Factor Premia and Factor Timing:
A Century of Evidence

ANTTI ILMANEN, RONEN ISRAEL, TOBIAS J. MOSKOWITZ, ASHWIN THAPAR, and FRANKLIN WANG*

Preliminary draft: January 2018
Current draft: June 2019

Abstract

We examine four prominent factor premia – value, momentum, carry, and defensive – over a century from six asset classes. First, we verify their existence with a mass of out-of-sample evidence across time and asset markets. We find a 30% drop in estimated premia out of sample, which we show is more likely due to overfitting than informed trading. Second, probing for potential underlying sources of the premia, we find little reliable relation to macroeconomic risks, liquidity, sentiment, or crash risks, despite adding five decades of global economic events. Finally, we find significant time-variation in factor premia that are mildly predictable when imposing theoretical restrictions on timing models. However, significant profitability eludes a host of timing strategies once proper data lags and transactions costs are accounted for. The results offer support for time-varying risk premia models with important implications for theory seeking to explain the sources of factor returns.

*Ilmanen, Israel, Thapar, and Wang are at AQR Capital, and Moskowitz is at Yale University, NBER, and AQR Capital. Corresponding email: tobias.moskowitz@yale.edu. We thank Ron Alquist, Cliff Asness, Jeff Dunn, Sarah Jiang, Michael Katz, Bryan Kelly, John Liew, Lasse Pedersen, and seminar participants at AQR for valuable comments and suggestions. Moskowitz thanks the International Finance Center at Yale University for financial support.

Empirical research in asset pricing finds a host of factors that predict the cross-section of returns. Much debate about their efficiency, the sources of the returns, and their variation over time remains. We examine four prominent factors in the asset pricing literature applied to six different asset classes using a unique dataset with a substantially long time series – in most cases almost a century – to shed new light on these issues. We first document the existence of these factor premia from our mass of out of sample data. Second, we examine proposed sources for these return premia from both risk-based and behavioral models, bringing new data to bear. Finally, we study time variation in factor premia and whether that variation is predictable.

The longer and broader sample has some advantages in addressing these questions. First, for assessing the existence and magnitude of premia, we provide more powerful tests of whether data mining biases have exaggerated documented factor premia, and whether arbitrage activity has changed factor return expectations. Second, the richer sample offers many more economic shocks, including macroeconomic, liquidity, sentiment, and crash events across markets that provide new data to test a number of proposed theories for the return premia. Third, longer sample provides a more powerful laboratory to measure conditional factor premia. These results provide new evidence on asset price dynamics across markets and time.

We study four factors that apply across multiple asset classes, are measureable over our century of data, and are most common to empirical asset pricing models: value, momentum, carry, and defensive. While the literature has produced a proliferation of hundreds of characteristics or "factors," mostly to explain the cross-section of U.S. equity returns, many have recently been questioned due to meager statistical support and lack of robust out-of-sample evidence (Harvey, Liu, and Zhu (2015), Mclean and Pontiff (2016), Hou, Xue, and Zhang (2017)). The handful of prominent factors we focus on have strong in- and out-of-sample evidence and apply to other markets and asset classes,[1] which is why they remain at the center of asset pricing research. However, even among this

[1] See Fama and French (2012), Gorton, Hayashi, and Rouwenhorst (2007), Asness, Liew, and Stevens (2002), Bhojraj and Swaminathan (2005), Asness, Moskowitz, and Pedersen (2013), Asness, Ilmanen, Israel, and Moskowitz (2014), Frazzini and Pedersen (2013), Koijen, Moskowitz, Pedersen, and Vrugt (2018), Brooks and Moskowitz (2018)), and Baltussen, Swinkels, and Van Vliet (2019). Most of the literature uses shorter return histories which start in the 1960s or later, except for U.S. stock selection strategy returns available since the 1920s. Some studies focus on a long history of one factor – e.g., Geczy and Samonov (2016) who study a 200-year track record on momentum – and study a narrower set of questions. Baltussen, Swinkels and Van Vliet (2019) is closest to our sample as it uses a 200-year history of several factor premia, some of which are different factors and different asset classes, but focuses mainly on statistical robustness. While our findings on factor premia existence complement their statistical robustness results, and for different factors, our focus is primarily on the sources of these returns and their conditional variation and predictability over time.

smaller subset of prominent factors, significant debate on their efficacy, underlying economic sources, and variation over time remains.

In addition to providing new out of sample evidence on these issues, we interpret the results through the lens of asset pricing theory to test competing theories for these factor premia:

1. **Overfitting:** There are two types of overfitting we test. The first is purely spurious data mining, where a factor's premium is spurious, and will disappear out of sample in previously unexplored time periods or other asset classes. The second is a factor premium exists, but its premium may be exaggerated in sample, and hence its magnitude may be lower (but not zero) out of sample.
2. **Behavioral finance and limits to arbitrage:** If investor behavioral biases create mispricings that drive the anomalous returns to these factors, then arbitrage activity may influence their efficacy over time. A natural implication is degradation in factor premia after their discovery and publication, if more arbitrage capital chases these factors as a result. In addition, correlations across factors and across asset classes for a given factor may also be higher as a consequence. Factor premia and correlations may also vary with investor sentiment and proxies for arbitrage costs like volatility and liquidity risks according to these theories.
3. **Rational risk-based theories:** There are a variety of models.
   a. **Unconditional risk premia** models imply that the factors are informative about the unconditional stochastic discount factor (sdf). In this case, a factor should have no reliable performance difference out of sample, stable correlation structure through time, and no time-varying return predictability.
   b. **Time-varying risk premia** models are informative about the conditional sdf. These models imply significant time-variation in risk premia and/or risk of the factors. Time-series predictability of factor returns may be present, though not necessarily, and should be more easily detectable with price-based measures.
   c. **Specific unconditional and conditional models** have been proposed that specify the types of risk these factors may be tied to, such as consumption-based, production-based, macroeconomic, downside, and rare disaster risks that are compensated in equilibrium. In some cases, we find reasonable proxies for these risks and exploit the longer and broader sample to test specific models. By studying the same factors applied to all asset classes, we also provide some evidence on the integration of these markets through time and whether a single asset pricing framework applies across markets.

We first confirm the existence, and estimate the magnitude, of value, momentum, carry, and defensive return premia in all six asset classes. We split the sample for each factor into three subperiods: the original sample period in which the factor was discovered, the pre-sample period before the original sample period's starting date (which for many asset classes is data never previously explored), and the post-publication sample after the factor's discovery. This analysis is similar in spirit to Mclean and Pontiff (2016), but with one important addition – the pre-sample evidence, which pre-dates the start of the original sample period in which the factor was discovered and goes back decades before the paper's publication, offers a period where arbitrage activity almost certainly did not trade on the factor.[2] By comparing the pre- versus post-sample performance for each factor (omitting the original sample performance of the factor), we identify the influence of informed trading, and by comparing the average of the pre- and post-samples to the original sample, we capture the influence of data mining and overfitting.

We find out-of-sample performance of the factors is about a third smaller, consistent with what Mclean and Pontiff (2016) find for U.S. equities over a much shorter sample period. However, we find no significant difference between factor performance in the original asset class (which is U.S. equities for value, momentum, and defensive and is currencies for carry) and other "out-of-sample" asset classes. While this evidence suggests some overfitting, the robust evidence of factor premia in the out-of-sample periods and other asset classes rejects spurious data mining as an explanation. Comparing the pre- versus post-sample periods, we find no evidence that arbitrage activity has weakened the premia. The post-sample evidence for the factors is at least as strong as the pre-sample evidence, inconsistent with smart trading eroding the profits of these factors.[3]

We also confirm out of sample that factors are mildly positively correlated across asset classes, consistent with more recent data (Asness, Moskowitz, and Pedersen (2013), Koijen, Moskowitz, Pedersen, and Vrugt (2018), and Brooks and Moskowitz (2018)). These correlations do not vary significantly over time, providing further evidence that arbitrage forces have not affected these factors, and also suggesting that these markets are not perfectly segmented, even a century ago. The

[2] While it is possible (or even likely) that some investors were aware of these factor/anomalies before they were published, and hence some arbitrage activity may have already been taking place before publication, our pre-sample definition ends *before the beginning of the original sample* of the paper that discovered the factor. So, if a paper was published in 1990 but used a sample period from 1963 to 1989, our "pre-sample" period would be from the 1920s until 1962. Hence, arbitrageurs would have had to know about the factor before the sample period of the original study even began, which is often many decades before the actual publication of the paper. Over this distant period it seems safe to assume that little if any arbitrage activity was taking place in the factor.

[3] Mclean and Pontiff (2016) come to a different conclusion using a different test that exploits the gap of a few years between the end of the original sample period in papers and their eventual publication date (on average this is about two years). Our pre-sample evidence is a more powerful test of the arbitrage hypothesis. We also examine four factors across six asset classes, while Mclean and Pontiff (2016) examine 97 factors in U.S. equities only.

existence of the same factor premia in other asset classes that share commonality over a century challenges some asset pricing models.[4]

The historical data across many asset markets provides a greater opportunity to measure economic shocks inspired by theory. We examine macroeconomic variables related to business cycles, growth, and interest rates; political risk, volatility risk, downside risk, tail risk, and crashes (Brunnermeier, Nagel, and Pedersen (2008), Lettau, Maggiori, and Weber (2014)); measures of market liquidity (Acharya and Pedersen (2005), Amihud (2014)); and investment sentiment (Baker and Wurgler (2008)). The additional 50 years of economic events aids in identifying low frequency shocks, such as rare risks. The breadth of asset classes also helps mitigate asset-class specific noise. Despite these advantages, however, we fail to find significant exposures of the asset pricing factors to economic activity or news, where previous claims do not hold up well out of sample.[5] The long-short factor premia we study are less sensitive to macroeconomic conditions than asset class premia.

Finally, we use the data to better capture conditional risk premia of the factors. Time-varying risk premia are notoriously difficult to estimate. Consequently, much debate exists on their identification and, relatedly, the efficacy of factor timing. While the existence of conditional return premia does not necessarily imply predictability in factor returns, for instance risk premia can shift based on unforecastable changes in risk or risk aversion of investors, the reverse is true. Time-series predictability of factor premia implies conditional return premia exist. We focus on a variety of factor timing signals (conditional information) and methods. Since the same unconditional factor premia exist in all asset classes, it is interesting to test whether timing of the same factors works similarly in other asset classes. Almost all of the evidence on factor timing comes from U.S. equity factors, hence our data facilitates a host of out of sample tests for factor timing.

We examine 12 different timing signals across 19 different methodologies for six asset classes and four factors within each asset class, using implementable factor timing trading strategies that serve several purposes. First, trading strategy returns put all timing models on equal footing, making them easily comparable. Second, they allow for economic magnitudes to be quantified. Third, they assess the marginal benefit of factor timing to static factor investing by examining the added

---

[4] For example, investment-based models (Cochrane (1991, 1996), Gomes, Kogan, and Zhang (2003), Carlson, Fisher, and Giammarino (2004), Zhang (2005), Xing (2008), Li, Livdan, and Zhang (2009), Liu, Whited, and Zhang (2009), Belo (2010), Li and Zhang (2010), Cooper and Priestly (2011), Liu and Zhang (2014), Hou, Xue, and Zhang (2015)) are equity-specific and do not seem to fit other asset classes (e.g., currencies or commodities). Likewise, behavioral models should grapple with how cognitive biases and limits to arbitrage apply in other markets.

[5] Our results are broadly consistent with the conclusions in Griffin, Ji, and Martin (2003), Asness, Moskowitz, and Pedersen (2013), and Koijen, Moskowitz, Pedersen, and Vrugt (2018), and inconsistent with the conclusions in Chordia and Shivakumar (2005), Lettau, Maggiori, and Weber (2014), and Hogan et al. (2017), who all study these relationships over much shorter samples and fewer markets.

contribution to an investor's optimal portfolio. Finally, investment returns circumvent issues with other metrics, such as *R*-squares, where degrees of freedom, precision of the parameters, and other statistical biases may matter (Stambaugh (1999)).

We find significant time-variation in realized factor return premia, but find weak and inconsistent evidence of factor timing. The strongest results for factor timing are based on valuation spreads and inverse volatility, which largely supports other findings in the literature (Asness, Friedman, Krail, and Liew (2000), Cohen, Polk, and Vuolteenaho (2003), and Moreira and Muir (2017)). We further find that imposing economic restrictions from theory on timing models (e.g., Campbell and Thompson (2007)) improves out-of-sample predictability. However, we find these results to be inconsistent across asset classes and vary across factors, questioning the robustness of the results. We find even weaker support for factor momentum timing (Arnott et al. (2019) and Gupta and Kelly (2018)) and other timing signals based on macroeconomic variables, factor spreads, inflation, growth, and market volatility. The weaker results stem from weaker out-of-sample evidence in other asset classes and time. In addition, using purely ex ante timing strategies (with all parameters estimated in real time) produces more modest results than prior studies that use in-sample moments to estimate timing parameters.[6] We also show that timing strategies often increase exposure to underlying static factors, which when taken into account, reduces performance. Finally, we find that timing strategies increase turnover and trading costs, which when combined with its meager returns, results in minimal net benefits. Despite limited practical appeal, variation in conditional return premia associated with the factors has important theoretical implications.

Our unique evidence over a century of data in six different asset classes points to the robustness of common factor premia out of sample, their lack of exposure to macroeconomic news and risks, and their significant time variation, which is difficult to capture in real time. Future theory should wrestle with the types of asset pricing models that can accommodate these facts.

The rest of the paper is organized as follows. Section I describes the data and factor construction. Section II measures the existence and magnitude of factor premia with the longer and broader sample, and conducts novel tests of the influence of data mining and informed trading on these premia. Section III tests a host of potential sources of these premia motivated by theory. Section IV analyzes time variation in factor returns through an array of factor timing models, methods, and signals. Section V concludes by relating these results to asset pricing theory.

I. Data and Factor Construction

[6] We also show that proper lagging of macroeconomic news is important to avoid look-ahead bias.

We describe our data and construction of factors.

A. Data

We collect asset returns and economic fundamental data going back as far as February 1877, though most series start in the 1920s. Our main data source is from Global Financial Data, supplemented by Bloomberg and DataStream. The data cover equity indices, government bonds, currencies, and commodities. We also examine nearly a century of returns on individual stocks in the U.S. from the Center for Research in Security Prices (CRSP). We also add individual stock return data from 21 international markets beginning in 1984.

In addition to examining a much longer time series in other asset classes than previously examined, we also study a broader cross-section of securities. Our sample contains equity indices from 43 countries, government bonds from 26 nations, 44 exchange rates, and 40 commodities, with many factor return series offering return histories of more than eighty years. For some perspective, Asness, Moskowitz, and Pedersen (2013), who examine value and momentum returns across some of the same markets, study 18 equity index futures, 10 fixed income securities, 10 exchange rates, and 27 commodities from 1972 to 2010. The data for each asset class and Table IA1 in the internet appendix reports summary statistics on all assets in our sample, including their sample periods. As the table shows, there is rich heterogeneity in returns and extreme events in our sample.

B. Factor definition

We construct cross-sectional value, momentum, carry, and defensive factor portfolios within each asset class. In order to reduce the specter of data mining, in constructing these factors we select the simplest, best-documented measures of each factor.

B.1 Value

We follow simple value measures used in the literature to capture “cheap” versus “expensive” securities within an asset class. For individual equities, we use the book-to-market ratio following Fama and French (1992, 2012) and Asness, Moskowitz, and Pedersen (2013). For global equity indices, we use the aggregate 10-year cyclically-adjusted price-to-earnings ratio CAPE (value-weighted average P/E ratio for all constituent firms in the index).[7] For global bonds, currencies, and commodities we follow Asness, Moskowitz, and Pedersen (2013): the 10-year real bond yield for fixed income, which is the difference between nominal yields and expected inflation, where we use

[7] We use CAPE for country equity indices instead of aggregate book-to-price since Campbell and Shiller (1998) find this to be a better description of equity market valuation. Using aggregate book-to-price of the index constituents does not change our results.

the 3-year trailing Consumer Price Inflation as a proxy for inflation expectations;[8] deviations from Purchasing Power Parity (PPP) exchange rates for currencies, compiled from the Penn World Tables, with additional information from OECD databases and reported inflation indices;[9] the negative of five-year changes in spot prices for commodities, motivated by evidence in DeBondt and Thaler (1985) and Fama and French (1996) that long-term reversal strategies are highly correlated with value strategies in equities.

B.2 Momentum

We use a uniform measure of momentum across all asset classes: the past 12-month cumulative excess-of-cash return on an asset, following Jegadeesh and Titman (1993) and skip the most recent month's return to avoid any microstructure effects, such as bid-ask bounce, that may induce negative short-term autocorrelation, following Asness (1994).[10]

B.3 Carry

We follow Koijen, Moskowitz, Pedersen, and Vrugt (2018) and define carry as the expected return on an asset assuming market conditions are unchanged. For equity indices, this is measured as the futures to spot discount of the front month contract. As futures discount data is generally not available for equity indices before 1990, we extend the sample using excess-of-cash dividend yield. For global currencies, carry equals the short-term interest rate differential between the two countries (specifically, the difference in the 3-month LIBOR rates or the closest 3-month equivalent unsecured lending rates in the two countries).[11] For global government bonds, the carry is the ten-year term spread (10-year yield minus 3-month interest rate). For commodity futures, carry is the return from holding a futures contract if there is no shift in the futures curve, we measured by the percent change in prices between the nearest and next-nearest to maturity contract.

---

[8] While alternate measures (such as survey forecasts) provide a measure of inflation expectations that is forward looking, these forecasts are only available over the recent time period (beginning 1990). Using historical 3-year moving averages of inflation has the benefit of being available consistently over our longer sample period and, as Asness, Moskowitz, and Pedersen (2013) show, delivers similar returns to using inflation survey forecasts over the same sample period.

[9] PPP exchange rates are equilibrium exchange rates that make a basket of goods equally expensive in two countries. The currency value portfolio buys currencies whose nominal exchange rate is lower than the PPP exchange rate and sells currencies whose nominal exchange rate is higher than PPP.

[10] Skipping a month in forming the momentum signal appears important for individual stocks (likely more so in our very early sample period where liquidity may be an issue), but is less important, if unnecessary, for other asset classes. Nevertheless, we use the same definition of momentum across asset classes.

[11] For USD, EUR, JPY, GBP, CHF and all legacy European currencies, we use the 3-month ICE LIBOR daily rate. For most other currencies, we find the local equivalent 3-month interbank offered rate that uses a methodology similar to that of ICE LIBOR. For example, we use the Prague Interbank Offered Rate for CZK. PRIBOR is sponsored by the Financial Markets Association of the Czech Republic, which selects a panel of reference banks and computes the average surveyed interest rate quotations. If there is no obvious LIBOR equivalent, we use 3-month bank bill/CD returns, deposit rates or swap rates (in this order of succession) as a substitute.

B.4 Defensive

Defensive is another well-documented factor that is long low volatility or low beta securities and is short high volatility or high beta securities, following Ang et al. (2006) and Frazzini and Pedersen (2013). We use a simple measure of defensive, which is the beta of the asset with respect to its local market index. For global equity indices and bonds, the betas are estimated based on a 36-month rolling regression of asset returns on the equal-weighted returns of all country indices and bonds, respectively. We do not construct a defensive strategy for currencies because there is no logical market index and an investor can take either side of a currency. We do not construct a defensive strategy for commodities because returns from different commodities (Agricultural, Metal, Energy, Softs) are not very correlated with each other and do not share a common market component.

C. Factor portfolio construction

We form zero-cost, long-short, (and constant leverage) factor portfolios for each asset class using their respective value, momentum, and carry characteristics as defined above. For defensive factors, we form constant-beta portfolios which are not zero-cost, because the dollar notional on the long side (lower-beta) needs to be higher than the short side (higher-beta) in order to stay beta-neutral. These are the same constructions used in the literature, for example Frazzini and Pedersen (2013)'s betting-against-beta, BAB, factor.

For each security i at time t with characteristic $s_{i,t} \in$ (value, momentum, carry, defensive) we first sort securities on the characteristic and assign weights based on each security's cross-sectional ranking within the asset class, where the weights sum up to zero and the portfolio is scaled to one dollar long and short (except for defensive portfolios which are scaled to beta one long and short to be beta-neutral rather than dollar-neutral).[12] Specifically, the weight on security i at time t is

$$w_{it}^{S} = c_t(\text{rank}\left(S_{it}\right) - \Sigma_i \text{rank}\left(S_{it}\right) / N), \tag{1}$$

representing a dollar-neutral long-short portfolio. We include a scaling factor c_t such that the overall portfolio is scaled to one dollar long and short (except for defensive). To ensure that these factor portfolios would be implementable in real time, we further lag the characteristic by a month so that an investor would not have to instantaneously traded based on the signal. An extra month lag ensures the signal or characteristic could be formed/computed before the portfolio is constructed. This is a very conservative choice. The return on the portfolio is

[12] This weighting scheme has been used in Moskowitz, Ooi, and Pedersen (2012), Asness, Moskowitz, and Pedersen (2013), Frazzini and Pedersen (2013), and Koijen, Moskowitz, Pedersen, and Vrugt (2018) and is shown to be similar to other methodologies for constructing factors, such as Fama and French (1993).

$$r_t^S = \sum_i w_{it-1}^S r_{it}.$$

We combine factor portfolios using equal risk weights, where we weight each asset class by the inverse of its standard deviation (estimated using the past 36 months of returns). We use this weighting scheme following Moskowitz, Ooi, and Pedersen (2012), Asness, Moskowitz, and Pedersen (2013), and Koijen, Moskowitz, Pedersen, and Vrugt (2018) to combine asset classes with very different volatilities so that the portfolio's returns are not dominated by one asset class.

II. Existence and Magnitude of Factor Premia

Panel A of Table 1 reports the returns of the factor portfolios over the last century by asset class. The first four columns report the annualized mean, standard deviation, Sharpe ratio, and *t*-statistic of the mean (from zero) of each factor in each asset class. The first four rows report the results combining all asset classes. The Sharpe ratios for value, momentum, carry, and defensive are 0.62, 0.67, 0.84, and 0.78, respectively. The *t*-statistics of the mean return easily reject that these factor premia are zero, ranging from a *t*-stat of 6.0 for value to more than 8.0 for carry. The fifth row reports results for the multifactor portfolio across all asset classes. The Sharpe ratio climbs to 1.59 with a *t*-stat of 14.72, indicating large diversification benefits from combining different factors. The results are consistent with Asness, Moskowitz, and Pedersen (2013) and Asness, Ilmanen, Israel, and Moskowitz (2014), but the evidence here spans a longer sample period (50 more years) and covers more asset classes.

The robustness of these factor premia using another half-century of data not previously explored in many of these asset classes, casts serious doubt on spurious data mining being an explanation for the large return premia associated with these factors. In addition, a *t*-statistic as large as 14.7 (for the returns of the multifactor, multi-asset portfolio), requires more than one trillion random searches to be generated purely by chance – a highly implausible data mining effort. Evidence on the existence of factor premia across asset classes over the last century is clear.[13]

Figure IA1 in the internet appendix plots the Sharpe ratios of each factor in each asset class decade-by-decade. The large premia associated with these factors are not driven by a few wildly profitable periods, and there are periods, as long as a decade, where factors fail to produce positive

[13] The Sharpe ratios of the factors are larger in equities than other asset classes, which a formal *F*-test confirms for all factors except value (Table IA2 in the internet appendix). This evidence could be consistent with overfitting the factor definitions in the original U.S. equity samples, or could be driven by greater breadth in individual equities that leads to better diversification benefits (the cross-section of individual stocks is in the thousands, whereas other asset classes have only a few dozen). In addition, trading costs and other implementation frictions are generally larger in individual equities, so net of cost Sharpe ratios may be about the same across asset classes.

returns. We investigate in Section IV whether we can predict the conditional returns of factors within and across asset classes.

A. Novel Tests of Data Mining and Informed Trading on Factor Premia

We address the potential role of data mining/overfitting and arbitrage activity using our unique sample. This exercise is related to work by Mclean and Pontiff (2016), but applied to a much longer time series and five additional asset classes, albeit on a smaller set of factors. The data offer a novel look at the potential influence of data mining and informed trading on the factors.

A.1 In-sample, post-publication, and pre-sample evidence

Mclean and Pontiff (2016) examine a variety of U.S. equity factors and anomalous characteristics from 1963 to 2014 and carve up their sample period for each factor into in-sample, post-sample, and post-publication subperiods. Across the 97 factors they examine, they find a 26% decline in performance post-sample and a 58% decline post-publication in average returns. Since the post-sample and post-publication subperiods largely overlap, they attribute the 26% decline to overfitting and the additional 32% decline from publication to informed trading.

We bring two new insights to this analysis. First, the century of data allows us to explore the "pre-sample" evidence on the efficacy of factors *before* their discovery that precedes the start of the original data sample of the study documenting the factor. This analysis is particularly useful in distinguishing data mining biases from informed trading since the sample period before the original sample even begins could not plausibly have been known to traders at the time, at least not widely.[14] Comparing the pre-sample evidence versus the original sample evidence then provides a pure test of overfitting biases. Second, we compare the post-sample evidence versus the original sample evidence as a measure of both data mining and informed trading. Looking at the difference between the post-sample evidence and pre-sample evidence should therefore be an unbiased estimate of the influence of informed trading on the factors that "differences out" the data mining bias (since both samples are out of sample). The pre-sample evidence uses a lot more data to tease out the influence of data mining from arbitrage activity compared to Mclean and Pontiff (2016)'s post-samples and post publication-samples, which have only a few years difference between them. Of course, one additional caveat to the pre-sample data is that the data may be of poorer quality, which could also affect the results. This, however, would make the pre-sample results weaker than the post-sample results, which is opposite of what the informed trading hypothesis predicts.

---

[14] It is possible that some traders knew of the value effect (or others) before its academic discovery, but it was not publicized or widely disseminated relative to the period after its publication. Moreover, our pre-sample period precedes the beginning of the original sample period, which is often several decades before the publication date.

As another out of sample test, we also examine the efficacy of factors across asset classes and compare the returns of factors in the original asset classes in which they were discovered to the efficacy of the same factors in other asset classes that the original studies did not examine. We further break down the performance of the factors in other asset classes into pre-, original, and post-publication samples as well. Figure 2 plots the annualized Sharpe ratios of each factor in each asset class over their respective pre-, original-, and post-sample periods.

A.2 Value

The first graph (upper left) in Figure 1 examines the value factor. We use the sample period of Fama and French (1992), from July 1963 to July 1990, as the original sample period for value, following Mclean and Pontiff (2016).[15] For the other asset classes, we use the same dates to define the subperiods. This is done for two reasons. First, most of the research on these factors in other asset classes has been recent (Asness, Moskowitz, and Pedersen (2013)) and hence there is very little post-sample data to work with. Second, we argue that once a factor is discovered in one asset class or market, it seems reasonable to believe that it was applied to other markets and asset classes at or near the same time by practitioners.[16] This choice is likely conservative because we ignore potential out-of-sample periods for other asset classes.

The Sharpe ratio for value from the original sample period in the original asset class (1963 to 1990, U.S. equities) is highlighted on the graph. For U.S. stocks, the graph shows that value performs better in the original sample period than either the pre- or post-sample periods, consistent with overfitting biases. However, there is a sizeable positive Sharpe ratio for U.S. stock value in the pre-sample period and positive (but smaller) performance in the post-sample period, indicating that the value premium is not driven by pure spurious data mining. That value does better in the pre-sample period than in the post-sample period is consistent with value degradation from informed trading. Looking at the other asset classes, however, paints a different picture. For international stocks, the post-sample period outperforms the original sample period by almost twofold. There is no pre-sample period for international stocks since the data start in 1984. For commodities, the original sample period outperforms both the pre- and post-sample periods, but the post-sample period outperforms the pre-sample period fairly significantly. The same pattern holds for equity indices and

---

[15] Although academic research also recognizes Rosenberg, Reid, and Lanstein (1985) as the first academic publication on value, the Fama and French (1992) paper is the one most often cited and the sample periods used in both papers are similar. Both of these papers apply solely to U.S. stocks.

