.

(e.g., the UNFCCC or the Kyoto protocol), or important global conferences to battle climate change (e.g., the 2009 UN Climate Change Conference in Copenhagen).

2.1.2 Crimson Hexagon's negative sentiment climate change news index. Implicit in our construction of the WSJ Climate Change News Index is the assumption that the number of climate change discussions increase when climate risk is elevated. In other words, the WSJ index embeds the view that, when it comes to climate change, no news is good news. While we view this as a plausible assumption, there is a risk of inaccurately capturing discussions of positive climate news (e.g., news about new mitigation technologies) as increases in climate risk. A separate potential shortcoming of the WSJ index is that, being based on a single source, it may be too narrow in its quantification of climate discourse among investors.

To address these possible concerns, we study a second news-based climate risk index that is designed to focus specifically on negative climate news, and that is drawn from a much more expansive collection of news articles. For this purpose, we use the services of the data analytics vendor Crimson Hexagon (CH). Starting in May 2008, Crimson Hexagon has collected a massive corpus of over one trillion news articles and social media posts. The underlying news sources cover over 1,000 outlets, including the WSJ, *The New York Times*, *The Washington Post*, Reuters, BBC, CNN, and Yahoo News. Coverage in terms of total articles available expands over time. Cross-sectionally, the distribution of article counts is fairly evenly distributed across news outlets, with the top-100 outlets accounting for approximately 14% of the total article count. For a given user-provided search term, CH applies a variety of proprietary natural language processing analytics, such as sentiment analysis and topic modeling, to construct time series of the sentiment of coverage of that term across the sources it collects.

We provide CH with the search phrase "climate change" and restrict our analysis to discussions in the news media (i.e., we exclude social media). Based on these choices for terms and content sources, CH provided us with an array of indices that summarize the total number of articles that include climate change news, as well as the fraction of those summarized to contain positive and negative climate change news. It also provided indices for further sentiment subcategories (e.g., fear, joy, anger), as well as a topic decomposition of climate-related articles. Thus, there are many potential degrees of freedom in using Crimson Hexagon data to construct a climate news series. For example, we could tune our choice of search terms, or optimize across each of the finer indices that CH supplies for any given set of search terms. As described above, given the brevity of our data sample, we need to guard against data mining, and we do so in this case by restricting ourselves to the most obvious search term ("climate change") and focusing on the most obvious category that resolves our desire for "signed" news, namely those that CH categorizes as basic "negative

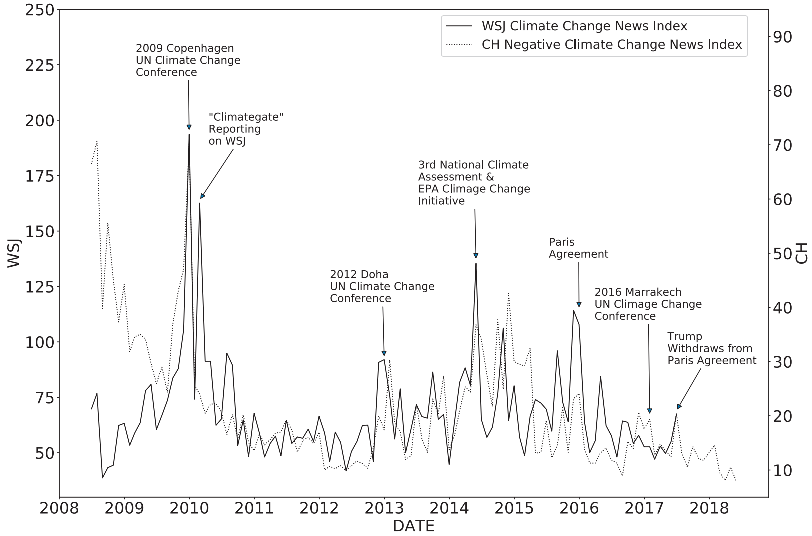


Figure 3
CH Negative Climate Change News Index
This figure shows the CH Negative Climate Change News Index from 2008 to 2017, overlaid against the WSJ Climate Change News Index, and annotated with climate-relevant news announcements.

sentiment." We calculate our CH Negative Climate Change News Index as the share of all news articles that are both about "climate change" and that have been assigned to the "negative sentiment" category; we multiply this measure by 10,000 in order to interpret the magnitudes of innovations in the index.

Figure 3 plots the time series of the CH Negative Climate Change News Index, in addition to that of the WSJ Climate Change News Index for comparison. Both indices regularly spike around salient climate events, such as climate conferences. The initial level of the CH index is somewhat higher than that of the WSJ index, though this is during a period for which Crimson Hexagon has relatively little data; this is also a period that will not be included in our final analysis (as we discuss below, our empirical analysis starts in September 2009, the first month for which we observe complete coverage of firm-level climate risk exposures). Interestingly, the WSJ index spikes in a number of instances in which the CH index does not. One of these was in early 2010, a period during which the WSJ extensively reported on the "Climategate" controversy.[8]

2.1.3 Constructing hedge targets. To measure innovations in climate news, we average the daily values for the WSJ Climate Change News Index and

[8] The Climategate controversy involved the publication of emails obtained through hacking a server at the Climatic Research Unit at the University of East Anglia. Several climate change "skeptics" alleged that these emails documented global warming to be a scientific conspiracy, with scientists manipulating data.

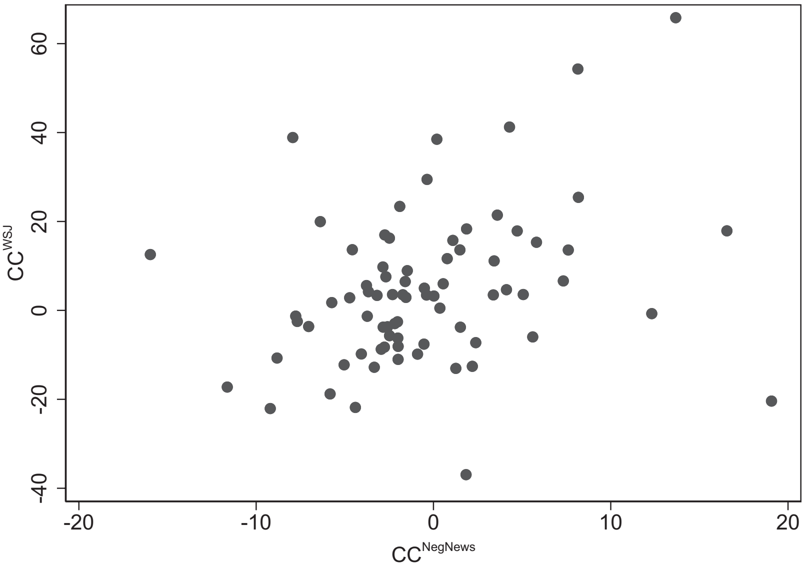


Figure 4
Correlation across CC_t measures
This figure shows a scatterplot highlighting the correlation across our two climate hedge targets, CC^{WSJ} and CC^{NegNes}. Each observation corresponds to 1 month between September 2009 to December 2016. The correlation coefficient is 0.30.

CH Negative Climate Change News Index to the monthly level, and then construct values of CC_t as residuals from an AR(1) model. This gives us our two monthly hedge targets: CC_t^{WSJ}, which captures innovations in the WSJ Climate Change News Index, and $CC_t^{NegNews}$, which captures innovations in the CH Negative Climate Change News Index. Figure 4 shows the correlation across these measures across the 88 months that will be included in our final analysis, September 2009 to December 2016. The correlation coefficient is 0.3, which suggests that, although both measures capture common elements of climate risk, they are by no means identical. As we have discussed above, which of the two series (or any one of the potential alternative series that we could have constructed) represents the ideal hedge target depends on the precise application; as a result, we view the construction of alternative hedge targets as an exciting area for further research.

2.2 Potential assets in hedge portfolios

After defining the hedge targets, the second step in implementing the mimicking portfolio hedge approach described in Section 1 is to determine the universe of assets used to build the hedge portfolio. In this project, we focus on constructing hedge portfolios using U.S. equities as the underlying assets. We obtain monthly individual U.S. stock return data from CRSP. We include only common equity securities (share codes 10 and 11) for firms traded on the NYSE, AMEX and

NASDAQ. Following Amihud (2002) and many others, we exclude penny stocks, defined as stocks with a price below $5 at the time of portfolio formation. This is to avoid including stocks whose returns are dominated by market microstructure issues. We also drop microcap stocks, defined as stocks with a market capitalization in the bottom 20% of the sample traded on the NYSE, following the observation in Fama and French (2008) that the returns of hedge portfolios obtained from long-short positions can be distorted by the inclusion of such microcaps (see also the discussion in Hou, Xue, and Zhang 2015).

2.3 Measuring climate risk exposures

Having identified the set of possible assets to include in the hedge portfolio, the next empirical challenge is to systematically measure different firms' exposures to climate risk, that is, to identify the characteristics in Z_t that drive such exposures. Our approach in this paper is to build on measures of firms' environmental exposures produced by third-party ESG data providers. Indeed, there has been a growing interest in ESG investing among investors who are increasingly demanding assets that fulfill certain environmental ("E"), social ("S"), and governance ("G") criteria.[9] Given this trend, measuring the ESG characteristics of firms has become an important task for investors, and firm-level ESG scores are available from numerous providers that collect raw data gathered from sources such as firms' disclosures, SEC filings, and reports by governments or NGOs. These raw data are then translated into numerical ESG scores using proprietary algorithms.[10]

Our study uses information on firm-level ESG scores from two leading data providers, MSCI and Sustainalytics.[11] Both data providers construct various subscores that evaluate firms on different aspects of their ESG performance. From these subscores, we choose the broadest scores that plausibly proxy for firms' exposure to climate risk.

2.3.1 MSCI. We obtained from MSCI a data set of annual firm-level ESG scores between 1995 and 2016.[12] MSCI evaluates firms along several subcategories that capture either positive or negative environmental performance; Appendix A.3 presents the full list of subcategories. Each

---

[9] According to The U.S. SIF Foundation, the dollar value of ESG assets owned by institutional investors grew to $4.73 trillion in 2016, an increase of 11% a year since 2005.

[10] As noted in the Introduction, ESG scores may capture specific notions of climate change exposure; for example, they may better capture exposure to regulatory risks than exposure to physical damages from climate risks. The methodology in this paper could be easily applied using other firm characteristics that may capture different types of climate risk exposures.

[11] The number of ESG data providers, including firms such as Arabesque and TruValue Labs, is growing. Analyzing which of these E-Scores results in the optimal hedge portfolio would be an interesting avenue for further research, but in the absence of longer time series is likely subject to concerns of data mining.

[12] These scores were formerly known as KLD scores. In 2010, following MSCI's acquisition of RiskMetrics, KLD scores were retooled into what are now known as MSCI KLD scores.

subcategory is either scored as a "1" when the firm satisfies a certain condition, or a "0" if the firm does not satisfy the condition. For instance, a "1" in the positive *"Climate Change - Energy Efficiency"* subcategory means that the company operates in a relatively energy-efficient way. The thresholds for satisfying each condition are determined by MSCI and are not disclosed with the data. Following Hong and Kostovetsky (2012), we calculate an overall environmental score for each firm by subtracting the total scores in the negative environmental subcategories from the total scores in positive environmental subcategories. We call the resultant variable the "MSCI E-Score," where a higher score suggests a firm is more environmentally friendly. In principle, it would be possible to also construct E-Scores from only a selection of all "E" subcategories, perhaps by focusing on those subcategories that are particularly relevant for climate change. The out-of-sample performance of hedge portfolios constructed using different combinations of "E" subcategories could then be compared to select the one with the best performance. However, given the relatively short time series to evaluate the performance of the resultant hedge portfolios, even such an "out-of-sample" approach of finding the "best" E-Scores is naturally subject to data mining concerns. We hence decided to restrict ourselves to only analyzing the relatively broad overall E-Score, following prior approaches in the literature; we leave a more detailed exploration of the various subcategories to future research.

2.3.2 Sustainalytics. Sustainalytics provided us with monthly firm-level ESG scores beginning in September 2009. The broadest score in the data is the "Total ESG Score," which is the average of the "Total Environment Score," the "Total Social Score," and the "Total Governance Score." To determine each of the "E," "S," and "G" scores, Sustainalytics uses a number of subcategories and evaluates each firm's score by comparing it to peers in the same industry (Sustainalytics uses a nonstandard industry classification). For instance, the fifty-seven subcategories for the "Total Environment Score" include evaluations of a firm's efforts to reduce greenhouse gas emissions, increase renewable energy use, and reduce water use; Appendix A.3 presents the full list of subcategories. The scores in the subcategories are then aggregated by weighting them according to how exposed each industry is to each ESG risk, though this aggregation procedure is not well documented. Final scores are between 0 and 100. As before, a higher score suggests a firm is more environmentally friendly. We use the "Total Environment Score" in our empirical analysis.

2.3.3 Summary statistics. Our analysis of climate hedge portfolios focuses on the period between September 2009 and December 2016. This is a period for which we observe both measures of innovations of climate news, CC_t^{WSJ} and $CC_t^{NegNews}$, and both the Sustainalytics and MSCI E-Scores. We can therefore conduct a direct comparison of the performance of the various hedge portfolios for the two climate news series over this time horizon. For the MSCI E-Score,

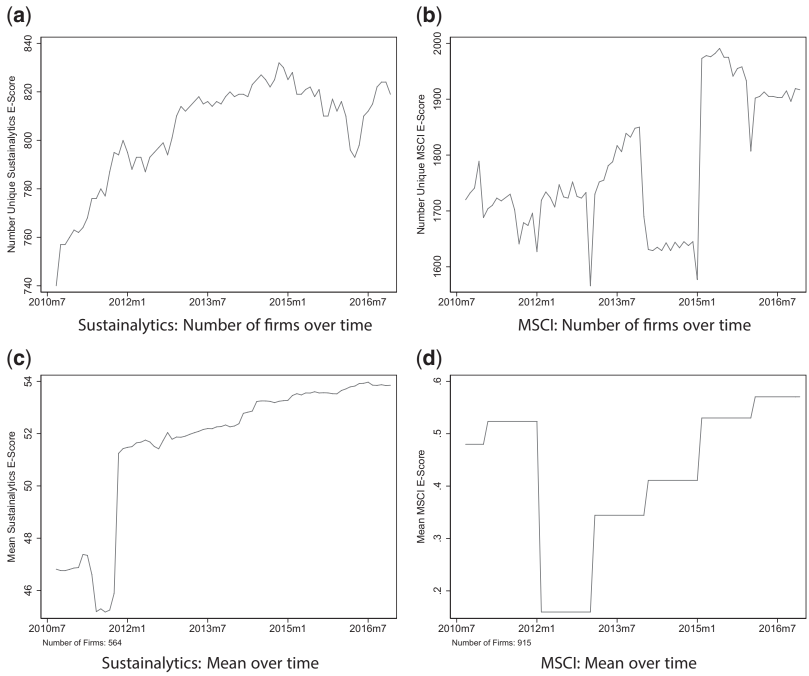


Figure 5
E-Scores: Summary statistics over time
This figure provides summary statistics for our two E-Scores. The top row shows the number of firms in our sample for which we observe E-Scores. The bottom row shows the average E-Score over time across those firms that we observe in every period in our sample. The left column shows these statistics for the Sustainalytics E-Score, and the right panel shows the statistics for the MSCI E-Score.

which is only reported annually, we assign the same score to all the months in the relevant year. Panels A and B of Figure 5 plot the number of firms in our pool of potential hedge assets for which we observe each E-Score over time. For Sustainalytics, we usually observe E-Scores for between 700 and 800 firms. MSCI E-Scores have broader coverage and are provided for between 1,700 and 1,900 firms.

Panels C and D of Figure 5 show the average values for each of the two E-Scores for a constant set of firms that we observe throughout the sample. The averages of each score contain a number of discontinuous breaks. For the MSCI E-Score, which is determined annually, these breaks could be either due to changes in firms' true ESG performance between years or due to changes in the modeling procedure. For Sustainalytics, which computes monthly scores, the discontinuous breaks are more likely due to changes in the modeling methodology over time, though we have been unable to obtain documentation on such changes that would allow us to verify this conjecture.[13] Such modeling changes would be problematic for building time-series models that perform well out of sample.

To minimize the complications from any modeling changes, we construct Z_t by cross-sectionally demeaning each E-Score in each month. However, this approach might still be problematic if changes to the model do not just shift the mean of the E-Scores over time, but also the cross-sectional dispersion. In that case, the meaning of absolute differences in the demeaned E-Score would change over time. As a second way to construct measures of Z_t, we therefore rank the E-Scores of all firms at each point in time, and then demean and rescale the ranked measure such that it ranges from -0.5 to +0.5. This approach preserves the ordinal content of the E-Scores but discards any information contained by the absolute differences between scores. Ranking-based approaches come with a number of issues. In particular, panels A and B of Figure 5 highlight that the number of firms for which E-Scores are available changes throughout the sample period. Firms added later in the sample are plausibly systematically different from those added earlier; for example, they might be less exposed to climate risk. The cross-sectional ranking of the same firm might therefore change over time without the true climate exposure of that firm changing. As a result, neither the demeaned absolute value nor the demeaned and rescaled ranked value of E-Scores are ex ante superior methods to construct climate exposures in Z_t. We will therefore present hedge portfolios using both approaches to constructing exposure measures and compare their relative performance.[14]

An interesting question is what firm characteristics are captured by the two E-Scores. A first hypothesis is that they primarily pick up industry-membership, whereby firms in “clean” industries, such as wind and solar energy, are assigned high E-Scores, and firms in “dirty” industries such as coal mining are assigned low E-Scores. To explore the extent to which the scores are primarily capturing a firm’s industry, we begin by taking the firm-level E-scores in December 2016 (the last period in our data) and regressing them onto industry fixed effects. When regressing the absolute value of the Sustainalytics E-Score on 2-digit SIC code fixed effects, the adjusted R-squared of the regression is .103; it is .184 when regressing on fixed effects for 4-digit SIC codes. The measures of R-squared were similar when using the ranked measure of the Sustainalytics E-Score. When regressing the absolute value of the MSCI E-Score on 2-digit SIC codes (4-digit SIC codes), the adjusted R-squared of the regression is .099

[13] Most uses of ESG scores by the financial services sector build on the cross-section of ESG scores at a given point in time, for example, by forming portfolios that have a relatively higher performance on these measures. Such use cases often do not require a stable meaning of the same numerical score over time.

[14] The climate exposure measures in Z_t can be constructed from the various raw E-Scores in other ways. For example, one could cross-sectionally standardize each absolute measure to have a constant standard deviation over time. Alternatively, one could rank firms’ E-Scores within industry rather than across all firms. However, in the absence of longer time series, a systematic analysis of which of these approaches obtains the best out-of-sample fit during our sample period is subject to the data mining concerns described earlier. As a result, we did not pursue these alternative approaches in this project.

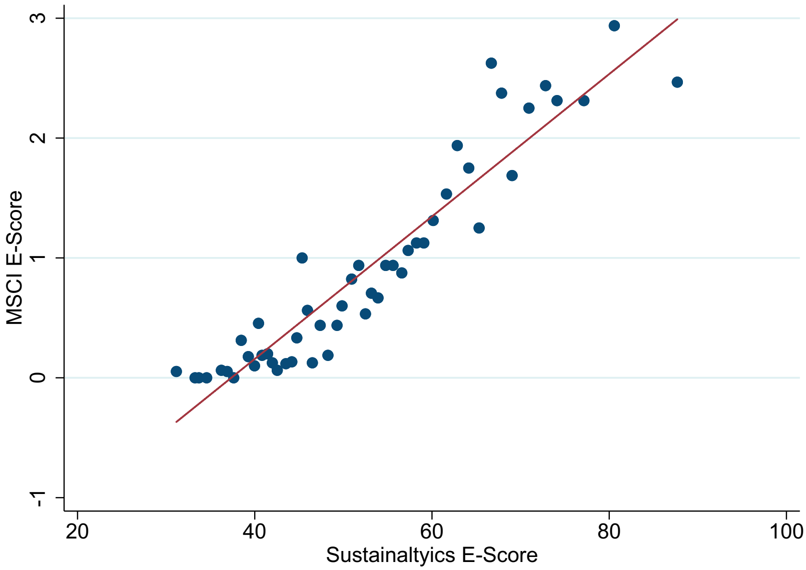


Figure 6
Correlation across E-Scores, December 2016
This figure shows a binned scatterplot that highlights the correlation across the Sustainalytics and MSCI E-Scores for all 796 firms in our sample that have both scores in December 2016. The correlation coefficient is 0.65.

(.203). These numbers show that, although there is some industry effect in determining E-Scores, most of the variation occurs within relatively narrow industries, rather than across industries.

Indeed, the three 2-digit SIC industries with the lowest Sustainalytics E-Scores are Personal Services (SIC code 72), Water Transportation (SIC code 44), and Motion Pictures (SIC code 78), probably not the first industries that come to mind when thinking of "dirty" industries. Similarly, the 2-digit SIC industries with the highest Sustainalytics E-Scores are Building Materials & Gardening Supplies (SIC code 52), Textile Mill Products (SIC code 22), and Furniture & Homefurnishings Stores (SIC code 57). When ranking by MSCI E-Scores, we similarly find that low-scoring firms are not necessarily those one would expect ex ante, such as those operating in the oil and gas sector.

A second question is the extent to which the MSCI and Sustainalytics E-Scores capture the same object. Figure 6 shows the correlation across the raw Sustainalytics and MSCI E-Scores in December 2016. They have a positive correlation of about 0.65, suggesting that they are both measuring aspects of the same object. However, enough independent variation occurs across the two measures to suggest that their usefulness in constructing climate hedge portfolios might vary. Indeed, we show below that the performance of the hedge portfolios varies noticeably when these hedge portfolios are constructed using the different E-Scores.

2.4 Forming hedge portfolios

In this section, we construct hedge portfolios for innovations in climate news, CC_t, using the mimicking portfolio approach described in Section 1.1. As discussed above, we use two different approaches to transform the raw E-Scores into the characteristic vector Z_t:

(1) Using firms' cross-sectionally demeaned absolute value of the E-Score ("absolute scores", e.g., $Z_t^{SUS\_A}$)

(2) Ranking the firms cross-sectionally by their E-Score, and then standardizing these rankings to range between -0.5 and +0.5 ("ranked scores", e.g., $Z_t^{SUS\_R}$).

Recall that one of the conditions for the mimicking portfolio approach to isolate climate change risk (and to avoid picking up other potentially correlated risks in the economy) is that the projection portfolios have to span all the risk factors driving returns. In addition to portfolios sorted on the climate characteristics, we therefore also include in regression 2 three additional factors that might be correlated with climate risk and that are known to be important in explaining the cross-section of returns: size (using cross-sectionally standardized market value to create Z_t, so that half the firms, sorted by market value, have positive weight, and half have negative weight; note that this portfolio will be long large firms and short small firms), value (using cross-sectionally standardized values of book-to-market to create Z_t), and the market (setting Z_t to equal the share of total market value).[15] For example, when we use the absolute Sustainalytics E-Score to measure firms' climate risk exposures, regression 3 becomes

$$CC_t = \xi + w_{SUS} Z_{t-1}^{SUS\_A'} r_t + w_{SIZE} Z_{t-1}^{SIZE'} r_t + w_{HML} Z_{t-1}^{HML'} r_t + w_{MKT} Z_{t-1}^{MKT'} r_t + e_t, \tag{4}$$

where $w_{SUS}, w_{SIZE}, w_{HML}$ and w_{MKT} are scalars that capture the weight of the corresponding portfolios in the mimicking (hedge) portfolio for CC_t. For comparability, we also analyze the performance of hedge portfolios constructed using returns of the exchange-traded funds (ETFs) XLE and PBD instead of the returns of portfolios of stocks sorted by their E-Scores. XLE is the ticker of the Energy Select Sector SPDR ETF, which represents the energy sector of the S&P 500. PBD is the ticker of the Invesco Global Clean Energy ETF, which is based on the WilderHill New Energy Global Innovation Index and comprises companies that focus on greener and renewable sources of energy and technologies facilitating cleaner energy. Constructing hedge portfolios based on those ETFs allows us to (a) analyze the extent to which our E-Score-based hedge portfolios simply represent a market tilt away from "brown energy"

15 To maximize the number of stocks used to construct the hedge portfolios, we include stocks even if some of the characteristics Z_t are missing for that stock. To do so, we set all missing characteristics equal to zero.

Table 1
Full-sample regression: WSJ Climate Change News Index

	(1)	(2)	(3)	(4)	(5)
$Z_{t-1}^{SUS\_A'} r_t$	1.416***				
	(0.436)				
$Z_{t-1}^{SUS\_R'} r_t$		67.789***			
		(17.834)			
$Z_{t-1}^{MSCI\_A'} r_t$			12.658*		
			(6.849)		
$Z_{t-1}^{MSCI\_R'} r_t$				53.743*	
				(27.401)	
r_t^{XLE}					0.085
					(0.810)
r_t^{PBD}					0.208
					(0.630)
$Z_{t-1}^{HML'} r_t$	1.221	2.309	−5.862	−5.941	−6.772
	(7.019)	(6.873)	(6.878)	(6.858)	(8.093)
$Z_{t-1}^{SIZE'} r_t$	−5.680**	−6.034**	−5.511*	−5.459**	−2.765
	(2.350)	(2.289)	(2.773)	(2.696)	(2.474)
$Z_{t-1}^{MKT'} r_t$	0.783	0.789	0.841	0.789	0.091
	(0.642)	(0.628)	(0.692)	(0.680)	(1.285)
Constant	2.894	2.673	4.659*	4.891*	5.959**
	(2.681)	(2.613)	(2.700)	(2.669)	(2.897)
R-squared	.153	.187	.083	.088	.047
N	88	88	88	88	88

This table shows results from regression 4. The dependent variable captures innovations for the WSJ-Based Climate News measure. The unit of observation is a month, and the sample runs between September 2009 and December 2016. Standard errors are presented in parentheses. *p <.1; **p <.05; ***p <.01.

and toward "green energy" and (b) explore whether hedge portfolios based on XLE and PBD would have performed better than our E-Score-based hedge portfolios.[16]

2.5 In-sample fit results

We begin by exploring the in-sample fit of various versions of regression 4 over the full sample period. Table 1 shows regressions when hedging innovations to the WSJ Climate Change News Index, CC_t^{WSJ}, described in Section 2.1. Columns 1 and 2 show that portfolios based on Sustainalytics E-Scores have a positive and significant relationship with CC_t^{WSJ}; in periods with more innovations in negative climate news, a portfolio that goes long firms with higher (more "green") E-Scores has relatively larger excess returns. The R-squared measures of these regressions show that the portfolios based on the Sustainaltyics E-Scores can hedge 15%–19% of the in-sample variation in

[16] As before, there are many degrees of freedom for how to compute hedge portfolios based on ETFs, and we do not want to suggest that portfolios constructed using XLE and PBD constitute the "best" ETF-based portfolios for hedging climate risk. Indeed, we view the analysis of which ETFs and other funds are most helpful in hedging climate risk to be an exciting area for future research.

CC_t. Columns 3 and 4 show that portfolios based on the MSCI E-Scores also have higher excess returns during periods with innovations in negative climate news; the R-squared measures of the regressions are lower than those in Columns 1 and 2. Portfolios based on ranked versions of both E-Scores have a slightly higher in-sample fit than portfolios based on absolute demeaned values. In addition to the ESG scores, size appears to correlate with climate change exposure: larger firms appear more exposed than smaller firms to climate change news, in the sense that they perform worse when the amount of news coverage of climate change in the WSJ increases. Column 5 includes the returns of XLE and PBD instead of the return of a characteristic-sorted portfolio. The in-sample fit of this regression is lower than that of any of the regressions in Columns 1–4, even though we have fewer explanatory variables in those regressions. This suggests that the characteristic-weighted portfolios might have some advantages over a hedge approach that creates industry tilts using energy-related ETFs.[17] It also shows that most of the R-squared in Columns 1–4 is the result of the characteristics-weighted portfolios, and not of the other portfolios, which are also included in Column 5.

Table 2 presents the same set of regressions as Table 1, but hedges innovations in the CH Negative Climate Change News Index, $CC_t^{NegNews}$. As before, the in-sample fits of the hedge portfolios based on Sustainalytics E-Scores are higher than the fits of the hedge portfolios based on MSCI E-Scores; similarly, the in-sample fits of the portfolios constructed using ranked E-Scores are marginally higher than those of the portfolios constructed using the absolute (demeaned) E-Score. Finally, the in-sample fits of all four portfolios based on E-Scores are somewhat higher than that of the portfolio based on XLE and PBD.[18] Overall, the relative performance of the various hedge portfolios is similar whether we are trying to hedge the WSJ Climate Change News Index or the CH Negative Climate Change News Index.

How would the hedge portfolios implied by these regressions look? To determine each firm i's weight in the hedge portfolio, we construct the following sum, where $Z_{i,t}$ values are taken as of December 2016: $\hat{w}_{SUS\_A} Z_{i,Dec16}^{SUS\_A'} + \hat{w}_{SIZE} Z_{i,Dec16}^{SIZE'} + \hat{w}_{HML} Z_{i,Dec16}^{HML'} + \hat{w}_{MKT} Z_{i,Dec16}^{MKT'}$, and where the various $\hat{w}$-terms represent the estimated coefficients from regression 4. This means that a firm's weight in the hedge portfolio is determined by its E-Score as well as its book-to-market ratio and its size. The resultant portfolio is the portfolio that an investor would form in December 2016 to hedge climate news in January 2017. Table 3 presents the average portfolio positions by 2-digit SIC code classification for

[17] The inclusion of the other factors in regression 4 make the resultant hedge portfolios in Column 5 of Table 2 different from a simple industry-tilt away from the market. Indeed, the resultant hedge portfolio will have a beta of 1 with CC_t, and a beta of zero with the other factors. Factor neutrality, not industry neutrality, is a desirable property of hedge portfolios.

[18] It is interesting to note that when hedging negative climate change news, the value-growth dimension seems to be aligned with the risk exposure. In particular, the table shows that value firms appear more exposed to climate news than growth firms.

Table 2
Full-sample regression: CH Negative Climate Change News Index

	(1)	(2)	(3)	(4)	(5)
$Z_{t-1}^{SUS\_A'} r_t$	0.266*				
	(0.141)				
$Z_{t-1}^{SUS\_R'} r_t$		12.286**			
		(5.864)			
$Z_{t-1}^{MSCI\_A'} r_t$			1.089		
			(2.173)		
$Z_{t-1}^{MSCI\_R'} r_t$				6.641	
				(8.696)	
r_t^{XLE}					−0.092
					(0.252)
r_t^{PBD}					0.036
					(0.196)
$Z_{t-1}^{HML'} r_t$	−4.536**	−4.390*	−5.934***	−5.919***	−5.520**
	(2.272)	(2.260)	(2.182)	(2.177)	(2.519)
$Z_{t-1}^{SIZE'} r_t$	−0.137	−0.179	0.210	0.100	0.501
	(0.761)	(0.753)	(0.880)	(0.856)	(0.770)
$Z_{t-1}^{MKT'} r_t$	0.315	0.314	0.287	0.295	0.297
	(0.208)	(0.206)	(0.219)	(0.216)	(0.400)
Constant	−0.115	−0.137	0.313	0.306	0.376
	(0.868)	(0.859)	(0.857)	(0.847)	(0.902)
R-squared	.125	.133	.090	.094	.089
N	88	88	88	88	88

This table shows results from regression 4. The dependent variable captures innovations for the *Newspaper-based negative climate news* measure. The unit of observation is a month, and the sample runs between September 2009 and December 2016. Standard errors are presented in parentheses. *p <.1; **p <.05; ***p <.01.

the industries with the six largest negative average portfolio weights and the industries with the six largest positive average portfolio weights. We only present the portfolio positions based on the absolute E-Scores, because they look very similar to the positions in the hedge portfolio constructed using the ranked E-Scores. For the portfolio constructed using Sustainlytics E-Scores to hedge innovations in the CH Negative Climate Change News Index, for example, the largest short position is "General Building Contractors," followed by "Water Transportation." The largest long positions are "Building Materials & Gardening Supplies" and "Tobacco Products." This analysis highlights that the resultant hedge portfolios will not necessarily conform with common priors that the optimal way to hedge climate change news involves primarily going long green energy stocks and short oil companies; this is consistent with our observation that industry membership can only explain a small amount of the cross-sectional variation in firm-level E-Scores.

2.6 Out-of-sample fit results

The most important test of the hedge portfolios is their ability to hedge out-of-sample innovations to climate news, that is, to hedge innovations in months that were not included in the estimation of the portfolio weights. To construct a first

Table 3
Largest average short and long positions (by 2-digit SIC code)

A. WSJ Climate Change News Index

Sustainalytics E-Score (absolute)		**MSCI E-Score (absolute)**	
Top negative portfolio weights	***SIC2***	***Top negative portfolio weights***	***SIC2***
Coal mining	12	Water transportation	44
Water transportation	44	Petroleum & coal products	29
Insurance agents, brokers, & service	64	Motion pictures	78
Mining nonmetallic minerals, except fuels	14	Communications	48
Transportation services	47	Security & commodity brokers	62
Security & commodity brokers	62	Oil & gas extraction	13
Top positive portfolio weights	***SIC2***	***Top positive portfolio weights***	***SIC2***
Building materials & gardening supplies	52	Pipelines, except natural gas	46
Tabacco products	21	Tabacco products	21
Food & kindred products	20	Miscellaneous manufacturing industries	39
Paper & allied products	26	Lumber & wood products	24
Textile mill products	22	Paper & allied products	26
Furniture & homefurnishings stores	57	Textile mill products	22

B. CH Negative Climate Change News Index

Top negative portfolio weights	***SIC2***	***Top negative portfolio weights***	***SIC2***
General building contractors	15	General building contractors	15
Water transportation	44	Nondepository institutions	61
Coal mining	12	Auto repair, services, & parking	75
Insurance agents, brokers, & service	64	Communications	48
Holding and other investment offices	67	Water transportation	44
Insurance carriers	63	Insurance carriers	63
Top positive portfolio weights	***SIC2***	***Top positive portfolio weights***	***SIC2***
Railroad transportation	40	Chemical & allied products	28
Transportation by air	45	Textile mill products	22
Furniture & homefurnishings stores	57	General merchandise stores	53
Textile mill products	22	Lumber & wood products	24
Building materials & gardening supplies	52	Building materials & gardening supplies	52
Tobacco products	21	Tobacco products	21

This table shows the industries (2-digit SIC code) with the largest average short and long positions in the estimated hedge portfolios resulting from regressions presented in Tables 1 and 2. Panel A explores hedge portfolios based on regression 4 using innovations in the WSJ Climate Change News Index as CC_t, and panel B explores hedge portfolios based using innovations in the CH Negative Climate Change News Index as CC_t. All portfolios are constructed using the absolute demeaned value of the E-Scores. Within each portfolio, industries are arranged in descending order of the absolute values of the portfolio weights.

measure of the out-of-sample performance of the hedge portfolios, for every period t we run regression 4 using data between periods t_{min} and $t-1$, where t_{min} corresponds to the first month for which we observe all climate exposures and CC_t series (September 2009). We then form the hedge portfolio based on these estimates and explore the correlation of the returns of that hedge portfolio in period t with CC_t. This corresponds to the approach one would have taken to hedge climate news in real time. Because we require a certain amount of data to estimate regression 4, we only compare the out-of-sample performance of the hedge portfolios starting in period $t_{min}+30$ (March 2012).[19]

[19] Further reducing the number of portfolios onto which to project CC_t may improve the out-of-sample performance of the hedging portfolio. Given the short sample size available, in this paper we decided to not optimize the hedge portfolio further along this dimension.

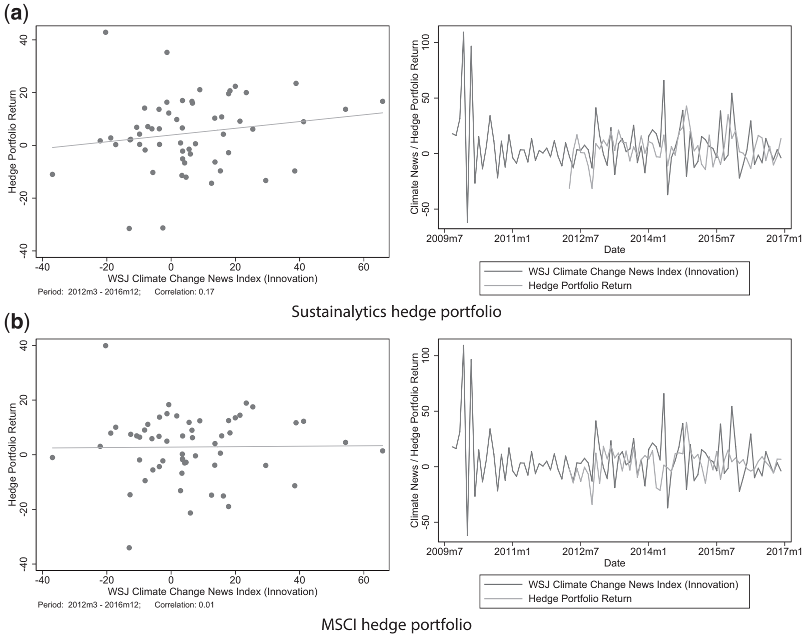


Figure 7
Out-of-sample fit: WSJ Climate Change News Index
This figure explores the out-of-sample performance of hedge portfolios constructed to hedge the *WSJ-Based Climate News Measure*. The top panel presents hedge portfolios built on the absolute values of the Sustainalytics E-Score, and the bottom panel presents portfolios built on the absolute values of the MSCI E-Score.

Figure 7 presents the out-of-sample performance of portfolios constructed to hedge innovations in the WSJ Climate Change News Index. The top panels show portfolios constructed using absolute values of the Sustainalytics E-Score, and the bottom panels show portfolios that build on the absolute values of the MSCI E-Score. The left columns present scatterplots of the out-of-sample returns of the hedge portfolios together with the realizations of the innovation of climate news. The right panels plot the time series of the climate news series and the return series of the hedge portfolios. There is a clear, positive out-of-sample correlation with CC_t of 0.17 for the Sustainalytics hedge portfolio. In other words, the hedge portfolios indeed have higher returns during periods with positive innovations to climate news. Portfolios based on MSCI E-Scores or ETFs, on the other hand, have very little ability to hedge innovations in the WSJ Climate Change News Index, with an out-of-sample correlation of just 0.01.

Panel A of Table 4 provides additional information about the out-of-sample performance of the various portfolios designed to hedge innovations in the WSJ Climate Change News Index. The first column is the most important one, showing the correlation between the realizations of CC_t^{WSJ} and the

Table 4
Cross-correlations: WSJ Climate Change News Index

A. Out-of-sample fit

	CC^{WSJ}	$H_{OOS}^{SUS\_A}$	$H_{OOS}^{SUS\_R}$	$H_{OOS}^{MSCI\_A}$	$H_{OOS}^{MSCI\_R}$	H_{OOS}^{ETF}	r_t^{XLE}	r_t^{PBD}
CC^{WSJ}	1.000							
$H_{OOS}^{SUS\_A}$	0.174	1.000						
$H_{OOS}^{SUS\_R}$	0.206	0.973	1.000					
$H_{OOS}^{MSCI\_A}$	0.013	0.688	0.621	1.000				
$H_{OOS}^{MSCI\_R}$	0.019	0.677	0.624	0.988	1.000			
H_{OOS}^{ETF}	−0.005	0.427	0.349	0.861	0.852	1.000		
r_t^{XLE}	0.068	−0.138	0.004	−0.097	−0.039	−0.141	1.000	
r_t^{PBD}	0.111	0.185	0.272	0.294	0.350	0.190	0.656	1.000

B. Cross-validation fit

	CC^{WSJ}	$H_{Cross}^{SUS\_A}$	$H_{Cross}^{SUS\_R}$	$H_{Cross}^{MSCI\_A}$	$H_{Cross}^{MSCI\_R}$	H_{Cross}^{ETF}	r_t^{XLE}	r_t^{PBD}
CC^{WSJ}	1.000							
$H_{Cross}^{SUS\_A}$	0.244	1.000						
$H_{Cross}^{SUS\_R}$	0.300	0.976	1.000					
$H_{Cross}^{MSCI\_A}$	0.039	0.742	0.671	1.000				
$H_{Cross}^{MSCI\_R}$	0.067	0.733	0.676	0.982	1.000			
H_{Cross}^{ETF}	−0.069	0.454	0.390	0.678	0.651	1.000		
r_t^{XLE}	0.068	0.041	0.072	−0.009	−0.034	0.297	1.000	
r_t^{PBD}	0.111	0.272	0.266	0.310	0.298	0.470	0.656	1.000

This table shows cross-correlations of different portfolios and innovations in the WSJ Climate Change News Index. Panel A focuses on the performance of hedge portfolios from our out-of-sample approach, and panel B focuses on the performance of hedge portfolios from our cross-validation approach.

returns of the various hedge portfolios (e.g., $R_{OOS}^{SUS\_A}$ corresponds to the out-of-sample returns of a hedge portfolio constructed using absolute values of the Sustainalytics E-Score). The hedge portfolios based on Sustainalytics E-Scores substantially outperform the hedge portfolios based on the MSCI E-Scores. In addition, hedge portfolios based on ranked E-Scores marginally outperform those based on absolute E-Scores, though the returns of portfolios based on absolute and ranked E-Scores from the same data provider are highly correlated. Finally, the out-of-sample performance of the Sustainalytics E-Score-based hedge portfolios is substantially better than that of portfolios based on ETFs. The returns of most hedge portfolios are *negatively* correlated with the returns to XLE, suggesting that these hedge portfolios are likely to hold short positions in the energy firms that constitute XLE. Similarly, we observe a positive correlation between the returns of all climate hedge portfolios and the returns of PBD, suggesting that the hedge portfolios likely hold long positions in many of the green energy firms that constitute PBD.

We also conduct a second test for the performance of the hedge portfolios based on a cross-validation approach. In particular, for every period t' we run regression 4 for all periods $t \neq t'$, and then use the resultant estimates to construct a hedge portfolio in a similar way as described above. The return of that hedge portfolio in period t' is then compared to $CC_{t'}$. Panel B of Table 5 explores the cross-validation performance of the various hedge portfolios. The hedge portfolios based on Sustainalytics E-Scores continue to outperform those based on MSCI E-Scores or ETFs substantially.

In sum, the hedge portfolios built using the Sustainalytics E-Score perform out of sample substantially better than any other hedge portfolio we have considered. The worse hedging performance of portfolios based on MSCI E-Scores highlights the importance of choosing characteristics that properly capture cross-sectional variation in exposure to climate change risks.

Figure 8 and Table 5 present results similar to those in Figure 7 and Table 4, but analyze the performance of portfolios designed to hedge innovations in the CH Negative Climate Change News Index. Portfolios based on Sustainalytics E-Scores have a similar ability to hedge this second climate news series as they had in hedging the CH Negative Climate Change News Index, both in the out-of-sample evaluation and in the cross-validation evaluation. The hedging ability of the MSCI indexes is in this case much higher than for the WSJ measure of climate change risks, suggesting that the MSCI E-Scores are more suited to capture negative climate change news as opposed to general coverage of climate change by the WSJ. Overall, the out-of-sample correlation between realization of climate change news and the hedge portfolios are 0.22 when using Sustainalytics E-Scores and 0.18 when using MSCI E-Scores.

3. Conclusion and Directions for Future Research

We demonstrate how a mimicking portfolio approach can be successful in hedging innovations in climate change news across a number of out-of-sample performance tests. Across our two indices for climate news, the hedge portfolios based on Sustainalytics E-Scores have the best in-sample fit as well as the best out-of-sample and cross-validation performance. Portfolios based on MSCI E-Scores and ETFs have a lower (but still positive) ability to hedge innovations in climate news. There are no systematic differences in the relative performance of hedge portfolios based on absolute or ranked versions of the raw E-Scores. In general, however, the differences between the out-of-sample and cross-validation performance of some of the portfolios highlight that the portfolios we construct are somewhat sensitive to the exact time series on which our models are trained. This is likely the result of only having a relatively few data points in each of our estimations. As we observe longer time series of E-Scores and climate news measures, our proposed method should deliver ever-better portfolios to hedge climate change news. Similarly, moving from hedging climate news that materializes over a monthly level to hedging on a daily level should allow researchers to substantially expand their training data, and thereby improve the out-of-sample performance of the hedge portfolios.

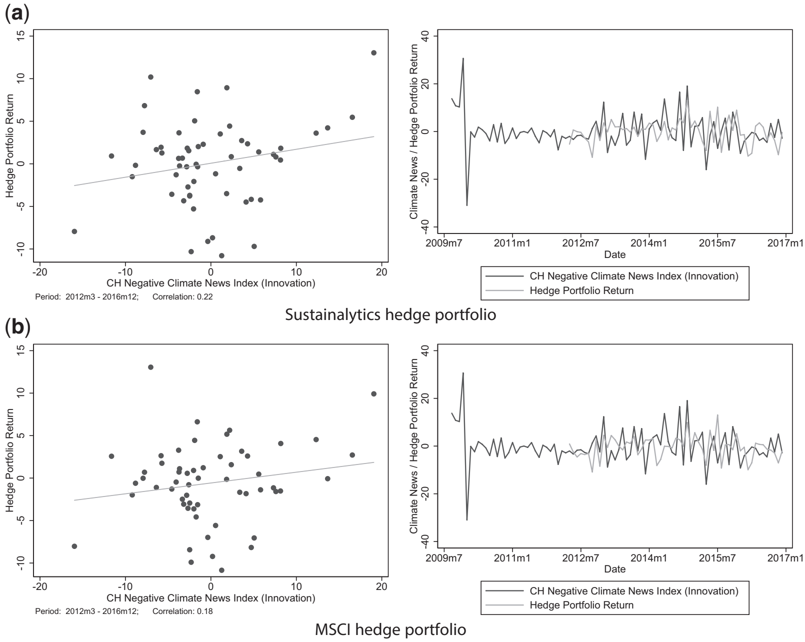


Figure 8
Out-of-sample fit: CH Negative Climate Change News Index
This figure explores the out-of-sample performance of hedge portfolios constructed to hedge CH-based negative climate news measure. The top panel presents hedge portfolios built on the absolute values of the Sustainalytics E-Score, and the bottom panel presents portfolios built on the absolute values of the MSCI E-Score.

More generally, we view this article as providing a rigorous methodology for constructing portfolios that hedge against risks that are otherwise difficult to insure. We do not view our resultant hedge portfolios as the definitive best hedges against climate change risk, but instead as a starting point for further exploration. Indeed, future research could consider many valuable directions for climate finance, and we discussed a number of the dimensions that should be explored further, including the addition of more assets to the hedge portfolios (such as international stocks) and the formation of hedge portfolios based on both characteristic-sorted portfolios and ETFs.

One additional important direction for future work is to integrate more and better data to measure firm-level climate risk exposures. These data could come from commercial data providers or could be constructed by researchers themselves, for example, by including information such as geographical proximity to potential climate disasters (e.g., rising sea levels or hurricane-prone regions). Indeed, articles in this volume, such as Choi, Gao, and Jiang (2018) and Kumar, Shashwat, and Wermers (2018) make valuable progress toward developing new ways to quantify climate risk exposures.

Table 5
Cross-correlations: CH Negative Climate Change News Index

A. Out-of-sample fit

	$CC^{NegNews}$	$H_{OOS}^{SUS\_A}$	$H_{OOS}^{SUS\_R}$	$H_{OOS}^{MSCI\_A}$	$H_{OOS}^{MSCI\_R}$	H_{OOS}^{ETF}	r_t^{XLE}	r_t^{PBD}
$CC^{NegNews}$	1.000							
$H_{OOS}^{SUS\_A}$	0.217	1.000						
$H_{OOS}^{SUS\_R}$	0.183	0.992	1.000					
$H_{OOS}^{MSCI\_A}$	0.179	0.869	0.852	1.000				
$H_{OOS}^{MSCI\_R}$	0.175	0.865	0.850	0.998	1.000			
H_{OOS}^{ETF}	0.157	0.780	0.767	0.961	0.960	1.000		
r_t^{XLE}	−0.066	−0.412	−0.353	−0.387	−0.367	−0.410	1.000	
r_t^{PBD}	0.063	0.061	0.112	0.096	0.127	0.119	0.656	1.000

B. Cross-validation fit

	$CC^{NegNews}$	$H_{Cross}^{SUS\_A}$	$H_{Cross}^{SUS\_R}$	$H_{Cross}^{MSCI\_A}$	$H_{Cross}^{MSCI\_R}$	H_{Cross}^{ETF}	r_t^{XLE}	r_t^{PBD}
$CC^{NegNews}$	1.000							
$H_{Cross}^{SUS\_A}$	0.148	1.000						
$H_{Cross}^{SUS\_R}$	0.154	0.991	1.000					
$H_{Cross}^{MSCI\_A}$	0.024	0.864	0.836	1.000				
$H_{Cross}^{MSCI\_R}$	0.048	0.885	0.861	0.993	1.000			
H_{Cross}^{ETF}	0.053	0.829	0.799	0.973	0.968	1.000		
r_t^{XLE}	−0.066	−0.208	−0.183	−0.205	−0.237	−0.223	1.000	
r_t^{PBD}	0.063	0.169	0.171	0.158	0.157	0.185	0.656	1.000

This table shows cross-correlations of different portfolios and innovations in the CH Negative Climate Change News Index. Panel A focuses on the performance of hedge portfolios from our out-of-sample approach, and panel B focuses on the performance of hedge portfolios from our cross-validation approach.

Another direction for follow-on work is to develop alternative definitions of the climate change risks. One interesting question is whether it is important to differentiate between physical and policy-oriented climate risks. For example, a tax on greenhouse gas emissions, if comprehensively applied at an appropriate level, would reduce the demand for climate hedge portfolios and consequently the cost of insuring against climate change. Thus, good regulation will mean less need for climate hedges. But regulation itself creates winners and losers from regulatory risk, and one might therefore want to construct regulatory hedge portfolios. The stability of such regulatory hedge portfolios may well be sensitive to the prevailing political environment.

A related question pertains to the expected returns of the various hedge portfolios. Indeed, an increasing use of climate hedge portfolios by investors will increase the price (and thus reduce the expected returns) of those firms whose stock provides the most effective hedge against innovations in climate change news. This lower expected return corresponds to the insurance premium paid for the climate hedge portfolio. An interesting avenue for future work will be to quantify the cost of the climate hedge portfolios by looking at the associated risk premiums.[20] It is also interesting to study the general equilibrium effects resulting from the fact that a lower cost of capital for firms with high E-Scores might actually have a direct effect on the climate trajectory. For example, to the extent that green energy firms see a reduction in their cost of capital, this might allow them to achieve efficient scale faster, and thereby affect the path of greenhouse gas emissions. The design of structural asset pricing models that feature such general equilibrium feedback loops seems a promising direction for research.

Appendix

A.1 Review of the Fama-MacBeth Approach

In this section, we review the Fama-MacBeth estimator for hedge portfolios in the context of our model. To apply the Fama-MacBeth procedure, the econometrician needs to take a stand on all the factors in the model: CC_t and v_t. Once the factors in the model are determined, the procedure follows two steps. In the first step, the risk exposures β_{CC} and β are estimated via time-series regressions of returns onto the factors, CC_t and v_t. In particular, for each asset i, $(\hat{\beta}^i_{CC}, \hat{\beta}^i)$ are estimated from the time-series regression:

$$r^i_t = \alpha^i + \beta^i_{CC} CC_t + \beta^i v_t + u_t.$$

In the second step, in each period t, hedge portfolios for all factors are obtained via cross-sectional regressions of returns r_t onto the estimated betas $(\hat{\beta}_{CC}, \hat{\beta})$:

$$r_t = h^{CC}_t \hat{\beta}_{CC} + h_t \hat{\beta} + e_t,$$

where $\hat{\beta}_{CC}$ and $\hat{\beta}$ are the betas estimated in the first step. The slopes of this regression in each period t are precisely the returns of the hedge portfolio in period t: h^{CC}_t (that hedges CC_t) and h_t (that hedges the remaining factors v_t). The hedge portfolios h^{CC}_t and h_t have, by construction, a beta of one with respect to the corresponding factors and zero with respect to all other factors. Their time-series means (the expected excess returns of the hedge portfolios) recover the risk premiums of the factors: $E[h^{CC}_t] = \gamma_{CC}$ and $E[h_t] = \gamma$. The Fama-MacBeth procedure for constructing hedge portfolios has two potential drawbacks. First, it requires knowing all the factors in the model, CC_t and v_t. Second, the procedure is not robust to measurement error in the factor of interest, CC_t, which is a natural concern in many settings, including in ours (see the further discussion of omitted factors and measurement error in Giglio and Xiu 2018).

A.2 Source of Climate Change Vocabulary

To create the Climate Change Vocabulary (CCV), we collect twelve climate change white papers from various sources including the Intergovernmental Panel on Climate Change (IPCC), Environmental Protection Agency (EPA), and the U.S. Global Change Research Program. We complement this with fifty-nine climate change glossaries from sources such as United Nations, NASA, IPCC, and EPA.

A.2.1 Twelve climate change white papers Table A1 reports the institution, title and published year of climate change white papers that we use to construct the CCV.

[20] Note that this requires substantial time-series data, because realizations of negative climate news in sample might actually lead the hedge portfolios to outperform over any given period.

Table A1
List of climate change white papers

Source	Title	Year
IPCC	IPCC Synthesis Report	1990, 1995, 2001, 2007, 2014
IPCC	IPCC Special Report: The Regional Impacts of Climate Change: an assessment of vulnerability	1997
IPCC	IPCC Special Report: Aviation and the Global Atmosphere	1999
IPCC	IPCC Special Report: Methodological and Technological Issues in Technology Transfer	2000
IPCC	IPCC Special Report: Safeguarding the Ozone Layer and the Global Climate System: Issues Related to Hydrofluorocarbons and Perfluorocarbons	2005
IPCC	IPCC Special Report: Carbon Dioxide Capture and Storage	2005
IPCC	IPCC Special Report: Renewable Energy Sources and Climate Change Mitigation	2011
IPCC	IPCC Special Report: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation	2012
American Association for the Advancement of Science	What We Know: The Reality, Risks, and Response to Climate Change	2014
UC Berkley	American Climate Prospectus	2015
U.S. EPA	Climate Change Indicators in the United States (4th edition)	2016
Science	Social and Economic Impacts of Climate	2016
IMF	The Effects of Weather Shocks on Economic Activity	2017
U.S. Global Change Research Program	Our Change Planet: The U.S. Global Change Research Program for Fiscal Year 2017	2017
U.S. Global Change Research Program	Climate Science Special Report (4th National Climate Assessment, Vol. I)	2017

IPCC reports scientific and technical assessments of the current state of climate change. Generally, these reports comprise three volumes: one for each of the Working Groups of the IPCC. In addition to the main reports, Summaries for Policymakers and Synthesis Reports are provided. A Synthesis Report integrates materials covered by Assessment Reports and Special Reports. It is a nontechnical report targeting policy makers and addressing a broad range of policy-relevant but policy-neutral questions. Summary for Policymakers is an abridged version of the full Synthesis Report. In addition, IPCC Special Reports provide an assessment of a specific issue relating to climate change. They are generally structured similar to a volume of an Assessment Report. IPCC, Intergovernmental Panel on Climate Change; EPA, Environmental Protection Agency; IMF, International Monetary Fund.

A.2.2 Fifty-nine climate change glossaries

We collect climate change glossaries, both words and their definition, from U.S. Environmental Protection Agency (EPA), BBC, United Nations(UN), Center for Climate and Energy Solutions Glossary of Key Terms, Intergovernmental Panel on Climate Change (IPCC), World Health Organization (WHO), European Climate Adaptation Platform, International Petroleum Industry Environmental Conservation Association(IPIECA), Lenntech, Wikipedia, Met Office, Integrated Regional Information Networks(IRIN), Climate Change in Australia, Guardian, International Rivers, Mekong River Commission, Exploratorium, *New York Times*, U.S. Forest Service, U.S. Department of Transportation, Durham Region, Classroom of the Future, Government of Canada, International Food Policy Research Institute (IFPRI), New Zealand Government, University of Miami, German Climate Finance, California Government, South West Climate Change Impacts Partnership (SWCCIP), Scent of Pine, Natural Climate Change, UN Climate Change Conference, Center for Strategic and International Studies(CSIS), Watts Up With That?, U.K. Climate Impacts Programme (UKCIP), Climate Change Zambia, Canadian Broadcasting Corporation(CBC), Auburn University, Global Warming Solved, REDD+, Climate Resilience Toolkit(CRT), What's Your Impact, The Nitric Acid Climate Action Group (NACAG), Garnaut Climate Change Review, Climate Policy Information Hub, Explaining Climate Change, Four Degrees Preparation, The European Initiative for Upscaling Energy Efficiency in the Music Event Industry (EE MUSIC), Regional Education and Information Centre (REIC), Ecology, Climate Reality Project, National Geographic, Agricultural Marketing Resource Center (AgMRC), Global Greenhouse Warming, and Conservation in a Changing Climate.

A.3 Subcategories for "E" scores

A.3.1 MSCI

Positive indicators are Environmental Opportunities - Clean Tech, Waste Management - Toxic Emissions and Waste, Waste Management - Packaging Materials & Waste, Climate Change - Carbon Emissions, Property/Plant/Equipment, Environmental Management Systems, Natural Resource Use - Water Stress, Natural Resource Use - Biodiversity & Land Use, Natural Resource Use - Raw Material Sourcing, Natural Resource Use - Financing Environmental, Environmental Opportunities - Green Buildings, Environmental Opportunities in Renewable Energy, Waste Management - Electronic Waste, Climate Change - Energy Efficiency, Climate Change - Product Carbon Footprint, Climate Change - Insuring Climate Change Risk, Environment - Other Strengths. Negative indicators are Regulatory Compliance, Toxic Emissions and Waste, Energy & Climate Change, Impact of Products and Services, Biodiversity & Land Use, Operational Waste, Water Stress, Environment - Other Concerns.

A.3.2 Sustainalytics

Subcategories are Formal Environmental Policy, Environmental Management System, External Certification of Environmental Management Systems (EMS), Environmental Fines and Non-monetary Sanctions, Participation in Carbon Disclosure Project, Scope of Corporate Reporting on GHG emissions, Programmes and Targets to Reduce GHG Emissions from Own Operations, Programmes and Targets to Increase Renewable Energy Use, Carbon Intensity, Carbon Intensity Trend, % of Primary Energy Use from Renewables, Operations Related Controversies or Incidents, Reporting Quality Non-Carbon Environmental Data, Programmes and Targets to Protect Biodiversity, Guidelines and Reporting on Closure and Rehabilitation of Sites, Environmental and Social Impact Assessments, Oil Spill Reporting and Performance, Waste Intensity, Water Intensity, Percentage of Certified Forests Under Own Management, Programmes & Targets to Reduce Air Emissions, Programmes & Targets to Reduce Air Emissions, Programmes & Targets to Reduce Water Use, Other Programmes to Reduce Key Environmental Impacts, GHG Reduction Programme, Programmes and Targets to Improve the Environmental Performance of Own Logistics and Vehicle Fleets, Programmes and Targets to Phase out CFCs and HCFCs[21]

[21] CFCs refers to chlorofluorocarbons, and HCFCs refers to Hydrochloroflourocarbons.

in Refrigeration Equipment, Formal Policy or Programme on Green Procurement, Environmental Supply Chain Incidents, Programmes to Improve the Environmental Performance of Suppliers, External Environmental Certification Suppliers, Programmes and Targets to Stimulate Sustainable Agriculture, Programmes and Targets to Stimulate Sustainable Aquaculture/Fisheries, Food Beverage & Tabacco Industry Initiatives, Programmes and Targets to Reduce GHG Emissions from Outsourced Logistics Services, Data on Percentage of Recycled/Reused Raw Material Used, Data on Percentage of Forest Stewardship Council (FSC) Certified Wood/Pulp as Raw Material, Programmes and Targets to Promote Sustainable Food Products, Food Retail Initiatives, Products & Services Related to Controversies or Incidents, Sustainability Related Products & Services, Revenue from Clean Technology or Climate Friendly Products, Automobile Fleet Average CO2 Emissions, Trend Automobile Fleet Average Fleet Efficiency, Products to Improve Sustainability of Transport Vehicles, Systematic Integration of Environmental Considerations at R&D Stage, Programmes and Targets for End-of-Life Product Management, Organic Products, Policy on Use of Genetically Modified Organisms (GMO) in Products, Environmental & Social Standards in Credit and Loan Business, Responsible Asset Management, Use of Life-Cycle Analysis(LCA) for New Real Estate Projects, Programmes and Targets to Increase Investment in Sustainable Buildings, Share of Property Portfolio Invested in Sustainable Buildings, Sustainability Related Financial Services, Products with Important Environmental/Human Health Concerns, Carbon Intensity of Energy Mix, Mineral Waste Management, Emergency Response Programme.

References

Amihud, Y. 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5:31–56.

Andersson, M., P. Bolton, and F. Samama. 2016. Hedging climate risk. *Financial Analysts Journal* 72:13–32.

Bakkensen, L., and L. Barrage. 2018. Flood risk belief heterogeneity and coastal home price dynamics: Going under water? Working Paper.

Baldauf, M., L. Garlappi, and C. Yannelis. 2018. Does climate change affect real estate prices? only if you believe in it. Working Paper.

Balduzzi, P., and C. Robotti. 2008. Mimicking portfolios, economic risk premia, and tests of multi-beta models. *Journal of Business and Economic Statistics* 26:354–68.

Bernstein, A., M. Gustafson, and R. Lewis. 2019. Disaster on the horizon: The price effect of sea level rise. *Journal of Financial Economics* Advance Access published March 23, 2019, 10.1016/j.jfineco.2019.03.013.

Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81:637–54.

Breeden, D., M. Gibbons, and R. Litzenberger. 1989. Empirical tests of the consumption-oriented capm. *Journal of Finance* 44:231–62.

Chen, N.-F., R. Roll, and S. Ross. 1986. Economic forces and the stock market. *Journal of Business* 59:383–403.

Choi, D., Z. Gao, and W. Jiang. 2018. Attention to global warming. Working Paper.

Daniel, K., R. Litterman, and G. Wagner. 2015. Applying asset pricing theory to calibrate the price of climate risk. Working Paper, Columbia University.

Fama, E., and K. French. 2008. Dissecting anomalies. *Journal of Finance* 63:1653–78.

Fama, F., and J. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–36.

Gentzkow, M., B. T. Kelly, and M. Taddy. 2018. Text as data. Working Paper.

Giglio, S., M. Maggiori, and J. Stroebel. 2015. Very long-run discount rates. *Quarterly Journal of Economics* 130:1–53.

Giglio, S., M. Maggiori, J. Stroebel, and A. Weber. 2018. Climate change and long-run discount rates: Evidence from real estate. Working Paper.

Giglio, S., and D. Xiu. 2018. Asset pricing with omitted factors. Working Paper.

Hong, H., and L. Kostovetsky. 2012. Red and blue investing: Values and finance. *Journal of Financial Economics* 103:1–19.

Hong, H., F. Li, and J. Xu. 2019. Climate risks and market efficiency. *Journal of Econometrics* 208:265–81.

Hou, K., C. Xue, and L. Zhang. 2015. Dissecting anomalies: An investment approach. *Review of Financial Studies* 28:650–705.

Huberman, G., S. Kandel, and R. Stambaugh. 1987. Mimicking portfolios and exact arbitrage pricing. *Journal of Finance* 42:1–9.

Kumar, N., A. Shashwat, and R. Wermers. 2018. Do fund managers misestimate climatic disaster risk? Working Paper.

Lamont, O. 2001. Economic tracking portfolio. *Journal of Econometrics* 105:161–84.

Lönn, R., and P. Schotman. 2017. Empirical asset pricing with many assets and short time series. Working Paper, Maastricht University.

Merton, R. 1973. An intertemporal capital asset pricing model. *Econometrica* 41:867–87.

Murfin, J., and M. Spiegel. 2018. Is the risk of sea level rise capitalized in residential real estate. Working Paper.

Roll, R., and A. Srivastava. 2018. Mimicking portfolios. Working Paper, California Institute of Technology.

THE JOURNAL OF FINANCE • VOL. LXI, NO. 4 • AUGUST 2006

Industry Concentration and Average Stock Returns

KEWEI HOU and DAVID T. ROBINSON*

ABSTRACT

Firms in more concentrated industries earn lower returns, even after controlling for size, book-to-market, momentum, and other return determinants. Explanations based on chance, measurement error, capital structure, and persistent in-sample cash flow shocks do not explain this finding. Drawing on work in industrial organization, we posit that either barriers to entry in highly concentrated industries insulate firms from undiversifiable distress risk, or firms in highly concentrated industries are less risky because they engage in less innovation, and thereby command lower expected returns. Additional time-series tests support these risk-based interpretations.

FIRMS GENERATE CASH FLOWS THROUGH their actions in product markets. These risky cash flows are in turn priced in financial markets. Yet the economic link between product markets and asset prices remains relatively unexplored. This paper explores the link between industry concentration and average stock returns, offering the first empirical evidence of the asset pricing implications of industry market structure.

There are a number of potential reasons why the structure of product markets may affect stock returns. Firms take operating decisions that may affect the riskiness of their cash flows. These operating decisions arise from an equilibrium in the product market that potentially reflects strategic interactions among market participants. Therefore, the structure of product markets may affect the risk of a firm's cash flows, and hence a firm's equilibrium rate of return.

Take, for example, innovation. According to Schumpeter (1912), innovation is a form of creative destruction that is more likely to occur in competitive industries or on the fringes of established industries. If innovation is risky, and this risk is priced, then this predicts that competitive industries or firms on the competitive fringe of established industries earn higher returns, all else

*Kewei Hou is from the Fisher College of Business at The Ohio State University, and David T. Robinson is from the Fuqua School of Business at Duke University. We thank Alon Brav, Anne-Marie Knott, Robert Stambaugh (the editor), Lu Zhang, seminar participants at the 2005 WFA meetings, Ohio State, The University of Arizona, George Mason University, and an anonymous referee for many helpful comments. This paper is a significantly revised version of a paper entitled, "Market Structure, Firm Size, and Expected Stock Returns." Judith Chevalier, Eugene Fama, Peter Hecht, Owen Lamont, Toby Moskowitz, Per Olsson, Per Strömberg, and seminar participants at The University of Chicago and the Chicago Quantitative Alliance provided many helpful comments on this earlier draft. Any remaining errors are our own.

equal. Hence, innovation is one channel through which the structure of product markets has implications for stock returns.

Or consider distress. If barriers to entry in product markets insulate some firms from aggregate demand shocks, while exposing others, then we would expect distress risk to vary with market structure. This predicts that industries with high barriers to entry are associated with lower equilibrium stock returns. Thus, distress is another way that market structure can impact stock returns.

Regardless of whether the link between market structure and stock returns is better characterized by distress risk, innovation risk, or yet some other channel, our message is simple. It is well understood from industrial organization that the structure of product markets affects managers' equilibrium operating decisions. If these operating decisions affect the risk of a firm's cash flows, then these decisions should impact stock returns.

The main finding in this paper is that firms in highly concentrated industries earn lower returns, even after controlling for size, book-to-market, momentum, and other known return predictors. This finding is true both of industry portfolio returns and individual firm-level returns, and it is robust to a variety of empirical specifications. Moreover, the economic magnitude of these effects is large. Our results indicate that firms in the quintile of the most competitive industries earn annual returns that are nearly 4% higher than those of similar firms in the quintile of the most concentrated industries. This difference is highly statistically significant.

To rule out chance or spurious correlation as a potential explanation for these findings, we explore a wide range of robustness tests and alternative explanations. Using the Davis, Fama, and French (2000) files, we extend our main results back to 1927. In addition, the results hold across a wide range of industry concentration measures. Our sample selection criteria ensure that the results are driven neither by regulated industries, nor by the de-listing bias documented by Shumway (1997).

Another possible explanation is that we are simply documenting differences in unexpected returns that arise from persistent in-sample cash flow shocks: These shocks may be correlated with industry concentration in sample, but are unlikely to persist in the future. To control for this explanation, we examine the relation between concentration, profitability, and returns. Our analysis shows that on average, highly concentrated industries have experienced positive abnormal profitability, while abnormal profitability for competitive industries has been negative. Thus, not only does unexpected profitability fail to account for our findings, it works in the opposite direction. This suggests that we may be understating the true relation between concentration and *expected* returns.

Given that we cannot easily dismiss our findings as arising from chance, spurious correlation, persistent in-sample cash flow shocks, or correlation with other known determinants of returns, our next step is to explore potential explanations for our results. We study the time-series properties of the concentration premium to explore its relation to risk-based explanations such as distress or innovation risk. Spanning tests of the cross-sectional concentration premium reveal that it is statistically and economically significant and not well explained by existing asset pricing factors. Moreover, the concentration premium exhibits sensible business cycle variation and is related to future economic activity: The premium grows as the economy contracts and it is high when current and near-term GDP growth are low. This indicates that when future economic conditions look bleak, investors raise the required rate of return for firms in relatively more competitive industries.

Finally, our concentration premium largely subsumes the size and market factors, but the premium on book-to-market grows when we control for concentration. This leads us to examine the book-to-market spread in returns across concentration quintiles. We show that the premium associated with the book-to-market ratio is larger in more concentrated industries. Through double-sort portfolios, we find that most of the spread in returns across concentration quintiles occurs for low book-to-market firms. This finding supports a risk-based explanation, since it shows that returns are high for low book-to-market firms in competitive industries (where book-to-market is low because expected growth is high), while returns are low for low book-to-market firms in concentrated industries (where book-to-market is low because capitalized future profitability is high).

Although we cannot rule out behavioral explanations for our results, these time-series findings suggest that industry concentration proxies for a risk factor sensitivity. Our findings are consistent with the view that innovation/distress risk, which is more pronounced in competitive industries, is a priced source of risk in the context of the multifactor asset pricing models of Merton (1973) and Ross (1976).

This paper is part of a larger literature that links industrial organization to issues in financial economics. Earlier work such as Titman (1984) studies how capital structure and product markets interact through the liquidation decision. A number of recent papers examine the link between capital structure and industry characteristics; see, for example, Mackay and Phillips (2005) or Almazan and Molino (2001). In addition, a series of papers, including Asness and Stevens (1996), Moskowitz and Grinblatt (1999), Cohen, Polk, and Vuolteenaho (2003), and Hou (2003), demonstrate that a wide range of asset pricing phenomena have important industry components. To our knowledge, ours is the first paper to link expected stock returns to industry product market characteristics through the channel we propose.

The remainder of the paper is structured as follows. Section I motivates our hypotheses linking industry concentration to stock returns. Section II describes the data and how we construct industry concentration measures. In addition, this section illustrates the relation between industry concentration and industry-level characteristics. Section III examines how industry concentration affects the cross-section of stock returns. Section IV examines profitability surprises as a potential explanation for our results, while Section V presents the time-series evidence. We explore the relation between value, growth, and concentration in Section VI. Section VII concludes.

I. The Link between Market Structure and Stock Returns

If the structure of product markets affects asset prices, then either market structure affects risk directly, or else it is somehow correlated with investor perceptions in a way that links it to behavioral phenomena. In this section, we focus on risk-based channels through which market structure affects stock returns.

For market structure to affect equilibrium stock returns through a risk-based channel, it must be that equilibrium operating decisions induced by a particular market structure are related to expected returns. While it is well understood that market structure affects equilibrium firm behavior, the industrial organization literature stops short of making predictions for stock returns. Our purpose in this section is to conjecture a possible mechanism through which industrial organization affects equilibrium stock returns.

Of course, whether or not existing asset pricing factors capture the risks brought about by market structure is an empirical question—one that we address later in this paper. Our goal in this section is not to argue whether a certain number of priced factors is correct for explaining stock returns, and thus whether existing asset pricing factors should or should not capture the risks associated with market structure (for more on this see Fama (1998)). Rather, our purpose here is to close the gap between industrial organization and asset pricing by generating testable predictions for stock returns based on theories from industrial organization.

We focus on two channels through which industry concentration can potentially affect stock returns. The first draws on Schumpeter's (1912) concept of creative destruction. The second, closely related, channel is through barriers to entry.

Creative destruction is the idea that innovation occurs in small firms on the fringes of established industries, and that these small challengers ultimately overturn the existing status quo and usher in a new technological paradigm. In short, innovation and technological progress involve unseating incumbent firms in industries.[1]

Recently this view has received renewed support in industrial organization. Knott and Posen (2003) show empirically that innovation increases with the degree of industry competition. He, Mørck, and Yeung (2003) present complementary evidence relating turnover in firm dominance to differences in economic growth across countries. They find that economic growth correlates positively

[1] Schumpeter is associated with two influential and opposing views of the link between market structure and innovation. His later view, discussed in Schumpeter (1942), argues that monopolistic firms have stronger incentives to innovate than firms in competitive industries, since monopolistic firms can enjoy the economic profits arising from their innovation, rather than have their supernormal profits competed away. This later view has received criticism. For instance, work by Geroski (1990) finds evidence against the hypothesis that competitive rivalry diminishes innovation, and Reinganum (1985) models an industry with a single incumbent and multiple challengers and shows that the challengers have stronger innovation incentives, suggesting that the level of innovation varies non-monotonically with the number of firms in the industry.

with firm turnover, suggesting that creative destruction is an important element of long-run growth.

If creative destruction describes the relation between market structure and risky innovative activities, then this predicts that more concentrated industries have lower average returns, all else equal, because firms in more concentrated industries engage in less innovation. We label this the *creative destruction hypothesis for stock returns*.

An alternative, but related, way to link market structure to stock returns is based on an old and influential paradigm in industrial organization known as the Structure/Conduct/Performance (S/C/P) paradigm. This work originates with Bain (1954), who links the exogenous production characteristics of an industry to a firm's pricing behavior, which in turn determines firm performance.

The observational starting point for the S/C/P paradigm is the nature of the production technology in an industry, which is taken to be exogenous. For example, the computer chip manufacturing industry has high fixed costs, since large, expensive plants must be built and customized to each new chip that is designed. The S/C/P paradigm would view these high fixed costs as a natural barrier that restricts competitive entry (*structure*). Since entry to this industry would be limited, the number of incumbent firms would be few, and each would be able to price significantly above marginal cost without fear of arousing entry (*conduct*). As a result, firms in this industry would earn supernormal economic profits (*performance*).

The S/C/P paradigm suggests that barriers to entry affect expected returns whenever differences in the number of competitors in an industry, or in the pricing practices they observe, change the risk characteristics of the firms in question. For example, barriers to entry may affect how firms optimally respond to aggregate demand shocks. Firms in high barriers-to-entry industries can respond to positive demand shocks by increasing prices or raising output without fearing competitive entry. All else equal, this raises their expected future profitability, giving them deeper pockets that help them weather downturns without facing industry exit. Thus, if exit in response to aggregate demand shocks is associated with priced distress risk, we would expect these firms to face less distress risk.[2]

Looking across industries, we would expect firms in high barriers-to-entry industries to earn lower average returns since the average distress risk would be lower in these industries. To test this prediction, one empirical approach is to measure barriers to entry directly and relate them with stock returns. However, recent work in industrial organization focuses on the fact that barriers to entry reflect the strategic choices of incumbent firms in addition to the inherent production characteristics of the industry. This is illustrated in a large body of work including Schmalensee (1978), Salop (1979), Schmalensee (1981),

[2] Industry exit could be priced if it changed the production possibilities of the economy and hence the investment opportunity set faced by investors. This would be the case if, for example, exit involved abandoning investments that are costly to reverse, or redeploying assets and human capital to production processes for which they were not originally specialized.

Sutton (1991), and Sutton (1998). The fact that barriers to entry reflect strategic choices of incumbent firms as well as the primitives of industry production technology makes it impractical to link stock returns directly to barriers to entry. In particular, the strategic nature of barriers to entry not only makes them difficult to measure, but introduces potential endogeneity with stock returns. For a variety of reasons, direct measures of barriers to entry are unattractive or incomplete.

Instead, we focus on industry concentration as a measure of barriers to entry, since it is a natural consequence of these barriers no matter how the barriers came to exist. Under the *barriers-to-entry hypothesis*, we hypothesize that firms in highly concentrated industries earn lower returns because, all else equal, they are better insulated from undiversifiable, aggregate demand shocks.

II. Data and Measures of Industry Concentration

A. Sample Selection

Our sample includes all NYSE-, AMEX-, and NASDAQ-listed securities with share codes 10 or 11 that are contained in the intersection of the CRSP monthly returns file and the COMPUSTAT industrial annual file between July 1963 and December 2001. Prior to January 1973, industry coverage is more sparse, since the CRSP sample includes NYSE and AMEX firms only. However, all of our findings hold for the 1963 to 2001 sample period as well as the 1973 to 2001 sample period. Throughout our analysis, we employ the corrections suggested in Shumway (1997) for the de-listing bias; however, these adjustments have no effect on our results.