[16] Indeed, the genesis of Asness, Moskowitz, and Pedersen (2013) came from a practitioner (Asness) implementing value and momentum strategies across many asset classes and markets shortly after their discovery in the early 1990's, but nearly 20 years prior to the publication of Asness, Moskowitz, and Pedersen (2013).

fixed income, where the original sample produces the largest returns, but the post-sample period produces better performance than the pre-sample period. Across all asset classes, a formal test of whether the Sharpe ratio of value is the same in the original period versus the out-of-sample periods is rejected (*p*-value of 0.048). Focusing exclusively on the two out-of-sample periods, a formal test of whether the post-sample performance is the same as the pre-sample performance, as a test of whether arbitrage activity diminished the profits to value, fails to reject (*p*-value of 0.924). We find no significant difference between the pre- and post-sample performance of value, and if anything, find stronger performance post-discovery than pre-discovery, inconsistent with arbitrage activity diminishing the value premium.

Aggregating all asset classes, a diversified value factor has a Sharpe ratio in the original sample of 1.0, in the post-sample period of 0.75, and is 0.30 in the pre-sample period. On average, the pre-sample results are 36% of the original sample, while the post-sample results are 78% of the original sample, consistent with overfitting in the original sample (but not pure spurious data mining), and inconsistent with informed trading causing a decline in performance after discovery.

We also compute the Sharpe ratios of value in each asset class relative to the Sharpe ratio of the original asset class in which the factor was discovered (U.S. equities). For value, the other asset classes show *greater* performance than the original U.S. equity sample (by 24%), providing another out-of-sample test that suggests the value factor was not overfitted to U.S. stocks and that the value premium is not smaller in other markets.

A.3 Momentum

The next graph (upper right) of Figure 1 repeats the analysis for momentum, where we define the subperiods using the original sample period of Jegadeesh and Titman (1993), which is January 1964 to December 1989. We find significant momentum premia out of sample, suggesting that the momentum premium is not a spurious accident, but we also find that momentum's performance is weaker out of sample: the pre-sample performance is 68.2% and the post-sample performance is 79.6% of the original sample. These results are consistent with overfitting biases, but not pure spurious data mining. However, for momentum, we fail to reject the null that the in- and out-of-sample performances are the same (*p*-value = 0.160). We also find no evidence that momentum is weaker in the post-sample period relative to the pre-sample period, which contradicts the idea that informed arbitrage activity drives down momentum's performance. This pattern is fairly consistent across asset classes. Relative to the original asset class for momentum – U.S. stocks – the other asset classes produce about 60% of the profits.[17]

A.4 Carry

The lower left graph of Figure 1 reports the results for the carry factor. The intuition behind the carry strategy in currencies originates in Meese and Rogoff (1983) and Fama (1984), who document a "forward premium puzzle" in currencies. The Meese and Rogoff (1983) sample is from March 1973 to June 1981 and the Fama (1984) sample from August 1973 to December 1982. We use the period from 1973 to 1982 to define the "original sample" for carry. Of course, for currencies this means we have no pre-sample period since exchange rates were pegged under Bretton Woods prior to 1973.[18]

We find that post-sample performance of currency carry is actually larger than its in-sample performance, suggesting neither data mining nor arbitrage activity impact its performance out of sample. Looking at the other asset classes, the pre- and post-sample efficacy of carry is about half that of the original sample, consistent with overfitting. A formal test of whether in- and out-of-sample performance of carry is the same is rejected (p-value = 0.004). However, there appears to be no difference between the pre- and post-sample performance of carry (p-value = 0.951), which is inconsistent with arbitrage degradation. The performance of carry outside of currencies, where it was originally discovered, is 42.2% larger, suggesting that carry is not overfitted to currencies.

A.5 Defensive

The lower right graph of Figure 1 shows the results for the defensive factor. To define the subperiods, we use the original sample period of Frazzini and Pedersen (2013) from 1960 to 2009. This is a debatable choice since Black (1972, 1992) originally looked at how leverage aversion might alter empirical tests of the CAPM, and Ang et al. (2006, 2009) showed that low volatility predicts stock returns. However, Frazzini and Pedersen (2013) establish the betting-against-beta (BAB) factor we use in this study and covers the longest sample period. For the defensive factor, both the pre- and post-samples outperform the original sample, which is inconsistent with lower out-of-sample performance due to data mining/overfitting or informed trading post-publication. Moreover, the post-sample period is greater than the pre-sample evidence, further contradicting the arbitrage story. Formal tests of in- versus out-of-sample performance, and pre- versus post-sample performance, fail

---

[17] Since we use a uniform measure of momentum, the past 12-month return on the asset skipping the most recent month, this may indicate some overfitting of the momentum measure to U.S. individual stocks. For example, skipping the most recent month for individual stocks avoids the one-month reversal effect of Jegadeesh (1990), which is not present for other, more liquid asset classes such as fixed income, currencies, and equity indices.

[18] Accominotti and Chambers (2014) find positive carry and momentum returns in currencies in the 1920s and 1930s that precede the fixed rate regime of Bretton-Woods, providing additional out of sample evidence.

to detect any significant differences. However, defensive strategies perform better in individual equities, where they were first discovered.

A.6 Multifactor Portfolio

The last graph in Figure 1 plots the results for a multifactor portfolio. Defining the subperiods is non-trivial since different factors have different sample periods of discovery and to make meaningful apples-to-apples comparisons, we need all four factors to be in the portfolio in every subperiod. We could take the union of original sample periods, which would be from 1960 to 2009, or their intersection, which would be from 1973 to 1981. There are pluses and minuses to using either as the in-sample period, since the former is likely too long a sample and would include factors both before and after their discovery, whereas the latter is too short a sample and missing significant parts of the original sample for most factors. As a compromise, we define the "original" in-sample period to be from 1960 to 1990, which covers the majority of the in-sample periods for all factors.

The results for the multifactor portfolios echo a summary of our previous findings, with the pre-sample Sharpe ratio about 64% of the original sample Sharpe ratio, and the post-sample Sharpe ratio about 69% of the original. A formal statistical test confirms a reliable reduction in performance out of sample (p-value = 0.009), but no difference in the pre- versus post-sample (p-value = 1.00).

Overall, we find that factor premia exist and are robust out of sample, and therefore not the result of spurious data mining. However, overfitting biases can account for about a 30% decline in out-of-sample efficacy of the factors, which is similar to the estimates Mclean and Pontiff (2016) get in their analysis of U.S. equity-only factors over a much shorter history. Contrary to Mclean and Pontiff (2016), we find little evidence that these factors were affected by informed trading after their publication. If anything, the post-sample efficacy of the factors is larger, and certainly no smaller, than the pre-sample performance.

III. Sources of Factor Premia

Having ruled out pure data mining as an explanation, we investigate a variety of risk-based and behavioral theories for the source of these premia and examine new evidence that may help shed light on these theories.

A. Common Variation

We begin by examining common variation across factors and asset classes. Panel A of Table 2 reports correlations (from monthly returns) across asset classes for a given factor. The first column shows the correlation of a value strategy in each asset class with a value strategy applied to all *other* asset classes, which on average is 0.15. The second column repeats this exercise for the momentum factor, where the average cross-asset correlation is 0.27. These results are consistent, in terms of sign and magnitude, with those from Asness, Moskowitz, and Pedersen (2013), but with an additional 50 years of out of sample data. For carry strategies, the correlations are much weaker across asset classes, consistent with Koijen, Moskowitz, Pedersen, and Vrugt (2018). Defensive strategies are positively 0.19 correlated across asset classes, which is novel to the literature.

Panel B of Table 2 reports the correlations of the factors within each asset class over the entire century. For U.S. stocks, value and momentum are strongly negatively correlated (-0.68), value and defensive are slightly negatively correlated (-0.17), and momentum and defensive are positively correlated (0.31). Similar results are obtained for international stocks, albeit over a more recent period. For non-equity asset classes, a consistent negative correlation between value and momentum is present (averaging -0.51), providing six decades of out-of-sample evidence for the ubiquitous negative correlation between value and momentum in other asset classes (Asness, Moskowitz, and Pedersen (2013)). Momentum and defensive factors are consistently positively correlated in each asset class. Value and defensive are consistently negatively correlated in each asset class, but the correlations are small. Carry is positively correlated to value in equity indices, fixed income, and currencies, but is negatively correlated to value in commodities. Carry and momentum are essentially uncorrelated in every asset class except commodities, where they are positively correlated (0.42).[19]

Table 3 reports time-series regressions of each factor's returns in each asset class on the other factor returns within the same asset class and the same factor's returns in all other asset classes over the full century of data. We also include an equity, bond, and commodity market index. The coefficients of each factor in each asset class on the same factor in all other asset classes are significantly positive, highlighting common variation for a given factor. However, we find positive and significant alphas for nearly every factor in every asset class after controlling for the other factors within that same asset class, the same factor in other asset classes, and the market portfolios in equities, bonds, and commodities, suggesting that only part of the expected return to each factor in each asset class is captured by this common variation.

B. Arbitrage Activity

[19] Figure 2 plots the time-series of pairwise correlations between the factors (across all asset classes) using rolling monthly return data over the prior 10 years from 1933 to 2018. Value and momentum are always negatively correlated over the sample period, ranging from a minimum correlation of -0.70 to a maximum of -0.18. Over the century, the correlation between value and carry ranges from -0.20 to 0.46 and for value and defensive varies from -0.44 to 0.31. Momentum and carry have correlations that range from -0.25 to 0.37, momentum and defensive from -0.13 to 0.55, and carry and defensive have correlations as low as -0.30 and as high as 0.51 over the century.

Although we do not find much support for arbitrage activity affecting the performance of these strategies, we can also look at whether correlations of the factors across asset classes, and across factors, changed over time after their discovery as another test of this theory. Arbitrage activity predicts higher correlations due to price pressure from trading (Lou and Polk (2015)), which should be greater post-discovery. Changing correlations can also provide another test of data mining. Linnainmaa and Roberts (2016) argue that data snooping biases can artificially lower correlations of factors in sample, since overfitting to get a high alpha or information ratio is achieved by maximizing means and minimizing correlations to existing factors. A purely data-mined factor that is just noise would not exhibit much correlation in- or out-of-sample to other asset classes, especially if overfitted to one asset class (the original one studied). To test these implications, we examine the correlations across factors and within a factor across asset classes in the pre-, original, and post-sample periods. Overfitting implies stronger correlations among factors out of sample, but weaker correlations across asset classes for a given factor. Informed trading, on the other hand, implies stronger correlations among factors and stronger correlations across asset classes for a given factor post-discovery. Finally, noisier data in the early part of the sample should imply lower correlations in the pre-sample period, which is another implication we can test.

The first graph in Figure 3 plots time-variation in the correlation *between* factors. There is little variation in the correlations across the subperiods, with a slight decrease in correlations post-sample. The bottom figure excludes currencies and international stocks since they have the shortest sample periods and thus may be limiting the analysis. The results are the same. There appears to be little evidence that correlations across factors are any higher in the out-of-sample periods relative to the in-sample period, and no evidence that correlations are higher post factor discovery.

The next two graphs of Figure 3 examine correlations across asset classes for a given factor. The first bars are the average correlation of the value factor in one asset class with value factors in all other asset classes, averaged across all pairwise value correlations. Average correlations across asset classes for value are lowest in the pre-sample period, higher in the in-sample period, and highest in the post-sample period. The same patterns are also found for momentum, but not for carry or defensive. These results suggest that value and momentum may have become more correlated across markets over time, possibly due to awareness of and active trading in these factors or to more market integration over time. However, the results are misleading because international stocks and currencies are missing from the pre-sample period due to shorter available histories. Since U.S. stocks and international stocks are highly correlated, the correlations in the pre-sample periods are likely to be biased downward relative to the correlations in the original and post-sample periods.

Confirming this conjecture, the bottom graph removes currencies and international stocks and finds no evidence of larger correlations among value, carry, or defensive strategies in any of the subperiods. Momentum has a slightly higher correlation across asset classes in the post-sample period, but the difference is small. The multifactor portfolio that is diversified across all four factors also shows no difference in average correlation across asset classes and markets. This also indicates that either arbitrage activity did not commence in a meaningful way post-publication or that such activity has little effect on correlations.

C. Economic Risks

Many asset pricing theories attempt to capture this common variation through systematic risks or state variables representing the changing investment opportunity set (Merton (1973)) in the economy. We examine whether factor returns relate to a variety of variables that capture economic activity and news to test a number of specific theories for the existence of factor premia. Attempts to link factors to economic risks have proven challenging due to limited time-series. Exploiting our much longer and broader sample, we examine more than 50 years of additional economic events across different markets to better identify these relationships and test previously documented relationships out of sample. The breadth of asset classes also helps reduce noise that may cloud these relationships.

C.1 Factor exposures to economic activity and news

Table 4 reports results from a time-series regression of each factor's returns over the last century on various economic measures. Panel A reports contemporaneous regressions between factor returns at time *t* and economic variables at time *t*. The first set of variables include a measure of illiquidity risk from Amihud (2014), the Baker and Wurgler (2008) sentiment index, and equity market volatility (realized volatility of the equal-weighted country indices, estimated over the prior 36 months). Illiquidity may capture the constraints on arbitrage activity that, when binding or loose, might amplify or dampen factor return premia. Sentiment is designed to capture investor behavior and optimism, which Baker and Wurgler (2008) argue and show is related to and predicts factor return premia. Market volatility could also capture arbitrage costs, but also proxies for risk and uncertainty in markets, which may impact factor returns.

The next set of variables capture macroeconomic activity in an attempt to link factor returns to macroeconomic models. One issue with this analysis is that there is little theoretical guidance on how factors should interact with the macroeconomy. For the aggregate stock or bond market we have solid intuition and theory for how economic growth or inflation is expected to impact their returns. But, what about market neutral, long-short factors like value, momentum, and carry? Should value returns rise or fall with various economic shocks? What about other factors? One set of theories posits that macroeconomic shocks represent states of the world investors wish to avoid or may pay to hedge. In this world, assets exposed to these shocks should carry a risk premium to compensate investors. However, since value, momentum, carry, and defensive all have positive abnormal returns and yet are lowly or even negatively correlated to each other, they must be exposed to different risks in this world. Absent strong theoretical predictions, we embark on an empirical exploration of factor exposure to macroeconomic events, recognizing that such an unguided exploration increases the dangers of data mining and fitting noise.

We use global GDP growth (real growth over the last year averaged over the U.S., UK, Germany, and Japan), global CPI inflation growth (growth over the last year in inflation averaged over the U.S., UK, Germany, and Japan), a tail risk indicator (if the developed equity market index is in the lower fifth percentile), a geopolitical risk index (from http://www.policyuncertainty.com/gpr.html), and three business cycle indicators. Specifically, we define periods into positive and negative growth based on the year-on-year GDP growth each quarter, and also define periods into "accelerating" and "decelerating" growth based on the change each quarter in the year-on-year GDP growth. The intersection of these two indicators creates four subperiods: contraction (negative growth and negative change in growth), recovery (negative growth and positive change in growth), expansion (positive growth and positive change in growth), and slowdown (positive growth and negative change in growth). Regime change is triggered by either GDP growth or changes in GDP growth hitting the +/-1 standard deviation bounds based on a 10-year rolling window. The trigger is designed to avoid frequent switches when growth or changes in growth dips into negative territory very briefly. For example, the transition from contraction to recovery would only occur if changes in growth are +1 standard deviation (relative to zero), and not just when they become positive. Unlike NBER recession and expansion dates, these subperiods are determined solely based on ex ante information. Table IA4 in the internet appendix provides more details on the construction of each of these variables and their data sources.

The macro variables seek to capture macroeconomic risks, state variables about the investment opportunity set, or proxies for the marginal utility of investors. While theory does not suggest how each factor might behave in different growth environments, there are theories (Lettau and Wachter (2007), Gormsen and Lazarus (2019)) that suggest different factors face different cash flow duration that may make them more or less sensitive to discount rate shocks, such as interest rates. For example, value and defensive factors may be short duration and hence may be impacted more by inflation shocks. The tail risk dummy seeks to capture rare disaster risks in the economy (Barro

(2006, 2009), Gabaix (2011), Tsai and Wachter (2015, 2016)) or pertaining to the factors themselves (Brunnermeier, Nagel, and Pedersen (2009), Daniel and Moskowitz (2016)). The geopolitical uncertainty measure is another global risk variable designed to capture policy uncertainty across a variety of issues globally, obtained from Baker, Bloom, and Davis (2015). To our knowledge, no one has linked policy uncertainty to factor premia.

An important aspect of some of the macroeconomic variables is their timing. Many macroeconomic variables are announced and reported after the actual quarter or month they pertain to. For example, the second quarter GDP number might only be announced in July (third quarter) of the same year. A question arises then as to whether we should match second quarter financial returns to the second quarter GDP number, which is what actual GDP was at that time, or whether we should match third quarter financial returns to the second quarter GDP number since the market learns about the second quarter growth only in the third quarter. The first choice measures the relationship between returns and actual economic activity. The second choice measures the relationship between returns and *news* of economic activity, which the market finds out later. In Panel A of Table 4 we make the first choice – the contemporaneous relationship between the factor returns and economic activity. In Panel B we make the second choice by lagging all macroeconomic variables by one period to capture their announcement lag, and the news associated with the economic activity.

The first column of Panel A of Table 4 reports the results for the value factor across all asset classes. Value loads positively and significantly on illiquidity risk, the Baker and Wurgler (2008) sentiment index, and the economy slowdown indicator. Value returns tend to be higher when illiquidity risk is high, sentiment is positive, and the economy is slowing down. The first two effects are broadly consistent with Asness, Moskowitz, and Pedersen (2013) and Baker and Wurgler (2008), respectively, over shorter sample periods. The rest of the variables show no significant effects on value returns. For momentum, nothing is significant except a negative coefficient on the slowdown dummy, although with a t-stat of only -2.20 and the multiple comparisons being made, this, too, is not very compelling. Momentum tends to exhibit the opposite-signed coefficients as value, because it is negatively correlated with value, but nothing significant emerges. Similarly, carry returns have no contemporaneous relation to any of the variables. The strongest effect appears to be the tail risk dummy (t-stat of -1.92), which is vaguely consistent with the carry crashes documented by Brunnermeier, Nagel, and Pedersen (2009). Defensive, like value, has higher returns when illiquidity risk and sentiment are high, but lower returns when inflation is high and during an expansion and slowdown in the economy. Defensive strategies, which are long low beta and short high beta, are contrarian in nature, so these contemporaneous relations are somewhat intuitive.

Overall, there is not much evidence that the factors vary in a meaningful way with the economic variables. The R-squares are small and the coefficients on the variables are mostly insignificant and oscillate sign. Even over the much longer sample and across six different asset classes, we fail to find any significant macroeconomic risk sensitivity of these factors. With all of the multiple tests being made, a Bonferroni-adjusted t-statistic suggests a t-stat of at least 3.5 is required for 5% significance. Under this criterion, none of the macroeconomic variables are significantly related to factor returns.

Panel B of Table 4 examines the same regressions using the same independent variables lagged an additional period (which can be a month or a quarter, depending on the frequency of the variable as detailed in Table IA4) to capture the news or announcement effect of these variables. For some variables, like illiquidity risk, volatility, and sentiment, lagging the variables represents "news" in the sense that it is the most recent information an investor could obtain in real-time about these variables.[20] For many of the macroeconomic variables, lagging ensures that the actual news was released by the portfolio formation date. The coefficients on these variables, therefore, represent predictive relationships; something we investigate in greater detail in the next section. The evidence for predictive relationships of economic news on the factor returns is even weaker than the evidence of contemporaneous economic activity, judging by the lower R-squares, and given the multiple tests being conducted, nothing is significant.

Another way to capture information in the economy is to examine the returns to a global market portfolio. Figure 4 plots time-varying exposures (correlations) of the multi-factor portfolio, diversified across value, momentum, carry, and defensive, to the returns of a market portfolio. We use three market proxies: the all-asset-class market portfolio, a global equity market portfolio, and a global fixed income portfolio, respectively. These market portfolios exhibit significant exposures to macroeconomic shocks, including the macroeconomic variables above. Hence, their returns should capture variation in economic news. As the figure shows, the average correlations to each of these market proxies is close to zero, consistent with random sampling variation. A formal test cannot reject that the correlations are zero.

Despite our long and broad sample of data that provides a rich set of macroeconomic movements and power to detect macroeconomic exposures, we do not find significant exposure of common asset pricing factors to economic activity or news. These results are broadly consistent with the conclusions reached in Griffin, Ji, and Martin (2003), Asness, Moskowitz, and Pedersen (2013)

[20] At least within a month, that is. We could obtain slightly more recent information by looking at these variables up until one day prior to the factor portfolio formation, but the small difference in time should not matter.

and are inconsistent with other studies (Chordia and Shivakumar (2005), Hodges, Hogan, Peterson, and Ang (2017)), all of whom study a much shorter sample period and equity-only factors.

C.2 Correlations in different economic environments

We also look at whether common variation among factors and asset classes varies with different macroeconomic environments. Figure 5 reports correlations computed separately for different economic environments: for the 20 percent worst and best months of global equity returns (using the MSCI index), the 20 percent worst and best months of global bond returns (using the Barclays Aggregate Bond Index), the 20 percent worst and best returns from a volatility-weighted average of all asset classes that includes stocks, bonds, and commodities, the highest and lowest 20 percent of months based on equity market volatility (realized volatility over the last 36 months), as well as during global recessions and expansions using the NBER's business cycle definitions applied to all developed markets in our sample.

The first chart plots the six pairwise correlations between factors across all asset classes, which shows little variation across economic environments. Despite anecdotal stories that correlations rise among factors during extreme, particularly negative events, we do not find that poor economic conditions change the correlation across factors. The second graph of Figure 5 reports average pairwise correlations across asset classes for a given factor. We find very little evidence that correlations across asset classes move in a meaningful way over different economic environments either, with one notable exception. During very low market volatility periods (lowest 20 percent), correlations across asset classes for a given factor are particularly low. Hence, precisely when volatility is of least concern, diversification benefits are greatest. From an investment perspective, an efficient portfolio would want these benefits to be greatest during the most turbulent times.

IV. Time-Varying Premia and Factor Timing

Another virtue of our long and broad sample is that it can provide a more powerful test of time-varying return premia and address the efficacy of factor timing. We examine the existence of time-varying factor premia through factor timing strategies with two motivations in mind. First, from a theoretical perspective factor timing can identify the conditional expected returns of factors and the optimal timing portfolio can inform us about the conditional stochastic discount factor in the economy (Haddad, Kozak, and Santosh (2018)).

Tests of time-varying premia and timing ability are inherently noisy and have low power due to the typically short sample periods of returns often analyzed. As a consequence, much debate exists on the efficacy of timing and identification of conditional risk premia. The additional 50+ years of data, across five more asset classes, provides a more powerful laboratory to detect factor timing as well as out-of-sample evidence on the robustness of previous timing studies. Since the same unconditional factor premia exist in other asset classes besides U.S. stocks, it is reasonable (and interesting) to ask whether conditional premia in other asset classes look similar and whether a unifying framework of conditional premia also applies across all asset classes.

We explore conditional factor returns and factor timing strategies through a variety of timing signals and methods, and expand the evidence on factor timing by looking across multiple factors in many asset classes over a much longer sample period. Timing studies typically fall into one of the following categories: 1) a single factor in a single asset class timed with a single predictor,[21] 2) multiple factors in a single asset class timed with one or many predictors,[22] and 3) a single factor in multiple asset classes timed with a related predictor.[23] We examine multiple factors across multiple asset classes and attempt to synthesize the results on factor timing across factors and asset classes by looking at them simultaneously.

To assess the efficacy of factor timing and make comparisons across signals and methods, we evaluate all factor timing models using a trading strategy. This serves several purposes. First, it puts all factor timing variables and methods on equal footing by comparing them based on out-of-sample return performance. Second, returns per dollar invested provide a measure of the economic magnitude of timing. Third, timing strategy returns allow us to assess the marginal benefit of factor timing by examining the added contribution of tactical timing to an investor's optimal static factor portfolio. Finally, focusing on the returns to an investment strategy circumvents many of the problems introduced by other timing metrics such as R-squares, where it is difficult to assess economic significance and where degrees of freedom, precision of the parameters, and other statistical biases matter (Stambaugh (1999)). The returns to an investment strategy encapsulate all of these issues because poor model estimation will generate poor out-of-sample performance.

A. Value spread timing

We begin with the simplest and best known factor-timing signal: value spreads. Valuation ratios have long been used to forecast equity market returns, dating back to Fama and French (1988) and

---

[21] For example, dividend yield (Fama and French (1988)) or CAPE (cyclically-adjusted P/E ratio in Campbell and Shiller (1998)) for the aggregate U.S. equity market, valuation spreads for the value factor (Asness, Friedman, Krail, and Liew (2000), Cohen, Polk, and Vuolteenaho (2003)), forward rates for bonds (Fama and Bliss (1987)), or more rarely timed with multiple predictors (Asness, Ilmanen, and Maloney (2017) for U.S. equity markets).

[22] See Cochrane and Piazzesi (2005), Cieslak and Povala (2015), and Brooks and Moskowitz (2018) for various factors in bond markets, and Stambaugh et al. (2012), Greenwood and Hanson (2012), Akbas et al. (2015), Kelly and Gupta (2018), Arnott et al. (2019) for multiple factors in individual equites.

[23] Baba, Boons, and Tamoni (2017), Koijen, Moskowitz, Pedersen, and Vrugt (2018), Asness, Liew, Pedersen, and Thapar (2018).

Campbell and Shiller (1988). The basic idea is to look for a valuation metric that identifies when the market portfolio looks “expensive” or “cheap.” A natural candidate for this metric is a value measure like book-to-price or CAPE. Aggregating the value measure across all stocks in the market provides a valuation metric for the market as a whole, which may indicate when expected returns are high or low and hence when an investor should be long or short the market. Variation in the expected return can be driven by time-varying risk premia or mispricing from market sentiment.

This idea has been applied to other portfolios as well (Vuolteenaho (2002), Cohen, Polk, and Vuolteenaho (2003), Lochstoer and Tetlock (2016), and Asness, Chandra, Ilmanen, and Israel (2017)). For example, applying this concept to a market-neutral factor that is long some assets and short others based on a characteristic (like value, momentum, carry, or defensive) by summing up the average valuations of the long positions versus the average valuations of the short positions, should be informative about whether the factor itself is trading cheaply. Rather than take the difference between the longs and shorts, we take the ratio or log difference between the average valuation of the longs and shorts, as most of the literature does. From a theoretical standpoint, it is not obvious whether differences or ratios is more appropriate, but the ratio is not contaminated by rising price levels over time that can distort the differences in valuations. On the other hand, ratios can be problematic if the denominator is small or negative. We use ratios of valuations for individual stock factors (following the literature and because the small or negative denominator problem is non-existent), but use differences in valuation metrics for non-equity class factors, where small or negative valuations can frequently occur. Results are not sensitive to using either method.

If valuations of the factors are indicative of time-varying factor premia (for risk or behavioral reasons), then we expect a positive relationship between value and future returns. The conditional expected return of the factor, conditional on its valuation metric, is expected to be positive.

Applying this concept to factors is intuitive, but not without issues. Asness (2016) argues that factor timing may be more difficult than market timing due to long-short factors often having higher turnover than the market, which makes long-run predictability more challenging. In addition, valuation-based timing signals, which are contrarian in nature, are often already embedded in the factor itself (e.g., the value factor) and hence may merely increase exposure to the static value factor. We investigate this latter point and find evidence consistent with that notion.

Asness, Friedman, Krail, and Liew (2000), and Cohen, Polk, and Vuolteenaho (2003) apply this concept to the value factor in U.S. stocks, and find that value spreads significantly predict expected returns. We begin our analysis by looking at the value spread’s ability to time the same U.S. equity value factor over a longer sample period from 1926 to 2018. We then apply the same concept of value spread timing to other factors in U.S. equities, namely momentum and defensive factors, over the same century of data. Finally, we apply value timing in other asset classes.

We start with a straightforward application of value timing. After computing the value spread of a factor, we standardize it over time to have mean zero and unit variance (a *z*-score). We then time the factor by increasing the weight on the factor in proportion to its *z*-score at each point in time, where positive *z*-scores assign positive weight to the factor and negative scores short the factor, with the absolute value of the *z*-score determining how many total dollars to be long or short. The *z*-score represents how much risk the strategy takes in the factor at each date by scaling up or down the dollar exposure to the factor, determined by how cheap or expensive the factor looks relative to history. To reduce the influence of outliers, we cap the weights at +2, -2. Also, to ensure no look-ahead bias, the *z*-scores are estimated using an expanding window from the first return observation in the sample to date *t*-1 and then applied to factor returns at time *t*, requiring at least 10 years of history. Since the value spreads are not distributed normally, and can experience extreme values, we use the historical median and the mean absolute deviation from the median to define our "*z*-scores." We will show a variety of alternative methods for timing based on value spreads below.

D.1 U.S. equity factors

The first panel of Table 5 reports results for value timing of U.S. equity factors. For comparison, we report in the first column the annualized Sharpe ratio and *t*-stat of the mean return (in parentheses) of the static factor that contains no timing. These numbers are slightly different than those in Table 1 due to the timing strategies skipping the first 10 years of the sample. The second column reports the annual Sharpe ratios of the raw timing strategy for each factor, using the *z*-score of the value spread of each factor to time it each period. The annualized Sharpe ratio of the value spread timing strategy for the U.S. equity value factor is only 0.17 with a *t*-stat of the mean return of 1.56, suggesting that value spread timing of the value factor does not yield significant profits. This result is weaker than Cohen, Polk, and Vuolteenaho (2003) and Asness, Friedman, Krail, and Liew (2000) find in their samples using slightly different timing methodologies. We investigate below the robustness of value spread timing to methodology and sample period.

Asness, Chandra, Ilmanen, and Israel (2017) and Asness, Liew, Pedersen, and Thapar (2018) show that value spread timing strategies implicitly load on the static value strategy itself. A value spread timing strategy that overweights a factor when value spreads are large and underweights when they are small may simply increase exposure to a static value factor. Buying factors that look cheap and shorting those that look expensive exposes an investor to static value, which we know carries a positive return premium. To account for this, we regress the returns of the value spread timing strategy of each factor to the static value factor in a univariate regression and compute its alpha. To compare to the Sharpe ratios reported in the other columns, we report the annualized information ratio of the alpha (alpha per unit of residual standard deviation). Information ratios are a better metric to evaluate the efficacy of timing strategies because the timing of long-short factors can exhibit substantial variation in leverage, which will arbitrarily vary the size of alphas across methodologies and factors. The information ratio puts everything on the same scale for ease of comparison. These two points – adjusting for underlying static factor exposure and looking at information ratios instead of alphas – we argue are better ways to assess and compare timing models.

As the third column reports, the information ratio for value timing the value factor is negative and insignificant (*t*-stat of -0.62). Value spread timing induces positive exposure to a static value factor, which when accounted for lowers the returns to a value timing strategy. The last column reports the information ratio after accounting for other static factors, representing pure timing profits after hedging out the returns to all unconditional factors. The timing alpha is statistically indistinguishable from zero.

The remaining rows report results for value timing the momentum and defensive factors. Value spread timing of the momentum factor results in a significantly positive 0.25 Sharpe ratio with a *t*-stat of 2.3 for the mean return. Regressing these timing returns on the static momentum factor, produces even larger profits, with an information ratio of 0.59 and a *t*-stat of 5.31. Value spread timing of momentum is particularly profitable after hedging out static momentum exposure. However, regressing these timing returns on all static factors, the information ratio drops to 0.40, primarily due to static value exposure that value spread timing induces, which is a particularly valuable diversifier to a momentum strategy. Hence, part of the additional performance to timing momentum using value spreads could have been replicated by simply adding static value exposure. For the defensive factor, we similarly find strong evidence for the efficacy of value spread timing. The raw Sharpe ratio is 0.75 (*t*-stat = 6.62), and the information ratios of the univariate and multivariate alphas are 0.47 and 0.43 (*t*-stats of 4.16 and 3.77), respectively, which are significantly lower after adjusting for the static defensive factor, which itself has an impressive Sharpe ratio of 0.73. The last row reports results for value spread timing on the multifactor portfolio diversified across value, momentum, and defensive factors, which shows significant timing profits with an information ratio with respect to all static factors of 0.33 (*t*-stat = 3.04).

Looking over the last century, we find little evidence of value spread timing for the value factor , but find significant evidence for value timing of momentum and defensive factors. These results pertain exclusively to U.S. equity factors, which is the testing ground for nearly every timing study.

D.2 Other asset markets

With our broad sample, we can examine value spread timing in other asset classes. Baba, Boons, and Tamoni (2017) show evidence that returns to value strategies in individual stocks, currencies, global equity indices, global government bonds, and commodities are predictable by the value spread. Asness, Liew, Pedersen, and Thapar (2018) show that "deep value" periods, where the valuation spread between cheap and expensive securities is in the extreme 20th percentile relative to its history, predicts value factor returns in global individual stocks, equity index futures, currencies, and global bonds. Brooks and Moskowitz (2018) show that value spreads predict the returns of global bond portfolios and Haddad, Kozak, and Santosh (2018) find evidence of timing bond and currency factors using value spreads.

The value spread timing strategies in other asset classes use the value characteristic for each asset class. For international equities, value spread timing produces negative returns relative to the static factors (although the sample period is short, beginning only in 1984). For global country indices, we find no evidence for value spread timing for any of the factors over the last century. For commodities, the results are mixed. There is positive and significant evidence of value spread timing for the commodity value factor, but insignificant timing profits for momentum, negative timing profits for carry, and no profits for multifactor timing in commodities. For global bonds, there is consistently negative factor timing, and for currencies there is positive and significant value spread timing for the carry factor only.

Finally, we look at value spread factor timing across all asset classes simultaneously. We combine the timing strategies across all asset classes using an equal risk combination of the asset classes, where each asset class is weighted in proportion to its inverse volatility (estimated over the last 36 months). Across all asset classes, we find positive value timing for the value, momentum, and defensive factors, and negative evidence for carry, when looking at raw returns. However, after adjusting for static factor exposure, only for defensive factors does value spread timing appear to deliver significantly positive alpha, and this is entirely driven by U.S. individual stocks. All of the other factors have insignificant timing alphas from value spreads. The last row of the all-asset-class panel reports the multifactor timing results, outlined in a frame box in Table 5. Timing all factors across all asset classes using the value spread of each factor in each asset class, results in positive and significant profits, even after accounting for static factor exposure across all factors and asset classes. The information ratio of the multifactor, all-asset-class value spread timing strategy is 0.28 with a *t*-statistic of 2.58. These gains are orthogonal and additive to the static multifactor portfolio across all asset classes, which has a Sharpe ratio of 1.64. However, the significance of the timing profits is entirely driven by U.S. individual stock factors. As the next panel shows, when we remove individual stock factors, all of the multivariate alphas/information ratios are insignificant. Hence, value spread timing of factors does not appear to be robust in other asset markets for any factor. Since individual stock factors are the most heavily mined with regard to timing, these results question the robustness of factor timing. Given the multiple comparisons being made, the evidence does not support the efficacy of value spread timing for factors.[24]

B. Other timing methodologies

The evidence in Table 5 suggests that value spread timing of factors delivers weak results when applied over the full century of data and across six different asset classes. However, the evidence in Table 5 examines only one possible way to use value spreads to time factors – the z-score timing using an expanding window of historical data to estimate the z-score parameters. We now consider other ways to use the information in value spreads to time factors. For instance, one could run a predictive regression of factor returns on lagged value spreads to measure the empirical relation between them and use that to time the factors. The regression can also impose economic restrictions on the coefficients as suggested by Campbell and Thompson (2007), who use economic constraints from the present value formula to better predict equity market returns out of sample. We also explore the method of Haddad, Kozak, and Santosh (2018) who impose a no near-arbitrage condition in the spirit of Ross (1976) to extract dominant principal components (PCs) of factors and apply value spreads to the PCs to time factors. Some studies also use in-sample moments to determine the timing parameters in order to reduce noise in the estimates, but which is not fully implementable in real time and may generate a look-ahead bias. We explore in this section various methodological choices for factor timing using value spreads as the timing signal, and then apply these methodologies to other timing signals in the next section.

The analysis of all of these different methodologies serves several purposes. First, by analyzing a bevy of timing models, we assess the robustness of the value spreads' ability to predict the returns of factors and how sensitive the timing results (or lack thereof) are to different specifications. Second, this exercise also serves as a grand specification search for the best way to extract predictive content from value spreads for factor returns. While this search raises the specter of data mining,

[24]Among the 26 different value spread timing strategies we examine across six asset classes and factors, less than 20% produce positive and significant alphas individually, and among the non-U.S. stock factors, less than 10% individually produce significant results when comparing to the standard t-statistic of 1.96 with 5% significance level. Taking into account the multiple comparisons, the Bonferroni-adjusted t-statistic is 3.1, which implies less than 10% of strategies produce positive profits and 0% of non-stock strategies produce positive returns.

given our earlier lack of findings, our aim is to show how robust those lackluster results are. If a massive search for value spread timing turns up little, then we feel more confident that the shortage of evidence is real and not unlucky. Put differently, if we fail to find anything significant despite intensive mining for a methodology to predict factor returns, then it is likely there is nothing to find. Finally, since many methodologies have been applied in the literature, this analysis serves as a robustness test of various studies using our longer and broader sample.

Table 6 reports value spread timing results for 19 different timing methodologies, which when applied to each factor in each asset class results in more than 500 timing strategies. To keep reporting manageable, rather than report 19 variants of Table 5, we report in Table 6 the timing results for the multifactor multi-asset class timing strategy, which summarizes the findings across all factors and asset classes. This strategy pertains to the framed box in Table 5 for reference. The first row of Table 6 repeats the results of this strategy for comparison. Table 6 lists the timing methodology used (e.g., *z*-score), whether the parameters of the model are estimated in- or out-of-sample, and any restrictions placed on the timing methodology. In the case of the *z*-score methodology used in Table 5, the parameters are estimated out of sample using an expanding historical window, are allowed to vary by asset, but impose an economic sign restriction in that value spreads are assumed to *positively* predict future expected returns, where we cap the weights on the timing signal to +2, -2. The table reports the annualized raw Sharpe ratio and the multivariate alpha information ratio of the timing strategy (we omit the univariate alpha information ratio for brevity).

The second row of Table 6 reports results for value spread timing using the same *z*-score methodology but instead of using completely out-of-sample data to estimate the parameters, we use the full sample of data. This type of methodology is often used in timing studies under the justification that a clearer picture of the distribution of value spreads is obtained from the full sample. However, this methodology is not implementable in real time since the researcher is using information he/she would not have had at the time of portfolio construction. Stated differently, using in-sample parameters may introduce a potential look-ahead bias. As the second row of Table 6 indicates, the results using the full sample moments are consistently better. The Sharpe and information ratios of the strategy improve to 0.42 and 0.33, respectively.

The next set of timing results employ a regression methodology to determine the relation between value spreads and conditional factor returns. We first use the expanding historical window of data to estimate a regression of returns in month t of each factor on the value spread of that factor at t-1. This exercise is wholly out of sample, as the regression estimates at each point in time only use data from time 1 to t-1 in the expanding historical sample. The product of the coefficient and the time *t* value spread provides the timing signal. We initially place no restrictions on the estimated coefficients, allowing them to vary by asset and by factor, including their sign. Although value spreads theoretically should positively predict returns for all factors, if the regression coefficient indicates a negative relation between value spreads and factor returns, we use that estimated negative coefficient. In other specifications below, we impose an economic sign restriction on the coefficients to be positive, consistent with theory.

The third row of Table 6 reports the results from this unrestricted out-of-sample regression timing strategy which is long/short a factor in proportion to its predicted conditional expected return based on the regression. The raw Sharpe ratio of this strategy is an impressive 1.36 per annum. However, the bulk of this performance is driven by static exposure to the underlying factors, which the regression picks up due to the large unconditional returns of the factors. Adjusting for static exposure to all factors, the multivariate alpha information ratio is 0.28.

The next set of results are from the same methodology, but rather than use the estimated regression coefficient to time the factors, we convert the regression coefficients to a *z*-score to scale them through time and across factors for comparison. Again, to keep the exercise purely out of sample and implementable in real time, the *z*-scores are determined using only the expanding window of historical data. As the fourth row of Table 6 shows, the Sharpe ratio from using the *z*-score of the regression coefficients is 0.27, indicating that rescaling the regression coefficients by the *z*-score eliminates the timing strategy's average exposure to the static factors. Consistent with this assessment, the multivariate alpha information ratio is no different than the raw Sharpe ratio at 0.29.

The fifth row of Table 6 adds another twist to the timing strategy by also imposing the +2 and -2 caps on the *z*-scores of the regression coefficients. As the table shows, the results are very similar, though there is some small improvement in the multivariate alpha information ratio to 0.35. This improvement is consistent with the caps reducing the influence of outliers on parameter estimation that aid predictability out of sample.

The sixth row of Table 6 imposes an economic restriction on the regression coefficients, requiring them to be positive. Economic intuition dictates that value spreads should positively predict returns. This economic restriction is similar to the one imposed by Campbell and Thompson (2007) when looking at timing the equity market with CAPE, where they find that economic restrictions improve out-of-sample timing performance. We find that economic restrictions on the value spread coefficients improve out-of-sample performance: the multivariate alpha information ratio improves to 0.41. The economic restriction binds 29% of the time.[25] Improvements from this specification are consistent across asset classes. Table IA5 in the internet appendix repeats Table 5 using this specification to show results by asset class, where 18 of the 26 timing strategies show out-of-sample improvement from imposing economic restrictions.

In the seventh row of Table 6 we repeat the timing specification without the economic restriction (same as row five), but estimate the regression coefficients and the *z*-scores using the full sample of data. The exercise shows how much look-ahead bias may or may not contribute to timing results when estimating model parameters in sample. This strategy also represents an upper bound on the timing performance from this specification knowing the full sample relation between value spreads and factor returns as well as the full sample distribution of value spreads. The Sharpe ratio of this timing strategy is expectedly larger at 0.42, compared to the 0.23 out-of-sample parameterization, but only moderately larger than the 0.31 Sharpe ratio we obtain by imposing economic restrictions on the out-of-sample exercise. However, the multivariate alpha information ratio is smaller for this strategy that employs full sample estimates, because the full sample timing strategy loads more on the static underlying factors (0.33 versus 0.41 information ratio). This result indicates that the economic restriction on value spreads produces better results than using the full sample regression coefficients that might have the wrong sign – an endorsement for using economic theory.

The next five rows (8 through 12) of Table 6 show results where we impose the same coefficient for every asset class for a given factor, allowing the relation between value spreads and future returns to vary by factor, but not by asset class. This restriction reduces the number of parameters to be estimated, thereby increasing the power of our tests, and the restriction may make economic sense if factors behave in the same manner across asset classes. The results show that this restriction leads to slightly smaller timing strategy profits, where information ratios are consistently 2 to 14 units smaller (that is, 2 to 14 basis points per unit of volatility smaller). Hence, allowing value spreads to time factors differently across asset classes yields slightly stronger out-of-sample return predictability, despite the additional degrees of freedom consumed.

The next five rows (13 through 17) of Table 6 impose an even tighter, and less economically appealing, constraint that the value spread coefficient be the same for all asset classes *and* for all factors. This is equivalent to estimating one pooled regression across all factors and asset classes. This specification has the fewest parameters to estimate. However, the assumption that value spreads

[25] The constraint binds for value 33% of the time, for momentum 31% of the time, for carry 50% of the time, and for defensive 1% of the time. These numbers are consistent with the strength of the simple *z*-score value spread timing across factors in Table 5.

relate to future returns in exactly the same way for every asset and factor seems less palatable and is inconsistent with our previous evidence. Table 6 shows that the pooled value spread timing regressions generally perform worse out of sample than the other specifications that allow for coefficients to differ by factor and by asset class. These results are consistent with meaningful economic variation across factors and asset classes being captured by value spreads, and suggest that the more flexible specification leads to a better out-of-sample performance.

Finally, we examine one other timing methodology using value spreads recently suggested by Haddad, Kozak, and Santosh (2018), who derive, under the absence of near-arbitrage conditions (Ross (1976)), a simple approximation for the maximum Sharpe ratio factor timing portfolio. We apply their methodology to our factors and asset classes.[26] Haddad, Kozak, and Santosh (2018) estimate the PCs and regressions in sample over their full sample period. We replicate their methodology in-sample over our full period, but also conduct an out-of-sample application of their methodology by estimating PCs and factor loadings over the expanding historical window of data.[27]

The last two rows of Table 6 report the results from the timing strategy of Haddad, Kozak, and Santosh (2018). The in-sample timing strategy produces significant profits with an information ratio relative to the static underlying factors of 0.45, which is consistent with the findings in Haddad, Kozak, and Santosh (2018) over their shorter sample period of U.S. equity factors. The out-of-sample timing results are meager, however, generating only a 0.13 information ratio, possibly indicating significant look-ahead bias when using full sample estimates of PCs and factor loadings. Moreover, the out-of-sample performance of this timing strategy is significantly worse than the regression-based approaches. Based on this evidence, and for simplicity and brevity, we focus on the regression-based

---

[26] Specifically, they extract the first two principal components (PCs) from a set of 50 long-short U.S. equity factors from 1963 to 2017 and use the average book-to-market ratio for each extracted PC as a measure of its valuation. They then run a predictive regression of each PC's future returns on its lagged valuation measure, and using the loadings of each factor on the PCs, they back out the implied factor-timing signals for each individual factor.

[27] It is important that the variation across factors and asset classes can be decomposed into a few dominant components as the procedure suggested by Haddad, Kozak, and Santosh (2018) assumes that factor risk premia are related to time-varying exposures to common risk and where quasi-arbitrage opportunities do not exist. The first two PCs from our four factors across the six asset classes capture 19.1% and 9.3% of the variation, respectively, among the 20 factor-by-asset class portfolios. These results are similar to the fraction of variation Haddad, Kozak, and Santosh (2018) explain among their 50 factors with the first two PCs (they find that PC1 explains 19% and PC2 explains 17% of the variation among their equity factors). The results are similar in our sample whether using PCs extracted from the full in-sample covariance matrix or the rolling expanding out of sample covariance matrix, and are robust to various subperiods.Table IA6 in the internet appendix reports results from PCA for different subsamples of our data. The first PC is intuitive, it is long momentum and short value (or vice versa) across all asset classes, with near zero weight on carry and defensive strategies. This is consistent with the strong negative ubiquitous correlation between value and momentum factors in every asset class as shown in Table 2 and Figure 3 and the close to zero correlations among the other factors. The second PC is less intuitive. Figure IA2 in the internet appendix plots the exposures of each factor portfolio to PC1 and PC2. The exposures are also stable over time.

timing methodology for the remainder of the paper, but provide results to other timing methodologies, including the Haddad, Kozak, and Santosh (2018) strategy, in the internet appendix.

C. Other timing signals

While value may be the most prominent timing signal in the literature, many other variables have been used to time factors. For example, Gupta and Kelly (2018), Ehsani and Linnainmaa (2017), and Arnott, et al. (2019) use factor momentum, or the past return on the factor itself, to time factors.

Similar in spirit to value timing, one can also use the spread in the factor signal to time factors. The idea is that when the difference or spread in characteristics between the average long and short positions is particularly wide, this may be when the factor's return premium is expected to be larger. Conversely, if there is little spread in characteristics, then expected returns should be feeble. The factor spread is a measure of the dispersion in signals across assets, which indicates how big a bet an investor can take on a particular factor characteristic. For the value factor, the "factor spread" is the same as the value spread used earlier. For other factors, the factor spread represents the momentum of the momentum factor (which is equivalent to factor momentum), the carry of the carry factor (as in Koijen, Moskowitz, Pedersen, and Vrugt (2018)), and the beta spread of the defensive factor.

We also look at five year reversals as a timing signal, which is the negative of the past five-year return on the factor, which can be viewed as a form of value timing (following the results of DeBondt and Thaler (1985) and Fama and French (1996)) or long-term factor momentum. In addition, we examine volatility timing following Moreira and Muir (2016) using the inverse of the standard deviation and variance of each factor to time it. We also examine a host of business cycle and macroeconomic variables to time factors using the business cycle indicators, GDP growth, and inflation growth variables from the previous section. We also add two general market timing variables – CAPE and the lagged realized volatility of the market, which we dub "VIX" for brevity – to see if general market conditions can predict factors returns.

In total, we analyze 11 timing signals across the 19 different specifications, for all 20 factor-by-asset class long-short portfolios plus the six multifactor and multi-asset portfolios, creating $11 \times 19 \times 26 = 5{,}434$ timing strategies. As stated previously, the aim of this large specification search is to test the robustness of various factor timing methodologies and the consistency of their performance across factors and asset classes. However, any significant results must be balanced against the number of comparisons being made. In this case, with 5,434 timing strategies, a Bonferroni correction to the normal 5% significance level results in a 0.09% threshold. In addition, we run a "full model timing" specification that uses all 11 timing signals simultaneously to combine their information. Rather than report all of these timing strategies, we focus on six particular specifications from Table 6, highlighted in bold: (1), (5), (6), (7), (10), and (11). Specification (1) is the simple z-score timing methodology we started with. Specification (5) is the regression methodology allowing the timing coefficient to vary across assets and factors using only out-of-sample estimates for all parameters. Specification (6) is the same as (5), but imposes a sign constraint on the timing coefficient. For value spreads that constraint is a non-negative coefficient. For other timing signals the sign constraint varies according to economic theory. Factor momentum, factor spread, inverse volatility and variance, and CAPE, should have positive signs. Five-year reversal and realized equity volatility ("VIX") should have negative signs. For the business cycle and macro variables, there is no economic sign prediction (since there is no theory predicting the sign), so we do not impose a constraint for these timing variables. Specification (7) uses in-sample estimates of the parameters over the full sample to estimate the regressions and imposes no constraints on the parameters, which introduces some look-ahead bias. This specification provides a non-implementable benchmark of how good timing would be with perfect foresight of future input of the model (but not of future asset returns). Specifications (10) and (11) are the same as (5) and (6) except we force the timing coefficients to be the same across asset classes for a given factor. These specifications are chosen based on the results from Table 5 to see if the same specifications and economic constraints that improve value timing have similar effects for other timing variables.

Figure 6 summarizes the results for all 11 factor timing variables across the six timing model specifications for the multifactor, across-all-asset-classes timing strategies. The figure displays the annualized information ratio of the multivariate alpha of the timing strategy to the static underlying factors. The first six bars on the graph pertain to value spread timing and repeat the numbers from the last column of Table 6 for rows (1), (5), (6), (10), (11), and (7), respectively. For value spread timing, we see again the out-of-sample forecasting benefit from imposing economic restrictions on the coefficients in specifications (6) and (11). However, none of these results passes the statistical threshold for significance, after accounting for the multiple tests.

C.1 Factor momentum

Looking at the other timing variables, we find that factor momentum (past 12-month return on the factor) delivers relatively weak timing performance. The simple z-score timing strategy for factor momentum fails to produce significantly positive profits and generates a negative alpha with respect to the static factors. This result is at odds with those in Arnott et al. (2019) and Gupta and Kelly (2018). The main differences between our study and theirs are 1) they focus on several dozen long-short U.S. equity factors, while we examine four factors across six asset classes and 2) we have an additional 50 years of data. Restricting our analysis to roughly the same sample period as these other studies, and focusing exclusively on U.S. equities, we find similarly large and positive factor momentum timing profits. However, looking at all asset classes over the same recent time period, we find much weaker results, and looking at U.S. equity factors only over the full sample period going back a century, we also find weaker results. This evidence suggests that factor momentum is weaker out of sample.[28]

Other timing specifications applied to factor momentum, such as the regression-based approach appear to deliver more positive results when the regression coefficients are unconstrained, although the alphas are still negligible. However, when imposing economic restrictions on the coefficients – in this case, forcing a positive relation between factor momentum and future returns – the out-of-sample results look worse. Finally, and relatedly, the only significantly positive profits we find for factor momentum are when we allow the regression coefficients to be estimated in-sample and unconstrained. In essence, this allows negative coefficients in sample to predict returns, which is the opposite of what factor momentum implies, but happens to work for some factors and asset classes in our century-long sample. The positive sign constraint for 12-month factor momentum is binding 38.4% of the time. Table IA7 in the internet appendix repeats Table 6 using factor momentum as the timing signal, which reports all 19 specifications for factor momentum to see if any specification provides stronger and significant results. None of the out of sample factor momentum timing strategies yield significant profits and many have negative returns. These results challenge the robustness of factor momentum as a timing signal, which across many specifications and methodologies, fails to deliver consistently positive timing results.

C.2 Value and momentum timing

Before moving to other timing signals, we investigate briefly if a combination of value and momentum timing can improve results. This idea is motivated by the strong negative correlation between value and momentum in the cross-section of assets (within and across markets as shown by Asness, Moskowitz, and Pedersen (2013)), that makes their combination a powerful investment tool. Table IA8 in the internet appendix repeats Table 6 using both value spreads and factor momentum to time the factors. As the table shows, there is not much improvement from combining value and

[28] To be fair, we also focus on the past 12-month factor momentum return signal, while the largest signal found in Arnott et al. (2019) and Gupta and Kelly (2018) is a shorter-term 1-month return signal. In addition, we skip an extra month in our signal, so that the past return is from month *t*-13 to *t*-2, which makes the results even weaker. Our claims on the robustness of factor momentum pertain only to this longer signal. That said, faster signals involve higher turnover and trading costs and will be costlier to implement. In addition, we also find that factor momentum in general, even at the 1-month level, is weaker in the pre-1963 period and is weaker for non-equity asset classes.

momentum timing. The results are mostly better than momentum timing alone, because there is some positive value timing evidence, but the findings are often worse than just value timing alone.

C.3 Factor spread

The next set of results in Figure 6 show that factor spreads are significantly weaker than value spread timing. Since factor spread timing includes value spreads as one of the signals, this suggests that the characteristics of the other factors (momentum, carry, and defensive) are not very useful in timing factors, and detract from value spreads.

C.4 Five-year reversals

Consistent with value spreads producing mildly positive timing results, five-year return reversals, often used as a "poor man's" measure of value because the past five year's returns often indicate which assets look expensive or cheap today, also produces mildly positive timing profits. Notably, the out-of-sample timing profits to five-year reversals are lower than they are using value spreads for every specification except for the in-sample regression results, which tend to overfit the sample.

C.5 Inverse volatility and variance

The next two sets of results pertain to inverse volatility and inverse variance as factor timing signals. Moreira and Muir (2017) find that inverse volatility and variance predict conditional Sharpe ratios to U.S. equity factors, including the market, value, momentum, and defensive factors, as well as the currency carry factor, under a variety of volatility estimates.[29] We examine how well inverse volatility and variance predict the Sharpe and information ratios of our factors across the six asset classes over the century. We measure realized volatility and variance of the factor's returns over the prior 36 months.[30] We then use these volatility and variance measures as timing signals. Figure 6 shows that the simple z-score timing model, which is closest to the method Moreira and Muir (2017) use, produces sizeable timing alphas, with information ratios above 0.50. The regression models for timing based on inverse volatility do not fare as well, although placing economic restrictions on the coefficients helps. The efficacy of inverse volatility timing appears to be sensitive to methodology, generating *negative* timing alphas out of sample when the regression coefficients are unconstrained.

C.6 Business cycle and macroeconomic timing

The next set of results focus on business cycle and macroeconomic timing signals. We use the business cycle variables – contraction, recovery, expansion, and slowdown indicators – described

---

[29] In a more narrow setting pertaining exclusively to the momentum factor, Barroso and Santa-Clara (2015) show effectiveness in timing U.S. equity momentum by the inverse of volatility and Daniel and Moskowitz (2016) show that a dynamic momentum strategy based on its forecasted conditional Sharpe ratio produces stronger performance and avoids the crashes associated with momentum.

[30] As in Moreira and Muir (2017), more sophisticated volatility and variance estimates lead to similar results.

earlier to time the factors. One issue, of course, is that theory provides no guidance on how these indicators should be related, if at all, to future returns. This makes timing based on macroeconomic variables particularly challenging because you have to be right twice – you have to first know what the factor exposures to macroeconomic indicators are, and secondly you have to be able to forecast the macroeconomic event. The latter is already enormously challenging. Put differently, if one is able to forecast macroeconomic growth, a much simpler and more effective trade would be to time the aggregate stock market, where we know what the predicted sign of the impact of growth will be. A large literature has documented how difficult this task is, so it would seem even more heroic to be able to use macro-timing for factors, where we are not sure what the predicted sign should be.