To ensure that accounting information is already impounded into stock prices, we match CRSP stock return data from July of year t to June of year $t + 1$ with accounting information for fiscal year ending in year $t - 1$, as in Fama and French (1992). To be included in our return tests, a firm must have CRSP stock price, shares outstanding and three-digit SIC classification data for June of year t.[3] Many of our tests require the presence of COMPUSTAT data on earnings, sales, book equity, market equity, and total assets for fiscal year $t - 1$. This data requirement probably biases our sample toward larger firms, which may in turn diminish the overall variation in the concentration measures.

Book equity is stockholder's equity plus balance sheet deferred taxes and investment tax credits minus the book value of preferred stock and post-retirement assets. The book-to-market ratio is calculated by dividing book equity by COMPUSTAT market equity, which is COMPUSTAT stock price times shares outstanding at fiscal year-end. Earnings is measured before interest, and equals income before extraordinary items plus interest expense plus income statement deferred tax. Leverage is defined as the ratio of book liabilities (total assets minus book equity) to total market value of firm (COMPUSTAT market

[3] Kahle and Walkling (1996) report problems between CRSP and COMPUSTAT with regard to SIC industry classifications. To minimize any impact this may have on our results, and to maintain internal consistency with our variable construction, we disregard COMPUSTAT SIC classifications.

equity plus total assets minus book equity). For size we use CRSP market equity for June of year t. We follow Fama and French (1992) to estimate market β by computing full-period βs for portfolios sorted by size and pre-ranking β and then assigning portfolios βs to stocks in those portfolios. The pre-ranking β is estimated as the sum of the coefficients of regressions of individual monthly stock returns on contemporaneous and lagged market returns over the past 3 years.

Throughout the paper, we use three-digit SIC classifications to define industry membership. This choice balances two offsetting concerns. On the one hand, we wish to use fine-grained industry classifications so that firms in unrelated lines of business are not grouped together. On the other hand, using too fine an industry classification results in portfolios that are statistically unreliable, with firms being grouped into distinct industries arbitrarily. Choosing three-digit classifications strikes a balance between these two concerns. Although all of the results in the paper are presented with three-digit SIC classifications, in unreported tables we replicate our findings at the two- and four-digit level, and the results are qualitatively identical.

Finally, we remove regulated industries from our sample.[4] Regulated industries may face lower costs of capital either because they have lower operating risks (due to regulated entry and exit), or because their capital structure and/or capital charges are legally constrained. If regulation is correlated with industry concentration, then this could potentially explain our findings without offering any fresh insights into the structure of asset prices. Removing these industries has no material effect on our findings.

B. Measuring Industry Concentration

We measure industry concentration using the Herfindahl index, which is defined as

$$\text{Herfindahl}_j = \sum_{i=1}^{I} s_{ij}^2, \tag{1}$$

where s_{ij} is the market share of firm i in industry j. We perform the above calculations each year for each industry, and then average the values over the past 3 years. This ensures that potential data errors do not have undue influence on our Herfindahl measure.[5]

The Herfindahl measure uses the entire distribution of industry market share information to obtain a complete picture of industry concentration. Small

[4] The industries are taken from Barclay and Smith (1995).

[5] In unreported robustness tests, we vary the averaging horizon of the Herfindahl calculation from 1 year (i.e., no averaging) to 10 years. We also skip multiple years between the Herfindahl calculation and the returns, and we relate Herfindahl in the beginning of the sample to late-sample returns. These robustness checks ensure that our results are not affected by industries with large swings in Herfindahl. Our findings hold under all of these alternative specifications.

values of the Herfindahl index imply that the market is shared by many competing firms, while large values imply that market share is concentrated in the hands of a few large firms.

A common way to measure Herfindahl is to use net sales to calculate market share. We call this variable H(Sales) in our analysis. We also define H(Assets) and H(Equity) using total assets and book equity, respectively, to compute market share. The H(Equity) measure allows us to use Davis et al. (2000) data and extend our results to time periods before net sales and asset data became widely available. The measures are only imperfectly correlated with true market share, but to ensure that they produce reasonable values, we compare the three measures over the 1963 to 2001 interval, during which time all three measures are available. As Panel A of Table I shows, they are highly correlated.

C. Characteristics of Concentration-Sorted Portfolios

In Panel B, we report characteristics averaged across concentration quintiles. The spread in H(Sales) is large: The most competitive quintile has an average H(Sales) of 0.133, while the most concentrated quintile has an average of 0.982. In addition, the production, risk, and profitability characteristics of the industry quintiles tell us much about the nature of industry concentration.

Average sales and assets are significantly larger for the most concentrated quintiles, but size is smaller for the most concentrated quintile. (Skewness in the within-industry size distribution of firms is responsible for the latter result.) Firm turnover, as proxied by the number of new listings and de-listings in an industry, is significantly higher in the quintile of the most competitive industries, which suggests that barriers to entry are higher in more concentrated industries.

Measures of risk and leverage are largely flat across concentration quintiles. The average book-to-market ratio is roughly constant, as is the average β. Leverage is roughly flat across the quintiles as well.

Unlike risk and leverage, profitability shows considerable variation across quintiles. We summarize profitability with four measures. Earnings to assets (E/A in Table I) averages 1.3% for the lowest concentration quintile, jumps to 2.9% for the second lowest quintile, and is above 3% for the remaining three quintiles. Similarly, earnings to sales (E/S) ranges from 11% for the lowest concentration quintile to 13.6% for the highest concentration quintile. More concentrated industries have higher profitability on average; this is consistent with the view that industry concentration is an indirect measure of barriers to entry.

The variable labeled *V/A* is our proxy for Tobin's Q, and is simply market value of assets over book value of assets. It exhibits behavior similar to that above, ranging from 1.29 for the lowest concentration quintile to 1.70 for the highest concentration quintile. The positive correlation between Tobin's Q and industry concentration suggests that high industry-concentration firms not only have higher current profitability, but expect this profitability to persist in the future.

Table I

Summary Statistics

The sample includes all NYSE/AMEX/NASDAQ-listed securities with share codes 10 or 11 that are contained in the intersection of the CRSP monthly returns file and the COMPUSTAT industrial annual file between July 1963 and December 2001. Panel A reports summary statistics of industry concentration measures for three-digit SIC industries. The H(Sales) for an industry is formed by first calculating the sum of squared sales-based market shares of all firms in that industry in a given year and then averaging over the past 3 years. H(Assets) and H(Equity) are computed analogously, using total assets and book equity in place of sales. The right-most columns present Spearman and Pearson correlations between industry concentration measures. Spearman (rank) correlations are presented below the main diagonal, Pearson above. Panel B reports average characteristics of quintile portfolios sorted by H(Sales). Quintile 1 corresponds to the 20% of industries with the lowest concentration, while Quintile 5 corresponds to the 20% of industries with the highest concentration. Newlist is the average number of newly listed firms per year in each quintile. Delists is the average number of de-listed firms per year. Size (market equity) is CRSP price times shares outstanding (in millions of dollars). Asset is COMPUSTAT Total Assets. Sales is COMPUSTAT Net Sales. E/A is earnings before interest (income before extraordinary items + interest expense + income statement deferred tax) divided by assets; E/S is earnings divided by sales. V/A is market value of firm (market equity + total assets − book equity) divided by total assets. D/B is the ratio of dividends to book equity. Book equity is stockholder's equity (or common equity + preferred stock par value, or asset − liabilities) plus balance sheet deferred taxes and investment tax credit minus the book value of preferred stock and post-retirement asset. R&D/A is the ratio of R&D expenditure to total assets. Lev. is the ratio of book liabilities (total assets − book equity) to total market value of firm. B/M is the ratio of book equity to market equity. Beta is post-ranking beta as in Fama and French (1992). Each of these characteristics is calculated at the firm level and then averaged within each H(Sales) quintile.

Panel A: Summary of Industry Concentration Measures

										Spearman–Pearson Correlation		
	Mean	Median	*SD*	Max	Min	20%	40%	60%	80%	H(Sales)	H(Assets)	H(Equity)
H(Sales)	0.544	0.490	0.310	1.000	0.025	0.231	0.385	0.611	0.944	1.000	0.976	0.951
H(Assets)	0.549	0.499	0.307	1.000	0.024	0.233	0.397	0.618	0.936	0.976	1.000	0.964
H(Equity)	0.546	0.502	0.308	1.000	0.024	0.230	0.405	0.609	0.931	0.953	0.966	1.000

Panel B: Characteristics of H(Sales) Sorted Quintile Portfolios

Rank	H(Sales)	Newlist	Delists	Size	Asset	Sales	E/A	E/S	V/A	D/B	R&D	R&D/A	Lev.	B/M	Beta
Low	0.133	267.40	214.60	531.3	1200.4	582.5	0.013	0.110	1.293	0.026	35.293	0.075	0.437	0.798	1.579
2	0.287	126.21	84.70	527.8	645.1	509.6	0.029	0.111	1.257	0.024	21.226	0.060	0.399	0.742	1.632
3	0.470	60.47	42.70	607.4	1204.8	786.7	0.036	0.116	1.327	0.031	21.759	0.040	0.432	0.809	1.595
4	0.745	41.51	23.82	606.4	1087.9	629.3	0.038	0.124	1.558	0.041	17.164	0.037	0.428	0.787	1.606
High	0.982	20.13	8.68	431.3	1604.9	717.6	0.037	0.136	1.695	0.036	13.059	0.027	0.421	0.767	1.609

The dividend payout ratio (D/B) also increases with industry concentration. Since Fama and French (2000) and many others relate dividend policy to expected profitability, we take this as further evidence that firms in high concentration industries are more profitable. We discuss this issue in further detail in Section IV.

To get a sense of how Schumpeter's prediction squares with our data, we also report two measures of R&D intensiveness. The first is simply gross R&D expenditure, which declines substantially as concentration increases, falling from an average of \$35 million per firm-year for the least concentrated quintile to \$13 million for the highest concentration quintile. When we scale by total assets, we see the same pattern, with the R&D to asset ratio falling from 7.5% for the lowest concentration quintile to 2.7% for the most concentrated quintile.

In Table II, we report Fama and Macbeth (1973, henceforth FM) regressions of the cross section of industry concentration measures on industry average characteristics. We estimate equations of the following form:

$$H(Sales)_{jt} = \alpha_t + \sum_{n=1}^{N} \lambda_{nt} X_{jt} + \varepsilon_{jt}, \tag{2}$$

where the X_{jt} are industry average characteristics. Regressions are run for every year t from 1963 to 2001, and the time-series means of annual cross-sectional coefficient estimates are reported along with the time-series t-statistics. This procedure allows for multivariate correlation analysis, and it is robust to cross-correlated error terms. Thus, the resulting coefficients can be interpreted as simple or conditional correlations between concentration and industry-average characteristics, and appropriate statistical inferences can be drawn about the magnitude of these relations.

The row labeled "Simple" reports results from FM regressions of concentration on each characteristic in isolation. (Thus, there are eleven separate univariate regressions reported in a single row.) Each row under the panel labeled "Multiple" reports a single regression in which multiple characteristics are included as independent variables simultaneously. This provides conditional correlations of H(Sales) on industry characteristics.

When we combine the correlations reported here with the descriptive statistics from Table I, a picture of industry concentration emerges that is consistent with the prior literature discussed in Section I and that is important for the interpretation of our findings. Measures of profitability are positively correlated with industry concentration. Earnings to assets, earnings to sales, and market-to-book ratios are all highly positively correlated with industry concentration, both unconditionally, and conditional on other industry characteristics.

Concentrated industries have large asset bases and high unit profitability. In addition, R&D to assets is much lower for these industries. Thus, highly concentrated industries have high capitalized future profitability but they do not engage in risky innovation (they do not have high levels of R&D). These descriptive statistics paint a picture of concentrated industries as innovation-poor, profit-rich industries with high barriers to entry.

Table II

Fama–MacBeth Regressions of H(Sales) on Industry Average Characteristics

This table presents Fama–MacBeth regressions of the H(Sales) index with other industry average characteristics. The variables are defined according to Table I. Every year, a cross-sectional regression is estimated. The time-series mean of the annual regression coefficients and the time-series t-statistics (appearing below) are reported. In Panel A, each coefficient is obtained from a simple (univariate) regression of H(Sales) on each characteristic alone. Panel B reports the results of multiple (multivariate) regressions of H(Sales) on a series of industry characteristics.

ln(Size)	ln(Assets)	ln(Sales)	E/A	E/S	V/A	D/B	R&D/A	Leverage	ln(B/M)	Beta
					Panel A: Simple Regressions					
−0.040	−0.032	−0.043	0.179	0.213	0.014	−0.002	−0.984	−0.033	−0.028	0.091
−13.45	−17.97	−22.81	4.77	7.15	3.71	−0.02	−4.97	−1.73	−5.34	4.92
					Panel B: Multiple Regressions					
−0.034			0.522				−1.443	−0.056	−0.057	−0.011
−3.89			3.28				−6.99	−0.84	−6.08	−0.23
	−0.027		0.525				−1.527	0.026	−0.044	0.021
	−3.26		3.32				−7.22	0.44	−3.75	0.46
		−0.027	0.489				−1.514	0.012	−0.036	0.021
		−4.09	2.50				−7.26	0.20	−3.35	0.55
−0.039				0.580			−1.399	−0.017	−0.056	−0.016
−4.36				7.57			−6.81	−0.24	−5.80	−0.38
−0.023					0.024		−1.487	−0.114		0.019
−1.91					4.23		−7.05	−1.80		0.36
−0.034			0.542			0.570	−1.361	−0.043	−0.044	0.011
−3.84			3.21			2.97	−6.75	−0.64	−3.83	0.22

III. Concentration and the Cross-Section of Returns

A. The Concentration Spread

Table III relates industry concentration to the cross section of average stock returns, measured both at the industry and firm level. In June of each year, industries are sorted into quintiles based on their Herfindahl index. We then report average monthly returns and t-statistics for these portfolios, as well as the difference between Quintile 5 (most concentrated) and Quintile 1 (least concentrated).

The first row in the left panel presents raw average returns computed by equally weighting firms within each concentration portfolio. Looking across Herfindahl quintiles, firms in the least concentrated (most competitive) industries earn an average return of 1.52% per month. This declines to 1.26% per month for firms in the most concentrated quintile. The spread between the two is -0.26% per month, which carries a statistically significant t-statistic of 2.14.

Because concentration is an attribute of an industry, not a firm, there is flexibility in how quintile returns are measured. The right panel reports returns calculated by first forming industry portfolios, and then equally weighting industry returns within each concentration quintile. These industry-level returns mirror the firm-level results. In each case, we see a large and statistically significant spread between the most concentrated and the most competitive quintiles.

Since Table I shows that industry concentration is associated with a number of known determinants of average returns, we also report characteristics-adjusted returns. We use the procedure in Daniel et al. (1997) to adjust individual stock returns for size, book-to-market, and momentum. All firms in our sample are first sorted each month into size quintiles, and then within each size quintile we further sort firms into book-to-market quintiles. Within each of these 25 portfolios, firms are again sorted into quintiles based on the firm's past 12-month return, skipping the most recent month. Stocks are averaged within each of these 125 portfolios to form a benchmark that is subtracted from each individual stock's return. The expected value of this excess return is zero if size, book-to-market, and past one-year return completely describe the cross section of expected returns.

The characteristic-adjusted average returns of the above quintile portfolios as well as the average spread between Quintile 5 and Quintile 1 are reported in the second row of each panel. Even after adjusting for these characteristics, we still see a significant spread in average returns across concentration quintiles. Interestingly, adjusted returns for the quintile of the most competitive industries (Q1) are positive and statistically significant, and they decrease monotonically to negative and statistically significant for the quintile of the most concentrated industries (Q5). However, the adjustments only serve to increase the spread in returns. The spread for the Herfindahl index jumps to -0.36% per month, 10 basis points higher in absolute value than the raw returns figure. Similarly, the spread for industry portfolios grows two basis points

Table III

Industry Concentration and the Cross-Section of Average Stock Returns

In June of each year, industries are grouped into quintiles based on their H(Sales) value. The average monthly returns (in percent) of the quintile portfolios are reported, as well as the difference between Quintile 5 (most concentrated) and Quintile 1 (least concentrated). We report t-statistics below average returns. Firm-level raw returns are unadjusted returns averaged across firms within the same concentration quintile. Firm-level adjusted returns are calculated by subtracting the return on a characteristic-based benchmark from each firm's return, then averaging within the same concentration quintile. Characteristic-based benchmarks are constructed following Daniel et al. (1997) to account for the premia associated with size, book-to-market, and momentum. Industry-level raw and adjusted returns are computed similarly, except that individual stock raw and adjusted returns are first averaged within each industry, and then averaged across industries within the same concentration quintile. During the 1927 to 1951 sample period, H(Sales) is replaced by H(Equity). This is constructed from Davis et al. (2000) data.

	Firm-Level Returns						Industry-Level Returns					
	Quintile						Quintile					
	1	2	3	4	5	5 − 1	1	2	3	4	5	5 − 1
Raw and adjusted returns, 63/07–01/12												
Raw	1.52	1.28	1.38	1.26	1.26	−0.26	1.35	1.29	1.30	1.24	1.09	−0.26
	5.02	4.37	4.55	4.3	4.43	−2.14	4.71	4.39	4.58	4.2	3.74	−2.35
Adjusted	0.26	0.00	0.02	−0.03	−0.10	−0.36	0.09	−0.02	−0.09	−0.15	−0.18	−0.28
	3.81	0.07	0.34	−0.76	−2.60	−3.8	2.59	−0.55	−1.77	−2.61	−1.80	−2.34
Adjusted returns, alternative sample periods												
27/01–01/12	0.11	−0.01	0.05	−0.02	−0.04	−0.15	0.06	0.00	−0.08	−0.07	−0.12	−0.17
	2.18	−0.26	1.21	−0.59	−1.33	−2.43	2.16	−0.03	−2.19	−1.74	−1.69	−2.18
51/07–01/12	0.21	0.00	−0.01	−0.03	−0.07	−0.28	0.07	−0.02	−0.08	−0.09	−0.13	−0.20
	3.94	−0.05	−0.21	−0.78	−2.29	−3.83	2.57	−0.64	−2.15	−2.05	−1.71	−2.29
74/01–01/12	0.31	0.02	0.01	−0.03	−0.13	−0.44	0.05	−0.05	−0.06	−0.15	−0.30	−0.35
	3.75	0.29	0.10	−0.62	−2.79	−3.8	2.56	−1.13	−0.97	−2.32	−2.62	−2.88
Adjusted returns, alternative concentration measures, 63/07–01/12												
H(Assets)	0.09	0.17	0.01	−0.02	−0.12	−0.20	0.10	−0.03	−0.13	−0.08	−0.16	−0.26
	1.19	2.64	0.39	−0.44	−2.95	−2.12	2.95	−0.75	−2.72	−1.42	−1.49	−2.12
H(Equity)	0.16	0.12	0.05	−0.10	−0.08	−0.24	0.11	−0.09	−0.09	−0.11	−0.16	−0.27
	2.30	2.20	0.95	−2.19	−2.09	−2.52	3.24	−2.16	−1.85	−1.80	−1.54	−2.28
Segment-level	0.18	−0.14	−0.26	−0.44	−0.45	−0.63	0.47	0.09	0.07	−0.20	−0.27	−0.74
H(Sales)	2.54	−1.83	−2.66	−3.83	−2.27	−2.75	2.89	0.94	0.86	−2.56	−3.40	−3.32

to −28 basis points. Together, this suggests that the return premium associated with industry concentration is independent from those of size, book-to-market, and momentum, and that controlling for industry concentration is important for understanding the cross section of stock returns.

In the second set of numbers, we take a number of steps to control for a variety of potential explanations of our results. We extend our results back to 1927 by using an H(Equity) concentration measure constructed from the Davis–Fama–French files. The row reporting results from 1951 to 2001 uses the entire length of the Compustat sample to compute Herfindahl indices, in spite of the fact that only NYSE-listed firms are present until 1963. Data from 1974 to 2001 provide our results for the subsample in which we have Nasdaq-, AMEX-, and NYSE-listed stocks. The concentration spread is robust to each of these specification choices.

The third set of numbers pushes the robustness question further with results from quintiles formed on alternative concentration measures.[6] The concentration premium is robust to using H(Assets) or H(Equity) to form concentration quintiles. The concentration premium also shows up significantly when we use net sales from the Compustat Business Segment file (available 1985 to 2001) to attribute sales of conglomerate firms to their respective industries.

B. Fama–MacBeth Cross-Sectional Regressions

To further examine the relation between industry concentration and average stock returns, we conduct Fama–MacBeth (FM) regressions of monthly stock returns on industry concentration and other characteristics. In Panel A of Table IV, we report regressions of industry portfolio returns regressed on industry characteristics and the H(Sales) measure. The time-series average of each cross-sectional regression loading is reported along with its time-series t-statistic. These regressions provide a robustness check of the relationship between industry concentration and average returns without imposing quintile breakpoints, and they allow us to control for additional alternative explanations.

The first column of Panel A shows that more concentrated industries earn lower average returns, consistent with our previous results from quintile portfolios. The cross-sectional regression coefficient on the H(Sales) index is negative and statistically significant at the 5% level. The next seven rows demonstrate that industry average returns are positively related to industry average book-to-market, leverage and momentum (past 1-year's industry return), negatively associated with industry average size, and insignificantly related to industry average market β.

The last two rows reexamine the industry concentration effect, controlling for the above characteristics. These rows show that taking these variables into

[6] In tables available from the authors, we also replicate our findings on the much smaller sample of observations for which Census of Manufactures definitions of industry concentration are available. In addition, we repeat our findings using the ratio of the sales of the top five firms in an industry to total industry sales (the five-firm ratio).

Table IV

Fama–MacBeth Cross-Sectional Regressions of Industry-Level and Firm-Level Returns

This table presents results from industry-level (Panel A) and firm-level (Panel B) Fama–MacBeth cross-sectional regressions estimated monthly between July 1963 and December 2001. In Panel A, industry average returns are regressed on industry H(Sales) measure, industry average values of ln(Size), ln(B/M), Leverage, Beta, and the past 1-year return on the industry portfolio (Momentum). In Panel B, individual stock returns are regressed on H(Sales) value of the industry to which each stock belongs, firm-level ln(Size), ln(B/M), Leverage, Beta, and the past 1-year stock return (Momentum). Time-series average values of the monthly regression coefficients are reported with time-series *t*-statistics appearing below.

H(Sales)	ln(Size)	ln(B/M)	Momentum	Beta	Leverage
		Panel A: Industry-Level Regressions			
−0.30					
−2.41					
	−0.12				
	−1.54				
		0.39			
		4.16			
			1.03		
			4.21		
				−0.18	
				−0.43	
					0.98
					2.96
	−0.24	0.28	0.95	−0.95	0.08
	−2.81	2.77	4.40	−2.56	0.24
−0.30	−0.12	0.29	0.90		
−2.58	−1.57	3.07	3.93		
−0.31	−0.25	0.27	0.94	−1.00	0.04
−2.85	−2.98	2.68	4.36	−2.73	0.12
		Panel B: Firm-Level Regressions			
−0.35					
−2.41					
	−0.14	0.35	0.56		
	−2.62	4.55	3.34		
−0.44	−0.14	0.35	0.55		
−3.75	−2.63	4.62	3.32		
				0.26	
				0.90	
					0.76
					2.81
	−0.18	0.38	0.60	−0.39	−0.30
	−3.78	6.41	3.81	−1.87	−1.56
−0.42	−0.18	0.39	0.59	−0.41	−0.33
−3.42	−3.81	6.62	3.78	−1.95	−1.70

account does not destroy the significance of the industry concentration effect. By including leverage in our regressions, we control for another possible explanation for our findings, namely, that competitive industries have higher leverage, thereby raising the required return on equity mechanically. In the univariate

FM regression, leverage works in the predicted direction, but controlling for other characteristics weakens leverage.

Thus, while the results of Table I suggest that industry concentration is correlated with other industry characteristics that describe average returns, the results from the top of Table IV suggest that those correlations are not the driving forces behind the inverse relationship between industry concentration and average stock returns.

Panel B of Table IV repeats the analysis described above, but replaces industry portfolio returns with firm-level stock returns and replaces industry characteristics with firm-level measures of size, book-to-market, leverage, and market β. These results mirror those obtained in the industry portfolio regressions. FM regressions of individual stock returns on the Herfindahl index alone produce an average slope coefficient of -0.35% with a t-statistic of -2.41. Accounting for the premia associated with known return predictors strengthens these results. Introducing size, book-to-market, past 1-year return, leverage and market β to the cross-sectional regressions raises both the point estimates and the t-statistics for industry concentration.

The conclusion that emerges from this section is that not only do industry returns vary with industry concentration, but so do individual stock returns: Firms in concentrated industries earn lower stock returns than firms in more competitive industries. The results hold under a variety of different empirical strategies, and are robust to whether or not we control for characteristics such as size, book-to-market ratio, and past returns, both at the firm and industry levels. These controls suggest that the industry concentration effect that we identify is not being driven by correlations with other determinants of expected returns, or through capital structure choice.

IV. Industry Concentration and Profitability Surprises

The preceding analysis demonstrates a statistically reliable and economically meaningful link between market structure and average stock returns. However, from a standard Campbell and Shiller (1988) decomposition we know that returns must, by their very definition, equal the sum of expected returns, shocks to cash flows, and shocks to discount rates. Thus, persistent differences in cash flow surprises across industries with different market structures could be responsible for our findings.

This section explores this issue. If differences in average returns across concentration quintiles are due to persistent in-sample cash flow shocks that need not persist in the future, then while industry concentration may happen to explain average returns during the period of our analysis, concentration would still be unrelated to true expected returns.

We extend the Fama and French (2000) profitability model by adding lagged profitability, following Vuolteenaho (2002). Specifically, we are interested in models of the form

$$\frac{E_t}{A_t} = \alpha_0 + \alpha_1 \frac{V_t}{A_t} + \alpha_2 DD_t + \alpha_3 \frac{D_t}{B_t} + \alpha_4 \frac{E_{t-1}}{A_{t-1}} + \varepsilon_t, \tag{3}$$

where E/A is earnings scaled by total assets, V/A is the ratio of market value of assets to book assets, DD is a dummy variable for non-dividend-paying firms, and D/B is the ratio of dividend payments to book equity. Expected profitability is the fitted value from this regression, and unexpected profitability is regression error. To estimate this model, we follow Fama and French (2000), and estimate cross-sectional regressions each year.

Panel A of Table V presents average coefficients from the cross-sectional regressions for three profitability specifications. The first row, labeled "Firm-Level," reports FM regressions of firm-level profitability on firm-level characteristics. The row labelled "Industry Total" computes a single-earnings measure for each industry, scales this by total industry assets, and then regresses it on the four independent variables that are constructed similarly. Finally, the row labeled "Industry Average" reports regressions of the industry average profitability on industry average values of the variables described above.

Our numbers closely match those reported in Fama and French (2000). Specifically, we obtain statistically positive loadings on D/B and statistically negative loadings on the dividend dummy. Profitability loads positively and significantly on V/A, suggesting that V/A captures differences across firms in expected profitability that are missed by the two dividend variables. Our regression R^2 values, ranging from 42% to 50%, are about twice as high as those reported in Fama and French (2000), due largely to the inclusion of lagged profitability as suggested in Vuolteenaho (2002).

In the rest of Table V we take the regression errors from Panel A and relate them to industry concentration. We do this for two measures of unexpected profitability. The variable UP_t is the in-sample regression error from the FM regression reported in Panel A. The variable UP_{t+1} is the one-period-ahead regression error: This is the error obtained by using the FM coefficients from a regression in year $t-1$ to forecast the profitability in year t, and treating this forecast error as unexpected profitability.

In Panel B, Quintiles 1 to 5 report the average unexpected profitability by concentration quintile. As in previous tables, Quintile 1 is the least concentrated and Quintile 5 the most concentrated quintile. If our results were driven by cash flow shocks, then we should expect to see large positive average profitability shocks for Quintile 1 and large negative shocks for Quintile 5.

Instead, we see the opposite. Concentrated industries have experienced better-than-expected profitability over the 1963 to 2001 period, while competitive industries have experienced poorer-than-expected profitability. Unexpected profitability is increasing as we move toward more concentrated quintiles. With firm-level UP_t and UP_{t+1}, and with industry-level UP_t measures, we can reject the null hypothesis that profitability is the same across all five concentration quintiles.

In the far-right column of Panel B, labeled "FM," we report FM regressions of UP on industry concentration. These results mirror the findings obtained by quintile breakdowns. In all but one specification, there is a statistically positive relation between unexpected profitability and industry concentration. In one case (industry average UP_{t+1}) we cannot reject the null of zero correlation;

Table V

Industry Concentration and Profitability Surprises

This table examines the relation between profitability surprises and industry concentration. Firm-level E/A is firm-level earnings to assets. Industry total (E/A) is the total earnings in the industry divided by total assets of the industry. Industry average E/A is the industry average earnings-to-assets ratio. Expected profitability is obtained from Fama–MacBeth regressions of the form

$$\frac{E_t}{A_t} = \alpha_0 + \alpha_1 \frac{V_t}{A_t} + \alpha_2 DD_t + \alpha_3 \frac{D_t}{B_t} + \alpha_4 \frac{E_{t-1}}{A_{t-1}} + \varepsilon_t,$$

following Fama and French (2000) and Vuolteenaho (2002). Unexpected profitability is the regression error from this regression. In Panel B, we group industries according to concentration quintiles and report average unexpected profitability. The variable UP_t is the unexpected profitability from in-sample regressions, while UP_{t+1} is the regression error obtained by predicting next year's profitability using next year's regressors but parameter values obtained from year t. The two "Tests" columns report the F-statistic for the joint equality of Quintiles 1 through 5, as well as the t-statistic for the equality of Quintiles 1 and 5. The final column reports the Fama–MacBeth coefficient from a regression of unexpected profitability on the H(Sales) index.

Panel A: Expected Profitability Regressions

Profitability Measure	α_0	V/A	DD	D/B	ROA_{t-1}	Avg. R^2
Firm-level E/A	0.0187	0.0120	−0.0164	0.0679	0.5341	0.4267
	11.91	7.46	−8.19	7.12	34.38	
Industry total E/A	0.0089	0.0136	−0.0027	0.0890	0.5684	0.5099
	5.69	9.53	−0.69	5.13	25.95	
Industry average E/A	0.0196	0.0139	−0.0132	0.0801	0.5130	0.4600
	10.69	7.34	−6.02	4.05	27.88	

Panel B: Unexpected Profitability by Concentration Quintile

	Quintiles					Tests		
Profitability:	1	2	3	4	5	$F\,(1=2=3=4=5)$	$t\,(1=5)$	FM
Firm-level								
UP_t	−0.0017	0.0014	0.0041	0.0025	0.0023	5.95	3.07	0.0101
	−4.14	2.45	3.96	2.90	1.88	0.0002	0.0036	4.49
UP_{t+1}	−0.0043	−0.0013	0.0029	0.0024	0.0002	2.81	1.65	0.0124
	−1.98	−0.76	1.97	1.47	0.14	0.0268	0.1035	2.85
Industry total								
UP_t	−0.0013	−0.0007	0.0007	0.0004	0.0025	3.06	2.58	0.0043
	−2.48	−1.25	1.06	0.49	1.81	0.0179	0.0131	2.66
UP_{t+1}	−0.0015	−0.0014	0.0004	−0.0008	0.0014	0.65	1.27	0.0038
	−1.07	−0.83	0.30	−0.61	0.77	0.6268	0.2065	2.01
Industry average								
UP_t	−0.0018	0.0002	0.0015	−0.0004	0.0013	2.95	2.31	0.0031
	−3.70	0.32	2.10	−0.60	1.05	0.0214	0.0254	2.21
UP_{t+1}	−0.0030	−0.0005	0.0001	−0.0015	0.0007	1.08	1.79	0.0026
	−2.19	−0.39	0.04	−1.16	0.47	0.3699	0.0773	1.36

however, we never see results going in the direction that would be required to support cash flow shocks as an explanation for our findings. In short, differences in unexpected profitability cannot explain our findings. In fact, since the shocks go in the opposite direction of the return spread, this suggests that the true spread in expected returns is more pronounced than the spread that we observe in the data.

V. Concentration and Time-Series Variation in Returns

A. Time-Series Variation of the Industry Concentration Premium

This subsection links changes in the concentration premium to various risk factors and business cycle indicators. This allows us to examine the question of whether the concentration premium remains significant after taking existing risk factors into account, and also whether it exhibits variation over the business cycle that is consistent with the risk-based explanations we argue are responsible for its existence.

In Table VI, we report results from the following time-series regressions of monthly concentration premia on risk factors and economic indicators:

$$\lambda_t^H = \alpha + \sum_{i=1}^{I} \beta_i F_{it} + \sum_{j=1}^{J} \gamma_j X_{jt} + \varepsilon_t, \tag{4}$$

where F_{it} are returns to factor-mimicking portfolios in month t, and X_{jt} are month-t values of the business cycle indicators. The dependent variable, λ_t^H, is the time series of H(Sales) risk premia generated from the FM regressions reported in Panel B of Table IV, in which the cross section of individual stock returns is regressed on industry concentration, controlling for other characteristics.

In the first row, the monthly concentration premium is regressed against the market excess return. The next row adds the factor-mimicking portfolios associated with the size effect (SMB) and book-to-market effect (HML). The following row adds a momentum factor-mimicking portfolio to the Fama–French factors as in Carhart (1997) to estimate a four-factor model. As the table indicates, the regression intercepts are both economically and statistically significant in the presence of various risk factors. The H(Sales) premium drops slightly (in absolute value) from −42 basis points (Table IV, Panel B, last row) to −40 per month when regressed on the market excess return. The adjusted R^2 from this regression is close to zero. Controlling for the Fama and French (1993) factors actually increases the premium to −0.46%, whereas the R^2 goes up to 16.1%. Adding the momentum factor decreases the premium to −0.33% (still significant), and there is a slight increase in the R^2 to 19%.

These three sets of regressions show that the concentration premium cannot be explained by known risk factors, which reinforces the finding in Section III that industry concentration contains independent information about the cross section of average returns.

Table VI

Time-Series Variation of the Concentration Premium

This table presents results from time-series regressions of the H(Sales) premium on various asset pricing factors and business cycle indicators. The H(Sales) premium is obtained from monthly Fama–MacBeth cross-sectional regressions of stock returns on industry H(Sales) index, controlling for other characteristics (see the last row of Table IV, Panel B). RMRF is the market excess return. SMB and HML are size and B/M factor-mimicking returns (see Fama and French (1993) for description). MOM is the momentum factor as in Carhart (1997). The factor data come from Ken French's website. INF is the monthly rate of inflation, obtained from the St. Louis Federal Reserve Economic Database (FRED). Term is the term spread, the difference between 10-year and 1-year treasury constant maturity rates. T-bill is the 30-day T-bill rate. g_t and g_{t+1} are the current and next four quarters' growth rates in GDP, also obtained from FRED. Alpha is the intercept from time-series regression. t-statistics are reported below parameter estimates.

Alpha	RMRF	SMB	HML	MOM	INF	Term	T-Bill	g_t	g_{t+1}	Adj. R^2
−0.40	−0.049									0.0072
−3.87	−1.99									
−0.46	0.0416	−0.2171	0.1729							0.1608
−4.27	1.32	−3.13	2.74							
−0.33	0.0341	−0.217	0.1225	−0.1129						0.1904
−3.4	1.14	−3.41	2.49	−2.83						
−1.02					1.44					0.0263
−4.37					3.75					
−0.59						−0.9313				0.0406
−4.49						−4.56				
−1.08							0.1011			0.0106
−4.16							3.27			
−2.02								0.0661	0.1324	0.0449
−3.79								0.71	3.15	
−0.84	0.0435	−0.2149	0.1202	−0.1118	1.2163					0.2086
−4.43	1.45	−3.46	2.48	−2.85	3.53					
−0.49	0.042	−0.2253	0.1078	−0.1118		−0.8871				0.227
−4.33	1.44	−3.63	2.3	−2.89		−4.82				
−0.90	0.0403	−0.216	0.1219	−0.1135			0.0877			0.1981
−3.63	1.35	−3.41	2.49	−2.86			2.61			
−2.04	0.0374	−0.2348	0.1069	−0.1106				0.1109	0.1358	0.2419
−5.33	1.27	−3.89	2.38	−3.02				1.15	4.47	
−0.50	0.0435	−0.2239	0.1081	−0.1111	0.6036	−0.8156	−0.0349			0.2263
2.47	1.48	−3.64	2.31	−2.89	1.55	−3.82	−0.88			
−2.36	0.0426	−0.2325	0.1064	−0.1108	0.2096	−0.0303	0.0576	0.0775	0.1282	0.2433
−3.36	1.45	−3.85	2.34	−3.01	0.54	−0.1	1.19	0.77	2.87	

In the next three rows we regress the concentration premium on the inflation rate, term spread, and T-bill rate. Inflation is measured by the growth rate of the consumer price index (CPI). The term spread is the difference between 10-year and 1-year treasury constant maturity rates. These variables are obtained from the Federal Reserve Bank of St. Louis, and are shown elsewhere in the literature to track business cycle fluctuations (see, e.g., Fama and French (1989)). The results indicate that the H(Sales) premium carries positive and statistically significant loadings on the inflation rate and the T-bill rate. Since the H(Sales) premium has a negative mean value, this means that the concentration premium grows (in absolute value) as the business cycle declines, since both the inflation rate and the T-bill rate rise during economic expansion and fall during economic contraction. The concentration premium also loads negatively on the term spread. Since the term spread tends to decrease as the business cycle moves from trough to peak, this finding is consistent with the loadings on the inflation rate and the T-bill rate, which indicates that the concentration premium diminishes as the economy takes an upturn. This is again consistent with a risk interpretation: The spread in returns between firms that are insulated from economic distress and those that are not grows as economic conditions deteriorate.[7]

In addition to examining inflation, the T-bill rate, and the term premium, we also include current and future GDP growth rates to directly examine the relation between the concentration premium and economic activities. The variable g_t is the current quarterly GDP growth rate, while g_{t+1} is the GDP growth rate over the next four quarters. There is a positive correlation with current GDP growth, but this correlation is not statistically significant. There is a much larger, and highly significant, correlation with GDP growth over the next year. This is in line with the result that short-horizon stock returns contain forward-looking information about the strength of economic activities over many future periods (see, e.g., Fama (1990), Kothari and Shanken (1992)). This supports a risk-based interpretation of the industry concentration effect, since it indicates that the magnitude of the concentration premium grows as the economic outlook deteriorates. Replacing GDP growth with industry production growth produces qualitatively similar, but slightly weaker, results.

In the remaining rows of this table, we regress the concentration premium on risk factors and business cycle indicators. Controlling for business cycle movement in addition to factor returns raises the regression intercept and R^2, but weakens the loading on the inflation rate, the term spread, and the T-bill rate. The loading on 1-year-ahead GDP growth remains highly significant. Nevertheless, the message remains largely unaltered: The premium associated with industry concentration is not spanned by existing factors and it exhibits sensible business cycle variation. The fact that the concentration premium grows during downturns, when economic distress is relatively greater, speaks in favor of the hypothesis that industry concentration is a mechanism through which aggregate shocks are propagated through the equity market.[8]

In an efficient market in which assets are priced rationally, industry concentration must proxy for sensitivity to a systematic risk factor in stock returns.

[7] The predictive power of these business cycle variables for the concentration premium is low. At most they explain 4% of the total variation in monthly concentration premium. However, this is not unusual given previous studies (e.g., Fama and French (1989), Lewellen (1999)), which show that ex ante instruments can only account for a small portion of the time-series variation in monthly stock returns.

[8] In unreported tables, we also include cash flow and discount rate news factors from Campbell and Vuolteenaho (2004). The concentration premium loads on them in a manner that is consistent with the interpretation offered above.

In unreported tables, we follow the logic offered in Fama and French (1993) and employ time-series regressions to explore this question. We find that a mimicking concentration factor captures substantial common variation in stock returns that is left unexplained by existing asset pricing factors. In addition, spread in returns across concentration portfolios is related to the spread in loadings on the concentration factor. Moreover, while existing factor models fail to price these concentration portfolios, including the concentration mimicking factor completely explains the industry concentration effect in average returns. The regression intercepts are all within 10 basis points of 0 and a Gibbons–Ross–Shanken (1989) test cannot reject the null that the constant terms are jointly 0.

B. Can the Concentration Premium Explain Existing Factors?

Since we show in Table VI that the concentration premium is not spanned by existing asset pricing factors, we now turn the tables and examine how much of existing asset pricing factors can be explained by the concentration premium. This is presented in Table VII, where we report time-series regressions of existing asset pricing factors on the conditional H(Sales) premium.

First we regress returns from a number of factor-mimicking portfolios on the conditional H(Sales) premium. This is presented in Panel A of Table VII. The first column, labeled "Mean," reports the unconditional mean of the mimicking factor returns over the 1963 to 2001 period. The remaining three columns report the conditional mean, the loading on the conditional H(Sales) premium, and the regression R^2.

The first row examines the excess market return. The unconditional mean of the excess market return is 47 basis points per month in our sample, but this drops to a statistically insignificant 36 basis points when we account for comovement with the H(Sales) premium. The regression R^2, however, reveals that very little of the variation in the market premium is explained by the concentration premium.

The concentration premium does a better job explaining the size factor, SMB. The size factor loads negatively and highly negatively significantly on the conditional H(Sales) premium, leaving a constant term that is statistically zero. The adjusted R^2 indicates that we explain 13% of the variation in the size premium with the conditional H(Sales) premium. This is only a modest success, however, as the unconditional size premium is a statistically insignificant 21 basis points in our sample.

The concentration premium does less well at explaining the book-to-market and momentum factors. The momentum factor does not load statistically significantly on the concentration premium, and its conditional mean is only slightly different from its unconditional mean.

Interestingly, controlling for the concentration premium actually strengthens the book-to-market factor. HML loads significantly positively on the conditional H(Sales) premium, and the concentration premium explains 12% of the variation in HML. But the mean of HML increases from 42 basis points per

Table VII

Can the Concentration Premium Explain Existing Factors?

This table reports results from time-series regressions of mimicking portfolio returns and cross-sectional premia of asset-pricing factors on the conditional H(Sales) premium. In Panel A, the dependent variables are factor mimicking returns obtained from Ken French's website. RMRF is the market excess return. SMB and HML are size and B/M factor mimicking returns (see Fama and French (1993) for description). MOM is the momentum factor as in Carhart (1997). In Panel B, the dependent variables are cross-sectional premia obtained from monthly Fama–MacBeth regressions of stock returns on ln(Size), ln(B/M), and Momentum. The column labeled "Mean" is the average value of the factor mimicking return or premium. Alpha is the regression intercept from regressing factor return/premium on H(Sales) Premium. The column labeled "H(Sales) Premium" reports the loading of the factor/premium on H(Sales) premium. The final column reports the adjusted R^2 of the regression. Point estimates are reported with t-statistics appearing below.

LHS	Mean	Alpha	H(Sales) Premium	Adj. R^2
Panel A: Explaining Factor Returns				
RMRF	0.47	0.36	−0.2827	0.02
	2.26	1.66	−3.31	
SMB	0.21	0.00	−0.5055	0.13
	1.35	0.00	−4.56	
HML	0.42	0.59	0.4410	0.12
	2.98	4.31	6.23	
MOM	0.87	0.82	−0.1430	0.01
	4.75	4.69	−0.82	
Panel B: Explaining Cross-Sectional Factor Premia				
Size premium	−0.13	−0.09	0.1194	0.07
	−2.65	−1.77	3.38	
B/M premium	0.35	0.44	0.2331	0.13
	4.83	6.48	5.41	
Momentum premium	0.56	0.58	0.0400	0.00
	3.54	3.81	0.42	

month unconditionally to 59 basis points per month conditionally. This conditional mean HML value is highly statistically significant (and different from the unconditional value).

In Panel B of Table VII we replace the asset pricing factors with the factor premia obtained from cross-sectional FM regressions. We focus on size, book-to-market, and a firm's lagged 1-year return. The results largely mirror those obtained in Panel A; namely, the size premium disappears, the momentum premium remains unexplained, and book-to-market premium increases. This points to an interesting interaction between concentration and the book-to-market premium, which we take up in more detail in the next section.

VI. Value, Growth, and Industry Concentration

Based on the fact that the book-to-market premium grows in magnitude when we control for industry concentration, we turn next to FM regressions that explore the interaction of book-to-market and concentration. These are presented in Panel A of Table VIII. Returns are at the firm level (as in Panel B of Table IV) and the final column is an interaction term between industry concentration and firm-level book-to-market. The coefficient on the interaction term is positive and statistically significant, suggesting that the premium associated with being a high book-to-market firm grows as industry concentration increases.

To get a sense of the economic magnitude involved here, next we examine returns to book-to-market portfolios for different levels of industry concentration in a five-by-five grid. At the end of June of each year, we sort industries into concentration quintiles according to their concentration (H(Sales)) value. Then within each industry, we calculate quintile breakpoints for book-to-market and place firms into five portfolios.[9] Finally, within each concentration group, we pool firms with the same book-to-market ranking into one portfolio and calculate the average returns from July to June of the following year. Our sorting approach also guarantees that one industry with particularly large variation in book-to-market does not dominate the tail portfolios of the five-by-five grid; instead, there is equal representation across industries in each book-to-market quintile. Panel B of Table VIII reports the average value-weighted monthly returns of the five book-to-market portfolios as well as the difference in returns between Quintile 5 and Quintile 1 for each H(Sales) group; Panel C reports equally weighted returns. Each row demonstrates the prevalence of the book-to-market effect within each concentration group.

As the table indicates, the spread in returns associated with the book-to-market ratio is largest among the most concentrated industries. For example, high book-to-market stocks outperform low book-to-market stocks by 47 basis points per month in the lowest Herfindahl quintile, and this number grows to 76 basis points per month for the highest Herfindahl quintile. These double-sorted portfolio results reinforce the findings from the cross-sectional regressions, which show that the book-to-market premium grows as industry concentration increases.

These double-sort portfolios yield insights into the relations among market structure, value, and growth. Prior research examines the book-to-market ratio as a risk proxy related to relative profitability or distress and yields mixed results.[10] Our results suggest an explanation for these mixed results: Firms with the same level of book-to-market are fundamentally different from one another depending on the market structure of the industry in which they operate. A low book-to-market firm in a concentrated industry is not well described as a "growth firm." This firm operates in an industry with a large asset base, high

[9] Using within-industry breakpoints addresses the fact that the cross-industry variation in the book-to-market ratio is not important for average stock returns; instead, most of the effect is the result of within-industry variation (Asness and Stevens (1996), Cohen, Polk, and Vuolteenaho (2003)).

[10] For evidence on the link between book-to-market and distress, see Fama and French (1995), Chen and Zhang (1998), Liew and Vassalou (2000), Shumway (1996), or Griffin and Lemmon (2002). For recent papers linking book-to-market to investment and costly reversibility, see Zhang (2005) and Cooper (2006).

Table VIII

Interaction between Industry Concentration and Book-to-Market Effects

Panel A presents results from monthly Fama–MacBeth cross-sectional regressions of individual stock returns on ln(Size), ln(B/M), Momentum, Beta, Leverage, H(Sales), and an interaction term between H(Sales) and ln(B/M). Time-series averages of the monthly regression coefficients, in percent, are reported with time-series t-statistics below. Panels B and C report value-weighted (Panel B) and equal-weighted (Panel C) average returns of B/M quintile portfolios, their t-statistics, and the difference in returns between Quintile 5 and Quintile 1 within each concentration quintile. Two-digit SIC Industries are first sorted into concentration quintiles based on their H(Sales) value. Firms within each industry are then sorted into quintiles based on their B/M. Finally, firms with the same B/M ranking from industries within the same concentration quintile are grouped into one portfolio to create the 5 × 5 double-sorted portfolios on H(Sales) and B/M. The row entitled "All" reports average returns of B/M quintiles formed across concentration quintiles.

Panel A: Fama–MacBeth Cross-Sectional Regressions

ln(Size)	ln(B/M)	Momentum	Beta	Leverage	H(Sales)	H(Sales) × ln(B/M)
−0.19	0.29	0.59	−0.41	−0.32	−0.37	0.25
−3.82	3.76	3.78	−1.99	−1.65	−3.14	2.38

H(Sales) Quintiles	B/M Quintiles 1 (Low)	2	3	4	5 (High)	5 − 1 Spread
Panel B: Value-Weighted Average Returns of H(Sales) and B/M Sorted Portfolios						
1 (Low)	1.11	1.26	1.24	1.42	1.58	0.47
	4.01	5.13	4.45	5.22	4.49	2.11
2	0.84	1.08	1.26	1.51	1.45	0.61
	3.23	4.33	4.93	5.95	5.23	2.80
3	0.98	1.04	1.13	1.32	1.64	0.65
	3.52	3.78	4.31	4.77	5.87	2.59
4	0.74	0.99	1.21	1.21	1.45	0.71
	2.85	3.93	5.27	4.74	5.07	3.35
5 (High)	0.62	1.03	1.05	1.33	1.38	0.76
	1.93	3.02	3.22	5.28	3.75	2.11
All	0.85	1.03	1.13	1.32	1.50	0.66
	3.31	4.65	5.11	5.73	5.78	3.19
Panel C: Equally Weighted Average Returns of H(Sales) and B/M Sorted Portfolios						
1 (Low)	0.81	1.34	1.31	1.61	1.83	1.02
	2.58	4.18	4.61	5.62	5.98	6.84
2	0.78	1.07	1.40	1.54	1.81	1.03
	2.47	3.80	5.07	5.64	6.28	6.99
3	0.70	0.99	1.34	1.56	1.82	1.13
	2.09	3.24	4.53	5.43	5.92	6.49
4	0.68	1.02	1.30	1.33	1.85	1.17
	1.89	3.40	4.41	4.50	5.91	4.70
5 (High)	0.15	0.97	1.10	1.17	1.59	1.43
	0.32	2.41	3.18	3.28	4.11	3.36
All	0.76	1.15	1.36	1.56	1.81	1.05
	2.47	3.97	4.92	5.68	6.22	8.12

unit profitability, and low R&D, and subsequently has high capitalized future profitability. Its book-to-market is low not because its growth prospects are high, but because its current and expected future profitability is high. High profitability, low-risk firms are thus being labeled growth firms, pulling down the average returns of low book-to-market stocks.

On the other hand, a low book-to-market stock in a competitive industry is indeed better characterized as a growth firm. These firms engage in more R&D on average and are less profitable, and thus the low market-to-book is not a reflection of high capitalized profitability, but rather of expected growth. Growth is risky, and this shows up in higher expected returns.

If we interpret book-to-market as a proxy for distress, then these findings help us to distinguish between the creative destruction and barriers-to-entry hypotheses offered in Section I. These findings favor the innovation risk interpretation, since the spreads in industry concentration are largest in the portion of the book-to-market spectrum where growth is most salient. Among high book-to-market firms, for which it is often argued that distress is more salient, we see a smaller spread in returns across concentration quintiles.

VII. Conclusion

What are the economic determinants of the cross section of stock returns? This is one of the fundamental questions in empirical asset pricing, and is especially important given the large body of recent research documenting return predictability based on a host of empirically motivated financial characteristics. We address this question from a new perspective, offering evidence that industry concentration—a feature of the product markets in which firms operate—is important for understanding stock returns.

Our main thesis is simple. We argue that the structure of product markets helps to determine a firm's risk by affecting the equilibrium operating decisions it makes. In particular, drawing on classic work in industrial organization from Schumpeter (1912) and Bain (1954), we link industry concentration to stock returns through innovation and distress risk. Industries in which innovation risk and distress risk are higher should command higher expected returns. Our analysis indicates that these are competitive industries.

We show that firms in competitive industries earn higher stock returns, even after controlling for the usual suspects that affect the cross section of average returns, such as size, book-to-market, and momentum. This holds both at the industry level and the firm level and is robust to alternative empirical specifications. Moreover, this result is not explained by differences in unexpected returns, and it has been a persistent feature of stock returns since the Great Depression.

These results suggest a number of fruitful areas for future research. First and foremost, our empirical evidence suggests a need for asset pricing models that explicitly incorporate features of product markets as determinants of asset returns. A more rigorous theory of why asset prices are affected by market structure will allow for a more careful exploration of the link that we demonstrate to be important in this paper.

Second, we argue in this paper that either innovation or distress risk is a likely culprit for the concentration premium. Our preliminary findings support an innovation risk interpretation, but clearly more work is needed to disentangle these potentially overlapping hypotheses. We find that the concentration spread in returns is larger for low book-to-market stocks than high book-to-market stocks. One interpretation is that low book-to-market stocks in concentrated industries have low returns because they have high capitalized future profitability and engage in less innovation, while low book-to-market stocks in less concentrated industries have higher returns because they engage in more innovative activity and thus have higher expected growth rates. This suggests that the links among value, growth, and product market structure are important questions for future work.

Finally, this paper primarily focuses on the unconditional roles played by industry characteristics for understanding the equilibrium trade-off between risk and return. However, much of the recent literature in empirical asset pricing uses industry membership as conditioning information, and explores whether certain asset pricing phenomena are attributable to industry effects. A better understanding of how industry characteristics affect expected returns can potentially yield insights into why many stylized facts about stock returns seem to contain important industry components.

We focus on risk-based explanations for the concentration premium. Of course, the alternative is that some behavioral bias causes investors to undervalue firms in less concentrated industries, producing high returns ex post. However, any behavioral explanation must stand up to a series of facts. The concentration premium exhibits sensible business cycle variation, growing in magnitude as expected future growth opportunities deteriorate. Moreover, there is substantial common variation in stock returns that is related to industry concentration.

The findings in this paper ultimately raise more questions than they answer. Are there other mechanisms through which market structure affects stock returns? Does the link between market structure and stock returns impact firms' investment and financing decisions? How does it impact the diffusion of information in the market? Is the geographic scope of the industry (national vs. local product markets) important? The story we propose in this paper is a reduced-form version of a more complicated analysis in which product markets affect investment opportunities and decisions about investment and capital structure. Ultimately, a better understanding of the precise mechanisms that link these phenomena is required. We leave these issues for future work.

REFERENCES

Almazan, Andres, and Carlos Molino, 2001, Intra-industry capital structure dispersion, Working paper, University of Texas.

Asness, Clifford, and Ross Stevens, 1996, Intra-industry and inter-industry factors in the cross-section of expected stock returns, Working paper, Goldman Sachs Asset Management.

Bain, Joe, 1954, Economies of scale, concentration and the condition of entry in twenty manufacturing industries, *American Economic Review* 44, 15–39.

Barclay, Michael J., and Clifford W. Smith, 1995, The maturity structure of corporate debt, *Journal of Finance* 50, 609–631.

Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y., and Tuomo Vuolteenaho, 2004, Good beta, bad beta, *American Economic Review* 94, 1249–1279.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chen, Nai-Fu, and Feng Zhang, 1998, Risk and returns of value stocks, *Journal of Business* 71, 501–535.

Cohen, Randolph, Christopher Polk, and Tuomo Vuolteenaho, 2003, The value spread, *Journal of Finance* 58, 609–641.

Cooper, Ilan, 2006, Asset pricing implications of nonconvex adjustment costs and irreversibility, *Journal of Finance* 61, 139–170.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristics-based benchmarks, *Journal of Finance* 52, 1035–1058.

Davis, Jim, Eugene F. Fama, and Kenneth French, 2000, Characteristics, covariances, and average returns: 1929–1997, *Journal of Finance* 55, 389–406.

Fama, Eugene F., 1990, Stock returns, expected returns, and real activity, *Journal of Finance* 45, 1089–1108.

Fama, Eugene F., 1998, Determining the number of priced state vectors in the ICAPM, *Journal of Financial and Quantitative Analysis* 33, 217–231.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F., and Kenneth R. French, 1992. The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns of bonds and stocks, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance* 50, 131–155.

Fama, Eugene F., and Kenneth R. French, 2000, Forecasting profitability and earnings, *Journal of Business* 73, 161–175.

Fama, Eugene F., and James MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Geroski, P. A., 1990, Innovation, technological opportunity, and market structure, *Oxford Economic Papers* 42, 586–602.

Gibbons, Michael R., Steven A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Griffin, John, and Michael Lemmon, 2002, Book-to-market equity, distress risk, and stock returns, *Journal of Finance* 57, 2317–2336.

He, Kathy, Randall Mørck, and Bernard Yeung, 2003, Corporate stability and economic growth, Working paper, University of Alberta.

Hou, Kewei, 2003, Industry information diffusion and the lead-lag effect in stock returns, Working paper, Ohio State University.

Kahle, Kathleen M., and Ralph A. Walkling, 1996, The impact of industry classifications on financial research, *Journal of Financial and Quantitative Analysis* 31, 309–335.

Knott, Anne Marie, and Hart Posen, 2003, Does competition increase innovation? New evidence from old industries, Working paper, Mack Center for Technological Innovation.

Kothari, S.P., and Jay Shanken, 1992, Stock return variation and expected dividends: A time-series and cross-sectional analysis, *Journal of Financial Economics* 31, 177–210.

Lewellen, Jonathan, 1999, The time-series relations among expected return, risk, and book-to-market, *Journal of Financial Economics* 54, 5–43.

Liew, Jimmy, and Maria Vassalou, 2000, Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics* 57, 221–245.

MacKay, Peter, and Gordon Phillips, 2005, How does industry affect firm financial structure? *Review of Financial Studies* 18, 1433–1466.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 127–147.

Moskowitz, Tobias J., and Mark Grinblatt, 1999, Do industries explain momentum? *Journal of Finance* 54, 1249–1290.

Reinganum, Jennifer, 1985, Innovation and industry evolution, *Quarterly Journal of Economics* 100, 81–99.

Ross, Steven A., 1976, The arbitrage theory of asset pricing, *Journal of Economic Theory* 13, 341–360.

Salop, Steven C., 1979, Strategic entry deterrence, *American Economic Review* 69, 335–338.

Schmalensee, Richard, 1978, Entry deterrence in the ready-to-eat breakfast cereal industry, *The Bell Journal of Economics* 9, 305–327.

Schmalensee, Richard, 1981, Economies of scale and barriers to entry, *Journal of Political Economy* 89, 1228–1238.

Schumpeter, Joseph, 1912, *The Theory of Economic Development* (Harvard University Press, Cambridge, Mass).

Schumpeter, Joseph, 1942, *Capitalism, Socialism, and Democracy* (Harper & Row, New York and London).

Shumway, Tyler, 1996, Size, overreaction, and book-to-market effect as default premia, working paper, University of Michigan.

Shumway, Tyler, 1997, The delisting bias in CRSP data, *Journal of Finance* 52, 327–340.

Sutton, John, 1991, *Sunk Cost and Market Structure: Price Competition, Advertising, and the Evolution of Concentration* (MIT Press, Cambridge).

Sutton, John, 1998, *Technology and Market Structure: Theory and History* (MIT Press, Cambridge).

Titman, Sheridan, 1984, The effects of capital structure and a firm's liquidation decision, *Journal of Financial Economics* 13, 137–151.

Vuolteenaho, Tuomo, 2002, What drives firm level stock returns? *Journal of Finance* 57, 233–264.

Zhang, Lu, 2005, The value premium, *Journal of Finance* 60, 67–103.

Product Market Competition and Industry Returns

M. Cecilia Bustamante
University of Maryland

Andres Donangelo
University of Texas at Austin

This paper studies how expected returns interact with product market competition. The model predicts that (1) competition erodes markups, such that firms are more exposed to systematic risk; (2) the threat of entry by new firms lowers exposure to systematic risk of incumbents; and (3) higher industry aggregate risk represents a barrier to entry, such that riskier industries become less competitive. We provide empirical evidence consistent with these three channels and for an overall *negative* relation between returns and competition. We also consider a sample selection correction for publicly listed firms and use it to construct an alternative concentration measure. (*JEL* L11, L22, G11, G12, G31)

Received June 3, 2014; editorial decision December 16, 2016 by Editor Andrew Karolyi.

This paper studies how the competitive environment in which firms operate and their exposure to systematic risk are interrelated. Earlier empirical studies have conjectured that firms in less competitive industries are insulated from aggregate shocks, which in turn leads to lower expected returns on average.[1] In this paper, we show that this mainstream conception is incomplete, since it misses two additional aspects of the connection between product market competition and asset prices. The first relates to the dynamic nature

We thank Aydogan Alti, Geert Bekaert, Alon Brav, Andrea Buffa, Anna Cieslak, Ilan Cooper, Laurent Frésard, John Griffin, Christian Julliard, Xiaoji Lin, Andrew Karolyi, Richard Lowery, Lorenzo Garlappi, Thomas Gilbert, Gerard Hoberg, Robert Pindyck, David Robinson, and Laura Starks. We would also like to thank the seminar participants at Boston University, Cornell University, Drexel University, Duke University, London School of Economics, the Norwegian Business School (Oslo), Tilburg University, University of Cambridge, University of Maryland, University of Minnesota (seminar series and Asset Pricing Conference), University of Rochester, University of Texas at Austin, Vienna Graduate School of Finance, as well as the 2013 Western Finance Association Meeting and the 2013 European Finance Association Meeting. All errors are our own. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to Andres Donangelo, Department of Finance, University of Texas at Austin, 1 University Station, B6600, Austin, TX 78712; telephone: (510) 232-6841. E-mail: andres.donangelo@mccombs.utexas.edu.

[1] For instance, Hou and Robinson (2006) state that "[i]f barriers to entry in product markets insulate some firms from aggregate demand shocks, while exposing others, then we would expect distress risk to vary with market structure. This predicts that industries with high barriers to entry are associated with lower equilibrium stock returns."

doi:10.1093/rfs/hhx033 Advance Access publication April 29, 2017

of competition: The negative effect of the entry threat on firm value is *procyclical*, which effectively lowers incumbents' exposure to systematic risk. The second aspect is that higher industry discount rates act as a barrier to entry.[2]

We formalize these arguments using an extension of the dynamic model of industry equilibrium by Pindyck (2009). Our model shows that the degree of competition is jointly and dynamically subsumed by the levels of concentration and average markup in the industry. Moreover, the model characterizes three alternative channels, each of which is closely related to the three aspects of competition that are relevant for asset pricing. These channels highlight that the link between competition and firms' exposure to systematic risk goes in two directions: the degree of competition affects firms' expected asset returns and, simultaneously, the industry exposure to systematic risk affects the likelihood of entry by new firms into the industry.

We first use the model to study how exogenously given levels of concentration and markup affect exposure to systematic risk in the industry. We find that competition affects expected asset returns through two distinct channels. Through the first channel, the *operating leverage channel*, competition affects firms' exposure to systematic risk through its effect on profit margins. Intuitively, since firms in more competitive industries have lower profit margins to buffer adverse shocks, these firms have higher operating leverage and hence higher exposure to risk.[3] The second channel, which we term the *entry threat channel*, reduces the expected returns of incumbents due to the value destruction associated with the expected entry of new firms. Given that entry is more likely during expansions rather than contractions, the value destruction due to the entry threat renders the value of incumbents less procyclical and lowers their exposure to systematic risk. In isolation, this channel leads to lower expected returns in more competitive industries with lower barriers to entry.[4]

We then use the model to study the reverse direction through which exposure to systematic risk and competition are related. We analyze the effect of industry exposure to systematic risk on the competitive environment, which has been studied in earlier works by Subrahmanyam and Thomadakis (1980) and Pindyck (2009). In our setting, this effect is subsumed by a third channel, which we define as the *risk feedback channel*. The model shows that industries with higher exposure to systematic risk are less attractive to new entrants,

---

[2] From this point forward, we refer to product market competition as competition, and to the level of industry concentration and the average industry markup as concentration and markup, respectively.

[3] Some examples from the literature that document how operating leverage translates into higher expected asset returns are Lev (1974), Mandelker and Rhee (1984), Danthine and Donaldson (2002), Carlson, Fisher, and Giammarino (2004), Zhang (2005), Garcia-Feijoo and Jorgensen (2010), Novy-Marx (2011), Belo, Lin, and Bazdresch (2014), Donangelo (2014), and Favilukis and Lin (2016b).

[4] See Bain (1956) and Tirole (1988) for seminal discussions on the link between entry, concentration, and firms' profit margin in an industry equilibrium.

given that the cash flows generated by new ventures in such industries are discounted at higher rates. Consequently, riskier industries are expected to remain relatively more concentrated and to have higher average markups over time. Lower competition partially compensates investors for bearing greater systematic risk.

In sum, our model shows that the relation between expected returns and competition is complex. While the operating leverage channel predicts a negative relation between expected returns and either concentration or markup, both the entry threat channel and the risk feedback channel predict the opposite. For this reason, we discriminate between the *conditional* and *unconditional* testable predictions of the model. The *conditional* testable prediction shows that the relation between expected returns and either concentration or markup is relatively more positive in industries with relatively lower operating leverage. This prediction does not depend on the absolute level of operating leverage in each industry. Moreover, the model suggests that, in a probabilistic sense, we should expect an *unconditional* positive relation between expected returns and either concentration or markup. Intuitively, due to the barriers to entry in each industry, the risk feedback channel reduces the likelihood of operating leverage becoming too high.

In order to test the predictions of our model, we employ different measures of concentration and markup. We consider two alternative measures of concentration: the Herfindahl–Hirshman index from the U.S. Census Bureau, and a new measure based on adjusted Compustat sales data, which we term characteristics-based concentration (CBC). The first measure, which we refer simply as HHI, is available only for manufacturing industries and every five years. A common workaround for these limitations is to construct concentration measures based on the subsample of publicly listed firms (e.g., firms in Compustat). Ali, Klasa, and Yeung (2009) document that concentration measures based on Compustat data have a low correlation with the HHI. We build on their finding and show that Compustat-based concentration measures are biased partly because the decision to be public is not random, and is significantly related to the degree of competition in the industry. We thus construct the CBC measure, which is an alternative concentration measure that covers both manufacturing and nonmanufacturing industries, and relies on Compustat sales data adjusted for the likelihood of observing a public firm in each industry.

We also employ two alternative measures of markup. The first measure of markup, MKP, is limited to manufacturing industries, and it is constructed from the Manufacturing Industry Database from the National Bureau of Economic Research and the U.S. Census Bureau's Center for Economic Studies (NBER/CES). To expand the sample to nonmanufacturing industries, we construct a second measure of markup, MKC, from Compustat data. Unlike the case of the measures of concentration using public firms' data, we find no evidence of a significant bias in the MKC measure constructed with Compustat data. We report that MKP and MKC have a positive and highly significant correlation in our working sample.

In addition to employing alternative measures of concentration and markup that are not subject to the sample selection bias of public listing, our empirical strategy also addresses the concern that we only observe accounting and market data in the subsample of publicly listed firms in each industry. This limitation would not represent a problem if systematic differences between the relative number of public and private firms across industries were unrelated to the degree of competition. Yet, there is evidence that a firm's decision to be publicly listed is significantly influenced by the competitive environment. In our working sample, the share of publicly listed firms in each industry is strictly greater in industries with either higher markup or higher concentration.[5]

In our empirical analyses, we find supporting evidence for the *unconditional* prediction that expected returns relate positively to concentration and markup. Using eight alternative proxies of expected returns, we report that the expected returns of the portfolio of firms in the top quintile of either concentration or markup is generally higher than the expected returns in the corresponding portfolio for the bottom quintile. For instance, the average excess realized stock return spread between these two quintile portfolios is between 5% and 7%. Using Fama–MacBeth regressions with controls, we also report that one cross-sectional standard deviation in either concentration or markup is associated with up to 4% higher realized stock returns. These findings are in contrast with those in Hou and Robinson (2006), who use Compustat-based measures of concentration to show that average realized returns are *lower* in more concentrated industries.

The empirical evidence is also consistent with the economic channels predicted by our model. Using our alternative proxies for expected returns, we provide evidence in line with the *conditional* prediction that the positive expected return spread between less and more competitive industries is higher in industries with lower operating leverage. Consistent with our model, we also find that operating leverage is indeed higher in more competitive industries with lower concentration and lower markup. Last, consistent with the value destruction mechanism that drives the entry threat channel, we find that firms in less competitive industries better preserve their value and thus have lower earnings-to-price and book-to-market ratios. Through the lens of the model, these findings jointly suggest that firms in more competitive industries are effectively safer, despite their greater level of operating leverage.

Beyond the testable predictions of the theory, our empirical strategy allows us to make inferences on the systematic differences between private and public firms in our working sample. For instance, we find that firms in less competitive

[5] See the discussion in Section 2 for more details. In a related study, Chemmanur, He, and Nandy (2010) show that the probability that a private firm undertakes an initial public offering is significantly affected by the product market.

industries are more likely to be publicly listed. We also find indirect evidence that private firms have, on average, higher expected returns than public firms. Put together, these two findings explain the fact that the positive relation between expected returns and either concentration or markup is generally biased upwards when we focus on the subsample of public firms in each industry.

The theoretical motivation of our paper is closely related to the models by Subrahmanyam and Thomadakis (1980) and Pindyck (2009), who argue that uncertainty acts as a barrier to entry in imperfectly competitive industries. Our model is also closely related to the model by Aguerrevere (2009), who also explores the opposing effects on expected returns of operating leverage and industry growth. The main departing point of our model is that it endogenizes firms' decision to enter the industry and the number of incumbents in equilibrium. This feature of our model allows us to take into account that competition both affects and is also affected by the industry systematic risk.

Our paper is related more generally to other theoretical studies in the literature. The entry threat channel in our model is related to the implications of the models of intra-industry strategic interaction proposed by Carlson et al. (2014) and Bustamante (2015). The models by Aguerrevere (2003), Garlappi (2004), Opp et al. (2014), Loualiche (2017), and V Binsbergen (2016) provide alternative hypotheses on the link between product markets and asset prices which complement our paper.

Our empirical analysis is closely related to the seminal empirical study by Hou and Robinson (2006) on the relation between stock returns and industry concentration. The empirical strategy in our paper builds on the empirical findings in Ali, Klasa, and Yeung (2009), who highlight the importance of considering both private and public firms in constructing measures of concentration. Our findings relate to the earlier study by Bulan, Mayer, and Somerville (2009), who show that firms in less competitive environments better preserve the value of their projects. Our evidence is also related to the findings by Hoberg and Phillips (2014), showing that firms have higher valuations when their products are more unique. Other related empirical studies include Cooper and Priestley (2016) on the riskiness of private and public firms, and the study by Hoberg and Phillips (2010) showing more pronounced industry booms and busts (i.e., higher cyclicality) in more competitive industries.

1. Model

In this section, we derive a dynamic model of industry equilibrium in which incumbents compete with one another and face the threat of new entrants. The model studies how concentration and markup interact in equilibrium with firms' exposure to systematic risk.

1.1 Model setup

We consider an exogenous pricing kernel, following Berk, Green, and Naik (1999). The dynamics of the pricing kernel Λ are given by

$$\frac{d\Lambda_t}{\Lambda_t} = -r\,dt - \eta\,dZ_t, \tag{1}$$

where $r > 0$ is the instantaneous risk-free rate, $\eta > 0$ is the market price of risk, and dZ is a Wiener process that represents the only source of priced risk in the economy.

The industry is populated by $N \geq 1$ incumbent firms, indexed by j, that produce and sell y_{jt} units of a single homogeneous industry good. New firms invest $\kappa > 0$ upon entry in the industry. Having entered the industry, incumbents face no costs to adjust their production.[6]

We consider a downward-sloping demand curve such that the price of the homogeneous industry good $\mathcal{P}$ is given by

$$\mathcal{P}_t = X_t - \alpha Y_t, \tag{2}$$

where $\alpha > 0$, y_j is the output of firm j, $Y \equiv \sum_j y_j$ is total industry output; and X is a procyclical industry shock that follows the diffusion

$$\frac{dX_t}{X_t} \equiv \mu\,dt + \sigma\left(\rho\,dZ_t + \sqrt{(1-\rho^2)}\,dZ_t^{\mathrm{X}}\right), \tag{3}$$

where $0 < \rho < 1$, and dZ^{X} is a Wiener process that captures the portion of the industry shock that is not priced (i.e., $\mathrm{E}[dZ^{\mathrm{X}}dZ]=0$). The volatility parameter $\sigma > 0$ represents the level of *total* instantaneous industry risk, both systematic and nonsystematic. Meanwhile, ρ represents the fraction of total instantaneous industry risk σ that is systematic.

Operating profits, defined as revenues minus the costs of production, are given by

$$\Pi_{jt} \equiv (\mathcal{P}_t - cX_t)y_{jt} - f, \tag{4}$$

where $f \geq 0$ is a fixed production cost, and cX is the marginal cost of production such that $0 < c < 1$. Note that the industry shock X affects both the industry's demand, as shown by Equation (2), as well as the marginal cost of the firm, as shown by Equation (4). In this sense, X captures both demand-side and supply-side shocks to the industry.

1.2 Concentration and markup in equilibrium

We follow Pindyck (2009) and assume Cournot competition at each point in time and assume that all firms within an industry are identical.[7] The quantity

[6] This specification is analogous to an alternative setting in which firms have production functions with the constant returns-to-labor property and with access to a perfect spot labor market.

[7] Given this assumption, we henceforth drop all subscripts j from our notation unless required.

produced by each firm, the number of incumbents N, and the price of the industry good $\mathcal{P}$ are determined endogenously in a Nash equilibrium. The concentration h that obtains in equilibrium is defined by the sales-based Herfindahl–Hirschman index, as given by

$$h_t \equiv \sum_j \left(\frac{y_{jt}}{Y_t}\right)^2 = \frac{1}{N}. \tag{5}$$

The profit margin of each firm in the industry is defined as price minus unit cost of production, $\mathcal{P}_t - cX_t$. In equilibrium, operating profits are given by

$$\Pi[p_t] = \frac{p_t^2}{\alpha} - f, \tag{6}$$

where $p > 0$ is the equilibrium profit margin (henceforth *profit margin*), which is an increasing function of both the industry shock X and the concentration h as given by

$$p_t = X_t(1-c)\left(\frac{h_t}{h_t+1}\right). \tag{7}$$

Note that at each point in time at which no entry occurs, the law of motion of the profit margin p is equal to that of the industry shock X, namely,

$$\frac{dp_t}{p_t} = \frac{dX_t}{X_t} = \mu dt + \sigma\left(\rho dZ_t + \sqrt{(1-\rho^2)} dZ_t^{\mathrm{X}}\right). \tag{8}$$

Finally, we define markup m as the ratio of operating profits to the value of sales so that, in equilibrium

$$m[h_t, p_t] = \underbrace{\left(\frac{1-c}{1+\frac{c}{h_t}}\right)}_{\text{Lerner Index}} \underbrace{\left(1 + \frac{f}{\Pi[p_t]}\right)^{-1}}_{\substack{\text{Fixed Cost} \\ \text{Adjustment Term}}}. \tag{9}$$

Equation (9) shows that markup $m \equiv \frac{\Pi_t}{\mathcal{P}_t y_t}$ is an increasing function of both the concentration h and the profit margin p. The markup m is expressed as the product of two economically meaningful factors. The first factor equals the ratio of the profit margin to the product price, commonly referred to as the Lerner index.[8] This first factor captures the average market power of incumbents, and it only varies when a new firm enters the industry. The second factor is continuously changing, since it is directly affected by profit margin p, which in turn depends on the industry shock X. The second factor reflects the fact that the markup m differs from the Lerner index as long as there are fixed operating costs, which introduce a wedge between marginal costs and average costs of production.

8 See Lerner (1934) for a discussion of this index. In our model, the Lerner index is such that $\frac{p_t}{\mathcal{P}_t} \equiv \frac{1-c}{1+\frac{c}{h_t}}$.

1.3 Operating leverage

We follow Gahlon and Gentry (1982) and Donangelo (2014) and define operating leverage as the local sensitivity of operating profit growth to industry shocks in the region where operating profits are strictly positive. Specifically, we define the level of operating leverage in an industry as the scaled slope of a regression of operating profit growth on the exogenous shocks that affect revenues.[9] In equilibrium, operating leverage equals

$$\Theta[p_t] = \frac{f}{\Pi[p_t]}, \quad \text{for} \quad \Pi[p_t] > 0. \tag{10}$$

Equations (6) and (10) show that, intuitively, operating leverage Θ is increasing in fixed costs f and decreasing in profit margin p.

A reduction in concentration, which is triggered only by the entry of new firms into the industry, raises operating leverage Θ by lowering operating profits. For a similar reason, operating leverage is strictly decreasing in markup m when firms enter the industry. The proposition below characterizes the links between operating leverage Θ, concentration h, and markup m at the profit margin threshold $p=\overline{p}[h]$ such that firms enter the industry.[10] We elaborate on the entry threshold $p=\overline{p}[h]$ in the next subsection.

Proposition 1 (Operating leverage, concentration, and markup). *Given* $\Pi[\overline{p}[h]] > 0$*:*

1. *Operating leverage is strictly decreasing in concentration at entry:* $\frac{\partial \Theta[\overline{p}[h]]}{\partial h} < 0$*, and*
2. *Operating leverage is strictly decreasing in markup at entry:* $\frac{\partial \Theta[\overline{p}[h]]}{\partial m} < 0$.