Since there is no ex ante prediction or economic intuition for whether economic growth should forecast the factors with any particular sign, we cannot run the simple *z*-score timing methodology. Instead, we focus on the regression methodologies, although again since there are no economic sign restrictions we can place on the coefficients, specifications (5) and (6) will be equivalent, as will specifications (10) and (11). Figure 6 shows that there is some timing alpha from using these business cycle indicators. Moreover, using the full sample regression to estimate the business cycle coefficients (e.g., using the results from Panel B of Table 3) the timing results look even better. This further highlights the danger of using in-sample parameters for timing, as the look-ahead bias appears substantial. In the case of business cycle variables, where there is no economic guidance on the sign of their relationship to factors, this is especially worrisome. In fact, the in-sample estimated coefficients from Table 3 on the business cycle variables are not significant, yet when using those in-sample parameters to time factors, the results in Figure 6 look good.[31]

Looking at growth momentum and inflation momentum, which similarly have no ex ante sign prediction for the long-short factors we examine, we find lackluster timing results. There is some small positive predictability for growth momentum, which is partly what the business cycle variables pick up, but no evidence that inflation momentum has any timing efficacy out of sample. Once again, the figure highlights the dangers of using in-sample parameter estimates, as the full sample regression model seems to generate positive returns, greater than anything achievable in real time. For all out-of-sample tests, there is no evidence for macroeconomic variables to time the factors.

[31] As alluded to earlier, macroeconomic variables can also suffer from look-ahead bias in the form of the macro data being revised ex post, and hence would not have been available ex ante to investors in real time. Moreover, often such revisions are designed to match markets better ex post, creating another form of look-ahead bias. We are careful in this study to avoid this latter bias, but many studies often embed this bias with the additional look-ahead bias that comes from using in-sample parameter estimates (Hodges, Hogan, Peterson, and Ang (2017)). Those choices tend to inflate substantially the efficacy of timing.

C.7 CAPE and VIX

The last two timing signals we examine – CAPE and "VIX" – are based off of the general equity market and are designed to capture changing risk, risk aversion, or sentiment in the market.[32] Forecasted equity volatility is the equal-weighted average of all equity markets' realized three-year volatility (estimated from monthly returns). As Figure 6 shows, neither CAPE nor aggregate equity volatility deliver much timing predictability for the factors. The timing alphas are positive, but are small, producing information ratios below 0.20, with nothing reliably different from zero. Results are a bit stronger when imposing economic restrictions, a feature that is robust across a variety of timing signals. Using in-sample parameter estimates biases the results significantly upward.

We also report in Figure 6 the average of all timing strategies, which is simply an equal-weighted average of all timing strategies based on each signal for each different method, to summarize the results. The average timing results are mediocre, producing positive but insignificant alphas with respect to the underlying static factors. On average, the economic sign constraints imposed on timing measures improve out-of-sample performance, suggesting theoretical guidance helps avoid overfitting. On the other hand, using in-sample moments estimated over the full sample for all timing models consistently overpromises timing ability.

C.8 Full model timing

While each of the timing variables might have some predictability for expected returns, it may be more powerful to combine timing signals if they each offer some independent glimpse of conditional return premia for the factors (although this introduces more estimation variance). A simple combination of value and momentum timing yielded lackluster results, but here we try to combine all 11 timing signals. The last set of results in Figure 6 examines a "full model" of factor timing using all timing signals simultaneously. For the *z*-score methodology, we rank all assets based on their average score across all eleven timing signals. For the regression methodologies, we run a multiple regression of future factor returns on all eleven timing signals and use the product of the estimated coefficients and the current timing signals to produce an expected return forecast for each factor in each asset class. Under certain specifications we also restrict the signs of some of the timing variables to match economic intuition and/or to be the same for a given factor.

The results for the full model show the most promising timing returns. The out-of-sample performance delivers information ratios above 0.40, that are orthogonal to the underlying static factors. Imposing economic constraints on the coefficients further improves out-of-sample

[32] We use these measures in lieu of Baker and Wurgler's (2008) sentiment index (from Table 4), because the latter only goes back to 1965.

performance, generating an information ratio of 0.62. Finally, using the full in-sample regression coefficient estimates for the timing model generates an information ratio of 1.1 or about double the best specification we could find out of sample, which highlights the dangers of using full sample information for predictability.[33]

D. Breaking Down Timing Strategies

Figure IA4 in the internet appendix breaks down the timing specification that provides the most consistently positive out-of-sample results, specification (6). Looking across the charts, there is little consistency in each asset class for the same factors contributing to the timing results, nor the same signals for a factor exhibiting consistency in performance. This inconsistency questions the robustness of the results and their economic interpretation, raising concerns that the positive full-model results could be driven by random chance found through a large specification search. The chart also highlights that the economic constraints bind on many of the timing variables frequently, indicating that for many of these variables there is an unstable relationship with future factor returns.

The lack of consistent results across asset classes also questions timing results from the literature that focus exclusively on U.S. individual stock factors. The evidence provides an array of out-of-sample tests of timing signals outside of U.S. equities and over a long period of time. Figure 7 plots the cumulative time-series of multivariate alphas for each timing strategy using specification (6) applied to all asset classes and factors. Each series is scaled to an ex-post 10% annual volatility to compare different timing strategies. As the first plot shows, the full timing model generates significant profits, and does not appear to be driven by a few extreme episodes. The remaining graphs plot cumulative alphas for each timing variable separately. The plots echo the earlier evidence that value spreads and inverse volatility provide some timing ability, though each exhibits substantial variation over time, experiencing occasional and large drawdowns. These strategies, along with factor spreads, had a particularly good run toward the end of our sample. Factor momentum suffered substantially in the early part of the sample but has positive predictability after the early 1970s, which coincides with the sample periods in Arnott et al. (2019) and Gupta and Kelly (2018). None of the macroeconomic or business cycle signals deliver much timing ability and exhibit substantial variation over time. Depending on the sample period chosen, one can make these signals look quite

[33] Figure IA3 in the internet appendix reports results for the PCA methodology of Haddad, Kozak, and Santosh (2018) applied to other timing signals besides value. In Figure IA3 we use their methodology on each of the 11 timing signals, using both out of sample PCA from the expanding historical window and in-sample PCA over the full sample. The performance of these timing strategies is weak and in several cases zero or negative. The PCA method of incorporating timing information through the covariance matrix as suggested by Haddad, Kozak, and Santosh (2018) does not appear to deliver robust out-of-sample timing performance.

positive or negative, and hence may explain some of the diversity of conclusions in the literature with regard to factor timing using macro variables.

E. Economic Impact of Factor Timing

The evidence on factor timing across our century of data and asset classes is modest. An open question is whether timing is economically significant and can meaningfully add to a portfolio that is diversified across factors and asset classes. We consider how much factor timing an investor would add to a static diversified factor portfolio to maximize the Sharpe ratio over the sample period. Table 7 reports the Sharpe ratios and information ratios with respect to the underlying static factors, of various timing strategies. We compute the optimal ex-post timing weight on the factor timing strategy that, when combined with the static multifactor, multi-asset portfolio, would maximize the in-sample Sharpe ratio. The first row reports statistics for the static diversified multifactor, multi-asset class portfolio that uses no timing. The Sharpe ratio of this portfolio, as reported previously, is 1.64. We also report the annual two-sided turnover of this portfolio per dollar of leverage, which is 4.3 (i.e. long 215% and short 215% of net asset value). Turnover per dollar levered measures the amount of trading taking into account leverage to compare portfolios on the same scale.

The remaining rows of Table 7 report these same statistics for various timing strategies. We start with the timing strategy that yields the greatest profits in our sample – the full timing model using a full-sample regression whose parameters are estimated in sample and imposes no restrictions on the coefficients. This is specification (7) from Table 6. We view this as the "best case scenario" for factor timing. As shown earlier, this strategy generates a sizeable information ratio with respect to the static underlying factors of 1.10. Combining this timing strategy with the static diversified factor portfolio, produces a Sharpe ratio of 2.00, where the ex-post optimal weight on the timing strategy is 40.2%. These results represent the best case scenario we can get from timing in our sample using the timing signals we study and full sample estimates of the parameters.

However, timing strategies also require additional turnover and trading. Table 7 reports that the timing strategy has turnover per dollar leverage of 6.2 (relative to 4.3 for the static portfolio). A question is whether this additional turnover is worth it? Rather than attempt to build a trading cost model that applies to all the asset classes we study, we instead back out the break-even costs from the *additional* turnover that would wipe out all of the gains from the timing strategy. Specifically, we calculate the additional turnover of the timing strategy (not turnover per dollar leverage, but total turnover) multiplied by the 40.2% timing weight and compute how large the costs would have to be per dollar traded to offset the performance increase from adding timing. In this particular case, that number is 9.8 basis points (bps) per dollar traded. Hence, as long as transactions costs are less than

9.8 bps per dollar traded, adding this timing strategy would improve the net returns of a diversified factor portfolio. Based on evidence in equities from Frazzini, Israel, and Moskowitz (2017), trading costs at a decent size (e.g., approximately one percent of daily volume) would be about this high and perhaps a little higher – for a cost-efficient arbitrageur over the past two decades.[34] Outside of equities, no good estimates of trading costs exist in the literature, but based on some proprietary data from AQR Capital, average trading costs in currencies and equity index futures might be a bit lower and for fixed income and commodities about the same or a bit higher than the equity estimates, for roughly the same fraction of trading volume. But, we caution that these estimates only reflect the current market and trading infrastructure, as opposed to what was available to investors decades ago which constitutes the majority of our sample and is likely much higher.

The next row reports the same statistics for the same timing model that estimates all of its parameters out of sample. This is specification (5) from Table 6, which is based off of the same regression but where the regression is estimated at each point in time using an expanding historical window of data. As Table 7 reports, the performance of the out of sample timing model is much weaker, generating a Sharpe ratio of 0.37 and information ratio of 0.41. This implementable timing strategy would only improve the unconditional gross Sharpe ratio of the diversified factor portfolio from 1.64 to 1.69, where the optimal weight on timing is 20 percent. This small increment in gross Sharpe ratio comes at a cost of an increase in turnover from 4.3 to 7.9 per dollar leverage.[35] Backing out the implied break-even trading cost from that increase in turnover, it would only take 2.5 basis points per dollar traded to wipe out the benefits of factor timing. Actual trading costs at a reasonable size almost certainly exceed this.

The next row reports results for our preferred out-of-sample specification of the full model timing regression, imposing economic restrictions on the coefficients consistent with theory (i.e., specification (6)). The information ratio of this timing strategy is 0.62 and when added to the static diversified multifactor portfolio increases the Sharpe ratio from 1.64 to 1.79, where the optimal weight on the timing strategy is 27.6%. This timing portfolio has a turnover per dollar leverage of 6.8, which is lower than the model that imposed no restrictions on the coefficients. Hence, another benefit of imposing economic restrictions is that they limit turnover of the portfolio in addition to

[34] For many investors, and for a more distant history, trading costs are likely to be a lot higher. Lacking good trading cost data going back more than two decades, we compare break-even costs to more recent estimates of trading costs.
[35] It is interesting that the out of sample regression approach increases turnover by much more than the in-sample approach did – 7.9 versus 6.2 bps per dollar leverage. This is due to the out-of-sample methodology allowing for time-variation in the regression coefficients, whereas the full-sample coefficients are of course constant. The out-of-sample methodology performs worse from a gross return perspective, since it relies purely on ex ante information, and fares even worse in terms of net of cost returns due to higher turnover.

improving out-of-sample performance. The break-even trading cost of adding this timing strategy to the underlying factor portfolio is 5.0 bps per dollar traded, which is about the average estimated cost for a reasonably sized portfolio *today*, if not for our full sample period.

The remaining 11 rows of Table 6 repeat this exercise for each timing signal. Consistent with our previous results, value spreads, inverse volatility, and business cycle timing are the only ones that seem to improve performance. Nevertheless, none of these signals individually get larger weight than the full timing model, and in all cases, the added turnover from applying factor timing can only be justified if trading costs are minimal; on the order of less than 4 bps per dollar traded.

The case for adding factor timing to an already diversified multifactor portfolio is tenuous. Despite looking at a plethora of timing strategies, methodologies, and signals, we find modest and inconsistent evidence of out-of-sample factor timing that is implementable in real time. Accounting for increased turnover and trading costs associated with factor timing, the net of cost returns are likely de minimis. On a more positive note, despite limited ability to profit from factor timing in real time, for some specifications we find significant conditional return premia associated with common factors across diverse asset classes. Given the number of timing methodologies we examine, the statistical strength of this predictability is debatable. Nevertheless, we find some evidence of value spreads and inverse volatility being able to capture conditional premia, consistent with economic theory in multiple asset classes and time periods and from multiple specifications. Moreover, imposing economic restrictions from theory on timing variables helps identify time-varying factor returns. This evidence offers hope for identifying conditional expected returns in the economy and the types of asset pricing models that can accommodate them. Studying a host of timing models with novel time-series data across multiple asset classes provides a new testing ground for these theories. Future research may well uncover a more powerful way to extract conditional information that yields more substantial economic returns.

B. Conclusion

A century of data across six diverse asset classes provides a rich laboratory to investigate the robustness of factor return premia, out-of-sample tests of existing theories, and conditional factor return premia. We find that return premia for value, momentum, carry, and defensive are robust and significant in every asset class over the last century. Their magnitudes, however, appear stronger in the original sample periods in which they were discovered, consistent with potential overfitting. We find no evidence that factor premia have been affected by informed arbitrage trading after their discovery. Common variation across asset classes for a given factor and across factors are stable over the century. Seeking to understand this variation and tie it to proposed asset pricing theories, we examine a century of economic news and shocks globally, but fail to find reliable or consistent evidence of macroeconomic, business cycle, tail risks, or sentiment driving factor premia. Finally, we analyze a host of timing models with varying specifications and signals and find some ability to capture conditional factor premia by valuation spreads and inverse volatility. However, trading profits to an implementable factor timing strategy are disappointing once we account for increased exposure to static factors and implementation issues such as real-time information and costs.

Our results shed light on unconditional and conditional asset pricing factors across many asset classes, with a wealth of out-of-sample results that challenge and support various theories and offer new features for asset pricing models to accommodate.

REFERENCES

Adrian, T., R. K. Crump, and E. Moench (2015), "Regression-based estimation of dynamic asset pricing models," *Journal of Financial Economics 118* (2), 211–244.

Akbas, F., W. J. Armstrong, S. Sorescu, and A. Subrahmanyam (2015). Smart money, dumb money, and capital market anomalies. *Journal of Financial Economics 118* (2), 355–382.

Arnott, R. D., N. Beck, and V. Kalesnik (2016a). Timing "smart beta" strategies? Of course! Buy low, sell high! Working paper, Research Affiliates.

Arnott, R. D., N. Beck, and V. Kalesnik (2016b). To win with "smart beta" ask if the price is right. Working paper, Research Affiliates.

Arnott, R. D., M. Clements, V. Kalesnik, and J. Linnainmaa, (2018), Factor momentum. Working paper, Research Affiliates.

Asness, Clifford, Tobias J. Moskowitz, and Lasse Pedersen (2013), "Value and Momentum Everywhere," *Journal of Finance* 68 (3), 929-985.

Asness, Clifford, Andrea Frazzini, and Lasse Pedersen (2016), "Quality Minus Junk," AQR Capital Working paper.

Asness, C., A. Ilmanen, and T. Maloney (2017), Market timing: Sin a little, *Journal of Investment Management*, 15 (3), 23-40.

Asness, C., S. Chandra, A. Ilmanen, and R. Israel (2017). Contrarian factor timing is deceptively difficult. *The Journal of Portfolio Management 43* (5), 72–87.

Asness, C. S. (2016). The siren song of factor timing aka "smart beta timing" aka "style timing". *The Journal of Portfolio Management 42* (5), 1–6.

Asness, C. S., J. A. Friedman, R. J. Krail, and J. M. Liew (2000). Style timing: Value versus growth. *Journal of Portfolio Management 26* (3), 50–60.

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance 61* (4), 1645–1680.

Bakshi, G. and G. Panayotov (2013). Predictability of currency carry trades and asset pricing implications. *Journal of Financial Economics 110* (1), 139–163.

Baltussen, G., L. Swinkels, and P. Van Vliet (2019), Global factors premiums. Robeco Working Paper.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015). X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics 115* (1), 1–24.

Belo, F., 'Production-based measures of risk for asset pricing', Journal of Monetary Economics, Vol. 57, 2010, pp. 146–163.

Brooks, Jordan and Tobias J. Moskowitz (2017), "Yield Curve Premia," AQR Capital and Yale University Working paper.

Campbell, J. and R. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies 1* (3), 195–228.

Campbell, J. and R. Shiller (1998). "Valuation Ratios and the Long-Run Stock Market Outlook," *Journal of Portfolio Management* 24 (2): 11–26.

Campbell, J. Y., C. Polk, and T. Vuolteenaho (2009). Growth or glamour? fundamentals and systematic risk in stock returns. *Review of Financial Studies 23* (1), 305–344.

Campbell, J. Y. and S. B. Thompson (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies 21* (4), 1509–1531.

Carlson, M., Fisher, A. and Giammarino, R., 'Corporate investment and asset price dynamics: Implications for the cross section of returns', Journal of Finance, Vol. 59, 2004, pp. 2577–2603.

Cieslak, A. and P. Povala (2015). Expected returns in Treasury bonds. *Review of Financial Studies*, hhv032.

Cochrane, J.H., 'Production-based asset pricing and the link between stock returns and economic fluctuations', Journal of Finance, Vol. 46, 1991, pp. 209–237.

Cochrane, J.H., 'A cross-sectional test of an investment-based asset pricing model', Journal of Political Economy, Vol. 104, 1996, pp. 572–621.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies 21* (4), 1533–1575.

Cochrane, J. H. and M. Piazzesi (2005). Bond Risk Premia. *American Economic Review*, 138–160.

Cochrane, J. H. and M. Piazzesi (2008). Decomposing the yield curve. *Graduate School of Business*, *University of Chicago, Working Paper*.

Cohen, R. B., C. Polk, and T. Vuolteenaho (2003). The value spread. *The Journal of Finance 58* (2), 609–641.

Cooper, I. and R. Priestley (2008). Time-varying risk premiums and the output gap. *The Review of Financial Studies 22* (7), 2801–2833.

Cooper, I. and Priestley, R., 'Real investment and risk dynamics', Journal of Financial Economics, Vol. 101, 2011, pp. 182–205.

Daniel, Kent and Tobias J. Moskowitz (2016), "Momentum Crashes," *Journal of Financial Economics* 122 (2), 221-247.

Ehsani, S. and J.T. Linnainmaa (2019), "Factor Momentum and the Momentum Factor." Working paper.

Estrella, A. and F. S. Mishkin (1997). The predictive power of the term structure of interest rates in europe and the united states: Implications for the european central bank. *European economic review 41* (7), 1375–1401.

Fama, E. and R. Bliss (1987). The information in long-maturity forward rates. *The American Economic Review*, 680–692.

Fama, E. and K. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics 22* (1), 3–25.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stock and bonds. *Journal of Financial Economics 33*, 3–56.

Fama, Eugene F. and Kenneth R. French (1996), Multifactor explanations of asset pricing anomalies. *Journal of Finance* 51, 55–84.

Fama, Eugene F. and Kenneth R. French (2012), Size, value, and momentum in international stock returns. *Journal of Financial Economics*.

Frazzini, Andrea, and Lasse H. Pedersen (2013), "Betting against beta," *Journal of Financial Economics* 111 (1) 1-25.

Geczy, C.G., and M. Samonov (2016), Two centuries of price return momentum. *Financial Analysts Journal*, 72 (5).

Gomes, J.F., L. Kogan. and L. Zhang, 'Equilibrium cross section of returns', Journal of Political Economy, Vol. 111, 2003, pp. 693–732.

Greenwood, R. and S. G. Hanson (2012). Share issuance and factor timing. *The Journal of Finance 67* (2), 761–798.

Gupta, T. and B. Kelly (2018), Factor Momentum Everywhere, *forthcoming Journal of Portfolio Management*.

Haddad, Valentin, Serhiy Kozak, and Shrihari Santosh, 2018, “Economic Factor Timing,” Working paper UCLA.

Harvey, C. R. and Y. Liu (2016). Lucky factors. Working paper, Duke University.

Hodges, Philip, Ked Hogan, Justin R. Peterson, and Andrew Ang, 2017, “Factor Timing with Cross-Sectional and Time-Series Predictors,” *The Journal of Portfolio Management* 44 (1) 30-43.

Hou, K., Xue, C. and Zhang, L., ‘Digesting anomalies: An investment approach’, Review of Financial Studies, Vol. 28, 2015, pp. 650–705.

Hou, K., C. Xue, and L. Zhang (2017). Replicating anomalies. Working paper, National Bureau of Economic Research.

Ilmanen, Antti, 2011, *Expected Returns*, Wiley.

Israel, Ronen, Diogo Palhares, and Scott A. Richardson (2016): “Common Factors in Corporate Bond and Bond Fund Returns,” AQR Capital Management Working paper.

Jegadeesh, Narasimhan and Sheridan Titman (1993), “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency,” *Journal of Finance*, vol. 48, pp. 65-91.

Koijen, Ralph, Tobias J. Moskowitz, Lasse Pedersen, and Evert Vrugt (2018), “Carry,” forthcoming *Journal of Financial Economics*.

Kozak, S., S. Nagel, and S. Santosh (2018). Shrinking the cross-section.

Lettau, M., M. Maggiori, and M. Weber (2014). Conditional risk premia in currency markets and other asset classes. *Journal of Financial Economics 114* (2), 197–225.

Li, D. and Zhang, L., ‘Does q-theory with investment frictions explain anomalies in the cross section of returns? Journal of Financial Economics, Vol. 98, 2010, pp. 297–314.

Li, E.X.N., Livdan, D. and Zhang, L., ‘Anomalies’, Review of Financial Studies, Vol. 22, 2009, pp. 4301–4334.

Liu, L.X.L. Whited, T.M. and Zhang, L., ‘Investment-based expected stock returns’, Journal of Political Economy, Vol. 117, 2009, pp. 1105–1139.

Liu, L.X.L. and Zhang, L., ‘A neoclassical interpretation of momentum’, Journal of Monetary Economics, Vol. 67, 2014, pp. 109–128.

Lochstoer, L. A. and P. C. Tetlock (2016). What drives anomaly returns?

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *The Review of Financial Studies 22* (12), 5027–5067.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the*

Econometric Society, 867–887.

Moreira, A. and T. Muir (2017). Volatility-managed portfolios. *The Journal of Finance 72* (4), 1611–1644.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Pedersen, 2012. "Time Series Momentum." *Journal of Financial Economics* 104 (2), 228-250.

Ohlson, J. A. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary accounting research 11* (2), 661–687.

Rosenberg, B., Reid, K. and Lanstein, R., 'Persuasive evidence of market inefficiency', Journal of Portfolio Management, Vol. 11, 1985, pp. 9–16.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13* (3), 341–360.

Shiller, R. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review 71*, 321 – 436.

Stambaugh, Robert F., 1999, "Predictive regressions," *Journal of Financial Economics* 54 375-421.

Stambaugh, R. F., J. Yu, and Y. Yuan (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics 104* (2), 288–302.

Vuolteenaho, T. (2002). What drives firm-level stock returns? *The Journal of Finance 57* (1), 233–264.

Xing, Y., 'Interpreting the value effect through the Q-theory: An empirical investigation', Review of Financial Studies, Vol. 21, 2008, pp. 1767–1795.

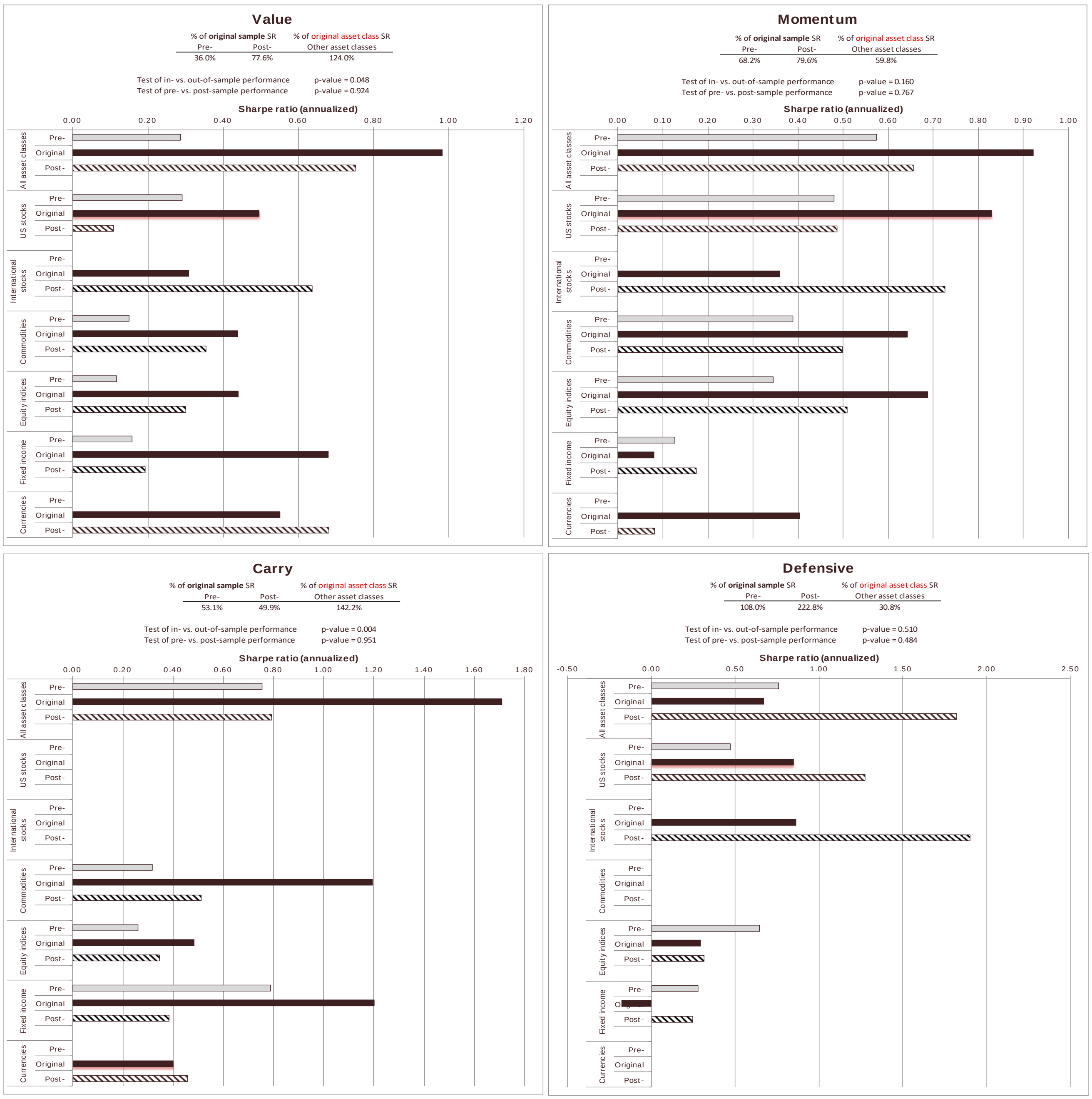
Zhang, L., 'The value premium', Journal of Finance, Vol. 60, 2005b, pp. 67–103.

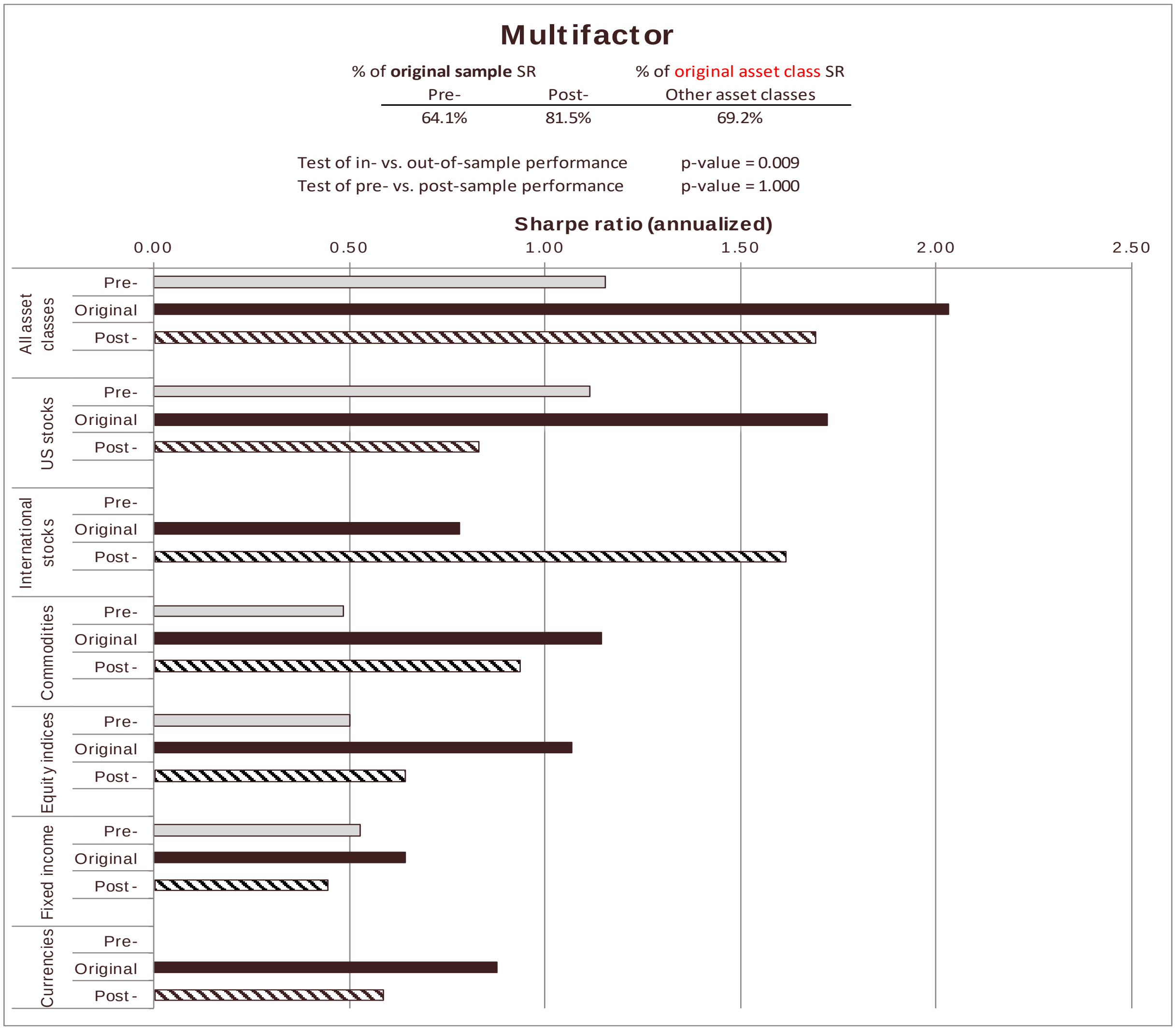
Table 1: Factor Return Premia Over a Century

Reported are the annualized mean, standard deviation, Sharpe ratio, t-statistic of the mean (from zero), skewness, kurtosis, and maximum drawdown of the returns to each factor portfolio in each asset class over the last century. The sample periods for each asset class are reported. Results are reported for all asset classes combined (equal-volatility-weighted average of the asset classes), and for U.S. stocks, international stocks, commodities, equity index futures, bonds, and currencies separately. The last column reports p-values of F-tests for whether each factor delivers the same Sharpe ratio across asset classes and whether a multifactor portfolio generates different Sharpe ratios across asset classes.

Asset class	Factor/Style	Mean	Stdev	Sharpe	t-stat	Skew	Kurt	Max drawdown	Start date	End date	p-value of equal Sharpe
All asset classes	Value	2.2%	3.6%	0.62	6.00	0.8	5.1	-16%	Jun-1923	Apr-2018	0.019
	Momentum	3.0%	4.5%	0.67	6.46	-0.9	4.9	-19%	Mar-1923	Apr-2018	0.002
	Carry	3.2%	3.8%	0.84	8.03	0.1	3.1	-11%	Mar-1923	Apr-2018	0.272
	Defensive	3.1%	3.9%	0.78	7.54	-0.6	4.6	-18%	Mar-1923	Apr-2018	0.000
	Multifactor	3.1%	1.9%	1.59	14.72	0.2	2.1	-4%	Mar-1923	Apr-2018	0.000
US stocks	Value	4.3%	14.8%	0.29	2.80	3.5	32.8	-48%	Aug-1926	Dec-2017	
	Momentum	8.4%	15.6%	0.54	5.08	-3.0	26.6	-64%	Jan-1927	Dec-2017	
	Defensive	8.0%	10.8%	0.74	6.82	-0.9	6.5	-48%	Dec-1930	Dec-2017	
	Multifactor	6.9%	5.9%	1.17	10.88	0.5	8.9	-22%	Aug-1926	Dec-2017	
International stocks	Value	5.1%	9.6%	0.53	3.07	0.1	6.6	-38%	Jul-1984	Dec-2017	
	Momentum	8.3%	12.5%	0.67	3.79	-1.1	5.9	-41%	Jan-1985	Dec-2017	
	Defensive	10.7%	9.8%	1.10	5.94	0.0	0.9	-33%	Feb-1987	Dec-2017	
	Multifactor	8.0%	5.4%	1.47	8.17	0.4	2.7	-11%	Jul-1984	Dec-2017	
Commodities	Value	5.5%	19.2%	0.29	2.79	0.1	1.2	-59%	Jan-1920	Apr-2018	
	Momentum	9.2%	18.8%	0.49	4.74	-0.3	2.2	-64%	Jan-1920	Apr-2018	
	Carry	8.0%	17.2%	0.46	4.51	0.0	1.4	-62%	Jan-1920	Apr-2018	
	Multifactor	7.5%	9.6%	0.79	7.60	0.2	3.5	-22%	Jan-1920	Apr-2018	
Equity indices	Value	3.2%	13.4%	0.24	2.30	0.3	4.9	-56%	Mar-1925	Apr-2018	
	Momentum	6.8%	15.4%	0.44	4.28	0.2	7.3	-49%	Mar-1923	Apr-2018	
	Carry	4.5%	15.9%	0.28	2.73	2.4	30.5	-76%	Mar-1923	Apr-2018	
	Defensive	5.9%	12.9%	0.46	4.48	0.2	5.2	-63%	Mar-1923	Apr-2018	
	Multifactor	5.3%	8.7%	0.61	5.86	2.3	38.3	-24%	Mar-1923	Apr-2018	
Fixed income	Value	1.4%	4.8%	0.30	2.91	0.3	8.9	-27%	Jun-1923	Apr-2018	
	Momentum	0.6%	4.8%	0.12	1.15	-0.8	6.4	-26%	Mar-1923	Apr-2018	
	Carry	3.0%	4.5%	0.66	6.40	0.5	8.5	-17%	Mar-1923	Apr-2018	
	Defensive	0.2%	4.3%	0.05	0.52	-0.3	5.5	-52%	Mar-1923	Apr-2018	
	Multifactor	1.3%	2.4%	0.54	5.19	-0.6	11.5	-13%	Mar-1923	Apr-2018	
Currencies	Value	3.4%	5.3%	0.64	4.21	0.2	1.6	-11%	Apr-1974	Apr-2018	
	Momentum	1.2%	6.7%	0.18	1.23	-0.5	0.8	-24%	Feb-1974	Apr-2018	
	Carry	2.8%	6.5%	0.44	2.91	-0.7	4.0	-28%	Feb-1974	Apr-2018	
	Multifactor	2.5%	3.8%	0.65	4.30	-0.5	2.0	-10%	Feb-1974	Apr-2018	

Figure 1. Factor Premia in the Original-Sample, Post-Publication, and Pre-Sample Periods Across Asset Classes. The figure plots the annualized Sharpe ratios of each factor in each asset class over their respective pre-, original-, and post-sample periods, defined using the dates in Mclean and Pontiff (2016) for U.S. equities but applied to our century of data in other asset classes. For carry strategies we use the original sample dates from Meese and Rogoff (1983) and Fama (1984). The original sample for the original asset class is highlighted in red. We also aggregate across all asset classes into a diversified factor. Computing the pre- and post-sample Sharpe ratios relative to the original sample Sharpe ratio for each asset class, we report at the top of each graph the equal-weighted average of these percentages across all asset classes. We also report the Sharpe ratios of each factor in each asset class relative to the Sharpe ratio of the original asset class in which the factor was first discovered. A formal *F*-test of the difference in Sharpe ratios for the original-sample versus out-of-sample periods and for the pre-versus post-publication sample periods is also reported at the top of each graph.





Multifactor
% of original sample SR
Pre-
Post-
64.1%
81.5%
% of original asset class SR
Other asset classes
69.2%
Test of in- vs. out-of-sample performance
p-value = 0.009
Test of pre- vs. post-sample performance
p-value = 1.000
Sharpe ratio (annualized)
0.00
0.50
1.00
1.50
2.00
2.50
All asset classes
US stocks
International stocks
Commodities
Equity indices
Fixed income
Currencies
Pre-
Original
Post-

Table 2: Full-Sample Correlations Across Factors and Asset Classes

Panel A reports correlations of the asset classes for a given factor, where we compute the correlation of a factor in each asset class with that same factor diversified across all *other* asset classes. Panel B reports the unconditional correlations of the factors within each asset class over the entire century of data.

Panel A: Correlations of asset classes by factor

	Correlation of each asset class with all other asset classes by factor				
	Value	Momentum	Carry	Defensive	Multifactor
Commodities	0.07	0.12	0.01		0.02
Equity indices	0.13	0.23	0.02	0.11	0.09
Fixed income	0.03	0.10	0.02	0.00	0.03
Currencies	0.03	0.21	0.05		0.03
US stocks	0.16	0.34		0.21	0.11
International stocks	0.49	0.62		0.44	0.13
Average	0.15	0.27	0.02	0.19	0.07

Panel B: Correlations of factors within an asset class

	Value	Momentum	Carry	Defensive	Value	Momentum	Carry	Defensive
	US Stocks				International Stocks			
Value	1.00	-0.68		-0.17	1.00	-0.60		-0.02
Momentum		1.00		0.31		1.00		0.26
Carry			1.00				1.00	
	Equity Indices				Fixed Income			
Value	1.00	-0.33	0.25	0.05	1.00	-0.17	0.27	-0.01
Momentum		1.00	-0.03	0.14		1.00	0.07	0.05
Carry			1.00	0.25			1.00	-0.01
	Currencies				Commodities			
Value	1.00	-0.17	0.25		1.00	-0.43	-0.28	
Momentum		1.00	0.13			1.00	0.42	
Carry			1.00				1.00	
	All Asset Classes							
Value	1.00	-0.51	0.09	-0.11				
Momentum		1.00	0.09	0.24				
Carry			1.00	0.10				

Figure 2. Time-Varying Correlations of Factors. The figure plots the pairwise correlations between the factors (across all asset classes) over time using rolling monthly return data over the prior 10 years to estimate the correlations at each point in time.

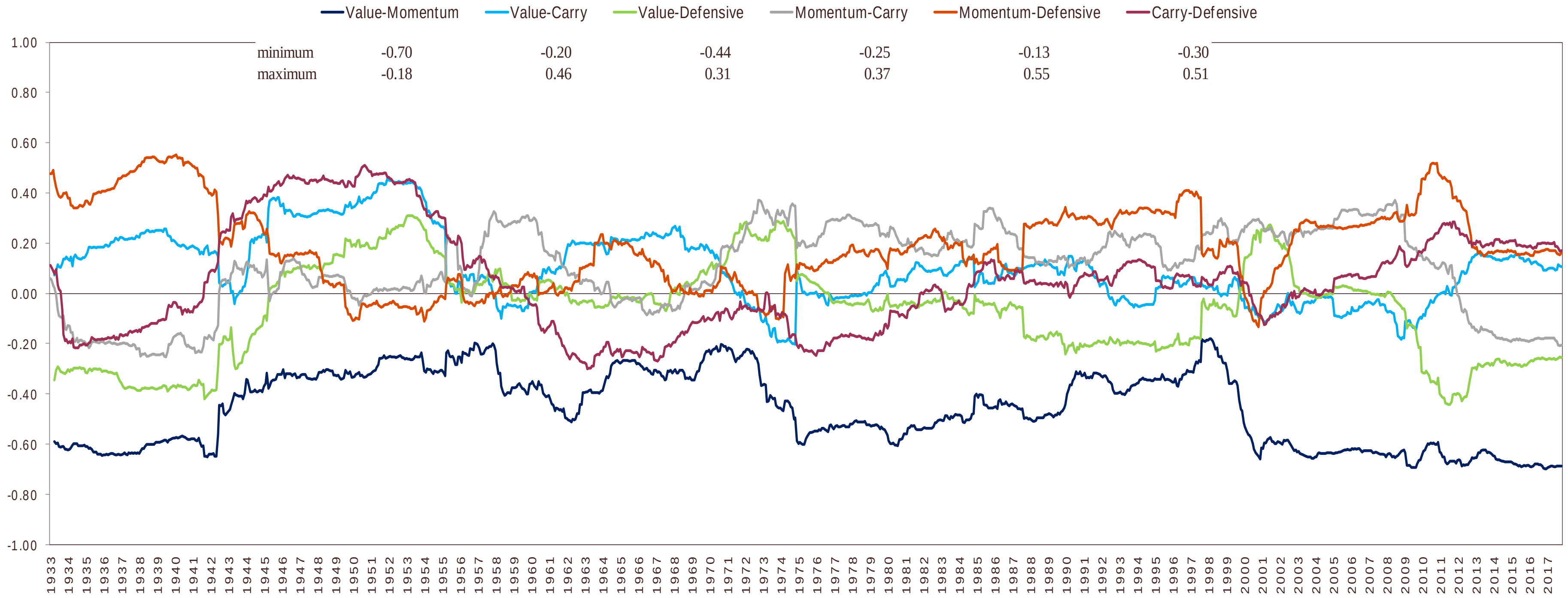
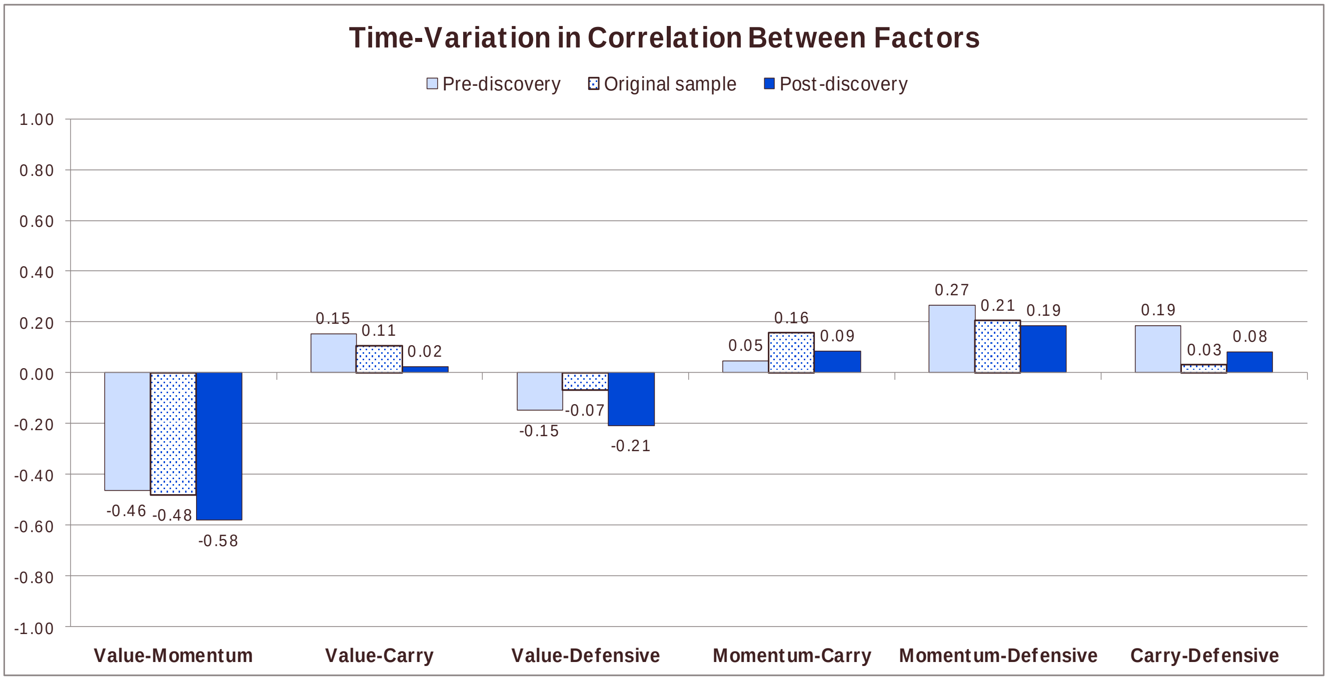


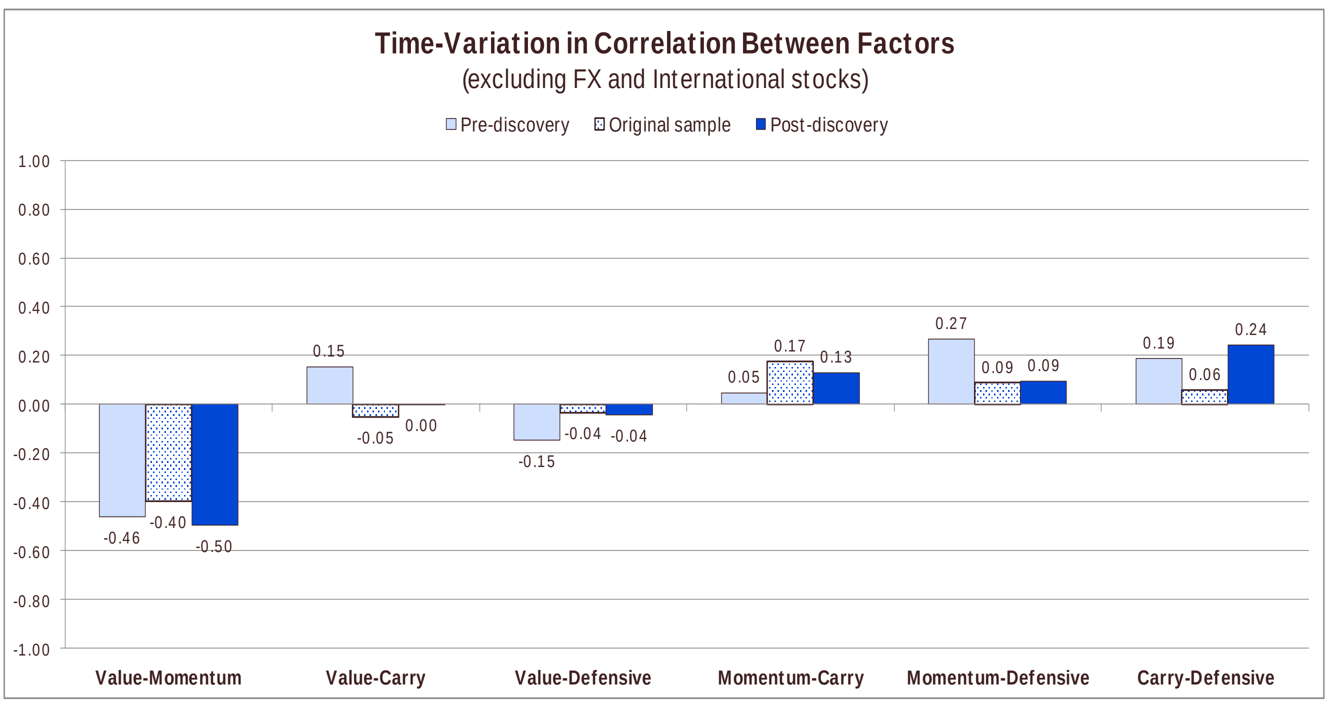
Table 3: Factor Loadings and Alphas

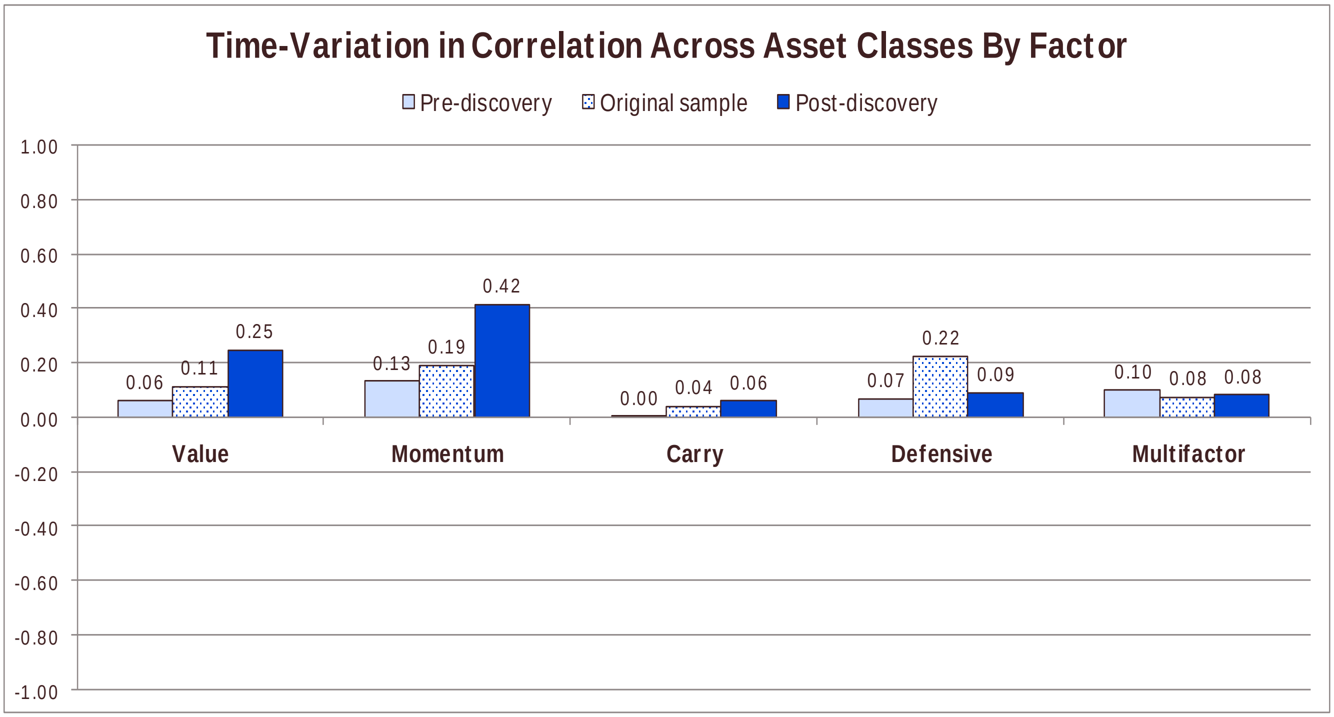
The table reports the results from time-series regressions of each factor's returns in each asset class on the other factor returns within the same asset class and the same factor's returns in all other asset classes over the full century of data. The factors diversified across all other asset classes are weighted by the inverse of their volatility using three-year rolling estimates of monthly returns to measure volatility. An equity, bond, and commodity market index (equal-weighted portfolios of all global equities, bonds, and commodities, respectively) are included as well.

			Other factors				Other asset classes	Market		
Asset class	**Factor**	**Alpha**	**Value**	**Momentum**	**Carry**	**Defensive**	**Same factor**	**Equity index**	**Bond index**	**Commodity index**
All asset classes	Value	**0.0024**		-0.42	0.13	-0.001		0.01	0.03	-0.01
		(8.71)		-(19.78)	(5.54)	-(0.03)		(1.75)	(1.26)	-(1.30)
	Momentum	**0.0029**	-0.62		0.14	0.19		-0.02	0.01	-0.01
		(8.38)	-(19.78)		(4.88)	(6.70)		-(2.10)	(0.43)	-(2.15)
	Carry	**0.0018**	0.20	0.14		0.08		-0.01	-0.06	0.02
		(5.12)	(5.54)	(4.88)		(2.72)		-(0.64)	-(1.71)	(3.26)
	Defensive	**0.0017**	0.00	0.20	0.08			-0.01	0.08	0.01
		(4.86)	-(0.03)	(6.70)	(2.72)			-(0.88)	(2.37)	(1.11)
	Multi-style	**0.0026**						-0.01	0.02	0.00
		(15.42)						-(2.57)	(0.88)	(1.13)
US stocks	Value	**0.0069**		-0.64		0.06	0.25	-0.04	0.08	0.16
		(6.79)		-(27.75)		(1.75)	(2.47)	-(1.43)	(0.79)	(7.36)
	Momentum	**0.0066**	-0.64			0.27	0.67	-0.16	-0.14	0.07
		(6.65)	-(27.02)			(8.58)	(8.02)	-(5.60)	-(1.37)	(3.11)
	Defensive	**0.0030**	0.06	0.25			0.58	0.06	0.44	0.00
		(3.11)	(2.01)	(8.94)			(6.80)	(2.21)	(4.73)	(0.10)
	Multi-style	**0.0045**					0.61	-0.05	0.16	0.04
		(8.10)					(6.28)	-(3.06)	(2.91)	(3.89)
International stocks	Value	**0.0040**		-0.42		0.19	0.84	-0.08	-0.02	0.01
		(3.46)		-(12.65)		(4.74)	(7.98)	-(3.39)	-(0.30)	(0.36)
	Momentum	**0.0054**	-0.52			0.21	1.00	-0.16	0.03	-0.09
		(4.12)	-(10.68)			(4.83)	(10.44)	-(5.56)	(0.28)	-(2.53)
	Defensive	**0.0053**	0.13	0.21			0.80	-0.09	0.13	0.18
		(3.72)	(2.20)	(4.36)			(7.16)	-(2.78)	(1.22)	(4.52)
	Multi-style	**0.0047**					0.90	-0.11	0.06	0.03
		(5.99)					(7.54)	-(6.68)	(0.97)	(1.59)
Commodities	Value	**0.0085**		-0.45	-0.13		0.23	-0.03	-0.29	-0.15
		(5.66)		-(14.74)	-(3.77)		(1.55)	-(0.69)	-(1.91)	-(4.65)
	Momentum	**0.0061**	-0.38		0.36		0.46	0.00	-0.04	-0.07
		(4.36)	-(15.00)		(12.30)		(4.32)	(0.06)	-(0.28)	-(2.21)
	Carry	**0.0047**	-0.10	0.35			0.00	-0.01	-0.26	-0.01
		(3.36)	-(3.70)	(12.30)			(0.02)	-(0.16)	-(1.92)	-(0.40)
	Multi-style	**0.0057**					0.23	-0.01	-0.17	-0.06
		(6.67)					(1.63)	-(0.45)	-(2.03)	-(3.19)
Equity indices	Value	**0.0027**		-0.26	0.20	0.07	0.40	-0.03	0.06	-0.05
		(2.39)		-(9.44)	(6.38)	(2.42)	(3.63)	-(1.00)	(0.53)	-(1.92)
	Momentum	**0.0045**	-0.29		-0.18	0.10	0.83	0.00	0.06	0.03
		(3.95)	-(9.87)		-(5.78)	(3.21)	(9.03)	-(0.10)	(0.52)	(1.34)
	Carry	**0.0022**	0.18	-0.16		0.09	0.19	-0.09	0.14	0.08
		(2.06)	(6.31)	-(6.09)		(3.02)	(1.88)	-(2.78)	(1.30)	(3.51)
	Defensive	**0.0030**	0.06	0.11	0.10		0.27	-0.05	0.08	-0.02
		(2.78)	(2.08)	(3.89)	(3.16)		(2.83)	-(1.44)	(0.76)	-(0.68)
	Multi-style	**0.0039**					0.46	-0.15	0.08	0.02
		(4.79)					(3.33)	-(6.76)	(1.01)	(1.13)
Fixed income	Value	0.0001		-0.18	0.30	0.00	0.03	0.06	0.08	-0.01
		(0.33)		-(6.44)	(10.03)	-(0.13)	(1.06)	(4.80)	(1.88)	-(1.11)
	Momentum	0.0000	-0.19		0.12	0.06	0.10	0.01	0.04	-0.02
		(0.04)	-(6.23)		(3.80)	(1.97)	(4.48)	(0.53)	(0.87)	-(2.15)
	Carry	**0.0021**	0.27	0.11		-0.01	0.04	-0.02	-0.01	0.01
		(5.44)	(10.07)	(4.04)		-(0.32)	(1.93)	-(1.61)	-(0.19)	(1.09)
	Defensive	0.0001	0.00	0.05	-0.01		0.00	0.03	-0.09	0.00
		(0.37)	-(0.12)	(1.95)	-(0.39)		(0.12)	(2.81)	-(2.32)	(0.13)
	Multi-style	**0.0008**					0.05	0.02	0.00	0.00
		(3.69)					(1.70)	(3.14)	-(0.13)	-(0.98)
Currencies	Value	**0.0023**		-0.16	0.26		-0.01	0.01	0.00	-0.06
		(3.52)		-(4.76)	(6.89)		-(0.15)	(0.68)	(0.00)	-(3.26)
	Momentum	0.0002	-0.23		0.16		0.29	0.06	-0.09	0.01
		(0.22)	-(4.31)		(3.37)		(6.25)	(2.73)	-(1.37)	(0.33)
	Carry	0.0010	0.32	0.11			0.10	0.10	-0.12	0.11
		(1.33)	(6.81)	(3.00)			(1.81)	(5.73)	-(2.05)	(5.67)
	Multi-style	**0.0014**					0.17	0.07	-0.09	0.03
		(2.75)					(2.51)	(6.15)	-(2.36)	(2.54)

Figure 3. Correlations in the Original-Sample, Post-Publication, and Pre-Sample Periods. The first graph plots time-variation in the correlation *between* factors (e.g., correlation between value and momentum). We examine all pairwise correlations between the four factors applied across all asset classes simultaneously, and estimate them separately over the pre-, original, and post-sample periods, which are all reported below. The bottom figure excludes currencies and international stocks that have the shortest sample periods. The last two graphs examine correlations *across* asset classes for a given factor, by averaging the pairwise correlations between asset classes for a given factor (e.g., the average correlation of the value factor across markets).