Proof. See Appendix B. ■

Proposition 1 states that firms in less concentrated industries have higher operating leverage, since they are less capable of buffering adverse shocks. Similarly, firms in industries with lower markup m have higher operating leverage. The first row in Figure 1 illustrates the negative relation between operating leverage and concentration and markup.

1.4 Firm value and entry

The value of an incumbent firm, V, is defined as the maximized expected discounted stream of operating profits. This value can be expressed as a function

9 In our case the level of operating leverage is given by $\Theta \equiv \frac{1}{2}\,\mathrm{Cov}\left[\frac{d\Pi}{\Pi}, \frac{dX}{X}\right] / \mathrm{Var}\left[\frac{dX}{X}\right] - 1$, for $\Pi > 0$.

10 Proposition 1 focuses on the relation between operating leverage and markup at the entry threshold for the sake of consistency, since the relation between operating leverage and concentration is fully determined at the entry threshold. It is straightforward to show that operating leverage is also strictly decreasing in markup for any $p < \overline{p}[h]$.

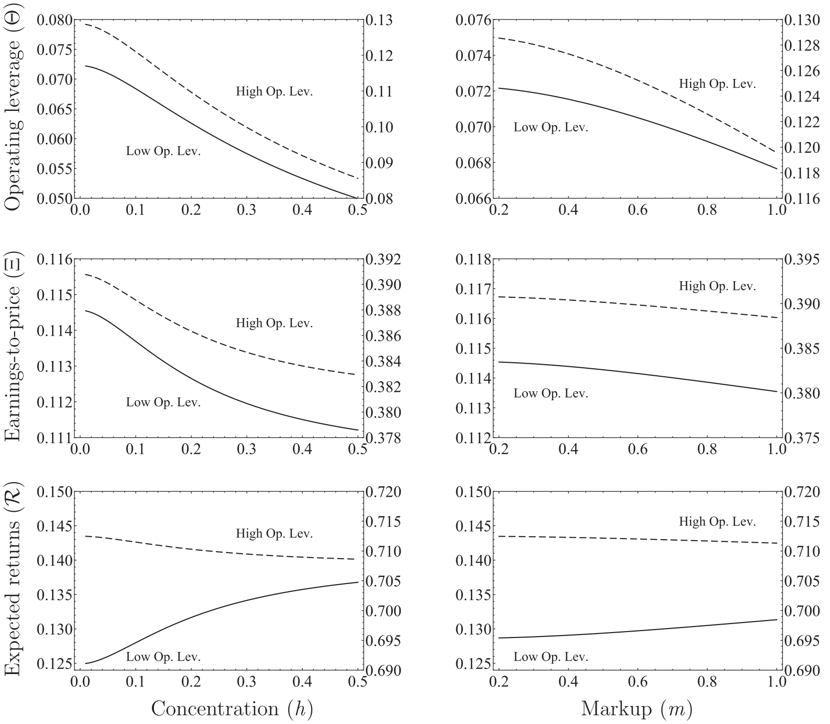


Figure 1
Model solution: Operating leverage, earnings-to-price ratio, and expected asset returns for various levels of concentration and markup
The baseline case ($f=0.01$) is represented by solid lines and left axes. The high–operating leverage case ($f=2.00$) is represented by dashed lines and right axes. Parameter values used in plots are: $\alpha=0.01$, $r=0.04$, $\eta=1.00$, $\mu=0.03$, $\sigma=0.1$, $\rho=0.5$, and $\kappa=100$.

of the current level of concentration h and profit margin p, as given by

$$V[h_t,p_t]=\underbrace{\frac{\Pi[p_t]}{\delta}\left(1-\Theta[p_t]\left(\frac{\delta-r}{r}\right)\right)}_{\substack{\text{value of assets in place}\\ \text{with no entry threat}}}-\underbrace{\frac{\Pi[p_t]}{\delta}(1+\Theta[p_t])\left(\frac{2}{\upsilon}\right)\left(\frac{p_t}{\overline{p}[h_t]}\right)^{\upsilon-2}}_{\substack{\text{value destruction}\\ \text{from entry threat}}}, \tag{11}$$

where $\delta>r$ and $\upsilon>2$ are constants defined in Appendix A, in which we also provide a formal definition of the entry threshold $\overline{p}[h]$.

Equation (11) highlights two key components of firm value. The first component represents the discounted value of the future operating profits of a firm in an industry that is identical in all other respects but has no threat of entry by new firms. The second component, which is strictly negative, represents the discounted value of the expected reduction in operating profits due to

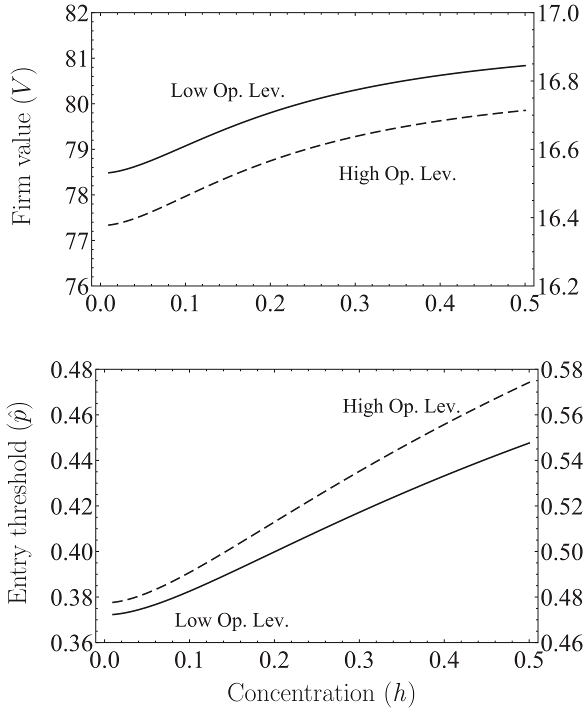


Figure 2
Model solution: Firm value and entry threshold for different levels of concentration
The baseline case ($f=0.01$) is represented by solid lines and left axes. The high–operating leverage case ($f=2.00$) is represented by dashed lines and right axes. Parameter values used in plots are: $\alpha=0.01$, $r=0.04$, $\eta=1.00$, $\mu=0.03$, $\sigma=0.1$, $\rho=0.5$, and $\kappa=100$.

the future entry of competing firms. Hence this second component represents the expected value destruction due to expected future entry by competing firms.

Equation (11) further shows that concentration affects firm value through the entry threshold $\overline{p}[h]$. The concentration h only changes at the time when a new firm enters the industry, which happens when operating margin p reaches the upper boundary $\overline{p}[h]$. When p reaches the threshold $\overline{p}[h]$, concentration decreases to $\frac{h}{h+1}$, which drives down the profit margin p and resets the threshold to the new level $\overline{p}[\frac{h}{h+1}]$. When there is no entry, all variation in the profit margin p is exclusively driven by fluctuations in the industry shock X.

Figure 2 summarizes the main predictions of the model with respect to firm value, the entry threshold, and concentration. The first panel in Figure 2 shows that firm value is strictly increasing in concentration. This result is consistent with previous studies in the literature (e.g., Grenadier (2002), as it implies that firms in more concentrated industries better preserve their value. The second panel of Figure 2 further illustrates that the entry threshold $\overline{p}[h]$ is strictly increasing in h.[11] This result implies that more concentrated industries have higher barriers to entry.

1.5 Earnings-to-price ratio

The earnings-to-price ratio Ξ is useful for disentangling the relative contributions of the two components of firm value from Equation (11). The reason is that while firm value is negatively affected by the *expected* entry into the industry, operating profits are not. In equilibrium, the earnings-to-price ratio is given by

$$\Xi[h_t,p_t] \equiv \frac{\Pi[p_t]}{V[h_t,p_t]} = \delta\left(1-\Theta[p_t]\left(\frac{\delta-r}{r}\right)-(1+\Theta[p_t])\left(\frac{2}{\upsilon}\left(\frac{p_t}{\overline{p}[h_t]}\right)^{\upsilon-2}\right)\right)^{-1}. \tag{12}$$

Equation (12) shows that, for a given level of profit margin p, the earnings-to-price ratio Ξ is increasing with respect to the likelihood of entry into the industry, given that the likelihood of entry is associated to the ratio $\frac{p}{\overline{p}[h]}$. Hence, Ξ is decreasing in the concentration h.

Proposition 2 (earnings-to-price ratio, concentration, and markup). *For any profit margin p such that* $V[h,p]>0$*:*

i. *The earnings-to-price ratio is strictly decreasing in concentration:* $\frac{\partial \Xi[h,p]}{\partial h}<0$, *and*
ii. *The earnings-to-price ratio is strictly decreasing in markup:* $\frac{\partial \Xi[h,p]}{\partial m}<0$.

Proof. See Appendix B. ■

Proposition 2 characterizes the comparative statics of the earnings-to-price ratio Ξ with respect to concentration h. A lower level of concentration reflects higher value destruction from the threat of entry, and it relates to a higher earnings-to-price ratio. Proposition 2 also characterizes the relation between earnings-to-price ratios and the markup m. As shown in Equation (9), the markup m is a noisy proxy for the empirically unobservable Lerner index, which is increasing in the concentration h. Consistent with economic intuition, the earnings-to-price ratio is decreasing in the markup m. The second row of Figure 1 illustrates the relation between the earnings-to-price ratio Ξ and concentration h and markup m.

The following lemma presents a lower bound for the earnings-to-price ratio Ξ. Lemma 1 follows directly from Equation (12).

---

[11] In Appendix A, we also show that the entry threshold $\overline{p}[h]$ converges to a minimum level as the industry becomes perfectly competitive (i.e., $h \to 0$). The existence of a lower bound for the entry threshold implies that, when concentration h is sufficiently low, the decision of prospective entrants has no material impact on the profit margin, p.

Lemma 1 (lower bound for the earnings-to-price ratio). *For any profit margin p and concentration h such that* $V[h,p]>0$*, the earnings-to-price ratio* $\Xi[h,p]$ *is such that* $\Xi[h,p]\geq\delta$*.*

1.6 Expected returns and systematic risk loadings

The instantaneous expected excess asset return $\mathcal{R}$ is defined as the drift of the gains process

$$\mathcal{R}_t \equiv \mathrm{E_t}\left[\frac{dV_t}{V_t}\right]\frac{1}{dt}+\frac{\Pi_t}{V_t}-r=\beta_t\eta, \tag{13}$$

where η is the market price of risk and β is the implied systematic risk loading. Equation (13) shows that the asset pricing implications for expected returns $\mathcal{R}$ are the same as those for the systematic risk loading β. Consequently, we discuss the asset pricing implications of the model by focusing on the systematic risk loading β given by

$$\beta[h_t,p_t]=\overbrace{\underbrace{2\gamma}_{\substack{\text{Risk loading of}\\ \text{revenues}}}\quad\underbrace{(1+\iota\Theta[p_t])}_{\substack{\text{Risk amplification}\\ \text{from operating leverage}}}}^{\substack{\text{Risk loading of assets in place}\\ \text{with no entry threat}}}\left(\frac{\Xi[h_t,p_t]}{\delta}\right)+\overbrace{\upsilon\gamma}^{\substack{\text{Risk loading of}\\ \text{entry threat}}}\left(1-\frac{\Xi[h_t,p_t]}{\delta}\right), \tag{14}$$

where $\gamma\equiv\rho\sigma>0$ and $\iota\equiv 1+\frac{\upsilon}{2}\left(\frac{\delta}{r}-1\right)>1$. Equation (14) decomposes a firm's systematic risk loading or *beta* into two terms: the beta of an otherwise identical firm in a hypothetical industry that has no entry threat, and the systematic risk loading due to the entry threat by competitors. The systematic riskiness of revenues equals 2γ and is multiplied by a term that captures the risk amplification due to operating leverage, $(1+\iota\Theta)>0$. The systematic riskiness of future expansions in the industry is captured by the term $\upsilon\gamma$. The two components are weighted by $\frac{\Xi}{\delta}$ and $\left(1-\frac{\Xi}{\delta}\right)$, where Ξ is the earnings-to-price ratio. Furthermore, note that Lemma 1 implies that the term $\left(1-\frac{\Xi}{\delta}\right)$ is always negative so that effect of the entry threat on expected returns is always negative.

We use Equation (14) to characterize the effect of an increase in concentration h on firms' exposure to systematic risk. The first term in Equation (14), joined with the result from Proposition 1, implies that the systematic risk amplification due to operating leverage is decreasing in concentration h. We refer to this effect as the *operating leverage channel*. Moreover, a reduction in concentration h lowers the entry threshold $\overline{p}[h]$ such that entry becomes more likely for any given level of profit margin p. Given the negative impact of entry on cash flows is more likely when the industry shock X is high, incumbents in industries with lower entry threshold (i.e, more competitive industries) are effectively less exposed to systematic risk. We refer to this effect as the *entry threat channel*.

A similar argument shows that the operating leverage and entry threat channels also apply to the relation between a firm's systematic risk loading β and markup m. The following proposition formalizes how concentration h and markup m affect a firm's systematic risk loading β.

Proposition 3 (effect of concentration *h* and markup *m* on systematic risk exposure). *Given $V[h,p]>0$, the effect of either concentration h or markup m on expected returns depends on the degree of operating leverage Θ such that*

i. $0<\Theta[p]<\hat{\Theta} \iff \frac{\partial\beta[h,p]}{\partial h}>0$ *and* $0<\Theta[p]<\hat{\Theta} \iff \frac{\partial\beta[h,p]}{\partial m}>0,$

ii. $\frac{\partial}{\partial\Theta[p]}\left(\frac{\partial\beta[h,p]}{\partial h}\right)<0$ *and* $\frac{\partial}{\partial\Theta[p]}\left(\frac{\partial\beta[h,p]}{\partial m}\right)<0,$

where the constant threshold on operating leverage $\hat{\Theta}>0$ is defined in Appendix B.

Proof. See Appendix B. ■

The first part of Proposition 3 predicts a positive relation between expected returns and concentration h and markup m as long as the degree of operating leverage in the industry is sufficiently low. Intuitively, when firms have sufficiently high profits to buffer adverse shocks such that $\Theta[p]<\hat{\Theta}$, the effect of the entry threat channel dominates, and the sensitivity of expected returns to either concentration h or to markup m is positive. Conversely, in the alternative case in which $\Theta[p]>\hat{\Theta}$, the operating leverage channel dominates, and expected returns become decreasing in concentration h and markup m. The third row of Figure 1 illustrates how the sign of the sensitivity of expected returns to concentration and markup reverses as operating leverage increases.

The second part of the Proposition 3 provides a conditional, albeit more general, statement on the effect of the operating leverage and entry threat channels on expected returns. This conditional statement is useful to derive testable implications, as it does not hinge on the unobservable threshold $\hat{\Theta}$ at which the sensitivity of expected returns to either concentration h or to markup m changes its sign. The second part of Proposition 3 predicts that, regardless of the region considered, the sensitivity of expected returns to either concentration h or to markup m becomes more positive in industries with lower operating leverage.

Proposition 3 characterizes the effect of changes in either concentration h or in markup m on expected returns. In this sense, the analysis so far implicitly takes as given the degree of competition in the industry. However, the relation between asset prices and the degree of competition in the industry is bidirectional, since the competitive environment is itself affected by the industry's exposure to systematic risk. The following proposition characterizes the effect of the industry's exposure to systematic risk on the competitive environment.

Proposition 4. (effect of systematic risk exposure on concentration *h* and markup *m*). *Given an industry A that is riskier than industry B such that* $\beta_{\mathrm{A}} - \beta_{\mathrm{B}} > 0$, *and all else being equal:*

i. *if* $p_{\mathrm{A},t} = p_{\mathrm{B},t} \Rightarrow h_{\mathrm{A},t} > h_{\mathrm{B},t},\ m_{\mathrm{A},t} > m_{\mathrm{B},t},$ *and* $\Pr[_{\mathrm{A,t,T}}] < \Pr[_{\mathrm{B,t,T}}]$,

ii. *if* $h_{\mathrm{A},t} = h_{\mathrm{B},t}$ *or* $m_{\mathrm{A},t} = m_{\mathrm{B},t} \Rightarrow p_{\mathrm{A},t} < p_{\mathrm{B},t},$ *and* $\Pr[_{\mathrm{A,t,T}}] < \Pr[_{\mathrm{B,t,T}}]$,

where $\Pr[_{\mathrm{i,t,T}}]$ *is the probability that a new firm enters industry i between dates t and* $T > t$.

Proof. See Appendix B. ■

Proposition 4 relates to the earlier studies by Subrahmanyam and Thomadakis (1980) and Pindyck (2009). It shows that the exposure to systematic risk is itself a barrier to entry that affects the degree of competition in the industry. Intuitively, the probability of entry $\Pr[_{\mathrm{i,t,T}}]$ is informative of future expected changes in the concentration h and markup m. If a given industry is riskier than another, the likelihood of entry is lower in the riskier industry. Part (i) of Proposition 4 shows that, controlling for the heterogeneity in the profit margin p, the riskier industry is more concentrated today and its future concentration is expected to remain higher since the probability of entry is lower in the industry. Similarly, Part (ii) of the proposition shows that, controlling for the heterogeneity in either concentration h or markup m, riskier industries have a lower likelihood of entry and thus a relatively higher expected concentration in the future. In sum, Proposition 4 implies that riskier industries remain relatively less competitive over time, since systematically riskier cashflows act as a barrier to entry. We refer to this effect as the *risk feedback channel*.

2. Empirical Strategy

2.1 Testable hypotheses

We begin the discussion of our empirical strategy by condensing the main implications of the model into three testable hypotheses. We start with the testable hypotheses for the unconditional bidirectional relation between expected returns, concentration, and markup.

Hypothesis 1 (unconditional implications for expected returns).

i. *For industries with "sufficiently low" levels of operating leverage, expected returns are positively related to concentration and markup.*
ii. *For most industries, the level of operating leverage is "sufficiently low."*

Proposition 3 suggests that the relation between expected returns and either concentration or markup is positive when the degree of operating leverage is below an endogenous industry-specific threshold, and it is negative otherwise. Moreover, the risk feedback channel (Proposition 4) favors a positive relation between expected returns and concentration and markup over time. The risk feedback channel prevents operating leverage from becoming too high in a given industry. Consequently, we conjecture that the negative effect of the operating leverage channel is mitigated over time, such that expected returns are positively related to concentration and markup, on average.[12]

Hypothesis 2 (conditional implications for expected returns). *The relation between expected returns and concentration or markup is relatively more positive in industries with relatively lower operating leverage.*

The positive unconditional relation between expected returns, concentration, and markup formalized by Hypothesis 1 is in fact a joint hypothesis that the theoretical predictions of our models are true and that the risk feedback channel is either "strong enough" or it has had "enough time" to keep operating leverage levels sufficiently low in the economy. Hypothesis 2 focuses on the relative ranking of operating leverage across industries. This hypothesis suggests a more direct test of the theoretical implications of the model involves the conditional relation between expected returns, concentration, and markup that does not rely on determining whether the degree of operating leverage of an industry is above or below an unobservable threshold.

Hypothesis 3 (operating leverage and relative valuation ratios).

i. *The average operating leverage is negatively related to concentration and markup.*
ii. *The average earnings-to-price ratios are negatively related to concentration and markup.*

Both the unconditional and conditional implications for the relation between expected returns, concentration, and markup formalized by Hypotheses 1 and 2 rely on the existence of the entry threat and operating leverage channels. We consider Hypothesis 3 to check empirically whether the earnings-to-price ratios, as well as the operating leverage of firms, are indeed decreasing in both concentration and markup, as predicted by the model.[13]

12 In the Online Appendix, we include a numerical illustration of the joint implication of Propositions 3 and 4, showing that average realized asset returns relate positively to concentration and markup in a large cross-section of industries simulated over extended periods.

13 While Proposition 2 explicitly refers to earnings-to-price ratios, it can be interpreted more generally as a prediction about firms' inverse valuation ratios and thus also to book-to-market ratios. In this sense, the first part of Hypothesis 3 relates to a prediction by Grenadier (2002) that less competitive firms better preserve their value.

2.2 Empirical measures of markup and concentration

2.2.1 Data sources and sample construction. Our working sample is constructed with data from multiple sources. For variables related to industry-level characteristics, we use data from the Bureau of Economic Analysis (BEA), the U.S. Census Bureau's Center for Economic Studies (CES), the survey of Statistics of U.S. Businesses (SUSB) by the U.S. Census Bureau, and the reports of the Census of Manufactures published by the U.S. Census Bureau. We use financial and accounting data from both the Compustat and CRSP/Compustat Merged data sets. The data from NBER–CES and the data from the Census of Manufactures publications only cover manufacturing industries, while the remainder of the data sources also cover nonmanufacturing industries.

Our working sample covers the period from 1992 to 2014, in which the starting year is determined by the availability of data on the number of firms by industry for both manufacturing and nonmanufacturing industries from the Statistics of U.S. Businesses (SUSB). We rely on SUSB data to implement our empirical methodology to correct for the sample selection bias of public listing and to construct our alternative measure of concentration (CBC). The SUSB data set reported by the U.S. Census Bureau covers both manufacturing and nonmanufacturing industries.

Throughout our analysis, we define an industry as the universe of firms within the same four-digit Standard Industrial Classification (SIC) code. We follow Hoberg and Phillips (2010) and use SIC codes from Compustat.[14] As discussed in earlier studies, the four-digit SIC codes in Compustat often end in zero or nine, masking a broader three-digit industry definition. We address this problem by replacing the SIC code of firms whose SIC ends in zero or nine with the SIC code of the main segment as reported in the Compustat segment tapes.[15] We then remove those firms whose four-digit SIC still ends in zero or nine after this adjustment. We also eliminate conglomerate firms from the sample. To accomplish this, we follow Gopalan and Xie (2011) and identify conglomerates as those firms in the Compustat segment tapes that have more than three segments. We provide further details of our sample construction in Appendix C.

2.2.2 Industry-level measures of markup. We broadly define average markup as the sum of the operating profits of all firms in a given industry and year, divided by the corresponding total volume of industry sales in that year. We construct two alternative measures: one based on the NBER–CES Manufacturing Industry Database, and another based on Compustat data. The first measure is limited to manufacturing industries, but it is arguably more

---

14 The studies by Guenther and Rosman (1994); Kahle and Walkling (1996); and Bhojraj, Lee, and Oler (2003) study the trade-offs between using historical SIC codes from CRSP and using SIC codes from Compustat. They conclude that Compustat-based SIC codes are generally more accurate.

15 We thank an anonymous referee for suggesting this procedure.

accurate, since it is based on both public and private firms in each industry. The second measure covers a wider cross-section of industries, but it is constructed with data from publicly listed firms only.

2.2.2.1 Markup based on U.S. Census data. We follow Allayannis and Ihrig (2001) and construct our first measure of average markup (hereafter, MKP) as

$$MKP_t \equiv \frac{Value\ of\ Sales_t + \Delta Inventories_t - Payroll_t - Cost\ of\ Materials_t}{Value\ of\ Sales_t + \Delta Inventories_t}, \tag{15}$$

where *Value of Sales*, *Inventories*, *Payroll*, and *Cost of Materials* are industry-level variables obtained from the NBER–CES Manufacturing Industry Database. This database is compiled from reports from the Annual Survey of Manufactures published by the U.S. Census Bureau. NBER–CES data, as of the date of this writing, are available at an annual frequency until 2011, so we forward-fill this measure up through 2012.

2.2.2.2 Markup based on Compustat data. We construct our second measure of average markup (hereafter MKC) with Compustat annual data as

$$MKC_t \equiv \frac{Value\ of\ Sales_t - Value\ of\ Costs\ of\ Good\ Sold_{it}}{Value\ of\ Sales_t}, \tag{16}$$

where the value of sales equals the sum of firms' sales by industry–year (i.e., Compustat annual data item SALE), and the value of the cost of goods sold equals the sum of firms' cost of goods sold by industry–year (i.e., Compustat annual data item COGS).

Despite the fact that MKC only covers public firms, we observe a significantly high correlation (64%) between MKP and MKC for manufacturing industries, as shown in panel A of Table 1. The table shows that the magnitude of this correlation is much higher than the correlation between alternative widely accepted measures of imperfect competition, such as MKP and HHI (18%). This finding suggests that a Compustat-based average industry markup measure is not subject to significant sample selection bias due to public listing.

2.2.3 Measures of concentration. A commonly used measure of industry concentration in the recent finance literature on competition and firm risk is the sales-based Herfindahl–Hirschman index, defined as

$$\text{Herfindahl–Hirschman Index}_t \equiv \sum_j s_{jt}^2, \tag{17}$$

where s_j is the market share of sales of firm j in the industry.

Table 1
Correlations

A. Correlations between the IPMC measures

IPMC measure	HHI	MKP	CBC	MKC	HHI comp.	CR8	$1/N^{All}$
MKP	0.18***						
	(0.01)						
CBC	0.39***	0.09***					
	(0.04)	(0.02)					
MKC	0.00	0.64***	0.09***				
	(0.01)	(0.02)	(0.02)				
HHIcomp.	0.09**	−0.10***	−0.10***	−0.06**			
	(0.04)	(0.03)	(0.02)	(0.02)			
CR8	0.75***	0.14***	0.37***	0.04***	0.14***		
	(0.00)	(0.00)	(0.05)	(0.00)	(0.01)		
$1/N^{All}$	0.34***	0.15**	0.41***	0.00	0.16***	0.26***	
	(0.01)	(0.06)	(0.01)	(0.02)	(0.01)	(0.06)	
$1/N^{Pub}$	−0.11***	−0.15***	−0.12***	−0.09***	0.72***	0.02	0.15***
	(0.03)	(0.00)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)

B. Correlations between the share of public firms and the IPMC measures

Share public	HHI	MKP	CBC	MKC	HHI comp.	CR8	$1/N^{All}$	$1/N^{Pub}$
N^{Pub}/N^{All}	0.19***	0.35***	0.42***	0.15***	−0.10***	0.13***	0.48***	−0.18***
	(0.01)	(0.04)	(0.02)	(0.01)	(0.02)	(0.03)	(0.02)	(0.01)

Newey-West standard errors are shown in parentheses. *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.
The table below shows estimates and standard errors of univariate Fama-MacBeth regressions where both dependent and independent variable are standardized (i.e., demeaned and rescaled to have standard deviation of one in each year of the sample). The estimates shown represent average cross-sectional correlation coefficients between the two variables. *HHI* is the logarithm of the Herfindahl-Hirschman index of sales of firms in the industry, *MKP* is average industry markup, CBC is the characteristic-based concentration measure, *MKC* is the average industry markup from Compustat data. *HHI comp.* is the logarithm of the Herfindahl-Hirschman index (x10,000) of sales of firms based on Compustat data, constructed as in Hou and Robinson (2006). *CR8* is the market share of the eight largest firms in the industry. N^{All} is the total number of firms in the industry. N^{Pub} is the number of firms in the industry in the Compustat sample. The sample period is 1992–2012.

2.2.3.1 Herfindahl–Hirschman index from the U.S. Census. We obtain Herfindahl–Hirschman index (HHI) measures from the U.S. Census of Manufactures publications from 1992, 1997, 2002, 2007, and 2012. In this paper, we represent the logarithm of the Herfindahl–Hirschman index from the U.S. Census simply as HHI.[16] The U.S. Census provides this measure at the four-digit SIC level in 1992 and provides it at the six-digit NAICS level in the remaining years. For this reason, we convert the measure to four-digit SIC levels using the methodology in Ali, Klasa, and Yeung (2009). We forward-fill the measure through 2012. We provide more details about the HHI measure in Appendix C.

2.2.3.2 Characteristics-based concentration. To extend our sample to nonmanufacturing industries, in this paper we propose an alternative sales

16 Maury and Pajuste (2005) argue that the use of the HHI in its logarithmic form helps to control for skewness in panel regressions. The log transformation of the HHI does not materially affect the main results of this paper, which is unsurprising since our analyses are cross-sectional. The log-transformation of the measure does not affect the ordering of industries in any given year.

concentration measure, which we term *vharacteristic-based concentration* (hereafter CBC). To construct the CBC measure, we start by noting that the Herfindahl–Hirschman index from Equation (17) can be re-expressed as a function of the number of firms in the industry (N), as well as the cross-sectional mean (μ_{Sales}) and variance (σ^2_{Sales}) of firm-level sales in the industry:

$$\text{Herfindahl–Hirschman Index}_t = \frac{1}{N_t}\left(\frac{\sigma^2_{\text{Sales},t}}{\mu^2_{\text{Sales},t}}+1\right). \tag{18}$$

The expression above shows that N, μ_{Sales}, and σ^2_{Sales} are sufficient statistics for the full vector of market share of firm-level sales in the construction of the Herfindahl–Hirschman index. As discussed in the next section, there is evidence that the cross-sectional industry means and the cross-sectional industry variances of the sales of public firms are biased. To overcome this problem, we use Compustat sales data of public firms and apply a sample selection correction methodology to obtain unbiased estimates of the average industry sales ($\hat{\mu}_{\text{Sales}}$), the cross-sectional variance of sales ($\hat{\sigma}_{\text{Sales}}$), and the total number of firms ($\hat{N}$) for all industry–years in the Compustat data set. We construct the CBC measure by replacing μ_{Sales}, σ_{Sales}, and N in Equation (18) with their unbiased estimates:

$$\text{CBC}_t = \log\left[\frac{1}{\hat{N}_t}\left(\frac{\hat{\sigma}^2_{\text{Sales},t}}{\hat{\mu}^2_{\text{Sales},t}}+1\right)\right]. \tag{19}$$

where we define the CBC measure in its logarithmic form so it is consistent with the HHI measure. We provide further details on the CBC measure in the appendix.

We construct the CBC measure for all industries in the Compustat data set since 1992. The correlation between CBC and HHI is informative of the goodness-of-fit in the second stage. Panel A of Table 1 shows that the correlation between these two measures is 39%, which is significantly higher than that between HHI and the Compustat-based concentration measure (9%). Similarly, the correlation between the CBC and the concentration ratio for the 8 largest firms in the industry reported by the U.S. Bureau (CR8) is 37%.[17]

2.3 Sample selection correction

An understated challenge in the empirical research is that we only observe the financial and accounting data of publicly listed firms. This limitation would not

[17] In the Online Appendix, we present results using the CR8 and the CR4 (i.e., the concentration ratio for the four largest firms in the industry) as alternative concentration measures. Similarly, the HHI and CBC measures of concentration abstract away from the role of importers competing with local incumbents in U.S. industries. For this reason, we also run our tests using an alternative measure, which we term the *concentration adjusted for Imports* (CAI); it adjusts the CR4 measure for import penetration following the methodology in Abdel-Raouf (2010). The results with the CAI measure are also reported in the Online Appendix. All results are qualitatively similar to those with the concentration measures in Section 3.

pose a significant concern in our empirical strategy if the subsample of publicly listed firms were random or unaffected by the competitive environment. However, the results in Table 1 indicate the existence of sample selection bias due to public listing in our data set. Panel B of Table 1 shows that firms are more likely to be publicly listed in less competitive industries. The correlation between the share of public firms and measures of imperfect competition considering private and public firms ranges from 19% to 42%. Moreover, Table 2 shows that the total number of publicly listed firms is higher in less competitive industries, which already suggests that a Compustat-based measure of concentration should be biased. The results in panel B of Table 1 confirm this intuition: the correlation between the share of public firms in an industry and the (unadjusted) Compustat-based measure of concentration is negative (-10%). As additional evidence, and consistent with the findings in Ali, Klasa, and Yeung (2009), panel A of Table 1 reports a low correlation between the HHI and the unadjusted Compustat-based measure of concentration (9%).[18]

Given these findings, we propose a methodology to control for the sample selection bias of public listing in two different aspects of our empirical analysis. First, we adjust the sales of publicly listed firms to construct the CBC measure. Second, we apply the sample-selection correction to unbias the variables relevant to test Hypotheses 1–3, given that we only observe them for publicly listed firms. The proposed sample selection correction consists of a two stage methodology adapted from Heckman (1979). We elaborate on each of these two stages below.

2.3.1 First stage: Selection model. The first stage of a selection model as in Heckman (1979) involves the estimation of a probit model in which the dependent variable equals 1 if the firm is publicly listed and 0 otherwise. Since we do not observe the characteristics of private firms at the firm level, we are unable to compute such a model. Instead, we propose an alternative methodology that follows the empirical approach to test selection models of proportions data discussed in Greene (2003). We use the results of the first stage to compute an inverse Mills ratio.

In particular, we compute the average probability that a firm is public in a given industry–year. The underlying assumption of our analysis is that, conditional on observable information at the industry level, all firms in the same industry have the same probability of being in the sample in any given year. The proposed methodology has therefore an advantage and a shortcoming. The advantage is that it enables us to compute inverse Mills ratios using available data. The shortcoming is that, to overcome a data limitation, we neglect the impact of firm-specific characteristics in explaining firm's public status. As we

[18] The table also shows that the Compustat-based concentration measure is *negatively* related to alternative measures of competition: MKP (−10%), MKC (−10%), and CBC (−6%).

elaborate below, the explanatory power of the first stage remains significant in spite of this limitation.

Let $p_{\mathrm{it}}^{\mathrm{Pub}} \equiv \frac{N_{\mathrm{it}}^{\mathrm{Pub}}}{N_{\mathrm{it}}^{\mathrm{All}}}$ be the proportion of public firms ($N_{\mathrm{it}}^{\mathrm{Pub}}$) to total firms ($N_{\mathrm{it}}^{\mathrm{All}}$) in industry i in year t. The selection model is given by

$$p_{\mathrm{it}}^{\mathrm{Pub}} = \Phi(\gamma_{\mathrm{t}} \mathbf{z}_{\mathrm{i,t-l}}) + \zeta_{\mathrm{it}}, \tag{20}$$

where $\mathbf{z}_i$ denotes the vector of industry-specific characteristics that determine the public status of firms in a given industry group, Φ denotes the normal cumulative density function, ζ_i is the sampling error, and l denotes the lag in years.[19] We treat the sampling of public firms within the same industry as a problem of sampling from a Bernoulli population. Hence ζ_i is such that $\mathrm{E}_t[\zeta_i]=0$, and $\mathrm{Var}_{\mathrm{t}}[\zeta_{\mathrm{i}}]=\Phi(\gamma_{\mathrm{t}}\mathbf{z}_{\mathrm{i,t-l}})(1-\Phi(\gamma_t \mathbf{z}_{\mathrm{i,t-l}}))\left(N_{\mathrm{it}}^{\mathrm{Pub}}\right)^{-1}$.

We follow Greene (2003) and estimate the selection model in Equation (20) for every year in our sample in two passes, where the first pass is used to construct weights used in the second pass to correct for heteroscedasticity.[20] The vector $\mathbf{z}_i$ includes variables that explain the public listing status of firms. We specify the vector $\mathbf{z}_i$ differently depending on the particular second stage application, which we discuss later. In all specifications, however, we include two instruments as explanatory variables in the vector $\mathbf{z}_i$.[21]

The instruments we include in the first stage are variables that are related to a firm's decision to be publicly listed but are unrelated to the dependent variables used in the second stage (i.e., the mean and variance of firms' sales). In particular, the first instrument is the average turnover of a firm's shares by industry, used as a proxy for the average amount of investor demand for stocks by industry. Chemmanur, He, and Nandy (2010) argue that the average stock turnover can be used as a proxy for the amount of information produced for a given stock. Similarly, Bernstein (2015) uses NASDAQ fluctuations during the book building phase as an instrument for initial public offering (IPO) completion.[22] We conjecture that the investor demand by industry group significantly affects the decision to go public, but it is orthogonal to the dependent variable in the second stage.

The second instrument used in the first stage regression is the share of public firms in the industry that are quoted at the New York Stock Exchange (NYSE).

---

[19] We consider lagged explanatory variables such that these are predetermined relative to p_{it}^{Pub}. We use a lag of one year in the CBC construction and a lag of two years in the empirical tests. We use 2-year lags throughout our empirical tests for consistency and to be conservative, since some of the empirical tests use annual returns (i.e., include 1-year lagged market values).

[20] See Appendix D for further details of the estimation methodology.

[21] To construct the CBC measure, the explanatory variables in the vector $\mathbf{z}_i$ that are not instruments are consistent with the empirical literature on the decision to go public. To test our hypotheses, the set of variables in vector $\mathbf{z}_i$ that are not instruments equals a set of controls that explain cross-sectional variation in the dependent variable in the second stage. We elaborate on the specification of vector $\mathbf{z}_i$ for each case in Appendix E.

[22] The acronym NASDAQ stands for National Association of Securities Dealers Automated Quotations.

Given that each stock exchange has different listing requirements and different listing fees, this variable captures the extent to which most public firms in a given industry meet the listing requirement to quote on the NYSE exchange relative to other exchanges. We conjecture that listing requirements affect a firm's decision to go public, but they are orthogonal to the dependent variable in the second stage.

We provide further details on the construction of the inverse Mills ratios and the goodness-of-fit of the first stage in Appendices D and E. In particular, Appendix Table E1 shows that the average R-squared is 32% the first stage of the CBC measure. The sign and significance of our instruments are also aligned with previous studies. Consistent with Chemmanur, He, and Nandy (2010), firms are more likely to go public in industries in which there is higher investor demand. Moreover, firms are more likely to go public in industries with higher proportions of public firms listed on NYSE. This is consistent with the evidence in Corwin and Harris (2001), who report that the initial size and age of firms conducting IPOs at NYSE are significantly higher than those of firms conducting IPOs at NASDAQ.

2.3.2 Second stage: Cross-sectional regressions. In the second stage of our sample selection model, in line with Heckman (1979), we use the inverse Mills ratio from the first stage to correct for the sample selection bias of public listing. We run the cross-sectional regressions of the dependent variable of interest v_{ij} on controls and on the inverse Mills ratio such that:

$$v_{\mathrm{ijt}} = c_{0,\mathrm{t}} + \mathbf{c}_{2,\mathrm{t}} \mathbf{x}_{\mathrm{ij,t-1}} + c_{3,t} \lambda_{\mathrm{i,t-1}} + \epsilon_{\mathrm{ijt}}, \tag{21}$$

where λ is the inverse Mills ratio derived in Appendix D, $\mathbf{x}$ is a vector with the same controls used in the first stage but excluding the two instruments, and l denotes the lag in years.[23] The dependent variable v_{ij} is determined by the specific application in which we apply the sample selection correction. For instance, the dependent variable v_{ij} is the logarithm of sales in the construction of the CBC measure, or the earnings-to-price ratio, the operating leverage ratio, or realized excess stock returns in the empirical tests of Hypotheses 1–3.

3. Empirical Evidence

3.1 Summary statistics

Table 2 reports average characteristics of quintile portfolios of firms sorted by our two measures of industry markup and our two measures of concentration. The first column of the table shows the quintile of the standardized measure of either concentration or markup. The next two columns of Table 2 show

[23] Consistent with the literature on sample selection, we specify vector $\mathbf{x}_i$ to include the variables in vector $\mathbf{z}_i$, but we exclude the instruments used during the first stage.

that, consistent with the results in Table 1, the total number of firms in the industry tends to decrease across the imperfect market competition portfolios. Also consistent with the findings of Table 1, Table 2 shows that the number of *publicly listed* firms in the sample tends to *increase* over the quintiles. These findings reinforce the ideas that the sample of publicly listed firms is not random, and that the decision to be publicly listed relates to the degree of competition in each industry.

Table 2 also shows that the average market value of assets of firms is generally higher in industries with high markup or high concentration. This is consistent with our model's prediction that firms in less competitive industries better preserve their value. Another possible reason for this trend could simply be the larger scale of operations in these industries. As seen in the next column, the average book value of assets of firms is generally increasing across HHI quintiles. However, there is no clear trend in the average book value of assets across the quintile portfolios formed with the three other measures of imperfect market competition (MKP, CBC, and MKC). Consistent with the prediction of Hypothesis (3), Table 2 shows that operating leverage is in fact greater in less competitive industries. We use the measure of operating measure from Novy-Marx (2011), who defines it as the ratio of the sum of operating costs and administrative expenses to total assets.

The trends in the subsequent columns in Table 2 are also consistent with Hypothesis (3). The average earnings-to-price and book-to-market ratios are generally lower in less competitive industries. As motivated by our model, the greater earnings-to-price ratios and book-to-market ratios found in more competitive industries is consistent with the existence of the entry threat channel, which lowers the risk exposure in these industries. The connection between low valuations and low risk among firms in more competitive industries is explained by the effect of the threat of entry on the value of incumbent firms: the greater the threat of entry, the more likely that future cash flows will be negatively affected, which in turn makes firms less valuable.

Table 2 shows that firms in high-markup or high-concentration industries have, on average, lower financial leverage ratios. One possible explanation for the lower financial leverage in these industries is their greater average exposure to systematic risk, consistent with our testable hypotheses. Another possible explanation is a possible trade-off between operating and financial leverage that would lead firms in more competitive industries (with greater operating leverage) to chose lower levels of financial leverage.[24] Given that financial leverage amplifies the exposure of a firm's equity holdings to systematic risk, we control for financial leverage in the analyses that follow.

The last column of Table 2 reports the average share of manufacturing industries across the quintile portfolios. By construction, HHI and MKP only

[24] See Mandelker and Rhee (1984) and Dugan, Minyard, and Shriver (1994) for discussions of the trade-off between operating and financial leverage.

Table 2
Summary statistics

Port.	Measure (SD)	Log N^{All}	Log N^{Pub}	Log size	Log asset	OL	E/P	B/M	Lev.	Share man.
A. HHI										
L	−1.40	6.59	1.92	5.14	5.41	1.20	0.21	0.88	0.37	1.00
2	−0.41	6.21	2.30	5.53	5.59	1.12	0.19	0.75	0.34	1.00
3	0.05	5.86	2.51	5.75	5.92	1.11	0.20	0.83	0.37	1.00
4	0.52	5.74	2.34	6.02	6.10	1.15	0.18	0.73	0.35	1.00
H	1.22	5.51	2.14	6.33	6.44	1.03	0.19	0.72	0.36	1.00
B. MKP										
L	−1.21	5.72	1.97	5.77	6.13	1.41	0.22	0.89	0.41	1.00
2	−0.57	6.08	2.13	5.61	5.83	1.21	0.20	0.82	0.38	1.00
3	−0.11	5.99	1.99	5.54	5.70	1.06	0.19	0.79	0.35	1.00
4	0.37	5.89	2.22	6.07	6.20	1.01	0.20	0.72	0.36	1.00
H	1.52	6.03	2.97	6.13	5.88	0.89	0.16	0.65	0.28	1.00
C. CBC										
L	−1.19	9.10	2.10	5.74	6.00	1.56	0.22	0.81	0.40	0.19
2	−0.53	7.44	2.16	5.79	6.04	1.34	0.22	0.81	0.40	0.61
3	−0.14	6.58	2.27	5.86	6.10	1.23	0.22	0.82	0.39	0.73
4	0.33	6.28	2.68	6.16	6.31	1.13	0.21	0.77	0.37	0.72
H	1.53	6.64	4.02	5.92	5.79	0.95	0.16	0.66	0.30	0.70
D. MKC										
L	−1.26	7.63	2.32	5.72	6.21	1.75	0.24	0.86	0.47	0.42
2	−0.61	7.17	2.22	5.71	6.08	1.36	0.24	0.85	0.43	0.59
3	−0.14	6.98	2.22	5.77	5.92	1.20	0.20	0.76	0.37	0.65
4	0.45	7.11	2.40	5.78	5.84	1.02	0.22	0.82	0.34	0.52
H	1.55	6.87	2.96	6.03	5.83	0.84	0.18	0.63	0.29	0.67

The table below reports average characteristics of portfolios of industries sorted by measures of concentration (*HHI* and *CBC*) and markup (*MKP* and *MKC*), described in Table 1. *Measure (SD)* is standardized measure of imperfect competition (i.e., demeaned and rescaled to have standard deviation of one in each year of the sample). *Log* N^{All} is the logarithm of the total number of firms in the industry from SUSB/Census. *Log* N^{Pub} is the logarithm of the number of firms in the industry in the Compustat sample. *Log size* is the logarithm of the market value of equity plus book value of total debt. *Log asset* is the logarithm of the book value of assets. *OL* is the measure of operating leverage from Novy-Marx (2011), defined as costs of goods sold plus sales, general, and administrative expenses over total assets. *E/P* is earnings divided by the market value of equity. *B/M* is shareholders equity divided by the market value of equity. *Lev.* is the ratio of book value of debt, adjusted for cash holdings (as reported in Compustat), divided by the assets. *Share man.* is the share of manufacturing firms (SIC codes from 2000 to 3999). The sample period is 1992–2012.

cover manufacturing industries, while CBC and MKC include nonmanufacturing industries. The average proportion of manufacturing industries is weakly increasing across the quintiles in the CBC measure.

3.2 Systematic risk loadings and expected returns

The testable asset pricing implications of our model are given by Hypotheses 1 and 2. Hypothesis 1 predicts a positive overall relation between expected returns, concentration, and markup. A test of Hypothesis 1 is in fact a test of the joint hypotheses that the proposed economic mechanisms embedded in our model are valid *and* that the effect of the risk feedback channel is economically relevant. Hypothesis 2 predicts a relatively more positive relation between expected asset returns and concentration and markup in industries with relatively lower operating leverage. The analysis of Hypothesis 2 thus offers a more direct test of the validity of our proposed economic mechanisms.

We present supporting empirical evidence on each of these predictions in steps. First, we report the empirical evidence on the average net relation between competition and exposure to risk (Hypothesis 1). We then show that the empirical evidence is also consistent with the conditional testable predictions of the model (Hypothesis 2).[25]

3.2.1 Expected returns, concentration, and markup: Unconditional tests. We investigate Hypothesis 1 using portfolio sorts and regression based analyses. We first consider portfolio sorts. We note that we face the challenge that expected returns are not directly observed. To address this challenge, which is common to most asset pricing studies, we use different proxies of expected returns. Table 3 reports the averages of each of eight different measures of expected returns of equally weighted quintile portfolios sorted on lagged measures of concentration and markup.

Panel A in Table 3 reports portfolio sorts using proxies for expected returns based on realized returns. The first two proxies used are the excess stock returns (R^{E}) and unlevered stock returns ($R^{E,U}$). The unlevered stock returns are stock returns adjusted for financial leverage. We consider excess stock returns since these are the most commonly used proxies for expected returns in the literature. However, unlevered stock returns are more closely related to the predictions of our model, which are about expected *asset* returns.[26] The last two proxies for expected returns in panel A in Table 3 are based on stock returns adjusted for firm characteristics: stock returns adjusted for size, book-to-market ratio, and past returns (R^{DGTW}), constructed according to the methodology in Daniel et al. (1997) (DGTW), and R^{ADJ}, which is a variation of R^{DGTW} in which, instead of the adjustment for past returns, we adjust returns for cross-sectional variation in profit margins (EBITDA / Assets). This specification allows us to disentangle the effect of concentration and markup on returns from the effect of profitability, discussed in Novy-Marx (2013).

Panel B in Table 3 reports portfolio sorts using proxies for expected returns based on traditional factor models. The first two proxies are based on the expected returns implied by the CAPM: unadjusted excess expected returns from the CAPM (R^{CPM}), as well as unlevered excess expected returns from the CAPM ($R^{CPM,U}$). The last two proxies for expected returns in panel A in Table 3 are based on expected returns implied by multifactor models: the

[25] We provide additional empirical evidence in support of our testable hypotheses using alternative measures of industry concentration (i.e., concentration adjusted for imports or CAI) in the Online Appendix.

[26] In the Online Appendix, we show that, in the context of our model, asset returns are equal to excess stock returns times one minus the leverage ratio. The result is shown under the assumption that debt is risk free. We follow this procedure to unlever equity returns and equity systematic risk loadings in our tests.

Fama-French three-factor model (R^{FF3}) from Fama and French (1993), and the Fama–French–Carhart four-factor model from Carhart (1997) (R^{FC4}).[27]

Overall, the empirical evidence in Table 3 shows that expected returns are generally increasing in concentration and markup. The results with excess stock returns, which are the most comparable to the rest of the literature, indicate that the relation between expected returns and our measures of competition is economically significant. Firms in the highest concentration or markup industry quintiles have excess stock returns that are, on average, 5% to 8% higher than those of firms in the lowest industry quintiles.

Results with proxies of expected returns based on excess stock returns adjusted for firm characteristics and with on proxies of expected returns implied by factor models further strengthen the evidence in support of the prediction that expected returns are unconditionally and positively related to concentration and markup. The results with R^{DGTW} and R^{ADJ} are relatively stronger than those with R^{E} and $R^{E,U}$, most likely because the adjustment for firm-level characteristics effectively controls for intra-industry differences in exposure to risk that are unrelated to our proposed mechanisms. Last, the sorts with measures of expected returns based the on factor models (panel B) provide the strongest support for our hypothesis that firms in less competitive industries are riskier than those in more competitive industries. We posit that the results in panel B are starker because of the known problems related to the use of realized returns as proxies for expected returns over short sample periods.

We next investigate Hypothesis 1 on the relation between expected returns and our measures of concentration and markup using regression analysis based on two alternative specifications. The first specification controls for the sample selection bias of public listing. We apply the two-stage methodology in Section 2.3 to the variable of interest v (e.g., realized returns) and we save the residual ϵ of the second stage. Note that the residual ϵ represents the unexplained variation in the variable of interest v after controlling for standard covariates *and* sample selection. With this insight, and to avoid multicollinearity problems between the CBC and the inverse Mills ratios, we run standard Fama-MacBeth regressions of the residual ϵ on each of our measures of concentration and markup such that:

$$\epsilon_{\mathrm{ijt}} = c_{0,\mathrm{t}}^{\mathrm{I,b}} + c_{1,\mathrm{t}}\mathrm{Measure}_{\mathrm{i,t-l}} + \tilde{\epsilon}_{\mathrm{ijt}}, \qquad (22)$$

where *Measure* refers to the measure of imperfect competition (i.e., HHI, MKP, CBC, or MKC) used in the test and l denotes the lag in years.[28]

27 See Appendix C for details of the construction of the eight alternative proxies for expected returns presented in Table 3.

28 The main advantage of using the residual ϵ instead of the variable v as a dependent variable is to avoid the multicollinearity problem arising due to the fact that, by construction, our CBC measure is itself a function of inverse Mills ratios. Using a residual as an independent variable eliminates this concern at the expense of reducing the efficiency of our inference.

Table 3
Cross-section of expected returns from portfolio sorts

A. Proxies for expected returns based on realized returns

Port.	HHI		MKP		CBC		MKC	
	R^{E}	$R^{E,U}$	R^{E}	$R^{E,U}$	R^{E}	$R^{E,U}$	R^{E}	$R^{E,U}$
L	12.39	7.35	12.70	5.78	12.09	5.28	11.32	4.51
2	12.28	8.11	13.41	7.13	10.24	4.51	14.21	6.93
3	16.96	9.19	16.22	9.46	14.43	6.24	12.80	5.81
4	17.93	11.19	16.11	10.33	18.32	11.46	15.52	9.30
H	18.35	9.72	19.00	13.00	17.44	11.73	18.66	12.39
H-L	5.96*	2.36	6.30	7.22**	5.35	6.45**	7.34*	7.89***
	(3.29)	(2.02)	(5.32)	(3.00)	(4.25)	(2.54)	(4.18)	(2.28)
	R^{DGTW}	R^{ADJ}	R^{DGTW}	R^{ADJ}	R^{DGTW}	R^{ADJ}	R^{DGTW}	R^{ADJ}
L	−0.87	−1.60	−0.58	−1.16	−0.34	−1.10	−1.18	−1.48
2	1.51	0.48	0.28	−0.20	−2.00	−2.50	1.00	−0.10
3	4.43	5.09	2.78	1.73	0.31	−0.40	−0.09	−0.28
4	4.48	5.26	5.02	7.19	5.31	5.63	2.90	2.86
H	5.43	6.24	7.70	8.62	6.19	7.27	6.50	7.36
H-L	6.30	7.84*	8.28**	9.79**	6.53*	8.37**	7.67**	8.83***
	(3.67)	(4.10)	(3.75)	(3.91)	(3.77)	(3.85)	(2.85)	(2.68)

B. Proxies for expected returns based on factor models

	R^{CPM}	$R^{CPM,U}$	R^{CPM}	$R^{CPM,U}$	R^{CPM}	$R^{CPM,U}$	R^{CPM}	$R^{CPM,U}$
L	11.03	7.59	11.55	7.07	10.72	6.40	11.22	6.15
2	11.40	8.58	11.52	7.63	11.07	6.84	10.74	6.75
3	11.69	8.39	11.31	8.27	11.75	7.72	11.70	8.02
4	12.15	8.89	12.17	10.04	12.22	9.46	12.13	9.33
H	12.67	9.51	12.43	10.30	12.19	9.87	12.19	9.95
H-L	1.64***	1.91***	0.88	3.23***	1.47***	3.48***	0.97	3.80***
	(0.35)	(0.35)	(0.94)	(0.80)	(0.39)	(0.42)	(0.63)	(0.48)
	R^{FF3}	R^{FC4}	R^{FF3}	R^{FC4}	R^{FF3}	R^{FC4}	R^{FF3}	R^{FC4}
L	11.47	10.94	11.88	11.21	11.15	10.60	11.76	11.23
2	11.61	11.28	11.79	11.28	11.52	10.92	11.37	10.87
3	12.22	11.67	11.64	11.25	11.94	11.39	11.91	11.38
4	12.72	12.33	12.86	12.48	12.51	12.17	12.39	12.05
H	12.99	12.52	12.65	12.30	12.71	12.53	12.42	12.07
H-L	1.52***	1.58***	0.77*	1.08***	1.56***	1.93***	0.65***	0.84***
	(0.39)	(0.43)	(0.41)	(0.35)	(0.19)	(0.29)	(0.19)	(0.25)

E is excess stock returns, E,U is unlevered excess stock returns, *DGTW* is excess stock returns adjusted for B/M, size, and past returns constructed as in Daniel et al. (1997), ADJ is excess stock returns adjusted for B/M, size, and profitability, *CPM* is expected excess returns from CAPM, CPM,U is unlevered expected excess returns from CAPM, *FF3* is expected excess returns from Fama and French's (1993) three-factor model, and $FC4$ is expected excess returns from the Fama-French-Carhart four-factor model. Newey-West standard errors are shown in parentheses. *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.
The table below presents averages of different measures of expected returns of quintile portfolios of firms sorted by measures of concentration (*HHI* and *CBC*) and markup (*MKP* and *MKC*), described in Table 1. H-L is the zero investment portfolio long the portfolio of firms with high concentration or markup (H) and short the portfolio of firms with low concentration or markup (L). Panel A presents results with the following proxies for expected returns based on realized returns, and panel B presents results with the following proxies for expected returns based on factor models. See Appendix C for more details about the construction of the eight measures of expected returns. The sample period is 1994–2014.

The second specification is based on standard Fama–MacBeth regressions of our variable of interest (i.e., realized stock returns) on a lagged measure of concentration or markup and controls, as given by

$$v_{ijt} = c_{0,t} + c_{1,t}\text{Measure}_{i,t-2} + \mathbf{c}_{2,t}\mathbf{x}_{ij,t-2} + \epsilon_{ijt}, \tag{23}$$

where v is the dependent variable of interest, *Measure* represents each of our measures of either concentration or markup, and $\mathbf{x}$ represents a vector of control variables.

The regression analyses are conducted both at the industry level and at the firm level. The main difference between the industry-level and firm-level analyses is that the latter reflects the distribution of the number of firms per industry of the Compustat sample, and not that of the economy as a whole.[29] For both industry-level and firm-level analyses, we include specifications with and without controlling for sample selection.

Each of the four alternative specifications provides inferences about different aspects of the relation between the dependent variable v and the measure of imperfect competition *Measure*. The specification more directly related to our model is that at the industry level with correction for sample selection. This specification is useful to make inferences about the relation between competition and the variable of interest for an average firm in the industry (i.e., considering both public and private firms). The firm-level analyses that do include the sample selection correction refer to the relation between competition and the variable of interest for the average firm (i.e., private or public), yet placing more weight on firms in less competitive industries as in the Compustat sample. We also include specifications that do not adjust for sample selection bias for completeness, although they are not fully aligned with our testable hypotheses. These analyses refer to the relation between competition *Measure* and the variable v only for the subsample of *public firms*, and not to the entire universe of firms in the economy.

Table 4 shows a positive relation between realized stock returns and the measures of concentration and markup, both at the industry level and at the firm level. We include four control variables in all our tests in Table 4. We include market size and book-to-market ratio as additional variables that control for other drivers of expected returns unrelated to our story. We also include the logarithm of the book value of assets and book leverage to control for the mismatch between our model, which is based on asset values, and stock returns, which represent equity returns. The remaining two variables, employment growth and average wage, are included to control for differences across industries in terms of either life cycle or cost structure.[30]

Table 4 shows that, in specifications that do not control for sample selection, and are thus more representative of a public firm in the industry, a one cross-sectional standard deviation in the measures of concentration or markup is associated with industry-level annual excess stock returns between 1.17% to 2.73% (panel A) and with firm-level annual excess stock returns between 2.17%

---

[29] As shown in Table 1, industries with lower concentration and lower markup generally have fewer publicly listed firms despite the fact that they have more firms in total. This finding suggests that the firm-level analyses place greater weight in observations from less competitive industries.

[30] We provide details about the construction of these variables in Appendix C.

Table 4
Annual excess stock returns and measures of concentration and markup

A. Industry-level regressions

		Measure							
		HHI		MKP		CBC		MKC	
Specification	I.a	I.b	II	I.b	II	I.b	II	I.b	II
Dep. variable	ret	ϵ^{ret}	ret	ϵ^{ret}	ret	ϵ^{ret}	ret	ϵ^{ret}	ret
Measure_{t-2}		1.73**	1.17	1.93***	2.25***	1.71***	2.73***	1.13***	1.45
		(0.66)	(0.99)	(0.51)	(0.66)	(0.53)	(0.90)	(0.37)	(1.08)
Log size_{t-2}	−14.23		−1.93		4.07		−12.52		−8.17
	(8.72)		(11.71)		(9.72)		(8.38)		(5.10)
Log B/M_{t-2}	−8.28		0.07		6.42		−5.63		−4.96
	(8.85)		(11.17)		(9.44)		(8.02)		(4.64)
Log asset_{t-2}	13.72		0.78		−4.85		11.71		7.00
	(9.08)		(11.32)		(9.44)		(8.56)		(5.64)
Leverage_{t-2}	−28.26		5.13		22.77		−19.42		−3.98
	(26.94)		(34.24)		(29.31)		(24.58)		(17.91)
Emp G_{t-2}	8.37		−76.71**		−80.92**		11.77		−5.34
	(34.79)		(34.79)		(31.04)		(39.23)		(43.34)
Wage_{t-2}	0.12		0.18*		0.18*		0.07		0.11
	(0.08)		(0.09)		(0.10)		(0.05)		(0.07)
λ_{t-2}	−2.16**								
	(0.80)								
R-sq.	0.10	0.01	0.12	0.02	0.13	0.01	0.10	0.01	0.09
Obs.	3,661	2,025	2,367	2,100	2,447	3,661	3,661	3,658	4,315

to 4.14% (panel B). In specifications that control for sample selection, and are thus more representative of a general firms in the industry, a one cross-sectional standard deviation in concentration or markup is associated with industry-level annual excess stock returns between 1.13% to 1.93% (panel A) and with firm-level annual excess stock returns between 1.35% to 2.15% (panel B).

Table 4 further provides evidence that the sample selection due to firms' decision to be publicly listed can materially affect the inference of asset pricing studies such as ours. Qualitatively, the results presented in Table 4 are remarkably consistent across the specifications with or without the sample selection correction. Quantitatively, however, we observe that the coefficients are generally lower with the sample selection correction. This suggests that the sample selection bias of public listing biases the coefficients upwards. To understand why, note that the negative and significant coefficients on the inverse Mills ratios imply that, on average, private firms are riskier than public firms. Also note that, as seen in Table 1, firms in less competitive industries are more likely to be publicly listed. Taken together, these two facts explain the upward bias in the estimates.

To summarize, the empirical tests shown in Tables 3 and 4 provide empirical support for Hypothesis 1: firms in less competitive industries are generally more exposed to systematic risk. Moreover, firms in less competitive industries earn expected returns that are economically and in most cases statistically significantly higher than their counterparts in more competitive industries. Alternatively, our results can also be interpreted as strong evidence that the

Table 4
Continued

B. Firm-level regressions

		Measure							
		HHI		MKP		CBC		MKC	
Specification	I.a	I.b	II	I.b	II	I.b	II	I.b	II
Dep. variable	ret	ϵ^{ret}	ret	ϵ^{ret}	Ret	ϵ^{ret}	Ret	ϵ^{ret}	Ret
Measure$_{t-2}$		2.02*	2.17*	2.15	3.87**	1.35**	4.14***	1.86**	3.55**
		(1.14)	(1.15)	(1.79)	(1.79)	(0.50)	(1.35)	(0.75)	(1.36)
Log size$_{t-2}$	−5.86***		−4.92		−2.91		−5.95***		−6.14***
	(1.89)		(3.01)		(2.44)		(1.75)		(1.74)
Log B/M$_{t-2}$	0.30		0.99		3.84**		0.38		0.17
	(1.90)		(2.35)		(1.72)		(1.67)		(1.81)
Log asset$_{t-2}$	4.59**		3.24		1.50		4.65**		4.84**
	(2.06)		(2.70)		(2.15)		(1.90)		(1.90)
Leverage$_{t-2}$	−2.48		0.38		7.38		−1.86		−1.48
	(5.74)		(9.82)		(8.28)		(5.17)		(4.96)
Emp G$_{t-2}$	58.69		23.34		26.31		30.20		23.18
	(72.48)		(74.11)		(64.19)		(61.84)		(53.61)
Wage$_{t-2}$	0.06		0.22*		0.20*		0.06		0.14
	(0.06)		(0.11)		(0.10)		(0.06)		(0.08)
λ_{t-2}	−6.58***								
	(1.95)								
R-sq.	0.04	0.00	0.05	0.01	0.05	0.00	0.04	0.00	0.05
Obs.	37,164	19,672	20,317	19,915	20,568	37,164	37,164	37,149	38,619

Newey-West standard errors are shown in parentheses. *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

The table shows estimates and standard errors of Fama-MacBeth regressions based on the two alternative specifications. Specification I, which controls for the sample-selection bias of public firms, is given by:

$$\text{Ret}_{\text{ijt}} = c_{0,\text{t}}^{\text{I.a}} + \mathbf{c}_{2,\text{t}}^{\text{I.a}} \mathbf{Controls}_{\text{ij,t}-2} + c_{3,\text{t}}^{\text{I.a}} \lambda_{\text{i,t}-2} + \epsilon_{\text{ijt}}^{\text{Ret}}, \tag{I.a}$$

$$\epsilon_{\text{ijt}}^{\text{Ret}} = c_{0,\text{t}}^{\text{I.b}} + c_{1}^{\text{I.b}} \text{Measure}_{\text{i,t}-2} + \tilde{\epsilon}_{\text{ijt}}, \tag{I.b}$$

where *Ret* is annual excess stock returns (of equally-weighted industry portfolios in panel A and of individual stocks in panel B), *Measure* denotes the measures of concentration (*HHI* and *CBC*) and markup (*MKP* and *MKC*), described in Table 1, and λ is the inverse Mills ratio. Specification II, which does not control for sample-selection bias, is given by:

$$\text{Ret}_{\text{ijt}} = c_{0,\text{t}}^{\text{II}} + c_{1,\text{t}}^{\text{II}} \text{Measure}_{\text{i,t}-2} + \mathbf{c}_{2,\text{t}}^{\text{II}} \mathbf{Controls}_{\text{ij,t}-2} + \epsilon_{\text{ijt}}. \tag{II}$$

The control variables used in both specifications are the logarithm of the market value of equity *Log size*, the logarithm of the book-to-market ratio *Log B/M*, the logarithm of the book value of assets (*Log asset*), the leverage ratio (*Leverage*), the employment growth in the industry, and the average wage level in the industry from the NIPA tables. Variables in panel A are industry averages. The sample period is 1994–2014.

relation between expected returns and the measures of concentration and markup is *not negative*. Our findings are thus in stark contrast with those by Hou and Robinson (2006), who find that firms in less concentrated industries are riskier using Compustat-based measures of concentration. Furthermore, when taken together, the evidence presented in Tables 1 and 4 suggests that the difference between our results and those in Hou and Robinson (2006) can be explained, at least partly, by the sample selection bias of public listing.

3.2.2 Expected returns, concentration, and markup: Conditional tests. The evidence presented so far is consistent with Hypothesis 1, focusing on the unconditional relation between competition and expected returns. In this section, we take a step further and investigate whether the observed relation between expected returns and concentration and markup is in fact consistent with our proposed economic channels. In particular, Hypothesis 2 predicts that the relation between our measures of expected returns and concentration and markup is relatively more positive in industries with relatively lower levels of operating leverage. Because Hypothesis 2 does not rely on the distribution of industries relative to an unobservable operating level threshold, it provides a complementary and more direct way to test the existence of these channels than 1.

We test the conditional prediction in Hypothesis 2 by double-sorting firms first into terciles (labeled "Bottom," "Middle," and "Top") of the industry average operating leverage, and then into terciles (labeled L, 2, and H) of concentration or markup. We calculate for each of the nine portfolios using averages of the eight measures of expected returns described in Table 3. We then estimate the differences in average expected returns between portfolios H and L in the bottom operating leverage tercile ($\mathrm{R}^{\mathrm{Bottom}}_{\mathrm{H\text{-}L}}$) and in the top portfolio ($\mathrm{R}^{\mathrm{Top}}_{\mathrm{H\text{-}L}}$). The prediction in Hypothesis 2 implies that, for each of our measures of concentration and markup, we should observe

$$\mathrm{R}^{\mathrm{Bottom}}_{\mathrm{H\text{-}L}} - \mathrm{R}^{\mathrm{Top}}_{\mathrm{H\text{-}L}} > 0. \tag{24}$$

Hypothesis 1 implies that the expected return spread over the entire sample $\mathrm{R}_{\mathrm{H\text{-}L}}$ is positive, a prediction confirmed by the results shown in Table 3.[31] The conditional prediction in Inequality (24) is that the expected return spread between less competitive and more competitive industries is *relatively more positive* among firms in industries with low average operating leverage ($\mathrm{R}^{\mathrm{Bottom}}_{\mathrm{H\text{-}L}}$) than among firms in industries with high average operating leverage ($\mathrm{R}^{\mathrm{Top}}_{\mathrm{H\text{-}L}}$), such that $\mathrm{R}^{\mathrm{Bottom}}_{\mathrm{H\text{-}L}} - \mathrm{R}^{\mathrm{Top}}_{\mathrm{H\text{-}L}} > 0$.

Table 5 shows empirical evidence that supports the prediction in Inequality (24). Panel A provides the corresponding evidence for the expected return spreads using the proxies for expected returns based on realized returns presented in panel A of Table 3: excess and unlevered stock returns, as well as two proxies based on stock returns adjusted for firm characteristics. In all cases but one, we observe that the return spread from the subsample (A) of low operating leverage industries is significantly positive, ranging from 2.36% to 13.50%. In contrast we observe that the return spread from the subsample (B) of high operating leverage industries ranges from −1.73% to 6.47%, and it is statistically insignificant in all but one case. The last rows of the panel,

[31] Note that Table 3 sorts firms into portfolios using quintiles of our measures of imperfect competition, while Table 5 focuses on portfolios using terciles. We group observations in terciles as opposed to quintiles in order to have a larger number of observations in each portfolio while considering double sorts.

Table 5
Cross-section of expected returns from portfolio sorts conditional on average operating leverage

A. Proxies for expected returns based on realized returns

	HHI		MKP		CBC		MKC	
OLtercile	$R^{E}_{H\text{-}L}$	$R^{E,U}_{H\text{-}L}$	$R^{E}_{H\text{-}L}$	$R^{E,U}_{H\text{-}L}$	$R^{E}_{H\text{-}L}$	$R^{E,U}_{H\text{-}L}$	$R^{E}_{H\text{-}L}$	$R^{E,U}_{H\text{-}L}$
Bottom (A)	8.72**	5.87**	9.71**	7.87***	2.36	5.74***	6.73*	7.64***
	(3.34)	(2.34)	(4.18)	(2.39)	(2.29)	(1.65)	(3.48)	(2.16)
Top (B)	5.84	4.61	−1.73	0.56	1.54	1.68	0.46	1.15*
	(6.13)	(3.56)	(2.79)	(1.00)	(1.44)	(1.13)	(0.77)	(0.64)
(A)-(B)	2.88	1.26	11.44***	7.31***	0.82	4.06***	6.27*	6.49***
	(5.56)	(4.16)	(2.21)	(1.78)	(2.24)	(1.34)	(3.45)	(2.11)
OLtercile	$R^{DGTW}_{H\text{-}L}$	$R^{ADJ}_{H\text{-}L}$	$R^{DGTW}_{H\text{-}L}$	$R^{ADJ}_{H\text{-}L}$	$R^{DGTW}_{H\text{-}L}$	$R^{ADJ}_{H\text{-}L}$	$R^{DGTW}_{H\text{-}L}$	$R^{ADJ}_{H\text{-}L}$
Bottom (A)	9.08**	11.07***	9.87**	12.10***	4.06**	5.26***	6.85***	8.06***
	(3.30)	(3.81)	(4.29)	(4.25)	(1.87)	(1.59)	(1.99)	(1.86)
Top (B)	5.38	6.47	−1.30	−1.40	0.51	0.50	−0.65	−0.66
	(4.09)	(4.27)	(1.50)	(1.49)	(1.10)	(1.29)	(0.52)	(0.65)
(A)-(B)	3.70	4.60	11.17***	13.50***	3.55**	4.76***	7.50***	8.72***
	(2.86)	(3.07)	(3.29)	(3.23)	(1.57)	(1.50)	(2.12)	(2.05)

B. Proxies for expected returns based on factor models

	HHI		MKP		CBC		MKC	
OLtercile	$R^{CPM}_{H\text{-}L}$	$R^{CPM,U}_{H\text{-}L}$	$R^{CPM}_{H\text{-}L}$	$R^{CPM,U}_{H\text{-}L}$	$R^{CPM}_{H\text{-}L}$	$R^{CPM,U}_{H\text{-}L}$	$R^{CPM}_{H\text{-}L}$	$R^{CPM,U}_{H\text{-}L}$
Bottom (A)	2.00***	1.79***	1.58*	2.79***	1.53***	3.50***	1.97	4.13***
	(0.57)	(0.56)	(0.86)	(0.78)	(0.47)	(0.35)	(1.28)	(0.89)
Top (B)	0.47***	0.60***	−0.34	0.47	1.00***	1.03***	0.15	1.35***
	(0.15)	(0.21)	(0.42)	(0.32)	(0.06)	(0.20)	(0.17)	(0.19)
(A)-(B)	1.53**	1.20***	1.92***	2.32***	0.53	2.47***	1.82	2.78***
	(0.56)	(0.40)	(0.55)	(0.51)	(0.45)	(0.18)	(1.20)	(0.72)
OLtercile	$R^{FF3}_{H\text{-}L}$	$R^{FC4}_{H\text{-}L}$	$R^{FF3}_{H\text{-}L}$	$R^{FC4}_{H\text{-}L}$	$R^{FF3}_{H\text{-}L}$	$R^{FC4}_{H\text{-}L}$	$R^{FF3}_{H\text{-}L}$	$R^{FC4}_{H\text{-}L}$
Bottom (A)	2.69***	2.56***	2.43***	2.46***	1.83***	2.12***	2.13***	2.04***
	(0.47)	(0.49)	(0.39)	(0.47)	(0.22)	(0.29)	(0.50)	(0.58)
Top (B)	0.38**	0.36**	−0.34	−0.31*	0.83***	0.61**	0.08	−0.26*
	(0.15)	(0.16)	(0.20)	(0.15)	(0.17)	(0.26)	(0.15)	(0.13)
(A)-(B)	2.31***	2.20***	2.77***	2.78***	1.00**	1.51***	2.05***	2.29***
	(0.58)	(0.58)	(0.31)	(0.39)	(0.38)	(0.37)	(0.51)	(0.60)

E is excess stock returns, E,U is unlevered excess stock returns, *DGTW* is excess stock returns adjusted for B/M, size, and past returns constructed as in Daniel et al. (1997), ADJ is excess stock returns adjusted for B/M, size, and profitability, *CPM* is expected excess returns from CAPM, *CPM*, U is unlevered expected excess returns from CAPM, *FF3* is expected excess returns from Fama and French's (1993) three-factor model, and $FC4$ is expected excess returns from the Fama-French-Carhart four-factor model. Newey-West standard errors are shown in parentheses. *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.
The table below presents the average measure of expected return of the H-L zero net investment portfolio long the top (H) tercile portfolio of firms and short (L) the bottom tercile portfolio of firms by concentration (*HHI* and *CBC*) or markup (*MKP* and *MKC*) formed in the subsamples firms in the bottom tercile (A) and top tercile (B) of industry average operating leverage (OL). OL is constructed as in Novy-Marx (2011). (A)-(B) denotes the difference in the estimates of the return of the H-L portfolio across the two subsamples. The sample period is 1994–2014.

labeled (A)-(B), show the difference in return spreads from the low (A) and high (B) operating leverage subsamples is always positive and in most cases also statistically significant, ranging between 0.82% to 13.50%.

In parallel with the unconditional portfolio sorts presented in Table 3, panel B in Table 5 considers proxies of expected returns implied by factor models. The evidence in this panel provides strong support for the prediction of Hypothesis 2 that the return spread of firms in industries with low operating leverage is significantly more positive than the corresponding spread for industries with high operating leverage. The last rows of panel B, denoted by (A)-(B), also show, in most cases, significantly positive differentials in across specifications. In particular, the expected return spread in terms of expected returns for FF3 and CF4 models in industries with lower operating leverage is between 1.00% and 2.77% higher than the corresponding expected return spread in industries with higher operating leverage.

3.3 Further evidence for the economic channels

Both the unconditional and conditional implications for the relation between expected returns, concentration, and markup formalized by Hypotheses 1 and 2 rely on the existence of the entry threat and operating leverage channels. To add further support that our findings in Tables 3–5 are consistent with the existence of these channels, we investigate in this section whether earnings-to-price ratios and operating leverage are in fact decreasing in concentration and markup as predicted by Hypothesis 3.

3.3.1 Earnings-to-price ratios. The first part of Hypothesis 3 predicts a negative relation between earnings-to-price ratios, book-to-market ratios, and either concentration or markup. We test this prediction using the same two alternative regression specifications discussed in Table 3 for expected returns. The controls used in the regressions are the same as those used in Table 4, with the exception of market size and book-to-market ratio, since these variables would be mechanically related to the dependent variables in these tests.

Table 6 reports the results with earnings-to-price ratios. For each measure of concentration and markup, the first specification controls for the sample selection bias of public listing, while the second relies on standard Fama–MacBeth regressions. Table 6 shows that earnings-to-price ratios are negatively related to concentration and markup. All specifications show negative and significant coefficients with respect to all measures of concentration and markup, both at the industry level and at the firm level.

In addition, we observe in Table 6 that the inverse Mills ratio in the first stage is positive and significant, both at the firm level and at the industry level. This suggests that firms in industries with a larger fraction of public firms tend to have significantly lower average earnings-to-price ratios. Intuitively, this explains why the second-stage coefficients on the measures of concentration and markup are, in general, lower when we control for the sample selection bias of public listing. Moreover, given that our sample selection correction operates under the working assumption that the likelihood of being public is solely driven by industry-specific characteristics, the finding that the inverse Mills ratios are positive and significant both at the firm level and industry level suggests that a significant industry component explains the likelihood of being public.