Time-Variation in Correlation Across Asset Classes By Factor
Pre-discovery
Original sample
Post-discovery
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
-0.80
-1.00
0.06
0.11
0.25
0.13
0.19
0.42
0.00
0.04
0.06
0.07
0.22
0.09
0.10
0.08
0.08
Value
Momentum
Carry
Defensive
Multifactor

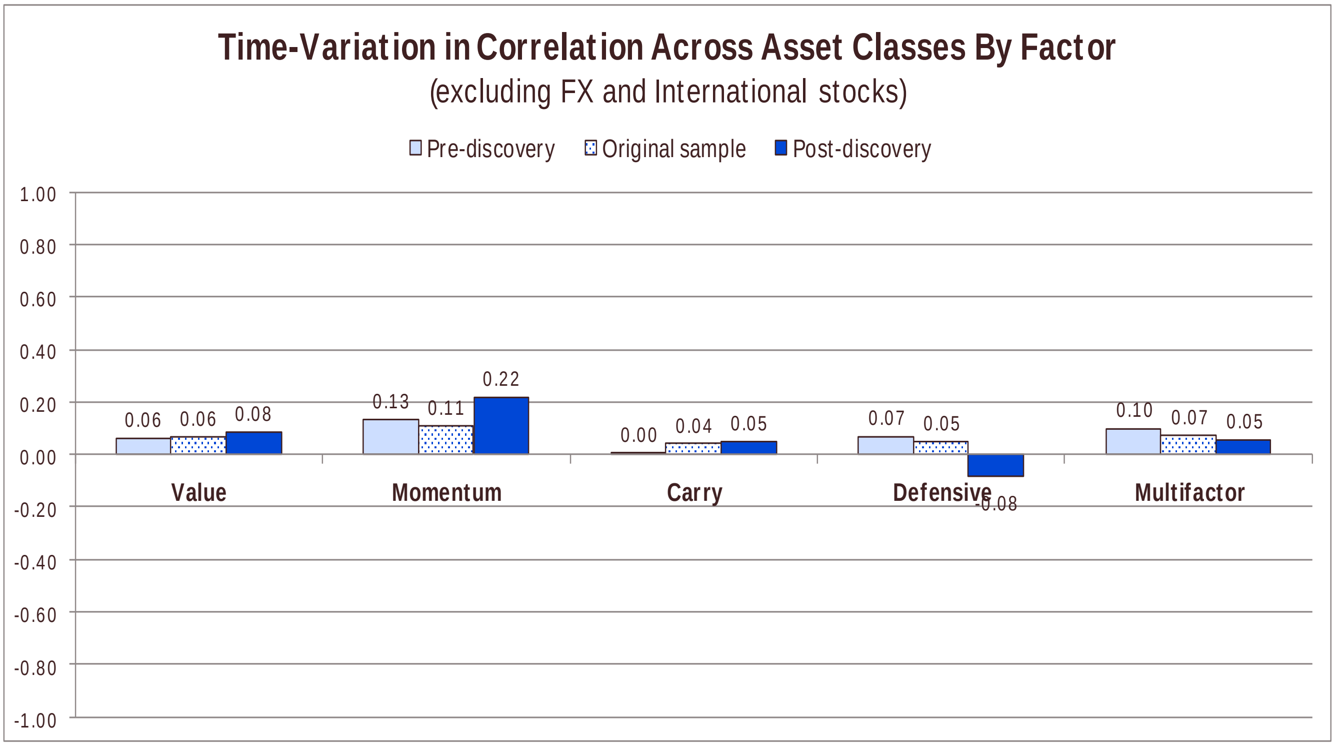

Time-Variation in Correlation Across Asset Classes By Factor
(excluding FX and International stocks)
Pre-discovery
Original sample
Post-discovery
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
-0.80
-1.00
0.06
0.06
0.08
0.13
0.11
0.22
0.00
0.04
0.05
0.07
0.05
-0.08
0.10
0.07
0.05
Value
Momentum
Carry
Defensive
Multifactor

Table 4: Contemporaneous and Predictive Marcroeconomic Exposures

The table reports results from a time-series regression of each factor's returns over the last century on various economic measures. Panel A reports contemporaneous regressions between factor returns at time *t* and the economic variables at the same time *t*. The variables include the illiquidity risk variable of Amihud (2014), the Baker and Wurgler (2008) sentiment index, equity market volatility (realized volatility of the equal-weighted country indices, estimated over the prior 36 months), global GDP growth (growth over the last year averaged over the U.S., UK, Germany, and Japan), global CPI inflation growth (growth over the last year in inflation averaged over the U.S., UK, Germany, and Japan), a tail risk indicator (if the developed equity market index is in the lower fifth percentile), geopolitical risk index (from http://www.policyuncertainty.com/gpr.html), and three business cycle indicators: contraction, expansion, and slowdown, which are determined using levels and changes in GDP growth. We define periods into positive and negative growth based on GDP growth each quarter, and also define periods into "accelerating" and "decelerating" growth each quarter based on the change in the change in GDP growth. The intersection of these two indicators creates four subperiods: contraction (negative growth and negative change in growth), recovery (negative growth and positive change in growth), expansion (positive growth and positive change in growth), and slowdown (positive growth and negative change in growth). Table IA2 in the internet appendix provides more details on the construction of each of these variables and their data sources. Panel B lags all macroeconomic variables by one period to capture their announcement lag, which captures the news associated with the economic variable and its contemporaneous impact on markets.

Panel A: Contemporaneous Economic Activity

	Value	Momentum	Carry	Defensive	Multifactor
Amihud illiquidity risk	0.0440	-0.0008	0.0117	0.0829	0.0363
	(3.03)	(-0.04)	(0.79)	(5.33)	(4.71)
Baker-Wurgler sentiment	0.0016	-0.0004	0.0001	0.0012	0.0007
	(3.10)	(-0.58)	(0.17)	(2.13)	(2.55)
Equity market volatility	0.0412	0.0157	0.0290	0.0161	0.0264
	(0.96)	(0.28)	(0.66)	(0.35)	(1.16)
GDP growth	-0.0445	0.0311	0.0000	-0.0617	-0.0212
	(-1.23)	(0.65)	(-0.00)	(-1.59)	(-1.10)
CPI inflation	0.0359	0.0041	0.0245	-0.0639	0.0052
	(1.95)	(0.17)	(1.31)	(-3.25)	(0.53)
Tail risk dummy	-0.0027	0.0034	-0.0035	-0.0034	-0.0006
	(-1.52)	(1.45)	(-1.92)	(-1.75)	(-0.61)
Geopolitical risk index	0.0001	-0.0004	-0.0008	-0.0015	-0.0007
	(0.06)	(-0.32)	(-0.76)	(-1.44)	(-1.31)
Contraction dummy	-0.0036	-0.0015	-0.0044	-0.0057	-0.0035
	(-0.93)	(-0.29)	(-1.10)	(-1.36)	(-1.70)
Expansion dummy	0.0011	0.0007	-0.0002	-0.0022	-0.0001
	(1.07)	(0.51)	(-0.21)	(-2.03)	(-0.14)
Slowdown dummy	0.0080	-0.0109	0.0014	-0.0098	-0.0037
	(2.12)	(-2.20)	(0.37)	(-2.43)	(-1.84)
R^2	6.0%	2.5%	1.6%	10.6%	6.4%

Panel B: Predictive Economic News

	Value	Momentum	Carry	Defensive	Multifactor
Amihud illiquidity risk	0.0017	0.0280	-0.0263	0.0455	0.0148
	(0.11)	(1.48)	(-1.78)	(2.93)	(1.91)
Baker-Wurgler sentiment	0.0013	0.0000	0.0001	0.0015	0.0008
	(2.50)	(-0.03)	(0.16)	(2.61)	(2.92)
Equity market volatility	-0.0057	0.0664	0.0218	0.0881	0.0421
	(-0.13)	(1.19)	(0.50)	(1.91)	(1.84)
GDP growth	-0.0687	0.0236	0.0150	-0.0449	-0.0263
	(-1.86)	(0.50)	(0.40)	(-1.15)	(-1.36)
CPI inflation	0.0373	-0.0011	0.0324	-0.0516	0.0085
	(2.00)	(-0.05)	(1.73)	(-2.62)	(0.87)
Tail risk dummy	-0.0007	-0.0006	-0.0024	-0.0094	-0.0032
	(-0.37)	(-0.26)	(-1.34)	(-4.92)	(-3.34)
Geopolitical risk index	0.0011	-0.0036	0.0001	-0.0020	-0.0013
	(1.11)	(-2.79)	(0.06)	(-1.93)	(-2.51)
Contraction dummy	-0.0042	-0.0003	-0.0019	-0.0043	-0.0027
	(-1.08)	(-0.06)	(-0.47)	(-1.04)	(-1.28)
Expansion dummy	0.0003	0.0015	-0.0005	-0.0013	0.0001
	(0.29)	(1.15)	(-0.46)	(-1.19)	(0.23)
Slowdown dummy	0.0056	-0.0131	0.0056	-0.0080	-0.0036
	(1.46)	(-2.67)	(1.46)	(-1.97)	(-1.81)
R^2	3.9%	4.6%	1.9%	10.7%	5.7%

Figure 4. Time-Varying Correlations of Multi-Factor Portfolio with the Market. The figure plots time-varying correlations of the multifactor portfolio, diversified across value, momentum, carry, and defensive and diversified across all asset classes, to the returns of a market portfolio. We use three market proxies: the equal-weighted all-asset-class market portfolio, a global equity market portfolio, and a global fixed income portfolio, respectively. Correlations are estimated from rolling monthly returns over the last 10 years of data over the period 1933 to 2018.

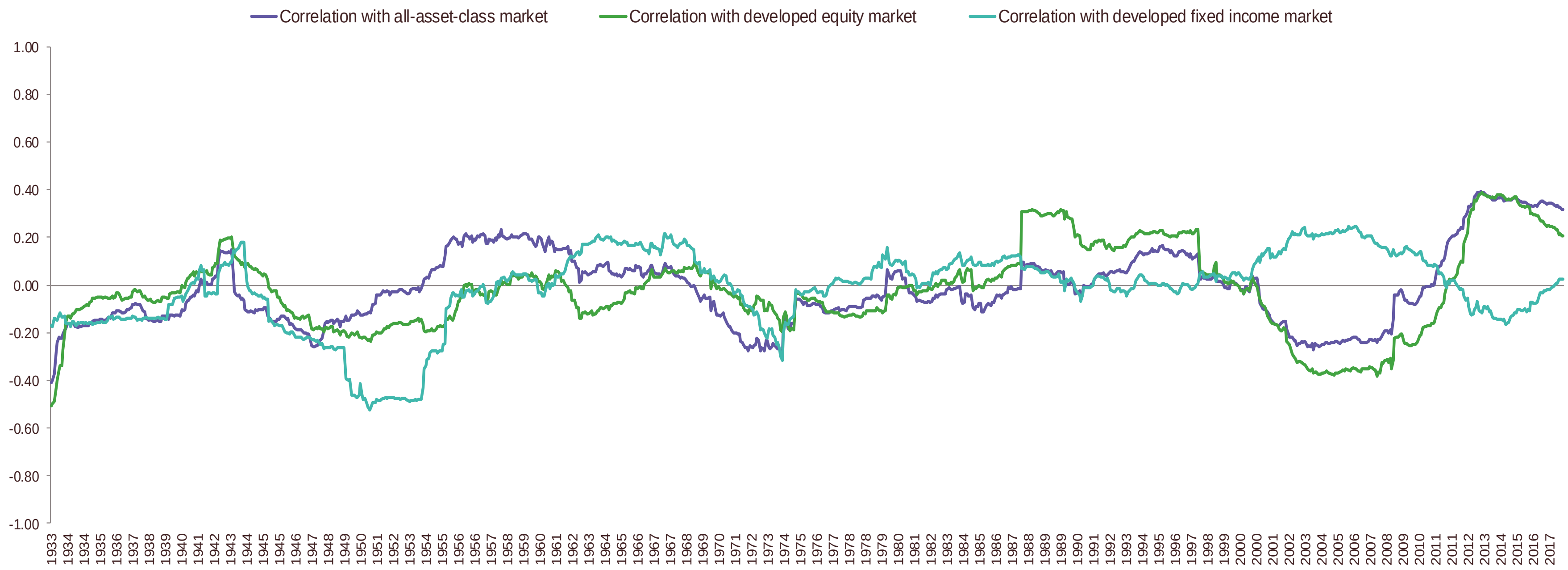
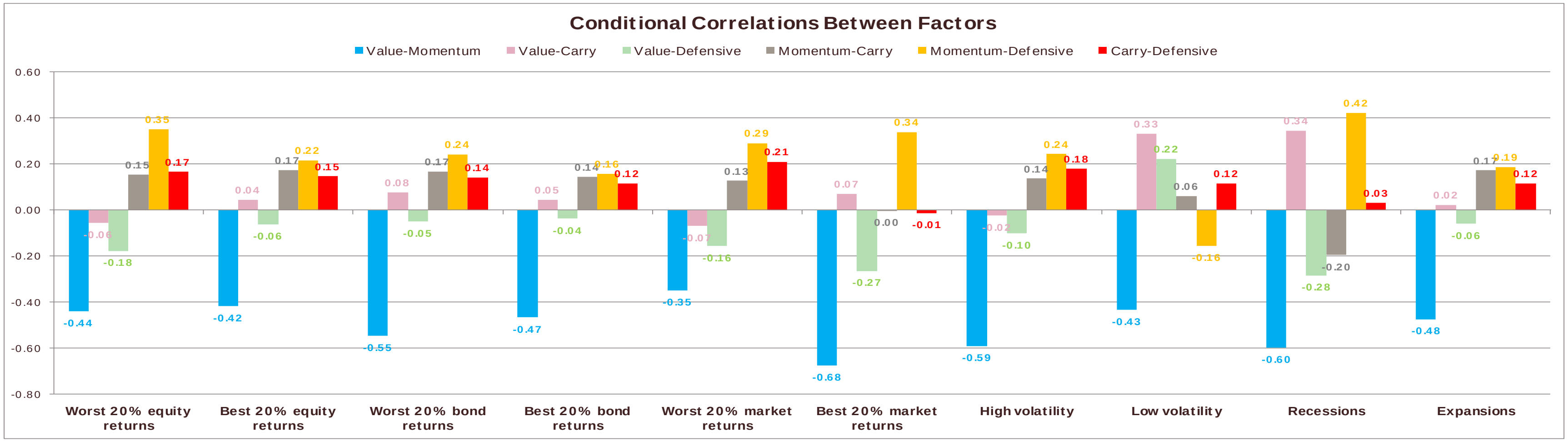


Figure 5. Correlations in Different Market Environments. The figure plots the correlations between factors and the correlations across asset classes for a given factor in different economic environments. We separately compute correlations for the 20 percent worst and best months of global equity returns (using the MSCI index), the 20 percent worst and best months of global bond returns (using the Barclays Aggregate Bond Index), the 20 percent worst and best global market returns (using a volatility-weighted average of all asset classes that includes stocks, bonds, stock indices, currencies, and commodities), the top and bottom 20 percent of months based on equity market volatility (realized volatility over the last 36 months), as well as during global recessions and expansions using the NBER's business cycle definitions applied to all developed markets in our sample. The first graph plots the six pairwise correlations between factors across all asset classes in each economic environment. The second graph repeats the same exercise looking at average pairwise correlations across asset classes for a given factor.



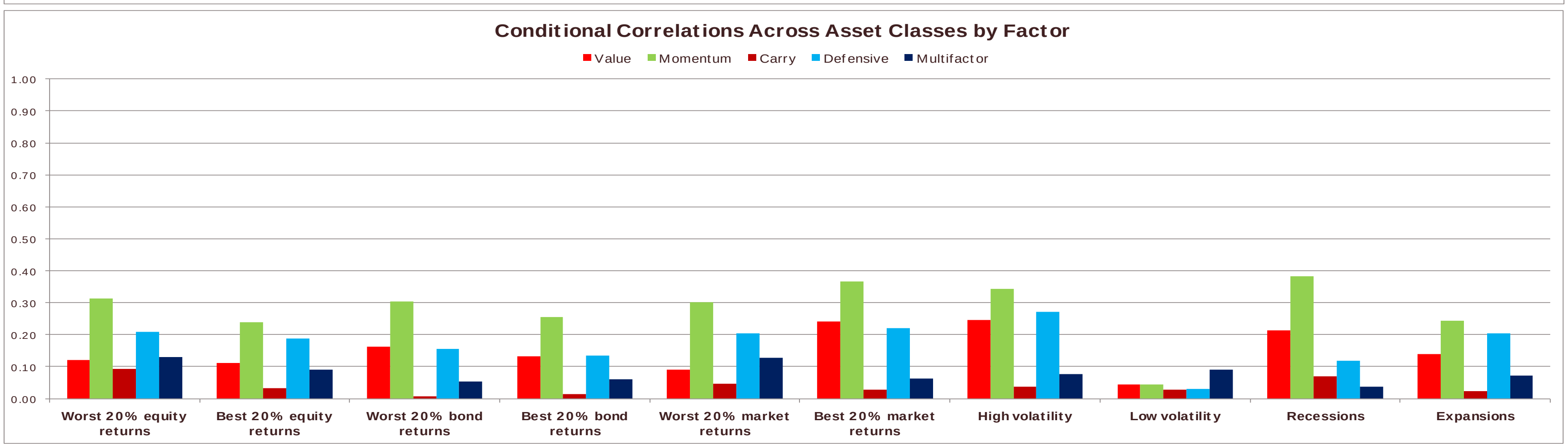


Table 5: Value Spread Factor Timing Across Asset Classes and Factors

The table reports out of sample performance of value spread timing strategies for each factor in each asset class, as well as across asset classes. The timing method is a simple *z*-score with caps at +2, -2. The *z*-scores are estimated using an expanding window to measure the median and absolute deviation from the median. The annual Sharpe ratio to the timing strategies are reported, along with univariate regression alpha and multivariate regression alpha information ratios, where we regress the returns of the timing strategy to the static underlying factors, with *t*-statistics of the mean in parentheses. For comparison, the first column in each panel reports the annualized Sharpe ratio and *t*-stat of the mean return (in parentheses) of the static factor strategy that contains no timing component. The last row of each panel reports results for value spread timing on the multifactor portfolio that is diversified across all factors. The all asset classes combine strategies using an equal risk combination of the asset classes, weighted by inverse volatility, estimated over the full sample period.

Sharpe and information ratios (*t*-statistics in parentheses)

	US stocks				International stocks				Equity index futures			
		Value spread timing				Value spread timing				Value spread timing		
Factor	No timing	Raw Sharpe ratio	Info ratio (univariate)	Info ratio (multivariate)	No timing	Raw Sharpe ratio	Info ratio (univariate)	Info ratio (multivariate)	No timing	Raw Sharpe ratio	Info ratio (univariate)	Info ratio (multivariate)
Value	0.30	0.17	-0.07	0.05	0.53	0.27	-0.45	-0.86	0.33	0.07	0.06	0.18
	(2.88)	(1.56)	-(0.62)	(0.46)	(3.17)	(1.38)	-(2.25)	-(4.31)	(3.17)	(0.64)	(0.51)	(1.65)
Momentum	0.50	0.25	0.59	0.40	0.68	-0.05	0.34	-0.48	0.46	0.12	0.19	0.09
	(4.77)	(2.27)	(5.31)	(3.65)	(4.00)	-(0.25)	(1.71)	-(2.38)	(4.53)	(1.08)	(1.72)	(0.86)
Carry									0.25	-0.08	-0.08	-0.04
									(2.43)	-(0.71)	-(0.72)	-(0.41)
Defensive	0.73	0.75	0.47	0.43	1.01	0.55	0.13	-0.16	0.40	-0.08	0.08	0.08
	(6.92)	(6.62)	(4.16)	(3.77)	(5.77)	(2.61)	(0.63)	-(0.77)	(3.92)	-(0.75)	(0.71)	(0.70)
Multifactor	1.14	0.42	0.33	0.40	1.43	0.21	-0.06	-0.74	0.72	0.02	0.08	0.14
	(11.00)	(3.84)	(3.04)	(3.61)	(8.49)	(1.04)	-(0.30)	-(3.72)	(7.04)	(0.20)	(0.78)	(1.32)

	Commodities				Fixed income				Currencies			
		Value spread timing				Value spread timing				Value spread timing		
Factor	No timing	Raw Sharpe ratio	Info ratio (univariate)	Info ratio (multivariate)	No timing	Raw Sharpe ratio	Info ratio (univariate)	Info ratio (multivariate)	No timing	Raw Sharpe ratio	Info ratio (univariate)	Info ratio (multivariate)
Value	0.26	0.32	0.18	0.25	0.36	0.11	-0.08	-0.25	0.64	0.02	0.02	0.04
	(2.59)	(2.99)	(1.72)	(2.29)	(3.55)	(1.03)	-(0.70)	-(2.35)	(4.26)	(0.12)	(0.09)	(0.22)
Momentum	0.51	-0.05	0.26	0.13	0.09	0.07	0.13	-0.17	0.18	0.17	0.21	-0.02
	(5.05)	-(0.45)	(2.41)	(1.21)	(0.88)	(0.64)	(1.19)	-(1.61)	(1.20)	(1.03)	(1.22)	-(0.15)
Carry	0.53	-0.32	-0.05	-0.23	0.63	0.04	-0.14	-0.24	0.43	0.16	0.53	0.39
	(5.32)	-(3.04)	-(0.46)	-(2.12)	(6.18)	(0.39)	-(1.29)	-(2.21)	(2.92)	(0.97)	(3.12)	(2.31)
Defensive					0.04	0.16	0.19	0.08				
					(0.41)	(1.50)	(1.79)	(0.71)				
Multifactor	0.86	0.01	0.14	0.08	0.55	0.13	-0.05	-0.23	0.64	0.20	0.43	0.22
	(8.56)	(0.11)	(1.34)	(0.72)	(5.37)	(1.21)	-(0.51)	-(2.09)	(4.32)	(1.20)	(2.52)	(1.30)

	All asset classes				Exclude individual stock factors			
		Value spread timing				Value spread timing		
Factor	No timing	Raw Sharpe ratio	Info ratio (univariate)	Info ratio (multivariate)	No timing	Raw Sharpe ratio	Info ratio (univariate)	Info ratio (multivariate)
Value	0.70	0.31	-0.10	0.13	0.65	0.25	-0.01	0.14
	(6.83)	(2.86)	-(0.89)	(1.22)	(6.34)	(2.34)	-(0.07)	(1.27)
Momentum	0.66	0.18	0.58	0.15	0.54	0.11	0.35	0.04
	(6.46)	(1.65)	(5.35)	(1.43)	(5.27)	(1.01)	(3.29)	(0.35)
Carry	0.85	-0.19	-0.04	-0.16	0.85	-0.19	-0.04	-0.16
	(8.35)	-(1.72)	-(0.36)	-(1.49)	(8.35)	-(1.72)	-(0.36)	-(1.49)
Defensive	0.75	0.57	0.61	0.51	0.32	0.07	0.20	0.11
	(7.39)	(5.30)	(5.68)	(4.72)	(3.10)	(0.61)	(1.88)	(1.00)
Multifactor	1.64	0.30	0.42	0.28	1.21	0.12	0.19	0.11
	(16.04)	(2.83)	(3.85)	(2.58)	(11.83)	(1.07)	(1.74)	(1.01)

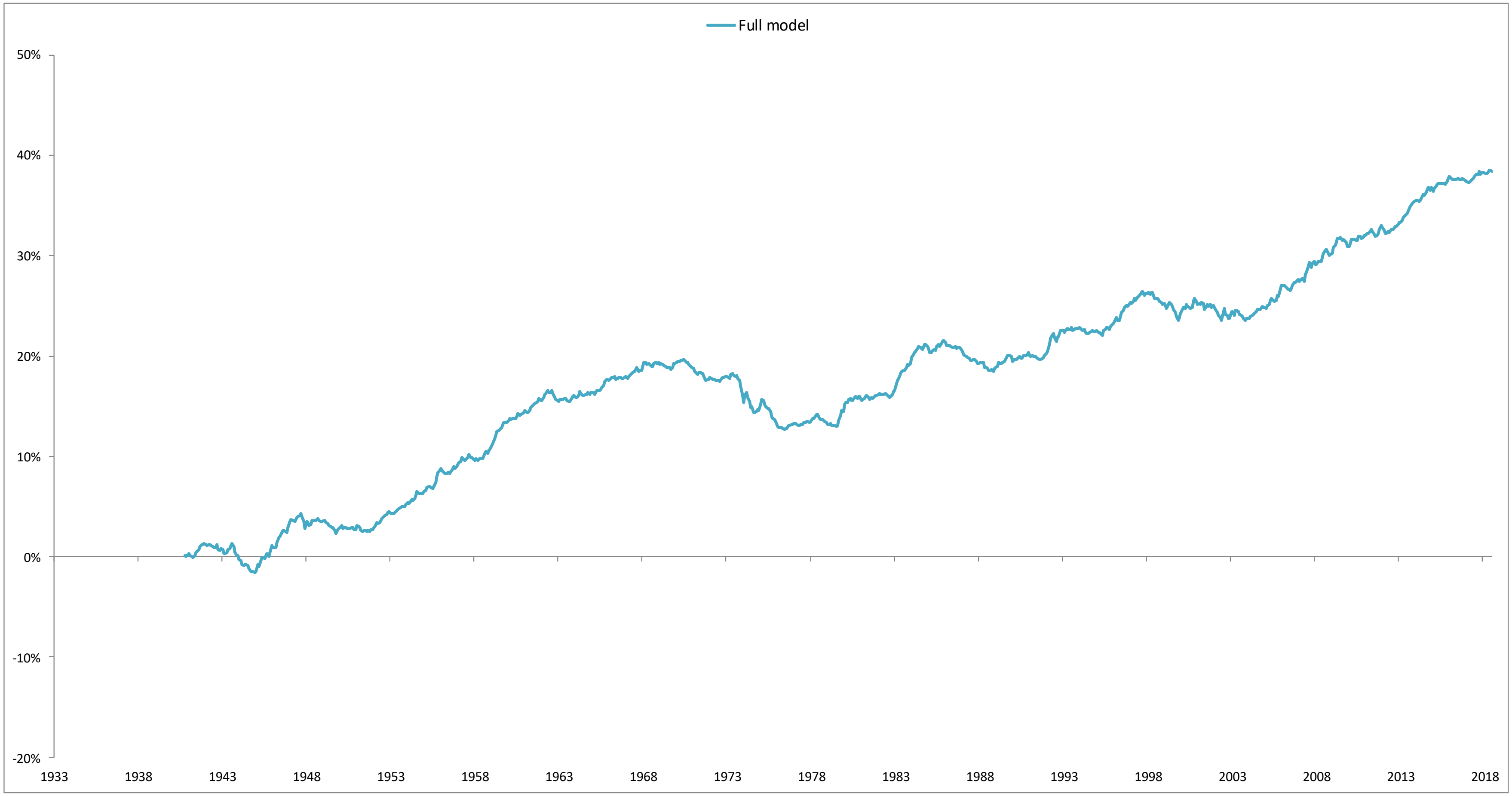
Table 6: Value Spread Factor Timing Using Different Methodologies