Table 6
Earnings-to-price ratios and measures of concentration and markup

		Measure							
		HHI		MKP		CBC		MKC	
Specification	I.a	I.b	II	I.b	II	I.b	II	I.b	II
Dep. variable	E/P	$\epsilon^{E/P}$	E/P	$\epsilon^{E/P}$	E/P	$\epsilon^{E/P}$	E/P	$\epsilon^{E/P}$	E/P
A. Industry-level regressions									
Measure$_{t-2}$		−0.03**	−0.05***	−0.09***	−0.09***	−0.06***	−0.11***	−0.05***	−0.06***
		(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
Log asset$_{t-2}$	0.09***		0.13***		0.09***		0.09***		0.07***
	(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
Leverage$_{t-2}$	1.27***		1.17***		1.13***		1.19***		1.35***
	(0.17)		(0.25)		(0.22)		(0.14)		(0.15)
Emp G$_{t-2}$	−0.85		0.17		0.30		−0.39		−0.35
	(0.69)		(0.34)		(0.38)		(0.61)		(0.45)
Wage$_{t-2}$	0.00**		−0.01***		−0.01***		0.00***		−0.01***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
λ_{t-2}	0.19***								
	(0.02)								
R-sq.	0.27	0.02	0.32	0.05	0.31	0.02	0.28	0.02	0.25
Obs.	3,596	1,990	2,304	2,064	2,383	3,596	3,596	3,594	4,192
B. Firm-level regressions									
Measure$_{t-2}$		−0.03**	−0.02**	−0.14***	−0.18***	−0.09***	−0.22***	−0.14***	−0.23***
		(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Log asset$_{t-2}$	0.06***		0.07***		0.06***		0.06***		0.05***
	(0.01)		(0.01)		(0.01)		(0.01)		(0.02)
Leverage$_{t-2}$	1.40***		1.33***		1.15***		1.37***		1.28***
	(0.16)		(0.18)		(0.18)		(0.18)		(0.15)
Emp G$_{t-2}$	−1.17		0.74		0.54		0.29		0.14
	(1.40)		(0.71)		(0.58)		(1.20)		(0.91)
Wage$_{t-2}$	0.00		−0.01***		−0.01***		0.00		0.00***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
λ_{t-2}	0.36***								
	(0.04)								
R-sq.	0.23	0.01	0.22	0.03	0.25	0.01	0.24	0.03	0.25
Obs.	28,175	13,864	14,433	14,085	14,659	28,175	28,175	28,168	29,382

Newey-West standard errors are shown in parentheses. *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.
The table shows estimates and standard errors of Fama-MacBeth regressions based on the two alternative specifications. Specification I, which controls for the sample-selection bias of public firms, is given by:

$$E/P_{ijt} = c^{I.a}_{0,t} + \mathbf{c}^{I.a}_{2,t}\mathbf{Controls}_{ij,t-2} + c^{I.a}_{3,t}\lambda_{i,t-2} + \epsilon^{E/P}_{ijt}, \tag{I.a}$$

$$\epsilon^{E/P}_{ijt} = c^{I.b}_{0,t} + c^{I.b}_{1}\text{Measure}_{i,t-2} + \tilde{\epsilon}_{ijt}, \tag{I.b}$$

where *E/P* is the earning-to-price ratio, *Measure* denotes the measures of concentration (*HHI* and *CBC*) and markup (*MKP* and *MKC*), described in Table 1, ***Controls*** is a vector of controls variables, and λ is the inverse Mills ratio. Specification II, which does not control for sample-selection bias, is given by:

$$E/P_{ijt} = c^{II}_{0,t} + c^{II}_{1,t}\text{Measure}_{i,t-2} + \mathbf{c}^{II}_{2,t}\mathbf{Controls}_{ij,t-2} + \epsilon_{ijt}. \tag{II}$$

The control variables used in both specifications are the logarithm of the book value of assets (*Log asset*), the leverage ratio (*Leverage*), the employment growth in the industry, and the average wage level in the industry from the NIPA tables. Variables in panel A are industry averages. The sample period is 1994–2014.

Taken together, the findings of Table 6 are consistent with Hypothesis 3. On average, firms in industries with higher markup or higher concentration better preserve their value and thus have lower earnings-to-price ratios. The analysis

Table 7
Operating leverage and measures of concentration and markup

		Measure							
		HHI		MKP		CBC		MKC	
Specification	I.a	I.b	II	I.b	II	I.b	II	I.b	II
Dep. variable	OL	ϵ^{OL}	OL	ϵ^{OL}	OL	ϵ^{OL}	OL	ϵ^{OL}	OL
A. Industry-level regressions									
Measure_{t-2}		−0.05**	−0.01	−0.35***	−0.18***	−0.09***	−0.11***	−0.33***	−0.29***
		(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Log asset_{t-2}	−0.09***		−0.08***		−0.08***		−0.09***		−0.10***
	(0.02)		(0.01)		(0.01)		(0.02)		(0.01)
Leverage_{t-2}	0.83***		0.67***		0.42***		0.79***		0.20*
	(0.24)		(0.08)		(0.08)		(0.25)		(0.10)
Emp G_{t-2}	−0.45**		0.24		0.35		0.35**		0.31
	(0.20)		(0.61)		(0.50)		(0.15)		(0.33)
Wage_{t-2}	−0.01***		0.00***		0.00***		−0.01***		−0.01***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
λ_{t-2}	0.25***								
	(0.02)								
R-sq.	0.22	0.01	0.13	0.18	0.29	0.01	0.21	0.16	0.34
Obs.	3,626	2,019	2,348	2,094	2,428	3,626	3,626	3,623	4,250
B. Firm-level regressions									
Measure_{t-2}		0.00	0.00	−0.21***	−0.18***	−0.07***	−0.14***	−0.19***	−0.22***
		(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)	(0.01)
Log asset_{t-2}	−0.08***		−0.08***		−0.08***		−0.08***		−0.09***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Leverage_{t-2}	0.64***		0.79***		0.63***		0.62***		0.50***
	(0.06)		(0.05)		(0.04)		(0.07)		(0.05)
Emp G_{t-2}	−0.18		0.37		0.19		0.76**		0.36
	(0.47)		(0.64)		(0.33)		(0.34)		(0.52)
Wage_{t-2}	−0.01***		0.00***		0.00***		−0.01***		−0.01***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
λ_{t-2}	0.25***								
	(0.01)								
R-sq.	0.28	0.00	0.14	0.06	0.24	0.01	0.28	0.05	0.33
Obs.	30,953	16,382	16,996	16,605	17,225	30,953	30,953	30,938	32,240

Newey-West standard errors are shown in parentheses. *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

The table shows estimates and standard errors of Fama-MacBeth regressions based on the two alternative specifications. Specification I, which controls for the sample-selection bias of public firms, is given by:

$$\mathrm{OL}_{\mathrm{ijt}} = c^{\mathrm{I.a}}_{0,\mathrm{t}} + \mathbf{c}^{\mathrm{I.a}}_{2,\mathrm{t}}\mathbf{Controls}_{\mathrm{ij,t-2}} + c^{\mathrm{I.a}}_{3,\mathrm{t}}\lambda_{\mathrm{i,t-2}} + \epsilon^{\mathrm{OL}}_{\mathrm{ijt}}, \tag{I.a}$$

$$\epsilon^{\mathrm{OL}}_{\mathrm{ijt}} = c^{\mathrm{I.b}}_{0,\mathrm{t}} + c^{\mathrm{I.b}}_{1}\mathrm{Measure}_{\mathrm{i,t-2}} + \tilde{\epsilon}_{\mathrm{ijt}}, \tag{I.b}$$

where *OL* is the measure of operating leverage from Novy-Marx (2011), *Measure* denotes the measures of concentration (*HHI* and *CBC*) and markup (*MKP* and *MKC*), described in Table 1, ***Controls*** is a vector of controls variables, and λ is the inverse Mills ratio. Specification II, which does not control for sample-selection bias, is given by:

$$\mathrm{OL}_{\mathrm{ijt}} = c^{\mathrm{II}}_{0,\mathrm{t}} + c^{\mathrm{II}}_{1,\mathrm{t}}\mathrm{Measure}_{\mathrm{i,t-2}} + \mathbf{c}^{\mathrm{II}}_{2,\mathrm{t}}\mathbf{Controls}_{\mathrm{ij,t-2}} + \epsilon_{\mathrm{ijt}}. \tag{II}$$

The control variables used in both specifications are the logarithm of the book value of assets (*Log asset*), the leverage ratio (*Leverage*), the employment growth in the industry, and the average wage level in the industry from the NIPA tables. Variables in panel A are industry averages. The sample period is 1994–2014 also shows that relative valuation ratios are significantly different between private and public firms.[32]

3.3.2 Operating leverage Hypothesis 3 predicts a negative relation between operating leverage and concentration and markup. In order to investigate this hypothesis, we use Novy-Marx's (2011) Compustat-based measure of operating leverage. This measure is defined as the ratio of the sum of a firm's selling expenses, general expenses, administrative expenses, and cost of goods sold to total assets. Intuitively, the measure proxies for the magnitude of fixed operating costs relative to the size of the firm.

Table 7 reports the regression analysis of operating leverage on our measures of concentration and markup. Consistent with our previous analyses, we use two alternative specifications for each measure of concentration and markup: one that controls for the sample selection bias of public listing, and one specification that does not. Consistent with the summary statistics, the tests in Table 7 show that operating leverage is generally higher among firms in industries with low markup or low concentration. This prediction holds in all specifications that consider markup measures and our CBC measure; results are weaker for the HHI measure, both at the industry level and at the firm level. The correction for sample selection is always significant, and its positive sign implies that private firms have higher operating leverage on average.

As a final remark, we are aware that Novy-Marx's measure of operating leverage is subject to the concern that a firm's selling expenses, general expenses, administrative expenses, and cost of goods sold include both fixed and variable operating costs. In the Online Appendix, we provide additional evidence that operating leverage is higher in more competitive industries using an alternative measure of operating leverage.[33] While this alternative measure is conceptually similar to our theoretical measure of operating leverage, its coverage is entirely cross-sectional and limited to manufacturing industries.

4. Conclusion

This paper contributes to the existing literature by showing, both theoretically and empirically, that the degree of competition in an industry has a nontrivial effect on expected returns. We construct a model through which we characterize three alternative economic channels that explain the relation between product market competition and expected returns. We then provide empirical evidence consistent with the theoretical predictions that (1) competition erodes markups,

[32] In the accompanying Online Appendix, we show that the results of tests using book-to-market ratios are very similar to the ones presented in Table 6.

[33] In particular, we construct an industry-level measure based on the sensitivity of operating profits to productivity shocks. To construct this measure, we consider time-series regressions of value-added growth on total factor productivity growth by industry, using the NBER–CES data for manufacturing industries.

such that firms are more exposed to systematic risk; (2) the threat of entry by new firms lowers incumbents' exposure to systematic risk; and (3) higher industry aggregate risk represents a barrier to entry.

As an additional contribution, our empirical findings offer an explanation for the discrepancy regarding the significant negative relation between concentration and expected returns documented by Hou and Robinson (2006) (using Compustat-based measures of concentration) and the lack of significance for this same relation documented by Ali, Klasa, and Yeung (2009) (using Census-based measures of concentration). We demonstrate a sample selection bias related to firms' decision to be publicly listed, which arises since firms are more likely to be public in less competitive industries. This selection bias leads to a weak relation between Compustat-based and Census-based measures of concentration. Using a sample selection correction, we propose a new measure of concentration that offers broad coverage in all industries in Compustat, without the sample selection bias inherent in Compustat-based measures.

More broadly, the results in this paper suggest that the industrial organization plays an important role for financial economics by affecting firms' relative valuations and expected asset returns. The industrial organization covers multiple aspects that go beyond the degree of product market competition explored in our study. Our empirical strategy also implies that empirical asset pricing papers that explore the link between firms' exposure to systematic risk and industry characteristics should control for the sample selection bias that arises if such industry characteristics are related to firms' decision to be publicly listed. We hope that these broad insights motivate future research on how industry characteristics interact with asset prices.

Appendix A: Model Solution

A.1 Cournot Equilibrium

We obtain the optimized operating profits of each firm by considering their optimal strategy in the symmetric Nash equilibrium of the Cournot competition. In an equilibrium with a level of concentration $h=1/N$, each firm produces the quantity

$$y_t = \frac{X_t}{\alpha}(1-c)\left(\frac{h_t}{h_t+1}\right). \tag{A1}$$

The equilibrium price of the industry is given by

$$\mathcal{P}_t = \frac{X_t(c+h_t)}{(1+h_t)}. \tag{A2}$$

Consequently, the optimal profit function for each firm is given by

$$\Pi_t = \frac{X_t^2}{\alpha}(1-c)^2\left(\frac{h_t}{h_t+1}\right)^2 - f. \tag{A3}$$

We can re-express the operating profits of the firm in terms of profit margin p as given in Equation (6). The profit margin p is a function of the concentration h and the industry shocks X as shown in Equation (7). Equations (A2), (A3), and (7) imply that the price of the homogeneous good ($\mathcal{P}$), the operating profits, and the profit margin (p) drop when concentration (h) decreases after a new firm enters the industry.

A.2 Value of the Firm

We term the values of an incumbent and a prospective firm by V and V_E, respectively. We assume there is competition among prospective entrants, so the value of a prospective firm is always zero (i.e., $V_E[h,p]=0$) for any levels of concentration $0<h<1$ and profit margin $p>0$. The value of an incumbent firm is a function of two state variables: concentration h and profit margin p. The variable h varies discretely each time a new firm enters the industry. A change in h does not affect the instantaneous return of the firm; rather, it affects the value matching condition. The variable p varies continuously due to the Brownian motion embedded in shocks. Consequently, a change in p affects the instantaneous drift and diffusion of firm value.

Applying the Martingale condition such that the drift of the gains process equals zero, we obtain the following ordinary differential equation (ODE):

$$\Pi[p]-rV[h,p]+(\mu-\eta\rho\sigma)pV_p[h,p]+\frac{1}{2}\sigma^2p^2V_{pp}[h,p]=0. \tag{A4}$$

The general solution to the ODE in the equation above is known and given by

$$V[h,p]=\frac{\Pi[p]+f}{\delta}-\frac{f}{r}+C_1[h]p^{\frac{\delta-r}{\sigma^2}+2-\upsilon}+C_2[h]p^{\upsilon}, \tag{A5}$$

where $\delta\equiv r+2\left(\eta\rho\sigma-\frac{\sigma^2}{2}-\mu\right)$ and $\upsilon\equiv\frac{\eta\rho\sigma-\mu+\sqrt{\left(\eta\rho\sigma+\frac{\sigma^2}{2}-\mu\right)^2+2r\sigma^2}}{\sigma^2}+\frac{1}{2}>2$. We focus on the parameter space that assures a positive risk premium (i.e., $\delta>r$).

A.3 The Case of Very Low Concentration ($h\to 0$)

This case is equivalent to the case in which the number of incumbents is low in the model by Pindyck (2009). In what follows, we use the informal notation $V[0,p]$ to represent the values of an incumbent firm when a very large number of incumbents already exists, so that concentration is arbitrarily low, $h\to 0$. $\overline{p}[0]$ denotes the boundary, in terms of profit margin, at which a marginal prospective entrant is indifferent between entering and not entering an industry with very low concentration. The boundary conditions are

$$V_E[0,\overline{p}[0]]=V[0,\overline{p}[0]]-\kappa=0, \text{ and} \tag{A6}$$

$$\left.\frac{\partial V[0,p]}{\partial p}\right|_{p=\overline{p}[0]}=0. \tag{A7}$$

The value matching condition (A6) states that, at entry, the value of the option to enter must be equal to the value of an incumbent firm minus the cost of entry κ. Consistent with Pindyck (2009), the condition (A7) ensures no arbitrage at the entry threshold $\overline{p}[0]$.

The specific solution of Equation (A5) for the case $h\to 0$ is given by

$$V[0,p]\equiv\frac{p^2}{\alpha\delta}-\frac{f}{r}-\frac{2\overline{p}[0]^2}{\upsilon\alpha\delta}\left(\frac{p_t}{\overline{p}[0]}\right)^{\upsilon}, \tag{A8}$$

where

$$\overline{p}[0]=\sqrt{\alpha\delta\left(\frac{f}{r}+\kappa\right)\left(\frac{\upsilon}{\upsilon-2}\right)}. \tag{A9}$$

The solution to the case of very low concentration is useful for deriving the general case discussed next.

A.4 General Case

We now derive the particular solution to Equation (A5) for the more general case in which concentration is not very low. This case is equivalent to the case in which the number of incumbents is very high in the model by Pindyck (2009). The main difference between this case and the case with very low concentration is that here a prospective entrant must take into account the impact its own entry decision will have on the industry. For instance, a simple inspection of Equations (5) and (7) shows that the entry of n new firms into an industry with current levels of concentration and profit margin of h and p will drive these to $\frac{h}{1+nh}$ and $\left(\frac{1+h}{1+(n+1)h}\right)p$, respectively. The non-negligible effect of entry on the industry implies that the entry boundary $\overline{p}[h]$ is no longer fixed as in Equation (A9).

The endogenous entry threshold $\overline{p}[h]$ is determined by the level of profit margin at which a marginal prospective entrant is indifferent between entering and not entering the industry, such that

$$V_{\mathrm{E}}[h,\overline{p}[h]]=V\left[\tfrac{h}{1+h},\overline{p}[h]\left(\tfrac{1+h}{1+2h}\right)\right]-\kappa=0,\text{ and} \tag{A10}$$

$$\left.\frac{\partial V[h,p]}{\partial p}\right|_{p=\overline{p}[h]}=0. \tag{A11}$$

The value matching condition in Equation (A10) takes into account that the levels of concentration and profit margin are reduced when a new firm enters the industry. For tractability, we follow Pindyck (2009) and assume that the no-arbitrage condition is similar to that of the very low concentration case, as given in Equation (A11). Numerical simulations show that the error from this approximation is consistently small and converges to zero as concentration shrinks.

Although no closed-form solution exists in this case, the solution for firm value can be expressed in terms of the endogenous entry threshold, namely

$$V[h,p]\equiv\frac{p^2}{\alpha\delta}-\frac{f}{r}-\frac{2\overline{p}[h]^2}{\upsilon\alpha\delta}\left(\frac{p_t}{\overline{p}[h]}\right)^{\upsilon}, \tag{A12}$$

where the entry threshold $\overline{p}[h]$ is determined from Equations (A10) and (A11), which imply the following recursive condition:

$$\frac{\overline{p}[\frac{h}{1+nh}]}{\overline{p}[\frac{h}{1+(n+1)h}]}=\left(\frac{1+(n+2)h}{1+(n+3)h}\right)^{\frac{2}{\upsilon-2}}\left(\frac{\upsilon}{2}\left(\frac{1+(n+2)h}{1+nh}\right)^{\upsilon-2}\left(\frac{\overline{p}[\frac{h}{1+(n-1)h}]}{\overline{p}[\frac{h}{1+nh}]}\right)^2-\left(\frac{\overline{p}[\frac{h}{1+(n-1)h}]}{\overline{p}[\frac{h}{1+nh}]}\right)^{\upsilon}\right)^{\frac{1}{\upsilon-2}}, \tag{A13}$$

where n represents the number of new firms that enter an industry with initial concentration h. The thresholds $\overline{p}[\frac{h}{1+nh}]$ can be solved numerically for all $n>0$ by solving the nonlinear difference equation system in Equation (A13). Note that $\overline{p}[0]$ from Equation (A9) represents the limiting case of $\overline{p}[h]$ defined recursively in Equation (A13) as $h\to 0$. In this sense, the solution given in Equation (A12) nests that of the special case of very low concentration given in Equation (A8).

A.5 Definition of Firm Beta

We begin by considering the asset pricing identity

$$\mathrm{E_t}\left[\frac{dV_t}{V_t}\right]\frac{1}{dt}+\frac{\Pi_t}{V_t}-r=-\mathrm{E_t}\left[\frac{dV_t}{V_t}\frac{d\Lambda_t}{\Lambda_t}\right]\frac{1}{dt}=\beta_t\eta, \tag{A14}$$

where we define a firm's beta or systematic risk loading as

$$\beta_t\equiv-\frac{\mathrm{E_t}\left[\frac{dV_t}{V_t}\frac{d\Lambda_t}{\Lambda_t}\right]\frac{1}{dt}}{\eta} \tag{A15}$$

Considering jointly Equations (A14) and (A15), we obtain Equation (13). Reordering terms and using the definition of firm value in Equation (A12), we then obtain an alternative expression for a firm's beta β_t which is given by:

$$\beta[h_t,p_t]=2\gamma\left(\frac{\Pi[p_t]/\delta}{V[h_t,p_t]}\right)+\upsilon\gamma\left(1-\frac{\Pi[p_t]/\delta}{V[h_t,p_t]}\right)+\frac{f}{V[h_t,p_t]}2\gamma\iota, \tag{A16}$$

where the constants γ and ι are defined as

$$\gamma\equiv\rho\sigma>0, \tag{A17}$$

and

$$\iota\equiv 1+\frac{\upsilon}{2}\left(\frac{\delta}{r}-1\right)>1. \tag{A18}$$

Using Equation (A16) and the definitions of operating leverage and earnings-to-price ratios in the body of the paper, we obtain Equation (14).

Appendix B: Derivation of Model Predictions

Proof of Proposition 1 (operating leverage Θ)

Here, we analyze the sensitivity of operating leverage to concentration h and markup m. We start by considering concentration h. The fact that concentration h only changes when there is entry implies that the profit margin p, operating profits Π, and, consequently, operating leverage Θ are only affected by changes in h at the time of entry. Firms enter the industry when profit margin reach a threshold $p=\overline{p}[h]$. Therefore, the sensitivity of operating leverage Θ to concentration h is given by:

$$\left.\frac{\partial\Theta[p[h]]}{\partial h}\right|_{p=\overline{p}[h]}=-\Theta[\overline{p}[h]]\frac{2\overline{p}[h]}{\alpha\Pi[\overline{p}[h]]}\frac{\partial\overline{p}[h]}{\partial h}<0, \tag{B1}$$

where the overall sign of Equation (B1) is negative since the second factor is strictly positive. By applying the implicit function theorem to the boundary condition (A11), it is straightforward to show that the derivative of the entry threshold $\overline{p}'[h]$ is weakly positive.

Consider now the sensitivity of operating leverage to the markup m, which is defined in Equation (9). As a preliminary step, note that at $p=\overline{p}[h]$ the sensitivity of m with respect to h is given by

$$\frac{\partial m[h,\overline{p}[h]]}{\partial h}=\frac{(1-c)\left(2\alpha f h(c+h)\frac{\partial\overline{p}[h]}{\partial h}-\alpha c f\overline{p}[h]+c\overline{p}[h]^3\right)}{(c+h)^2\overline{p}[h]^3}>0, \tag{B2}$$

which is strictly positive given $\frac{\partial\overline{p}[h]}{\partial h}>0$ and $\Pi[p]>0$. Applying the chain rule, we infer that at $p=\overline{p}[h]$ (i.e., at the time of entry), operating leverage Θ is strictly decreasing in the industry markup, as given by

$$\frac{\partial\Theta[\overline{p}[h]]}{\partial m}=\frac{\partial\Theta[\overline{p}[h]]}{\partial h}\left(\frac{\partial m[h,\overline{p}[h]]}{\partial h}\right)^{-1}<0, \tag{B3}$$

where the first factor on the right-hand side of Equation (B3) is strictly negative given Equation (B1), and the second factor, which follows from the inverse function theorem, is strictly positive given Equation (B2).

While Equation (B3) focuses on the instance at which firms enter the industry, operating leverage is decreasing in markup at any point in time, with or without entry of new firms. Using the chain rule, Equation (9) and our previous findings, it is straightforward to show that:

$$\frac{\partial\Theta[p]}{\partial m}=-\left(\frac{2p\Theta[p]}{\alpha\Pi[p]}\right)\left(\frac{(1-c)X}{(1+h)^2}\frac{c(1+h)^2p^2}{(1-c)(f\alpha+p^2)}+\frac{(1-c)h}{(1+h)}\frac{c(1+h)^2p^3}{2(1-c)^2f\alpha h^2}\right)<0 \tag{B4}$$

which is strictly negative for any parameter value.

Proof of Proposition 2 (Earnings-to-price ratio Ξ)

We solve for the partial derivative of the earnings-to-price ratio Ξ with respect to concentration h for a profit margin p such that $\Pi[p]>0$. The corresponding partial derivative with respect to h is given by

$$\frac{\partial\,\Xi[h,p]}{\partial h}=-\frac{2\Pi[p](\upsilon-2)\overline{p}[h]\left(\frac{p}{\overline{p}[h]}\right)^{\upsilon}}{\alpha\delta\upsilon V[h,p]^2}\overline{p}'[h]<0, \tag{B5}$$

where the expression above is always negative given that $\frac{\partial\overline{p}[h]}{\partial h}\geq 0$ as discussed in the proof of Proposition 1.

Next, we investigate the partial derivative of the earnings-to-price ratio Ξ with respect to markup m. We show that $\frac{\partial\,\Xi[h,p]}{\partial m}<0$, and we begin by applying the chain rule:

$$\frac{\partial\,\Xi[h,p]}{\partial m}=\frac{\partial\,\Xi[h,p]}{\partial h}\left(\frac{\partial m[h,p]}{\partial h}\right)^{-1}<0. \tag{B6}$$

Equation (B5) implies that the first factor in the equation above is strictly negative. From the definition of markup in Equation (9), it follows that $\frac{\partial m[h,p]}{\partial h}>0$. Hence, the partial derivative of the earnings-to-price ratio Ξ with respect to markup m is strictly negative.

Auxiliary Result: Firm Value, V, and Upper Bound for Operating Leverage, Θ

We start by showing that $\Theta[p]>0$ is a sufficient condition for $V[h,p]>0$. We use a proof by contradiction: Assume that $\Pi[p]<0$. Equation (11) shows that $V[h,p]<0$. As a result, it also holds that $V[h,p]>0$ implies $\Pi[p]>0$. Hence $V[h,p]>0$ implies $\Xi[h,p]>0$ and also $\Theta[p]>0$.

As a second step, we consider the definition of firm value given in Equation (12). Given Equation (11), the condition $V[h,p]>0$ implies that

$$0<\Theta[p]<\frac{\frac{\upsilon}{2}-\left(\frac{p}{\overline{p}[h]}\right)^{\upsilon-2}}{\iota-1+\left(\frac{p}{\overline{p}[h]}\right)^{\upsilon-2}}, \tag{B7}$$

where $\iota>1$. Hence the condition $V[h,p]>0$ effectively imposes an upper bound on the operating leverage of the firm. Given $0<p\leq\overline{p}[h]$, the constraint $0<\Theta[p]<\frac{\upsilon-2}{2\iota}$ is a sufficient yet not necessary condition such that $V[h,p]>0$.

Proof of Proposition 3 (expected returns $\mathcal{R}$)

Given Equation (13), the asset pricing implications for expected returns $\mathcal{R}$ are the same as those for the systematic risk loadings β up to a constant. Hence, we derive all implications in terms of the systematic risk loading β. We first solve for the partial derivative of the systematic risk loading β with respect to concentration h. The corresponding partial derivative, with respect to h, is given by

$$\frac{\partial\beta[h,p]}{\partial h}=\gamma(2-\upsilon+2\iota\Theta[p])\frac{\partial\,\Xi[h,p]}{\partial h} \tag{B8}$$

where $\Xi>0$ is strictly decreasing in concentration h as shown in Proposition 2, and the sign of the derivative in Equation (B8) is strictly positive if and only if $0<\Theta[p]<\hat{\Theta}$, where

$$\hat{\Theta}\equiv\frac{\upsilon-2}{2\iota}>0.$$

Next, consider the partial derivative of β with respect to the markup m. We differentiate β with respect to m for a given $p>0$. We apply the chain rule to obtain

$$\frac{\partial\beta[h,p]}{\partial m}=\frac{\partial\beta[h,p]}{\partial h}\left(\frac{\partial m[h,p]}{\partial h}\right)^{-1} \tag{B9}$$

Equation (B8) implies that the first factor in the expression above is strictly positive if and only if $0<\Theta[p]<\hat{\Theta}$. The second factor is positive, as previously discussed. Hence the partial derivative of the systematic risk loading $\beta[p,h]$ with respect to markup m is strictly positive if and only if operating leverage is in the region defined by $0<\Theta[p]<\hat{\Theta}$.

Last, consider the conditional implication in part (ii) of Proposition 2. The functional form of operating leverage $\Theta[p]>0$ implies that there exists a continuously differentiable inverse function g of Θ such that $p\equiv g[\Theta]$. Hence, we replace $p\equiv g[\Theta]$ in Equation (B8) and differentiate the resulting expression with respect to Θ to obtain:

$$\frac{\partial\beta[h,p[\Theta]]}{\partial h\partial\Theta}=-\frac{\gamma\overline{p}[h]p^2(\upsilon-2)\left(V\alpha\delta\upsilon+2\overline{p}[h]^2(p/\overline{p}[h])^{\upsilon-2}\right)(p/\overline{p}[h])^{\upsilon-2}}{r^2(\alpha\delta V)^3\Theta(1+\Theta)\upsilon(\upsilon-2\Theta(\iota-1))(1+\Theta)\upsilon}\overline{p}'[h]\Psi^h[h,\Theta] \quad \text{(B10)}$$

where $\overline{p}'[h]>0$, and the function $\Psi^h[h,\Theta]$ is defined as:

$$\Psi^h[h,\Theta]\equiv\delta^2\Theta^2\upsilon^3+r^2(1+\Theta)^2(\upsilon-2)^2\left(\upsilon+2(p/\overline{p}[h])^{\upsilon-2}\right)-\dots$$

$$\dots-2r\delta\Theta(1+\Theta)\upsilon^2\left(\upsilon+(p/\overline{p}[h])^{\upsilon-2}-\frac{2}{\upsilon}(\upsilon+1)\right),$$

such that the overall sign of Equation (B10) depends on the sign of the factor $(\upsilon-2\Theta(\iota-1))$ in the denominator of Equation (B10), and the sign of the function $\Psi^h[h,\Theta]$.

To solve for the sign of the factor $(\upsilon-2\Theta(\iota-1))$ in Equation (B10), note that in the extreme case in which p tends to zero from the right (i.e., $p\to 0$), the threshold in Equation (B7) tends to $\upsilon/2(\iota-1)$ (i.e., $\Theta\to\upsilon/2(\iota-1)$). Hence the factor $(\upsilon-2\Theta(\iota-1))$ is strictly positive as long as $\Xi[h,p]>0$. Similarly, the function $\Psi^h[h,\Theta]$ is strictly positive for any value of $\Theta[p]$, as long as $\Theta[p]$ is below the threshold in Equation (B7). Put together, these results imply that Equation (B10) is strictly negative, as stated in Proposition 2.

Consider now the analogous prediction in Equation (B10) for the case of markup m. Using the chain rule and the property $p\equiv f[\Theta]$, we obtain:

$$\frac{\partial\beta[h,p[\Theta]]}{\partial m\partial\Theta}=\left(\frac{\partial m[h,p[\Theta]]}{\partial h}\right)^{-1}\Psi^m[h,\Theta], \quad \text{(B11)}$$

where the function $\Psi^m[h,\Theta]$ is defined as:

$$\Psi^m[h,\Theta]\equiv\frac{\partial\beta[h,p[\Theta]]}{\partial h\partial\Theta}-\left(\frac{\partial m[h,p[\Theta]]}{\partial h}\right)^{-1}\frac{\partial m[h,p[\Theta]]}{\partial h\partial\Theta}\frac{\partial\beta[h,p[\Theta]]}{\partial h}.$$

Given the previous derivations in this subsection, it is straightforward to see that Equation (B11) is strictly negative if operating leverage is high enough such that $\Theta[p]>\hat{\Theta}$ and hence $\partial\beta/\partial h<0$. To see this, note that the first factor in Equation (B11) is always positive. The first term of the function $\Psi^m[h,\Theta]$ is negative since Equation (B10) is negative. Moreover, the second term of the function $\Psi^m[h,\Theta]$ is also strictly negative since $\partial\beta/\partial h<0$ and also:

$$\frac{\partial m[h,p[\Theta]]}{\partial h\partial\Theta}=-\frac{c(1-c)}{(\Theta+1)^2(c+h)^2}<0,$$

which proves our claim.

Consider now the alternative case in which $0<\Theta[p]<\hat{\Theta}$. In this case, it is also possible to show that Equation (B11) is strictly negative. To prove this, we replace each of the terms in function $\Psi^m[h,\Theta]$ by our previous findings, and reorder terms to conclude that $\Psi^m[h,\Theta]<0$ holds if and only if:

$$\Psi^h[h,\Theta]>\frac{\Theta}{r}\left(\delta\Theta\upsilon+(\Theta+1)r\left(2\left(p/\overline{p[h]}\right)^{\upsilon}-\upsilon\right)\right)-2((\Theta+1)r(\upsilon-2)-\delta\Theta\upsilon),$$

where the term on the left-hand side of the inequality is our previous function $\Psi^h[h,\Theta]$, which is strictly positive as long as $\Theta[p]$ is below the threshold in Equation (B7). The term on the right-hand side of the inequality is strictly negative given $0<\Theta[p]<\hat{\Theta}$. Hence, we conclude that $\Psi^m[h,\Theta]$ and Equation (B11) are strictly negative given $0<\Theta[p]<\hat{\Theta}$.

Proof of Proposition 4 (risk-feedback channel)

Here, we formalize the intuition that riskier industries (i.e., industries with higher betas) are less likely to experience entry of new firms over a finite future time horizon. For this sake, we consider two alternative industries indexed by $i \in \{A, B\}$. Without loss of generality, let $\beta_A - \beta_B \geq 0$. By using the definition of a firm's beta in Equation (14), the inequality $\beta_A - \beta_B \geq 0$ implies:

$$\frac{\Xi_A}{\delta}(2+2\iota\Theta_A - \upsilon) - \frac{\Xi_B}{\delta}(2+2\iota\Theta_B - \upsilon) > 0. \tag{B12}$$

We also express Ξ_i as a function of $q_i \equiv \left(p_i / \overline{p[h_i]}\right)^{\upsilon}$ and operating leverage Θ_i to obtain:

$$\Xi_i \equiv \delta\upsilon(\upsilon - 2q_i + 2\Theta_i(1 - q_i - \iota))^{-1} \tag{B13}$$

Further replacing Ξ_i in Equation (B12) by the expression in Equation (B13), we obtain an inequality in terms of q_A, q_B, Θ_A and Θ_B, namely,

$$\frac{(2+2\iota\Theta_A - \upsilon)}{(\upsilon - 2q_A + 2\Theta_A(1 - q_A - \iota))} - \frac{(2+2\iota\Theta_B - \upsilon)}{(\upsilon - 2q_B + 2\Theta_B(1 - q_B - \iota))} > 0. \tag{B14}$$

Since asset prices in our model are driven by changes in the state variables p and h, we assess the implications of the inequality in Equation (B14) in two alternative configurations: either when p is the same across industries, or when h is the same across industries. We consider the alternative case in which $\beta_i < 0$ later on. Put together, the assumptions $\beta_A > \beta_B > 0$ and $\Xi_i > 0$ imply that $0 < \Theta_i < \hat{\Theta}$.

Consider the case in which $p_A = p_B = p$ at time $t = 0$ (i.e., part (i) of Proposition 4). This initial condition already implies that $\Pi_A = \Pi_B$ and $\Theta_A = \Theta_B$ at time $t = 0$. Replacing by $\Theta_A = \Theta_B = \Theta$ in Equation (B14), it is straightforward to see that the inequality holds if and only if $q_A < q_B$. Paraphrasing, Equation (B14) holds when industry A has higher barriers to entry since $h_A > h_B$ and also $m_A > m_B$. In sum, given $p_A = p_B = p$ and $\beta_A > \beta_B > 0$, we conclude that entry is less likely in industry A such that $q_A < q_B$.

The alternative case in which $h_A = h_B = h$ at time $t = 0$ (i.e., part (ii) of Proposition 4) is the situation in which industries A and B have the same number of firms at time $t = 0$. It is straightforward to see that the inequality in Equation (B14) holds if and only if $\Pi_A < \Pi_B$, which implies $\Theta_A > \Theta_B$ and $q_A < q_B$. In sum, given $h_A = h_B = h$ and $\beta_A > \beta_B > 0$, we prove that entry is less likely in industry A such that $q_A < q_B$.

Note that all derivations so far considered the assumptions $\beta_A > \beta_B > 0$. Consider now the alternative case in which $0 > \beta_A > \beta_B$ at $t = 0$. In this case, firms in industry A are actually *safer* than firms in industry B. Moreover, the assumptions $0 > \beta_A > \beta_B$, and $\Xi_i > 0$ jointly imply that $\Theta_i > \hat{\Theta}$ (and yet Θ_i is below the threshold in Equation (B7)). In this case, it is straightforward to show that all relations implied by Equation (B14) for cases (i) and (ii) are actually reversed, such that $q_A > q_B$. This is consistent with Proposition 4 since industry A is now safer than industry B. Similarly, the case $\beta_A > 0$, $\beta_B < 0$ and $\beta_A - \beta_B > 0$ where industry A is riskier implies $q_A < q_B$. Finally, note that the case in which $\beta_A < 0$ and $\beta_B > 0$ is not relevant for our analysis as it violates the initial assumption $\beta_A - \beta_B > 0$.

To complete the proof, we note that the probability of entry $\Pr[\mathrm{i,t,T}]$ in Proposition 4 is a continuously increasing function of the term q_i for $i = A, B$, and is defined as:

$$\Pr[\mathrm{i,t,T}] \equiv \Phi\left[\frac{-ln[\frac{\overline{p[h_i]}}{p_i}] + \hat{\mu}(T-t)}{\sigma\sqrt{T-t}}\right] + e^{\frac{2\hat{\mu}}{\sigma^2}}\frac{\overline{p[h_i]}}{p_i}\Phi\left[\frac{-ln[\frac{\overline{p[h_i]}}{p_i}] - \hat{\mu}(T-t)}{\sigma\sqrt{T-t}}\right] \tag{B15}$$

where $\hat{\mu} \equiv \mu - \delta - \frac{\sigma^2}{2}$ is the risk-adjusted drift of p, Φ is the cumulative distribution of the standard normal density, and $T > t$. Equation (B15) equals the probability of the first hitting time of a geometric Brownian motion as in Musiela and Rutkowski (2011).

Appendix C: Details of Data Sources and Sample Construction

For variables related to industry-level characteristics, we use data from the Bureau of Economic Analysis (BEA), and from the Center for Economic Studies (CES), the Statistics of U.S. Businesses (SUSB), the Economic Census, and the Census of Manufactures of the U.S. Census Bureau. Whenever needed, we forward-fill data by repeating industry–year observations in subsequent years. In all of our analyses, we lag both right-hand side and sorting variables by two years, because we do not have access to the exact release date of each industry–year observation from BEA, NBER–CES, or the U.S. Census.[34] Six-digit NAICS measures are aggregated into corresponding four-digit SIC measures by weighting them according to their squared share of the broader industry classification. We bridge NAICS and SIC codes using the concordance tables available at the U.S. Census Web site.

Monthly common stock and accounting data are obtained from firms covered in the CRSP/Compustat merged files that are listed on the NYSE, AMEX, and NASDAQ. We exclude industries with abnormal competitive environments, namely financial industries (SIC codes from 6000 to 6999) and regulated industries (SIC codes from 4900 to 4999). We exclude firm–year observations that have at least one missing monthly return observation in the year and those that have a missing size, book-to-market, or book leverage observation in the previous two years. Firm-level accounting variables and size measures are Winsorized at the 0.5% level in each sample year to reduce the influence of possible outliers. For the same reason, we exclude from the sample firms with market values lower than the NYSE 5% bottom size percentile of firms to avoid anomalies driven by micro-cap firms, as discussed by Fama and French (2008).

Size is defined as the market value of equity (Compustat variables PRC times SHROUT). Book value of equity is defined as in Daniel and Titman (2006). Book-to-market ratio is defined as the book value of equity over size. We require that the measures of book-to-market and size are available at least seven months before the year. Leverage ratios are calculated as the book value of debt adjusted for cash holdings, as reported in Compustat, divided by the sum of market value of equity and the book value of debt (i.e., the market-valued leverage ratio).

Excess stock returns (R^{E}) are constructed as stock returns in excess of the 1-month Treasury bill rate. Annualized excess stock returns are constructed by compounding monthly excess stock returns. Unlevered excess returns ($R^{E,U}$) are defined as excess stock returns times one minus the ratio of the book value of debt to assets minus the book value of equity plus the market value of equity.[35] Excess stock returns adjusted for size, book-to-market ratio, and past returns (R^{DGTW}) are constructed according to the methodology in Daniel et al. (1997). Excess stock returns adjusted for size, book-to-market ratio, and profitability (R^{ADJ}) follow the same construction procedure of R^{DGTW} except that, instead of the adjustment for past returns, we adjust returns for cross-sectional variation in profit margins (EBITDA/Assets).

Conditional betas are estimated every year using monthly return data following the methodology in Lewellen and Nagel (2006). The construction of conditional betas follows the methodology in Dimson (1979) to take into account nonsynchronous stock return data. Conditional betas on a given factor are constructed as the sum of the slope coefficients of regressions of monthly

[34] The use of 2-year lags in regressions and sorting exercises, as well as the use of forward-filling instead of back-filling for missing observations, is a compromise between the need to minimize forward-looking information in our measures, which could introduce spurious pricing anomalies, and the need to avoid the use of outdated information, which would reduce the statistical power of our analyses. We should note, however, that the mechanism discussed in this paper is based on the real interaction between firms in a given industry and not on the reaction of investors that receive news about these firms. In other words, the main testable implications in this paper are not related to pricing anomalies but instead relate to the explanation of "where betas come from," as discussed by Cochrane (2011), and an explanation of how these betas affect, and are affected by, the industry structure.

[35] That is, $R^{E,U} \equiv R^{E}\left(1-\frac{D}{A-BE+ME}\right)$, where D, A, BE, and ME are the book value of debt, book value of assets, book value of equity, and market value of equity, respectively.

returns on contemporaneous and lagged factor returns. Conditional betas of multifactor models are constructed in multiple variate regressions and are also robust to nonsynchronous stock return data. The returns of the market, value, size, and momentum risk factors are obtained from Kenneth French's Web site and described in Fama and French (1993) and Carhart (1997). Expected returns from the CAPM (R^{CPM}), Fama and French (1993) three-factor model (R^{FF3}), and Fama-French-Carhart four-factor model from Carhart (1997) (R^{FC4}) are constructed as the product of conditional betas and the average factor returns over the period 1927 to 2015. Unlevered expected returns from the CAPM model ($R^{CPM,U}$) are constructed as R^{CPM} times one minus the ratio of the book value of debt to assets minus the book value of equity plus the market value of equity.

We employ additional variables to use in the first stage of our sample selection correction procedure. We compute the share of public firms by industry–year by considering the ratio of the total number of public firms as reported in Compustat over the total number of private and public firms reported by the survey of Statistics of U.S. Businesses (SUSB).[36] The average industry annual growth in sales and the volatility in the industry growth are constructed using the item sale in Compustat. The volatility in industry growth is computed using a span of four years. The share of firms in the industry with positive expenses in R&D is constructed using the item XRD in Compustat. The employment growth by industry is measured as the growth rate in employment as reported by the Bureau of Labor Statistics (BLS). The average wage by industry is also obtained from the BLS. The average turnover of firms' shares by industry is constructed as the ratio of the volume of stock traded to the number of shares outstanding for each firm. We then compute the average turnover by industry year. We also construct the share of public firms in a given industry that are registered in the NYSE in that year. To achieve this, we extract the main exchange for each firm as reported in CRSP.

Appendix D: Construction of the Inverse Mills Ratios

We start by motivating the need to control for sample selection bias in our setting. Let d_{ij} be an unobservable variable that determines the decision of firm j operating in industry i to be publicly listed. Specifically, if $d_{ij} \geq \bar{d}$, then the firm chooses to be publicly listed, and if $d_{ij} < \bar{d}$, the firm chooses to be private. Furthermore, let us assume that d_{ij} is partially explained by some observable characteristics, as given by

$$d_{ijt} = \gamma_t \mathbf{z}_{ij,t-l} + u_{ijt}, \tag{D1}$$

where $\mathbf{z}_{ij}$ is a vector of observable characteristics predetermined relative to d_{ij}, and u_{ij} is the firm-specific deviation from the industry mean, such that $\mathrm{E}[u_{ij}]=0$ and $\mathrm{Var}\left(u_{ij}\right) \equiv \sigma^2_{\mathrm{u}i}$. Let v_{ij} denote a variable of interest (e.g., excess stock returns or sales) for firm j in industry i. With no loss of generality, v_{ijt} can be decomposed as

$$v_{ijt} = c_t \mathbf{x}_{ij,t-l} + \epsilon_{ijt}, \tag{D2}$$

where $\mathbf{x}_{ij}$ is a vector of observable characteristics that are predetermined relative to v_{ij}, and ϵ_{ij} is the firm-specific deviation from the industry mean, such that $\mathrm{E}[\epsilon_{ij}]=0$ and $\mathrm{Var}\left(\epsilon_{ij}\right) \equiv \sigma^2_{\epsilon i}$.[37] The challenge is that one can only observe v_{ij} for the subsample of firms that decide to be publicly

[36] The SUSB data set covers both manufacturing and nonmanufacturing industries with annual frequency since 1992. See http://www.census.gov/programs-surveys/susb/about.html for further details on this survey. The U.S. Census of Manufactures also provides data on the total number of firms by industry, although it is restricted to manufacturing industries at a 5-year frequency and thus less suited to implement the sample selection correction methodology proposed in this paper.

[37] In practice, we use lagged explanatory variables in the vectors $\mathbf{z}$ and $\mathbf{x}$ when estimating Models (D1) and (D2) so that they are predetermined relative to d_{ijt} and v_{ijt}, respectively.

listed (i.e., those firms for which $d_{ij} \geq \bar{d}$). Because we only observe data for publicly listed firms, the conditional mean of v_{ij} is given by

$$\mathrm{E}_t[v_{ij}|v_{ij} \text{ is observable}] = \underbrace{\mathrm{E}_t[v_{ij}]}_{\text{Population mean}} + \underbrace{\mathrm{E}_t[\epsilon_{ij}|u_{ij} > \bar{d} - \gamma_t \mathbf{z}_{ij,t-l}]}_{\text{Sample selection bias}}. \tag{D3}$$

Equation (D3) shows that the estimation is biased if ϵ_{ij} is correlated with u_{ij} (i.e., if ϵ_{ij} is correlated with the underlying determinants of a firm's public status). In our setting, controlling for sample selection bias relies on the premise that concentration is related to both the decision to be public d_{ij} in Equation (D1) and the variable of interest v_{ij} in Equation (D2). For instance, the inference from the relation between stock returns and product market competition may be biased if the concentration of the industry in which firms operate affects both their exposure to risk and their public status.

We compute the inverse Mills ratios of manufacturing and nonmanufacturing industries by industry and by year using a methodology similar to that used in the selection models of proportions data, which is discussed in Greene (2003). This methodology relies on the assumption that, conditional on observable information, all firms in the industry have the same probability of being in the sample in any given year. Let $p_{it}^{\text{Pub}} \equiv \frac{N_{it}^{\text{Pub}}}{N_{it}^{\text{All}}}$ be the proportion of public to total firms in industry i in year t. We construct the variable p_{it}^{Pub} in our data set by considering the number of public firms as reported by Compustat and considering the total number of firms (i.e., total number of both private and public firms) as reported by the SUSB data set of the U.S. Census Bureau. The selection model is given by

$$p_{it}^{\text{Pub}} = \Phi\left(\gamma_t \mathbf{z}_{i,t-l}\right) + \zeta_{it}, \tag{D4}$$

where $\mathbf{z}_i$ denotes the vector of industry-specific characteristics that determine the public status of firms in a given industry group, Φ denotes the normal cumulative density function, and ζ_i is the sampling error. We treat the sampling of public firms within the same industry as a problem of sampling from a Bernoulli population. Hence ζ_i is such that $\mathrm{E}_t[\zeta_i] = 0$ and $\mathrm{Var}_t[\zeta_i] \equiv \sigma_{\zeta it} = \Phi\left(\gamma_t \mathbf{z}_{i,t-l}\right)\left(1 - \Phi\left(\gamma_t \mathbf{z}_{i,t-l}\right)\right)\left(N_{it}^{\text{Pub}}\right)^{-1}$.

The model in Equation (D4) can be estimated using nonlinear least squares. However, as discussed in Greene (2003), there is a simpler approach using linear least squares. Given that the function Φ has an inverse, we use the alternative specification

$$\Phi^{-1}\left(p_{it}^{\text{Pub}}\right) \approx \gamma_t \mathbf{z}_{i,t-l} + \frac{\zeta_{i,t}}{\phi\left(\gamma_t \mathbf{z}_{i,t-l}\right)}. \tag{D5}$$

We estimate Model (D5) for every year of our sample in two passes. The resulting standard errors from the first pass are heteroskedastic. For this reason, we use the results of the first pass to generate sample weights to correct for heteroscedasticity. In particular, we use estimates of the coefficient γ of Model (D5), which we represent as $\hat{\gamma}$, to generate the sample weights w_i given by

$$w_{it} = \frac{N_{it}^{\text{Pub}} \phi\left(\hat{\gamma}_t \mathbf{z}_{i,t-l}\right)^2}{\Phi\left(\hat{\gamma}_t \mathbf{z}_{i,t-l}\right)\left(1 - \Phi\left(\hat{\gamma}_t \mathbf{z}_{i,t-l}\right)\right)}. \tag{D6}$$

In the second pass, we estimate Model (D5) using the weights above as p-weights, to obtain the unbiased estimates of the coefficient γ, which we term $\hat{\hat{\gamma}}$. We use our estimates $\hat{\hat{\gamma}}$ to construct the inverse Mills ratio for all industries in our first-stage sample, as given by

$$\lambda_{it} \equiv \frac{\phi\left(\hat{\hat{\gamma}}_t \mathbf{z}_{i,t-l} / \sigma_{it}^{\zeta}\right)}{\Phi\left(\hat{\hat{\gamma}}_t \mathbf{z}_{i,t-l} / \sigma_{it}^{\zeta}\right)}. \tag{D7}$$

The last step is to compute the inverse Mills ratio λ_i for those industries in our working sample of CRSP/Compustat firms for which the total number of firms is not available in the SUSB data set of the U.S. Census Bureau. Since we do not observe p_i^{Pub} for these industries, we use the vector of instruments $\mathbf{z}_i$ and our estimates $\hat{\hat{\gamma}}$ to compute λ_i for these industries. The working assumption here is that the estimates $\hat{\hat{\gamma}}$ are the same for all industries in our CRSP/Compustat data set.

Appendix E: Construction of the CBC Measure

We construct the CBC measure in two stages: In the first stage, we estimate the inverse Mills ratios, as given by Equation (D7). In the second stage, we use these inverse Mills ratios to estimate the adjusted means and the variances of sales of public firms as well as the adjusted number of firms in the industry for those industries in which we do not observe the total number of firms. We then use these estimates in our definition of the CBC measure, as given in Equation (19).

To estimate the inverse Mills ratio, we first define the variables used in the vector $\mathbf{z}_i$ from Equation (D4). We purposely do not include any variables in $\mathbf{z}_i$ that contain market valuations. We do this to avoid any mechanical relation in our tests between the CBC measure and a firm's earnings-to-price ratios or stock returns. We specify the vector $\mathbf{z}_i$ and closely follow the variables that explain the decision to go public, as discussed in Chemmanur, He, and Nandy (2010). In particular, the vector $\mathbf{z}_i$ includes the average firm growth sales by industry, the standard deviation in the industry growth sales over the past four years, and the share of firms in the industry with positive expenses in R&D. While these variables are based on public firm data only, the average firm growth in sales by industry from Compustat is highly correlated with the growth in the value of shipments reported for manufacturing industries in the ASM: We use the Compustat measure to extend our analysis to nonmanufacturing industries. Similarly, R&D expenditures are highly industry specific. The remainder of the variables in vector $\mathbf{z}_i$ are obtained from sources that report industry averages, including both private and public firms, for both manufacturing and nonmanufacturing industries. The employment growth by industry captures variation in industry size and in industry life cycle. The average industry wage captures differences in operating leverage not driven by differences in industry performance.[38] All these variables are also used in the vector $\mathbf{x}_i$ from Equation (D2), which is used in the second stage.

In addition to the variables above, the vector $\mathbf{z}_i$ also includes two instruments that are not included as explanatory variables in the second stage of the estimation. We consider these variables as instruments that relate to a firm's decision to be publicly listed, but these variables do not explain the dependent variable in the second stage (i.e., the mean and variance of firms' sales). The first instrument is the average turnover of a firm's shares by industry. To construct this variable, we compute the ratio of the volume of stock traded divided by the number of shares outstanding for each firm; we then compute the average turnover by industry group. The second instrument is the share of public firms in the industry listed at the New York Stock Exchange (NYSE).

Appendix Table E1 summarizes the results of the first stage. The first column reports the baseline specification, which we use to construct the CBC. The remainder of the columns report alternative specifications to assess the robustness of our findings. The goodness-of-fit of our baseline specification is particularly high: The time-series average of the R-squared is between 25% and 32% across all cross-sectional regressions. The sign and significance of the coefficients of our instruments also align with previous studies. Consistent with Chemmanur, He, and Nandy (2010), firms are more likely to go public in industries with higher investor demand. Moreover, firms are more likely to go public in industries in which the fraction of public firms quotes predominantly at the NYSE. This claim is justified, given the evidence in Corwin and Harris (2001), who report that the initial firm size and firm age of IPOs at NYSE are significantly higher than the corresponding firm characteristics of IPOs at NASDAQ.

The sample selection correction is conducted at the industry level and relies on the distributional assumptions that the error terms of the selection and main equations of the model are bivariate normal such that $(\zeta_{it}, \epsilon_{it})$ bivariate normal $\sim \left[0, 0, \sigma^{\zeta}_{it}, \sigma^{\epsilon}_{it}, \rho_i\right]$. During the second stage, we consider v_{ij} to be the logarithm of a firm's Compustat sales in a given industry. Given that the empirical methodology to correct for selection bias relies on the normality assumption, we use log sales in the OLS regressions, since sales are highly skewed and the goodness-of-fit is higher when we use the same variable in logs.

[38] See, for instance, Favilukis and Lin (2016a).

Table E1
Estimates from the first stage

Specification	1	2	3	4	5	6
$NYSE_{t-1}$	0.53***	0.57***	0.59***	0.41***	0.53***	0.48***
	(0.05)	(0.07)	(0.06)	(0.05)	(0.06)	(0.05)
$Turnover_{t-1}$	4.47**	3.08*	3.65*	6.70***	4.29**	4.55**
	(1.79)	(1.59)	(1.93)	(2.24)	(1.76)	(1.74)
Emp G_{t-1}	−4.29***		−4.44***	−5.53***	−4.09***	−4.09***
	(0.65)		(0.72)	(0.75)	(0.65)	(0.65)
$Wage_{t-1}$	0.01***	0.01***		0.01***	0.01***	0.01***
	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)
RD_{t-1}	0.58***	0.65***	0.73***		0.58***	0.59***
	(0.03)	(0.04)	(0.03)		(0.02)	(0.03)
Sale G_{t-1}	−0.04	−0.11	−0.05	−0.06		0.14**
	(0.09)	(0.07)	(0.09)	(0.11)		(0.06)
Std Sale G_{t-1}	0.53***	0.49***	0.72***	0.59***	0.47***	
	(0.08)	(0.07)	(0.12)	(0.10)	(0.06)	
R-sq.	0.32	0.29	0.25	0.25	0.31	0.31
Obs.	4,792	4,802	4,792	4,792	4,792	4,792

The table below shows estimates and standard errors of cross-sectional regressions of the first stage of the sample selection correction procedure (Specification 1), as well as alternative specifications. In each year of the sample, we regress the inverse normal of the share of public firms in the industry on our first-stage instruments (i.e., the share of firms listed on NYSE and the average stock turnover in the industry) and second-stage controls. Newey–West standard errors, estimated with one lag, are shown in parentheses. *, ** and *** indicate significance levels of 10%, 5%, and 1%, respectively. The sample period is 1992–2012.

We know that the conditional industry mean of v_{ij} for a public firm is given by

$$\begin{aligned} \mathrm{E}_t\left(v_{ij}\,|\,j \text{ is public}\right) &= c_t \mathbf{x}_{i,t-1} + \mathrm{E}_t\left(\epsilon_{ij}\,|\,j \text{ is public}\right) \\ &= c_t \mathbf{x}_{i,t-1} + \rho_{it}\sigma^{\epsilon}_{it}\lambda_{it}. \end{aligned} \tag{E1}$$

Similarly, the conditional variance of v_{ijt} in industry i is given by

$$\begin{aligned} \mathrm{Var}_t\left(v_{ij}\,|\,j \text{ is public}\right) &= \mathrm{E}\left(\epsilon^2_{ij}\,|\,j \text{ is public}\right) \\ &= (\sigma^{\epsilon}_{it})^2\left[1-\rho^2_{it}(1-\upsilon_{it})\right], \end{aligned} \tag{E2}$$

where $\upsilon_{it} \equiv 1-\lambda_{it}\left(\lambda_{it}+\hat{\hat{\gamma}}_t \mathbf{z}_{i,t-1}\right)$.

We run the second stage of the CBC by regressing the log of industry sales on the controls from the first stage and the inverse Mills ratio from the first stage. We use results of these regressions to construct the adjusted average log sales for public and private firms by industry–year $\hat{\mu}_{lnsales}$ and the adjusted industry variance in log sales of public and private firms by industry–year, $\hat{\sigma}_{lnsales}$.

Given the definition of the mean and the variance of the log normal distribution function, we then use the estimates of $\hat{\mu}_{lnsales}$ and $\hat{\sigma}_{lnsales}$ to compute $\hat{\mu}_{sales}$ and $\hat{\sigma}_{sales}$. The underlying logic is that, since lnsales is normally distributed, sales are lognormally distributed. We define $\hat{\mu}_{sales} \equiv e^{\left(\hat{\sigma}_{lnsales}+\frac{1}{2}\hat{\sigma}^2_{lnsales}\right)}$ and $\hat{\sigma}_{sales} \equiv \hat{\mu}_{sales}\left(e^{\hat{\sigma}_{lnsales}}-1\right)^{\frac{1}{2}}$.

We construct the CBC measure using the estimates $\hat{\mu}_{sales}$ and $\hat{\sigma}_{sales}$ and the definition of the CBC in Equation (19). For those industries that are not covered in the SUSB data set and thus are industries for which we do not observe the total number of firms, we use the adjusted number of firms in the industry $\hat{N}_{it}$ given by

$$\hat{N}_{it} \equiv \frac{N^{\text{Pub}}_{it}}{\Phi(\hat{\hat{\gamma}}_t \mathbf{z}_{it})}. \tag{E3}$$

Appendix Table E2 reports the 15 top and the 15 bottom industries in 2012 sorted by the CBC measure. The list shows that the lowest-CBC industries are service-based, while most of the highest-CBC industries are manufacturing-based. Later, we will show that the share of manufacturing

Table E2
Top and bottom industries sorted by CBC

SIC	Industry title	CBC
	Least concentrated industries	
8011	Offices and Clinics of Doctors of Medicine	−1.2
8741	Management Services	−1.1
5141	Groceries, General Line	−0.3
8721	Accounting, Auditing, and Bookkeeping Services	−0.2
8093	Specialty Outpatient Facilities	−0.1
5013	Motor Vehicle Supplies and New Parts	0.2
7381	Detective, Guard, and Armored Car Services	0.5
8734	Testing Laboratories	0.7
5531	Auto and Home Supply Stores	0.8
1731	Electrical Work	0.9
7311	Advertising Agencies	1.0
7011	Hotels and Motels	1.3
5084	Industrial Machinery and Equipment	1.4
5122	Drugs, Drug Proprietaries, and Druggists Sundries	1.4
5063	Electrical Apparatus and Equipment Wiring Supplies	1.4
	Most concentrated industries	
2834	Pharmaceutical Preparations	12.2
1311	Crude Petroleum and Natural Gas	10.3
3674	Semiconductors and Related Devices	10.2
7372	Prepackaged Software	10.1
2836	Biological Products, Except Diagnostic Substances	8.5
3845	Electromedical and Electrotherapeutic Apparatus	8.3
3841	Surgical and Medical Instruments and Apparatus	5.9
3663	Radio and TV Broadcasting Equipment	5.3
5812	Eating Places	4.4
2835	In Vitro and In Vivo Diagnostic Substances	4.3
3714	Motor Vehicle Parts and Accessories	4.1
2911	Petroleum Refining	3.9
7373	Computer Integrated Systems Design	3.2
3842	Orthopedic, Prosthetic, and Surgical Appliances	2.9
7374	Computer Processing and Data Preparation Services	2.8

The table presents the bottom 15 and top 15 four-digit SIC industries sorted by the CBC measure in 2012.

industries over all CBC-sorted industries is in fact increasing, but not as starkly as the extremes shown in the table.

References

Abdel-Raouf, F. 2010. Trade-adjusted concentration ratios in the US manufacturing sector. *International Journal of the Economics of Business* 17:385–403.

Aguerrevere, F. L. 2003. Equilibrium investment strategies and output price behavior: A real-options approach. *Review of Financial Studies* 16:1239–72.

———. 2009. Real options, product market competition, and asset returns. *Journal of Finance* 64:957–83.

Ali, A., S. Klasa, and E. Yeung. 2009. The limitations of industry concentration measures constructed with Compustat data: Implications for finance research. *Review of Financial Studies* 22:3839–71.

Allayannis, G. and J. Ihrig. 2001. Exposure and markups. *Review of Financial Studies* 14:805–35.

Bain, J. 1956. *Barriers to New Competition*. Cambridge, MA: Harvard University Press.

Belo, F., X. Lin, and S. Bazdresch. 2014. Labor hiring, investment and stock return predictability in the cross section. *Journal of Political Economy* 122:129–77.

Berk, J. B., R. C. Green, and V. Naik. 1999. Optimal investment, growth options, and security returns. *Journal of Finance* 54:1553–607.

Bernstein, S. 2015. Does going public affect innovation? *Journal of Finance* 70:1365–403.

Bhojraj, S., C. M. C. Lee, and D. K. Oler. 2003. What's my line? A comparison of industry classification schemes for capital market research. *Journal of Accounting Research* 41:745–74.

Bulan, L. T., C. J. Mayer, and C. T. Somerville. 2009. Irreversible investment, real options, and competition: Evidence from real estate development. *Journal of Urban Economics* 65:237–51.

Bustamante, M. C. 2015. Strategic investment and industry risk dynamics. *Review of Financial Studies* 28:297–341.

Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52:57–82.

Carlson, M., A. Fisher, and R. Giammarino. 2004. Corporate investment and asset price dynamics: Implications for the cross-section of returns. *Journal of Finance* 59:2577–603.

Chemmanur, T., S. He, and D. Nandy. 2010. The going public decision and the product market. *Review of Financial Studies* 23:1855–908.

Cochrane, J. H. 2011. Discount rates. *Journal of Finance* 4:1047–108.

Cooper, I. and R. Priestley. 2016. The expected returns and valuations of private and public firms. *Journal of Financial Economics* 120:41–57.

Corwin, S. A. and J. H. Harris. 2001. The initial listing decisions of fims that go public. *Financial Management* 30:35–55.

Daniel, K. and S. Titman. 2006. Market reactions to tangible and intangible information. *Journal of Finance* 61:1605–43.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52:1035–58.

Danthine, J. and J. B. Donaldson. 2002. Labor relations and asset returns. *Review of Economic Studies* 69:41–64.

Dimson, E. 1979. Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics* 7:197–226.

Donangelo, A. 2014. Labor mobility: Implications for asset pricing. *Journal of Finance* 69:1321–46.

Dugan, M. T., D. H. Minyard, and K. A. Shriver. 1994. A re-examination of the operating leverage-financial leverage tradeoff hypothesis. *Quarterly Review of Economics and Finance* 34:327–34.

Fama, E. F. and K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3–56.

———. 2008. Dissecting anomalies. *Journal of Finance* 63:1653–78.

Favilukis, J. and X. Lin. 2016a. Does wage rigidity make firms riskier? Evidence from long-horizon return predictability. *Journal of Monetary Economics* 78:80–95.

———. 2016b. Wage rigidity: A quantitative solution to several asset pricing puzzles. *Review of Financial Studies* 29:148–92.

Gahlon, J. M. and J. A. Gentry. 1982. On the relationship between systematic risk and the degrees of operating and financial leverage. *Financial Management* 11:15–23.

Garcia-Feijoo, L. and R. D. Jorgensen. 2010. Can operating leverage be the cause of the value premium? *Financial Management* 39:1127–53.

Garlappi, L. 2004. Risk premia and preemption in R&D ventures. *Journal of Financial and Quantitative Analysis* 39:843–72.

Gopalan, R. and K. Xie. 2011. Conglomerates and industry distress. *Review of Financial Studies* 24:3642–87.

Greene, W. 2003. *Econometric analysis*. Pearson, Upper Saddle River, NJ.

Grenadier, S. R. 2002. Option exercise games: An application to the equilibrium investment strategies of firms. *Review of Financial Studies* 15:691–721.

Guenther, D. A. and A. J. Rosman. 1994. Differences between Compustat and CRSP SIC codes and related effects on research. *Journal of Accounting and Economics* 18:115–28.

Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47:153–61.

Hoberg, G. and G. M. Phillips. 2014. Product market uniqueness, organizational form and stock market valuations. Working Paper.

Hoberg, G. and G. Phillips. 2010. Real and financial industry booms and busts. *Journal of Finance* **65**:45–86.

Hou, K. and D. T. Robinson. 2006. Industry concentration and average stock returns. *Journal of Finance* 61:1927–56.

Kahle, K. M. and R. A. Walkling. 1996. The impact of industry classifications on financial research. *Journal of Financial and Quantitative Analysis* 31:309–35.

Lerner, A. 1934. The concept of monopoly and the measurement of monopoly power. *Review of Economic Studies* 3:157–75.

Lev, B. 1974. On the association between operating leverage and risk. *Journal of Financial and Quantitative Analysis* 9:627–41.

Lewellen, J. and S. Nagel. 2006. The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics* 82:289–314.

Loualiche, E. 2017. Asset pricing with entry and imperfect competition. Working Paper.

Mandelker, G. N. and S. G. Rhee. 1984. The impact of the degrees of operating and financial leverage on systematic risk of common stock. *Journal of Financial and Quantitative Analysis* 19:45–57.

Maury, B. and A. Pajuste. 2005. Multiple large shareholders and firm value. *Journal of Banking & Finance* 29:1813–34.

Musiela, M. and M. Rutkowski. 2011. *Martingale Methods in Financial Modelling*. Springer, New York.

Novy-Marx, R. 2011. Operating leverage. *Review of Finance* 15:103–34.

———. 2013. The other side of value: the gross profitability premium. *Journal of Financial Economics* 108:1–28.

Opp, M., C. A. Parlour, and J. Walden. 2014. Markup cycles, dynamic misallocation, and amplification. *Journal of Economic Theory* 154:126–61.

Pindyck, R. S. 2009. Sunk costs and risk based barriers to entry. NBER Working Paper.

Subrahmanyam, M. G. and S. B. Thomadakis. 1980. Systematic risk and the theory of the firm. *Quarterly Journal of Economics* 94:437–51.

Tirole, J. 1988. *The Theory of Industrial Organization*. Cambridge: MIT Press.

V Binsbergen, J. H. 2016. Good-specific habit formation and the cross section of expected returns. *Journal of Finance* 99:1–10.

Zhang, L. 2005. The value premium. *Journal of Finance* 60:67–103.

The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXX, NO. 2 • APRIL 2015

Size Anomalies in U.S. Bank Stock Returns

PRIYANK GANDHI and HANNO LUSTIG*

ABSTRACT

The largest commercial bank stocks, ranked by total size of the balance sheet, have significantly lower risk-adjusted returns than small- and medium-sized bank stocks, even though large banks are significantly more levered. We uncover a size factor in the component of bank returns that is orthogonal to the standard risk factors, including small minus big, which has the right covariance with bank returns to explain the average risk-adjusted returns. This factor measures size-dependent exposure to bank-specific tail risk. These findings are consistent with government guarantees that protect shareholders of large banks, but not small banks, in disaster states.

BANKS ARE DIFFERENT FROM NONFINANCIAL FIRMS in many ways. One of the most salient distinctions is that banks are subject to bank runs during banking panics and crises, not just by depositors, but also by other creditors (see Gorton and Metrick (2012) and Duffie (2010)). Because financial crises are high marginal utility states for the average investor, the expected return on bank stocks should be especially sensitive to variation in the anticipated financial disaster recovery rates of bank shareholders related to bank size, the regulatory regime, implicit government guarantees, and other characteristics. For example, if a bank is deemed too big to fail, the expected return on its stock is lower in equilibrium than that of smaller banks holding the exact same assets in their portfolio because the government absorbs some of the large bank's tail risk. We find evidence that the pricing of bank-specific tail risk in the stock market depends on all of these bank characteristics.

To explore the asset pricing implications of financial disasters, our paper studies historical bank stock returns in the United States. We find that there

*Gandhi is with Mendoza College of Business, University of Notre Dame and Lustig is with Anderson School of Management, University of California at Los Angeles. The authors thank the Editors, Cam Harvey and Ken Singleton, as well as the Associate Editor and several referees for excellent suggestions. The authors also thank Viral Acharya, Martin Bodenstein, Markus Brunnermeier, John Campbell, John Cochrane, Ron Feldman, Cesare Fracassi, Etienne Gagnon, Mark Garmaise, Amit Goyal, Mark Grinblatt, Jennifer Huang, John Krainer, Arvind Krishnamurthy, David Laibson, Jonathan Parker, Lasse Pedersen, Andrea Raffo, Richard Roll, Jan Schneider, Martin Schneider, Clemens Sialm, Skander Vandenheuvel, Rob Vishny, and Baolian Wang for many for detailed and helpful comments, as well as seminar participants at University of Texas at Austin, Chicago Booth, New York University Stern, Harvard Economics, the Federal Reserve Board of Governors, the Federal Reserve Bank of San Francisco, the Society for Economic Dynamics 2010 Annual Meeting in Montreal, and the 2010 Stanford Institute for Theoretical Economics for detailed and helpful comments.

DOI: 10.1111/jofi.12235

is a size effect in bank stock returns that is different from the market capitalization effects that have been documented in nonfinancial stock returns (see Banz (1981) and many others). All else equal, a 100% increase in a bank's book value lowers its annual return by 2.23% per annum. For nonfinancial stocks, there is no similar relation between book value and returns (Berk (1997)).

These return differences cannot be imputed to differences in standard risk exposure. A long position in the stock portfolio of the largest commercial banks, measured by deciles of total book value, and a short position in the stock portfolio of the smallest banks underperforms an equally risky portfolio of all (nonbank) stocks and bonds by more than 7% per annum. The average alphas are small but positive for commercial banks in the first five deciles and then decrease for the largest banks in the top three deciles.

Small banks differ from large banks in many ways, but these differences should not lead to differences in average risk-adjusted returns on bank portfolios unless there is bank-specific tail risk that is priced but not spanned by the traded returns on other stocks in the sample. We find evidence of such a risk factor in bank stock returns: the second principal component of the risk-adjusted returns on size-sorted portfolios of commercial banks is a size factor that has exactly the right covariance with the portfolio returns to account for most of this pricing anomaly. By construction, this size factor is orthogonal to the stock and bond risk factors.

This highly levered size portfolio, determined by the second principal component, which goes long in small bank stocks and short in large bank stocks, loses an average of 41 cents during National Bureau of Economic Research (NBER) recessions per dollar invested at the start, after hedging out exposure to standard stock and bond risk. We attribute the cyclical banking size factor in the data to size-dependent differences in the perceived shareholder recovery rates on these bank portfolios during financial disasters.

In a version of the Barro (2006), Rietz (1988), and Longstaff and Piazzesi (2004) asset pricing model with a time-varying probability of rare events, developed by Gabaix (2012), Wachter (2013), and Gourio (2008), financial disasters that disproportionately impact bank cash flows contribute an additional bank-specific risk factor. These rare events are priced into the expected returns on portfolios of banks, but are not fully spanned by the returns on other assets in a small sample. A general equilibrium version of our model can match the average alphas in a sample without disasters if the financial disaster recovery rate is 35 cents higher for large banks, in line with the failure rate of banks in the lowest decile during the latest crisis.

Historically, the probability of a financial disaster increases during recessions. Because of the size-contingent nature of the recovery rate for bank stockholders in the case of a financial disaster, the variation in the probability of a financial disaster generates a common business cycle factor in the normal risk-adjusted returns of size-sorted bank stock portfolios; the loadings of bank stock portfolio returns on this size factor are determined by the recovery rates and hence by size. Small banks have positive loadings while large banks have negative loadings. As the probability of a financial disaster increases, the expected return gap between small and large banks grows.

In the United States, shareholder recovery rates for banks depend on size. During financial disasters, large banks fare much better, even though they are more levered than their smaller counterparts. A total of 30% of publicly traded commercial banks in the first size decile were delisted in 2009 but there were no delistings in the last decile.

Why study the effect of bailouts on bank equity? The anticipation of future bailouts of bondholders and other creditors always benefits shareholders (see Kareken and Wallace (1978)) ex ante. Furthermore, during a crisis, there may be massive uncertainty about the resolution regime, especially for large financial institutions. As a result, government guarantees will inevitably tend to benefit shareholders ex post as well. Clearly, the U.S. government and regulators are willing to let small banks fail, but not large banks. Acharya and Yorulmazer (2008) point out that bailouts may be ex post efficient if a sufficiently large fraction of banks is impacted. Of course, ex ante, one could have expected that the government would wipe out shareholders of large financial institutions in the case of a bailout. Our evidence suggests that this is not what market participants expect.

Government guarantees essentially grant stockholders of large banks a menu of path-dependent put options that can only be exercised after large declines in a broad index of stocks. This essentially reduces the negative co-skewness of large bank stock returns, but not of small banks. In our sample, large bank stock returns are indeed less negatively skewed and feature less co-skewness, even though the Harvey and Siddique (2000) skewness factors constructed from nonfinancial stocks cannot fully account for the variation in average returns on size-sorted bank portfolios.

To back out the implicit financial tail risk premium or discount charged by the shareholders of commercial banks, we multiply the loadings on the size factor by its risk price. The implicit insurance provided against financial disaster risk lowers the expected equity return for the largest U.S. commercial banks by 1.97%, but the additional exposure to bank-specific tail risk increases the expected return on the smallest bank stocks by 2.85%, compared to a portfolio of nonbank stocks and bonds with the same standard risk characteristics. The largest banks have an average market capitalization of $140 billion in 2005 dollars. For the largest commercial banks, this amounts to an annual savings of $2.76 billion per bank. The market imposes large financial tail risk “subsidies” (“taxes”) on large (small) bank stocks compared to a portfolio of stocks and bonds with the same observed risk profile. There is direct evidence from option markets to support this conclusion: Kelly, Lustig, and Nieuwerburgh (2011) find that out-of-the-money put options on large banks were cheap during the crisis.

The pricing of financial tail risk depends not only on bank size. We relate the financial disaster premium of banks to the regulatory regime. Commercial banks, which have access to the discount window and benefit from deposit insurance, and government-sponsored enterprises (GSEs), which benefit from an explicit guarantee, are imputed a large financial tail risk subsidy while investment and foreign banks are not. On the other hand, hedge funds are imputed a financial tail risk tax, just like small banks.

After the repeal of key provisions in the Glass-Steagall Banking Act in 1999, we find large across-the-board increases in the size of the subsidy for large commercial banks, investment banks, and GSEs. For example, the Fannie Mae subsidy tripled to 5.93% over the period 2000 to 2005. This period also coincides with dramatic growth in securitization, which allows financial institutions to benefit from the collective bailout option more aggressively by eliminating idiosyncratic risk exposure (see Brunnermeier and Sannikov (2014) for a clear description of this effect of securitization).

Furthermore, we provide a direct link to bailouts. We show that the financial disaster subsidy of the largest 10 banks increases immediately after bailout announcements. O'Hara and Shaw (1990) document large positive wealth effects for shareholders of banks that were declared "too big to fail" by the Comptroller of the Currency in 1984, and negative wealth effects for other banks. Consistent with this result, we document large increases in the implicit financial disaster subsidy to the too-big-to-fail banks after this announcement, and six other bailout announcements prior to the recent financial crisis identified by Kho, Lee, and Stulz (2000). Furthermore, we find large increases after announcements that benefited large banks during the recent financial crisis.

The rest of this paper is organized as follows. Section I discusses the related literature. In Section II, we construct portfolios of commercial U.S. bank stocks sorted by size as measured by the market capitalization and the book value. Presumably, the government cares about the size of the entire balance sheet. Section III describes the size effect in bank stock returns. Section IV establishes that there is a procyclical size factor in the normal risk-adjusted returns of these portfolios. Section V relates the pricing of bank tail risk to government announcements and the regulatory environment. We use a calibrated version of the model to back out the implied differences in recovery rates in Section VI, and we use the model to explain how government puts for the largest banks may also impact the expected returns of smaller banks. Section VII concludes.

I. Related Literature

A large literature in finance considers size effects in stock returns (see Banz (1981), Basu (1983), Lakonishok, Shleifer, and Vishny (1993), Fama and French (1993), Berk (1995), and others), but most of these papers do not include financial stocks, presumably because of their high leverage. Our paper is the first to document that the size effect in financial stocks is really about size, rather than market capitalization. We attribute the size effect to how tail risk is priced in financial stocks.