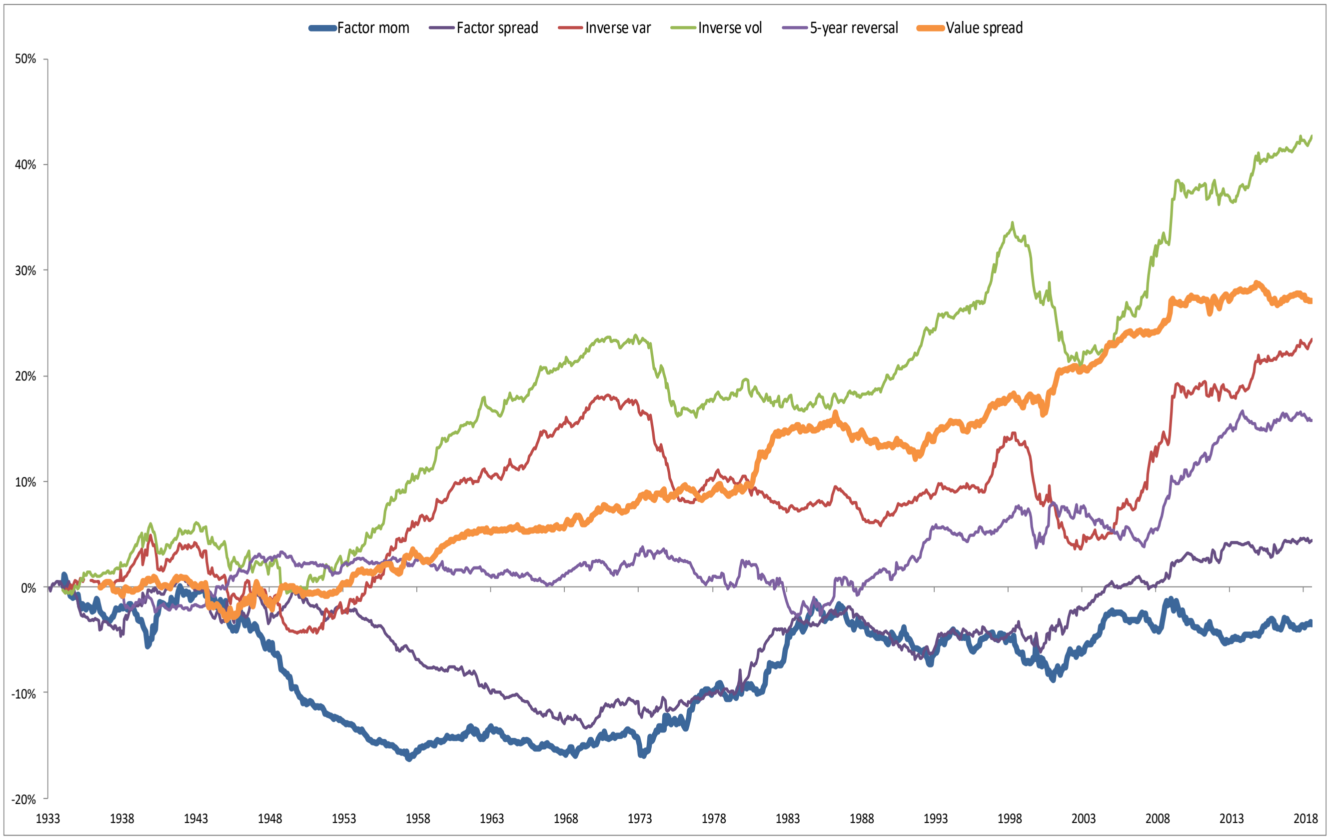
The table reports in and out of sample value spread factor timing for the multi-factor, all asset class implementation of these timing strategies using various timing methodologies that include: a *z*-score to time the strategy, a fitted regression of the strategy's returns on value spreads that uses the coefficient from that regression to time the factor out of sample, and the PCA timing strategy of Haddad, Kozak, and Santosh (2018). The first two rows report the results for the *z*-score timing methodology both out of sample and in-sample, where the first row corresponds to the last row of Table 5. The next 15 rows report results from timing strategies based off of regressions which are run both over the full sample as well as an expanding window that relies only on ex ante information (out of sample) as indicated below. Regressions are run allowing the coefficient on the value spread to vary across factors and assets ("individual asset") as well as forcing the coefficients to be the same for a given factor ("factor") and forcing the coefficients to be the same for every asset and factor ("pooled"). In some cases, an economic sign constraint is imposed on the coefficient (e.g., Campbell and Thompson (2007)), where value spreads should positively predict returns, so that if the regression coefficient is negative it is set to zero. In some cases, the standardized measures, whether z-scores or regression weights, are capped at +2, -2 (identified in the "Weight" column). We report annualized Sharpe ratios of the raw returns, the information ratio of the alpha of each timing strategy with respect to all static factors (value, momentum, carry, and defensive) applied across all asset classes.

	Timing methodology	Variation by	Sample	Economic constraints	Weight	Raw Sharpe ratio	Multivariate alpha info. ratio
(1)	Z-score	Individual asset	Out of sample	Yes	Capped	0.30	0.28
(2)	Z-score	Individual asset	Full sample	Yes	Capped	0.42	0.33
(3)	Regression	Individual asset	Out of sample	No	Raw	1.36	0.28
(4)	Regression	Individual asset	Out of sample	No	Z-score	0.27	0.29
(5)	Regression	Individual asset	Out of sample	No	Z, capped	0.23	0.35
(6)	Regression	Individual asset	Out of sample	Yes	Z, capped	0.31	0.41
(7)	Regression	Individual asset	Full sample	No	Z, capped	0.42	0.33
(8)	Regression	Factor	Out of sample	No	Raw	1.39	0.16
(9)	Regression	Factor	Out of sample	No	Z-score	0.31	0.16
(10)	Regression	Factor	Out of sample	No	Z, capped	0.28	0.21
(11)	Regression	Factor	Out of sample	Yes	Z, capped	0.41	0.39
(12)	Regression	Factor	Full sample	No	Z, capped	0.41	0.30
(13)	Regression	Pooled	Out of sample	No	Raw	1.49	0.10
(14)	Regression	Pooled	Out of sample	No	Z-score	0.36	0.24
(15)	Regression	Pooled	Out of sample	No	Z, capped	0.26	0.19
(16)	Regression	Pooled	Out of sample	Yes	Z, capped	0.32	0.29
(17)	Regression	Pooled	Full sample	No	Z, capped	0.36	0.26
(18)	PCA	Individual asset	Out of sample	No	Raw	-0.08	0.13
(19)	PCA	Individual asset	Full sample	No	Raw	0.12	0.45

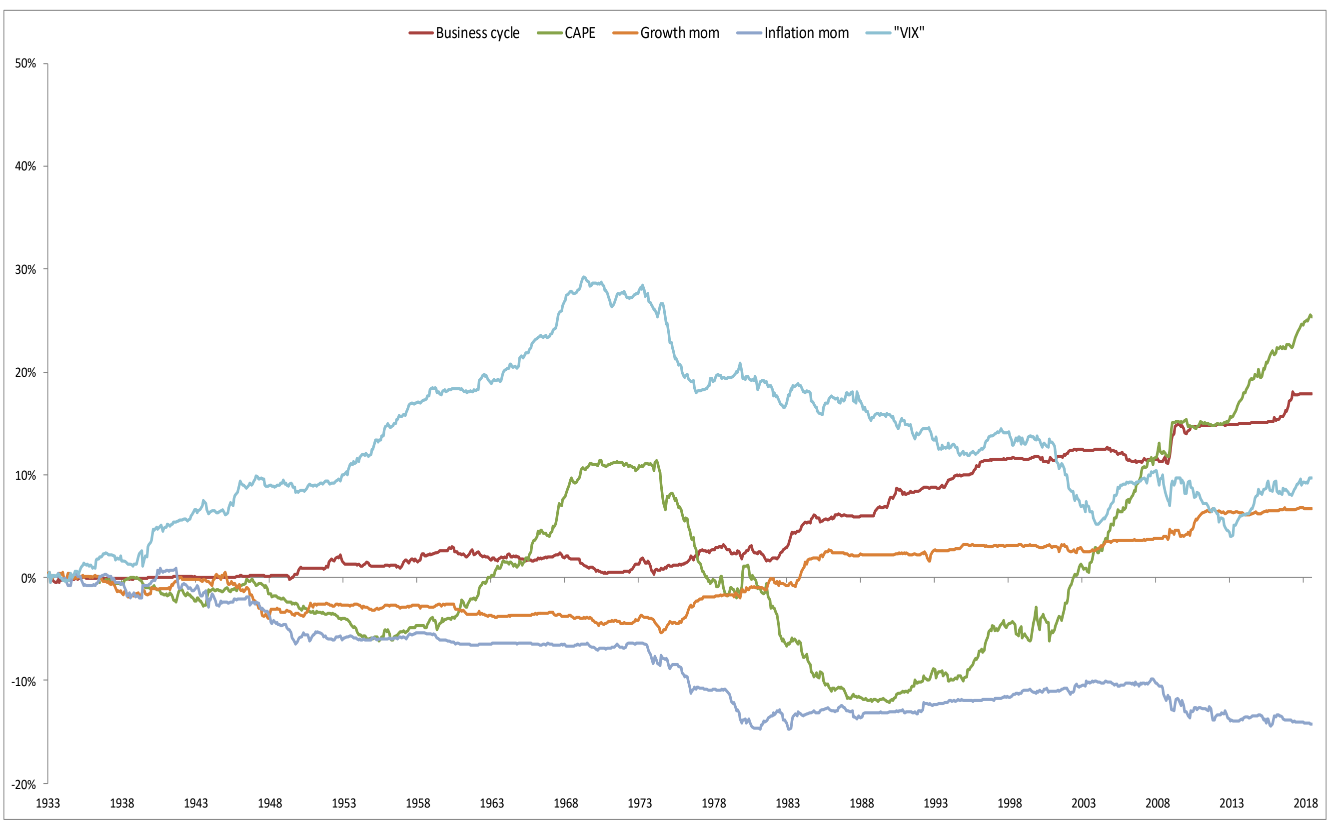
Figure 6. Other Factor Timing Variables. The figure reports timing portfolio information ratios relative to the multivariate static factors in each asset class for different timing variables using different timing methodologies. We use the following timing methodologies: z-score, out-of-sample regression with no sign restrictions, out-of-sample regression with the restriction that a given factor have the same sign across all asset classes for a given timing variable, out-of-sample regression with both an economic sign and factor restriction (e.g., value spreads should have a positive coefficient and be the same for a given factor across all asset classes), and finally a full sample regression (in-sample) that places no restrictions on any coefficients. The timing variables are the value spread, past 12-month return on each factor (factor momentum), average characteristic of the factor itself (factor spread), which is the value spread for value portfolios, the momentum for momentum portfolios, the average carry for the carry portfolios, and the average negative beta for the defensive portfolios, five-year reversals (negative of past five year return on the factor), inverse volatility (where volatility is estimated over the prior 36 months of returns), inverse variance, business cycle (an ex ante measure that seeks to identify stages of the business cycle – contraction, recovery, slowdown, expansion – where we use both the level and change of GDP growth, compute a rolling 10-year z-score of level and changes in GDP growth, and identify the turning point of a business cycle as whether the z-score breaks +/-1.0 to identify each of the four periods), growth momentum (moving average of annual GDP growth), inflation momentum (moving average of inflation growth), CAPE, and "VIX" (realized volatility of the market over the last 36 months). We also report a simple average of all the timing strategies as well as a timing strategy based on the "full model" that incorporates all timing variables into one model, under each methodology, to time the factors. Highlighted in red is the out-of-sample regression approach using sign restrictions on the timing variables that we use for the full model going forward.

Figure 7. Time-Series of Returns to Timing Strategies. The figure plots the cumulative time-series of multivariate alphas to the out-of-sample regression timing methodology using the full model of timing variables with economic sign restrictions (specification (6)), applied to all factors and asset classes and scaled to ex post 10% volatility, over time. The second graphs plot the cumulative returns to the same timing methodology applied to each of the timing variables individually and applied to all factors and asset classes and scaled to ex post 10% annual volatility.





Factor mom
Factor spread
Inverse var
Inverse vol
5-year reversal
Value spread
50%
40%
30%
20%
10%
0%
-10%
-20%
1933
1938
1943
1948
1953
1958
1963
1968
1973
1978
1983
1988
1993
1998
2003
2008
2013
2018



Business cycle
CAPE
Growth mom
Inflation mom
"VIX"
50%
40%
30%
20%
10%
0%
-10%
-20%
1933
1938
1943
1948
1953
1958
1963
1968
1973
1978
1983
1988
1993
1998
2003
2008
2013
2018

Table 7: Economic Impact of Factor Timing

The table reports the Sharpe ratio, information ratio to the static factor portfolio, ex-post optimal Sharpe ratio of combining the static and timing portfolio, ex-post weight placed on the timing portfolio, and the turnover per dollar leverage and break-even trading cost per dollar traded for adding the timing portfolio to the static portfolio at the ex-post optimal weight. These statistics are reported for timing strategies that use the full model with no restrictions estimated over the full sample (in-sample), for the same model estimated out of sample using an expanding window of data up to time *t-1*, for the full model out of sample with economic sign restrictions on the coefficients as well as a restriction that the same factor across asset classes have the same coefficient, and the full model out of sample with only economic sign restrictions on the timing variable coefficients, but allowing different coefficients for the same factor across asset classes.. This last timing strategy (the one we like best) is further broken down by each individual timing variable separately and the statistics reported for each timing variable. The static (no timing) factor portfolio across all asset classes is also reported for comparison in the first row.

	In/Out-of-sample	Sharpe ratio	Information ratio	Sharpe ratio static + timing (ex-post optimal)	Ex-post optimal weight on timing	Turnover per $ leverage	Break-even trading cost per dollar traded for adding timing (in bps)
Static (no timing)	OOS	1.64	0.00	1.64	0.0%	4.3	
Timing strategies							
Full model, no restrictions	IS	1.28	1.10	2.00	40.2%	6.2	9.83
Full model, no restrictions	OOS	0.37	0.41	1.69	20.0%	7.9	2.49
Full model, economic sign restrictions	OOS	0.39	0.62	1.79	27.6%	6.8	5.03
Value spread	OOS	0.31	0.41	1.71	20.1%	4.7	3.92
Factor momentum	OOS	0.36	-0.04	1.64	0.0%	5.6	0.00
Factor spread	OOS	0.34	0.05	1.64	3.2%	5.1	0.04
Five year reversal	OOS	-0.03	0.21	1.67	11.4%	5.2	2.64
Inverse volatility	OOS	0.34	0.40	1.69	19.4%	6.1	2.41
Inverse variance	OOS	0.27	0.24	1.65	12.6%	6.2	1.32
Business cycle	OOS	0.35	0.45	1.71	21.6%	5.4	3.41
Growth momentum	OOS	0.12	0.16	1.65	8.9%	6.0	1.25
Inflation momentum	OOS	-0.27	-0.25	1.64	0.0%	7.0	0.00
CAPE	OOS	0.62	0.29	1.64	15.1%	4.6	0.31
"VIX"	OOS	-0.41	0.11	1.65	6.3%	6.2	2.93

Internet Appendix: Supplemental Tables and Figures

Data description and sources

Global equity indices

We obtain returns on equity indices from 43 equity markets internationally from Global Financial Data, which include all countries covered in the MSCI World Index as of April 2018. Since most countries have multiple equity indices at each point in time, we select the index that is investible, has the most representative coverage of the total stock market of that country, and has the longest history.

We use monthly index total returns (including returns from dividends) from Global Financial Data and subtract the local currency cash returns to get excess returns. We also use monthly futures returns from Bloomberg and Datastream, which covers a shorter history, to supplement these data.

Nine of the stock indices have data going back to the 1920s, most of the rest of the developed equity markets have index data going back prior to the 1950s, 1960s, or early 1970s, and emerging markets go back in some cases to the 1970s, most in the mid to late 1980s, and two countries in 1991. All countries have returns up through April 2018, so the minimum history is 27 years (Poland) and the maximum is 98 years (AU, BD, FN, FR, SD, US, UK) of monthly returns.

Panel A of Table IA1 in the internet appendix reports summary statistics on the country indices covered, including the start dates for each country's index returns, the annualized mean and standard deviation of returns, and the worst 12-month return for each index over the sample period. The latter highlights some of the extreme events these markets have experienced (e.g., Germany, Brazil) over the last century that provides more tail events and downside risks to examine.

Global fixed income

We use nominal yield and total returns data of 10-year local currency government bonds as well as 3-month interest rates from Global Financial Data and supplement it with Bloomberg and Datastream. The cross-section of government bond indices includes 26 countries, covering North America, Western and Northern Europe, Japan, and the Antipodeans.

Given the evolving nature of bond issuance historically, the inputs into our bond yield and returns can vary over time. In general, 10-year bond yield and returns only become available between 1960 and 1980. Between 1920 and 1960, the database uses the closest available tenor to 10-year, while before 1920, the yield and returns are typically for individual bonds. Using the Netherlands as an example, a series of individual bonds were used for the 10-year yield and return series before December 1917. Yields on an index of public bonds are used from 1919 through 1925. The 3% consol bond is used between January 1926 and May 1940. The 2 1/2% consol is used from 1946 until 1954, the 3 1/4% issue of 1948 is used from 1955 until October 1964, and an index of the three to five longest running issues of the Dutch government bonds begins in November 1964. Data for the 10-year bond begins in 1978. A similar splicing of various government bonds is done for each of the 26 countries we examine. Panel B of Table IA2 reports summary statistics on the bonds. Again, there is both rich heterogeneity in bond returns across countries and significant extreme events that occur over our sample period.

Global currencies

We use spot and 1-, 2-, 3-, and 6-month forward exchange rates from AQR's production data base and interpolate the forward exchange rate for the next quarterly International Money Market (IMM) date. The return series simulate a strategy of buying and holding the forward contract maturing at the nearest IMM date and rolling to the far contract five days before the maturity date. Before 1990, when the forward contract data is not available, we use changes in spot exchange rates plus the carry of the currency (difference in local interest rates) for the total return.

We cover 20 developed market currencies, including the G10 (legal tenders of Australia, Eurozone, Canada, Japan, Norway, New Zealand, Sweden, Switzerland, United Kingdom, United States) and 10 legacy European currencies (legal tenders of Belgium, Spain, Finland, France, Germany, Ireland, Italy, Netherlands, Austria, Portugal) before the Euro in 1999. The dataset starts in August 1971 when the US severed its link between the value of the U.S. dollar and gold. Panel C of Table IA1 reports summary statistics on the currency returns. In addition to large heterogeneity in mean and volatility of currency-pair returns, we also see evidence of currency crashes, which many have claimed are related to carry strategies in currencies (Brunnermeier, Nagel, and Pedersen (2008), Burnside, Eichenbaum, and Rebelo (2010), Koijen, Moskowitz, Pedersen, and Vrugt (2018)).

Commodity futures

We obtain monthly futures prices for 40 commodities starting in February 1877. The source of the data until 1951 is the Annual Report of the Trade and Commerce of the Chicago Board of Trade. Between 1951 and 2012, the futures prices across various contracts are provided by Commodity Systems, Inc. After 2012, the futures prices are from Bloomberg. For base metals and platinum, rolled return series from the S&P, Goldman Sachs, and Bloomberg are used.

The total returns are the sum of spot returns and the "roll-down" on the futures curve. The methodology for computing total returns is as follows. At each month end, we calculate the return on each contract from the previous month end. For each month, we hold the nearest of the contracts whose delivery month is at least two months away.[36] For months in which the desired contract does not have a return, we move to the next contract and follow the same procedure until there is a return or until we reach the fifth contract. If there is still no return, we then hold the contract in front of the desired contract. Note that there are days with limit moves in various grains contracts, and we assume all contiguous limit moves are incorporated into the first move price.[37] This methodology for calculating commodity returns is the same as that used in Moskowitz, Ooi, and Pedersen (2012) and Koijen, Moskowitz, Pedersen, and Vrugt (2016). The cross section of commodities covers energy, base metal, precious metal, agricultural, and livestock sectors, where Panel D of Table IA1 in the internet appendix reports their summary statistics.

U.S. stock selection

We also supplement the data above with nearly a century's worth of factor return data in U.S. individual equities. The data come from CRSP and begin in July 1926. The U.S. equity data is well known from many studies, so we do not report summary statistics here.

[36] For example, we hold an April contract through the end of February. An exception is Brent oil, whose delivery month needs to be at least three months away, i.e. we hold the April contract through the end of January. This methodology is chosen to coincide with the procedure employed by the popular Goldman Sachs Commodity Index.

[37] For limit day periods, we incorporate all the limit day returns into the first limit day following Roll (1984) and Boudoukh, Richardson, Shen, and Whitelaw (2007). Limit days are determined by whether on that day (i) the maximum price shift across contracts of the same commodity is a round amount (before closing prices are available, the largest positive shift from high price and the largest negative shift from low price are used) (ii) two or more contracts move by this amount, (iii) if maximum price shift is from the front contract and does not meet above conditions, maximum shift of the other contracts meets above conditions (since sometimes the front contract is not subject to limits if it is considered "spot"), and (iv) this shift is equal to or higher than the official price limit set by the exchange (when available).

International stock selection

We also examine international individual equities across 21 developed stock markets (those from Frazzini, Israel, and Moskowitz (2018)), but note that the longest sample of international individual equity returns begins in 1972 and for our factors (described below) the earliest data point is July 1984. Despite the limited time-series, the international equity data provide another asset class to examine the robustness of many of our results.

Cash returns

We use the 3-month local-currency T-bill yield and returns from Global Financial Data as the risk-free cash returns. For a more recent period when LIBOR rates become available, we use the 3-month ICE LIBOR rates or the closest equivalent.

Table IA1: Summary Statistics on Assets

Summary statistics on every asset in our sample, excluding individual stocks, are reported. The in-sample mean return and standard deviation, worst 12-month return, and sample start dates are reported for country equity indices (Panel A), fixed income (Panel B), currencies (Panel C), and commodities (Panel D).

Panel A: Country Equity Indices				
Country	**Mean (annualized)**	**Stdev (annualized)**	**Worst 12-month return**	**Start date**
AR	26.2%	64.2%	-68.3%	1/29/1988
AU	8.2%	14.8%	-43.9%	2/27/1920
BD	9.9%	43.5%	-94.2%	2/27/1920
BG	5.9%	15.9%	-66.1%	1/31/1951
BR	-50.4%	71.8%	-100.0%	7/30/1965
CB	7.7%	32.7%	-92.6%	1/29/1988
CH	6.0%	35.4%	-80.3%	2/29/1996
CL	18.6%	35.1%	-59.7%	2/28/1975
CN	6.6%	14.8%	-47.2%	1/31/1934
DK	7.2%	17.8%	-45.5%	1/30/1970
ES	7.0%	18.5%	-43.4%	4/30/1940
FN	8.0%	23.8%	-61.8%	2/27/1920
FR	8.5%	21.6%	-50.6%	2/27/1920
GR	3.0%	34.7%	-66.8%	1/31/1977
HK	15.7%	32.6%	-77.2%	1/30/1970
HN	7.9%	32.0%	-63.0%	1/31/1991
ID	12.6%	39.3%	-64.2%	1/29/1988
IN	15.1%	29.9%	-52.2%	1/29/1988
IR	3.6%	21.1%	-72.1%	1/29/1988
IS	8.7%	20.3%	-47.3%	2/28/1986
IT	7.2%	26.5%	-54.4%	1/30/1925
JP	8.7%	21.4%	-47.6%	1/31/1921
KO	17.2%	46.6%	-76.0%	2/28/1962
MX	11.1%	24.6%	-45.7%	1/29/1988
MY	8.3%	26.3%	-56.7%	12/29/1972
NL	7.9%	17.5%	-54.1%	1/31/1951
NW	7.7%	24.1%	-56.1%	1/30/1970
NZ	0.5%	19.5%	-58.9%	7/31/1986
OE	4.4%	20.4%	-68.8%	1/30/1970
PH	8.9%	28.8%	-56.7%	1/29/1982
PO	10.0%	41.6%	-77.9%	5/31/1991
PT	-0.3%	20.3%	-51.7%	1/29/1988
RS	7.6%	45.3%	-92.7%	1/31/1995
SA	10.5%	21.4%	-42.9%	2/29/1960
SD	7.7%	17.4%	-52.4%	2/27/1920
SG	8.7%	26.9%	-57.9%	1/30/1970
SW	6.2%	16.3%	-40.9%	2/28/1966
TA	8.2%	33.4%	-74.5%	10/30/1987
TH	9.8%	30.7%	-68.1%	5/30/1975
TK	15.3%	53.3%	-74.2%	2/28/1986
UK	6.5%	16.4%	-61.8%	2/27/1920
US	8.0%	18.6%	-68.3%	2/27/1920
VE	31.3%	50.1%	-75.1%	1/29/1988

Panel B: Global Fixed Income				
Country	**Mean (annualized)**	**Stdev (annualized)**	**Worst 12-month return**	**Start date**
AR	-15.5%	19.3%	-76.7%	5/31/1991
AU	1.9%	7.5%	-26.6%	2/27/1920
BD	2.4%	6.1%	-19.3%	1/31/1924
BG	1.8%	5.5%	-19.9%	2/27/1920
CN	2.1%	5.8%	-23.0%	1/31/1934
DK	1.9%	6.9%	-20.7%	2/27/1920
ES	1.4%	6.4%	-21.2%	2/27/1920
FR	1.8%	6.5%	-24.2%	2/27/1920
HK	2.0%	6.2%	-14.6%	6/30/1993
IN	-0.7%	6.9%	-20.0%	2/27/1920
IS	1.8%	3.9%	-10.8%	11/30/1993
IT	0.9%	8.5%	-36.2%	2/27/1920
JP	2.5%	7.4%	-19.7%	2/27/1920
KO	9.0%	18.0%	-28.2%	1/31/1957
MX	3.9%	7.7%	-12.0%	1/31/1995
MY	2.2%	7.2%	-19.5%	1/31/1961
NL	2.1%	7.3%	-21.5%	2/27/1920
PH	8.6%	15.7%	-34.6%	9/30/1996
PO	2.9%	9.4%	-24.6%	5/31/1999
SA	2.9%	10.3%	-35.0%	2/29/1960
SD	1.8%	5.3%	-17.2%	2/27/1920
SG	1.1%	3.9%	-8.1%	12/31/1987
TA	2.9%	5.1%	-11.4%	1/31/1995
TH	3.9%	12.1%	-30.7%	12/31/1979
UK	1.3%	3.8%	-15.3%	1/31/1933
US	2.0%	6.2%	-19.7%	2/27/1920