There is direct evidence from option markets that tail risk in the financial sector is priced differently. Kelly, Lustig, and Nieuwerburgh (2011) find that the out-of-the-money index put options of bank stocks were relatively cheap during the recent crisis, as a consequence of the government absorbing sector-wide tail risk. In related work on bank stock returns, Fahlenbrach, Prilmeier, and Stulz (2012) document that those banks that incurred substantial losses during previous crises were more likely to incur losses during the recent crisis. If some banks benefit from a larger perceived tail risk subsidy, they have an incentive to load up on this type of risk. In fact, shareholder value maximization requires that they do so, as pointed out by Panageas (2010a), who analyzes optimal risk management in the presence of guarantees. Interestingly, Fahlenbrach, Prilmeier, and Stulz (2012) find some evidence that banks whose managers' interests were more aligned with shareholders actually performed worse during the recent financial crisis.

Our work contributes to the important task of measuring systemic risk in the financial sector. Acharya, Brownless, et al. (2011), Acharya, Pedersen, et al. (2011), Adrian and Brunnermeier (2010), and Huang, Zhou, and Zhou (2009) develop novel methods for measuring systemic risk. Our measure of the banking tail risk premium is determined by the bank's loading on the size factor, which gauges a firm's systemic risk exposure. Firms that are deemed systemically important have large negative loadings on the size factor, because these are less likely to be allowed to fail in the event of a financial disaster, and they trade at a premium as a result. As far as we know, our paper is the first to link the subsidy that accrues to banks deemed systemically important with exposure to systemic risk. To the extent that these differences in bank tail risk pricing are directly attributable to government policies, they are an ex ante measure of the distortion created by the implicit guarantee extended to some U.S. financial institutions. Estimating the entire ex post realized cost of the various measures implemented by the U.S. Treasury, the Federal Reserve system, the FDIC, and other regulators in the face of the recent crisis is hard. Veronesi and Zingales (2010) estimate the cost to be between \$21 billion and \$44 billion, with a benefit of more than \$86 billion.

II. Size Effect in Bank Stock Returns

This section reports returns on size-sorted portfolios of commercial bank stocks. We also show the results of a cross-sectional regression of returns on firm characteristics that confirm the portfolio results.

We collect data on equity returns from the Center for Research in Security Prices (CRSP). There is no unique, fullproof way of identifying all of the U.S. commercial banks in CRSP. Many papers in the literature identify these banks manually over short samples. This is not feasible in our study. Instead, we define commercial banks as all firms with header Standard Industrial Classification (SIC) codes 60 or historical SIC code 6712. This definition ensures that bank holding companies are consistently included in our sample. Bank holding companies need to be included in this definition because the banks that belong to a holding company are not publicly traded themselves.

We use the header SIC codes (HSICCD) on December 2013 rather than the historical SIC codes (SICCD) to identify these firms in the data.[1] When we screen the CRSP database for SICCDs equal to 60 or 67, several of the largest U.S. banks drop out of the sample, including BB&T Corp, Banc One Corp, Barclays, Citigroup, First Bancorp, and Sterling Bancorp. All of these banks are identified by the Federal Reserve as one of the largest 100 deposit-taking institutions in the United States. These banks clearly belong in our sample. The HSICCD screen correctly identifies these commercial banks throughout the sample, because the coding conventions were changed in the late 1990s. Conversely, we define nonfinancials as all firms excluding those with two-digit HSICCDs ranging from 60 to 67. The Internet Appendix includes a detailed discussion of these choices.[2]

We exclude data for all financial firms that are inactive and we also exclude financial firms that are not incorporated in the United States because these financial firms are subject to regulations both in the country of operation and the country of incorporation. Since these policies vary across countries, our focus on financial firms operating and incorporated inside the United States ensures that all firms in our analysis are subject to a uniform regulatory regime. Foreign firms are identified by share codes ending in 2 or 5.

We start by building portfolios of domestic commercial bank stocks. We employ the standard portfolio formation strategy of Fama and French (1993). We rank all bank stocks by market capitalization as of January of each year. The stocks are then allocated to 10 portfolios based on their market capitalization. We calculate value-weighted returns for each portfolio for each month over the next year. At the end of this exercise, we have monthly value-weighted returns for each size-sorted portfolio of banks.

The data start in January 1970 and end in December 2013. Only a small fraction of all banks that operate in the United States are publicly listed. For instance, for the years 2000 to 2008, data are available from CRSP for approximately 630 banks, as compared to more than 5,300 FDIC-insured banks operating in the United States over the same period. However, the largest 600 banks control more than 88% of all commercial bank assets in the United States. Most of these large banks are publicly listed. To the extent that small banks that are not publicly listed are very different from those that are, some of our results need to be qualified.

Presumably, the government cares about the entire balance sheet of banks, not just their equity. As a result, book value may be the better measure of size. We follow a similar strategy for forming portfolios of commercial bank stocks sorted by book value. We rank all bank stocks by book value as of December of each year. Book values for all bank stocks are obtained from the merged CRSP-COMPUSTAT database. We calculate value-weighted returns for each portfolio

[1] The complete panel of commercial banks that fall under our definition is available from the authors.

[2] The Internet Appendix is available in the online version of the article on the *Journal of Finance* website.

for each month over the next year. While our market capitalization results are based on 17,594 bank-years from 1980 to 2013, the book value results are based on only 14,403 bank-years. The reduction in the number of banks is primarily due to missing balance sheet data in the CRSP-COMPUSTAT merged data set.

While the CRSP data are available from 1926, our main sample of banks begins only in 1970, as there are not enough publicly traded commercial banks prior to 1970. In addition, data for book value of commercial banks are not available for a substantial number of banks in our portfolio in COMPUSTAT prior to 1980. Hence, our main sample for book value size-sorted portfolios begins only in 1980.

III. Size Effect in Normal Risk-Adjusted Bank Stock Returns

We start by adjusting the portfolio returns for exposure to the standard risk factors that explain cross-sectional variation in average returns on other portfolios of nonfinancial stocks and bonds. We do so by comparing the performance of the bank portfolio to the performance of a portfolio of nonbank stocks and bonds with the same exposure to normal risk factors. To do so, we use the Fama and French (1993) three-factor model. We find that small banks, measured either by market cap or book value, outperform the benchmark portfolio of bonds and stocks, while large banks underperform.

A bank manages a portfolio of bonds of varying maturities and credit risk.[3] Therefore, we also include two bond risk factors in addition to three stock risk factors,

$$\boldsymbol{f}_t = [market \;\; smb \;\; hml \;\; ltg \;\; crd], \tag{1}$$

where f_t is 5×1. The terms *market*, *smb*, and *hml* represent the returns on the three Fama-French stock factors, namely, the market, small minus big, and high minus low factors, respectively. The Fama-French stock factors are constructed using the six value-weighted portfolios of all stocks on NYSE, Amex, and NASDAQ (including financials) formed on size and book-to-market. We capture *market* using the value-weighted return on all NYSE, Amex, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate (from Ibbotson Associates). We use *ltg* to denote the excess returns on an index of 10-year bonds issued by the U.S. Treasury as our first bond risk factor. The U.S. 10-year Government Bond Total Return Index (*ltg*) is downloadable from Global Financial Data. We use *crd* to denote the excess returns on an index of investment grade corporate bonds, maintained by Dow Jones, as our second

[3] Flannery and James (1984) were the first to demonstrate a link between interest rate changes and common stock returns of commercial banks that depends on the maturity structure of their assets and liabilities. Longstaff and Myers (2009) also show that banks can be treated as active managers of fixed income portfolios.

bond risk factor. To compute excess returns, we use the one-month risk-free rate.[4]

A. Risk-Adjusted Returns on Commercial Bank Stock Portfolios

We regress monthly excess returns for each size-sorted portfolio on the three Fama-French stock factors and two bond factors. For each portfolio i we run the following time-series regression to estimate the vector of betas $\boldsymbol{\beta}_i$:

$$R^i_{t+1} - R^f_{t+1} = \alpha^i + \boldsymbol{\beta}^{i,\prime} \boldsymbol{f}_{t+1} + \varepsilon^i_{t+1}, \tag{2}$$

where R^i_{t+1} is the monthly return on the i^{th} size-sorted portfolio. Since all of the risk factors in $\boldsymbol{f}_t$ are traded returns, the estimated residuals in the time series-regression are estimates of the normal risk-adjusted returns $\widehat{R}^i_{t+1}$.

Market Capitalization. Table I provides the results of the regression specified in equation (2). The portfolios are ranked from smallest (1) to largest (10). Panel A reports the results based on sorting by market capitalization into deciles. The table reports the regression coefficients for each size-sorted portfolio, along with their statistical significance and adjusted R^2. Table I excludes the recent financial crisis.

The estimated intercepts decrease nearly monotonically with bank size from 1.94% for the first decile to −5.09% for the tenth decile. The negative alpha on the tenth decile is significantly different from zero at the 1% level. A significant share of this alpha is due to the very largest banks. We also split the highest decile in two bins, 10A and 10B. In the top half of the tenth decile (10B) the alpha is −5.60% (statistically significant at the 1% level), while it is only −3.13% in the bottom half of the tenth decile. If we split the top decile into three bins instead, the top 3.33% earns −6.78% per annum (not reported in the table, statistically significant at the 1% level). The top 3.33% accounts for more than 90% of the industry's market capitalization. Clearly, the largest U.S. commercial banks earn significantly negative risk-adjusted returns.

A long-short position that goes long \$1 in a portfolio of the largest market capitalization banks in decile 10 and short \$1 in a portfolio of the smallest market capitalization banks in decile 1 loses 7.03% over the entire sample. This return spread is statistically significant at the 5% level. The difference between 10A and 10B is 2.47% and accounts for 35% of the entire 7.03% spread between the first and last deciles. The total difference between 10B and 1 is 7.54%. When we split the top decile into three bins, the spread between the first decile and the top 3.33% in the market cap distribution is even larger: 8.72% (not reported in the table). When we exclude the largest banks, the differences in risk-adjusted returns are much smaller. The average normal risk-adjusted return on a 9-minus-2 position is −4.26% per annum, and −5.29% per annum for the 8-minus-3 portfolio.

[4] Data for the risk-free rate and the Fama-French factors were collected from Kenneth French's website. The Dow Jones Corporate Bond Return Index (*crd*) is downloadable from Global Financial Data.

Table I

Risk-Adjusted Returns on Size-Sorted Portfolios of Commercial Banks

This table presents estimates from OLS regression of monthly value-weighted excess returns on each size-sorted portfolio of domestic commercial banks on the three Fama and French (1993) stock and two bond risk factors. U.S. commercial banks are defined as all firms with HSICCD equal to 60 or historical SICCD equal to 6712. We exclude foreign banks with share codes ending in 2 or 5. *market*, *smb*, and *hml* are the three Fama-French stock factors: market, small minus big, and high minus low, respectively. *ltg* is the excess return on an index of long-term government bonds and *crd* is the excess return on an index of investment-grade corporate bonds. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Alphas are annualized by multiplying by 12 and are expressed in percentages. Standard errors are adjusted for heteroskedasticity and autocorrelation using Newey-West (1987) with three lags. The sample is 1970 to 2013 in Panel A, 1980 to 2013 in Panel B.

	Small	2	3	4	5	6	7	8	9	Large	10A	10B	10-1	10A-1	10B-1
							Panel A: Market Capitalization								
α	1.94	1.12	3.28*	−0.16	1.04	−0.14	−1.49	−2.01	−3.14*	−5.09***	−3.13	−5.60***	−7.03**	−5.06*	−7.54***
market	0.46***	0.55***	0.50***	0.59***	0.57***	0.68***	0.78***	0.81***	0.92***	1.22***	0.98***	1.24***	0.76***	0.52***	0.78***
smb	0.40***	0.42 ***	0.38***	0.41***	0.39***	0.44***	0.49***	0.44***	0.42***	−0.13*	0.14**	−0.15**	−0.53***	−0.26**	−0.55***
hml	0.50***	0.48***	0.44***	0.54***	0.49***	0.59***	0.69***	0.66***	0.65***	0.70***	0.58***	0.73***	0.19*	0.08	0.22**
ltg	−0.19	−0.07	−0.05	0.09	0.15*	0.06	0.14	0.24**	0.39***	0.19*	0.24**	0.24**	0.38**	0.43**	0.43***
crd	0.36*	0.23	0.29**	0.14	0.11	0.21*	0.04	−0.06	−0.16	−0.14	−0.04	−0.19	−0.49**	−0.40*	−0.55**
R^2	25.37	43.68	43.20	49.84	51.96	59.74	60.89	62.16	66.50	68.01	62.02	65.73	27.42	14.32	27.59
							Panel B: Book Value								
α	2.22	1.93	2.43	2.62	1.30	1.07	−1.40	−3.53	−5.21**	−5.70**	−4.59*	−6.14**	−7.92***	−6.81**	−8.36**
market	0.44***	0.46***	0.54***	0.54***	0.65***	0.79***	0.89***	0.88***	0.90***	1.07***	1.07***	0.83***	0.64***	0.64***	0.40***
smb	0.38***	0.38***	0.30***	0.38***	0.46***	0.54***	0.58***	0.60***	0.32***	0.02	0.01	0.04	−0.36***	−0.37***	−0.34***
hml	0.36***	0.39***	0.44***	0.51***	0.60***	0.75***	0.85***	0.85***	0.73***	0.68***	0.69***	0.42***	0.32**	0.33**	0.07
ltg	0.11	0.11	0.11	0.06	−0.02	0.19*	0.42***	0.45***	0.35***	0.05	0.04	0.22 *	−0.06	−0.07	0.11
crd	0.11	0.15	0.15	0.21*	0.32*	0.08	−0.13	−0.22	−0.14	0.20	0.17	0.17	0.09	0.07	0.06
R^2	34.01	39.85	39.29	47.66	49.13	57.53	62.47	57.68	63.47	56.53	50.03	40.27	22.54	19.90	9.73

The differences in risk-adjusted portfolio returns tend to be larger than the differences in raw portfolio returns, because larger banks are more levered and hence impute higher market betas to large bank stock portfolios. The market beta increases from 0.46 for the first decile to 1.22 in the last decile. However, this effect is attenuated by the lower credit risk exposure for the larger banks.

The second row of Table I reports the coefficient on excess market return, *market*, for each size-sorted portfolio. The market beta increases monotonically with bank size. Over the entire sample, a portfolio of large banks has a market beta of 1.22, as compared to a beta of 0.46 for a portfolio of the smallest banks. The largest banks were 2.65 times more exposed to market risk as compared to the smallest banks. This difference can be attributed largely to differences in leverage.

The loadings on *smb* and *hml* also depend systematically on size. We first look at the exposure to the size factor. Contrary to what one expects to find, over the entire sample the loading on *smb* increases from 0.40 in the first size decile to 0.42 in the ninth decile, while in the tenth decile the loading is −0.13. Clearly, the common variation in stock returns of banks along the size dimension is very different from that in other industries. A similar pattern holds true for the loadings on *hml*, which increase from 0.50 for the first portfolio to 0.70 for the last portfolio.

There is a clear size pattern in the loadings on the bond risk factors. The slope coefficient on the excess return on an index of 10-year bonds issued by the U.S. Treasury, *ltg*, is negative and statistically insignificant for small banks, and positive and almost always statistically significant for large banks. The loadings vary monotonically in size. A $1 long position in large banks and a $1 short position in small banks results in a net exposure of 30 cents to long-term government bonds over the entire sample. The results for the portfolios of large bank stocks seem largely consistent with the findings of Flannery and James (1984) for a value-weighted portfolio of large bank stocks. They interpret this bond factor loading as a measure of interest rate sensitivity resulting from the maturity mismatch between assets and liabilities. Small and mid-sized banks seem to be different.

On the other hand, the loadings on the credit risk factor, *crd*, are surprisingly small for large banks and positive for small banks. A long-large-banks-short-small-banks position delivers a net negative exposure to credit markets of 49 cents per dollar invested.

Book Value. Market cap measures size, but it also measures expected returns. Firms that generate more cash flows will tend to have higher market capitalization, but firms with lower expected returns, holding cash flows constant, also have larger market capitalization. As a result, Berk (1995) argues that there should be a relation between expected returns and market capitalization. Of course, this argument does not apply to other measures of size such as book value. For example, while market cap sorts are likely to be picking up liquidity effects, book sorts are not. A priori, there is no reason to expect a relation between book values and expected returns.

Panel B reports the results obtained by sorting by book value. The pattern in risk-adjusted returns is similar to that obtained when sorting by the market capitalization of banks. The risk-adjusted returns remain around 100 to 200 bps for the first six portfolios. The seventh portfolio posts average risk-adjusted returns of −140 bps. After that, the average risk-adjusted returns decline to −353 bps for portfolio 8, −521 bps for portfolio 9, and −5.70% for portfolio 10, which is significantly different from zero at the 5% level. The top 5% of banks by book value earn risk-adjusted returns that are even lower: −6.14% per annum.

A long-short position that goes long $1 in a portfolio of the largest banks in decile 10 and short $1 in a portfolio of the smallest banks in decile 1 loses 7.92% over the entire sample. This return spread is statistically significant at the 1% level. The difference between 10A and 10B is 1.55%, and accounts for 1/5 of the entire 7.92% spread between the first and last deciles. The average normal risk-adjusted return on a 9-minus-2 position is −7.14% per annum, and −5.96% per annum for the 8-minus-3 portfolio. These results are statistically significant at the 1% and 5% levels, respectively.

Larger banks have higher market betas, consistent with leverage increasing in size, although the increase is smaller than the difference in leverage suggests. However, the negative effect of higher market betas on risk-adjusted returns is partly offset by a strong inverse U-shaped pattern in the credit risk loading. The loading increases from 0.11 in the first portfolio to 0.32 in the fifth portfolio, and then declines to −0.14 in the ninth portfolio. Clearly, there is a strong connection between average risk-adjusted returns on bank stocks and the actual size of these banks as measured by book value. This is not the case for nonfinancials.

This size anomaly is quite robust. However, if we exclude bank holding companies from the sample, we do not find evidence of a size anomaly for the banking sector. That is not surprising, because these firms include some of the largest U.S. commercial banks.

B. Risk-Adjusted Returns on Portfolios of Nonfinancial Stocks

To make the results easily comparable, we sort all banks and nonfinancials into 10 size bins using the NYSE market capitalization decile breakpoints available from Ken French's web site. Table II reports the results, Panel A for commercial banks and Panel B for nonfinancials. By design, the banks and nonfinancials in each portfolio are roughly of the same size. The value-weighted risk-adjusted returns on banks in the last size bin are 7.35% lower than those in the first bin. The difference between the ninth and tenth bank portfolios is 3.91%, which confirms that the very largest banks that exceed the tenth NYSE decile breakpoint earn much lower risk-adjusted returns.

For nonfinancials, there is no evidence of a size anomaly. In fact, the value-weighted risk-adjusted returns on the last portfolio are now 3.38% higher than those on the first portfolio. Finally, by comparing Panel A and Panel B, we note

Table II

Risk-Adjusted Returns on Size-Sorted Portfolios Using the NYSE Market Capitalization Decile Breakpoints

Panel A (B) presents estimates from OLS regression of monthly value-weighted excess returns on each size-sorted portfolio of commercial banks (nonfinancials) on the three Fama and French (1993) stock and two bond risk factors. We use the NYSE market capitalization breakpoints available from Ken French's website. U.S. commercial banks are defined as all firms with HSICCD equal to 60 or historical SICCD equal to 6712. We exclude foreign banks with share codes ending in 2 or 5. Nonfinancials are defined as all firms excluding those with two-digit HSICCDs ranging from 60 to 67. *market*, *smb*, and *hml* are the three Fama-French stock factors: market, small minus big, and high minus low, respectively. *ltg* is the excess return on an index of long-term government bonds and *crd* is the excess return on an index of investment-grade corporate bonds. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Alphas are annualized by multiplying by 12 and are expressed in percentages. Standard errors are adjusted for heteroskedasticity and autocorrelation using Newey-West (1987) with three lags. The sample is 1970 to 2013.

	Small	2	3	4	5	6	7	8	9	Large	10-1
					Panel A: Commercial Banks						
α	0.88	−0.31	−0.08	−1.95	−2.53	−4.16**	−1.28	−4.29*	−3.44*	−7.35***	−8.23***
market	0.58***	0.67***	0.73***	0.83***	0.83***	0.92***	1.00***	1.09***	1.15***	1.03***	0.45***
smb	0.43***	0.49***	0.51***	0.45***	0.38***	0.15**	0.13*	−0.08	−0.10	−0.16**	−0.58***
hml	0.54***	0.61***	0.65***	0.61***	0.65***	0.67***	0.61***	0.66***	0.60***	0.75***	0.22*
ltg	−0.05	0.20**	0.13	0.29**	0.11	0.17	0.29***	0.16	0.35***	0.09	0.14
crd	0.23*	−0.03	0.07	−0.10	0.10	0.13	−0.05	0.10	−0.15	−0.42*	−0.65***
R^2	56.39	58.48	60.50	61.85	62.36	63.17	61.36	59.55	62.07	47.29	17.83
					Panel B: Nonfinancials						
α	−0.77	−1.82*	0.76	−0.83	0.04	−0.94	1.50	1.92	1.99	2.61	3.38**
market	0.94***	1.02***	1.03***	1.03***	0.99***	1.01***	1.03***	1.08***	1.02***	0.94***	−0.01
smb	0.95***	0.94***	0.90***	0.81***	0.68***	0.50***	0.44***	0.32***	0.20***	0.11*	−0.85***
hml	0.25***	0.12**	0.00	0.08*	−0.08*	−0.12***	−0.11**	−0.24***	−0.32***	−0.32***	−0.57***
ltg	−0.28***	−0.08	0.03	0.03	0.07	0.01	0.06	0.01	−0.04	−0.15*	0.13
crd	0.17	0.04	−0.04	−0.12**	−0.05	0.01	−0.09	−0.09	−0.06	0.05	−0.12
R^2	86.07	91.08	91.74	91.18	89.24	87.18	84.49	80.66	74.13	68.64	32.12

that the risk-adjusted returns on the largest banks are a full 9.96% per annum lower than those of nonfinancials of the same size.[5]

C. Characteristics Regression

The portfolio sorts reveal that the actual size of a bank measured by its book value seems to be a key determinant of bank stock returns: larger banks have lower returns. This is confirmed by running standard characteristics regressions. When we run a cross-sectional regression of average annual returns on firm characteristics (the log of market capitalization, the log of book value, book-to-market, and leverage), we obtain a large and significant negative coefficient for log book value (−2.23) and a positive coefficient for market capitalization (2.79). Thus, a 100% increase in book value above the sample average lowers annual returns by 223 bps for a typical bank, holding all variables, including market capitalization, fixed. These coefficients are significant at the 1% level. The detailed results are in the Internet Appendix.

This pooled regression explains 0.42% of the variation in annual returns. Leverage seems to have no additional explanatory power for returns. We obtain identical results when we exclude leverage from the regression. When we drop book value, the regression only accounts for 0.09% of the variation in annual returns. Hence, this size effect in bank stocks is very different from the "market capitalization" effect first documented by Banz (1981).

IV. Size Factor in Bank Stock Returns

The key to activating the collective bailout clause is common variation in bank payoffs.[6] We use principal component analysis to study the common variation. We uncover a bank-specific size factor that can help us understand and interpret these anomalies. The second principal component of normal risk-adjusted returns on size-sorted portfolios of bank stocks has loadings that depend monotonically on size. The covariance between the returns on size-sorted

[5] When sorting the nonfinancials into standard size deciles, there is a large positive risk-adjusted return of 15% per annum on the first portfolio that is driven entirely by the nonfinancial firms with the smallest market capitalization. This is not surprising. In 1980, the average market capitalization of a firm in the first portfolio is only \$22.8 million, compared to \$75.9 million for the banks in the first portfolio in 1980. Illiquid stocks earn abnormal returns (see, for example, Brennan and Subrahmanyam (1996)). However, when we use the NYSE breakpoints, this effect disappears completely, because the smallest nonfinancials do not significantly affect the value-weighted returns in the first portfolio. These size decile results are reported in the separate Internet Appendix.

[6] In a recent paper, Acharya and Yorulmazer (2007) and Farhi and Tirole (2012) explore the incentives for banks in this type of environment to seek exposure to similar risk factors. The government's guarantee creates complementarities in firm payoffs. In earlier work, Schneider and Tornell (2004) explain the currency mismatch on firm balance sheets in emerging markets endogenously by means of a bailout guarantee for the nontradeables sector. Ranciere and Tornell (2011) discuss how to design regulation in the context of government bailout guarantees. Panageas (2010b) explores the optimal taxation implications of bailouts.

Table III
Principal Components of Size-Sorted Commercial Bank Stock Returns

This table presents the loadings for the first and second principal components ($\boldsymbol{w}_1$, $\boldsymbol{w}_2$) extracted from the residuals of the regression specified in equation (3). U.S. commercial banks are sorted into deciles by market capitalization. The last row indicates the percentage of variation explained by each principal component. Standard errors in brackets are generated by bootstrapping from the data 10,000 times. First, we bootstrapped the returns for each size-sorted portfolio and the risk factors 10,000 times. For each bootstrapped sample, we regress the returns on the standard risk factors. We then compute the first and second principal components from the residuals of this regression. This results in 10,000 samples of the first and second principal components, which we use to compute the standard errors.

	Market Cap 1970–2013		Book Value 1980–2013	
Portfolio	$\boldsymbol{w}_1$	$\boldsymbol{w}_2$	$\boldsymbol{w}_1$	$\boldsymbol{w}_2$
Small	0.47	0.53	0.25	0.34
	[0.02]	[0.05]	[0.02]	[0.10]
2	0.34	0.24	0.27	0.28
	[0.02]	[0.03]	[0.02]	[0.07]
3	0.31	0.20	0.32	0.27
	[0.02]	[0.05]	[0.02]	[0.08]
4	0.32	0.12	0.31	0.19
	[0.02]	[0.02]	[0.01]	[0.06]
5	0.31	−0.02	0.39	0.16
	[0.01]	[0.04]	[0.01]	[0.05]
6	0.30	−0.12	0.39	−0.02
	[0.02]	[0.03]	[0.03]	[0.05]
7	0.31	−0.31	0.36	−0.16
	[0.01]	[0.04]	[0.01]	[0.09]
8	0.25	−0.41	0.37	−0.27
	[0.01]	[0.05]	[0.02]	[0.12]
9	0.24	−0.43	0.27	−0.27
	[0.01]	[0.04]	[0.01]	[0.11]
Large	0.24	−0.37	0.18	−0.71
	[0.01]	[0.05]	[0.02]	[0.17]
%	49.19	18.86	47.93	15.15

portfolios of bank stocks and the size factor can explain the size pattern in average risk-adjusted returns.

A. Constructing the Size Factor

We compute the residuals from the time-series regression of returns of each size-sorted portfolio on the equity and bond risk factors in equation (2). We extract the loadings for the principal components ($\boldsymbol{w}_1$, $\boldsymbol{w}_2$) and report the results in Table III. This table shows the loadings for the first two principal components. Together, these two principal components explain 66% of the residual variation over the entire sample. The numbers in brackets are standard errors generated by bootstrapping 10,000 samples. The first two columns in the table show results for market capitalization sorts; the last two columns show results for book sorts. The two sets of results are similar. We therefore focus on the results obtained using the market capitalization sort, as this sort provides more observations and hence the loadings are estimated more precisely.

The first principal component is a banking industry ("level") factor with roughly equal weights on all 10 portfolios. The second principal component is a size factor that loads positively on portfolios of small banks and negatively on portfolios of large banks. The loadings vary monotonically in size. This is a candidate risk factor because the loadings align with the average normal risk-adjusted returns that we want to explain.

Next, we take our $(T \times 10)$ matrix of estimated residuals, $\boldsymbol{\epsilon}_t$, and multiply it by the (10×10) loadings of the principal components to construct the asset pricing factors. The weights $(\boldsymbol{w}_1, \boldsymbol{w}_2)$ are renormalized to $(\widehat{\boldsymbol{w}}_1, \widehat{\boldsymbol{w}}_2)$ so that they sum to one. This results in a $(T \times 10)$ linear combination of the residuals. We focus on the first two principal components, denoted $PC_t^1 = \widehat{\boldsymbol{w}}_1' \boldsymbol{\epsilon}_t$ and $PC_{2,t} = \widehat{\boldsymbol{w}}_2' \boldsymbol{\epsilon}_t$, where $\widehat{\boldsymbol{w}_2}$ is given by

$$\begin{bmatrix} 0.53 & 0.24 & 0.20 & 0.12 & -0.02 & -0.12 & -0.31 & -0.41 & -0.43 & -0.37 \end{bmatrix}$$

This is a highly levered portfolio with a long position of \$53 in small banks and a short position of \$37 in large banks. The return on this portfolio investment has a monthly standard deviation of 8.25%.

The size factor is a natural candidate for explaining the size pattern in normal risk-adjusted returns, because the average normal risk-adjusted returns align with the covariance between the size factor (second principal component) and the returns on the portfolios. This is not the case for any of the other principal components. The second principal component is the only candidate factor, because the second principal component is the only one for which the covariances line up with the average excess returns, suggesting that the common variation in banks' stock returns captured by the second principal component can explain the size anomaly in bank stock returns.

To check whether the size factor actually explains the average normal risk-adjusted returns, we define a new independent variable. We take the $(T \times 10)$ matrix of returns for each of the size-sorted portfolio of banks and multiply it by the (10×1) loading of the second principal component. We denote the results of our multiplication by $R[PC_2]_{t+1} = \widehat{\boldsymbol{w}}_2 \boldsymbol{R}_t$, which is a $(T \times 1)$ vector of the returns weighted by the second principal component. Thus, for each month, the returns of each of the 10 portfolios are multiplied by their corresponding weights in the second principal component and added together. This portfolio is long in small banks and short in large banks. The weights of the portfolio are given by the second principal component loadings, renormalized to sum to one. We then run a time-series regression of the returns on the size-sorted bank portfolios on the equity and bond factors, as well as the size factor $R[PC_2]$:

$$R_{t+1}^i - R_{t+1}^f = \alpha^i + \boldsymbol{\beta}^{i,\prime} \boldsymbol{f}_{t+1} + \beta_{PC,2}^i R[PC_2]_{t+1} + \varepsilon_{t+1}^i. \tag{3}$$

The tail and normal risk-adjusted returns or alphas from this regression are presented in Table IV. Panel A corresponds to the precrisis sample (1970 to 2005) and Panel B to the whole sample (1970 to 2013). In Panel A, we want to use an ex ante measure of the risk price that excludes the effects of the crisis; the disaster model rationalizes this approach. The risk-adjusted returns on all portfolios are smaller than 250 bps over the entire sample once we account for the size factor. Not only does the magnitude of the alphas change, but nearly all of them are statistically insignificant. In addition, there is no discernible size-related pattern in these normal risk-adjusted returns. In Panel B, we show the risk-adjusted returns that obtain over the entire sample, which includes the crisis. For the very largest banks, we still see significantly negative risk-adjusted returns over the sample that includes the crisis.

B. What Is the Size Factor?

We define $PC_{2,t} = \widehat{\boldsymbol{w}}_2' \boldsymbol{\epsilon}_t$ as the normal risk-adjusted return on a portfolio that is long small banks and short large banks. The weights of the portfolio are given by the second principal component. Figure 1 plots the 12-month moving average (months $t-11$ through t) of the PC_2 series along with a plot of the index for industrial production. The units are monthly returns. Recall that this portfolio is levered almost 10-to-1. The gray-shaded regions represent NBER recessions and the light-shaded regions represent banking crises. The NBER recession dates are published by the NBER Business Cycle Dating Committee. The dates for the Mexico and long-term capital management (LTCM) crises come from Kho, Lee, and Stulz (2000) and the FDIC (for the developing country debt crisis of 1982).

The size factor, which by construction is orthogonal to the bond and equity pricing factors, declines during recessions and financial crises. Moreover, it is very sensitive to large slowdowns in the growth rate of industrial production. We plot a backward-looking 12-month moving average, which explains why the returns appear to drop a couple of months after the start of the NBER recessions. The returns also tend to increase before the end of the NBER recession.

There are two exceptions to this cyclical pattern. The first is the double-dip recession in the early 1980s. Small bank stocks were already recovering from the huge declines suffered relative to large bank stocks, and hence starting from very low valuations, when the second recession started. The second is the 2001 recession in the wake of the LTCM crisis. Moreover, in 2001, the NBER chose the starting point of the recession well after the decline in industrial production started (in other recessions, the starting date coincides with the decline in industrial production). On average, during recessions, this normal risk-adjusted return drops by an average of 3.34% per month or 40.08% per annum. During the most recent recession, which coincides with a financial crisis, the levered size factor lost more than 100% of its value after adjusting for risk exposure.

Table V, Panel A shows the value at the trough of the NBER cycle (the end of the banking crisis) of \$100 invested at the peak of the NBER cycle (the start

Table IV

Size-Factor-Adjusted Returns for Size-Sorted Portfolios of Commercial Banks

This table presents estimates from OLS regression of monthly value-weighted excess returns on each size-sorted portfolio of U.S. commercial banks on the three Fama and French (1993) stock and two bond risk factors, and the second principal component weighted returns. *mkt*, *smb*, and *hml* are the three Fama-French factors: the market, small minus big, and high minus low, respectively. *ltg* is the excess return on an index of long-term government bonds and *crd* is the excess return on an index of investment-grade corporate bonds. R^{PC_2} is the time-series of the returns of the size-sorted portfolios weighed by the loadings of the second principal component $\widehat{w}_2$. The weights of the second principal component have been renormalized so that they sum to one. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The alphas have been annualized by multiplying by 12 and are expressed in percentages. Standard errors are adjusted for heteroskedasticity and autocorrelation using Newey-West (1987) with three lags. In Panel A, the weights of the second principal component are computed over 1970 to 2005 and the sample is 1970 to 2005. The annualized price of risk is 11.14% in the sample ending in 2005. In Panel B, the weights of the second principal component are computed over 1970 to 2013 and the sample is 1970 to 2013. The annualized price of risk is 9.36% in the sample ending in 2013. The last line in each panel shows the loadings on the size factor PC_2 times the risk price.

	Small	2	3	4	5	6	7	8	9	Large	10A	10B
Panel A: Precrisis Sample (1970–2005)												
Risk-adjusted returns	0.55	0.62	3.92**	0.94	2.49	1.66	1.67	0.75	−0.02	−1.23	−0.77	−1.76
PC_2	0.15***	0.07***	0.07**	0.05***	−0.01	−0.03*	−0.06***	−0.09***	−0.11***	−0.12***	−0.12***	−0.12***
Size	1.71	0.82	0.78	0.50	−0.07	−0.29	−0.69	−1.00	−1.20	−1.37	−1.34	−1.37
Panel B: Full Sample (1970–2013)												
Risk-adjusted returns	0.06	0.26	2.56**	−0.58	1.12	0.27	−0.38	−0.57	−1.62	−3.78**	−1.72	−4.31**
PC_2	0.30***	0.14***	0.12***	0.07***	−0.01	−0.07***	−0.18***	−0.24***	−0.25***	−0.21***	−0.23***	−0.21***
Size	2.85	1.30	1.10	0.65	−0.12	−0.62	−1.69	−2.20	−2.32	−2.00	−2.15	−1.97

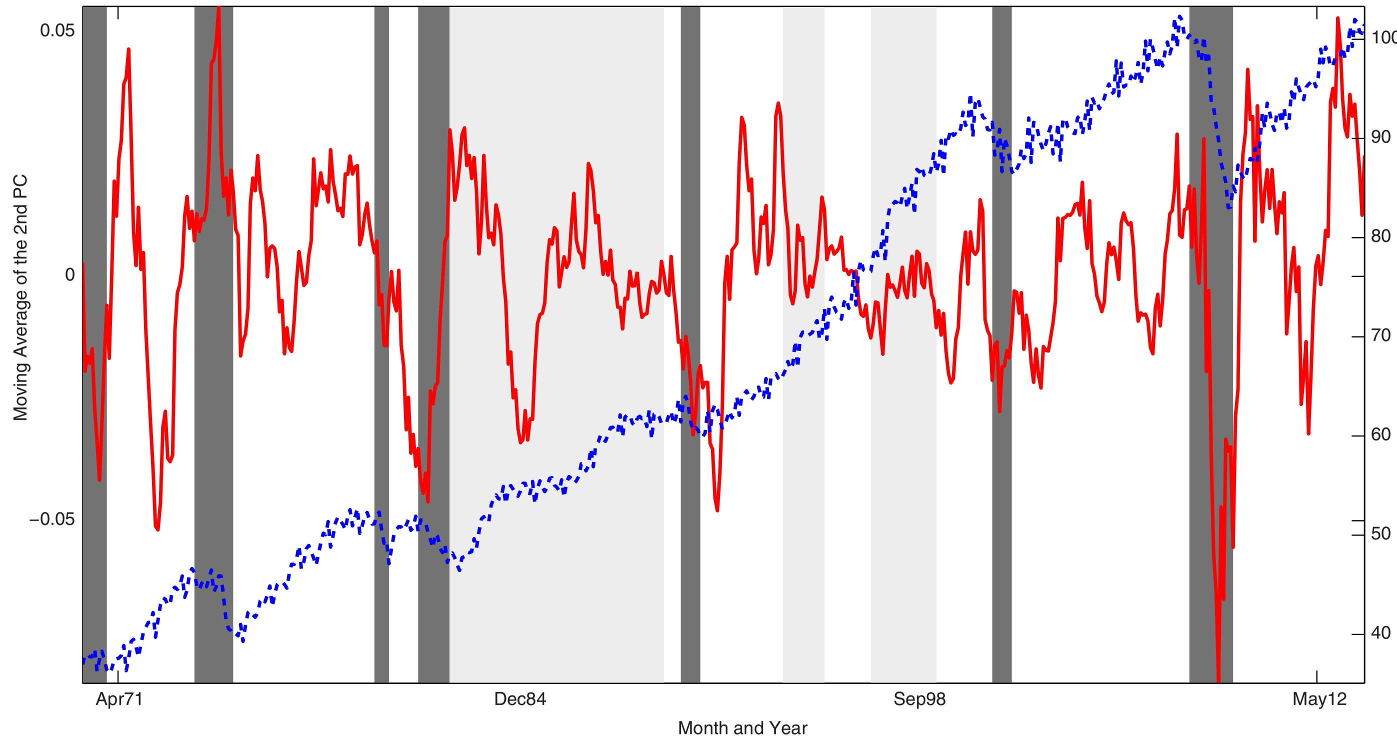


Figure 1. Size factor in normal risk-adjusted returns of commercial banks The solid line plots the 12-month (backward-looking) moving average (months $t - 11$ through t) of the time-series of the weighted sum of the residuals from the OLS regression of monthly value-weighted excess stock returns for each size-sorted portfolio of U.S. commercial banks on the Fama-French and bond risk factors. The weights are given by the second principal component and sum to one. The dashed line represents the index of industrial production. The gray-shaded regions represent NBER recessions and the light-shaded regions represent banking crisis. The NBER recession dates are published by the NBER Business Cycle Dating Committee. The dates for the Mexico and LTCM crisis are obtained from Kho, Lee, an Stulz (2000) and the FDIC (for the developing country debt crisis of 1982). The left axis references the moving average of the residuals and the right axis references the index of industrial production.

Table V

Cumulative Return on Second Principal Component Portfolio in Recessions and Financial Crises

This table shows the value of $100 invested in a portfolio that goes long in small commercial banks and short in large commercial banks. The weights of the portfolio are given by the second principal component, renormalized so that they sum to one ($\widehat{w}_2$). $100 is invested in this portfolio at the "Start" date and its value, given in the third and fourth columns, is measured on the "End" date. The column labeled *Value* represents the value of $100 invested at the peak (or start of the crisis) as of the trough (or end of the crisis) on this portfolio and the column labeled *Hedged Value* represents the normal-risk-adjusted returns on this portfolio. The average is computed for all NBER recessions using NBER dating conventions. The bottom panel shows the value of a $100 investment n months into the recession.

Panel A: Portfolio Value at NBER Trough

Start	End	Value	Hedged Value
01: 1970	11: 1970	14.00	55.14
11: 1973	03: 1975	91.14	115.07
01: 1980	11: 1982	24.18	36.75
07: 1990	03: 1991	66.13	60.85
03: 2001	11: 2001	110.29	82.67
12: 2007	06: 2009	−12.58	−1.44
Average		48.86	58.17

Panel B: Average Portfolio Value n Months after NBER Peak

	Value	Hedged Value
Month 1	113.12	101.84
Month 2	94.42	93.31
Month 3	112.51	88.19
Month 4	98.11	84.30
Month 5	94.05	83.15
Month 6	101.77	81.45
Month 12	63.39	59.92

of the banking crisis) in the size portfolio. The third column reports the dollar value without risk adjustment. The fourth column reports the dollar value after subtracting the performance of a benchmark portfolio with the same exposure to the bond and equity factors ($\$100 + x$, the cumulative return of $x\%$ in excess of the benchmark portfolio). This is the return on a portfolio that is hedged to have zero betas with respect to the standard risk factors. On average, the unhedged size portfolio loses $36.61 during a recession or banking crisis. The hedged strategy loses more than $40.08 per recession. As is clear from Panel B, the largest losses are concentrated in the first 12 months of the NBER recessions. Moreover, this portfolio (both hedged and unhedged) experienced steep declines during the developing country debt and LTCM crises. Panel B in Table V shows the average value of the portfolio n months into a recession. The hedged portfolio gradually drops more in value. Twelve months after the peak it has lost almost $41 dollars of its value.

The size factor appears to be a reliable measure of bank-specific tail risk. During the most recent U.S. recession, a full-fledged banking crisis, the hedged size portfolio of commercial banks lost more than 100 cents on the dollar (see Table V). This is not a surprise. In 2008, 18% of the commercial banks in the first market capitalization decile were delisted, followed by another 30% in 2009. We also went back to 1926 by including all financial firms in our sample. During the Great Depression (NBER recession dates), the hedged size portfolio of all financials was trading at −44 cents at the end of the recession per $100 invested at the peak. We did not find a similar cyclical pattern in the second principal component of nonfinancials.

In the data, there is a strong connection between the business cycle and the incidence of banking panics. We examine U.S. banking panics starting in 1873, as well as NBER business cycle peaks and troughs. Except for the first banking panic, all of these occur during the contraction phase of the U.S. business cycle. The dates of the banking panics come from Gorton (1988, p. 223) and Wicker (1996, p. 155). Details are provided in the Internet Appendix. This is not the case for nonfinancials. Giesecke et al. (2011) examine 150 years of U.S. corporate history and find a weak relation between the business cycle and corporate bond defaults.

C. Alternative Explanations

Large idiosyncratic shocks can cause bank failures. If the volatility of these shocks increases more in recessions for small banks, that could explain some of our findings. Smaller banks are much more exposed to idiosyncratic risk than large banks, but the amount of idiosyncratic risk exposure of small banks does not seem to increase very much during recessions. During NBER recessions, the standard deviation ranges from 30.11% for the smallest banks to 23.86% for the largest banks as compared to 36.06% and 19.19%, respectively, in the full sample. Details are in Appendix A. Hence, the largest percentage point increase in volatility during recessions—from 19.19% to 23.86%—is noted for the largest banks. For the smallest banks, the idiosyncratic volatility decreases by 5.85%. There is no evidence to suggest that the cyclicality of the size factor is due to idiosyncratic bank risk. While smaller banks are more exposed to idiosyncratic risk, we do not see large increases in this type of risk during recessions.

There is no evidence that business cycle variation in cash flows can explain our findings. If anything, the evidence suggests that large financial institutions are more exposed to business cycle risk. Boyd and Gertler (1993) analyze the impact of size on the performance of banks as measured by accounting data. They show that increased competition and financial innovation have induced the largest banks to participate in riskier investments. We examine bank performance during the last two recessions by studying the Quarterly Banking Reports issued by the FDIC, and find that small banks tend to outperform large banks during recessions along several dimensions: return on equity, returns on assets, loan losses, and several other measures. Appendix B contains the details. We analyze the data in the report for the first three quarters of 2001, which corresponds to the recession dates provided by NBER.

D. Size and Co-Skewness

By granting the shareholders of large bank stocks a menu of out-of-the-money put options, the government reduces the negative co-skewness of large bank stock returns. Consistent with our interpretation of the size factor, we find that large bank stock returns have significantly less co-skewness with the market than small banks. We measure co-skewness by adding the squared market return as a risk factor. Table VI reports the results. We find large and statistically significant positive differences in the loadings on the squared market return between the upper and lower deciles. Given that the largest commercial banks use more leverage, this finding is surprising, unless we consider the effect of government guarantees. Harvey and Siddique (2000) show that co-skewness is priced in stocks.[7] Finally, we also find that small bank stock returns are significantly more exposed to the Fama-French momentum factor than large bank stock returns. This is not surprising given that Harvey and Siddique (2000) relate the momentum effect to systematic skewness.

V. The Pricing of Bank Tail Risk and the Government

The average return of this size factor is the price of banking tail risk insurance. For individual banks, we measure the effect on the cost of equity capital as the loading on this factor times this risk price. When the total effect is negative, we refer to this as a tail risk subsidy; otherwise, we refer to it as a tail risk tax. Of course, this only measures the impact on equity. Since these institutions are highly levered, the direct effect on the overall cost of capital may be small, but the indirect effect is not: since shareholders are last in line, the implied subsidy to other bank creditors is even larger.

This section examines how bank-specific tail risk is priced in the stock market, and relates it to both the regulatory regime and government announcements.

A. Size of Largest Banks

The events immediately after the collapse of Lehman in September 2008 are in line with the commonly held view that the U.S. government and monetary authorities are reluctant to let large financial institutions fail collectively, even though they may occasionally be willing to let individual institutions fail. For example, over the course of the recent financial crisis, the Federal Reserve made emergency loans totaling about $9.99 trillion to 10 of the largest U.S. financial institutions, which accounted for 83% of the emergency credit extended to

[7] However, the Harvey and Siddique (2000) skewness factors constructed from nonfinancial as well financial stocks cannot account for the variation in risk-adjusted returns on banks.

Table VI
Betas to Market Squared

This table presents the estimates from an OLS regression of monthly value-weighted excess returns on each size-sorted portfolio of U.S. commercial banks on the three Fama and French (1993) stock and two bond risk factors, and $market^2$. *market*, *smb*, and *hml* are the three Fama-French stock factors: the market, small minus big, and high minus low, respectively. *ltg* is the excess return on an index of long-term government bonds and *crd* is the excess return on an index of investment-grade corporate bonds. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The alphas have been annualized by multiplying by 12 and are expressed in percentages. Standard errors are adjusted for heteroskedasticity and autocorrelation using Newey-West (1987) with three lags. The sample is 1970 to 2013.

	Small	2	3	4	5	6	7	8	9	Large	10A	10B	10-1	10A-1	10B-1
Panel A: Market Capitalization															
market	0.44***	0.54***	0.49***	0.57***	0.56***	0.67***	0.77***	0.81***	0.92***	1.22***	0.99***	1.25***	0.78***	0.55***	0.80***
smb	0.39***	0.42***	0.37***	0.39***	0.38***	0.44***	0.48***	0.44***	0.43***	−0.12*	0.15**	−0.14*	−0.51***	−0.23**	−0.53***
hml	0.50***	0.48***	0.43***	0.53***	0.49***	0.59***	0.69***	0.66***	0.66***	0.70***	0.59***	0.73***	0.21*	0.09	0.23**
ltg	−0.18	−0.07	−0.05	0.10	0.16*	0.06	0.15	0.24**	0.39***	0.19*	0.23**	0.24**	0.37**	0.42**	0.42***
crd	0.37**	0.23	0.30**	0.15	0.13	0.21*	0.05	−0.06	−0.17	−0.14	−0.05	−0.20	−0.51**	−0.42*	−0.57***
$market^2$	−0.86*	−0.29	−0.54	−0.93**	−0.86**	−0.17	−0.53	0.11	0.44	0.36	0.61	0.38	1.22**	1.46*	1.23**
R^2	25.60	43.63	43.32	50.39	52.48	59.68	60.97	62.10	66.54	68.00	62.12	65.72	27.88	15.08	28.03
Panel B: Book Value															
market	0.41***	0.42***	0.50***	0.52***	0.64***	0.78***	0.91***	0.92***	0.92***	1.07***	1.07***	0.85***	0.67***	0.66***	0.44***
smb	0.36***	0.35***	0.26***	0.36***	0.45***	0.54***	0.59***	0.63***	0.34***	0.02	0.00	0.06	−0.34***	−0.36***	−0.30**
hml	0.34***	0.36***	0.40***	0.49***	0.59***	0.75***	0.86***	0.88***	0.74***	0.68***	0.68***	0.44***	0.34**	0.35**	0.10
ltg	0.12	0.12	0.13	0.07	−0.01	0.19*	0.42***	0.44***	0.34***	0.05	0.04	0.21*	−0.08	−0.08	0.09
crd	0.12	0.18	0.17	0.23*	0.32*	0.08	−0.13	−0.24	−0.15	0.20	0.18	0.16	0.08	0.06	0.04
$market^2$	−1.03**	−1.55***	−1.69**	−0.80*	−0.53	−0.12	0.47	1.36***	0.71	0.00	−0.32	0.74	1.03	0.71	1.77***
R^2	34.82	41.97	41.43	48.09	49.17	57.43	62.48	58.40	63.65	56.42	49.94	40.35	22.82	19.90	10.84

all U.S. institutions.[8] Moreover, even if regulators are willing to let these large banks fail, uncertainty about the resolution regime for distressed banks clearly favors the creditors and shareholders of large financial institutions.

Consistent with this view, even within the highest market capitalization decile of commercial banks, we find a strong negative relation between the market capitalization of individual firms relative to GDP and the loading on the size factor. We choose banks that are in portfolio 10 in each year of our sample and then compute the loadings on PC_2 over the subsequent five-year window. As individual banks grow larger relative to GDP over time, their loadings on this size factor clearly tend to increase. The slope coefficient in the regression of PC_2 loadings on market capitalization/GDP is 0.032, meaning that a 100% increase in the size of market capitalization relative to GDP raises the loading by 0.032 (t-stat is 5.9) in absolute value, or equivalently, increases the tail risk subsidy by 35 bps per annum, using the precrisis market price of 11.14%. We find a similar relation in the ninth decile, but not in the other deciles. The PC_2 itself is computed over the full 1970 to 2013 sample.

B. Regulatory Regime

We want to relate the pricing of tail risk in the precrisis sample, as captured by the size factor, to the regulatory regime of different banks. Commercial banks and GSEs benefit from special provisions: deposit insurance,[9] access to the discount window at the Federal Reserve and other special lending facilities in the case of commercial banks, and widely acknowledged debt guarantees in the case of GSEs. Foreign banks and investment banks do not enjoy the same level of protection.

Table VII compares the results for a value-weighted index of commercial banks, investment banks, foreign banks, and GSEs. The first row reports the value-weighted average market capitalization for each index. For foreign banks, this only includes the market capitalization of U.S. listed shares.[10] Investment and foreign banks do not benefit from the tail risk subsidy to commercial banks, but the GSEs (Fannie Mae and Freddie Mac) clearly do. Over the entire sample, the subsidy to commercial banks is 1.18% and the subsidy to GSEs is 2.58%. This subsidy is computed as the loading on PC_2 times the risk price (11.14%). The loadings on PC_2 are computed over 1970 to 2005. The loadings on $R[PC_2]$ are negative and statistically significantly different from zero for commercial banks and GSEs at the 1% level, but the loadings on

[8] Data are from the Term Auction Facility (TAF), which provided emergency loans to commercial banks, the Primary Dealer Credit Facility (PDCF), which provided emergency loans to investment banks and other broker-dealers and typically do not have access to Fed funds, and the Term Securities Lending Facility (TSLF), which allowed financial firms to borrow Treasury securities.

[9] The FDIC Improvement Act of 1991 limits the protection of creditors, but it provides a systemic risk exception.

[10] The worldwide market-cap for just the six largest banks included in the index of foreign banks is $330.21 billion in 2010.

Table VII

Bank Tail Risk Pricing for Investment Banks, Foreign Banks, and GSEs

This table presents the estimates from OLS regression of monthly excess returns on a value-weighted index of U.S. commercial banks, U.S. investment banks, and foreign banks on the Fama-French stock factors, bond factors, and the second principal component weighted returns. The table also reports results for individual banks. U.S. commercial banks are defined as all firms with HSICCD equal to 60 or historical SICCD equal to 6712. U.S. Investment banks are those with two-digit HSICCD of 62. We exclude foreign banks with share codes ending in 2 or 5. For individual banks, the longest available sample for each bank till 2005 is selected. The starting year for each bank is mentioned in parentheses under the name of the bank. PC_2 is the time-series of the returns of the size-sorted portfolios weighed by the loadings of the second principal component $\widehat{w}_2$. The weights of the second principal component have been renormalized so that they sum to one. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are adjusted for heteroskedasticity and autocorrelation using Newey-West (1987) with three lags. The implicit subsidy is the risk price (the 1970 to 2005 average return on PC_2 11.14%) times the negative of the loading on PC_2. The risk price is fixed at 11.14% in the different subsamples, but we recompute the loading on PC_2 in Panel B over shorter subsamples. The full sample is 1970 to 2005.

	Index of Banks			Individual Banks								
	Commercial	Investment	Foreign	BoA	Citi	WFC	FNM	FRE	GS	LEH	ML	MS
Mkt Cap (Jan 05)	118.57	24.12	44.71	187.30	254.56	52.22	34.33	55.78	61.25	103.71	62.48	44.95
start				(1970)	(1986)	(1970)	(1970)	(1989)	(1999)	(1994)	(1971)	(1986)
Panel A: Precrisis Sample (1970–2005)												
market	0.78***	1.53***	1.16***	0.90***	1.38***	0.65***	0.79***	0.72***	1.46***	1.36***	1.70***	1.51***
smb	0.01	0.24**	0.65***	−0.14	−0.44***	−0.46***	−0.31**	−0.15	0.00	−0.28	0.18	−0.17
hml	0.42***	0.05	0.58***	0.40**	0.15	0.30**	0.18	0.34*	−0.35*	−0.33	0.07	−0.23
ltg	−0.03	0.06	−0.18	0.03	−0.08	−0.10	1.30***	1.11**	0.85	−0.01	−0.30	−0.13
crd	0.25**	−0.21	0.36	0.35	0.32	0.35	−0.23	−0.35	−0.45	0.74	0.37	−0.25
PC_2	−0.11***	−0.04**	0.04	−0.32***	−0.14*	−0.32***	−0.16***	−0.30***	−0.18	−0.33**	−0.10	−0.13*
size	1.18	0.48	−0.42	3.51	1.52	3.61	1.78	3.37	2.03	3.63	1.06	1.50
Panel B: Precrisis												
1990–2005												
PC_2	−0.09***	−0.06***	0.04	−0.26**	−0.13*	−0.35***	−0.28***	−0.32***	−0.18	−0.33**	−0.12*	−0.15**
size	1.00	0.65	−0.50	2.91	1.48	3.89	3.15	3.51	2.03	3.63	1.30	1.70
2000–2005												
PC_2	−0.16***	−0.01***	0.00	−0.35***	−0.17*	−0.34***	−0.53***	−0.35***	−0.21	−0.46**	−0.17	−0.32**
size	1.75	0.12	−0.02	3.91	1.89	3.82	5.93	3.87	2.39	5.09	1.92	3.55

$R[PC_2]$ are much smaller (investment banks) or positive (foreign banks) and not statistically significant.

Table VII also shows the same results for some of the largest commercial banks and investment banks in the United States. Panel A reports the results for the entire sample excluding the crisis. The tail risk subsidy is largest for the large commercial banks. For BoA (from 1973 to 2005), we estimate a tail risk subsidy of 3.51% per annum, for Wells Fargo (from 1970 to 2005) it is 3.61%, and for Citibank (from 1986 to 2005) it is 1.52%. For investment banks, the loadings on PC_2 are mostly not statistically significant, except for Lehman. In contrast, BoA, Citi, Wells Fargo, Freddie Mac, and Fannie Mae have loadings that are negative and statistically significantly different from zero at the 5% level or better.

We also report the subsidies over 1990 to 2005 in Panel B. As above, the loadings on PC_2 are computed over 1970 to 2005. The subsidies are computed as the loading on PC_2 times the precrisis risk price (11.14%). Over this period the subsidy to commercial banks is 1.00%, but the subsidy to GSEs is 3.33%. This number is the unweighted average of the subsidy for Fannie Mae and Freddie Mac.

As a benchmark, we also compute the loading on $R[PC_2]$ for an index of hedge fund returns. Hedge funds do not benefit from the umbrella extended to large banks. We use the HFRI fund-weighted hedge fund index. These results are not reported. Over the entire sample (1991 to 2005) the loading for hedge fund returns on $R[PC2]$ is 0.02 (t-statistic 2.31,) and it is 0.05 (t-statistic 2.02) over 2000 to 2005. Hence, as expected, hedge funds face a tail risk tax, because the loadings are positive, just like small banks.

These results lend some support to a government-based interpretation of the size factor, as commercial banks and GSEs benefit from more extensive government guarantees than other financial institutions.[11]

C. Elimination of Glass-Steagall Act

The Glass-Steagall Act of 1933 effectively separated U.S. commercial banking from investment banking. The provisions of this act preventing bank holding companies from owning financial companies were repealed in 1999. Its repeal allowed large commercial banks to originate and trade collateralized debt obligations.

After 2000, the tail risk subsidy to commercial banks and GSEs nearly doubled to 1.75% and 4.90%, respectively. These numbers are derived by multiplying the loadings with the same risk price (11.14%) computed over the entire sample, and hence are valid only if the risk price is constant across these subsamples. These changes are large even when taking into account the statistical uncertainty. For example, the loading for commercial banks increased by almost three standard errors from −0.09 (with a standard error of 0.03) to −0.16.

[11] The GSEs and foreign banks were suggested to us by Martin Bodenstein.

The loadings for the largest banks increased dramatically in the last decade. The largest subsidies are collected by Fannie Mae (5.93% per year), Lehman, (5.09%), and Freddie Mac (3.82%), in spite of their smaller size. All of these banks were building up substantial exposure to the subprime mortgage market during this period. Note that there is no mechanical connection between our size factor and the subprime exposure, since we exclude the financial crisis. In addition, Fannie Mae, Lehman, and Freddie Mac are themselves excluded from the sample when we compute the size factor. Exposure to the size factor seems to be a good yardstick of systemic risk exposure.

D. Announcement Effects

In September 1984, the Comptroller of the Currency announced a list of 10 banks deemed too big to fail. We examine the pricing of the financial tail risk embedded in the stocks of these 10 banks around this announcement date. The Internet Appendix lists all the announcement dates.

Precrisis Announcement Dates. We also look at six other announcement dates listed by Kho, Lee, and Stulz (2000). Table VIII reports the results. We report regressions for windows of 30, 45, 60, 90, and 105 days around the announcement date. Panel A reports results from a pooled regression for all seven announcement dates. In the 30-day window after the Comptroller announcements, the loading increases by 0.12. This amounts to an annualized 1.33% tail risk subsidy per year. This effect gradually decreases as we increase the event window. We find slightly smaller effects for the LTCM, Brazilian, Mexican, and South Korean crises. The average effect in a 30-day window is a 33 bps per annum (0.03 times 11.14%) increase in the tail risk subsidy. This average effect is roughly constant across the windows. These effects are economically and statistically significant.

Crisis Announcement Dates. In the crisis sample, we identify announcements that increased the likelihood of a bailout for all banks, and for large banks, and we also look at events that decreased the likelihood of a bailout. These are listed in the Internet Appendix.

Table VIII, Panel B looks at the financial crisis announcements. Only the positive announcements for large banks have an economically and statistically significant effect on the pricing of tail risk. The tail risk subsidy for the too-big-to-fail banks increases by 78 bps (per annum) in a 30-day window around these announcements. The other announcements have small or negative effects that are statistically insignificant.

VI. Recovery Rates and Equilibrium Pricing of Tail Risk in the Banking Sector

To help us interpret our empirical findings, we use a stylized dynamic asset pricing model with time-varying probability of banking panics that reproduces the size anomalies, as well as the size factor in returns. The driving force is the size variation in recovery rates. The model yields a key testable prediction: a

Table VIII

Bailout Announcements

This table presents the results of OLS regression $R_t^{TBTF} - R_t^f = \alpha + \beta_1 PC_2 + \beta_2 PC_2 D + \epsilon$, where $TBTF$ represents the value-weighted return of the 10 banks that were declared too big to fail by the Comptroller of Currency in September of 1984, PC_2 represents the daily return of the portfolio that goes long in small U.S. commercial banks and short in large U.S. commercial banks, the weights for the portfolio are given by the second principal component and sum to one, and D represents a dummy variable that equals one after the announcement date and zero otherwise. The regression is estimated over a 30-, 60-, 90-, and a 105-day window around the announcement date. A seven-day window around the exact announcement date is excluded from the sample while estimating coefficients. Dates for the announcements are from O'Hara and Shaw (1990) and Kho, Lee, and Stulz (2000). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are adjusted for heteroskedasticity and autocorrelation using Newey-West (1987).

Coeff	30D	45D	60D	90D	105D
			Panel A: Precrisis Announcements		
			9/19/1984; Comptroller of Currency		
PC_2	-0.19^{***}	-0.19^{***}	-0.20^{***}	-0.21^{***}	-0.20^{***}
PC_2D	-0.12^{*}	-0.05	-0.03	-0.02	0.00
			9/24/1998; LTCM		
PC_2	-0.22^{***}	-0.23^{***}	-0.24^{***}	-0.23^{***}	-0.24^{***}
PC_2D	-0.05	-0.05	-0.05	-0.05	-0.05^{*}
			9/15/1998; Brazilian Crisis		
PC_2	-0.24^{***}	-0.25^{***}	-0.25^{***}	-0.26^{***}	-0.26^{***}
PC_2D	-0.03	-0.03	-0.03	-0.02	-0.03
			10/08/1998; Brazilian Crisis		
PC_2	-0.24^{***}	-0.24^{***}	-0.25^{***}	-0.25^{***}	-0.25^{***}
PC_2D	-0.08^{*}	-0.09^{**}	-0.09^{***}	-0.08^{***}	-0.06^{**}
			11/13/1998; Brazilian Crisis		
PC_2	-0.27^{***}	-0.26^{***}	-0.27^{***}	-0.25^{***}	-0.25^{***}
PC_2D	-0.06	-0.05	-0.03	-0.05	-0.03
			11/14/1997; South Korean Crisis		
PC_2	-0.27^{***}	-0.27^{***}	-0.27^{***}	-0.26^{***}	-0.26^{***}
PC_2D	-0.01	0.00	0.00	0.00	0.00
			01/25/1995; Mexico Crisis		
PC_2	-0.17^{*}	-0.11^{*}	-0.12^{***}	-0.15^{***}	-0.14^{***}
PC_2D	-0.06	-0.12^{*}	-0.08	-0.05	-0.05
			Pooled Regression		
PC_2	-0.24^{***}	-0.24^{***}	-0.24^{***}	-0.24^{***}	-0.24^{***}
PC_2D	-0.03^{**}	-0.04^{***}	-0.04^{***}	-0.04^{***}	-0.04^{***}
			Panel B: Crisis Announcements		
			Positive Announcements: All Banks		
PC_2	-0.17^{***}	-0.17^{***}	-0.17^{***}	-0.16^{***}	-0.16^{***}
PC_2D	0.00	-0.00	-0.01	-0.01	-0.01
			Positive Announcements: Large Banks		
PC_2	-0.11^{***}	-0.15^{***}	-0.16^{***}	-0.15^{***}	-0.14^{***}
PC_2D	-0.07^{***}	-0.04^{**}	-0.02	-0.02	-0.03^{*}
			Negative Announcements		
PC_2	-0.15^{***}	-0.15^{***}	-0.16^{***}	-0.16^{***}	-0.16^{***}
PC_2D	-0.01	-0.01	0.00	-0.01	-0.01

size factor in normal risk-adjusted returns on banking portfolios that is tied to the U.S. business cycle.

A. A Simple Model of the Size Anomaly in Bank Stock Returns

We adopt a version of models with time-varying probabilities of financial disasters proposed by Gabaix (2012) and Wachter (2013), which are extensions of the rare event models developed by Barro (2006) and Rietz (1988). The model produces a one-to-one relation between the average risk-adjusted returns and the financial disaster recovery rates. In our model, the stochastic discount factor has two components, a standard normal component and a disaster component:

$$M_{t+1} = M_{t+1}^{G} \times 1 \text{ in states without a financial disaster} \tag{4}$$

$$M_{t+1} = M_{t+1}^{G} \times M_{t+1}^{D} \text{ in states with a financial disaster,}$$

where M_{t+1}^{G} denotes the representative investor's intertemporal marginal rate of substitution in normal times, that is, in states without a disaster. We use p_t to denote the probability of a financial disaster.

In the simplest consumption-based capital asset pricing model (CCAPM) version of his model, Gabaix (2012) defines

$$\Delta \log C_{t+1} = g_C + \sigma \eta_{t+1} \text{ in states without a financial disaster} \tag{5}$$

$$\Delta \log C_{t+1} = g_C + \sigma \eta_{t+1} + \log F^{c} \text{ in states with a financial disaster,}$$

where $1 \geq F^{c} > 0$, and η_{t+1} is Gaussian white noise. We assume that agents have standard power utility defined over consumption with coefficient of relative risk aversion γ.

The dividend process of a portfolio of bank stocks of size i is given by

$$\Delta \log D_{t+1}^{i} = \Delta \log D_{t+1}^{i,G} \text{ in states without banking crisis}$$

$$\Delta \log D_{t+1}^{i} = \Delta \log D_{t+1}^{i,G} + \log F_{t}^{i} \text{ in states with banking crisis,}$$

where $\Delta \log D_{t+1}^{i,G}$ is the i.i.d. Gaussian component of dividend growth, and $1 \geq F^{i} > 0$ can be thought of as the recovery rate, that is, in the case a rare event is realized, a fraction F^{i} of the dividend gets wiped out (as in Longstaff and Piazzesi (2004) and Barro (2006)). The recovery rate varies across banks depending on size, in part because the realization of the rare event can trigger a collective bailout of larger banks, but not necessarily of smaller banks.

Following Gabaix (2012), the resilience of banks is defined as the marginal utility-weighted recovery rate in disaster states: $H_t^{i} = p_t E_t \left[M_{t+1}^{D} F^{i} - 1\right]$. In the CCPAM case, this would be $H_t^{i} = p_t E_t \left[(F^{c})^{-\gamma} F^{i} - 1\right]$. As the economy enters into a recession, p_t increases and the resilience of large banks H_t^{B} increases relative to that of small banks H_t^S if $F^B > F^S$. Indeed, we assume that the recovery rate $F^n > F^{n-1}$ increases monotonically in size.

The log expected return on asset i conditional on no disaster realization after adjusting for normal risk exposure becomes $\log E_t[\widehat{R}^i_{t+1}] = (r - h^i_t)$, where r denotes the rate of return on an asset with zero resilience, $\log R_t = \log E_t[M^G_{t+1}]^{-1}$, and h^i_t denotes $\log(1 + H^i_t)$. The proof is in Appendix IV.A. This implies that, in a sample without a disaster realization, the average normal risk-adjusted return will be given by

$$\log E\left[\widehat{R}^i_{t+1}\right] \approx (\bar{r} - \bar{h}^i), \tag{6}$$

where $\bar{h}^i = E[\log(1 + H^i)]$ denotes the average resilience and $\bar{r}$ denotes the average r. The difference in alphas in a sample without a rare event realization measures the differences in average resilience between different bank stock portfolios: $\log \alpha^B - \log \alpha^S = \bar{h}^S - \bar{h}^B$. Hence, we can interpret the difference between small and large bank portfolios in the normal risk-adjusted returns as measuring differences in the resilience of these bank portfolios to banking crises.

Quantitative Implications of CCAPM Model. We set the coefficient of relative risk aversion γ to five. We consider a per annum consumption drop of 5% ($F^C =$ 0.95) in the financial disaster state. This scenario matches the experience of all developed economies considered by Reinhart and Rogoff (2009) during banking crises. The authors document a cumulative drop of 5%. We set the average probability of a banking crisis to 13%, because the United States spent 13% of all years since 1800 in a banking panic according to Reinhart and Rogoff (2009).[12]

If the spread in recovery rates is 35 cents per dollar, then the difference in risk-adjusted returns ($\log \alpha^S - \log \alpha^B = \bar{h}^B - \bar{h}^S = 3.70\% - (-2.1\%)$) in a sample without disasters is equal to 5.8%. When the coefficient of relative risk aversion increases to 15, the spread increases to 8.9%.

Recovery Rates in the Data. There is strong empirical evidence for size-dependent variation in financial disaster recovery rates. In our sample (from 1970 to 2009), the average delisting rate for banks in the first market capitalization decile is 1.77%, compared to 0.018% for the ninth decile and 0% for the tenth decile. During 2008 alone, 18% of banks in the first decile were delisted, another 30% were delisted in 2009 and 10% were delisted in 2010. None of the commercial banks in the highest decile were delisted.

Size Factor. A key prediction of this model is that the variation in the probability of a financial disaster imputes common variation to the normal risk-adjusted stock returns along the size dimension, since we assumed that the recovery rate depends on size, even in a sample without disasters. The loadings on this common factor are proportional to $F^i - 1$. To see why, note that $\log(1 + H^i_t) \approx p_t E_t[M^D_{t+1} F^i - 1]$. This is a size factor because the loadings

[12] This matches 13 U.S. financial crises over 210 years with an average length of 2.1 years.

depend on the recovery rates and hence (by assumption) on size. The conditional normal risk-adjusted multiplicative risk premium on a long-short portfolio is given by the expression $\log E_t[\widehat{R}_{t+1}^B] - \log E_t[\widehat{R}_{t+1}^S] = h_{t+1}^S - h_{t+1}^B$. As p_t increases during recessions, the risk premium on this long-short portfolio becomes more negative. This variation in risk premia is the driving force. The size factor tracks the variation in p_t.

B. Mergers, Acquisitions, and Risk-Adjusted Returns

Suppose that only the very largest banks directly benefit from government guarantees. Our model does not predict that only those banks will have lower expected risk-adjusted returns. Because of the possibility of mergers, some of the effects will contaminate the expected risk-adjusted returns on smaller bank stocks.

In our model, a characteristic (the size of the bank) actually determines the financial disaster risk premium, because of the collective bailout guarantee for large banks. This creates an opening for arbitrage opportunities. Let us assume that there is a single critical size threshold. In this case, the low recovery rate ($F^i = \underline{F}$) applies for all bank portfolios with size below the cutoff. Also, suppose banks do not switch between portfolios as a result of growth, mergers, or acquisitions. For banks in portfolios above the cutoff, the higher recovery rate applies: $F^i = \overline{F}$. The baseline model predicts large, positive, but constant alphas for all the banks in size-sorted portfolios below the threshold, and much smaller negative alphas for all banks above the threshold. In that sense, the pattern we find in the data is surprising. However, this stark $(\underline{\alpha}, \overline{\alpha})$ outcome can only be an equilibrium if there are prohibitively large costs associated with merging and acquiring banks.

Suppose there are no such costs. Consider two banks (A and B) just below the threshold with recovery rates $F^A = F^B = \underline{F}$. By bundling the cash flows of these two banks (A and B), the recovery rate increases to $F^{A+B} = \overline{F}$, and the value of a claim to the cash flows of A and B will exceed the sum of the value of these cash flows: $P(A) + P(B) \leq P(A + B)$. In the absence of costs, this represents an arbitrage. However, if there are positive costs C, then the value of A and B has to increase such that $P(A) + P(B) \geq P(A + B) - C[A, B]$ to eliminate the arbitrage opportunities. This increase reflects the probability that these banks end up crossing the size threshold because of growth or because of a future merger or acquisition. Hence, the alphas for these banks (A and B) will decrease, as their value rises, even though they do not directly benefit from the guarantee yet. Alternatively, A and B will actually merge right away.

There has been a large amount of consolidation in the banking sector in the last few decades, with the share of total market capitalization accounted for by the top decile of commercial banks increasing from 50% in the 1970s to 90% in the last decade. Similarly, the share of the total balance sheet accounted for by the top decile has increased from 52% to 95%. Kane (2000) and Brewer and Jagtiani (2007) document that acquiring banks are willing to pay larger premiums for banks that put them over critical size thresholds, consistent with our hypothesis. By backward induction, the same argument applies to smaller banks in other portfolios. However, the costs of bundling the cash flows ($C[D, E, F, \ldots, Z]$) of many smaller banks to reach this critical threshold increase, which mitigates the effect on the average risk-adjusted returns. This can account for the decreasing pattern in the alphas that we find in the data.

VII. Conclusion

Our paper documents a size anomaly in bank stock returns that is different from the size effect that has been documented for nonfinancials. This size effect can be explained by the covariance with a new size factor that we extract from that component of bank stock returns that is orthogonal to standard risk factors. This size factor is a measure of bank-specific tail risk. Our evidence from bank stock returns reveals how the pricing of bank-specific tail risk in financial markets depends on which bank is holding the risk.

To the extent that these effects reflect implicit bailout guarantees in financial disasters, the government subsidizes large financial institutions to take on tail risk. To mitigate this distortion, the government could consider taxing size in the banking sector. Our paper is the first to develop asset price-based measures of the resulting distortion to the equity component of the balance sheet. Our findings suggest that cost of capital distortions might have contributed to the precrisis growth in the size of the financial sector relative to the overall economy. Philippon (2008) argues that much of the variation in the size of the U.S. financial sector can be imputed to standard corporate finance forces. However, he notes the 2002 to 2007 period as an exception, which is exactly when we identify the largest distortions.

Initial submission: January 29, 2012; Final version received: November 20, 2014
Editor: Campbell Harvey

Appendix A. Derivation of Tail Risk Premium Expression

Consider the investor's Euler equation for asset i:

$$E_t\left[M_{t+1}R^i_{t+1}\right] = 1. \tag{A1}$$

The stand-in investor's SDF M_{t+1} is described in equation (6). This Euler equation can be decomposed as

$$(1-p_t)E_t\left[M^G_{t+1}R^i_{t+1}\right] + p_t E_t\left[M^G_{t+1}M^D_{t+1}R^{G,i}_{t+1}R^{D,i}\right] = 1. \tag{A2}$$

We assume that the distribution of the Gaussian factors is (conditionally) independent of the realization of the disaster,

$$\left((1-p_t) + p_t E_t\left[M^D_{t+1}R^{D,i}\right]\right) E_t\left[M^G_{t+1}R^{G,i}_{t+1}\right] = 1. \tag{A3}$$

Given these assumptions, this expression can be further simplified to yield

$$\left(1 + p_t E_t\left[M^D_{t+1} F^i - 1\right]\right) E_t\left[M^G_{t+1} R^i_{t+1}\right] = 1, \tag{A4}$$

where we have substituted the recovery rate F^i for $R^{D,i}$. To see why, note that the Gaussian return on stock i can be stated as

$$R^{G,i}_{t+1} = \frac{(P_{t+1}/D_{t+1}) + 1}{P_t/D_t} \frac{D_{t+1}}{D_t}, \tag{A5}$$

which in the case of no disaster can be stated as $R^{G,i}_{t+1} = \frac{(P_{t+1}/D_{t+1})+1}{P_t/D_t} \exp(g_D + \Delta \log D^{i,G}_{t+1})$. In the case of a disaster, the return is given by

$$R^i_{t+1} = R^{G,i}_{t+1} F^i_{t+1}, \tag{A6}$$

which only reflects the effect of the recovery rate on the dividend growth realization. Using the definition of resilience $p_t E_t[M^D_{t+1} F^i - 1]$, this yields the expression

$$\left(1 + H^i_t\right) E_t\left[M^G_{t+1} R^{G,i}_{t+1}\right] = 1. \tag{A7}$$

Decomposing this expectation into a covariance term and a cross-product yields

$$E_t\left[M^G_{t+1}\right] E_t\left[R^i_{t+1}\right] + cov_t\left[M^G_{t+1}, R^{G,i}_{t+1}\right] = \left(1 + H^i_t\right)^{-1}. \tag{A8}$$

Given the linear specification of the stochastic discount factor, this equation can in turn be written in the conditional beta representation:

$$E_t\left[R^{G,i}_{t+1}\right] = E_t\left[M^G_{t+1}\right]^{-1} \left(1 + H^i_t\right)^{-1} - \frac{cov_t\left[M^G_{t+1}, R^{G,i}_{t+1}\right]}{var_t\left[M^G_{t+1}\right]} \frac{var_t\left[M^G_{t+1}\right]}{E_t\left[M^G_{t+1}\right]}. \tag{A9}$$

We let $R_t = E_t[M^G_{t+1}]^{-1}$. Note that the variation in the price/dividend ratios induced by the variation in the case of a disaster does not co-vary with the normal risk factors—by assumption—and hence is not priced in the normal risk premium. The expected return on asset i, conditional on no disaster realization, after adjusting for Gaussian risk exposure, becomes

$$E_t\left[\widehat{R}^i_{t+1}\right] = \exp\left(r_t - h^i_t\right), \tag{A10}$$

where r_t denotes $\log R_t$ and h^i_t denotes $\log(1 + H^i_t)$.

Appendix B. Other Explanations

Business Cycle Variation in Common and Idiosyncratic Risk. Factors other than financial disasters could explain the cyclicality in the size factor. In particular, large idiosyncratic shocks can cause bank failures. If the volatility of

Table BI

Measuring Residual Risk Exposure

This table presents the standard deviation of residuals from OLS regression of monthly value-weighted excess returns of each size-sorted portfolio of U.S. commercial banks on the three Fama and French (1993) stock and two bond risk factors. In Panel A the row labeled "Recession" computes the (time-series) standard deviation of residuals during recession months and the row labeled "Full Sample" computes the (time-series) standard deviation for the 1970 to 2013 sample. In Panel B we examine the cross-sectional standard deviation of the residuals of banks in each bin for each period t. Panel B reports the time-series average of the cross-sectional standard deviation for each bin. The row labeled "Recession" lists the standard deviation of residuals during recession months and the row labeled "Full Sample" lists the standard deviation for the full sample. The standard deviations have been annualized by multiplying by $\sqrt{12}$ and are expressed in percentages.

Period	1	2	3	4	5	6	7	8	9	10
Panel A: Portfolios										
Recession	19.02	15.57	11.25	14.46	13.15	12.44	15.25	15.50	15.65	18.45
Full Sample	14.64	11.68	11.36	11.34	10.73	10.72	11.70	11.37	11.21	12.70
Panel B: Individual Banks										
Recession	30.11	29.85	20.34	22.91	23.89	23.68	23.66	26.53	24.31	23.86
Full Sample	36.06	27.88	26.21	25.24	24.29	23.28	23.29	22.95	21.12	19.16

these shocks increases more in recessions for small banks, that could explain some of our findings. Table BI measures the standard deviation of normal risk-adjusted returns both at the portfolio level (Panel A) and the bank level (Panel B). The first one measures the quantity of residual common risk. The second one measures the quantity of residual idiosyncratic risk. The portfolio-level measure in Panel A is the time-series standard deviation of normal risk-adjusted returns, reported separately for NBER expansions and recessions. The bank-level measure in Panel B is the average over time of the cross-sectional standard deviation within each portfolio of normal risk-adjusted returns.

During recessions, the exposure of the largest banks to residual common risk increases from 14.2% to 21.6%. For the smallest banks, the increase is only 3%. As expected, smaller banks are much more exposed to idiosyncratic risk than large banks, but the amount of idiosyncratic risk exposure of small banks does not seem to increase very much during recessions. The standard deviation ranges from 38% for the smallest banks to 26% for the largest banks during recessions, and from 36% to 20% in the whole sample. However, the largest percentage point increase in volatility, from 20% to 26%, during recessions is noted for the largest banks. For the smallest banks, the increase is less than 2%. There is no evidence to suggest that the cyclicality of the size factor is due to idiosyncratic bank risk.

Business Cycle Variation in Cash Flows. We analyze the data in the report for the first three quarters of 2001, which corresponds to the recession dates provided by NBER. During this period, small banks outperform large banks on almost all 13 performance parameters measured. Small banks had a higher return on equity (14.00% versus 13.80%), a higher return on assets (1.15 times that of large banks), a higher net interest margin (4.34% versus 3.62%), and comparable cost of funds (approximately 3.75% for both). During this recession, 70% of small banks and 60% of large banks reported earnings gains.

In 2008, large banks are again unable to match the performance of small banks on most measures. For the first half of 2008, small banks' return on equity is 1.5 times and yield on assets is 50 bps higher than corresponding values for large banks. Further, 14.16% of the 558 small banks and 26.72% of the 114 large banks were unprofitable, and 41.22% of small banks reported an earnings gain as compared to 24.14% of large banks.

For the full year of 2008, 28.70% of small banks and 40.35% of large banks reported losses. Small banks do have lower return on assets and return on equity for the full year, but it is not obvious if this is due to higher cash flow risk. During the second half of 2008, small banks not only earned a higher yield on assets and a higher net interest margin, but also provisioned more conservatively for losses. The ratio of loan loss provisions to assets increases to 1.93% for small banks by 4Q 2008 from 0.76% during 1Q 2008, but this ratio hardly changes for the largest banks.

REFERENCES

Acharya, Viral V., Christian Brownlees, Robert Engle, Farhang Farazmand, and Matthew Richardson, 2011, *Measuring Systemic Risk, Regulating Wall Street: The Dodd-Frank Act and the New Architecture of Global Finance* (John Wiley and Sons, Inc.).

Acharya, Viral V., Lasse Pedersen, Thomas Philippon, and Matthew Richardson, 2011, *Taxing Systemic Risk, Regulating Wall Street: The Dodd-Frank Act and the New Architecture of Global Finance* (John Wiley and Sons, Inc., New York, NY).

Acharya, Viral V., and Tanju Yorulmazer, 2007, Too many to fail: An analysis of time-inconsistency in bank closure policies, *Journal of Financial Intermediation* 16, 1–31.

Acharya, Viral V., and Tanju Yorulmazer, 2008, Cash-in-the-market pricing and optimal resolution of bank failures, *Review of Financial Studies* 21, 2705–2742.

Adrian, Tobias, and Markus K. Brunnermeier, 2010, Covar, Working paper, Federal Reserve Bank of New York.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Barro, Robert, 2006, Rare disasters and asset markets in the twentieth century, *Quarterly Journal of Economics* 121, 823–866.

Basu, Sanjoy, 1983, The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence, *Journal of Financial Economics* 12, 129–156.

Berk, Jonathan, 1995, A critique of size-related anomalies, *Review of Financial Studies* 8, 275–286.

Berk, Jonathan, 1997, Does size really matter? *Financial Analysts Journal* 53, 12–18.

Boyd, J. H., and M. Gertler, 1993, U.S. commercial banking: Trends, cycles, and policy, *NBER Macroeconomics Annual* 8, 319–368.

Brennan, Michael J., and Avanidhar Subrahmanyam, 1996, Market microstructure and asset pricing: On the compensation for illiquidity in stock returns, *Journal of Financial Economics* 41, 441–464.

Brewer, Elijah, and Julapa Jagtiani, 2007, How much would banks be willing to pay to become "too-big-to-fail" and to capture other benefits? Research Working paper RWP 07-05, Federal Reserve Bank of Kansas City.

Brunnermeier, Markus K., and Yuliy Sannikov, 2014, A macroeconomic model with a financial sector, *American Economic Review* 104, 379–421.

Duffie, Darrell, 2010, The failure mechanics of dealer banks, *Journal of Economic Perspectives* 24, 51–72.

Fahlenbrach, Rüdiger, Robert Prilmeier, and René M. Stulz, 2012, This time is the same: Using bank performance in 1998 to explain bank performance during the recent financial crisis, *The Journal of Finance* 67, 2139–2185.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Finanicial Economics* 33, 3–56.

Farhi, Emmanuel, and Jean Tirole, 2012, Collective moral hazard, maturity mismatch and systemic bailouts, *The American Economic Review* 102, 60–93.

Flannery, Mark J., and Christopher M. James, 1984, The effect of interest rate changes on the common stock returns of financial institutions, *The Journal of Finance* 39, 1141–1153.

Gabaix, Xavier, 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, *Quarterly Journal of Economics* 127, 645–700.

Giesecke, Kay, Francis A. Longstaff, Stephen Schaefer, and Ilya Strebulaev, 2011, Corporate bond default risk: A 150-year perspective, *Journal of Financial Economics* 102, 233–250.

Gorton, Gary, 1988, Banking panics and business cycles, *Oxford Economic Papers* 40, 751–781.

Gorton, Gary, and Andrew Metrick, 2012, Securitized banking and the run on repo, *Journal of Financial Economics* 104, 425–451.

Gourio, Francois, 2008, Time-series predictability in the disaster model, *Finance Research Letters* 5, 191–203.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263–1295.

Huang, Xin, Hao Zhou, and Haibin Zhu, 2009, A framework for assessing the systemic risk of major financial institutions, *Journal of Banking and Finance* 33, 2036–2049.

Kane, Edward J., 2000, Incentives for banking megamergers: What motives might regulators infer from event-study evidence? *Journal of Money, Credit and Banking* 32, 671–701.

Kareken, John H., and Neil Wallace, 1978, Deposit insurance and bank regulation: A partial-equilibrium exposition, *Journal of Business* 51, 413–438.

Kelly, Bryan T., Hanno N. Lustig, and Stijn Van Nieuwerburgh, 2011, Too-systemic-to-fail: What option markets imply about sector-wide government guarantees, Working paper, UCLA Anderson School of Management.

Kho, Bong-Chan, Dong Lee, and René M. Stulz, 2000, U.S. banks, crises, and bailouts: From Mexico to LTCM, *American Economic Review* 90, 28–31.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1993, Contrarian investment, extrapolation, and risk, Working paper, University of Chicago.

Longstaff, Francis A., and Brett Myers, 2009, Valuing toxic assets: An analysis of CDO equity, Working paper, UCLA.

Longstaff, Francis A., and Monika Piazzesi, 2004, Corporate earnings and the equity premium, *Journal of Financial Economics* 74, 401–421.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

O'Hara, Maureen, and Wayne Shaw, 1990, Deposit insurance and wealth effects: The value of being "too big to fail", *Journal of Finance* 45, 1587–1600.

Panageas, Stavros, 2010a, Bailouts, the incentive to manage risk, and financial crises, *Journal of Financial Economics* 95, 296–311.

Panageas, Stavros, 2010b, Optimal taxation in the presence of bailouts, *Journal of Monetary Economics* 57, 101–116.

Philippon, Thomas, 2008, The evolution of the U.S. financial industry from 1860 to 2007, Working paper, NYU Stern.

Ranciere, Romain, and Aaron Tornell, 2011, Financial black-holes: The interaction of financial regulation and bailout guarantees, Working paper, UCLA.

Reinhart, Carmen M., and Kennth Rogoff, 2009, *This Time Is Different* (Princeton University Press, Princeton, NJ and Cambridge, UK).

Rietz, Thomas A., 1988, The equity risk premium: A solution, *Journal of Monetary Economics* 22, 117–131.

Schneider, Martin, and Aaron Tornell, 2004, Balance sheet effects, bailout guarantees and financial crises, *Review of Economic Studies* 71, 883–913.

Veronesi, Pietro, and Luigi Zingales, 2010, Paulson's gift, *Journal of Financial Economics* 97, 339–368.

Wachter, Jessica A., 2013, Can time-varying risk of rare disasters explain aggregate stock market volatility?, *Journal of Finance* 68, 987–1035.

Wicker, Elmus, 1996, *The Banking Panics of the Great Depression* (Cambridge University Press, Cambridge, UK).

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publishers website:

Appendix S1: Internet Appendix.

Journal of Financial Economics 105 (2012) 18–36

Contents lists available at SciVerse ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Illiquidity or credit deterioration: A study of liquidity in the US corporate bond market during financial crises [☆]

Nils Friewald [a,*], Rainer Jankowitsch [a], Marti G. Subrahmanyam [b]

[a] WU (Vienna University of Economics and Business), Heiligenstädter Straße 46–48, Vienna 1190, Austria
[b] New York University, Stern School of Business, 44 West Fourth, New York, NY 10012, USA

A R T I C L E   I N F O

Article history:
Received 8 October 2010
Received in revised form
11 July 2011
Accepted 8 August 2011
Available online 10 February 2012

JEL classification:
G01
G12
G14

Keywords:
Liquidity
Corporate bonds
Financial crisis
OTC markets

A B S T R A C T

We investigate whether liquidity is an important price factor in the US corporate bond market. In particular, we focus on whether liquidity effects are more pronounced in periods of financial crises, especially for bonds with high credit risk, using a unique data set covering more than 20,000 bonds, between October 2004 and December 2008. We employ a wide range of liquidity measures and find that liquidity effects account for approximately 14% of the explained market-wide corporate yield spread changes. We conclude that the economic impact of the liquidity measures is significantly larger in periods of crisis, and for speculative grade bonds.

1. Introduction

The global financial crisis had its origins in the US subprime mortgage market in 2006–2007, but has since spread to virtually every financial market around the world. The most important aspect of this crisis, which sharply distinguishes it from previous crises, is the rapidity and degree to which both the liquidity and credit quality of several asset classes deteriorated. While clearly both liquidity and credit risk are key determinants of asset prices, in general, it is important to quantify their relative effects and, particularly, how much they changed during the crisis. It is also relevant to ask if there are interactions between these factors, and whether these relations changed substantially in magnitude and quality from prior periods. In this paper, we study liquidity effects in the US corporate bond market for the period October 2004 to December 2008, including the GM/Ford

☆ We thank INQUIRE Europe for their support of this project and Markit for providing us with valuation data. We also thank Viral Acharya, Engelbert Dockner, Peter Feldhütter, Alois Geyer, Frank de Jong, Robert Korajczyk, David Lando, David Lesmond, Stefan Pichler, Jian Wang, Jason Wei, and especially the referee, Paul Schultz, and the editor, William Schwert, for helpful comments on prior drafts. We are grateful to conference participants at the 2009 German Finance Association, the 2010 Midwest Finance Association, the 2010 Swiss Society for Financial Market Research, the Annual Finance Conference on Recent Advances in Corporate Finance of the Wilfrid Laurier University, the Fourth Annual Risk Management Conference of the National University of Singapore, the International Risk Management Conference of the University of Florence, the Third Erasmus Liquidity Conference of the Rotterdam School of Management, the 2010 China International Conference in Finance, the WU Gutmann Center Symposium 2011, and seminar participants at the Copenhagen Business School, University of Melbourne, ESSEC, NYU Stern School, the Italian Treasury, and Standard & Poor's for useful comments and suggestions.

* Corresponding author.
E-mail addresses: nils.friewald@wu.ac.at (N. Friewald), rainer.jankowitsch@wu.ac.at (R. Jankowitsch), msubrahm@stern.nyu.edu (M.G. Subrahmanyam).

0304-405X/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jfineco.2012.02.001

downgrades and the subprime crisis, using a unique data set covering basically the whole US corporate bond market. We employ a wide range of liquidity measures to quantify the liquidity effects in corporate bond yield spreads.

Our analysis explores the time-series and cross-sectional aspects of liquidity for the whole market, as well as various important segments, using panel and Fama-MacBeth regressions, respectively.

Most major financial markets, including those for equity, foreign exchange, credit, and commodities, were severely affected in terms of price and liquidity in the subprime crisis. However, the impact has been disproportionately felt in the fixed income markets, including the markets for collateralized debt obligations (CDO), credit default swaps (CDS), and corporate bonds. An important point to note is that these securities are usually traded in over-the-counter (OTC) markets, where there is no central market place, or even a clearing house. Indeed, this aspect has come under regulatory scrutiny since the near collapse of the CDS market, which was an opaque OTC market. It is the OTC structure of fixed income markets that makes research, especially on liquidity effects, difficult as traded prices and volumes are not readily available, and important aspects of the markets can only be analyzed based on quotations from individual dealers, which are not necessarily representative of the market as a whole.

US corporate bonds trade in an important OTC market. This market is an ideal laboratory to examine liquidity and credit factors because of the following reasons: First, in contrast to most other OTC markets, detailed transaction data are available on prices, volumes, and other market variables since 2004, through an effort of the Financial Industry Regulatory Agency (FINRA), known as the Trade Reporting and Compliance Engine (TRACE). This database aggregates virtually all transactions in the US corporate bond market, which is unusual for any OTC market. Second, the US corporate bond market bore the brunt of the subprime crisis in terms of credit deterioration, almost to the same extent as the credit derivatives market, to which it is linked by arbitrage and hedging activities. Third, there is considerable variation in credit quality as well as liquidity in this market, both over time and across bonds, providing researchers with the opportunity to examine the differences arising out of changes in liquidity.

For our empirical analysis, we use all traded prices from TRACE, along with market valuations from Markit, bond characteristics from Bloomberg, and credit ratings from Standard & Poor's. Our combined data set is perhaps the most comprehensive one of the US corporate bond market that has been assembled to date, covering 23,703 bonds and 3,261 firms. This data set enables us to study liquidity effects for virtually the whole bond market, including bond segments that show very low trading activity.

The main focus of our research in this paper is to determine the quantitative impact of liquidity factors, while controlling for credit risk, based on credit ratings and other risk characteristics. In our analysis, we focus on the yield spread of a corporate bond, defined as its yield differential relative to that of a risk-free benchmark of similar duration. The benchmark could be either the Treasury bond or the swap rate curve.

To measure liquidity, we consider several alternative proxies for liquidity. We employ *bond characteristics* that have been used as liquidity proxies in many studies. We use directly observable *trading activity variables* (e.g., the number of trades) and, most important, we employ several alternative *liquidity measures* proposed in the literature, i.e., the Amihud, Roll, zero-return, and price dispersion measure.

First, we explore the hypothesis that liquidity is priced in the US corporate bond market. We find that the liquidity proxies account for about 14% of the explained time-series variation of the yield spread changes over time for individual bonds, while controlling for credit quality. Most of the liquidity proxies exhibit statistically as well as economically significant results. While the trading activity variables are important in explaining the bond yield spread changes, the liquidity measures exhibit even stronger effects in terms of economic impact. In particular, measures estimating trading costs based on transaction data show the strongest effects.

Second, our main research question is whether the effect of liquidity is stronger in times of crises. Our hypothesis is that in crises, when capital constraints become binding and inventory holding costs and search costs rise dramatically, liquidity effects are more pronounced. Therefore, we analyze credit and liquidity effects for three different regimes during our sample period, i.e., the GM/Ford crisis, the subprime crisis, and the period in between, when market conditions were more normal. Based on time-series analysis, we find that the effect of the liquidity measures is far stronger in both the GM/Ford crisis and the subprime crisis: the economic significance of the liquidity proxies increased by 30% in the GM/Ford crisis compared to the normal period, and more than doubled in the subprime crisis. We also examine the cross-sectional behavior of the yield spread using Fama-MacBeth regressions in the three different time periods. In general, the cross-sectional results paint a picture similar to the time-series analysis. Moreover, we find in the cross-section that time-invariant bond characteristics, e.g., amount issued, show significant effects as well.

Third, we analyze the interaction between credit and liquidity risk. We expect to find higher liquidity in the investment grade sector if liquidity concerns cause investors to abandon the junk bond market in favor of investment grade bonds in a *flight-to-quality*. We present descriptive statistics providing evidence for a flight-to-quality during financial distress and the regression analysis indeed shows lower liquidity for speculative grade bonds as well as a stronger reaction to changes in liquidity. In general, these results indicate that the liquidity component is far more important in explaining the change in the yield spread for bonds with high credit risk.

The remainder of the paper is organized as follows: we present a survey of the relevant literature in Section 2 of the paper, focusing mainly on papers relating to liquidity effects in corporate bond markets. Section 3 discusses the hypotheses being tested in the paper and the economic motivation behind them. In Section 4, we explain, in detail, the composition of our data set and the filters and matching procedures we employ in combining data from four different data sources. Section 5 discusses the alternative measures of liquidity that have been proposed and used in the literature and their pros and cons. We focus, in particular, on the relevance of these measures for a relatively illiquid OTC market. In Section 6 we outline the methodology. Section 7 presents the time-series results, based on panel regressions, and the results for the cross-sectional analysis based on the Fama-MacBeth procedure, used to test our hypotheses. Section 8 concludes.

2. Literature survey

The academic literature on liquidity effects on asset prices is vast. An early paper was by Amihud and Mendelson (1986), who first made the conceptual argument that transaction costs result in liquidity premiums in asset prices in equilibrium, due to different trading horizons of investors. This conclusion has been extended and modified in different directions and also been tested in a host of asset markets. This literature, focusing mainly on equity markets, is surveyed by Amihud, Mendelson, and Pedersen (2006). In the context of OTC markets, Duffie, Garleanu, and Pedersen (2007) show that transaction costs are driven by search frictions, inventory holding costs, and bargaining power in this particular market structure. A related argument is presented in Jankowitsch, Nashikkar, and Subrahmanyam (2011). In a recent paper, Acharya, Amihud, and Bharath (2009) argue that these frictions change over time and are higher in times of financial crises, due to binding capital constraints and increased holding and search costs.

The literature on credit risk modeling provides evidence of liquidity effects in the corporate bond market and shows that risk-free interest rates and credit risk are not the only factors that drive corporate bond prices. This result has been established based on reduced-form models (see, for example, Longstaff, Mithal, and Neis, 2005; Nashikkar, Subrahmanyam, and Mahanti, 2011), and structural models (see, for example, Huang and Huang, 2003), i.e., neither credit risk measured by the prices of CDS contracts nor asset value information from the equity market, can fully explain corporate bond yields.

Several authors study the impact of liquidity, based on corporate bond yields or yield spreads over a risk-free benchmark. Most of these papers rely on indirect proxies based on *bond characteristics* such as the coupon, age, amount issued, industry, and bond covenants; some papers additionally use market-related proxies based on *trading activity* such as trade volume, number of trades, number of dealers, and the bid-ask spread, see, e.g., Elton, Gruber, Agrawal, and Mann (2001), Collin-Dufresne, Goldstein, and Martin (2001), Perraudin and Taylor (2003), Eom, Helwege, and Huang (2004), Liu, Longstaff, and Mandell (2004), Houweling, Mentink, and Vorst (2005), Longstaff, Mithal, and Neis (2005), De Jong and Driessen (2006), Edwards, Harris, and Piwowar (2007), and Acharya, Amihud, and Bharath (2009). Essentially, all these papers find that liquidity is priced in bond yields. However, they find different magnitudes and varying importance of these basic liquidity proxies, but mostly at the market-wide level.

In the more recent literature, several alternative *liquidity measures* that are estimators of transaction costs, market impact, or turnover, have been proposed and applied to analyze liquidity in the corporate bond market at the level of individual bonds. The *Roll measure* (see Roll, 1984; Bao, Pan, and Wang, 2011) interprets the subsequent prices as arising from the "bid–ask bounce": thus, the autocovariance in price changes provides a simple liquidity measure. A similar idea to measure transaction costs is proposed and implemented in the *LOT measure* proposed by Lesmond, Ogden, and Trzcinka (1999). The *Amihud measure* (see Amihud, 2002) relates the price impact of a trade to the trade volume. Trading activity itself is used in the *zero-return measure* based on the number of unchanged sequential prices and the *no-trade measure* based on time periods without trading activity (see, e.g., Chen, Lesmond, and Wei, 2007). Mahanti, Nashikkar, Subrahmanyam, Chacko, and Mallik (2008) propose another measure known as *latent liquidity* that is based on the institutional holdings of corporate bonds, which can be used even in the absence of transaction data. Jankowitsch, Nashikkar, and Subrahmanyam (2011) develop the *price dispersion measure*, which is based on the dispersion of market transaction prices of an asset around its consensus valuation by market participants.

Most of the early papers on bond market liquidity are based only on quotation data as reasonably complete transaction data were not available until a few years ago. However, some papers use restricted samples of the transaction data for certain parts of the corporate bond market to analyze liquidity, including Chakravarty and Sarkar (1999), Hong and Warga (2000), Schultz (2001), and Hotchkiss and Ronen (2002). Many more researchers focused on the issue of liquidity in the corporate bond market since the TRACE data on US corporate bond transactions started to become available in 2002. This new source of bond price information allows researchers to analyze many different aspects of the US corporate bond market; see, e.g., Edwards, Harris, and Piwowar (2007), Goldstein and Hotchkiss (2007), Mahanti, Nashikkar, Subrahmanyam, Chacko, and Mallik (2008), Ronen and Zhou (2009), Nashikkar, Subrahmanyam, and Mahanti (2011), Lin, Wang, and Wu (2011), and Jankowitsch, Nashikkar, and Subrahmanyam (2011).

It is especially interesting to examine how liquidity affects the corporate bond market in times of financial crisis. While much of the research on the current financial crisis is probably in progress, two recent papers do provide some early evidence on the impact of liquidity in the US corporate bond market. These include Bao, Pan, and Wang (2011) and Dick-Nielsen, Feldhütter, and Lando (2012).

Bao, Pan, and Wang (2011) use the TRACE data to construct the Roll measure as a proxy for liquidity. Using a sample of around 1,000 bonds that existed prior to October 2004, they show that illiquidity measured by the Roll measure is quite significant in this market and much larger than would be predicted by the bid–ask bounce. They also show that their measure exhibits commonality across bonds, which tends to go up during periods of market crisis. Further, they relate the Roll measure to bond yield spreads in a cross-sectional regression setup and provide evidence that part of the yield spread differences across bonds is due to illiquidity.

Dick-Nielsen, Feldhütter, and Lando (2012) combine the TRACE data using straight bullet bonds (around 4,000 bonds), with accounting data and equity volatility, as proxies for credit risk. They use a panel regression based on quarterly data to study the effects of five different liquidity measures and the defined credit risk variables. In general, they find a significant effect of liquidity, which increased with the onset of the subprime crisis. However, their multivariate regression results show somewhat mixed results for different rating classes.

There are several important differences between these prior papers and our own research in this paper. First, we employ a much larger data set on transaction data on US corporate bonds than any prior papers, as our sample of 23,703 bonds basically covers the whole traded market. This is a major difference even compared with the recent work of Bao, Pan, and Wang (2011) and Dick-Nielsen, Feldhütter, and Lando (2012), who focus only on a certain, generally the more liquid, subsegment of the market. Second, our research explicitly covers two crisis periods, which are analyzed separately: the broader subprime crisis and the earlier, GM/Ford crisis, which affected particular segments of the US corporate bond market. We contrast the behavior of liquidity and its pricing in bond yield spreads during periods of crisis with more normal periods and analyze the interaction of credit and liquidity risk. Third, we include the additional information on the market's consensus valuation of bonds provided by Markit. These data permit us to estimate the price dispersion measure for the bonds in our sample and, thus, include an important additional measure of transaction costs. This liquidity proxy is particularly relevant for our research question, as transaction cost measures appear to be especially important in explaining liquidity in OTC markets.

3. Hypotheses

In this section, we provide an overview of the research questions we pose and the hypotheses we test in our research. Our approach is to examine the validity of specific arguments regarding the effect of liquidity in the US corporate bond market.

H1: Liquidity is an important price factor in the US corporate bond market.

As argued by Amihud and Mendelson (1986), investors with different trading horizons have different expected returns, after taking into account the transactions costs they will incur over their respective horizons. This phenomenon translates into a clientele effect (for securities in positive net supply) by which the more illiquid assets are cheaper and are held by investors with longer horizons relative to their liquid counterparts, which are held by those with shorter horizons. Duffie, Garleanu, and Pedersen (2007) and Jankowitsch, Nashikkar, and Subrahmanyam (2011) argue that in OTC markets the liquidity premium is driven by transaction costs due to search frictions, inventory holding costs, and bargaining power. In the corporate bond market context, these frictions are reflected in the bond prices, whereby liquid bonds earn a lower expected return than illiquid bonds which are similar on other dimensions, such as bond features and risk characteristics.

The US corporate bond market is especially interesting in this respect, as liquidity differences across individual bonds seem to be rather pronounced: very few bonds are traded frequently, while most other bonds are hardly ever traded at all (see Mahanti, Nashikkar, Subrahmanyam, Chacko, and Mallik, 2008 for details of a cross-sectional comparison for the US corporate bond market). Moreover, trading in the US corporate bond market involves much higher transaction costs compared to related markets such as the stock market. Thus, we would expect a significant liquidity premium, as argued in Amihud and Mendelson (1986), and expect that our liquidity proxies can explain a significant part of bond yield spreads. Our aim is to quantify these liquidity effects as a priced factor.

H2: Liquidity effects are more important in periods of financial distress.

The liquidity premium in the corporate bond market can be expected to change over time depending on market conditions, especially during a financial crisis. Several arguments have been proposed in the literature regarding the behavior of agents in a crisis. For example, Duffie, Garleanu, and Pedersen (2007) propose that liquidity is more important in crisis periods, since inventory holding costs and search costs are higher, and also asymmetric information is a more important issue. Acharya, Amihud, and Bharath (2009) provide empirical support for this hypothesis arguing that banks face more stringent capital requirements when they hold illiquid assets and could find it more difficult to access liquidity during a crisis. Moreover, a greater proportion of investors could have shorter horizons in a crisis. For example, bond mutual funds and hedge funds could face the possibility of redemptions or are forced to meet value-at-risk requirements and margin calls and, therefore, wish to hold more liquid assets to address this eventuality; see, e.g., Sadka (2010). Individual investors could shift more of their portfolios from illiquid to liquid assets as they turn more risk averse. For all these reasons, the gap in pricing between liquid and illiquid bonds, that are otherwise similar, may widen, resulting in a higher liquidity premium.

Thus, the second and main research question of this paper is whether the effect of liquidity is stronger during times of financial crises. We expect a particularly strong effect in the subprime crisis, when capital constraints became binding and inventory holding costs and search costs rose dramatically for all market participants.

H3: Liquidity effects are more important for bonds with high credit risk.

We study whether a bond's credit rating is related to liquidity effects by focusing on the difference between investment grade and speculative grade bonds. Acharya, Amihud, and Bharath (2009) show that liquidity is substantially different between investment grade and speculative grade bonds using a regime switching model. They argue that in periods of financial crisis, all bond prices decline due to an increase of illiquidity. At the same time, a *flight-to-quality* effect is expected, which leads to lower price reactions among investment grade bonds. Chen, Lesmond, and Wei (2007) also provide empirical support for this argument. Thus, we expect stronger liquidity effects for speculative grade bonds and to find flight-to-quality effects in periods of crisis.

4. Data description

In this section, we present the unique data set we have at hand for this liquidity study covering basically the whole US corporate bond market. Our data are drawn from several different sources:

1. Transaction data from the Trade Reporting and Compliance Engine (TRACE).
2. Consensus market valuations from Markit.
3. Credit ratings from Standard & Poor's.
4. Bond characteristics from Bloomberg.
5. Treasury and swap data from Bloomberg.

Our time period starts with the date when TRACE was fully implemented on October 1, 2004, and covers the period until December 31, 2008. TRACE provides detailed information about all transactions in the US corporate bond market, i.e., the actual trade price, the yield based on this price, as well as the trade volume measured in US dollars for each transaction.[1] Phase I of TRACE was launched by the Financial Industry Regulatory Agency (FINRA) in July 2002, with the aim of improving transparency in the US corporate bond market. This phase covered only the larger and generally higher credit quality issues. Phase II expanded the coverage and dissemination of information to smaller investment grade issues. Since the final Phase III was implemented on October 1, 2004, transactions of essentially all US corporate bonds have been reported. Hence, the TRACE database has been reasonably complete since its final implementation. This data source is almost unique for an OTC market, since in many other cases, price information usually must be obtained either from an individual dealer's trading book, which provides a very limited view of the market, or by using bid–ask quotations instead. In the US corporate bond market, reporting of any transaction to TRACE is obligatory for broker-dealers and follows a set of rules approved by the Securities and Exchange Commission (SEC), whereby all transactions must be reported within a time frame of 15 minutes.

We use the filters proposed by Dick-Nielsen (2009) for the TRACE data to eliminate potentially erroneous data points.[2] In addition, we follow Edwards, Harris, and Piwowar (2007) and apply a *median filter* and a *reversal filter* to eliminate further potential data errors. While the median filter identifies potential outliers in reported prices within a certain time period, the reversal filter identifies unusual price movements.[3] Eliminating any potential errors in the reported transactions reduces the number of reported trades by roughly 5.5% to 23.5 million trades. This results in a TRACE data sample consisting of 34,822 bonds from 4,631 issuers.

An important additional source for the market's valuation of a bond is obtained from Markit Group Limited, a leading data provider, specialized in security and derivatives pricing. One of its services is to gather, validate, and distribute end-of-day composite bond prices from dealer polls. Up to 30 contributors provide data from their books of record and from feeds to automated trading systems (see Markit Group Limited, 2006). These reported valuations are averaged for each bond after eliminating outliers, using their proprietary methodology. Hence, this price information can be considered as a market-wide average of a particular bond price, reflecting the market consensus. The Markit valuations are used by many financial institutions to mark their portfolios to market and have credibility among practitioners. In total, we have 5,522,735 Markit quotes, covering 28,145 bonds in our database.

To control for default risk, we use credit ratings from Standard & Poor's (S&P). We focus on long-term, issue credit ratings as the market's current judgment of the obligor's creditworthiness with respect to a specific financial obligation. It should be noted, that in our descriptive statistics of the rating variable, we assign integer numbers to ratings, i.e., AAA=1, AA+=2, etc., to measure the "average" rating of certain groups of bonds or time periods. Our time period contains 25,464 bonds, which have at least one S&P credit rating each. Note that credit risk could be measured using alternative approaches. Two prominent examples come to mind: using CDS spreads in the context of a reduced-form credit risk model, as in Longstaff, Mithal, and Neis (2005) and Nashikkar, Subrahmanyam, and Mahanti (2011), or using accounting-based and equity-related data in a structural model context, as in Huang and Huang (2003). We do not incorporate such proxies as this information is generally only available for a very small (presumably more liquid) segment of the market and our intention is to explicitly analyze liquidity effects for the whole market. In addition, the impact of the liquidity on these data inputs would also have to be taken into account, rendering the analysis far more complex, and hence, prone to additional error. This issue is particularly true during periods of crisis when liquidity and counterparty risk considerations are exacerbated in the pricing of CDS as well as equity contracts. Hence, we apply the more parsimonious approach of using only the credit ratings, with their admitted

[1] The reported trade volume is capped at $1 million for high yield and unrated bonds and at $5 million for investment grade bonds.

[2] As pointed out by Dick-Nielsen (2009), care should be exercised when accounting for order cancelations or corrections in the TRACE data. To mitigate the errors that result from these issues, Dick-Nielsen (2009) suggests that the trade data need to be "cleaned up" using error filters.

[3] The median filter eliminates any transaction where the price deviates by more than 10% from the daily median or from a nine-trading-day median centered at the trading day. The reversal filter eliminates any transaction with an absolute price change deviating from the lead, lag, and average lead/lag price change by at least 10%.

shortcomings, in terms of their own error and failure to anticipate changes in credit risk.

For each of the bonds available in TRACE, we additionally obtain bond characteristics from Bloomberg. These bond characteristics include the issue date, maturity, age, coupon, amount issued, industry sector, and bond covenants. Most of these characteristics have been considered as simple liquidity proxies by previous studies. Furthermore, we use swap rates and Treasury rates for various maturities retrieved from Bloomberg as the benchmark for the risk-free interest rate curve to compute the corporate bond yield spreads.

Given these data sets, we generate a sample that is representative of the whole market by merging the daily trade observations from TRACE with end-of-day Markit-quotations, the available S&P ratings, and the bond characteristics. This sample covers 23,703 bonds of 3,261 firms. On average per day, we observe 5,423 traded bonds, 21,254 trades, and $7.563 billion in volume. Thus, our panel data set covers approximately 80% of the overall trading activity in the US corporate bond market. We find that the market coverage is at this high level throughout the observation period and, hence, is highly representative of the whole US corporate bond market including bonds with very low trading activity, which is a major difference compared to most other studies. Nevertheless, a considerable number of bonds is traded only very rarely. However, data limitations caused by this lack of trading activity should actually bias us against finding any clear liquidity effects at all.

5. Liquidity proxies

This section presents the various liquidity proxies that we use in the regression analysis as explanatory variables. A number of liquidity proxies have been proposed in the literature (see Section 2) which are not all equally viable, given the challenges of obtaining detailed and sufficiently frequent data in the relatively illiquid corporate bond market. Our data set allows us to compare the efficiency of most of these proposed proxies in this empirical study. We classify the available proxies into three groups: *bond characteristics*, *trading activity variables*, and *liquidity measures*.

Bond characteristics, such as the amount issued, are simple liquidity proxies which provide a rough indication about the potential liquidity of a bond. Trading activity variables, such as the number of trades, provide bond-specific information based on transaction data. Liquidity measures, such as the price dispersion and Amihud measure, are alternative estimators of transaction costs or market impact.[4]

All these liquidity proxies can either be calculated on a daily basis, if price information is observable for a particular bond, or are time-invariant (e.g., coupon), or change linearly with time (e.g., age). In the following subsections, we present the definitions of the various liquidity proxies that we use in our analysis and discuss the details of their computation.

5.1. Bond characteristics

The bond characteristics we consider as liquidity proxies are the *amount issued*, *coupon*, *maturity*, and *age*. These proxies, while admittedly crude measures, make intuitive sense. In general, we expect bonds with a larger amount issued to be more liquid and bonds with a larger coupon to be less liquid.[5] Bonds with long maturities (over 10 years) are generally considered to be less liquid since they are often bought by ‘‘buy-and-hold’’ investors, who trade infrequently. Similarly, we expect recently issued (on-the-run) bonds to be more liquid. We consider these measures to be important only for our cross-sectional analysis, as most of these are either constant (e.g., coupon) or change linearly (e.g., maturity) over time.

5.2. Trading activity variables

A bond’s trading activity provides information about liquidity. In this sense, higher trading activity generally indicates higher liquidity. We consider the following trading activity variables: *number of trades*, *trade volume*, and *trading interval*. We compute the number of trades and the trade volume of a particular bond on each day from the trading information given by TRACE. The trading interval is the elapsed time (measured in days) since the last day a given bond was traded. Longer trade intervals indicate less trading activity and, thus, lower liquidity. Therefore, we expect liquidity to be higher for bonds with shorter time intervals between trading days.

5.3. Liquidity measures

5.3.1. Amihud measure

This liquidity proxy is a well-known measure originally proposed for the equity market by Amihud (2002), which is conceptually based on Kyle (1985). It relates the price impact of trades, i.e., the price change measured as a return, to the trade volume measured in US dollars. The Amihud measure at day t for a certain bond over a particular time period with N_t observed returns is defined as the average ratio between the absolute value of these returns r_j and its trading volumes v_j, i.e.,

$$\text{Amihud}_t = \frac{1}{N_t}\sum_{j=1}^{N_t}\frac{|r_j|}{v_j}. \tag{1}$$

[4] We are aware that many studies without access to transaction data use the quoted bid–ask spread as a liquidity proxy. However, bid–ask spreads are, in general, only available for a small subsample representing the relatively larger issues. In a robustness check we find that bid–ask spreads from Bloomberg are of minor importance once transaction-based measures are considered, in the empirical specifications we investigate below. These results are not reported in this paper, but are available from the authors upon request.

[5] Note that the coupon *per se* is rather a crude proxy for credit risk. Once we adjust for credit risk (e.g., by using ratings), bonds with different coupons but with identical credit risk exhibit different levels of liquidity. However, as we are certainly not able to perfectly adjust for credit risk, the coupon cannot be viewed as a pure liquidity proxy.

A larger Amihud measure implies that trading a bond causes its price to move more in response to a given volume of trading, in turn, reflecting lower liquidity. We use the daily volume-weighted average TRACE prices to generate the returns r_j and calculate the Amihud measure on a day-by-day basis.

5.3.2. Price dispersion measure

A new liquidity measure recently introduced for the OTC market is the price dispersion measure of Jankowitsch, Nashikkar, and Subrahmanyam (2011). This measure is based on the dispersion of traded prices around the market-wide consensus valuation. A low dispersion around the valuation indicates that the bond can be bought close to its fair value and, therefore, represents low trading costs and high liquidity, whereas high dispersion implies high transaction costs, and hence, low liquidity. This measure is derived from a market microstructure model and shows that price dispersion is the result of market frictions such as inventory risk for dealers and search costs for investors. It presents a direct estimate of trading costs based on transaction data. As in Jankowitsch, Nashikkar, and Subrahmanyam (2011), the traded prices are obtained from TRACE and the market valuations from Markit. The price dispersion measure is defined as the root mean squared difference between the traded prices and the respective market-wide valuation weighted by volume, i.e., for each day t and a particular bond, it is given by

$$\text{Price dispersion}_t = \sqrt{\frac{1}{\sum_{k=1}^{K_t} v_k} \sum_{k=1}^{K_t} (p_k - m_t)^2 v_k}, \tag{2}$$

where p_k and v_k represent the K_t observed traded prices and their trade volumes on date t and m_t is the market-wide valuation for that day. Hence, the price dispersion indicates the potential transaction cost for a trade.

5.3.3. Roll measure

This measure developed by Roll (1984) shows that, under certain assumptions, adjacent price movements can be interpreted as a bid–ask bounce which, therefore, allows us to estimate the effective bid–ask spread. This bid-ask bounce results in transitory price movements that are serially negatively correlated and the strength of this covariation is a proxy for the round-trip costs for a particular bond, and hence, a measure of liquidity. More precisely, the Roll measure is defined as

$$\text{Roll}_t = 2\sqrt{-\text{Cov}(\Delta p_t, \Delta p_{t-1})}, \tag{3}$$

where Δp_t is the change in prices from $t-1$ to t. We compute the Roll measure based on the daily volume-weighted bond prices p_t from the TRACE data set, where we use a rolling window of 60 days and require at least eight observations to determine the covariance.[6]

5.3.4. Zero-return measure

The zero-return measure indicates whether we observe a zero price movement between trading days. The zero-return measure is set to one, if we find an unchanged price, and is set to zero, otherwise. Bond prices that stay constant over long time periods are likely to be less liquid, as the information could be stale. Obviously, such a measure can only be based on price quotations or valuations, such as Markit quotes in our case. Constant price information in these data sources reveal illiquidity as unchanged quotations could indicate an incomplete coverage of the bond.

6. Methodology

This section outlines our general approach to measuring the impact of liquidity and credit risk on pricing in the US corporate bond market. We present here our definitions of the bond yield spread and define the subperiods of interest to test our hypotheses about financial crises. We then present the specifications for our panel data regressions to explore the time-series properties, and the Fama-MacBeth regressions to explore the cross-sectional properties of our data. We use these specifications to study market-wide liquidity effects and analyze subsegments of the market, where we compare investment grade and speculative grade bonds.

6.1. Bond yield spread

The dependent variable in our setup is the corporate bond yield spread, represented by the yield differential relative to that of a risk-free benchmark. We define this benchmark as the yield of a risk-free zero-coupon bond with a maturity equal to the duration of the corporate bond. We compute this duration based on the reported yield in the TRACE database and the corporate bond's cash flow structure. Note that we do not incorporate adjustments for optionalities or covenants included in the bond structure to determine the duration. Overall, yield spreads based on this duration adjustment can be considered as a proxy for the zero-coupon yield spread taken from a more complete pricing model.[7]

We use both the Treasury yield curve and the swap curve as risk-free benchmarks to calculate the bond yield spreads. We find that the general structure of the resulting yield spread is basically identical for both benchmarks. However, as expected, the yield spread based on the swap curve is shifted downwards compared to the spread based on the Treasury curve, indicating that the swap curve represents market participants with AA

[6] If positive covariances occur, we set the Roll measure to zero. Since we interpret the Roll measure as a transaction cost metric, we think it is quite reasonable to bound this measure at zero. However, we also compared the results with two alternatives: preserving the sign in the spirit of Roll (1984) (i.e., positive covariance translates into a negative Roll measure) and not using these observations at all. Neither of these changes affects the qualitative nature of our results.

[7] Given the complexity of these models and the limited information available for their calibration, we presume that the resulting zero-coupon yield spread would not improve the economic interpretation of our results, in general. To test this assumption, we have employed regression analyses for a subsample of straight coupon, bullet bonds without any option features. For this subsample, we find similar results, confirming our conjecture.

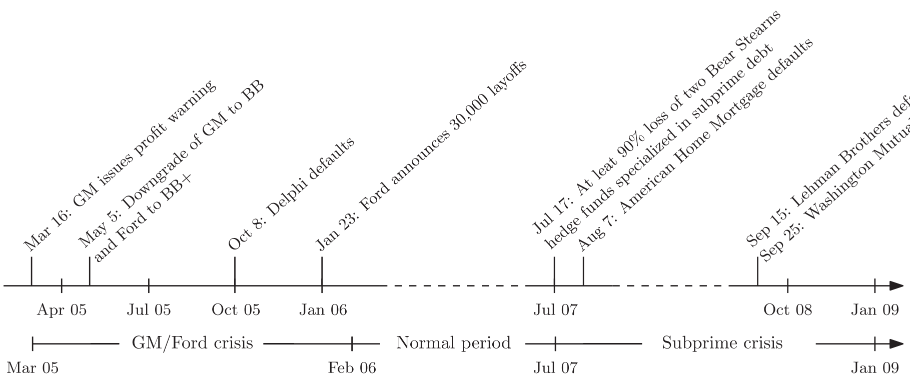


Fig. 1. Timeline showing important events in the US corporate bond market, since March 2005. Based on these events, we identified three different regimes: the GM/Ford crisis between March 2005 and January 2006, the normal period from February 2006 to June 2007, with no exceptional events, and the subprime crisis that started in July 2007.

ratings with greater credit risk, while the Treasury curve represents lower credit risk. We conduct all our regression analysis on both spread series; however, as the results are basically identical, we report only the results for the spreads against the Treasury benchmark in the empirical results section.

We calculate the bond yield spread for every price observation in the TRACE data set. Thus, we can have more than one spread observation for a given bond on a particular day, since there can be multiple trades for the bond on that day. Hence, to get a single value for the yield spread for each day, we estimate the bond spread from the individual observations by calculating a volume-weighted average for the day, i.e., we implicitly assume that the spread information is reflected more strongly in large trades.

6.2. *Subperiods of interest*

We are interested in how the explanatory power of the independent variables differs in financial crises compared to normal market environments. Therefore, we define the following three subperiods: The *GM/Ford crisis* (March 2005–January 2006) when a segment of the corporate bond market was affected, the *subprime crisis* (July 2007–December 2008), which was much more pervasive across the corporate bond market, and the *normal period* in between (February 2006–June 2007). We choose the start and end dates of the subperiods based on exceptional events that are believed to have affected market conditions (see Fig. 1).[8]

6.3. *Panel data regression*

We rely on a panel data regression approach to analyze bond yield spread changes. We use first differences, as we observe that yield spreads are integrated. Since we observe autocorrelated yield spread changes, we add one autoregressive parameter to our specifications.[9] Of course, in this difference specification, the static bond characteristic variables drop out. Thus, our panel consists of the pooled time-series of the first differences of the bond yield spread as the dependent variable and the trading activity variables and liquidity measures as the explanatory variables. Furthermore, we add changes in rating class dummies to the regression to consider credit risk-related effects on the yield spread:

$$\begin{aligned}\Delta(\text{Yield spread})_{i,t} = a_0 + a_1 \cdot \Delta(\text{Yield spread})_{i,t-1} &+ \boldsymbol{a_2} \cdot \Delta(\text{Trading activity variables})_{i,t} \\ &+ \boldsymbol{a_3} \cdot \Delta(\text{Liquidity measures})_{i,t} \\ &+ \boldsymbol{a_4} \cdot \Delta(\text{Rating dummies})_{i,t} + \epsilon_{i,t}. \end{aligned} \tag{4}$$

Our basic time-series data are at a daily frequency. However, because of computational restrictions due to the large sample size, we create weekly averages of all variables from the daily data for each bond. Thus, all the time-series regression results presented in the empirical results section below are based on weekly data. Note that we use logarithmic values of the traded volume in the regressions, as is common practice.

6.4. *Fama-MacBeth cross-sectional regression*

These regressions are in levels rather than in changes and, therefore, allow a cross-sectional analysis. In particular, we can test for the importance of static bond characteristics in explaining the cross-sectional differences in yield spread. The regressions are performed with the following structure:

$$\begin{aligned}(\text{Yield spread})_{i,t} = a_0 + \boldsymbol{a_1} \cdot (\text{Bond characteristics})_{i,t} &+ \boldsymbol{a_2} \cdot (\text{Trading activity variables})_{i,t} \\ &+ \boldsymbol{a_3} \cdot (\text{Liquidity measures})_{i,t} \\ &+ \boldsymbol{a_4} \cdot (\text{Rating dummies})_{i,t} + \epsilon_{i,t}. \end{aligned} \tag{5}$$

We run this regression based on weekly averages from the daily data of all variables. Thus, we have the

[8] Alternative definitions of these subperiods could have been used. Therefore, as a stability test, we varied the start and end dates of the subperiods by up to one month. However, we find similar results, and hence, report only results for the three subperiods defined above.

[9] We investigated alternative specifications of the time-series model, including different lags of the autoregressive parameters, and find that the results are very similar for these specifications.

Table 1
This table reports the cross-sectional descriptives statistics (5th, 25th, 50th, 75th, and 95th quantiles, mean, and standard deviation) for the yield spread, credit rating, bond characteristics, and liquidity proxies. The corporate bond yield spread is measured relative to the US Treasury bond yield curve and given in percentage points. We use credit ratings from Standard & Poor's where we assign integer numbers to ratings, i.e., AAA=1, AA+=2, etc., to measure the average rating. The liquidity proxies are classified into three groups: bond characteristics (amount issued, coupon, maturity, and age), trading activity variables (traded volume, number of trades, and time interval between trades), and liquidity measures (Amihud, price dispersion, Roll, and zero-return measure). For time-varying variables, the statistics are first averaged across time for each individual bond. The data set consists of 23,703 US corporate bonds traded over the period October 2004 to December 2008.

		$Q_{0.05}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.95}$	Mean	Std. dev.
	Yield spread (%)	0.52	1.17	1.92	3.67	7.67	2.87	2.95
	Rating	1.30	5.00	7.03	11.50	15.46	8.00	4.14
Bond characteristics	Amount issued (bln)	0.00	0.01	0.20	0.40	1.25	0.32	0.50
	Coupon (%)	3.03	5.00	5.97	7.00	9.13	5.98	1.87
	Maturity (yr)	0.45	1.83	5.20	9.74	24.87	7.62	7.63
	Age (yr)	0.47	1.52	2.77	4.53	10.36	3.80	3.61
Trading activity variables	Volume (mln)	0.02	0.05	0.39	1.74	5.23	1.35	2.53
	Trades	1.45	2.00	2.46	3.25	8.44	3.47	4.50
	Trading interval (dy)	1.50	2.72	4.48	5.72	7.80	4.46	2.18
Liquidity measures	Amihud (bp per mln)	0.68	9.99	38.33	103.01	260.68	78.38	137.21
	Price dispersion (bp)	1.69	16.08	33.64	58.00	106.84	41.53	35.56
	Roll (bp)	24.48	82.99	155.98	259.69	420.86	185.12	144.69
	Zero-return (%)	0.00	0.00	0.01	0.02	0.16	0.03	0.08

cross-sectional regression result for each week and we use the Fama-MacBeth procedure to report the regression parameters and *t*-statistics. We present the results of this procedure for the subperiods defined earlier. This approach allows us to analyze liquidity effects in times of regular market conditions and financial crises, across bonds. Again, we use logarithmic values of the traded volume and the amount issued in the regressions.

7. Results

7.1. Descriptive statistics

This section provides summary statistics for the US corporate bond market based on our matched data sample of 23,703 bonds for the period October 2004 to December 2008 (see Section 4). Table 1 reports the cross-sectional variation of the main variables used in our empirical analysis, i.e., the yield spread, the credit rating, and the liquidity proxies (bond characteristics, trading activity variables, and liquidity measures). For time-varying explanatory variables, the statistics are computed as the time averages for each individual bond. The table reports the 5th, 25th, 50th, 75th, and 95th percentiles, as well as the mean and standard deviation of each variable. It provides an aggregate picture of the substantial cross-sectional variation of the variables.

The yield spread between the 5th and 95th percentiles ranges from 52 to 767 basis points (bp) with a mean of 287 bp. Part of this enormous variation is obviously due to credit risk given that our sample contains bonds with credit ratings all the way from AAA (=1) to C (=21). The average credit rating is roughly eight which corresponds to BBB+ and a standard deviation of approximately four rating notches.

As is to be expected, there is a reasonable variation in the bond characteristics of amount issued, coupon, maturity, and age across bonds, e.g., the amount issued varies from just below $5 million to $1.25 billion between the 5th and 95th percentiles. Regarding trading activity variables, we find that the average frequency of bond trading is every 4.5 days. For a bond that is traded on a particular day, we observe an average of 3.5 trades with an average trade size of roughly $1.4 million dollars, with substantial cross-sectional variation.

Regarding the liquidity measures, the mean value of the Amihud measure is 78.4 bp per million, which indicates that trading one million dollars in a particular bond shifts the price by 78.4 bp, on average. The variation in liquidity across bonds is remarkably high and ranges between 0.7 and 260.7 bp, a factor of around 400 for the 5th and the 95th percentiles. The price dispersion indicates the trading cost of a single transaction for which we observe a mean of around 41.5 bp with high variation across bonds as well. For the Roll measure, which corresponds to the round-trip costs, we observe an average value of 185.1 bp. Interestingly, this mean value is more than twice as large as the mean value of the price dispersion measure. Considering the zero-return measure, we find that these are mostly zero, indicating only very few observations of stale prices or quotations.

Table 2 presents the correlations between the various liquidity proxies within our panel data. Overall, we find the expected patterns: in general, there is positive correlation among the trading activity variables (e.g., the correlation between volume and number of trades is 0.51) and among the liquidity measures estimating trading costs (e.g., the correlation between Amihud, Roll, and price dispersion measure is between 0.02 and 0.2). However, the general level of correlation appears to be relatively low, especially for the liquidity measures. Thus, correlation measured at the individual bond level over time shows that the liquidity proxies have substantial idiosyncratic movements. This result suggests that the

Table 2
This table presents the correlation matrix of bond characteristics (amount issued, coupon, maturity, and age), trading activity variables (traded volume, number of trades, and time interval between trades), and liquidity measures (Amihud, price dispersion, Roll, and zero-return measure) based on the panel data for pairwise complete observations. The data set consists of weekly averages of all variables and consists of 23,703 US corporate bonds traded over the period October 2004 to December 2008.

	Amount issued	Coupon	Maturity	Age	Volume	Trades	Trading interval	Amihud	Price dispersion	Roll	Zero-return
Amount issued	1.00										
Coupon	−0.01	1.00									
Maturity	−0.05	0.14	1.00								
Age	−0.12	0.20	−0.03	1.00							
Volume	0.46	0.03	0.05	−0.13	1.00						
Trades	0.46	−0.02	−0.03	−0.01	0.51	1.00					
Trading interval	−0.28	−0.02	0.03	0.03	−0.15	−0.17	1.00				
Amihud	−0.13	−0.01	0.09	0.03	−0.09	−0.07	0.11	1.00			
Price dispersion	0.06	0.08	0.28	0.01	0.01	0.14	−0.11	0.02	1.00		
Roll	−0.29	0.03	0.29	0.08	−0.21	−0.13	0.15	0.19	0.20	1.00	
Zero-return	−0.04	0.14	−0.01	−0.04	−0.02	−0.06	−0.01	−0.03	−0.01	−0.05	1.00

various liquidity proxies are measuring somewhat different aspects of liquidity empirically, although at a conceptual level they are related.[10] Therefore, for our empirical work, the issues of multicollinearity may not be as severe as one may suspect, at first glance. Note that once correlations in our sample are measured at a more aggregate level (e.g., averaging across time or across bonds), the correlations are much higher. Thus, it is important not to analyze the bond market based solely on aggregated data, but also at the level of individual bonds, as we do here, to distinguish between the effects of the various liquidity proxies.

To gain a better understanding of the time-series behavior of the bond yield spread over the whole time period, we compute the count-weighted average of the daily yield spreads over all bonds in our sample.[11] Fig. 2 shows this time-series of the market-wide average corporate bond yield spread, indicating the dramatic increase of the spread during the two crisis periods. Especially during the subprime crisis, we observe a sharp increase in the yield spread, which rose, on average, from around 2% to 10%, most likely indicating a far higher risk premium for illiquidity and credit risk.

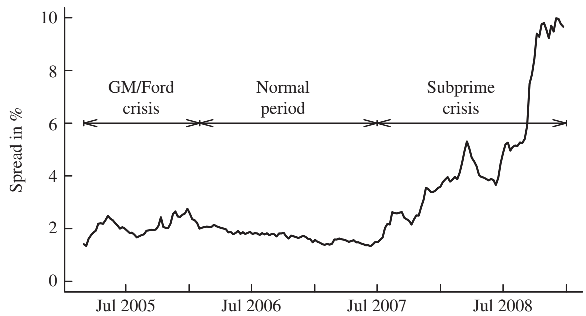


Fig. 2. This figure shows the market-wide corporate bond yield spread between October 2004 and December 2008 computed by averaging the bond yield spreads across bonds traded. The corporate bond yield spread is measured relative to the US Treasury bond yield curve and given in percentage points. The data set consists of 23,703 US corporate bonds traded over the period October 2004 to December 2008.

7.2. Liquidity effects in corporate bond yield spreads

In this section, we examine whether liquidity effects are priced in the US corporate bond market. As argued in Section 3, we expect to find a significant liquidity premium in bond yield spreads. We base our conclusions in this section on our overall sample covering the whole market for the time period of our sample. We present the empirical results explaining the time-series properties of the bond yield spread changes with the credit ratings and the liquidity proxies introduced in Section 5, and using the panel data regression methodology presented in Section 6.

The regressions are based on a sample of data consisting of 691,016 bond-week observations. The results are shown in Table 3. This table presents four different specifications. In Regression 1, we use a specification without the liquidity proxies, which is a base case that can be compared to the other specifications, allowing us to explore the increase in explanatory power after including liquidity proxies. Note that there is reasonable explanatory power even in this specification, which includes the information contained in the dummy variables based on the credit ratings and the persistence of bond yield spreads in terms of first differences measured by the lagged term. The next three specifications present the results of the panel regressions using the liquidity proxies (i.e., trading activity variables and liquidity measures).[12] Regression 2 reports the results with the trading activity

[10] Along the same lines, a principal component analysis (not reported here) shows that the liquidity proxies can only be represented by a relatively large number of components.

[11] We also examine the behavior of bond yield spreads, weighted by the volume of trading and by the amount outstanding of the individual bonds, both of which show a similar pattern.

[12] Since the regressions are based on the *change* in the bond yield spread, the static bond characteristics, such as coupon, drop out of the specification since they are fixed effects. Others, such as age, vary linearly with time and are absorbed in the constant term.

Table 3
This table reports the panel data regression models explaining the yield spread changes based on weekly averages of all variables:

$$\Delta(\text{Yield spread})_{i,t} = \alpha_0 + \alpha_1 \cdot \Delta(\text{Yield spread})_{i,t-1} + \alpha_2 \cdot \Delta(\text{Volume})_{i,t} + \alpha_3 \cdot \Delta(\text{Trades})_{i,t} + \alpha_4 \cdot \Delta(\text{Trading interval})_{i,t}$$
$$+ \alpha_5 \cdot \Delta(\text{Amihud})_{i,t} + \alpha_6 \cdot \Delta(\text{Price dispersion})_{i,t} + \alpha_7 \cdot \Delta(\text{Roll})_{i,t} + \alpha_8 \cdot \Delta(\text{Zero-return})_{i,t} + \sum_{k=1}^{21} \beta_k \cdot \Delta(\text{Rating dummy})_{i,t,k} + \epsilon_{i,t}.$$

The yield spread change is explained by the change in the lagged yield spread, trading activity variables (traded volume, number of trades, and time interval between trades), liquidity measures (Amihud, price dispersion, Roll, and zero-return measure), and rating dummies to control for credit risk. The corporate bond yield spread is measured relative to the US Treasury bond yield curve. In Regression (1) we use a specification without the liquidity proxies. Regression (2) reports the results with the trading activity variables only, while Regression (3) reports them with the liquidity measures only. In Regression (4) we add both types of liquidity proxies. The *t*-statistics are given in parentheses and are calculated from Newey and West (1987) standard errors, which are corrected for heteroskedasticity and serial correlation. In addition, the table also reports each model's R^2 and the number of observations. The data set consists of 23,703 US corporate bonds traded over the period October 2004–December 2008.

	(1)	(2)	(3)	(4)
Intercept	0.0731***	0.0726***	0.0721***	0.0717***
	(73.6195)	(76.4741)	(76.0680)	(77.1532)
$\Delta(\text{Yield spread})_{i,t-1}$	−0.2853***	−0.2825***	−0.2816***	−0.2797***
	(73.6195)	(−44.1891)	(−43.7139)	(−44.0573)
$\Delta(\text{Volume})_{i,t}$		−0.0204***		−0.0108***
		(−23.2748)		(−12.9274)
$\Delta(\text{Trades})_{i,t}$		0.0067***		0.0054***
		(16.0245)		(12.8891)
$\Delta(\text{Trading interval})_{i,t}$		0.0068***		0.0070***
		(19.4614)		(20.2071)
$\Delta(\text{Amihud})_{i,t}$			0.0502***	0.0477***
			(35.0938)	(33.8126)
$\Delta(\text{Price dispersion})_{i,t}$			0.0744***	0.0702***
			(26.7098)	(25.4506)
$\Delta(\text{Roll})_{i,t}$			0.0510***	0.0512***
			(16.0256)	(16.1036)
$\Delta(\text{Zero-return})_{i,t}$			−0.0774***	−0.0696***
			(−8.4762)	(−7.6153)
Δ(Rating dummies)	Yes	Yes	Yes	Yes
R^2	0.0735	0.0766	0.0839	0.0856
Observations	691,016	691,016	691,016	691,016

variables only, while Regression 3 reports them with the liquidity measures only. Regression 4 includes both types of liquidity proxies.

Focusing on the model that includes both types of proxies, the results of Regression 4 show that all the liquidity proxies are statistically significant in explaining the changes in the bond yield spreads. Among the trading activity variables, changes in the volume and trading interval have the highest *t*-statistics, while among the liquidity measures, changes in the Amihud measure and the price dispersion measure are most important. Interestingly, despite the correlation across the liquidity measures, they are sufficiently different across bonds and time that they are all incrementally relevant in explaining the changes in the bond yield spreads. All variables have the expected signs except for two liquidity proxies: the number of trades and the zero-return measure. Regarding the zero-return measure, the economic significance of this measure is very low, as discussed below. Thus, we assume that this measure might not be meaningful. As for the number of trades, we find that an increase in the number of trades increases the yield spread. This result can arise if, in times of crisis, regular trades are split up in smaller trades due to a general reduction in trade size or sellside pressure, as more institutional orders are broken up to be placed in the market. This aspect will be analyzed in the next section.

In terms of R^2, we find a relative improvement of about 4.2% when the trading activity variables are added to Regression 1. When we add the liquidity measures to Regression 1, we find an additional relative improvement of 9.5% in the R^2, showing that liquidity measures are more important compared to the trading activity variables. Overall, we find that liquidity effects account for approximately 14% of the explained market-wide corporate yield spread changes.

The Amihud measure turns out to be the most important explanatory variable in these regressions in economic terms. A one standard deviation change in the Amihud measure explains about 6.1 basis points of the change in the bond yield spread in Regression 4.[13] Similar statistics for the price dispersion measure and the Roll measure are 3.4 bp and 3.1 bp, respectively. The impact of the trading

[13] Note that the calculation of the economic significance is based on the standard deviation of the first differences of the variables. Due to space limitations we do not report details concerning these statistics for the first differences.

interval, volume, and number of trades are 2.5, 1.8, and 1.5 bp, respectively. The smallest impact is provided by the zero-return measure (0.8 bp), which seems not to be particularly relevant given its low economic significance. Considering all liquidity proxies together, a one standard deviation move in the direction of greater illiquidity in all proxies would increase the yield spread by 19.2 bp. This effect is important when compared with the volatility of the yield spread changes of 75.6 bp.[14,15]

Overall, we find in this analysis that liquidity is an important factor driving yield spread changes. Liquidity measures as well as trading activity variables can explain a fair proportion of bond yield spread changes; in particular, liquidity measures estimating trading costs seem to be more important than pure trading activity measures.

7.3. Liquidity effects in periods of financial distress

In this section, we explore whether the effect of liquidity is stronger during times of financial crises. As argued in Section 3, we expect that liquidity is an even more important factor in times of distress. To focus on the role of liquidity in financial crises, we analyze three different subperiods of our overall sample. We present the results for the two different crisis periods (the GM/Ford crisis and the subprime crisis) and compare them with those for the period in between, which can be considered as a period with more normal market conditions. We first provide evidence on the descriptive statistics of the key variables for the three subperiods, and then draw our main conclusions based on the panel data and Fama-MacBeth regressions introduced in Section 6.

The analysis of the averages of the variables in these three subperiods allows us to gain some important insights into the causes of the variation (see Table 4). The top panel of the table presents the average yield spread and the credit rating as well as information about the average daily market-wide trading activity (i.e., number of traded bonds, trades, and volume). The bottom panel provides the liquidity proxies computed for each subperiod.

The average yield spread in the normal period of 1.9% is less than in the GM/Ford crisis with 2.3%, and even less so than in the subprime crisis with 5.0%, documenting the strong impact of this crisis on yield spreads for the whole market. This evidence is also visible in Fig. 2. The averages of the market-wide trading activity variables are also illustrative. During both crises, trading activity is lower, in terms of the number of traded bonds and trade volume, than in the normal period. This reduction is more severe in the subprime crisis. For example, the number of bonds traded each day dropped during the subprime crisis, from roughly 6,000 on average, to a little under 5,200. The volume of trading showed a similar decline. Interestingly, during the subprime crisis, we find a larger number of trades indicating relatively smaller trade sizes for this period. Overall, the impact on trading activity is more severe in the subprime crisis, indicating that the liquidity changes that occurred during the two crisis periods were different. During the GM/Ford crisis, there was some shuffling of bond portfolios to account for the shifts in credit ratings, particularly in the automobile sector, resulting only in a minor reduction of trading activity. In contrast, during the subprime crisis, overall market liquidity was affected. This point is also evidenced by the changes in the average credit rating in the different subperiods. The credit rating of the average bond traded during the GM/Ford crisis was somewhat worse than during normal times. In contrast, the credit rating of the average bond traded during the subprime crisis was better than during normal times, indicating a *flight-to-quality* during the subprime crisis: the average rating is 8.8 (close to BBB) for the GM/Ford crisis, 8.4 (between BBB and BBB+) for the normal period, and 7.6 (between BBB+ and A−) for the subprime crisis.

The bottom panel of Table 4 presents similar evidence for the averages of the daily bond-level liquidity proxies. All liquidity measures indicate lower liquidity in times of crisis, especially for the subprime crisis. Considering the average price dispersion measure, as one example, we find that the average value is higher in both crises (46.4 bp in the GM/Ford crisis and 70.0 bp in the subprime crisis) compared to the normal period (39.8 bp). With regard to the trading activity variables, we find that the average daily volume and the trade interval at the bond-level stay approximately at the same level. However, the number of trades increases in both crises. These results are consistent with the level of market-wide trading activity, where we find that, in crises, trading takes place in fewer bonds, with a larger number of smaller size trades.

We next analyze the behavior of the changes in the yield spreads in the different subperiods, using the panel data regression, as in the previous section, incorporating dummy variables for the subperiods. More importantly, we include interaction terms between the liquidity proxies with the dummy variables for the two crisis subperiods. This setup allows us to analyze whether the

[14] Since credit ratings might adjust slowly compared to changes in credit risk, part of the explanatory power of the liquidity variables in our regressions could result because these variables might be proxies for changes in credit risk. As a robustness check, we test whether adding future rating changes (i.e., assuming perfect foresight) to the regressions affects the coefficients of the liquidity variables. In our tests (the results of which are not reported here to conserve space), we add weekly rating changes for each of the next 12 weeks to the regression equation in column 4 of Table 3. We find that future rating changes are statistically significant, i.e., ratings indeed change slowly. More importantly, however, we find the same results as in the original regression for the liquidity variables, i.e., "perfect-foresight" rating information does *not* take explanatory power away from the liquidity variables. Thus, our original results are confirmed and there is no evidence that the liquidity variables are proxies for credit risk information. Rather, future rating changes explain part of the current changes in yield spreads, *in addition to* changes in the liquidity variables. Furthermore, we test whether the liquidity variables can forecast future rating changes. Again, we find no evidence for this conjecture.

[15] As a robustness check for the use of ratings as a credit risk proxy, we instead use CDS spreads. We are able to match a small sample (representing the rather more liquid issues) with 5-year CDS spreads obtained from Markit. We then repeat our regression analysis using this CDS spread variable. The R^2 is only marginally improved in these regressions and for our liquidity measures, we find essentially the same results as in the analysis based on ratings, i.e., the coefficients and statistical significance stay at the same levels, thus strengthening the robustness of our results.

Table 4

Panel A shows the mean and standard deviation for the yield spread, credit rating, and daily market-wide trading activity in the three regimes (GM/Ford crisis, normal period, and subprime crisis). The corporate bond yield spread is measured relative to the US Treasury bond yield curve and given in percentage points. We use credit ratings from Standard & Poor's where we assign integer numbers to ratings, i.e., AAA = 1, AA+ = 2, etc., to measure the average rating. The market-wide trading activity variables represent the number of traded bonds and trades, and the total trading volume per day. Panel B shows the mean and the standard deviation for the bond characteristics (amount issued, coupon, maturity, and age), trading activity variables (traded volume, number of trades, and time interval between trades), and liquidity measures (Amihud, price dispersion, Roll, and zero-return measure). The data set consists of 23,703 US corporate bonds traded over the period October 2004–December 2008.

Panel A: Yield-spread, rating, and market-wide trading activity

	Mean			Standard deviation		
	GM/Ford crisis	Normal period	Subprime crisis	GM/Ford crisis	Normal period	Subprime crisis
Yield spread (%)	2.34	1.88	5.00	0.43	0.25	2.38
Rating	8.82	8.38	7.63	0.15	0.36	0.28
Traded bonds (thd)	5.23	5.92	5.19	0.54	0.42	0.53
Market-wide trades (thd)	20.43	20.71	22.77	2.34	2.05	4.67
Market-wide volume (bln)	7.65	8.06	6.99	1.32	1.41	1.56

Panel B: Liquidity proxies

	Mean			Standard deviation		
	GM/Ford crisis	Normal period	Subprime crisis	GM/Ford crisis	Normal period	Subprime crisis
Amount issued (bln)	0.43	0.45	0.54	0.01	0.03	0.06
Coupon (%)	6.26	6.24	6.23	0.06	0.05	0.05
Maturity (yr)	7.57	7.75	8.31	0.13	0.20	0.18
Age (yr)	3.91	4.36	4.76	0.09	0.12	0.14
Volume (mln)	1.51	1.44	1.53	0.26	0.22	0.32
Trades	4.48	4.06	5.33	0.48	0.24	1.31
Trading interval (dy)	3.31	3.38	3.37	0.44	0.49	0.48
Amihud (bp per mln)	66.48	53.21	89.20	6.13	7.42	35.75
Price dispersion (bp)	46.36	39.75	70.02	4.33	1.83	21.84
Roll (bp)	164.28	142.82	209.77	9.57	10.57	52.34
Zero-return (%)	0.02	0.02	0.03	0.01	0.01	0.01

yield spread changes are more sensitive to liquidity changes in times of crisis. The results are presented in Table 5.

Overall, we find that liquidity is far more important in times of crisis. During the subprime crisis period, we find that nearly all the liquidity proxies have a statistically significantly higher impact on the changes in the bond yield spreads. Again, this result suggests that the various liquidity proxies are measuring somewhat different aspects of liquidity, as already indicated by the low level of correlation (see Section 7.1). The most important ones are the price dispersion and the Amihud measure, where both coefficients basically increase by around 100%. A similar result can be found for the GM/Ford period, although the effects are not quite as strong. We do not observe a statistically significant increase in all of the proxies for the GM/Ford period, and also, the magnitude of the increase seems to be smaller. However, an F-test shows that we can reject at a 1% level the hypothesis that the interaction terms for each period of crisis are jointly zero.

In terms of the improvement in R^2, we find that the inclusion of the interaction terms leads to an increase from 8.56% to 10.14%, compared to the analysis for the whole time-series, highlighting the importance of adding these terms. Considering the economic significance, a one standard deviation move in all proxies in the direction of greater illiquidity would increase the spread by 11.6 bp in the normal period compared to 15.2 bp and 25.9 bp in the GM/Ford and subprime crisis periods, respectively. Thus, we find a far higher impact of the liquidity proxies in the crisis periods: the economic significance more than doubles during the subprime crisis and increases by approximately 30% in the GM/Ford crisis. The ranking of the economic importance of the individual liquidity proxies in the different time periods stays approximately the same, with the Amihud measure showing the highest impact in all periods (4.3 bp in the normal period, 5.2 bp in the GM/Ford period, and 7.7 bp in the subprime period). In sum, we find a significant increase, in both statistical and economic terms, of the liquidity component in the crisis periods.

To widen the scope of the analysis, we explore the cross-sectional differences in explaining the bond yield spreads considering all liquidity proxies using the Fama-MacBeth procedure to report the results for the three subperiods.[16] Again, rating class dummies are used to explain credit risk-related differences in spreads across bonds.

[16] Since the t-statistics of the Fama-MacBeth regression could potentially be biased due to serial correlation, we additionally calculate results for a cross-sectional regression on time-series averages. This robustness check (not presented here) shows that the variables are, again, statistically significant.

Table 5

This table reports the panel data regression model explaining the yield spread changes based on weekly averages of all variables:

$$\begin{aligned}\Delta(\text{Yield spread})_{i,t} &= \alpha_0+\alpha_1\cdot\Delta(\text{Yield spread})_{i,t-1}+\alpha_2\cdot\Delta(\text{Volume})_{i,t}+\alpha_3\cdot\Delta(\text{Trades})_{i,t}+\alpha_4\cdot\Delta(\text{Trading interval})_{i,t}\\ &+\alpha_5\cdot\Delta(\text{Amihud})_{i,t}+\alpha_6\cdot\Delta(\text{Price dispersion})_{i,t}+\alpha_7\cdot\Delta(\text{Roll})_{i,t}+\alpha_8\cdot\Delta(\text{Zero-return})_{i,t}+(\text{GM/Ford dummy})_t\times[\beta_1\cdot\Delta(\text{Yield spread})_{i,t-1}\\ &+\beta_2\cdot\Delta(\text{Volume})_{i,t}+\beta_3\cdot\Delta(\text{Trades})_{i,t}+\beta_4\cdot\Delta(\text{Trading interval})_{i,t}+\beta_5\cdot\Delta(\text{Amihud})_{i,t}+\beta_6\cdot\Delta(\text{Price dispersion})_{i,t}\\ &+\beta_7\cdot\Delta(\text{Roll})_{i,t}+\beta_8\cdot\Delta(\text{Zero-return})_{i,t}]+(\text{Subprime dummy})_t\times[\gamma_1\cdot\Delta(\text{Yield spread})_{i,t-1}+\gamma_2\cdot\Delta(\text{Volume})_{i,t}+\gamma_3\cdot\Delta(\text{Trades})_{i,t}\\ &+\gamma_4\cdot\Delta(\text{Trading interval})_{i,t}+\gamma_5\cdot\Delta(\text{Amihud})_{i,t}+\gamma_6\cdot\Delta(\text{Price dispersion})_{i,t}+\gamma_7\cdot\Delta(\text{Roll})_{i,t}+\gamma_8\cdot\Delta(\text{Zero-return})_{i,t}]\\ &+\sum_{k=1}^{21}\delta_k\cdot\Delta(\text{Rating dummy})_{i,t,k}+\epsilon_{i,t}.\end{aligned}$$

The yield spread change is explained by the change in the lagged yield spread, trading activity variables (traded volume, number of trades, and time interval between trades), liquidity measures (Amihud, price dispersion, Roll, and zero-return measure), and rating dummies to control for credit risk. Additionally, we add interaction terms between the subperiod dummies and the liquidity proxies. The corporate bond yield spread is measured relative to the US Treasury bond yield curve. The t-statistics are given in parentheses and are calculated from Newey and West (1987) standard errors, which are corrected for heteroskedasticity and serial correlation. We provide an F-test to test whether the interaction terms of the dummy variable with the liquidity proxies are jointly zero. The standard errors of the F-statistics are also Newey and West (1987) corrected. In addition, the table also reports the model's R^2 and the number of observations. The data set consists of 23,703 US corporate bonds traded over the period October 2004–December 2008.

Intercept	0.0644***	(70.5469)
Δ(Yield spread)$_{i,t-1}$	−0.4318***	(−47.6400)
Δ(Volume)$_{i,t}$	−0.0137***	(−15.8406)
Δ(Trades)$_{i,t}$	0.0030***	(7.3703)
Δ(Trading interval)$_{i,t}$	0.0032***	(7.8087)
Δ(Amihud)$_{i,t}$	0.0332***	(19.7728)
Δ(Price dispersion)$_{i,t}$	0.0417***	(13.0552)
Δ(Roll)$_{i,t}$	0.0080***	(2.7972)
Δ(Zero-return)$_{i,t}$	−0.0495***	(−4.5916)
(GM/Ford dummy)$_t$ × Δ(Yield spread)$_{i,t-1}$	0.0366***	(3.6263)
(GM/Ford dummy)$_t$ × Δ(Volume)$_{i,t}$	−0.0028**	(−1.9884)
(GM/Ford dummy)$_t$ × Δ(Trades)$_{i,t}$	0.0019***	(2.6333)
(GM/Ford dummy)$_t$ × Δ(Trading interval)$_{i,t}$	−0.0040***	(−6.1069)
(GM/Ford dummy)$_t$ × Δ(Amihud)$_{i,t}$	0.0069***	(2.6422)
(GM/Ford dummy)$_t$ × Δ(Price dispersion)$_{i,t}$	0.0046	(0.9846)
(GM/Ford dummy)$_t$ × Δ(Roll)$_{i,t}$	−0.0029	(−0.7340)
(GM/Ford dummy)$_t$ × Δ(Zero-return)$_{i,t}$	0.0239	(1.3503)
(Subprime dummy)$_t$ × Δ(Yield spread)$_{i,t-1}$	0.2260***	(19.5475)
(Subprime dummy)$_t$ × Δ(Volume)$_{i,t}$	0.0135***	(6.5583)
(Subprime dummy)$_t$ × Δ(Trades)$_{i,t}$	0.0039***	(4.4352)
(Subprime dummy)$_t$ × Δ(Trading interval)$_{i,t}$	−0.0070***	(−8.2507)
(Subprime dummy)$_t$ × Δ(Amihud)$_{i,t}$	0.0266***	(9.4860)
(Subprime dummy)$_t$ × Δ(Price dispersion)$_{i,t}$	0.0529***	(9.4447)
(Subprime dummy)$_t$ × Δ(Roll)$_{i,t}$	0.0777***	(13.6655)
(Subprime dummy)$_t$ × Δ(Zero-return)$_{i,t}$	−0.0824***	(−3.5895)
Δ(Rating dummies)	Yes	
F-stat. $\mathcal{H}_0$: (GM/Ford dummy) × Δ(Liquidity proxies) = 0		7.8864
F-stat. $\mathcal{H}_0$: (Subprime dummy) × Δ(Liquidity proxies) = 0		64.3814
Observations		691,016
R^2		0.1014

Table 6 provides the detailed results. The findings for the individual measures basically confirm the results of the panel data analysis, i.e., based on the t-statistics, liquidity measures are more important than trading activity variables; among the liquidity measures, the Amihud measure and the price dispersion measures are the most important proxies. As in the panel data analysis, we find an unexpected sign for the number of trades. Interestingly, the bond characteristics are important liquidity proxies in explaining the cross-section, as well. The most important one is the amount issued with a high overall t-statistic. Thus, high outstanding amounts indicate higher liquidity. The coefficient of the coupon variable indicates higher liquidity for bonds with lower coupons. As expected, a longer time-to-maturity indicates lower liquidity for bonds in the normal period and in the GM/Ford crisis. However, the effect is negative for the subprime period. This result could indicate that, for "buy-and-hold" bonds with long maturities, the selling pressure was not as high as for bonds with shorter maturities resulting in lower spreads.

We find that a large part of the cross-sectional differences in the yield spread across bonds can be explained by our specification, indicated by an R^2 ranging between

Table 6
This table reports the cross-sectional regression models explaining the weekly averages of yield spreads based on the Fama-MacBeth procedure, estimated for the three regimes (GM/Ford crisis, normal period, and subprime crisis):

$$\begin{aligned}(\text{Yield spread})_{i,t} &= \alpha_0+\alpha_1\cdot(\text{Amount issued})_{i,t}+\alpha_2\cdot(\text{Coupon})_{i,t}\\ &+\alpha_3\cdot(\text{Maturity})_{i,t}+\alpha_4\cdot(\text{Age})_{i,t}+\alpha_5\cdot(\text{Volume})_{i,t}\\ &+\alpha_6\cdot(\text{Trades})_{i,t}+\alpha_7\cdot(\text{Trading interval})_{i,t}\\ &+\alpha_8\cdot(\text{Amihud})_{i,t}+\alpha_9\cdot(\text{Price dispersion})_{i,t}\\ &+\alpha_{10}\cdot(\text{Roll})_{i,t}+\alpha_{11}\cdot(\text{Zero-return})_{i,t}\\ &+\sum_{k=1}^{21}\beta_k\cdot(\text{Rating dummy})_{i,t,k}+\epsilon_{i,t}.\end{aligned}$$

The level of the yield spread is explained by bond characteristics (amount issued, coupon, maturity, and age), trading activity variables (traded volume, number of trades, and time interval between trades), liquidity measures (Amihud, price dispersion, Roll, and zero-return measure), and rating dummies to control for credit risk. The corporate bond yield spread is measured relative to the US Treasury bond yield curve. The *t*-statistics are given in parentheses and are calculated from Newey and West (1987) standard errors, which are corrected for heteroskedasticity and serial correlation. The table also reports each model's R^2, and the number of observations, representing the average number of bonds in the weekly cross-sectional regressions. The data set consists of 23,703 US corporate bonds traded over the period October 2004–April 2008.

	GM/Ford crisis	Normal period	Subprime crisis
Intercept	1.8476***	1.4437***	4.4413***
	(16.2425)	(24.7543)	(3.7724)
$(\text{Amount issued})_{i,t}$	−0.2539***	−0.1824***	−0.3250***
	(−27.2371)	(−14.5819)	(−10.0361)
$(\text{Coupon})_{i,t}$	0.1567***	0.1142***	0.3506**
	(5.8644)	(27.9038)	(2.3741)
$(\text{Maturity})_{i,t}$	0.0110***	0.0177***	−0.0599***
	(3.6594)	(11.6510)	(−3.6448)
$(\text{Age})_{i,t}$	0.0053**	−0.0038	−0.0430***
	(2.4118)	(−0.8124)	(−3.2945)
$(\text{Volume})_{i,t}$	0.0013	−0.0113**	0.0432**
	(0.2851)	(−2.4204)	(2.4711)
$(\text{Trades})_{i,t}$	0.0452***	0.0316***	0.0335*
	(16.7888)	(13.4338)	(1.8092)
$(\text{Trading interval})_{i,t}$	0.0073***	0.0025**	0.0062
	(6.0842)	(1.9338)	(1.0638)
$(\text{Amihud})_{i,t}$	0.0864***	0.0718***	0.1696***
	(28.4551)	(29.4427)	(12.5740)
$(\text{Price dispersion})_{i,t}$	0.3496***	0.2742***	0.4523***
	(14.0045)	(18.8816)	(7.8913)
$(\text{Roll})_{i,t}$	0.0721***	0.0808***	0.1133*
	(8.2594)	(23.4407)	(1.9265)
$(\text{Zero-return})_{i,t}$	0.2371***	0.0387	0.6128
	(4.1320)	(0.7780)	(1.5735)
(Rating dummies)	Yes	Yes	Yes
R^2	0.5905	0.6016	0.4966
Observations	3,815	3,845	3,187

49.7% and 60.2% in the three subperiods. The relative improvement in R^2 when considering the liquidity proxies (not presented in the tables) is around 10%. Interestingly, this ratio stays at the same level in all three subperiods. Thus, we cannot observe an increase of explanatory power due to liquidity proxies in the crisis periods. It seems that especially in the subprime crisis, the spread levels of *all* bonds increased, and thus, the cross-sectional variation did not change dramatically. Considering the credit risk component of the yield spreads, the results clearly show the importance of the rating class dummies in the cross-section, as the remaining 90% of the explanatory power stems from this credit risk proxy. However, the lower R^2 in the subprime crisis results from a decrease in the explanatory power of the credit ratings, indicating that ratings could have become stale and reacted rather slowly to the increase in credit risk.

When analyzing the economic effect, we find that the cross-sectional variation of the yield spread measured by the standard deviation is 200.5 bp. With regard to the economic effect of the liquidity proxies based on the Fama-MacBeth regressions, we find statistically significant results, in terms of the coefficients of the relevant dummy variables: e.g., the Amihud measure and the price dispersion measure show strong effects; a one standard deviation change explains around 12.1 bp and 17.1 bp, respectively. The effects are more pronounced in the crisis periods compared to the normal period, e.g., for the price dispersion measure, the economic significance is 11.3 bp in the normal period vs. 15.4 bp and 24.4 bp in the GM/Ford and subprime crisis, respectively. Again, the zero-return measure shows the lowest economic effect of around 2.5 bp. A one standard deviation move in all the liquidity proxies in the direction of greater illiquidity would increase the spread by 98.9 bp in the normal period, compared to 111.9 bp and 153.1 bp, respectively, in the GM/Ford and subprime crisis periods. Thus, we find a higher impact of the liquidity proxies in the crises periods.[17]

Overall, the panel data and Fama-MacBeth regressions show a significant increase, in both statistical and economic terms, of the liquidity component in the crisis periods. We observe a dramatic increase in the liquidity premium, especially during the subprime crisis. Furthermore, we find that beyond liquidity measures and trading activity variables, simple bond characteristics, such as the amount issued, are also of importance in explaining liquidity.

7.4. Interaction effects between liquidity and credit ratings

In this section, we explore whether the effect of liquidity is related to credit risk measured by credit ratings. We divide the bonds into investment grade (AAA to BBB−) and speculative grade (BB+ to C/CCC), expecting the liquidity effects of speculative grade bonds to be more pronounced. This analysis allows us to explore the interaction between credit and liquidity risk. We expect to find lower liquidity effects for investment grade bonds compared to speculative grade bonds, as argued in Section 3.

Fig. 3 shows the yield spreads for the two time-series at the market-wide level. As expected, the bond yield

[17] As a robustness test for causality between liquidity proxies and yield spreads, we estimated all cross-sectional regressions using liquidity variables lagged by one week instead of contemporaneous ones. We find that the lagged liquidity proxies show basically the same explanatory power as the contemporaneous proxies in the cross-sectional regressions.

spread for investment grade bonds is always lower than that for speculative bonds. However, we stress three important points here: First, the GM/Ford crisis is mainly reflected in the speculative grade yield spreads, as the GM/Ford bonds were downgraded to junk bond status and probably had spillover effects in the whole corporate bond market. Second, in the normal period, the difference between the spreads of investment and speculative grade bonds systematically shrank over time reflecting decreasing risk premiums, a phenomenon that has received widespread attention in the popular press. Third, in the subprime crisis, the spread series for both investment and speculative grade bonds increased dramatically.

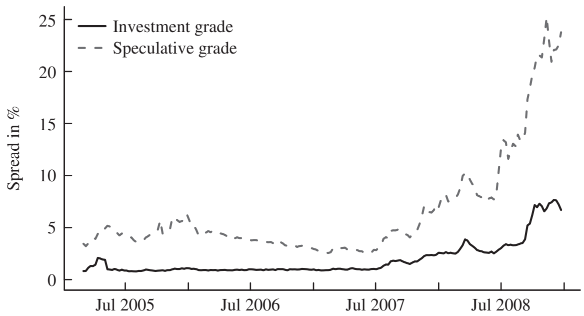


Fig. 3. This figure shows the corporate bond yield spread of investment grade and speculative grade bonds computed by averaging the bond yield spreads across bonds traded. The corporate bond yield spread is measured relative to the US Treasury bond yield curve and given in percentage points. The data set consists of 23,703 US corporate bonds traded over the period October 2004 to December 2008.

Table 7 (Panel A) presents the descriptive statistics of the yield spread, credit rating, and market-wide trading activity for the two subsegments in the three different time periods. We find that, in general, trading is focused on the investment grade segment. In the GM/Ford crisis, we observe a higher level of trading activity for the speculative grade segment compared with the normal period, perhaps due to the trade volume caused by a shuffling of bonds, due to clientele preferences in anticipation of, and as a consequence of, the downgrades. In the subprime crisis, we observe a lower market-wide volume for both segments. Furthermore, we find a significant reduction in the number of traded bonds and trades for the speculative grade segment, whereas we observe approximately the same number of bonds and more trades in the case of investment grade bonds. Thus, we find a flight-to-quality indicated by trading in better rated bonds compared to the normal period.

Table 7 (Panel B) presents the descriptive statistics of the liquidity proxies for the two subsegments. In general, we find that the liquidity proxies clearly indicate lower liquidity for speculative grade bonds, e.g., the price dispersion measure is 44.1 bp vs. 38.8 bp for investment grade bonds in the normal period. In the crisis periods, the liquidity of bonds in both groups deteriorates, e.g., the price dispersion measure for speculative grade bonds is 55.8 bp and 68.2 bp in the GM/Ford and the subprime crisis, respectively. Interestingly, the difference in the liquidity proxies between the two groups is less

Table 7

This table reports the mean of the yield spread, the credit rating, and the daily market-wide trading activity in Panel A for investment grade and speculative grade bonds for the three different regimes (GM/Ford crisis, normal period, and subprime crisis). The corporate bond yield spread is measured relative to the US Treasury bond yield curve and given in percentage points. We use credit ratings from Standard & Poor's where we assign integer numbers to ratings, i.e., AAA=1, AA+=2, etc., to measure the average rating. The market-wide trading activity variables represent the number of traded bonds and trades, and the total trading volume per day. Panel B provides the averages for the trading activity variables (traded volume, number of trades, and time interval between trades) and liquidity measures (Amihud, price dispersion, Roll, and zero-return measure). The data set consists of 23,703 US corporate bonds traded over the period October 2004–December 2008.

Panel A: Yield-spread, rating, and market-wide trading activity

	Investment grade			Speculative grade		
	GM/Ford crisis	Normal period	Subprime crisis	GM/Ford crisis	Normal period	Subprime crisis
Yield spread (%)	1.19	0.97	3.21	4.41	3.48	10.82
Rating	6.50	6.01	5.88	13.10	13.44	14.12
Traded bonds (thd)	3.23	3.90	3.96	1.99	2.02	1.22
Market-wide trades (thd)	11.53	13.10	18.11	8.90	7.61	4.66
Market-wide volume (bln)	5.26	6.01	5.66	2.38	2.04	1.33

Panel B: Liquidity proxies

	Investment grade			Speculative grade		
	GM/Ford crisis	Normal period	Subprime crisis	GM/Ford crisis	Normal period	Subprime crisis
Volume (mln)	1.99	1.91	1.77	1.29	0.96	1.08
Trades	4.37	4.12	5.86	4.77	3.94	4.24
Trading interval (dy)	3.28	3.30	3.28	3.18	3.37	3.49
Amihud (bp per mln)	61.27	50.47	75.14	68.76	54.69	147.94
Price dispersion (bp)	44.74	38.78	72.72	55.79	44.09	68.22
Roll (bp)	154.70	128.83	206.68	183.47	164.95	215.64
Zero-return (%)	0.01	0.01	0.01	0.04	0.04	0.06

Table 8
This table reports the panel data regression model explaining the yield spread changes based on weekly averages of all variables:

$$\begin{aligned}\Delta(\text{Yield spread})_{i,t} &= \alpha_0+\alpha_1\cdot\Delta(\text{Yield spread})_{i,t-1}+\alpha_2\cdot\Delta(\text{Volume})_{i,t}+\alpha_3\cdot\Delta(\text{Trades})_{i,t}+\alpha_4\cdot\Delta(\text{Trading interval})_{i,t}\\ &+\alpha_5\cdot\Delta(\text{Amihud})_{i,t}+\alpha_6\cdot\Delta(\text{Price dispersion})_{i,t}+\alpha_7\cdot\Delta(\text{Roll})_{i,t}+\alpha_8\cdot\Delta(\text{Zero-return})_{i,t}+(\text{Speculative grade dummy})_t\cdot[\beta_1\cdot\Delta(\text{Yield spread})_{i,t-1}\\ &+\beta_2\cdot\Delta(\text{Volume})_{i,t}+\beta_3\cdot\Delta(\text{Trades})_{i,t}+\beta_4\cdot\Delta(\text{Trading interval})_{i,t}+\beta_5\cdot\Delta(\text{Amihud})_{i,t}+\beta_6\cdot\Delta(\text{Price dispersion})_{i,t}\\ &+\beta_7\cdot\Delta(\text{Roll})_{i,t}+\beta_8\cdot\Delta(\text{Zero-return})_{i,t}]+\sum_{k=1}^{21}\delta_k\cdot\Delta(\text{Rating dummy})_{i,t,k}+\epsilon_{i,t}.\end{aligned}$$

The yield spread change is explained by the change in the lagged yield spread, trading activity variables (traded volume, number of trades, and time interval between trades), liquidity measures (Amihud, price dispersion, Roll, and zero-return measure), and rating dummies to control for credit risk. Additionally, we add interaction terms between the subsegment of speculative grade bonds and the liquidity proxies. The corporate bond yield spread is measured relative to the US Treasury bond yield curve. The *t*-statistics are given in parentheses and are calculated from Newey and West (1987) standard errors, which are corrected for heteroskedasticity and serial correlation. We provide an *F*-test to test whether the interaction terms of the dummy variable with the liquidity proxies are jointly zero. The standard errors of the *F*-statistics are also Newey and West (1987) corrected. In addition, the table also reports the model's R^2 and the number of observations. The data set consists of 23,703 US corporate bonds traded over the period October 2004–December 2008.

Intercept	0.0757***	(73.5243)
Δ(Yield spread)$_{i,t-1}$	−0.3034***	(−35.2591)
Δ(Volume)$_{i,t}$	−0.0038***	(−3.6157)
Δ(Trades)$_{i,t}$	0.0057***	(8.9298)
Δ(Trading interval)$_{i,t}$	0.0070***	(15.1603)
Δ(Amihud)$_{i,t}$	0.0458***	(25.4481)
Δ(Price dispersion)$_{i,t}$	0.0803***	(20.8335)
Δ(Roll)$_{i,t}$	0.0602***	(13.5645)
Δ(Zero-return)$_{i,t}$	−0.0833***	(−3.5542)
(Speculative grade dummy)$_t$ × Δ(Yield spread)$_{i,t-1}$	0.0377***	(3.2552)
(Speculative grade dummy)$_t$ × Δ(Volume)$_{i,t}$	−0.0122***	(−4.9216)
(Speculative grade dummy)$_t$ × Δ(Trades)$_{i,t}$	0.0015	(1.3913)
(Speculative grade dummy)$_t$ × Δ(Trading interval)$_{i,t}$	−0.0097***	(−9.5900)
(Speculative grade dummy)$_t$ × Δ(Amihud)$_{i,t}$	0.0246***	(7.6564)
(Speculative grade dummy)$_t$ × Δ(Price dispersion)$_{i,t}$	0.0315***	(4.5624)
(Speculative grade dummy)$_t$ × Δ(Roll)$_{i,t}$	−0.0010	(−0.1301)
(Speculative grade dummy)$_t$ × Δ(Zero-return)$_{i,t}$	−0.0085	(−0.3184)
Δ(Rating dummies)	Yes	
F-stat. $\mathcal{H}_0$: (Speculative grade dummy) × Δ(Liquidity proxies) = 0	28.8800	
Observations	637,814	
R^2	0.0954	

pronounced in the subprime crises for the price dispersion and Roll measure, i.e., the average trading cost increases relatively more for the investment grade segment. However, for the Amihud measure we find a large difference between investment and speculative grade bonds in the subprime crisis (i.e., 75.1 bp vs. 147.9 bp) indicating that large trades in speculative grade bonds have a high price impact.

Table 8 presents the results for the panel data regressions using a dummy variable for speculative grade bonds and, more important, including interaction terms between this dummy and the liquidity proxies. Overall, we find that speculative grade bonds react more strongly to changes in liquidity. The Amihud measure, the price dispersion measure, and the trading activity parameters are significantly higher (in absolute terms) for speculative grade bonds. Thus, we find a significant interaction between credit and liquidity risk. On average, bonds with higher credit risk are less liquid and react more strongly to liquidity changes. The most important ones are the Amihud and the price dispersion measure, for which both coefficients basically increase by 50%. An *F*-test reveals that we can reject at a 1% level the hypothesis that the interaction terms between credit and liquidity risk are jointly zero.

As for the improvement in R^2, we find that the inclusion of the interaction terms leads to an increase from 8.56% to 9.54% compared to the analysis for the whole time-series, highlighting the importance of adding these terms. Considering the economic significance, a one standard deviation move in all proxies in the direction of greater illiquidity would increase the spread by 13.8 bp for investment grade bonds, compared to 37.6 bp for speculative grade bonds, respectively. Thus, we find a higher impact of the liquidity proxies for bonds with high credit risk. The ranking of the economic importance of the individual liquidity proxies for investment grade and speculative grade bonds stays approximately the same, with the Amihud measure showing the highest impact (4.6 bp for investment grade bonds and 10.4 bp for speculative grade bonds).[18]

[18] In other tests, not reported here due to space constraints, we also estimated a panel data regression separately for each subsegment using

Overall, we find that the liquidity effects are far more pronounced for speculative grade bonds and, thus, indicating an interaction between credit and liquidity risk. This effect is particularly important for the subprime crisis, when we observe a clear flight-to-quality effect.

8. Conclusion

Financial economists have been concerned with the impact of liquidity and liquidity risk on the pricing of assets for at least two decades. During this period, several issues relating to liquidity effects in asset prices have been analyzed at a theoretical and empirical level by academic researchers, particularly in the context of US equity markets. More recently, the focus on liquidity has been broadened to include a wider class of assets such as derivatives and fixed income securities. This trend has accelerated since the onset of the subprime crisis, as the discussion of liquidity has attracted much interest among academics, practitioners, and regulators. While the crisis has manifested itself in almost every financial market in the world, the most stressed markets, by far, have been those for fixed income securities and their derivatives, particularly those with credit risk, including corporate bonds, CDSs, and CDOs. These developments require that the scope of the discussion of liquidity be extended to include the interplay between liquidity and credit.

Corporate bond markets are far less liquid than related equity markets, since only a very small proportion of the universe of corporate bonds trades even as often as once a day. In addition, corporate bonds trade in an over-the-counter market, where there is no central market place. Hence, conventional transaction metrics of liquidity such as bid-offer quotes do not have the same meaning in this market compared to exchange traded markets. The issue of liquidity in this relatively illiquid, OTC market is fundamentally different from that in exchange traded markets: Thus, it is necessary to use measures of liquidity that go beyond the standard transaction-based measures common in research in more liquid, exchange traded markets.

We employ a wide range of liquidity measures to quantify the liquidity effects in corporate bond yield spreads. Our analysis explores the time-series and cross-sectional aspects of liquidity using panel and Fama-MacBeth regressions, respectively. We find that the liquidity proxies in the specified regression models account for about 14% of the explained time-series variation of the yield spread changes. Furthermore, we find that the effect of the liquidity measures is far stronger in both the GM/Ford crisis and the subprime crisis, most remarkably the economic effect more than doubles in the subprime crisis. All the liquidity proxies considered exhibit statistically as well as economically significant results. In particular, measures estimating trading costs based on transaction data show the strongest effects.

Comparing investment grade to speculative grade bonds, we find lower liquidity for speculative grade bonds as well as a stronger reaction to changes in liquidity. These results show that bonds with higher credit risk also are more exposed to liquidity risk.

These results are useful for many practical applications, particularly pricing and risk management, and also have implications for regulatory policy. They also highlight the importance of transparency of trades for OTC markets, with reporting to a central authority being a crucial element for price discovery.

(*footnote continued*)
dummy variables with interaction terms for the subperiods. This allows us to compare liquidity effects in different time periods between the investment grade and speculative grade bonds. For the subprime crisis, we find a significant increase in the liquidity proxies for both subsegments, where the increase is particularly strong for speculative grade bonds. This result reinforces the findings of the previous analysis.

References

Acharya, V.V., Amihud, Y., Bharath, S., 2009. Liquidity risk of corporate bond returns. Unpublished working paper. New York University and Arizona State University.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid–ask spread. Journal of Financial Economics 17, 223–249.

Amihud, Y., Mendelson, H., Pedersen, L.H., 2006. Liquidity and asset prices. Foundations and Trends in Finance 1, 269–364.

Bao, J., Pan, J., Wang, J., 2011. Liquidity and corporate bonds. Journal of Finance 66, 911–946.

Chakravarty, S., Sarkar, A., 1999. Liquidity in US fixed income markets: a comparison of the bid–ask spread in corporate, government and municipal bond markets. Federal Reserve Bank of New York Staff Report No. 73.

Chen, L., Lesmond, D.A., Wei, J., 2007. Corporate yield spreads and bond liquidity. Journal of Finance 62, 119–149.

Collin-Dufresne, P., Goldstein, R.S., Martin, J.S., 2001. The determinants of credit spread changes. Journal of Finance 56, 2177–2207.

De Jong, F., Driessen, J., 2006. Liquidity risk premia in corporate bond markets. Unpublished working paper. University of Tilburg.

Dick-Nielsen, J., 2009. Liquidity biases in TRACE. Journal of Fixed Income 19, 43–55.

Dick-Nielsen, J., Feldhütter, P., Lando, D., 2012. Corporate bond liquidity before and after the onset of the subprime crisis. Journal of Financial Economics 103, 471–492.

Duffie, D., Garleanu, N., Pedersen, L.H., 2007. Valuation in over the counter markets. Review of Financial Studies 20, 1865–1900.

Edwards, A., Harris, L., Piwowar, M., 2007. Corporate bond market transaction costs and transparency. Journal of Finance 62, 1421–1451.

Elton, E.J., Gruber, M., Agrawal, D., Mann, C., 2001. Explaining the rate spread on corporate bonds. Journal of Finance 56, 247–277.

Eom, Y., Helwege, H.J., Huang, J.Z., 2004. Structural models of corporate bond pricing: an empirical investigation. Review of Financial Studies 17, 499–544.

Goldstein, M.A., Hotchkiss, E.S., 2007. Dealer behavior and the trading of newly issued corporate bonds. Unpublished working paper. Babson College and Boston College.

Hong, G., Warga, A., 2000. An empirical study of bond market transactions. Financial Analysts Journal 56, 32–46.

Hotchkiss, E.S., Ronen, T., 2002. The informational efficiency of the corporate bond market: an intraday analysis. Review of Financial Studies 15, 1325–1354.

Houweling, P., Mentink, A., Vorst, T., 2005. Comparing possible proxies of corporate bond liquidity. Journal of Banking and Finance 29, 1331–1358.

Huang, J., Huang, M., 2003. How much of the corporate-treasury yield spread is due to credit risk? Unpublished working paper. Pennsylvania State University and Cornell University.

Jankowitsch, R., Nashikkar, A., Subrahmanyam, M., 2011. Price dispersion in OTC markets: a new measure of liquidity. Journal of Banking and Finance 35, 343–357.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Lesmond, D.A., Ogden, J., Trzcinka, C., 1999. A new estimate of transaction costs. Review of Financial Studies 12, 1113–1141.

Lin, H., Wang, J., Wu, C., 2011. Liquidity risk and expected corporate bond returns. Journal of Financial Economics 99, 628–650.

Liu, J., Longstaff, F.A., Mandell, R.E., 2004. The market price of credit risk: an empirical analysis of interest rate swap spreads. Journal of Business 79, 2337–2359.
Longstaff, F., Mithal, S., Neis, E., 2005. Corporate yield spreads: default risk or liquidity? New evidence from the credit-default swap market. Journal of Finance 60, 2213–2253.
Mahanti, S., Nashikkar, A., Subrahmanyam, M., Chacko, G., Mallik, G., 2008. Latent liquidity: a new measure of liquidity, with an application to corporate bonds. Journal of Financial Economics 88, 272–298.
Markit Group Limited, 2006. Markit.com User Guide—Version 8.0 March 2006. User guide, Markit Group Limited.
Nashikkar, A., Subrahmanyam, M., Mahanti, S., 2011. Liquidity and arbitrage in the market for credit risk. Journal of Financial and Quantitative Analysis 46, 627–656.
Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Perraudin, W., Taylor, A., 2003. Liquidity and bond market spreads. Unpublished working paper. Bank of England.
Roll, R., 1984. A simple implicit measure of the effective bid–ask spread in an efficient market. Journal of Finance 39, 1127–1139.
Ronen, T., Zhou, X., 2009. Where did all the information go? Trade in the corporate bond market. Unpublished working paper. Rutgers University.
Sadka, R., 2010. Liquidity risk and the cross-section of hedge-fund returns. Journal of Financial Economics 98, 54–71.
Schultz, P., 2001. Corporate bond trading costs and practices: a peek behind the curtain. Journal of Finance 56, 677–698.

Journal of Financial Economics 131 (2019) 619–642

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Common risk factors in the cross-section of corporate bond returns☆

Jennie Bai, Turan G. Bali, Quan Wen

McDonough School of Business, Georgetown University, 3700 O St., Washington, DC, 20057, USA

ARTICLE INFO

Article history:
Received 25 October 2017
Revised 5 February 2018
Accepted 7 February 2018
Available online 16 August 2018

JEL Classification:
G11
G12
O13

Keywords:
Corporate bond
Risk factors
Downside risk
Credit risk
Liquidity risk

ABSTRACT

We investigate the cross-sectional determinants of corporate bond returns and find that downside risk is the strongest predictor of future bond returns. We also introduce common risk factors based on the prevalent risk characteristics of corporate bonds—downside risk, credit risk, and liquidity risk—and find that these novel bond factors have economically and statistically significant risk premiums that cannot be explained by long-established stock and bond market factors. We show that the newly proposed risk factors outperform all other models considered in the literature in explaining the returns of the industry- and size/maturity-sorted portfolios of corporate bonds.

1. Introduction

Over the past three decades, financial economists have identified a large number of risk factors that explain the cross-sectional variation in stock returns. In contrast, far less studies are devoted to the cross-section of corporate

☆ We are grateful to the editor, Bill Schwert, and an anonymous referee for their extremely helpful comments and suggestions. We thank Alex Butler, Pierre Collin-Dufresne, Robert Engle, Zhiguo He, Stefan Nagel, Raghavendra Rau, Sheridan Titman, Pietro Veronesi, Jianfeng Yu, and Hao Zhou for their insightful and constructive comments. We also benefited from discussions with Laurent Barras, Sandeep Dahiya, Andrea Gamba, Anurag Gupta, Nishad Kapadia, George Kapetanios, Gi Kim, Bart Lambrecht, Tao Li, Michael Neumann, Lee Pinkowitz, George Skiadopoulos, Rohan Williamson, and seminar participants at New York Univerity, the University of Texas at Austin, Rice University, Ecole Polytechnique Federale de Lausanne, University of Cambridge, University of Warwick, Vanderbilt University, Case Western Reserve University, the City University of Hong Kong, Georgetown University, Koc University, PBC Tsinghua University, Peking University, Queen Mary University of London, Southern Methodist University, Tulane University, the Federal Reserve Board, the Federal Reserve Bank of Richmond, the HKUST finance symposium, the Institute for Financial Research (SIFR) conference on in Stockholm, the Second Annual Moody's Credit Risk Conference, and the Seventh NYU Annual Volatility Institute Conference for their extremely helpful comments and suggestions. We also thank Kenneth French, Lubos Pastor, and Robert Stambaugh for making a large amount of historical data publicly available in their online data library.
We merged the findings of our earlier working paper "Do the distributional characteristics of corporate bonds predict their future returns?" with this paper so that Bai et al. (2016) cited in the paper will remain as a permanent working paper.

E-mail addresses: jennie.bai@georgetown.edu (J. Bai), turan.bali@georgetown.edu (T.G. Bali), quan.wen@georgetown.edu (Q. Wen).

https://doi.org/10.1016/j.jfineco.2018.08.002
0304-405X/

bond returns.[1] Compared to the size of the US equity market ($19 trillion), the corporate bond market is relatively small, with a total amount outstanding of $12 trillion.[2] However, the issuance of corporate bonds is at a much larger scale than the issuance of stocks for US corporations: an annual average of $1.3 trillion for corporate bonds compared to $265 billion for stocks since 2010. Moreover, corporate bonds play an increasingly important role in institutional investors' portfolios, evidenced by the recent influx to bond funds.[3] Both corporate bonds and stocks are important financing channels for corporations, and both are important assets under management for fund managers. Thus, it is pivotal to enhance our understanding of the common risk factors that determine the cross-sectional differences in corporate bond returns.

Earlier studies on corporate bonds generally rely on long-established stock and bond market factors to predict contemporaneous or future bond returns, including the stock market factors of Fama and French (1993), Carhart (1997), and Pastor and Stambaugh (2003): excess stock market return, the size factor (SMB), the book-to-market factor (HML), the momentum factor (MOM), and the liquidity factor (LIQ), along with the bond market factors of Fama and French (1993), Elton et al. (1995), and Bessembinder et al. (2009): excess bond market return, the default spread (DEF), and the term spread (TERM). However, these commonly used factors are either constructed from stock-level data or aggregate macroeconomic variables; hence, their cross-sectional predictive power is limited for bond-level returns. When we test these existing models in terms of their ability to explain the industry-sorted and size/maturity-sorted portfolios of corporate bonds, their empirical performance turns out to be poor. In this paper, we show that it is crucial to rely on the prominent features of corporate bonds when constructing bond-implied risk factors to explain the cross-sectional differences in corporate bond returns.

Although corporate bonds and stocks both reflect firm fundamentals, they differ in several key features. First and foremost, bondholders, compared to stockholders, are more sensitive to downside risk.[4] Second, it is well known that firms issuing corporate bonds suffer from potential default risk given legal requirements on the payment of coupons and principal, whereas firms issuing stocks have relatively lower exposure to bankruptcy. This feature makes credit risk particularly important in determining corporate bond returns. Third, the corporate bond market, due to its over-the-counter trading mechanism and other market features, bears higher liquidity risk. Bond market participants are dominated by institutional investors such as insurance companies, pension funds, and mutual funds.[5] Many bondholders are long-term investors who often follow a buy-and-hold strategy. Therefore, liquidity in the corporate bond market is lower compared to the stock market in which active trading is partially attributable to the existence of individual investors.

Given these significant differences in market features and the types of investors in the equity and bond markets, we endeavor to identify bond-implied risk factors that provide an accurate characterization of the cross-sectional variation in bond returns. Following Bessembinder et al. (2006) who highlight the importance of using Trade Reporting and Compliance Engine (TRACE) transaction data, we calculate bond returns at the monthly frequency using the intraday transaction records from the Enhanced TRACE data for the period July 2002 to December 2016. Our proxy for downside risk is the 5% value at risk (VaR) estimated from the lower tail of the empirical return distribution; that is, the second lowest monthly return observation over the past 36 months. Our proxy for credit quality is bond-level credit rating. Our proxy for illiquidity is the bond-level measure of Bao et al. (2011). In addition to these three economically sensible risk characteristics for corporate bonds, we take into account bond exposure to the market risk factor (market beta).

First, we test the significance of a cross-sectional relation between downside risk and future returns on corporate bonds using portfolio-level analysis. We find that bonds in the highest downside risk quintile generate 11.88% per annum higher return than bonds in the lowest downside risk quintile. After controlling for ten well-known stock and bond market factors, the risk-adjusted return difference between the lowest and highest downside risk quintiles (downside risk premium) is economically large and statistically significant: 8.64% per annum with a *t*-statistic of 2.82, suggesting that loss-averse bond investors prefer high expected return and low downside risk. We also examine the average portfolio characteristics of VaR quintiles and find that bonds with high VaR have higher market risk, higher credit risk, lower liquidity, longer maturity, and smaller size. Thus, we test whether the positive relation between downside risk and future returns holds after controlling for bond characteristics. Bivariate portfolio-level analyses indicate that downside risk remains a significant predictor of future bond returns after controlling for credit rating, illiquidity, maturity, and size.

Having established the evidence that downside risk is a strong predictor of future bond returns, we investigate the source of downside risk premium. Specifically, we dissect downside risk into volatility, skewness, and kurtosis components and find that bond return volatility (skewness) is a significantly positive (negative) predictor

[1] This is partly because of the dearth of high-quality corporate bond data and the complex features of corporate bonds such as optionality, seniority, changing maturity, and risk exposure to a number of financial and macroeconomic factors.

[2] Source: Table L.213 and L.223 in the Federal Reserve Board Z1 Flow of Funds, Balance Sheets, and Integrated Macroeconomic Accounts, as of the fourth quarter of 2016.

[3] See Feroli et al. (2014) and the Investment Company Institute Annual Report (2014).

[4] Bondholders gain the cash flow of fixed coupon and principal payment, thus hardly benefit from the euphoric news in firm fundamentals. Since the upside payoffs are capped, bond payoffs become concave in the investor beliefs about the underlying fundamentals, whereas equity payoffs are linear in investor beliefs regarding fluctuations in the underlying factors (e.g., Hong and Sraer, 2013).

[5] Source: Financial Accounts of the United States, Release Z1, Table L.213.

of future bond returns after controlling for skewness (volatility) and kurtosis. Moreover, volatility and skewness contribute strongly to the significance of downside risk in the corporate bond market, whereas kurtosis makes a weak incremental contribution to the downside risk premium after volatility and skewness are controlled for.

Then, we investigate the cross-sectional relation between downside risk and expected returns at the bond level using Fama-MacBeth (1973) regressions in which we control for multiple factors simultaneously. Specifically, we present the time-series averages of the slope coefficients from the regressions of one-month-ahead excess returns on downside risk controlling for past bond risk/return characteristics, including credit rating, illiquidity, market beta, maturity, size, lagged return, and bond exposures to the default and term factors. The results indicate that downside risk remains a strong predictor of future bond returns after controlling for a large number of bond characteristics. Among the control variables, only the short-term reversal effect is found to be strong and robust across different regression specifications. Thus, in addition to the three risk factors (downside, credit, and liquidity risk), we construct a bond return reversal factor and examine its empirical performance in predicting the cross-sectional variation in corporate bonds.

Finally, we introduce novel risk factors based on the above prevalent risk characteristics. In a similar spirit to Fama and French (2015) and Hou et al. (2015), we rely on the independently sorted portfolios using credit rating as the main sorting variable and downside risk, illiquidity, and past one-month return as the other sorting variables when constructing the new bond factors; namely the downside risk factor (DRF), liquidity risk factor (LRF), and return reversal factor (REV). These independent sorts also produce three credit risk factors so that the final credit risk factor (CRF) is defined as the average of the three factors of credit risk. We run time-series regressions to assess the predictive power of these new risk factors. The intercepts (alphas) from the regressions represent the abnormal returns not explained by standard stock and bond market factors. When using the most general ten-factor model that combines all of the commonly used stock and bond market factors, we find that the alphas for the DRF, CRF, LRF, and REV factors are all economically and statistically significant, indicating that the existing factors are not sufficient to capture the information content in these newly proposed bond factors.

Motivated by the findings in Daniel and Titman (1997) and Brennan et al. (1998), we further examine if the exposures to the new bond factors predict future bond returns. For each bond and each month in our sample, we estimate the factor betas from the monthly rolling regressions of excess bond returns on the DRF, CRF, LRF, and REV factors over a 36-month fixed window while controlling for the bond market factor (MKT^{Bond}). After we obtain the factor exposures, namely, the downside risk beta (β^{DRF}), the credit risk beta (β^{CRF}), the liquidity risk beta (β^{LRF}), and the return reversal beta (β^{REV}), we investigate the significance of the bond factor betas in predicting the cross-sectional differences in corporate bond returns using bond-level cross-sectional regressions. Our results show that all three factor betas (β^{DRF}, β^{CRF}, β^{LRF}) are positively related to future bond returns, lending further support to the finding that the newly proposed factors capture systematic variations in bond returns and common risk premiums in the corporate bond market. However, the bond exposure to the return reversal factor (β^{REV}) turns out to be statistically insignificant with and without controlling for bond characteristics. Thus, we conclude that one-month lagged return (REV) is a strong cross-sectional determinant of future bond returns, but it can be viewed as a nonrisk bond characteristic instead of a common risk factor in the bond market.

One important critique in asset pricing tests, as pointed out by Lewellen et al. (2010), is that characteristic-sorted portfolios (used as test assets) do not have sufficient independent variation in the loadings of factors constructed with the same characteristics. To improve the power of asset pricing tests, Lewellen et al. (2010) suggest that the empirical performance of risk factors should be tested based on alternative test portfolios. Following their insight, we form two sets of test portfolios that do not necessarily relate to the aforementioned risk characteristics: (i) 5×5 independently sorted bivariate portfolios of size and maturity and (ii) 30 industry-sorted portfolios. Then, we examine the relative performance of factor models in explaining the time-series and cross-sectional variations in these test portfolios. We find that the newly proposed four-factor model with the market, downside, credit, and liquidity risk factors substantially outperforms all other models considered in the literature in predicting the returns of the industry- and size/maturity-sorted portfolios of corporate bonds.[6]

Specifically, our model produces an average 56% adjusted R^2 for the 25 size/maturity-sorted portfolios of corporate bonds, whereas the existing models can explain up to 18%. Our model also remains its high explanatory power for the 30 industry-sorted portfolios of corporate bonds, with an average adjusted R^2 of 37%, in contrast to the weak performance of existing models with average adjusted R^2 values of only 13% to 18%. Consistent with these findings, the new model has markedly smaller and insignificant alphas in explaining the cross-section of bond returns, generating economically and statistically insignificant alphas for all 25 size/maturity-sorted portfolios of corporate bonds, with an average alpha of 0.04% per month. In contrast, the existing models generate significant alphas for all 25 portfolios, with an average alpha of 0.33% to 0.42% per month. Similarly, the new model generates insignificant alphas for all of the 30-industry portfolios, with an average alpha of 0.14%, whereas the existing models produce significant alphas with a monthly average of 0.41% to 0.55%. These results indicate that the new factors of corporate bonds significantly outperform

[6] Note that the test portfolios constructed based on size, maturity, and industry characteristics do not have a direct link to downside risk, credit risk, or illiquidity. At an earlier stage of the study, we form test portfolios based on downside risk, credit risk, and illiquidity, and, as anticipated, the empirical performance of the newly proposed four-factor model is even higher in predicting the time-series and cross-sectional variations in the returns of the downside risk/credit risk/illiquidity-sorted portfolios.

all existing factor models, and hence the new model serves as a proper and higher benchmark in evaluating the risk-return tradeoff in the corporate bond market.

This paper proceeds as follows. Section 2 sets forth a literature review. Section 3 describes the data and main variables. Section 4 examines the cross-sectional relation between downside risk and expected returns of corporate bonds. Section 5 introduces new risk factors for corporate bonds and compares their relative performance with long-established stock and bond market factors. Section 6 conducts a battery of robustness checks and Section 7 concludes the paper.

2. Literature review

Our empirical findings contribute to the literature in several important ways. The foremost contribution is to identify bond-implied new risk factors that significantly predict the cross-sectional variation in future bond returns. The earlier literature on corporate bond returns focuses on aggregate indices (see, e.g., Fama and French, 1993; Elton et al., 1995) and bond portfolios (e.g., Blume et al., 1991).[7] Subsequent studies have investigated the bond returns at the firm level, mainly with quoted price data (see, e.g., Kwan, 1996; Gebhardt et al., 2005)[8] and recently with transaction data (see, e.g., Bessembinder et al., 2009; Lin et al., 2011; Acharya et al., 2013; Jostova et al., 2013; Chordia et al., 2017; Choi and Kim, 2018).[9] Our paper also uses transaction data but differs from the literature by deriving bond-implied risk factors. Our downside, credit, and liquidity risk factors together have superior predictive power over the long-established risk factors, outperforming the existing models in explaining the cross-sectional differences in individual bond returns as well as the industry-sorted and size/maturity-sorted portfolios of corporate bonds.

The idea of linking credit and liquidity to bond pricing is by no means new. Our paper, however, advances the literature by showing that credit risk and liquidity risk have significant pricing power for the cross-section of future corporate bond returns. The literature on the credit spread puzzle well documents the evidence that credit and illiquidity can explain contemporaneous bond yield spreads (see, e.g., Longstaff et al., 2005; Chen et al., 2007). In a recent paper, Culp et al. (2018) show that a risk premium for idiosyncratic tail risk is the primary determinant of corporate spreads, whereas bond market illiquidity, investors' overestimation of default risks, and corporate frictions do not explain credit spreads. The main theme, focus, and methodological approaches of all these papers, however, are very different from ours, as we do not use any parametric/structural model or option data to back out our risk measures. More importantly, our paper differs from earlier studies by analyzing the cross-section of future corporate bond returns (not yield spreads) and introducing a novel risk factor model that measures abnormal returns on corporate bond portfolios.

The second contribution of this paper is to demonstrate the empirical performance of downside risk in predicting the cross-sectional differences in future returns of corporate bonds. There is a large body of literature on safety-first investors who minimize the chance of disaster (or the probability of failure). The portfolio choice of a safety-first investor is to maximize expected return subject to a downside risk constraint. The safety-first investor in Roy (1952), Baumol (1963), Levy and Sarnat (1972), and Arzac and Bawa (1977) uses a downside risk measure that is a function of value at risk. Roy (1952) indicates that most investors are principally concerned with avoiding a possible disaster and that the principle of safety plays a crucial role in the decision-making process. Thus, the idea of a disaster exists and a risk averse, safety-first investor will seek to reduce the chance of such a catastrophe occurring insofar as possible.

Our work is also related to Lettau et al. (2014) who show that downside risk capital asset pricing model (DR-CAPM) can price the cross-section of currency returns and several other assets' returns, but they find no evidence that downside beta is positively related to corporate bond returns (see pp. 222–223). Our work is different from Lettau et al. (2014) by focusing on the extreme total downside risk as measured by value at risk, instead of systematic downside risk as measured by downside beta along the lines of Bawa and Lindenberg (1977) and Ang et al. (2006).

The use of VaR techniques in risk management has exploded over the past two decades. Financial institutions now routinely use VaR and expected shortfall in managing their risk, and nonfinancial firms adopt this technology for their risk management as well. There is an extensive literature on risk management and VaR per se; however, only a few studies investigate the time-series or cross-sectional relation between VaR and expected returns on individual stocks or equity portfolios (e.g., Bali et al., 2009; Huang et al., 2012). The predictive power of VaR or expected shortfall has not been investigated for alternative asset classes. This paper provides the first evidence on the theoretically consistent positive and significant relation between left-tail risk and future corporate bond returns.

[7] Fama and French (1993) use five corporate bond indices from the module of Ibbotson for rating groups AAA, AA, A, Baa, and LG (low-grade, that is, below Baa). Elton et al. (1995) study 20 bond indices across Treasury bonds, corporate bonds, mortgage securities from Ibbotson, Merrill Lynch, and Lehman Brothers. Blume, Keim, and Pate (1991) study the Salomon (Lehman) Brothers index of corporate bonds, Ibbotson long-term government bond index as well as bonds below BBB listed in the S&P Bond Guide. Note that quite a few papers, though they study bonds, are indeed limited to Treasury bonds or a combination of Treasury and corporate bonds.

[8] Gebhardt et al. (2005) test the cross-sectional predictive power of default and term spread beta and find that they are significantly related to corporate bond returns.

[9] Bessembinder et al. (2009) find that using the daily bond returns generated from the TRACE data increases the power of the test statistics designed to detect abnormal bond returns in corporate event studies. Lin et al. (2011) construct the market liquidity risk factor and show that it is priced in the cross-section of corporate bond returns. Acharya et al. (2013) show that corporate bonds are exposed to liquidity shocks in equity and Treasury markets. Jostova et al. (2013) investigate whether the momentum anomaly exists in the corporate bond market. There are also two recent papers, Chordia et al. (2017) and Choi and Kim (2018), that examine whether equity market predictors are priced in the cross-section of corporate bond returns.

3. Data and variable definitions

3.1. Corporate bond data

For corporate bond data, we rely on the transaction records reported in the enhanced version of the TRACE for the sample period July 2002 to December 2016. Ideally, we would prefer to investigate the cross-section of corporate bond returns using a longer sample period. However, one critical risk factor of corporate bond returns, illiquidity, requires daily bond transaction prices that are not provided in such datasets as the Lehman Brothers fixed income database, Datastream, or Bloomberg.[10] Therefore, we focus on the TRACE dataset that offers the best quality of corporate bond transactions with intraday observations on price, trading volume, and buy and sell indicators. We then merge corporate bond pricing data with the Mergent fixed income securities database to obtain bond characteristics such as offering amount, offering date, maturity date, coupon rate, coupon type, interest payment frequency, bond type, bond rating, bond option features, and issuer information.

In the online Internet Appendix, we also expand the TRACE data by including alternative bond datasets, mainly those containing quoted prices, for a longer sample period starting from January 1977. For this longer sample, we construct downside risk factor and credit risk factor (but not the liquidity risk factor) and replicate our main analysis in the online Internet Appendix.

For TRACE intraday data, we adopt the following filtering criteria:

1. Remove bonds that are not listed or traded in the US public market, which include bonds issued through private placement, bonds issued under the 144A rule, bonds that do not trade in US dollars, and bond issuers not in the jurisdiction of the United States.
2. Remove bonds that are structured notes, mortgage backed or asset backed, agency backed, or equity linked.
3. Remove convertible bonds since this option feature distorts the return calculation and makes it impossible to compare the returns of convertible and nonconvertible bonds.[11]
4. Remove bonds that trade under \$5 or above \$1000.
5. Remove bonds that have a floating coupon rate, which means the sample comprises only bonds with a fixed or zero coupon. This rule is applied based on the consideration of the accuracy in bond return calculation, given the challenge in tracking a floating-coupon bond's cash flows.
6. Remove bonds that have less than one year to maturity. This rule is applied to all major corporate bond indices such as the Barclays Capital Corporate Bond Index, the Bank of America Merrill Lynch Corporate Master Index, and the Citi Fixed Income Indices. If a bond has less than one year to maturity, it will be delisted from major bond indices; hence, index-tracking investors will change their holding positions. This operation will distort the return calculation for bonds with less than one year to maturity; thus, we remove them from our sample.
7. For intraday data, we also eliminate bond transactions that are labeled as when-issued, locked-in, or have special sales conditions, and that have more than a two-day settlement.
8. Remove transaction records that are canceled and adjust records that are subsequently corrected or reversed.
9. Remove transaction records that have trading volume less than \$10,000.[12]

3.2. Corporate bond return

The monthly corporate bond return at time t is computed as

$$r_{i,t} = \frac{P_{i,t} + AI_{i,t} + C_{i,t}}{P_{i,t-1} + AI_{i,t-1}} - 1, \tag{1}$$

where $P_{i,\,t}$ is transaction price, $AI_{i,\,t}$ is accrued interest, and $C_{i,\,t}$ is the coupon payment, if any, of bond i in month t. We denote $R_{i,\,t}$ as bond i's excess return, $R_{i,t} = r_{i,t} - r_{f,t}$, where $r_{f,\,t}$ is the risk-free rate proxied by the one-month Treasury bill rate.

Using TRACE intraday data, we first calculate the daily clean price as the trading volume-weighted average of intraday prices to minimize the effect of bid-ask spreads in prices, following Bessembinder et al. (2009). We then convert the bond prices from daily to monthly frequency. Specifically, our method identifies two scenarios for a return to be realized at the end of month t: (i) from the end of month $t-1$ to the end of month t and (ii) from the beginning of month t to the end of month t. We calculate

[10] The National Association of Insurance Commissioners (NAIC) database also includes daily prices, but given the fact that it covers only a part of the market and it contains more illiquid observations and transactions only by the buy-and-hold insurance companies, combining this data with TRACE does not make a compatible sample. For consistency, we focus on the TRACE data.

[11] Bonds also contain other option features such as being putable, redeemable/callable, exchangeable, and fungible. Except callable bonds, bonds with other option features are a relatively small portion in the sample. However, callable bonds constitute approximately 67% of the whole sample. Hence, we keep the callable bonds in our final sample. As a robustness check, we also replicate our main analyses by using a smaller sample excluding bonds with any option feature. The main findings remain robust.

[12] Bessembinder et al. (2009) test the power of test statistics to detect abnormal bond returns and suggest that eliminating noninstitutional trades (daily volume smaller than \$100,000) from the TRACE data helps increase the power of the tests to detect abnormal performance, relative to using all trades or the last price of the day. Here we include more bonds with relatively smaller trading volume, which only makes our tests more stringent, that is, it becomes harder to detect abnormal bond alphas. In unreported results, we use two alternative samples; one is smaller by keeping bonds with trading volume larger than \$100,000, following Bessembinder et al. (2009), and the other is larger by keeping all bonds regardless of trading volume (we do apply the rule of using trading-volume-weighted price as the daily price, which vastly mitigates the impact of trades with smaller trading volume, mainly from individual investors). In both of these alternative samples, our main findings remain intact. As expected, the smaller sample gives us greater power to detect significant alphas. To make our results more generally applicable to a wide range of bonds, we adopt the current rule, which is to eliminate bonds with trading volume smaller than \$10,000.

Table 1
Descriptive statistics.

Panel A reports the number of bond-month observations, the cross-sectional mean, median, standard deviation and monthly return percentiles of corporate bonds, and bond characteristics including credit rating, time to maturity (Maturity, year), amount outstanding (Size, $ million), downside risk (5% VaR), illiquidity (ILLIQ), and the CAPM beta based on the corporate bond market return, β^{Bond}. Ratings are in conventional numerical scores, where 1 refers to an AAA rating and 21 refers to a C rating. Higher numerical score means higher credit risk. Numerical ratings of 10 or below (BBB- or better) are considered investment grade, and ratings of 11 or higher (BB+ or worse) are labeled high yield. Downside risk is the 5% VaR of corporate bond return, defined as the second lowest monthly return observation over the past 36 months. The original VaR measure is multiplied by −1 so that a higher VaR indicates higher downside risk. Bond illiquidity is computed as the autocovariance of the daily price changes within each month, multiplied by −1. β^{Bond} is the corporate bond exposure to the excess corporate bond market return, constructed using the value-weighted average return of all corporate bonds in our sample. The betas are estimated for each bond from the time-series regressions of bond excess returns on the excess bond market return using a 36-month rolling window estimation. Panel B reports the time-series average of the cross-sectional correlations. The sample period is from July 2002 to December 2016.

Panel A: Cross-sectional statistics over the sample period of July 2002–December 2016

					Percentiles					
	N	Mean	Median	SD	1st	5th	25th	75th	95th	99th
Bond return (%)	1,243,543	0.68	0.50	3.13	−7.46	−3.66	−0.68	1.86	5.59	10.33
Rating	1,243,543	8.32	7.65	4.05	1.56	2.25	5.52	10.35	16.30	19.09
Time to maturity (maturity, year)	1,243,543	9.49	6.60	8.26	1.11	1.51	3.59	12.81	26.69	31.63
Amount out (size, $million)	1,243,543	393.73	269.59	478.63	1.60	5.17	76.99	504.15	1353.24	2480.32
Downside risk (5% VaR)	579,333	5.84	4.08	5.78	0.70	1.17	2.46	6.96	16.75	29.42
Illiquidity (ILLIQ)	977,011	2.14	0.46	5.17	−1.17	−0.23	0.07	1.99	10.16	24.13
Bond market beta (β^{bond})	584,223	1.12	1.01	1.15	−0.24	0.15	0.58	1.67	3.72	5.38

Panel B: Average cross-sectional correlations

	Rating	Maturity	Size	VaR	ILLIQ	β^{Bond}
Rating	1	−0.138	−0.021	0.383	0.117	0.089
Maturity		1	−0.042	0.171	0.106	0.356
Size			1	−0.108	−0.160	0.076
VaR				1	0.323	0.195
ILLIQ					1	0.098
β^{Bond}						1

monthly returns for both scenarios, where the end (beginning) of month refers to the last (first) five trading days within each month. If there are multiple trading records in the five-day window, the one closest to the last trading day of the month is selected. If a monthly return can be realized in both scenarios, the realized return in scenario one (from month-end $t-1$ to month-end t) is selected.

Our final sample includes 38,957 bonds issued by 4079 unique firms, for a total of 1,243,543 bond-month return observations during the sample period July 2002 to December 2016.[13] On average, there are approximately 7147 bonds per month over the whole sample. Panel A of Table 1 reports the time-series average of the cross-sectional bond return distribution and bond characteristics. The average monthly bond return is 0.68%. The sample contains bonds with an average rating of 8.32 (i.e., BBB+), an average issue size of 393 million dollars, and an average time to maturity of 9.49 years. Among the full sample of bonds, 75% are investment-grade and the remaining 25% are high-yield bonds.

[13] Our key variable of interest, downside risk proxied by the 5% VaR is estimated using monthly returns over the past 36 months. A bond is included in VaR calculation if it has at least 24 monthly return observations in the 36-month rolling window before the test month. Thus, the final sample size that involves downside risk reduces from 1,243,543 to 579,333 bond-month return observations for the period July 2002–December 2016.

3.3. Cross-sectional bond risk characteristics

The literature that investigates the cross-section of corporate bond returns relies on commonly used stock market factors. This is a natural starting point since the rational asset pricing models suggest that risk premiums in the equity market should be consistent with the corporate bond market to the extent that the two markets are integrated. First, both bonds and stocks are contingent claims on the value of the same underlying assets; thus, stock market factors, such as the size and book-to-market equity ratio, should share common variations in stock and bond returns (e.g., Merton, 1974). Second, the expected default loss of corporate bonds changes with equity price. Default risk decreases as the equity value appreciates, and this induces a systematic risk factor that affects corporate bond returns.

However, the corporate bond market has its own unique features. First, credit risk is particularly important in determining corporate bond returns because firms that issue corporate bonds suffer from potential default risk given legal requirements on the payment of coupons and principal. Second, bondholders are more sensitive to downside risk than stockholders. Third, the corporate bond market is less liquid than the equity market, with most corporate bonds trading infrequently. Thus, both the level of liquidity and liquidity risk are serious concerns for investors in the corporate bond market. Fourth, corporate bond market participants have been dominated by institutional investors, such as insurance companies, pension funds, and mutual funds, whose attitudes toward risk differ significantly from individual investors.[14] Finally, there is some evidence that shows the discrepancy in return premiums between equity and corporate bond markets (e.g., Chordia et al., 2017; Choi and Kim, 2018), suggesting potential market segmentation.

Thus, it is important to identify common risk factors based on the broad risk characteristics of corporate bonds rather than relying on stock market factors or aggregate bond market factors (e.g., DEF, TERM). As discussed below, we introduce three new risk factors originated from the cross-section of individual bond returns.

3.3.1. Downside risk

Extraordinary events, such as stock market crashes and bond market collapses, are major concerns in risk management and financial regulation. Regulators are concerned with the protection of the financial system against catastrophic events, which can be a source of systematic risk. A central issue in risk management has been to determine capital requirement for financial and nonfinancial firms to meet catastrophic market risk. This increased focus on risk management has led to the development of various methods and tools to measure the risks companies face. A primary tool for financial risk assessment is VaR.

Hence, we measure downside risk of corporate bonds using VaR, which determines how much the value of an asset could decline over a given period of time with a given probability as a result of changes in market rates or prices. For example, if the given period of time is one day and the given probability is 1%, the VaR measure would be an estimate of the decline in the asset's value that could occur with 1% probability over the next trading day. Our proxy for downside risk, 5% VaR, is based on the lower tail of the empirical return distribution, that is, the second lowest monthly return observation over the past 36 months. We then multiply the original measure by -1 for convenience of interpretation.[15] As shown in Table 1, the average downside risk is 5.84% in the whole sample, implying that there is only a 5% probability that an average corporate bond would lose more than 5.84% over the next one month (or the maximum loss expected on a typical bond, at the 95% confidence level, is 5.84% over the next month).

VaR as a risk measure is criticized for not being subadditive. To alleviate this problem, Artzner et al. (1999) introduce an alternative measure of downside risk, "expected shortfall," defined as the conditional expectation of loss given that the loss is beyond the VaR level. In our empirical analyses, we use the 10% expected shortfall (ES) defined as the average of the four lowest monthly return observations over the past 36 months (beyond the 10% VaR threshold). In the online Internet Appendix, we reexamine the cross-sectional relation between downside risk and future bond returns using the 10% VaR and 10% ES measures and show that our main findings are not sensitive to the choice of a downside risk measure.

3.3.2. Credit quality

We measure credit quality of corporate bonds via their credit ratings that capture information on bond default probability and the loss severity. Ratings are assigned to corporate bonds on the basis of extensive economic analysis by rating agencies such as Moody's and S&P's. Bond-level ratings synthesize the information on both the issuer's financial condition, operating performance, and risk-management strategies, along with specific bond characteristics like coupon rate, seniority, and option features, hence making ratings a natural choice to measure credit risk of a corporate bond.

We collect bond-level rating information from Mergent Fixed Income Securities Database (FISD) historical ratings. All ratings are assigned a number to facilitate the analysis; for example, 1 refers to a AAA rating, 2 refers to AA+, ... , and 21 refers to CCC. Investment-grade bonds have ratings from 1 (AAA) to 10 (BBB-). Noninvestment-grade bonds have ratings above 10. A larger number indicates higher credit risk or lower credit quality. We determine a bond's rating as the average of ratings provided by Standard & Poor (S&P) and Moody's when both are available or as the rating provided by one of the two rating agencies when only one rating is available.

Although credit rating is the widely used, traditional measure of credit quality, earlier studies also use other credit risk proxies such as the distance-to-default measure developed by KMV or the credit default spread (Longstaff et al., 2005). Different from bond-level credit rating, all alternative proxies can only be constructed at the firm level, as the calculation requires firm balance sheet information. In addition, the CDS spread is available only for a limited number of firms that are usually large, liquid, and important. Our objective is to investigate the cross-section of corporate bond returns, which differs across firms and even bonds issued by the same firm may have different returns.[16] Therefore, we adopt credit rating to measure bond-level credit risk.

In the online Internet Appendix, we reexamine the cross-sectional relation between credit quality and future bond returns using the firm-level distance-to-default and implied CDS measures in Bai and Wu (2016) and show that our main findings are not sensitive to the choice of a credit quality measure.

3.3.3. Bond illiquidity

The literature shows the importance of illiquidity and liquidity risk in the corporate bond market. For example,

[14] Institutional investors in particular make extensive use of corporate bonds in constructing their portfolios. According to Flow of Fund data during the 1986–2012 period, about 82% of corporate bonds were held by institutional investors including insurance companies, mutual funds, and pension funds. The participation rate of individual investors in the corporate bond market is very low.

[15] Note that the original maximum likely loss values are negative since they are obtained from the left tail of the return distribution. After multiplying the original VaR measure by -1, a positive regression coefficient and positive return/alpha spreads in portfolios are interpreted as the higher downside risk being related to the higher cross-sectional bond returns.

[16] Bonds issued by the same firm may have similar probability of default but not necessarily have the same recovery rate, liquidity risk, market risk, or downside risk. Thus, bonds issued by the same firm often have different returns.

the empirical results in Chen et al. (2007) and Dick-Nielsen et al. (2012) establish the relation between corporate bond yield spreads and bond illiquidity. Using transactions data from 2003 to 2009, Bao et al. (2011) show that the bond-level illiquidity explains a substantial proportion of cross-sectional variations in bond yield spreads. Lin et al. (2011) construct a liquidity risk factor for the corporate bond market and show that the market liquidity beta is priced in the cross-section of corporate bond returns.[17] Given the importance of the transaction-based data, such as TRACE, for measuring bond illiquidity, we follow Bao, Pan, and Wang (2011) to construct bond-level illiquidity measure, *ILLIQ*, which aims to extract the transitory component from bond price. Specifically, let $\Delta p_{itd} = p_{itd} - p_{itd-1}$ be the log price change for bond i on day d of month t. Then, *ILLIQ* is defined as

$$ILLIQ = -Cov_t(\Delta p_{itd}, \Delta p_{itd+1}). \quad (2)$$

In the online Internet Appendix, we reexamine the cross-sectional relation between illiquidity and future bond returns using two additional proxies of liquidity risk: Roll (1984) and Amihud (2002) illiquidity measures.

3.3.4. Bond market β

We compute the bond market excess return (MKT^{Bond}) as the value-weighted average returns of all corporate bonds in our sample minus the one-month Treasury bill rate.[18] We estimate the bond market beta, β^{Bond}, for each bond from the time-series regressions of individual bond excess returns on the bond market excess returns using a 36-month rolling window. As shown in Table 1, the bond market beta has a wide range from 0.15 in the 5th percentile to 3.72 in the 95th percentile, with a mean (median) of 1.12 (1.01).

3.3.5. Summary statistics

Table 1 presents the correlation matrix for the bond-level characteristics and risk measures. As shown in Panel B, downside risk is positively associated with β^{bond}, illiquidity, and rating, with respective correlations of 0.195, 0.323, and 0.383. The bond market beta, β^{bond}, is also positively associated with rating and illiquidity, with respective correlations of 0.089 and 0.098. Bond maturity is positively correlated with all risk measures, except credit rating, implying that bonds with longer maturity (i.e., higher interest rate risk) have higher β^{bond}, higher VaR, higher ILLIQ, and lower rating. Bond size is negatively correlated with VaR and ILLIQ, indicating that bonds with smaller size have higher VaR and higher ILLIQ. The correlations between size and rating and between size and maturity are economically weak.

[17] Choi and Kronlund (2018) examine reaching for yield by corporate bond mutual funds and find that reaching for yield is stronger for retail-oriented mutual funds when corporate bond liquidity is high.

[18] We also consider alternative bond market proxies such as the Barclays Aggregate Bond Index and Merrill Lynch Bond Index. The results from these alternative bond market factors turn out to be similar to those reported in our tables.

4. Downside risk and expected corporate bond returns

We investigate the distributional characteristics of corporate bonds and find that the empirical distribution of bond returns is skewed, peaked around the mode, and has fat tails, implying that extreme returns occur much more frequently than predicted by the normal distribution. Hence, ignoring nonnormality features of the return distribution significantly understates downside risk in bond portfolios, potentially posing a solvency risk for bond investors. We argue for a pricing framework for corporate bonds that builds in nonnormality up front because, beyond its pure statistical merit, the framework offers a significant, practical benefit for investors: the potential to improve portfolio efficiency and reduce its risk relative to unpredictable, extreme negative events.

In this section, we first present the empirical results from testing whether the time-series and cross-sectional returns of corporate bonds are normally distributed. Then, we provide comprehensive empirical evidence supporting the positive relation between downside risk and the cross-section of future bond returns.

4.1. Normality test for corporate bond returns

For each bond in our sample from July 2004 to December 2016, we compute the volatility, skewness, and kurtosis of monthly returns. Panel A of Table A.1 in the online Internet Appendix shows their summary statistics. Panel A tests whether these high-order moments are significantly different from zero based on the time-series distribution of bond returns. Among 38,957 bonds, 84.6% of them have significant volatility at the 10% level or better. In addition, 19,548 bonds exhibit positive skewness, and 19,409 bonds exhibit negative skewness. Among the bonds with positive (negative) skewness, 48.0% (49.5%) are statistically significant at the 10% level or better. Finally, the majority of bonds (26,493) exhibit positive excess kurtosis, and among these bonds, 67.7% are statistically significant at the 10% level or better. We also conduct the Jarque-Bera (JB) normality test, and the last column of Panel A shows that 79.9% of the bonds in our sample exhibit significant JB statistics, rejecting the null hypothesis of normality at the 10% level or better.[19]

Panel B of Table A.1 tests whether these high-order moments are significantly different from zero based on the cross-sectional distribution of bond returns. For each month from July 2004 to December 2016, we compute the volatility (%), skewness, and excess kurtosis of the cross-sectional observations of bond returns and test whether these distributional moments are significantly different from zero. We find that the JB statistics are significant for all months in the sample period, rejecting the null hypothesis of normal distribution of the cross-sectional bond returns.[20]

[19] For 68% of the corporate bonds in our sample, the JB statistics are significant at the 5% level or better, rejecting the null hypothesis of normality.

[20] Bai et al. (2016) test, for the first time in the literature, nonnormality of the return distribution of corporate bonds and also investigate whether

Table 2
Univariate portfolios of corporate bonds sorted by downside risk.
Quintile portfolios are formed every month from July 2004 to December 2016 by sorting corporate bonds based on the 5% VaR, defined as the second lowest monthly return observation over the past 36 months. The original VaR measure is multiplied by - 1 so that a higher VaR indicates higher downside risk. Quintile 1 is the portfolio with the lowest VaR and Quintile 5 is the portfolio with the highest VaR. The portfolios are value weighted using amount outstanding as weights. Table reports the average VaR, the next-month average excess return, the five-factor alpha from stock market factors, the five-factor alpha from bond market factors, and the ten-factor alpha for each quintile. The last five columns report average portfolio characteristics including bond beta (β^{Bond}), illiquidity (ILLIQ), credit rating, time to maturity (years), and amount outstanding (size, in $billion) for each quintile. The last row shows the differences in monthly average returns, the differences in alphas with respect to the factor models. The five-factor model with stock market factors includes the excess stock market return (MKT^{Stock}), the size factor (SMB), the book-to-market factor (HML), the momentum factor (MOM^{Stock}), and the stock liquidity factor (LIQ^{Stock}). The five-factor model with bond market factors includes the excess bond market return (MKT^{Bond}), the default factor (DEF), the term factor (TERM), the bond momentum factor (MOM^{Bond}), and the bond liquidity factor (LIQ^{Bond}). The ten-factor model combines the five stock and five bond market factors. Average excess returns and alphas are defined in monthly percentage terms. Newey-West adjusted *t*-statistics are given in parentheses. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively.

Quintiles	Average	Average	Five-factor stock	Five-factor bond	Ten-factor	Average portfolio characteristics				
	VaR	return	alpha	alpha	alpha	β^{Bond}	ILLIQ	Rating	Maturity	Size
Low VaR	1.59	0.21	0.19	0.03	0.03	0.55	0.57	7.02	4.43	0.56
		(1.10)	(1.28)	(1.05)	(1.09)					
2	2.95	0.34	0.30	0.04	0.05	0.82	1.15	7.69	7.07	0.46
		(2.99)	(2.78)	(1.17)	(1.35)					
3	4.38	0.44	0.37	0.04	0.04	1.06	1.82	7.91	10.39	0.43
		(2.77)	(2.58)	(0.65)	(0.79)					
4	6.71	0.62	0.54	0.17	0.19	1.44	2.72	8.64	13.26	0.41
		(3.02)	(3.32)	(1.82)	(1.98)					
High VaR	15.72	1.20	0.99	0.81	0.75	2.52	5.20	12.16	12.15	0.34
		(4.18)	(4.41)	(4.32)	(3.16)					
High - Low	14.13***	0.99***	0.79***	0.78***	0.72***					
Return/Alpha diff.	(9.94)	(3.95)	(3.82)	(3.90)	(2.82)					

Since the empirical distribution of bond returns is skewed, peaked around the mode, and has fat tails, downside risk—defined as a nonlinear function of volatility, skewness, and kurtosis—is expected to play a major role in the cross-sectional pricing of corporate bonds.

4.2. Univariate portfolio analysis

We first examine the significance of a cross-sectional relation between VaR and future corporate bond returns using portfolio-level analysis. For each month from July 2004 to December 2016, we form quintile portfolios by sorting corporate bonds based on their downside risk (5%VaR), where quintile 1 contains bonds with the lowest downside risk and quintile 5 contains bonds with the highest downside risk. The portfolios are value weighted using amount outstanding as weights. Table 2 shows the average 5% VaR of bonds in each quintile, the next-month value-weighted average excess return, and the alphas for each quintile. The last five columns report the average bond characteristics for each quintile, including the bond market beta, illiquidity, credit rating, time to maturity, and bond size. The last row displays the differences of average returns and the alphas between quintile 5 and quintile 1. Average excess returns and alphas are defined in terms of monthly percentages. Newey and West (1987) adjusted *t*-statistics are reported in parentheses.

Moving from quintile 1 to quintile 5, the average excess return on the downside risk portfolios increases monotonically from 0.21% to 1.20% per month. This indicates a monthly average return difference of 0.99% between quintiles 5 and 1 with a Newey-West *t*-statistic of 3.95, showing that this positive return difference is economically and statistically significant. This result also indicates that corporate bonds in the highest-VaR quintile generate 11.88% per annum higher return than bonds in the lowest-VaR quintile.

In addition to the average excess returns, Table 2 presents the intercepts (alphas) from the regression of the quintile excess portfolio returns on the well-known stock and bond market factors—the excess stock market return (MKT^{Stock}), a size factor (SMB), a book-to-market factor (HML), a momentum factor (MOM^{Stock}), and a liquidity risk factor (LIQ^{Stock}), following Fama and French (1993), Carhart (1997), and Pastor and Stambaugh (2003).[21] The third column of Table 2 shows that, similar to the average excess returns, the five-factor alpha on the downside risk portfolios also increases monotonically from 0.19% to 0.99% per month, moving from the low-VaR to the high-VaR quintile, indicating a positive and significant alpha difference (downside risk premium) of 0.79% per month (*t*-stat.= 3.82). This result suggests that loss-averse bond investors prefer high expected return and low VaR.

Beyond the well-known stock market factors (size, book to market, momentum, and liquidity risk), we also test whether the significant return difference between high-VaR bonds and low-VaR bonds can be explained

the higher order moments of corporate bonds predict their future returns. In this paper, we merge the main findings of Bai et al. (2016) with our empirical analyses on downside risk so that Bai et al. (2016) remains a permanent working paper.

[21] The factors MKT^{Stock} (excess market return), SMB (small minus big), HML (high minus low), MOM (winner minus loser), and LIQ (liquidity risk) are described in and obtained from Kenneth French's and Lubos Pastor's online data libraries: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ and http://faculty.chicagobooth.edu/lubos.pastor/research/.

by prominent bond market factors. Following Fama and French (1993), Elton et al. (1995), and Bessembinder et al. (2009), we use the aggregate corporate bond market, default spread, and term spread factors. The excess bond market return (MKT^{Bond}) is proxied by the value-weighted average return of all corporate bonds in our sample in excess of the one-month T-bill return. The default factor (DEF) is defined as the difference between the return on a market portfolio of long-term corporate bonds (the composite portfolio on the corporate bond module of Ibbotson Associates) and the long-term government bond return. The term factor (TERM) is defined as the difference between the monthly long-term government bond return (from Ibbotson Associates) and the one-month Treasury bill rate. In addition to MKT^{Bond}, DEF, and TERM, we use the momentum factor for the corporate bond market. Following Jostova et al. (2013), the bond momentum factor (MOM^{Bond}) is constructed from 5×5 bivariate portfolios of credit rating and bond momentum, defined as the cumulative returns over months from $t-7$ to $t-2$ (formation period). We also use the liquidity risk factor (LIQ^{Bond}) of Lin, Wang, and Wu (2011), constructed for the corporate bond market. Specifically, we follow Lin, Wang, and Wu (2011) and estimate the liquidity beta over a five-year rolling window for each individual bond. We then sort individual bonds into ten decile portfolios each month by the preranking liquidity beta. The liquidity factor used in Lin, Wang, and Wu (2011) is defined as the average return difference between the high liquidity beta portfolio (decile 10) and the low liquidity beta portfolio (decile 1).[22]

Similar to our earlier findings from the average excess returns and the five-factor alphas from stock market factors, the fourth column of Table 2 shows that, moving from the low-VaR to the high-VaR quintile, the five-factor alpha from bond market factors increases monotonically from 0.03% to 0.81% per month. The corresponding five-factor alpha difference between quintiles 5 and 1 is positive and highly significant; 0.78% per month with a t-statistic of 3.90.

The fifth column of Table 2 presents the ten-factor alpha for each quintile from the combined five stock and five bond market factors. Consistent with our earlier results, moving from the low-VaR to the high-VaR quintile, the ten-factor alpha increases monotonically from 0.03% to 0.75% per month, generating a positive and highly significant risk-adjusted return spread of 0.72% per month with a t-statistic of 2.82.

Finally, we examine the average bond characteristics of VaR-sorted portfolios. As shown in the last five columns of Table 2, bonds with high downside risk have a higher market beta, lower liquidity, higher credit risk, longer time to maturity, and smaller size. This creates a potential concern about the interaction between downside risk and bond characteristics. We provide several ways to handle this concern. Specifically, in the following sections, we test whether the positive relation between VaR and the cross-section of bond returns holds once we control for the market beta, credit rating, maturity, liquidity, and size based on bivariate portfolio sorts and Fama-MacBeth (1973) regressions.

4.3. Bivariate portfolio analysis

Table 3 presents the results from the bivariate sorts of VaR and bond characteristics. Quintile portfolios are formed every month from July 2004 to December 2016 by first sorting corporate bonds into five quintiles based on their credit ratings (Panel A), maturity (Panel B), size (Panel C), or illiquidity (Panel D); then within each quintile portfolio, bonds are sorted further into five subquintiles based on their VaR. This methodology, under each characteristic-sorted quintile, produces subquintile portfolios of bonds with dispersion in downside risk but that have nearly identical characteristics, such as rating, maturity, size, and illiquidity. The portfolios are value weighted using amount outstanding as weights. VaR,1 represents the lowest VaR-ranked bond quintiles within each of the five bond characteristic-ranked quintiles. Similarly, VaR,5 represents the highest VaR-ranked quintiles within each of the five bond characteristic-ranked quintiles.

Panel A of Table 3 shows that the ten-factor alpha increases monotonically from VaR,1 to VaR,5 quintile. More importantly, after controlling for credit rating, the ten-factor alpha difference between high- and low-VaR bonds remains positive, 0.46% per month, and highly significant with a t-statistic of 2.60. We further investigate the interaction between VaR and credit rating by sorting investment-grade and noninvestment-grade bonds separately into bivariate quintile portfolios. As expected, the positive relation between VaR and expected returns is stronger for noninvestment-grade bonds with the alpha spread of 0.78% per month (t-stat.= 3.38), but the positive downside risk premium remains significant for investment-grade bonds even after controlling for credit ratings, with the alpha spread of 0.38% per month (t-stat.= 2.46).

Panel B of Table 3 reports the results from the bivariate sorts of downside risk and maturity. After controlling for bond maturity, the ten-factor alpha difference between high- and low-VaR bonds remains positive, 0.91% per month, and highly significant with a t-statistic of 4.10. We further examine the interaction between VaR and maturity by sorting short-maturity bonds (one year $\leq$ maturity $\leq$ five years), medium-maturity bonds (five years $<$ maturity $\leq$ ten years), and long-maturity bonds (maturity $>$ ten years) separately into bivariate quintile portfolios based on their VaR and maturity. After controlling for maturity, the alpha spread between the VaR,1 and VaR,5 quintiles is 0.64% per month (t-stat.= 2.66) for short-maturity bonds, 0.81% per month (t-stat.= 2.88) for medium-maturity bonds, and 0.99% per month (t-stat.= 4.59) for long-maturity bonds. Although the economic significance of these alpha spreads is similar across the three maturity groups, the statistical significance of the alpha spread is greater for medium- and long-maturity bonds.

[22] We thank Junbo Wang for providing us with the data on LIQ1 and LIQ2 used by Lin, Wang, and Wu (2011). The monthly data on LIQ1 and LIQ2 are available from January 1999 to March 2009. We extend their liquidity risk factors up to December 2016 and use LIQ1 to calculate the risk-adjusted returns (alpha) of VaR-sorted portfolios. The results from LIQ2 are very similar to those reported in Table 2.

Table 3
Bivariate portfolios of corporate bonds sorted by downside risk controlling for bond characteristics.

Quintile portfolios are formed every month from July 2004 to December 2016 by first sorting corporate bonds based on credit rating (Panel A), maturity (Panel B), size (Panel C), or illiquidity (Panel D). Then, within each control quintile, corporate bonds are further sorted into subquintiles based on their 5% VaR, defined as the second lowest monthly return observation over the past 36 months multiplied by - 1. "VaR,1" is the portfolio of corporate bonds with the lowest VaR within each quintile portfolio, and "VaR, 5" is the portfolio of corporate bonds with the highest VaR within each quintile portfolio. The portfolios are value weighted using amount outstanding as weights. Table shows the ten-factor alpha for each quintile. The last row shows the differences in alphas with respect to the ten-factor model, which combines the five stock and five bond market factors. Newey-West adjusted t-statistics are given in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Panel A: Controlling for credit rating			Panel B: Controlling for maturity			
	All bonds	Investment grade	Noninvestment grade	All bonds	Short maturity	Medium maturity	Long maturity
VaR,1	0.04	0.02	0.32	−0.05	−0.08	−0.06	0.01
	(1.07)	(0.96)	(3.01)	(−1.19)	(−1.22)	(−1.34)	(0.44)
VaR,2	0.11	0.05	0.31	−0.02	−0.14	−0.01	0.07
	(3.75)	(2.25)	(3.15)	(−0.57)	(−2.47)	(−0.35)	(3.06)
VaR,3	0.12	0.03	0.43	0.12	0.08	0.11	0.18
	(3.11)	(1.23)	(3.68)	(3.23)	(1.37)	(2.03)	(4.73)
VaR,4	0.18	−0.03	0.42	0.20	0.16	0.23	0.27
	(3.02)	(−1.16)	(2.61)	(2.67)	(1.42)	(1.71)	(3.77)
VaR,5	0.51	0.40	1.10	0.86	0.56	0.74	1.00
	(3.41)	(1.39)	(4.46)	(4.36)	(2.78)	(2.91)	(4.83)
VaR,5 - VaR,1	0.46**	0.38**	0.78***	0.91***	0.64***	0.81***	0.99***
Return/Alpha diff.	(2.60)	(2.46)	(3.38)	(4.10)	(2.66)	(2.88)	(4.59)

	Panel C: Controlling for size			Panel D: Controlling for illiquidity		
	All bonds	Small bonds	Large bonds	All bonds	Investment grade	Noninvestment grade
VaR,1	0.06	0.04	0.05	−0.01	−0.01	0.13
	(0.96)	(0.55)	(0.96)	(−0.20)	(−0.28)	(1.15)
VaR,2	0.08	0.18	0.06	0.07	0.02	0.30
	(1.20)	(1.94)	(1.57)	(2.69)	(0.89)	(2.54)
VaR,3	0.19	0.38	0.08	0.09	0.03	0.39
	(4.16)	(4.09)	(1.63)	(1.98)	(1.41)	(3.03)
VaR,4	0.26	0.49	0.10	0.19	0.01	0.65
	(3.10)	(3.05)	(1.56)	(2.50)	(0.45)	(3.43)
VaR,5	0.71	0.83	0.65	0.72	0.37	1.29
	(4.17)	(2.77)	(3.50)	(4.26)	(2.03)	(4.46)
VaR,5 - VaR,1	0.65***	0.79**	0.60***	0.72***	0.38**	1.16***
Return/Alpha diff.	(3.48)	(2.37)	(3.08)	(3.90)	(2.48)	(3.55)

This result makes sense because longer term bonds usually offer higher interest rates but may entail additional risks.

Panel C of Table 3 presents the results from the bivariate sorts of downside risk and bond size measured by bond outstanding value. After controlling for size, the ten-factor alpha difference between high- and low-VaR bonds remains positive, 0.65% per month, and statistically significant. In the same panel, we further investigate the interaction between VaR and bond size by sorting small and large bonds separately into bivariate quintile portfolios. As expected, the positive relation between VaR and expected returns is stronger for bonds with low market value, but the positive link remains strong for bonds with high market value as well.

Panel D of Table 3 demonstrates the results from the bivariate sorts of downside risk and bond illiquidity defined in Eq. (2). After controlling for illiquidity, the ten-factor alpha difference between high- and low-VaR bonds remains positive, 0.72% per month, and statistically significant. In the same panel, we further investigate the interaction between VaR and bond illiquidity by sorting investment-grade and noninvestment-grade bonds separately into bivariate quintile portfolios. As expected, the positive relation between VaR and expected returns is stronger for noninvestment-grade bonds, but the significantly positive link remains strong for investment-grade bonds even after controlling for illiquidity.

4.4. Bond-level Fama-MacBeth regressions

We have thus far tested the significance of downside risk (5% VaR) as the cross-sectional determinant of future bond returns based on the univariate and bivariate portfolio-level analyses. We now examine the cross-sectional relation between risk characteristics and expected returns at the bond level using Fama and MacBeth (1973) regressions. We present the time-series averages of the slope coefficients from the regressions of one-month-ahead excess bond returns on VaR, rating, ILLIQ, β^{Bond} and the control variables, including years-to-maturity (MAT), the natural logarithm of bond amount outstanding (SIZE), lagged excess return (REV), and bond exposure to the default and term factors (β^{DEF}, β^{TERM}). Monthly cross-sectional regressions are run for the following econometric

Table 4
Bond-level Fama-MacBeth cross-sectional regressions.

This table reports the average intercept and slope coefficients from the Fama and MacBeth (1973) cross-sectional regressions of one-month-ahead corporate bond excess returns on the VaR, credit rating, illiquidity (ILLIQ), bond market beta (β^{Bond}) with and without control variables. Bond characteristics include time to maturity (years) and the natural logarithm of amount outstanding (Size). Ratings are in conventional numerical scores, where 1 refers to an AAA rating and 21 refers to a C rating. Higher numerical score means higher credit risk. Other control variables are the default beta (β^{DEF}), the term beta (β^{TERM}), and bond return in previous month (REV). The Fama and MacBeth regressions are run each month for the period from July 2004 to December 2016. Newey and West (1987) t-statistics are reported in parentheses to determine the statistical significance of the average intercept and slope coefficients. The last column reports the average adjusted R^2 values. Numbers in bold denote statistical significance at the 5% level or better.

	Intercept	5% VaR	Rating	ILLIQ	β^{Bond}	β^{DEF}	β^{TERM}	Maturity	Size	REV	Adj. R^2
(1)	−0.011	**0.064**									0.086
	(−0.10)	(4.88)									
(2)	0.127	**0.052**				−0.006	0.002	−0.002	−0.012	**−0.122**	0.173
	(0.96)	(4.36)				(−1.03)	(0.23)	(−0.46)	(−0.84)	(−9.24)	
(3)	−0.182		**0.068**								0.054
	(−1.32)		(3.84)								
(4)	−0.130		**0.064**			−0.008	0.018	**0.015**	−0.001	**−0.119**	0.155
	(−1.23)		(2.84)			(−1.46)	(1.28)	(2.12)	(−1.00)	(−9.36)	
(5)	0.463			**0.081**							0.028
	(3.41)			(6.45)							
(6)	0.304			**0.066**		−0.007	0.041	0.007	0.030	**−0.079**	0.152
	(2.68)			(6.32)		(−0.90)	(1.37)	(1.17)	(0.99)	(−5.29)	
(7)	0.209				**0.486**						0.055
	(1.72)				(3.15)						
(8)	0.224				**0.318**	**−0.023**	0.026	0.004	−0.053	**−0.069**	0.156
	(2.50)				(2.14)	(−2.76)	(1.63)	(0.72)	(−1.13)	(−3.27)	
(9)	−0.195	**0.111**	0.031	**0.047**	−0.097						0.144
	(−1.37)	(5.29)	(1.50)	(6.22)	(−0.94)						
(10)	−0.178	**0.106**	0.030	**0.041**	−0.097	−0.002	0.003	0.002	0.000	**−0.132**	0.217
	(−1.55)	(4.72)	(1.48)	(5.25)	(−0.95)	(−0.31)	(0.31)	(0.30)	(3.22)	(−8.51)	

specification and nested versions thereof:

$$R_{i,t+1} = \lambda_{0,t} + \lambda_{1,t}VaR_{i,t} + \lambda_{2,t}Rating_{i,t} + \lambda_{3,t}ILLIQ_{i,t} + \lambda_{4,t}\beta^{Bond}_{i,t} + \sum_{k=1}^{K}\lambda_{i,k,t}Control_{i,k,t} + \epsilon_{i,t+1}, \tag{3}$$

where $R_{i,t+1}$ is the excess return on bond i in month t+1.

Table 4 reports the time-series average of the intercept and slope coefficients (λ) and the average adjusted R^2 values over the 149 months from July 2004 to December 2016. The Newey-West adjusted t-statistics are reported in parentheses. The univariate regression results show a positive and statistically significant relation between VaR and the cross-section of future bond returns. In Regression (1), the average slope $\lambda_{1,\,t}$ from the monthly regressions of excess returns on VaR alone is 0.064 with a t-statistic of 4.88. The economic magnitude of the associated effect is similar to that shown in Table 2 for the univariate quintile portfolios of VaR. The spread in average VaR between quintiles 5 and 1 is approximately 14.13 (= 15.72 - 1.59); multiplying this spread by the average slope of 0.064 yields an estimated monthly downside risk premium of 90 basis points.

Regressions (3), (5), and (7) show that the average slopes on credit risk (Rating), bond-level illiquidity (ILLIQ), and the bond market beta (β^{Bond}) from the univariate regressions of excess bond returns on these risk characteristics are all positive and statistically significant.[23] Regression specifications (2), (4), (6) and (8) in Table 4 show that after controlling for maturity, size, lagged excess return, β^{DEF}, and β^{TERM}, the average slope coefficients on VaR, rating, ILLIQ, and β^{Bond} remain positive and statistically significant. In other words, controlling for bond characteristics and other risk factors does not affect the positive cross-sectional relation between the individual risk proxies and future bond returns.

Regression (9) tests the cross-sectional predictive power of VaR, rating, ILLIQ, and β^{Bond} simultaneously. The average slopes on VaR and ILLIQ are significantly positive at 0.111 (t-stat.= 5.29) and 0.047 (t-stat.= 6.22), respectively. However, the average slope coefficients on rating and β^{Bond} become insignificant in this general specification, implying that credit rating and the market beta lose their predictive power for future bond returns after VaR and ILLIQ are controlled for.

The last specification, Regression (10), presents the results from the multivariate regression with all bond risk proxies (VaR, Rating, ILLIQ, and β^{Bond}) after controlling for maturity, size, lagged bond return, β^{DEF}, and β^{TERM}. Similar to our findings in Regression (9), the cross-sectional relations between future bond returns and VaR and ILLIQ are positive and highly significant. However, the predictive power of rating and β^{Bond} disappears, indicating that downside risk and liquidity risk have a more pervasive effect on future bond returns than credit risk and market risk.

Among the control variables, only the short-term reversal effect is found to be strong and robust across regression specifications. Thus, in Section 5, we construct a new bond return reversal factor and investigate its performance in predicting the cross-sectional and time-series variations in future bond returns.

[23] These findings are also consistent with the univariate portfolio results reported in Table A.2 of the online Internet Appendix. The average return and alpha spreads between quintiles 5 and 1 of the rating-, ILLIQ-, and β^{Bond}-sorted portfolios are all positive and highly significant.

Table 5
The source of downside risk premium.

In Panel A, all corporate bonds in the sample are grouped into 27 portfolios based on trivariate dependent sorts of volatility (VOL), skewness (SKEW), and kurtosis (KURT). Panel A reports the next-month average returns and the ten-factor alpha for i) high- minus low-volatility portfolio controlling for skewness and kurtosis, ii) high- minus low-skewness portfolio controlling for volatility and kurtosis, and iii) high- minus low-kurtosis portfolio controlling for volatility and skewness. The portfolios are value weighted using amount outstanding as weights. VOL, SKEW, and KURT are calculated using a 36-month rolling window estimation. Panel B reports the average intercept and slope coefficients from the Fama and MacBeth (1973) cross-sectional regressions of one-month-ahead corporate bond excess returns on the VOL, SKEW, and KURT with and without control variables. The Fama and MacBeth regressions are run each month for the period from July 2004 to December 2016. Newey and West (1987) *t*-statistics are reported in parentheses to determine the statistical significance of the average intercept and slope coefficients. The last column reports the average adjusted R^2 values. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively. Numbers in bold denote statistical significance at the 5% level or better.

Panel A: Trivariate dependent-sort portfolios by VOL, SKEW, and KURT

	Average return	Ten-factor alpha		Average return	Ten-factor alpha		Average return	Ten-factor alpha
Low VOL	0.27	0.03	Low SKEW	0.78	0.45	Low KURT	0.49	0.01
	(1.29)	(0.91)		(4.56)	(3.65)		(3.03)	(0.10)
2	0.47	0.05	2	0.65	0.26	2	0.53	0.10
	(2.05)	(0.92)		(3.11)	(2.29)		(3.35)	(1.73)
High VOL	0.91	0.37	High SKEW	0.53	0.23	High KURT	0.55	0.21
	(3.66)	(2.56)		(2.79)	(2.39)		(3.89)	(4.23)
High - Low	0.64***	0.34**	High - Low	−0.25**	−0.22**	High - Low	0.06	0.20**
t-stat	(3.32)	(2.48)	*t*-stat	(−2.47)	(−2.35)	*t*-stat	(0.81)	(2.34)

Panel B: Cross-sectional regressions with VOL, SKEW, and KURT

	Intercept	VOL	SKEW	KURT	Rating	Maturity	Size	β^{DEF}	β^{TERM}	REV	Adj. R^2
(1)	0.231	**0.011**									0.076
	(2.03)	(2.92)									
(2)	0.128	**0.012**			0.025	0.004	−0.011	−0.001	−0.010	**−0.126**	0.180
	(1.15)	(2.66)			(1.22)	(0.84)	(−0.55)	(−0.10)	(−1.06)	(−9.28)	
(3)	0.206	**0.011**	**−0.128**								0.083
	(1.80)	(2.96)	(−2.81)								
(4)	0.127	**0.012**	**−0.099**		0.025	0.005	−0.014	−0.000	−0.010	**−0.125**	0.184
	(1.16)	(2.74)	(−2.88)		(1.21)	(0.85)	(−0.73)	(−0.09)	(−1.01)	(−9.24)	
(5)	0.375			**0.047**							0.012
	(2.26)			(2.73)							
(6)	0.071			**0.016**	**0.052**	**0.012**	−0.034	−0.005	0.013	**−0.108**	0.159
	(0.54)			(2.34)	(1.97)	(2.02)	(−1.13)	(−0.73)	(1.07)	(−7.95)	
(7)	0.214	**0.011**	**−0.102**	−0.001							0.087
	(1.81)	(2.89)	(−2.43)	(−0.05)							
(8)	0.129	**0.012**	**−0.081**	0.002	0.025	0.005	−0.016	−0.000	−0.010	**−0.125**	0.186
	(1.16)	(2.57)	(−2.53)	(0.26)	(1.27)	(0.94)	(−0.86)	(−0.07)	(−1.05)	(−9.36)	

4.5. The source of downside risk premium

Our results thus far show that downside risk is a strong predictor of future bond returns. In this section, we investigate the source of downside risk premium since this measure is a nonlinear function of high-order moments of the return distribution. Specifically, we test if the high-order moments of bond returns—volatility, skewness, and kurtosis—contribute to the predictive power of downside risk. First, we examine the significance of a cross-sectional relation between volatility/skewness/kurtosis and future returns on corporate bonds using portfolio-level analysis. We then investigate the predictive power of volatility, skewness, and kurtosis simultaneously using bond-level cross-sectional regressions.

Table 2 shows that the average return and ten-factor alpha spreads between the high- and low-VaR quintiles are 99 and 72 basis points per month, respectively. To understand the source of downside risk premium, we dissect downside risk into volatility, skewness, and kurtosis components and conduct trivariate dependent-sort portfolio analyses. Specifically, for each month from July 2004 to December 2016, all bonds in the sample are grouped into portfolios based on ascending sorts of volatility, skewness, and kurtosis. To determine the contribution of volatility to the magnitude of downside risk premium, we group all bonds into 27 portfolios using a trivariate dependent-sort on skewness, kurtosis, and then volatility, with the breakpoints for each sort determined by the 33rd and 67th percentile of the sort variable. We then calculate the value-weighted average return for each of the 27 portfolios, as well as the difference in average returns between the high and low (3-1) volatility-sorted portfolio, for each skewness and kurtosis group. To examine the relation between skewness (kurtosis) and future bond returns, we repeat the analysis, sorting first on kurtosis (skewness), then volatility (volatility), and then skewness (kurtosis).[24]

[24] In all analyses, the portfolio sorts are designed to examine the relation between the last sort variable and future bond returns after controlling for the effects of each of the first two sort variables.

Panel A of Table 5 shows that after controlling for skewness and kurtosis, the average return and alpha spreads between the low- and high-volatility sorted tercile portfolios are, respectively, 0.64% and 0.34% per month and highly significant with corresponding t-statistics of 3.32 and 2.48. Panel A also provides evidence for significant skewness premium after controlling for volatility and kurtosis; the average return and alpha spreads between the low- and high-skewness sorted tercile portfolios are, respectively, - 0.25% and - 0.22% per month and statistically significant with corresponding t-statistics of - 2.47 and - 2.35. After controlling for volatility and skewness, the predictive power of kurtosis turns out to be weak both economically and statistically; the average return and alpha spreads between the low- and high-kurtosis sorted tercile portfolios are, respectively, 0.06% and 0.20% per month with corresponding t-statistics of 0.81 and 2.34.[25]

These results indicate that volatility contributes the most to downside risk premiums; 64 basis points out of the 99 basis points per month, and skewness the second; 25 basis points per month, while kurtosis contributes only 6 basis points per month. Consistent with the findings for raw returns, the ten-factor alphas exhibit similar patterns, except that kurtosis contributes somewhat higher, at 20 basis points per month, which is still lower than volatility and skewness premiums. As expected, volatility contributes the most to the alpha spread in VaR-sorted portfolios; 34 basis points out of the 72 basis points per month, and skewness is again the second, 22 basis points per month.[26]

Finally, we examine the cross-sectional relation between volatility, skewness, and kurtosis and expected returns at the bond level using Fama and MacBeth (1973) regressions. Panel B of Table 5 reports the time-series average of the intercept and slope coefficients and the adjusted R^2 values over the 149 months from July 2004 to December 2016. The results show a positive (negative) and statistically significant relation between volatility (skewness) and the cross-section of future bond returns, in both univariate and multivariate regressions. However, the average slope on KURT is not statistically significant after controlling for volatility and skewness, suggesting that kurtosis does not make a robust incremental contribution to predictability. Overall, Table 5 shows that bond return volatility and skewness contribute significantly to the predictive power of downside risk on future bond returns.

5. Common risk factors in the corporate bond market

In this section, we first introduce novel risk factors based on downside risk, credit quality, bond illiquidity, and return reversal and test whether the newly proposed factors are explained by well-established stock and bond market factors. Then, we investigate if the new factors capture systematic variation in bond returns or common risk premiums in the bond market. Finally, we form alternative test assets based on the industry- and the size/maturity-sorted portfolios of corporate bonds and compare the relative performance of the new factors with the commonly used factor models in predicting the cross-sectional dispersion of corporate bond returns.

5.1. New risk factors: DRF, CRF, LRF, and REV

As discussed previously, corporate bonds with high credit risk also have higher downside risk and higher illiquidity both at the bond level and portfolio level, indicating a positive cross-sectional relation between credit risk and bond illiquidity and downside risk. More importantly, default/credit risk is one of the most frequently monitored barometers, closely followed by rating agencies, financial regulators, and investors. Thus, it is natural to use credit risk (proxied by credit rating) as the first sorting variable in the construction of these new bond market factors.

We construct the bond factors in a similar vein to Fama and French (2015) and rely on independent sorts. To construct the downside risk factor for corporate bonds, for each month from July 2004 to December 2016, we form bivariate portfolios by independently sorting bonds into five quintiles based on their credit rating and five quintiles based on their downside risk (measured by 5% VaR). The downside risk factor, *DRF*, is the value-weighted average return difference between the highest-VaR portfolio and the lowest-VaR portfolio across the rating portfolios. The credit risk factor, CRF_{VaR}, is the value-weighted average return difference between the lowest-rating (i.e., highest credit risk) portfolio and the highest-rating (i.e., lowest credit risk) portfolio across the VaR portfolios.

The liquidity risk and the return reversal factors are constructed similarly using independent sorts. The liquidity risk factor, *LRF*, is the value-weighted average return difference between the highest-illiquidity and the lowest-illiquidity portfolios across the rating portfolios. The return reversal factor, *REV*, is the value-weighted average return difference between the short-term loser and the short-term winner portfolios (losers-minus-winners) across the rating portfolios.[27] The above independent sorts used to construct *LRF* and *REV* produce two additional credit risk

[25] At an earlier stage of the study, we sort corporate bonds into univariate quintile portfolios based on kurtosis and find that the average return and alpha spreads between the low- and high-kurtosis quintile portfolios are 0.38% (t-stat. = 2.56) and 0.32% per month (t-stat. = 2.30), respectively. Although kurtosis itself is a significant predictor of future bond returns, its incremental contribution to downside risk premium is much lower after controlling for volatility and skewness.

[26] When we focus on the alpha spreads reported in Table 5, Panel A, the sum of the volatility, skewness, and kurtosis premiums is 0.76% per month (= 0.34% + 0.22% + 0.20%), which is similar to downside risk premium of 0.72% per month reported in Table 2. Similarly, when we focus on the average return spreads reported in Table 5, Panel A, the sum is 0.95% per month (= 0.64% + 0.25% + 0.06%), which is somewhat lower than the downside risk premium of 0.99% per month reported in Table 2. These results indicate that since VaR is a function of volatility, skewness, kurtosis, and even higher order moments of the return distribution, moments higher than kurtosis may contribute (though very small) to downside risk premium.

[27] Table A.3 of the online Internet Appendix reports the average monthly excess returns for the 5 × 5 portfolios independently sorted on *Rating* and *VaR*, *Rating* and *ILLIQ*, and *Rating* and *REV*.

Table 6
Summary statistics for corporate bond factors.

Panel A reports the descriptive statistics for the excess bond market return and the newly constructed bond factors. MKT^{Bond} is the corporate bond market excess return constructed using the value-weighted average return of all corporate bonds in the sample (in excess of one-month T-bill rate). Downside risk factor (*DRF*) is constructed by independently sorting corporate bonds into 5×5 quintiles based on the 5% VaR and credit rating. *DRF* is the value-weighted average return difference between the highest-VaR portfolio minus the lowest-VaR portfolio within each rating portfolio. Liquidity risk factor (*LRF*) is constructed by independently sorting corporate bonds into 5×5 quintiles based on illiquidity and credit rating. *LRF* is the value-weighted average return difference between the highest-illiquidity portfolio minus the lowest-illiquidity portfolio within each rating portfolio. Return reversal factor (*REV*) is constructed by independently sorting corporate bonds into 5×5 quintiles based on the previous month return and credit rating. *REV* is the value-weighted average return difference between the short-term loser and the short-term winner portfolio within each rating portfolio. Credit risk factor (*CRF*) is the average of the *CRF* obtained from forming the *DRF*, *LRF*, and *REV*, and $CRF = 1/3(CRF_{VaR} + CRF_{ILLIQ} + CRF_{REV})$. Panel B reports the intercepts (alphas) and *t*-statistics (in parentheses) from time-series regressions of the factors on the commonly used stock and bond market factors. Model 1 includes the five stock market factors defined in Table 2. Model 2 includes the five bond market factors defined in Table 2. Model 3 is the ten-factor model that combines the five stock and five bond market factors. DRF and CRF covers the period from July 2004 to December 2016. LRF and REV cover the period from August 2002 to December 2016.

Panel A: Summary statistics on the value-weighted bond factors

	Mean	*t*-stat
MKT^{Bond}	0.39	3.58
Downside risk factor (DRF)	0.70	3.60
Credit risk factor (CRF)	0.43	2.78
Liquidity risk factor (LRF)	0.52	5.02
Return reversal factor (REV)	0.41	4.05

Panel B: Factor alpha from the ten-factor model

	Model 1	Model 2	Model 3
DRF alpha	0.83	0.79	0.80
t-stat	(2.90)	(3.19)	(2.76)
CRF alpha	0.44	0.34	0.35
t-stat	(2.92)	(2.01)	(1.89)
LRF alpha	0.37	0.32	0.32
t-stat	(3.15)	(2.79)	(2.45)
REV alpha	0.48	0.49	0.46
t-stat	(4.10)	(4.46)	(4.74)

factors, CRF_{ILLIQ} and CRF_{REV}. The credit risk factor *CRF* is defined as the average of CRF_{VaR}, CRF_{ILLIQ}, and CRF_{REV}.[28]

Panel A of Table 6 reports the summary statistics for the new bond factors (DRF, CRF, LRF, and REV). Over the period from August 2002 to December 2016, the corporate bond market risk premium, MKT^{Bond}, is 0.39% per month with a *t*-statistic of 3.58. The value-weighted DRF factor has an economically and statistically significant risk premium of 0.70% per month with a *t*-statistic of 3.60. The value-weighted CRF, LRF, and REV factors also have significant premiums of 0.43% per month (*t*-stat.= 2.78), 0.52% per month (*t*-stat.= 5.02), and 0.41% per month (*t*-stat. = 4.05) respectively. Fig. 1 plots the monthly time series of the new factors (DRF, CRF, LRF, and REV).

Since risk premiums are expected to be higher during financial and economic downturns, we examine the average risk premiums for the newly proposed factors, DRF, CRF, LRF, and REV, during recessionary versus nonrecessionary periods, determined by the Chicago Fed National Activity Index (CFNAI).[29] As expected, we find that the average risk premium on the DRF factor is higher at 0.84% per month during recessionary periods (CFNAI $\leq$ - 0.7), whereas it is 0.67% per month during nonrecessionary periods (CFNAI $>$ - 0.7). The average risk premiums on the CRF and LRF factors are 0.75% and 1.17% per month during recessionary periods, and the corresponding values are lower during nonrecessionary periods; 0.36% and 0.40%, respectively. These magnitudes provide clear evidence that the newly proposed DRF, CRF, and LRF risk factors generate economically large risk premiums during economic downturns.[30]

Finally, we examine whether conventional stock and bond market factors explain the newly proposed factors of corporate bonds. For each of the new factors (DRF, CRF, LRF, and REV), Panel B of Table 6 presents the alphas from (i) the five-factor stock market model of Fama-French (1993), Carhart (1997), and Pastor and Stambaugh (2003) with the stock market (MKT^{Stock}), SMB, HML, MOM^{Stock}, and LIQ^{Stock} factors; (ii) the five-factor bond market model of Fama-French (1993), Elton, Gruber, and Blake (1995), Bessembinder et al. (2009), Jostova et al. (2013), and Lin, Wang, and Wu (2011) with the bond market (MKT^{Bond}), DEF, TERM, MOM^{Bond}, and LIQ^{Bond} factors; and (iii) the ten-factor model that combines the aforementioned five stock and five bond market factors.

Table 6, Panel B, shows that the alphas from the five-factor stock market model, the five-factor bond market model, and the combined ten-factor model are all positive and highly significant for the DRF, CRF, LRF, and REV factors. Overall, these results indicate that the existing stock and bond market factors are not sufficient to capture the information content in the newly proposed bond factors so that these novel factors capture an important source of common return variation in corporate bonds missing from long-established stock and bond market factors.[31]

---

[28] We rely on the independently sorted 5×5 portfolios to construct the factors to be consistent with our univariate and bivariate portfolio results from quintile portfolios. However, we also follow Fama and French (2015) and Hou et al. (2015) and construct 2×3 and $2 \times 2 \times 2 \times 2$ factors. The results are presented in Table A.4. As shown in Panel B of Table A.4, the correlations between different versions of the same factors (5×5, 2×3, and $2 \times 2 \times 2 \times 2$ factors) are very high.

[29] CFNAI is a monthly index designed to assess overall economic activity and related inflationary pressure (see, e.g., Allen et al., 2012). CFNAI is a weighted average of 85 existing monthly indicators of national economic activity. It is constructed to have an average value of zero and a standard deviation of one. An index value below (above) - 0.7 corresponds to a recessionary (nonrecessionary) period.

[30] Contrary to our findings for the DRF, CRF, and LRF factors, the bond return reversal factor has lower (higher) average return during recessionary (nonrecessionary) periods. Specifically, the average return on the value-weighted REV factor is 0.29% per month during recessionary periods and 0.43% per month during nonrecessionary periods. This result suggests that REV is a nonrisk characteristic of corporate bonds.

[31] Following Fama and French (2015), we also conduct factor spanning tests by running time-series regressions of each of the four factors on the other three factors. Table A.5 of the online Internet Appendix shows that the regression intercepts (alphas) for the DRF and LRF remain economi-

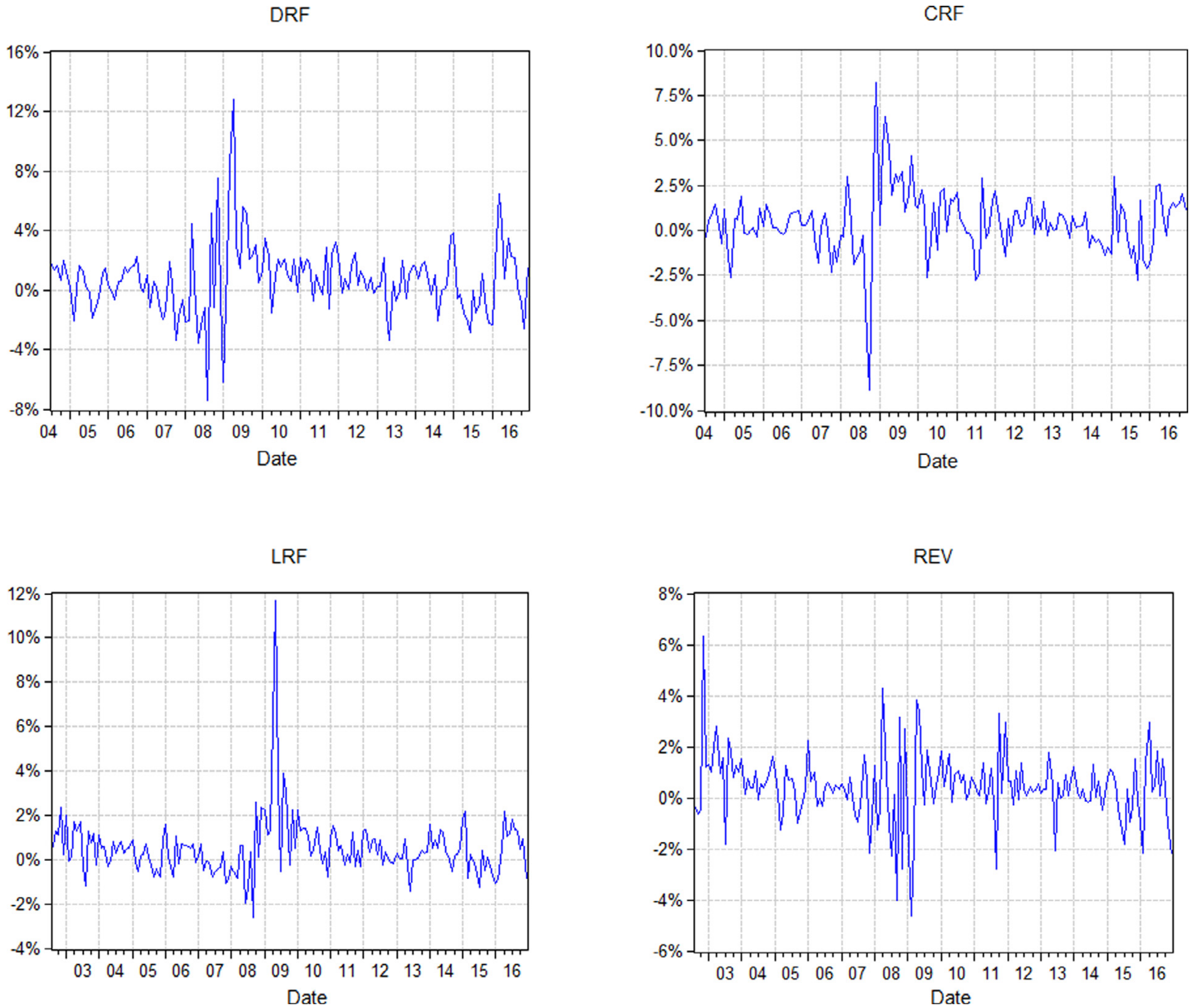


Fig. 1. DRF, CRF, LRF, and REV Factors: 2002–2016. This figure plots the monthly time series of the value-weighted downside risk factor (DRF), credit risk factor (CRF), liquidity risk factor (LRF), and return reversal factor (REV). DRF and CRF cover the period from July 2004 to December 2016. LRF and REV cover the period from August 2002 to December 2016.

5.2. Are exposures to bond risk factors priced?

If the newly proposed DRF, CRF, LRF, and REV factors truly capture systematic variation in bond returns or common risk premiums in the corporate bond market, exposures of corporate bonds to these factors (factor betas) are supposed to predict cross-sectional differences in future bond returns. Motivated by Daniel and Titman (1997) and Brennan et al. (1998), we investigate this issue using bond-level cross-sectional regressions. Specifically, for each bond and each month in our sample, we estimate the factor betas from the monthly rolling regressions of excess bond returns on the DRF, CRF, LRF, and REV factors over a 36-month fixed window after controlling for the bond market factor (MKT^{Bond}):

$$R_{i,t} = \alpha_{i,t} + \beta_{i,t}^{MKT} \cdot MKT_t^{Bond} + \beta_{i,t}^{Factor} \cdot Factor_t + \epsilon_{i,t}, \quad (4)$$

where $Factor_t$ is one of the four value-weighted bond market factors: DRF, CRF, LRF, and REV, and $\beta_{i,t}^{Factor}$ is one of the four factor betas: $\beta_{i,t}^{DRF}$, $\beta_{i,t}^{CRF}$, $\beta_{i,t}^{LRF}$, and $\beta_{i,t}^{REV}$ of bond i in month t.

We examine the cross-sectional relation between β^{DRF}, β^{CRF}, β^{LRF}, and β^{REV} and expected returns at the bond level using Fama and MacBeth (1973) regressions. Regression (1) in Table 7 presents positive and statistically significant relations between all three factor betas (β^{DRF}, β^{CRF}, β^{LRF}) and the cross-section of future bond returns, whereas the bond exposure to the return reversal factor (β^{REV}) turns out to be statistically insignificant. The results indicate that the DRF, CRF, and LRF factors capture common risk premiums in the corporate bond market, instead of proxying for bond characteristics. Another notable point in Table 7 is

cally and statistically significant, whereas the CRF alpha becomes smaller and statistically indistinguishable from zero.

Table 7

Are exposures to new bond factors priced?

This table reports the average intercept and slope coefficients from the Fama and MacBeth (1973) cross-sectional regressions of one-month-ahead corporate bond excess returns on the bond market betas, with and without control variables. The bond market betas (β^{Bond}, β^{DRF}, β^{CRF}, β^{LRF}, and β^{REV}) are estimated for each bond from the time-series regressions of bond excess returns on the excess bond market return and the associated bond factors (DRF, CRF, LRF, or REV) using a 36-month rolling window estimation. Bond characteristics include VaR, credit rating, illiquidity (ILLIQ), bond return in previous month (REV), time to maturity (years), and the natural logarithm of bond amount outstanding (Size). Numbers in bold denote statistical significance at the 5% level or better.

	Intercept	β^{Bond}	β^{DRF}	β^{CRF}	β^{LRF}	β^{REV}	VaR	Rating	ILLIQ	REV	Maturity	Size	Adj. R^2
(1)	0.513	**0.301**	**0.399**	**0.816**	**0.267**	−0.174							0.103
	(3.32)	(2.85)	(2.99)	(4.89)	(2.39)	(−1.27)							
(2)	−0.088	**0.202**	**0.275**	**0.118**	**0.272**	−0.049	**0.158**						0.152
	(−0.97)	(2.34)	(2.78)	(2.23)	(2.31)	(−0.54)	(4.61)						
(3)	−0.237	**0.269**	**0.403**	**0.430**	**0.321**	−0.095		**0.095**					0.128
	(−1.74)	(2.63)	(2.37)	(4.36)	(2.74)	(−0.71)		(3.27)					
(4)	0.315	**0.282**	**0.443**	**0.545**	**0.331**	−0.076			**0.095**				0.136
	(2.94)	(2.66)	(2.82)	(3.86)	(2.33)	(−0.54)			(5.59)				
(5)	0.466	**0.261**	**0.321**	**0.627**	**0.162**	−0.234				**−0.035**			0.125
	(3.90)	(2.74)	(2.69)	(4.01)	(2.22)	(−1.91)				(−3.69)			
(6)	0.411	**0.296**	**0.357**	**0.799**	**0.255**	−0.165					0.009		0.129
	(3.12)	(2.71)	(2.75)	(4.83)	(2.28)	(−1.16)					(1.44)		
(7)	0.975	0.318	**0.434**	**0.737**	**0.305**	−0.132						−0.090	0.112
	(2.31)	(1.28)	(2.92)	(4.73)	(2.36)	(−0.93)						(−1.68)	
(8)	−0.346	0.172	**0.351**	**0.493**	**0.289**	−0.102	**0.126**	0.044	**0.055**	**−0.118**	−0.005	0.007	0.234
	(−2.90)	(1.05)	(2.51)	(2.60)	(2.85)	(−1.17)	(5.60)	(1.11)	(6.81)	(−7.40)	(−0.88)	(0.36)	

that β^{REV} does not predict future bond returns in any of the regression specifications with and without controlling for bond characteristics, whereas one-month lagged return (REV) remains a strong cross-sectional determinant of future bond returns. Thus, we conclude that REV is a nonrisk bond characteristic instead of a common risk factor in the bond market.[32]

Brennan et al. (1998) investigate the extent to which expected equity returns can be explained by risk factors (e.g., SMB, HML) rather than by nonrisk firm characteristics (e.g., firm size and book-to-market ratio).[33] Following Brennan et al. (1998), Regressions (2) to (7) in Table 7 control for the risk and nonrisk characteristics of corporate bonds (VaR, rating, illiquidity, lagged return, maturity, and size) one-by-one. Regression (2) shows that downside risk, both as a risk characteristic of individual bonds (VaR) and a common risk factor (DRF), remains a strong predictor of future bond returns because the average slope coefficients on both β^{DRF} and VaR are positive and highly significant with t-statistics of 2.78 and 4.61, respectively.

Regression (8) in Table 7 presents results from the multivariate regression with all factor betas while simultaneously controlling for all risk and nonrisk bond characteristics. Similar to our findings from Regressions (1) through (7), the cross-sectional relations between future bond returns and all three factor betas (β^{DRF}, β^{CRF}, β^{LRF}) are positive and highly significant. Regression (8) provides evidence that the DRF, CRF, and LRF remain significant risk factors along with downside risk, illiquidity, and one-month lagged return as significant characteristics in the cross-section of bond returns.

5.3. Alternative test portfolios

Lewellen et al. (2010) provide evidence that the low power of asset pricing tests is driven by characteristic-sorted portfolios (used as test assets) that do not have sufficient independent variation in the factor loadings. To improve the power of asset pricing tests, Lewellen et al. (2010) suggest testing risk factors based on alternative test portfolios. Thus, we consider two sets of test portfolios that are not related to the risk characteristics examined in previous sections, that is, downside risk, credit rating and illiquidity.

The first set of test portfolios is based on 5×5 independently sorted bivariate value-weighted portfolios of size and maturity. The second set of test portfolios is based on 30 value-weighted industry-sorted portfolios. We examine the relative performance of factor models in explaining the time-series and cross-sectional variations in the 25 size/maturity-sorted and 30 industry-sorted portfolios of corporate bonds. We investigate the empirical performance of the following five different models:

- Model 1: The five-factor model with the stock market factors of Fama and French (1993), Carhart (1997), and Pastor and Stambaugh (2003), including the excess stock market return (MKT^{Stock}), the size factor (SMB), the book-to-market factor (HML), the stock momentum factor (MOM^{Stock}), and the stock liquidity risk factor (LIQ^{Stock}).

[32] Bali et al. (2017) propose return-based factors based on the short-term reversal, momentum, and long-term reversal effects in the bond market. They provide an illiquidity-based explanation of the short-term reversal effect, but they do not test whether REV captures systematic variation in bond returns or common risk premiums in the corporate bond market.

[33] Daniel and Titman (1997) show that portfolios of firms that have similar characteristics of size and book-to-market ratio, but different loadings on the SMB and HML factors of Fama-French (1993), have similar average returns. They use this result to conclude that these firm characteristics (size and book-to-market ratio) have an independent influence on expected stock returns.

- Model 2: The five-factor model with the bond market factors of Fama and French (1993), Elton et al. (1995), Bessembinder et al. (2009), Jostova et al. (2013), and Lin et al. (2011), including the bond market factor (MKT^{Bond}), the default factor (DEF), the term factor (TERM), the bond momentum factor (MOM^{Bond}), and the bond liquidity factor (LIQ^{Bond}).
- Model 3: The three-factor model introduced in the paper, including the excess bond market return (MKT^{Bond}), the credit risk factor (CRF), and the bond liquidity risk factor (LRF).
- Model 4: The four-factor model introduced in the paper, including the excess bond market return (MKT^{Bond}), the downside risk factor (DRF), the credit risk factor (CRF), and the bond liquidity risk factor (LRF).
- Model 5: The five-factor model introduced in the paper, including the excess bond market return (MKT^{Bond}), the downside risk factor (DRF), the credit risk factor (CRF), the bond liquidity risk factor (LRF), and the return reversal factor (REV).

Panel A of Table 8 shows that the adjusted R^2, averaged across the 25-size/maturity-sorted portfolios, is only 7% for Model 1, implying that a large proportion of the variance in 25 bond portfolio returns is not explained by the commonly used stock market factors. Panel B shows that the average adjusted R^2 from Model 2 improves to 18% mainly because of the stronger predictive power of the aggregate bond market factor. Compared to the results in Panels A and B, the average R^2 from Model 3 is stronger. As shown in Panel C of Table 8, when we augment MKT^{Bond} with our newly proposed credit and liquidity risk factors (CRF and LRF), the average adjusted R^2 further increases from 18% to 27%, suggesting that these new credit and liquidity risk factors of corporate bonds capture significant cross-sectional information about the portfolio returns that is not fully picked up by the aggregate bond market factor. Moreover, the average alpha of the 25 size/maturity-sorted portfolios reduces from 0.33% per month to 0.14% per month when we replace the existing four bond market factors (DEF, TERM, MOM^{Bond}, and LIQ^{Bond}) with our two new bond factors (CRF and LRF).

We also investigate the relative performance of our CRF and LRF factors with the existing credit and liquidity risk factors proposed by earlier studies. Fama and French (1993) introduce a bond factor to capture the credit risk component of corporate bond returns. In Model 2, we use the default factor (DEF) of Fama-French (1993) defined as the difference between the returns on aggregate corporate bond index and aggregate government bond index. We should note that the average return on the DEF factor is economically and statistically insignificant for the period 2002–2016, 0.03% per month with a t-statistic of 0.19, whereas the average return on the CRF factor is highly significant, both economically and statistically, 0.43% per month (t-stat. $= 2.78$). These results along with the increase in average R^2 moving from Model 2 (average $R^2 = 18\%$) to Model 3 (average $R^2 = 27\%$) indicate that the DEF factor used in the literature is constructed too coarsely and there is a scope for defining a better credit risk factor, CRF, as the difference between returns on low-rated and high-rated corporate bonds.

The literature has also shown the importance of a liquidity factor in corporate bond returns. Lin et al. (2011) propose a bond factor to capture the liquidity risk component of corporate bond returns. As detailed in Section 4.2, we construct a tradable, return-based liquidity factor following Lin, Wang, and Wu (2011) and find that LIQ^{Bond} has a mean of 0.13% per month (t-stat.$= 2.45$) over the period from July 2002 to December 2016, whereas the average return on our LRF factor has a higher premium of 0.52% per month with a t-statistic of 5.02. Again, these results along with the improvement in average R^2 moving from Model 2 to Model 3 suggest that there is an opportunity to propose a superior liquidity risk factor, LRF, as the difference between returns on illiquid and liquid corporate bonds.

We now investigate the incremental performance of the DRF in predicting the cross-sectional variation in corporate bond portfolios. Compared to the remarkable results in Panel C obtained from Model 3, the average R^2 from Model 4 is even stronger. As shown in Panel D of Table 8, when we augment Model 3 with our newly proposed DRF, the average adjusted R^2 substantially increases from 27% to 56%, suggesting that the DRF captures significant incremental information about the cross-sectional variation in bond portfolio returns. However, Panel E of Table 8 shows that when we augment Model 4 with the bond REV, the average adjusted R^2 increases only by 1% (from 56% to 57%).[34] Overall, the results in Table 8 indicate that the newly proposed four-factor model with the market, downside, credit, and liquidity risk factors outperforms the existing factor models in explaining the returns of the size/maturity-sorted portfolios of corporate bonds.

As an alternative way of evaluating the relative performance of the factor models, we focus on the magnitude and statistical significance of the alphas on the 25-size/maturity portfolios generated by Models 1 through 5. Panel A of Table 8 shows that the five-factor model with the stock market factors (Model 1) generates economically significant alpha for all 25 portfolios, ranging from 0.14% to 0.58% per month. Consistent with the economic significance, the alphas are statistically significant for all 25 portfolios. As shown in the last row of Panel A in Table 8, the average alpha across the 25 portfolios is very large, 0.42% per month, and highly significant with a p-value less than 0.01 according to the Gibbons et al. (1989, GRS) test. Panel B of Table 8 shows that the magnitude and statistical significance of the alphas decrease when moving from Model 1 to Model 2. However, the five-factor model with the existing bond market factors (Model 2) still generates economically and statistically significant alphas, ranging from 0.12% to 0.51% per month, for 23 out of 25 portfolios. Similar to our findings in Panel A, the last row of Panel B shows that the average alpha across the 25

[34] This result is consistent with the factor spanning test results in Table A.5 that the REV factor is closely related to the LRF and in line with the illiquidity-based explanation of the short-term reversal effect, proposed by Bali, Subrahmanyam, and Wen (2017).

Table 8
Explanatory power of alternative factor models for 25-size/maturity-sorted bond portfolios.

The table reports the intercepts (alphas), the *t*-statistics, and the adjusted R^2 values for the time-series regressions of the test portfolios' excess returns on alternative factors. The 25 value-weighted test portfolios are formed by independently sorting corporate bonds into 5×5 quintile portfolios based on size (amount outstanding) and maturity and then constructed from the intersections of the size and maturity quintiles. Five alternative factor models are considered. Model 1 is the five-factor model with stock market factors, including the excess stock market return (MKT^{Stock}), the size factor (SMB), the book-to-market factor (HML), the stock momentum factor (MOM^{Stock}), and the stock liquidity factor (LIQ^{Stock}). Model 2 is the five-factor model with bond market factors: the excess bond market return (MKT^{Bond}), the default spread factor (DEF), the term spread factor (TERM), the bond momentum factor (MOM^{Bond}), and the bond liquidity factor (LIQ^{Bond}). Model 3 is the three-factor model with the excess bond market return (MKT^{Bond}), credit risk factor (CRF), and liquidity risk factor (LRF). Model 4 is the four-factor model with the excess bond market return (MKT^{Bond}), the downside risk factor (DRF), the credit risk factor (CRF), and the liquidity risk factor (LRF). Model 5 is the five-factor model with the excess bond market return (MKT^{Bond}), the downside risk factor (DRF), the credit risk factor (CRF), the liquidity risk factor (LRF), and the return reversal factor (REV). The sample covers the period from July 2004 to December 2016.

Panel A: Model 1

	Alpha (α)						*t*-statistics						Adj. R^2				
	Short	2	3	4	Long		Short	2	3	4	Long		Short	2	3	4	Long
Small	0.38	0.53	0.58	0.42	0.55	Small	3.44	3.74	3.49	2.63	3.26	Small	0.10	0.09	0.07	0.07	0.09
2	0.31	0.47	0.51	0.52	0.51	2	4.06	4.06	3.60	2.60	3.13	2	0.10	0.10	0.05	0.05	0.07
3	0.24	0.38	0.41	0.43	0.53	3	4.18	4.16	3.25	3.18	3.04	3	0.17	0.13	0.10	0.05	0.01
4	0.23	0.31	0.41	0.37	0.50	4	3.71	3.50	3.15	2.57	2.50	4	0.13	0.08	0.07	0.04	0.02
Big	0.14	0.31	0.40	0.41	0.52	Big	2.30	3.14	2.91	2.70	2.35	Big	0.05	0.04	0.05	0.03	0.02
Average $\lvert\alpha\rvert$	**0.42**											Average R^2	**0.07**				
p-GRS	< 0.01																

Panel B: Model 2

	Short	2	3	4	Long		Short	2	3	4	Long		Short	2	3	4	Long
Small	0.35	0.48	0.51	0.40	0.49	Small	3.07	3.33	3.03	2.50	2.83	Small	0.13	0.15	0.13	0.15	0.12
2	0.28	0.39	0.42	0.45	0.40	2	3.52	3.38	3.03	2.21	2.39	2	0.17	0.19	0.19	0.07	0.12
3	0.20	0.29	0.27	0.30	0.38	3	3.57	3.32	2.40	2.41	2.24	3	0.30	0.28	0.33	0.24	0.17
4	0.18	0.24	0.28	0.24	0.37	4	3.17	2.98	2.35	1.77	1.86	4	0.30	0.30	0.29	0.20	0.14
Big	0.12	0.28	0.31	0.32	0.39	Big	2.00	2.87	2.27	2.12	1.74	Big	0.13	0.14	0.14	0.11	0.10
Average $\lvert\alpha\rvert$	**0.33**											Average R^2	**0.18**				
p-GRS	< 0.01																

Panel C: Model 3

	Short	2	3	4	Long		Short	2	3	4	Long		Short	2	3	4	Long
Small	0.15	0.22	0.25	0.14	0.20	Small	1.45	1.65	1.58	0.88	1.21	Small	0.29	0.29	0.22	0.17	0.20
2	0.13	0.17	0.20	0.16	0.18	2	1.90	1.77	1.69	0.82	1.13	2	0.29	0.29	0.22	0.17	0.20
3	0.07	0.12	0.06	0.14	0.24	3	1.38	1.55	0.64	1.11	1.45	3	0.48	0.47	0.50	0.27	0.18
4	0.04	0.08	0.08	0.08	0.18	4	0.79	1.02	0.71	0.56	0.92	4	0.45	0.40	0.41	0.21	0.13
Big	0.02	0.12	0.12	0.17	0.20	Big	0.39	1.29	0.89	1.07	0.87	Big	0.20	0.17	0.21	0.07	0.07
Average $\lvert\alpha\rvert$	**0.14**											Average R^2	**0.27**				
p-GRS	0.03																

(*continued on next page*)

Table 8 (*continued*)

Panel D: Model 4

Alpha (α)	Short	2	3	4	Long
Small	0.02	0.04	0.03	−0.08	−0.02
2	0.05	0.06	0.06	−0.08	−0.03
3	0.01	0.04	−0.05	0.01	0.06
4	−0.01	0.00	−0.04	−0.06	−0.04
Big	−0.05	0.01	−0.05	−0.02	−0.08
Average $\lvert\alpha\rvert$	**0.04**				
p-GRS	0.06				

t-statistics	Short	2	3	4	Long
Small	0.23	0.47	0.28	−0.71	−0.17
2	0.88	0.79	0.67	−0.57	−0.21
3	0.37	0.59	−0.57	0.12	0.44
4	−0.32	0.05	−0.45	−0.49	−0.22
Big	−1.12	0.15	−0.51	−0.15	−0.43

Adj. R^2	Short	2	3	4	Long
Small	0.63	0.64	0.66	0.62	0.61
2	0.65	0.65	0.64	0.56	0.56
3	0.65	0.64	0.68	0.48	0.46
4	0.62	0.56	0.61	0.43	0.43
Big	0.55	0.50	0.58	0.45	0.46
Average R^2	**0.56**				

Panel E: Model 5

Alpha (α)	Short	2	3	4	Long
Small	0.02	0.05	0.04	−0.06	−0.01
2	0.05	0.06	0.06	−0.07	0.00
3	0.02	0.04	−0.04	0.02	0.07
4	−0.01	0.01	−0.03	−0.05	−0.02
Big	−0.04	0.02	−0.03	0.01	−0.04
Average $\lvert\alpha\rvert$	**0.03**				
p-GRS	0.06				

t-statistics	Short	2	3	4	Long
Small	0.26	0.49	0.37	−0.55	−0.05
2	0.94	0.83	0.68	−0.49	−0.04
3	0.44	0.63	−0.52	0.16	0.50
4	−0.20	0.12	−0.33	−0.42	−0.10
Big	−0.94	0.29	−0.31	0.07	−0.24

Adj. R^2	Short	2	3	4	Long
Small	0.62	0.64	0.66	0.62	0.61
2	0.65	0.65	0.64	0.56	0.57
3	0.65	0.64	0.67	0.48	0.45
4	0.62	0.56	0.61	0.42	0.43
Big	0.57	0.50	0.59	0.47	0.48
Average R^2	**0.57**				

portfolios is large, 0.33% per month, and highly significant according to the GRS test.

Panel D of Table 8 presents substantially different results compared to Panels A and B. The newly proposed four-factor model with DRF, CRF, and LRF (Model 4) generates economically and statistically insignificant alphas for all 25 portfolios. As shown in the last row of Panel D, the average alpha across the 25 portfolios is very low, economically insignificant at 0.04% per month (p-value = 0.06), and it is not statistically significant at the 5% level.[35]

Overall, these results confirm the superior performance of the newly proposed factors in predicting the cross-sectional variation in the returns of the 25-size/maturity-sorted portfolios of corporate bonds. Thus, the 4-factor model with DRF, CRF, and LRF factors provides a more accurate characterization of the abnormal returns on portfolios of corporate bonds, which has important practical implications. For example, a typical bond portfolio manager using a traditional factor model (such as Model 1 or 2) thinks that he or she outperforms the standard benchmark with economically large alphas. However, the results in Panel D of Table 8 indicate that these significantly large abnormal returns generated by the existing factor models are in fact compensation for downside, credit, and liquidity risks. Therefore, institutional investors in the corporate bond market should account for bond exposure to the DRF, CRF, and LRF factors to accurately determine the risk-adjusted performance of their bond portfolios.

We also test the relative performance of the factor models using the 30-industry portfolios based on the Fama-French (1997) industry classification. Table 9 shows that the adjusted R^2, averaged across the 30-industry portfolios, is 13% for Model 1, 18% for Model 2, 31% for Model 3, and 37% for Model 4. These results show that the newly proposed four-factor model performs better than the existing stock and bond market factors in explaining the returns of the industry-sorted portfolios of corporate bonds.

We then focus on the magnitude and statistical significance of the alphas for the 30-industry portfolios. As shown in Table 9, Model 1 generates economically significant alphas for 25 out of the 30 portfolios, ranging from 0.28% to 1.33% per month. Consistent with their economic significance, the alphas are also statistically significant for 24 out of 30 portfolios. As shown in the last row of Table 9, the average alpha across the 30 portfolios is very large, 0.55% per month, and highly significant. The results from Model 2 are somewhat better. As shown in Table 9, Model 2 generates economically significant alphas for most of the 30 industry portfolios, ranging from 0.19% to 1.08% per month. As shown in the last row of Table 9, the average alpha across the 30 portfolios is economically large, 0.41% per month, and highly significant.

Similar to our findings from the 25-size/maturity portfolios, Table 9 presents considerably different results from the new four-factor model for the 30-industry portfolios.

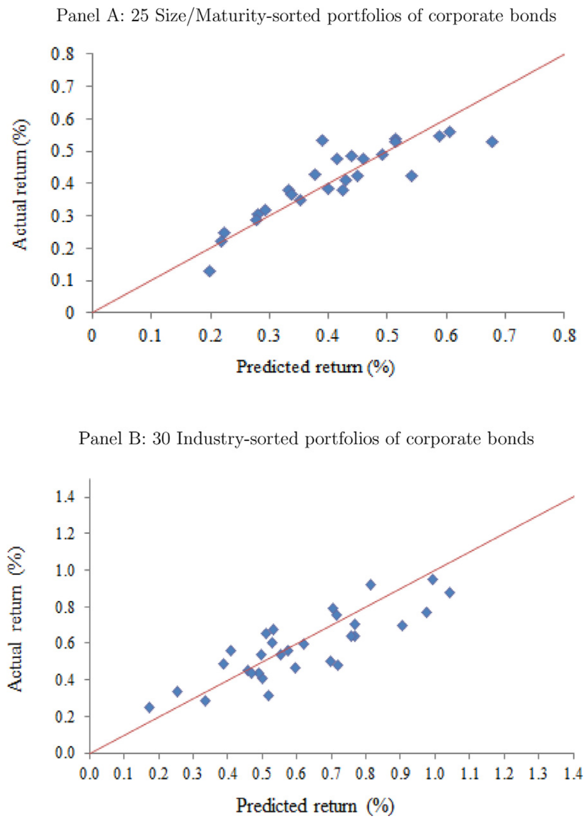


Fig. 2. Average performance of the four-factor model with DRF, CRF, and LRF. The figure plots the monthly mean excess return versus the predicted excess return in percent for the four-factor model with MKT^{Bond}, DRF, CRF, and LRF factors. Test assets are the value-weighted 25 size/maturity-sorted portfolios of corporate bonds in the top panel and the 30 industry-sorted portfolios of corporate bonds in the bottom panel. The sample covers the period from July 2004 to December 2016.

Model 4 with DRF, CRF, and LRF generates statistically insignificant alphas (at the 10% level) for all 30 portfolios, with only one economically significant alpha for one of the 30-industry portfolios. As shown in the last row of Model 4, the average alpha across the 30 portfolios is very low and economically insignificant, at 0.14% per month.[36] Overall, these results provide supporting evidence for the remarkable performance of the newly proposed factors in predicting the cross-sectional variation in the returns of the 30-industry portfolios of corporate bonds.[37]

Finally, Fig. 2 plots the monthly mean excess return (i.e., actual return) versus the predicted excess return for the four-factor with MKT^{Bond}, DRF, CRF, and LRF. The test assets are the 25-size/maturity- and 30-industry-sorted portfolios of corporate bonds. Consistent with our earlier findings,

[35] Panel E of Table 8 shows that when we augment Model 4 with the bond REV, the average alpha reduces only by one basis point per month, indicating low incremental contribution of the REV factor to portfolio return predictability over the four-factor model with the DRF, CRF, and LRF factors.

[36] Note that the average alpha is only 14 basis points (bps) per month but is statistically significant with a p-value of 0.03 according to the GRS test.

[37] Tables A.6 and A.7 of the online Internet Appendix report the explanatory power of the 2×3 and $2 \times 2 \times 2 \times 2$ factors for the 25 size/maturity- and 30 industry-sorted bond portfolios, respectively. The results from the 2×3 and $2 \times 2 \times 2 \times 2$ factors are similar to those obtained from the 5×5 factors presented in Tables 8 and 9.

Table 9
Explanatory power of alternative factor models for 30 industry-sorted bond portfolios.
The table reports the intercepts (alphas), the *t*-statistics, and the adjusted R^2 values for the time-series regressions of the test portfolios' excess returns on alternative factors. The value-weighted industry portfolios are formed by sorting corporate bonds into 30 portfolios based on the Fama-French (1997) industry classifications. Five alternative factor models are defined in Table 8.

Industry #	Industry description	Model 1			Model 2			Model 3			Model 4			Model 5		
		Alpha (α)	$t(\alpha)$	R^2	Alpha (α)	$t(\alpha)$	R^2	Alpha (α)	$t(\alpha)$	R^2	(α)	$t(\alpha)$	R^2	(α)	$t(\alpha)$	R^2
1	Food	0.37	3.33	0.10	0.25	2.42	0.23	0.10	0.94	0.25	0.08	0.79	0.27	0.16	1.51	0.30
2	Beer	0.28	3.19	0.05	0.22	2.59	0.14	0.14	1.69	0.18	0.11	1.43	0.30	0.14	1.66	0.30
3	Tobacco	0.43	2.40	0.08	0.39	2.11	0.03	0.21	1.12	0.10	0.14	0.83	0.21	0.24	1.33	0.23
4	Games	0.79	2.24	0.13	0.69	1.92	0.13	0.11	0.33	0.30	0.02	0.05	0.36	0.05	0.15	0.36
5	Books	0.55	1.86	0.38	0.44	1.40	0.32	−0.22	−0.77	0.46	−0.31	−1.17	0.52	−0.22	−0.79	0.53
6	Household	0.45	2.17	0.10	0.39	1.85	0.09	0.22	1.02	0.14	0.14	0.71	0.26	0.18	0.87	0.26
7	Clothes	0.68	2.24	0.26	0.40	1.43	0.37	−0.18	−0.60	0.36	−0.22	−0.78	0.38	0.00	0.01	0.41
8	Health	0.48	2.69	0.02	0.41	2.32	0.09	0.18	1.04	0.16	0.16	0.90	0.17	0.22	1.23	0.18
9	Chemicals	0.52	2.42	0.35	0.40	1.79	0.29	−0.08	−0.37	0.42	−0.17	−0.88	0.54	−0.05	−0.24	0.55
10	Textiles	0.66	1.59	0.03	0.48	1.13	0.04	0.19	0.44	0.05	0.13	0.31	0.07	0.27	0.61	0.07
11	Construction	0.70	3.17	0.17	0.56	2.75	0.31	0.18	0.97	0.47	0.14	0.78	0.50	0.19	0.99	0.49
12	Steel	0.75	2.72	0.18	0.77	2.60	0.09	0.21	0.81	0.32	0.12	0.48	0.42	0.24	0.95	0.43
13	Fabric	1.33	2.49	0.02	1.08	1.97	0.01	0.83	1.48	0.01	0.79	1.40	0.01	0.86	1.46	0.01
14	Electrical equipment	0.65	1.94	0.11	0.38	1.09	0.06	−0.02	−0.06	0.13	−0.05	−0.15	0.14	0.17	0.47	0.16
15	Automobiles	0.76	2.52	0.25	0.54	1.71	0.20	−0.17	−0.68	0.53	−0.24	−1.03	0.57	−0.11	−0.43	0.58
16	Transportation equipment	0.47	1.30	0.03	0.33	0.91	0.03	0.13	0.35	0.06	0.06	0.16	0.10	0.04	0.10	0.09
17	Mines	0.35	1.14	0.06	0.17	0.56	0.10	−0.25	−0.85	0.18	−0.30	−1.00	0.19	−0.08	−0.25	0.22
18	Coal	0.48	1.27	0.02	0.30	0.83	0.10	−0.11	−0.31	0.15	−0.23	−0.67	0.25	−0.12	−0.33	0.25
19	Oil	0.65	1.04	0.01	0.48	0.76	0.02	0.18	0.28	0.04	0.08	0.13	0.06	0.08	0.12	0.06
20	Utilities	0.28	2.42	0.07	0.19	1.95	0.33	0.04	0.37	0.31	−0.01	−0.08	0.46	0.05	0.57	0.47
21	Communication	0.37	2.43	0.13	0.21	1.49	0.30	−0.06	−0.49	0.49	−0.10	−0.90	0.56	−0.03	−0.25	0.57
22	Services	0.43	2.42	0.17	0.28	1.65	0.29	−0.13	−0.96	0.56	−0.17	−1.32	0.60	−0.07	−0.52	0.62
23	Business equipment	0.39	2.59	0.17	0.26	1.83	0.28	0.03	0.22	0.39	−0.02	−0.14	0.47	0.07	0.57	0.49
24	Paper	0.50	2.08	0.24	0.37	1.52	0.24	−0.12	−0.60	0.51	−0.19	−1.01	0.57	−0.06	−0.29	0.59
25	Transportation	0.54	3.40	0.13	0.44	3.10	0.34	0.14	1.11	0.47	0.10	0.80	0.55	0.17	1.35	0.56
26	Wholesale	0.46	2.59	0.12	0.30	1.86	0.28	0.07	0.47	0.43	0.03	0.20	0.48	0.11	0.72	0.49
27	Retail	0.54	2.36	0.12	0.37	1.64	0.17	0.02	0.09	0.34	−0.04	−0.19	0.39	0.02	0.12	0.39
28	Restaurant	0.40	1.55	0.14	0.28	1.09	0.16	−0.25	−1.14	0.43	−0.32	−1.52	0.49	−0.20	−0.92	0.50
29	Finance	0.43	3.30	0.08	0.36	2.85	0.15	0.03	0.30	0.53	−0.01	−0.08	0.61	−0.01	−0.06	0.60
30	Other	0.73	3.27	0.15	0.55	2.55	0.25	−0.05	−0.30	0.53	−0.07	−0.43	0.54	0.03	0.15	0.55
Average $\lvert\alpha\rvert$		**0.55**		**0.13**	**0.41**		**0.18**	**0.15**		**0.31**	**0.14**		**0.37**	**0.12**		**0.38**
p-GRS		< 0.01			< 0.01			0.03				0.03		0.03		

the scatter plots in Fig. 2 are most dense around the 45-degree line, indicating that the newly proposed four-factor model provides a good fit of the actual portfolio returns.

6. Robustness check

In this section, we conduct a battery of robustness checks, but we present and discuss these findings in the online Internet Appendix to save space. As discussed earlier, bond risk characteristics are correlated. To address a potential concern about what unique information each risk characteristic carries, in Section A.1, we construct orthogonalized risk characteristics by running contemporaneous cross-sectional regressions of one risk characteristic on the remaining three variables for each month in our sample.[38] Then, we repeat the Fama-MacBeth regressions in Table A.8 of the online Internet Appendix with the orthogonalized risk characteristics and find that with and without the control variables, the orthogonalized rating and orthogonalized market beta lose their significance, whereas the orthogonalized measures of downside risk and illiquidity remain highly significant in predicting the cross-sectional dispersion of bond returns.

Downside risk has so far been proxied by the 5% VaR. The results remain intact when we use two alternative measures of downside risk: the 10% VaR and the 10% ES that are described in Section A.2 of the online Internet Appendix. Table A.9 of the online Internet Appendix shows that the average returns and alphas on these alternative factors of downside risk, constructed based on the 10% VaR and 10% ES, are positive and highly significant.

We reexamine the properties of the LRF based on two alternative proxies of liquidity: the Roll (1984) and Amihud (2002) illiquidity measures that are described in Section A.3 of the online Internet Appendix. As presented in Table A.10 of the online Internet Appendix, the average returns and alphas on these alternative factors of liquidity risk turn out to be economically and statistically significant.

In Section A.4 of the online Internet Appendix, we provide evidence from alternative measures of credit risk: the distance to default (DD) and implied CDS. Table A.11 of the online Internet Appendix presents Fama-MacBeth regressions using DD and CDS to substitute for credit rating. The results from the firm-level measures of credit risk (DD, CDS) turn out to be similar to those obtained from the bond-level measure of credit risk (rating).

Our empirical analyses are so far based on the Enhanced TRACE transaction data from July 2002 to December 2016. To check whether our results are sensitive to different datasets, we use an extended sample of corporate bonds gathered from a range of data sources covering a longer time period from January 1977 to December 2016. Section A.5 of the online Internet Appendix describes the construction of this comprehensive dataset. As shown in Table A.12 of the online Internet Appendix, our main findings are robust to an extended sample of corporate bond data compiled from different sources including the quoted- and transaction-based bond data.

7. Conclusion

An extensive literature examines the cross-sectional determinants of stock returns. There is, however, surprisingly little research on the common risk factors that explain the cross-section of corporate bond returns. This paper aims to fill this gap by identifying common risk factors that predict the cross-sectional differences in corporate bonds.

In contrast to the commonly used stock market factors and aggregate macroeconomic variables that have been investigated in the literature for bond returns, the common risk factors we identify are motivated by the unique features of individual corporate bonds. Specifically, we find that downside risk, credit risk, and liquidity risk positively predict the cross-sectional variation in future bond returns. We then introduce novel risk factors based on these prevalent bond risk characteristics. We show that all new factors have economically and statistically significant risk premiums, which cannot be explained by the existing stock and bond market factors. We also find a strong short-term reversal effect in the cross-section of corporate bond returns and hence introduce a bond return reversal factor. However, a detailed investigation of the reversal factor indicates that one-month lagged return is a strong nonrisk bond characteristic instead of a common risk factor in the bond market.

We further examine the explanatory power of the newly proposed risk factors for alternative test portfolios sorted by bond size, maturity, and industry. We find that the four-factor model with the bond market factor and our new factors (DRF, CRF, LRF) outperforms all models considered in the literature in explaining the returns of the industry/size/maturity-sorted portfolios of corporate bonds. The results also indicate that the significantly large abnormal returns (alphas) on corporate bond portfolios, generated by the existing factor models, are in fact compensation for downside, credit, and liquidity risks. Thus, institutional investors in the corporate bond market should account for bond exposure to the newly proposed DRF, CRF, and LRF factors to accurately estimate the risk-adjusted performance of bond portfolios.

References

Acharya, V., Amihud, Y., Bharath, S., 2013. Liquidity risk of corporate bond returns: conditional approach. J. Financ. Econ. 110, 358–386.
Allen, L., Bali, T.G., Tang, Y., 2012. Does systemic risk in the financial sector predict future economic downturns? Rev. Financ. Stud. 25, 3000–3036.
Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. J. Financ. Mark. 5, 31–56.
Ang, A., Chen, J., Xing, Y., 2006. Downside risk. Rev. Financ. Stud. 19, 1191–1239.
Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Math. Finance 9, 203–228.
Arzac, E.R., Bawa, V.S., 1977. Portfolio choice and equilibrium in capital markets with safety-first investors. J. Financ. Econ. 4, 277–288.
Bai, J., Bali, T. G., Wen, Q., 2016. Do the distributional characteristics of corporate bonds predict their future returns?Unpublished working paper,Georgetown University.
Bai, J., Wu, L., 2016. Anchoring credit spreads to firm fundamentals. J. Financ. Quant. Anal. 51, 1521–1543.

[38] Each one of these risk characteristics (VaR, Rating, ILLIQ, and β^{Bond}) is orthogonalized with respect to the remaining three variables by running separate contemporaneous cross-sectional regressions for each characteristic so that the results do not depend on the order of orthogonalization.

Bali, T.G., Demirtas, K.O., Levy, H., 2009. Is there an intertemporal relation between downside risk and expected returns? J. Financial and Quantitative Analysis 44, 883–909.
Bali, T. G., Subrahmanyam, A., Wen, Q., 2017. Return-based factors for corporate bonds. Unpublished working paper,Georgetown University and University of California at Los Angeles.
Bao, J., Pan, J., Wang, J., 2011. The illiquidity of corporate bonds. J. Finance 66, 911–946.
Baumol, W.J., 1963. An expected gain-confidence limit criterion for portfolio selection. Manag. Sci. 10, 174–182.
Bawa, V.S., Lindenberg, E.B., 1977. Capital market equilibrium in a mean-lower partial moment framework. J. Financ. Econ. 5, 189–200.
Bessembinder, H., Kahle, K.M., Maxwell, W.F., Xu, D., 2009. Measuring abnormal bond performance. Rev. Financ. Stud. 22, 4219–4258.
Bessembinder, H., Maxwell, W.F., Venkataraman, K., 2006. Market transparency, liquidity externalities, and institutional trading costs in corporate bonds. J. Financ. Econ. 82, 251–288.
Blume, M., Keim, D., Patel, S., 1991. Return and volatility of low-grade bonds 1977–1989. J. Finance 46, 49–74.
Brennan, M., Chordia, T., Subrahmanyam, A., 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. J. Financ. Econ. 49, 345–373.
Carhart, M.M., 1997. On persistence in mutual fund performance. J. Finance 52, 57–82.
Chen, L., Lesmond, D., Wei, J., 2007. Corporate yield spreads and bond liquidity. J. Finance 62, 119–149.
Choi, J., Kim, Y., 2018. Anomalies and market (dis)integration. J. Monetary Econ. forthcoming.
Choi, J., Kronlund, M., 2018. Reaching for yield by corporate bond mutual funds. Rev. Financ. Stud 31, 1930–1965.
Chordia, T., Goyal, A., Nozawa, Y., Subrahmanyam, A., Tong, Q., 2017. Are capital market anomalies common to equity and corporate bond markets? J. Financ. Quant. Anal. 52, 1301–1342.
Culp, C., Nozawa, Y., Veronesi, P., 2018. Option-based credit spreads. Am. Econ. Rev. 108, 454–488.
Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross-sectional variation in stock returns. J. Finance 52, 1–33.
Dick-Nielsen, J., Feldhutter, P., Lando, D., 2012. Corporate bond liquidity before and after the onset of the subprime crisis. J. Financ. Econ. 103, 471–492.
Elton, E.J., Gruber, M.J., Blake, C., 1995. Fundamental economic variables, expected returns, and bond fund performance. J. Finance 50, 1229–1256.
Fama, E.F., French, K.R. , 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33, 3–56.
Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116, 1–22.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. J. Polit. Econ. 81, 607–636.
Feroli, M., Kashyap, A. K., Schoenholtz, K., Shin, H. S., 2014. Market tantrums and monetary policy. Unpublished working paper, New York University and University of Chicago.
Gebhardt, W.R., Hvidkjaer, S., Swaminathan, B., 2005. The cross-section of expected corporate bond returns: betas or characteristics? J. Financ. Econ. 75, 85–114.
Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.
Hong, H., Sraer, D., 2013. Quiet bubbles. J. Financ. Econ. 110, 596–606.
Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: an investment approach. Rev. Financ. Stud. 28, 650–705.
Huang, W., Liu, Q., Rhee, S.G., Wu, F., 2012. Extreme downside risk and expected stock returns. J. Bank. Finance 36, 1492–1502.
Jostova, G., Nikolova, S., Philipov, A., Stahel, C., 2013. Momentum in corporate bond returns. Rev. Financ. Stud. 26, 1649–1693.
Kwan, S.H., 1996. Firm-specific information and the correlation between individual stocks and bonds. J. Financ. Econ. 40, 63–80.
Lettau, M., Maggiori, M., Weber, M., 2014. Conditional risk premia in currency markets and other asset classes. J. Financ. Econ. 114, 197–225.
Levy, H., Sarnat, M., 1972. Safety first: an expected utility principle. J. Financ. Quant. Anal. 7, 1829–1834.
Lewellen, J., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset pricing tests. J. Financ. Econ. 96, 175–194.
Lin, H., Wang, J., Wu, C., 2011. Liquidity risk and the cross-section of expected corporate bond returns. J. Financ. Econ. 99, 628–650.
Longstaff, F.A., Mithal, S., Neis, E., 2005. Corporate yield spreads: default risk or liquidity? new evidence from the credit-default-swap market. J. Finance 60, 2213–2253.
Merton, R.C., 1974. On the pricing of corporate debt: the risk structure of interest rates. J. Finance 29, 449–470.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Pastor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. J. Polit. Econ. 111, 642–685.
Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. J. Finance 39, 1127–1139.
Roy, A.D., 1952. Safety first and the holding of assets. Econometrica 20, 431–449.