Panel C: Currencies				
Country	Mean (annualized)	Stdev (annualized)	Worst 12-month return	Start date
AR	7.4%	17.9%	-56.0%	4/30/1991
AU	1.8%	11.1%	-28.7%	1/31/1972
BD	0.6%	10.9%	-31.0%	1/29/1971
BG	1.4%	12.3%	-34.9%	12/30/1977
BR	10.2%	18.8%	-36.4%	7/29/1994
BU	-0.2%	10.6%	-22.1%	3/31/2005
CB	2.8%	11.5%	-35.4%	2/28/1992
CH	1.1%	2.1%	-3.7%	2/29/1996
CL	-8.5%	12.2%	-52.1%	6/30/1982
CN	0.1%	6.7%	-22.5%	1/31/1972
CZ	1.8%	11.8%	-24.4%	4/30/1993
ES	0.1%	10.9%	-29.3%	12/30/1977
FN	1.2%	11.3%	-29.8%	12/30/1977
FR	2.2%	11.2%	-30.3%	1/31/1973
GR	5.9%	10.5%	-13.2%	10/30/1987
HK	-5.5%	3.8%	-32.1%	12/30/1977
HN	3.3%	12.6%	-25.8%	8/31/1989
ID	0.0%	20.6%	-80.7%	12/31/1981
IN	0.7%	7.1%	-19.3%	1/29/1993
IR	2.7%	10.7%	-19.5%	1/31/1989
IS	0.0%	8.2%	-23.6%	1/31/1986
IT	2.1%	11.0%	-27.5%	6/30/1978
JP	0.8%	11.4%	-29.0%	1/29/1971
KO	-1.9%	12.1%	-51.1%	1/31/1980
MX	3.5%	23.1%	-78.7%	1/31/1980
MY	-4.0%	8.0%	-44.6%	11/28/1986
NL	1.0%	11.4%	-32.2%	1/31/1974
NW	0.6%	10.9%	-25.5%	12/30/1977
NZ	2.5%	12.2%	-34.2%	12/30/1977
OE	1.3%	11.5%	-30.3%	2/28/1974
PH	-2.3%	10.6%	-40.1%	1/31/1980
PO	3.8%	12.8%	-34.9%	1/31/1992
PT	4.2%	10.9%	-22.4%	3/31/1989
RS	6.3%	17.4%	-55.5%	9/30/1994
SA	-5.0%	14.2%	-41.3%	11/30/1972
SD	-0.4%	11.0%	-32.0%	12/30/1977
SG	0.3%	5.5%	-18.3%	3/31/1986
SW	1.1%	12.0%	-29.1%	1/29/1971
TA	-1.8%	5.8%	-19.2%	1/31/1985
TH	0.4%	9.8%	-49.3%	5/31/1988
TK	8.7%	15.6%	-32.2%	9/30/1996
UK	0.7%	10.0%	-27.3%	1/29/1971
US	0.0%	0.0%	0.0%	1/29/1971
VE	-6.2%	26.0%	-61.2%	1/31/1989

Panel D: Commodities				
Commodity	**Mean (annualized)**	**Stdev (annualized)**	**Worst 12-month return**	**Start date**
ALUMINUM	-0.3%	22.8%	-60.4%	2/28/1979
BRENTOIL	9.7%	31.5%	-63.8%	7/29/1988
CATTLE	4.0%	16.7%	-40.6%	12/31/1964
COCOA	4.0%	31.4%	-53.9%	2/26/1965
COFFEE	4.7%	37.4%	-61.9%	9/29/1972
COMEXCOPPER	9.2%	25.8%	-60.3%	8/31/1988
COPPER	7.2%	26.5%	-58.5%	2/28/1977
CORN	3.4%	26.1%	-60.7%	02/28/1877
COTTON	4.3%	24.3%	-59.5%	2/27/1925
CRUDE	6.6%	33.6%	-70.0%	4/29/1983
FEEDERCATTLE	3.7%	16.6%	-50.3%	12/31/1971
FLAX	6.9%	23.5%	-17.4%	09/30/1890
GASOIL	7.6%	30.8%	-64.2%	5/29/1981
GOLD	5.0%	20.8%	-45.7%	1/30/1970
HEATOIL	7.0%	32.4%	-62.4%	12/29/1978
HOGS	2.3%	25.8%	-53.0%	3/31/1966
KANSASWHEAT	1.8%	25.9%	-56.0%	6/30/1966
LARD	0.5%	24.6%	-59.0%	02/28/1877
LEAD	7.6%	28.7%	-69.3%	2/28/1995
LONDONCOCOA	5.8%	31.2%	-49.1%	2/29/1968
LUMBER	-6.6%	30.6%	-63.1%	2/28/1979
NATGAS	-14.6%	48.9%	-81.8%	5/31/1990
NICKEL	7.6%	34.0%	-69.3%	2/26/1993
NYMEXPLATINUM	6.6%	27.1%	-56.0%	2/28/1964
OATS	6.4%	31.4%	-60.4%	02/28/1877
ORANGEJUICE	1.2%	31.0%	-61.2%	2/28/1979
PALLADIUM	11.5%	32.0%	-67.0%	1/31/1986
PLATINUM	3.6%	23.2%	-52.8%	2/29/1984
PORK	7.1%	30.2%	-51.1%	02/28/1877
RYE	2.4%	33.1%	-68.7%	02/28/1877
SHORTRIBS	12.1%	24.9%	-53.8%	02/28/1877
SILVER	3.4%	31.0%	-70.8%	7/31/1963
SOYBEANS	9.1%	25.6%	-44.7%	2/26/1937
SOYMEAL	12.4%	30.6%	-60.9%	9/28/1951
SOYOIL	9.6%	30.3%	-58.1%	8/31/1950
SUGAR	3.6%	41.6%	-75.1%	2/26/1965
TIN	8.0%	24.0%	-47.6%	2/28/1995
UNLEADED	12.6%	35.3%	-61.5%	1/31/1985
WHEAT	2.2%	24.9%	-57.6%	02/28/1877
ZINC	1.7%	24.8%	-62.0%	2/28/1991

Table IA2: F-tests of Return Premia Across Assets and Factors

Panel A reports an *F*-test for whether each individual factor delivers the same premium per unit of risk across asset classes and whether a multifactor portfolio generates different Sharpe ratios across asset classes, where we both include and exclude international stock and currencies since the latter two have more limited samples (starting in 1974 and 1984, respectively). We also report *F*-tests for whether factor premia are different between the stock and non-stock asset classes. Panel B reports *F*-tests of equal Sharpe ratios across factors within each asset class.

Panel A: *F*-test of equal Sharpe ratios across asset classes for a given factor

	Value	Momentum	Carry	Defensive	Multifactor
Across all asset classes					
F-stat	2.7	3.7	1.3	7.6	7.0
p-value	0.019	0.002	0.272	0.000	0.000
Excluding currencies and international stocks					
F-stat	0.4	6.5	4.5	18.9	9.2
p-value	0.747	0.000	0.011	0.000	0.000
Stock versus non-stock					
F-stat	0.8	13.5		20.6	17.5
p-value	0.361	0.000		0.000	0.000
Excluding currencies and international stocks					
F-stat	1.0	12.1		32.7	26.5
p-value	0.321	0.000		0.000	0.000

Panel B: *F*-test of equal Sharpe ratios across factors within asset class

	Commodities	Equity indices	Fixed income	Currencies	US stocks	International stocks	All asset classes
F-stat	2.1	2.0	9.2	2.7	15.0	3.4	1.5
p-value	0.120	0.117	0.000	0.065	0.000	0.032	0.202

Figure IA1: Factor Return Premia by Decade. The figure plots the Sharpe ratios of each factor in each asset class decade-by-decade. The last two graphs plot Sharpe ratios by decade for the portfolio diversified across factors within each asset class ("multifactor") and for portfolios diversified across asset classes for each factor ("all assets"), respectively. The multifactor across all assets portfolios are highlighted in the last graph.

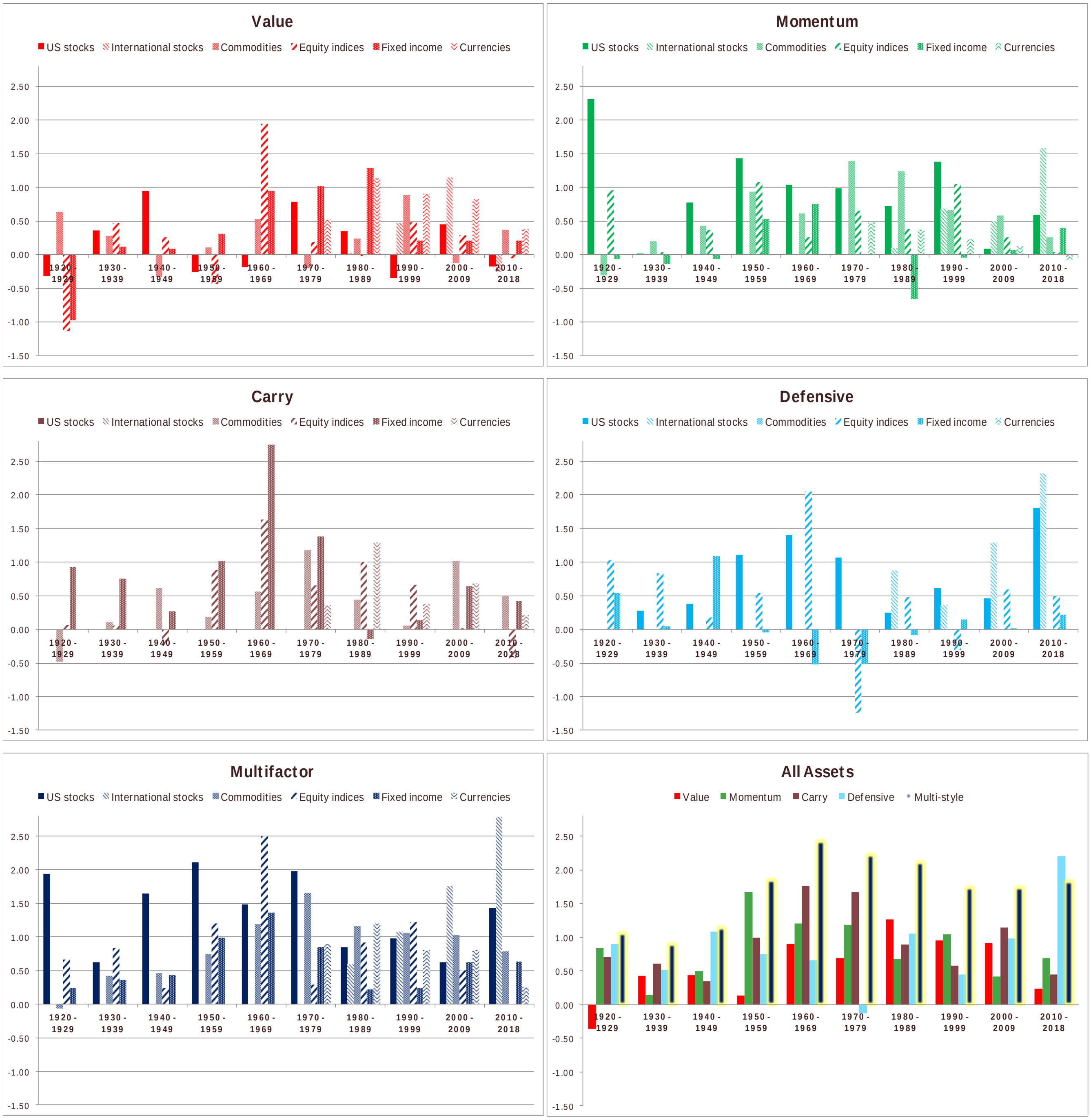


Table IA3: Test of Equal Sharpe Ratios Across Decades by Factor and Asset Class

Reported are formal F-tests of whether factor premia differ across decades in our sample for each asset class.

Tests of Equal Sharpe Ratios Across Decades			
Asset class	Factor	F-stat	p-value
All asset classes	Value	1.56	0.122
	Momentum	2.01	**0.035**
	Carry	3.39	**0.000**
	Defensive	3.11	**0.001**
	Multifactor	2.80	**0.003**
Equity indices	Value	1.36	0.199
	Momentum	1.37	0.196
	Carry	2.47	**0.009**
	Defensive	2.46	**0.009**
	Multifactor	2.52	**0.007**
Fixed income	Value	1.15	0.325
	Momentum	0.71	0.699
	Carry	2.32	**0.014**
	Defensive	1.64	0.100
	Multifactor	1.13	0.340
Currencies	Value	0.78	0.536
	Momentum	0.50	0.732
	Carry	0.95	0.437
	Multifactor	1.12	0.347
US stocks	Value	1.34	0.213
	Momentum	2.41	**0.011**
	Defensive	2.28	**0.020**
	Multifactor	5.40	**0.000**
International stocks	Value	2.51	0.058
	Momentum	0.56	0.640
	Defensive	8.25	**0.000**
	Multifactor	4.18	**0.006**
Commodities	Value	1.50	0.142
	Momentum	0.99	0.446
	Carry	1.89	0.050
	Multifactor	1.90	**0.048**

Table IA4: Macroeconomic Variable Definitions

We run two versions of analysis: 1) contemporaneous where no variable is lagged; 2) predictive where all RHS variables are lagged by 1 month.

Exposure item	Description
GDP growth	y-o-y GDP growth, averaged across the US, UK, Germany and Japan
CPI inflation	y-o-y CPI inflation, averaged across the US, UK, Germany and Japan
EQ volatility	36-month realized vol of developed EQ market returns (equal-weighted average country returns)
Tail risk dummy	1 if developed EQ market returns is in its lower 5th percentile; 0 otherwise
Geopolitical risk index	http://www.policyuncertainty.com/gpr.html
Business cycle dummies	3 dummies for 4 stages of business cycle; business cycles based on level and changes in US GDP growth
Amihud Illiquidity Index	n.a.
Baker-Wurgler Sentiment	n.a.

Table IA5: Value Spread Factor Timing Across Asset Classes and Factors Using Timing Specification (5) – Regression with Economic Sign Constraints and Variation Across Assets

The table repeats Table 5 in the paper using timing specification (5) from Table 6. Reported are the out of sample value spread timing for each factor in each asset class, as well as across asset classes, where the timing method is based off an out of sample regression of future factor returns on value spreads, where the coefficient is constrained to be positive and is allowed to differ by asset class and factor. The regression coefficients are estimated using an expanding historical window of data, converted to z-scores using the median and absolute deviation estimated from the same expanding window, and the timing weights are the z-scores of those regression coefficients, capped at +2 and -2. The all asset classes combine the timing strategies across asset classes using an equal risk combination of the asset classes, where each asset class is weighted in proportion to its inverse volatility, estimated over the full sample period. Results are reported for the static "no timing" portfolio for each factor in each asset class, the raw returns of the timing strategy, and the alpha of the timing strategy to the no timing static portfolio in each asset class (univariate alpha) as well as the alpha of the strategy to the no timing static portfolio for all factors (value, momentum, carry, and defensive) in each asset class (multivariate alpha), with *t*-statistics in parentheses.

	US stocks				International stocks				Equity index futures			
		Value spread timing				Value spread timing				Value spread timing		
Factor	No timing	Raw	Univariate alpha	Multivariate alpha	No timing	Raw	Univariate alpha	Multivariate alpha	No timing	Raw	Univariate alpha	Multivariate alpha
Value	0.30	0.06	-0.10	0.09	0.53	0.19	0.03	-0.09	0.33	-0.28	-0.17	-0.19
	(2.88)	(0.52)	-(0.88)	(0.84)	(3.17)	(0.96)	(0.17)	-(0.47)	(3.17)	-(2.59)	-(1.54)	-(1.76)
Momentum	0.50	0.26	0.62	0.41	0.68	-0.56	-0.91	-0.34	0.46	-0.16	-0.23	-0.16
	(4.77)	(2.34)	(5.59)	(3.69)	(4.00)	-(2.81)	-(4.54)	-(1.69)	(4.53)	-(1.51)	-(2.15)	-(1.49)
Carry									0.25	0.09	0.08	0.06
									(2.43)	(0.83)	(0.73)	(0.52)
Defensive	0.73	0.83	0.40	0.44	1.01	0.67	0.18	-0.07	0.40	-0.10	0.11	0.12
	(6.92)	(7.36)	(3.51)	(3.90)	(5.77)	(3.21)	(0.85)	-(0.33)	(3.92)	-(0.96)	(1.00)	(1.10)
Multi-factor	1.14	0.48	0.39	0.48	1.43	-0.02	-0.30	-0.28	0.72	-0.22	-0.02	-0.11
	(11.00)	(4.38)	(3.53)	(4.35)	(8.17)	-(0.08)	-(1.50)	-(1.41)	(7.04)	-(2.05)	-(0.14)	-(0.99)

	Commodities				Fixed income				Currencies			
		Value spread timing				Value spread timing				Value spread timing		
Factor	No timing	Raw	Univariate alpha	Multivariate alpha	No timing	Raw	Univariate alpha	Multivariate alpha	No timing	Raw	Univariate alpha	Multivariate alpha
Value	0.26	0.15	0.00	0.08	0.36	0.05	0.09	0.09	0.64	-0.12	-0.10	-0.13
	(2.59)	(1.44)	(0.00)	(0.72)	(3.55)	(0.48)	(0.86)	(0.79)	(4.26)	-(0.71)	-(0.57)	-(0.76)
Momentum	0.51	-0.19	0.03	-0.02	0.09	0.03	0.05	-0.04	0.18	0.26	0.31	0.09
	(5.05)	-(1.80)	(0.29)	-(0.23)	(0.88)	(0.24)	(0.45)	-(0.41)	(1.20)	(1.56)	(1.81)	(0.51)
Carry	0.53	0.27	0.09	0.11	0.63	0.03	0.17	0.15	0.43	0.07	0.25	0.19
	(5.32)	(2.57)	(0.84)	(1.00)	(6.18)	(0.30)	(1.55)	(1.36)	(2.92)	(0.43)	(1.49)	(1.11)
Defensive					0.04	0.05	0.01	-0.06				
					(0.41)	(0.48)	(0.10)	-(0.59)				
Multi-factor	0.86	0.13	0.10	0.09	0.55	0.06	0.09	0.05	0.64	0.16	0.30	0.14
	(8.56)	(1.24)	(0.95)	(0.84)	(5.37)	(0.52)	(0.86)	(0.45)	(4.32)	(0.98)	(1.78)	(0.83)

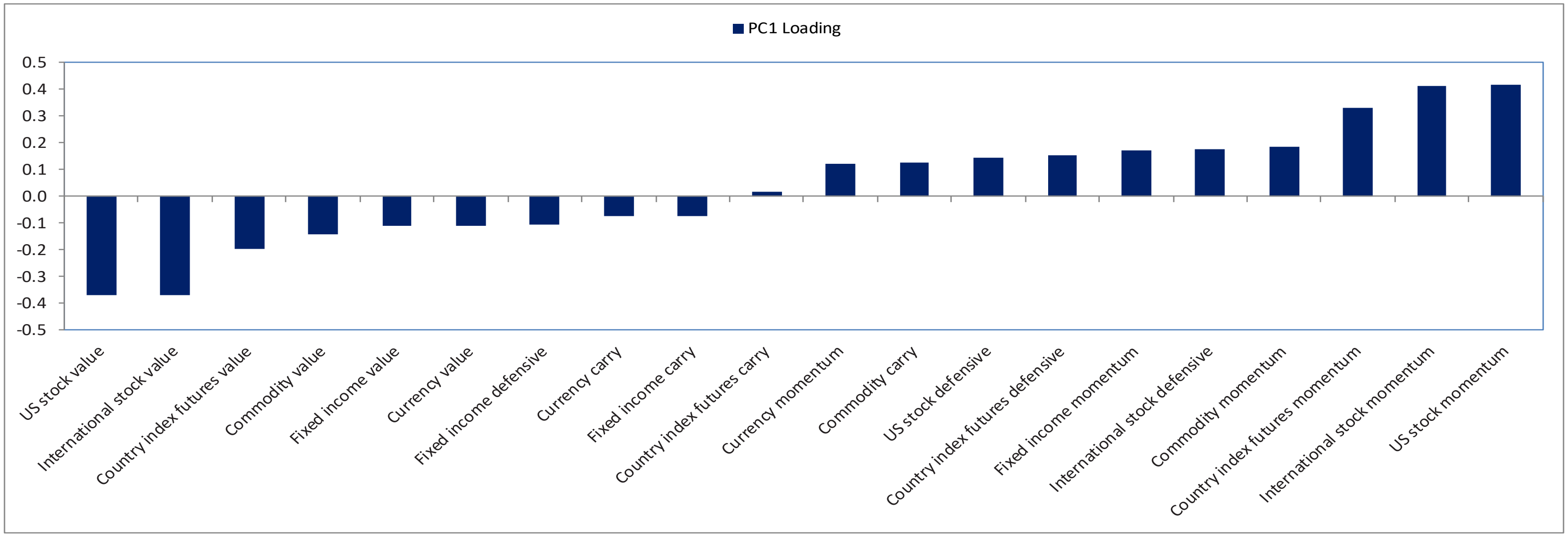
	All asset classes					Summary of value spread timing strategies		
		Value spread timing				Multivariate alpha		
Factor	No timing	Raw	Univariate alpha	Multivariate alpha		% significant	% right sign	% significant and right sign
Value	0.70	-0.01	-0.17	-0.01		0.0%	50.0%	0.0%
	(6.83)	-(0.11)	-(1.55)	-(0.10)				
Momentum	0.66	-0.07	0.12	0.06		16.7%	33.3%	16.7%
	(6.46)	-(0.65)	(1.16)	(0.55)				
Carry	0.85	0.26	0.23	0.37		0.0%	100.0%	0.0%
	(8.35)	(2.46)	(2.11)	(3.44)				
Defensive	0.75	0.56	0.60	0.53		25.0%	50.0%	25.0%
	(7.39)	(5.18)	(5.60)	(4.92)				
Multi-factor	1.64	0.23	0.39	0.35		16.7%	66.7%	16.7%
	(16.04)	(2.09)	(3.64)	(3.20)				
					Overall	11.5%	57.7%	11.5%

Table IA6: Principal Components Analysis of Factor Portfolios Across Asset Classes

The table reports results from PCA of our 20 factor by asset class portfolios. The percentage of the covariance matrix explained by each of the first five PCs, as well as the cumulative variance explained, is reported. Panel A reports results for the period when all asset classes and factors are present (1984 to 2018), and Panel B reports results over the longer period that excludes international stocks and currencies (1926 to 2018).

	Panel A: All asset classes				
	PC1	PC2	PC3	PC4	PC5
Variance explained	19.1%	9.3%	8.6%	7.2%	7.0%
Cumulative	19.1%	28.4%	37.1%	44.2%	51.2%
	Panel B: Excluding international stocks and currencies				
	PC1	PC2	PC3	PC4	PC5
Variance explained	16.0%	11.6%	9.6%	9.1%	8.8%
Cumulative	16.0%	268.1%	268.1%	268.1%	268.1%

Figure IA2: Factor Loadings on PCs. The figure reports the loadings of each factor portfolio to PC1 and PC2 over the common sample where all asset classes are represented.



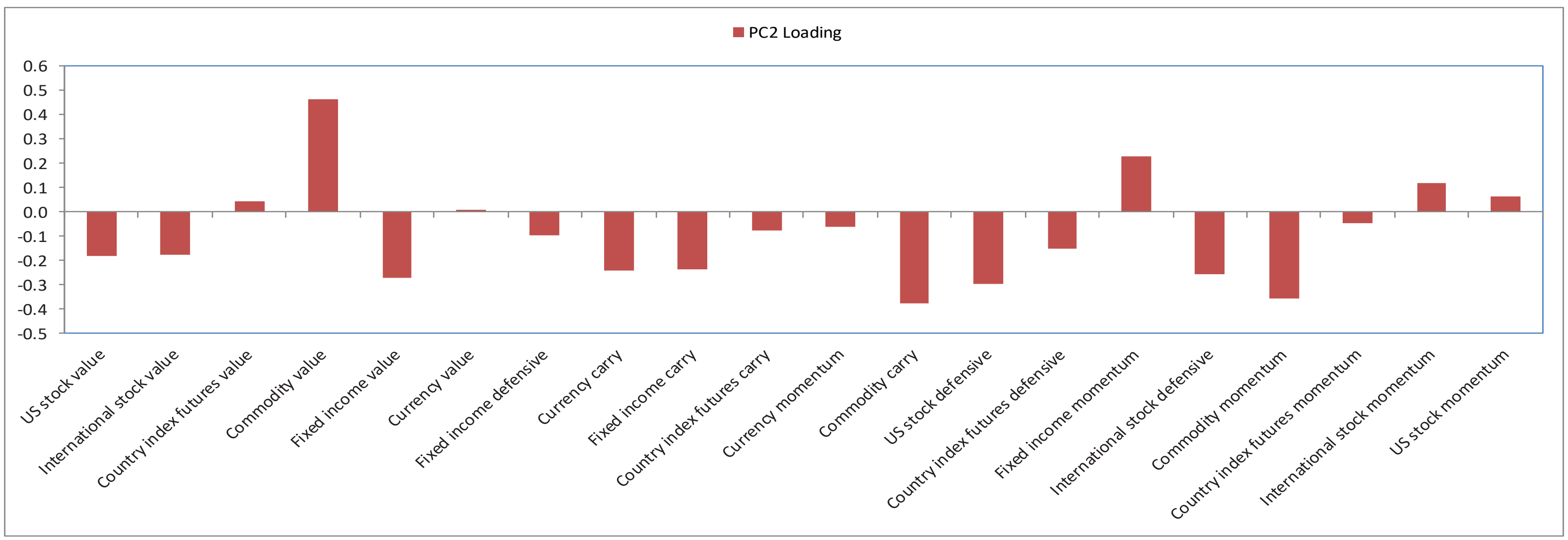


Table IA7: Momentum Factor Timing Using Different Methodologies

The table reports in and out of sample momentum factor timing for the multi-factor, all asset class implementation of these timing strategies using various timing methodologies that include: a z-score to time the strategy, a fitted regression of the strategy's returns on past 12-month returns that uses the coefficient from that regression to time the factor out of sample, and the PCA timing strategy of Haddad, Kozak, and Santosh (2018). The first two rows report the results for the z-score timing methodology both out of sample and in-sample, where the first row corresponds to the last row of Table 5. The next 15 rows report results from timing strategies based off of regressions which are run both over the full sample as well as an expanding window that relies only on ex ante information (out of sample) as indicated below. Regressions are run allowing the coefficient on momentum to vary across factors and assets ("individual asset") as well as forcing the coefficients to be the same for a given factor ("factor") and forcing the coefficients to be the same for every asset and factor ("pooled"). In some cases, an economic sign constraint is imposed on the coefficient (e.g., Campbell and Thompson (2007)), where value spreads should positively predict returns, so that if the regression coefficient is negative it is set to zero. In some cases, the standardized measures, whether z-scores or regression weights, are capped at +2, -2 (identified in the "Weight" column). We report annualized Sharpe ratios of the raw returns, the information ratio of the alpha of each timing strategy with respect to the static single factor being timed, and the information ratio of the alpha of each timing strategy with respect to all static factors (value, momentum, carry, and defensive) applied across all asset classes.

	Timing signal	Variation by	Sample	Constraints	Weight	Raw Sharpe ratio	Univariate alpha info. ratio	Multivariate alpha info. ratio
(1)	Z-score	Individual asset	Out of sample	Economic sign	Capped	0.17	-0.24	-0.09
(2)	Z-score	Individual asset	Full sample	Economic sign	Capped	0.49	0.31	0.34
(3)	Regression	Individual asset	Out of sample	None	Raw	1.34	0.12	0.28
(4)	Regression	Individual asset	Out of sample	None	Z-score	0.26	-0.04	0.00
(5)	Regression	Individual asset	Out of sample	None	Z, capped	0.38	0.07	0.10
(6)	Regression	Individual asset	Out of sample	Economic sign	Z, capped	0.36	-0.14	-0.04
(7)	Regression	Individual asset	Full sample	None	Z, capped	0.49	0.31	0.34
(8)	Regression	Factor	Out of sample	None	Raw	1.25	-0.09	0.09
(9)	Regression	Factor	Out of sample	None	Z-score	0.44	0.02	0.06
(10)	Regression	Factor	Out of sample	None	Z, capped	0.58	0.12	0.14
(11)	Regression	Factor	Out of sample	Economic sign	Z, capped	0.44	-0.12	0.03
(12)	Regression	Factor	Full sample	None	Z, capped	0.49	0.16	0.16
(13)	Regression	Pooled	Out of sample	None	Raw	1.42	-0.17	0.06
(14)	Regression	Pooled	Out of sample	None	Z-score	0.14	-0.25	-0.11
(15)	Regression	Pooled	Out of sample	None	Z, capped	0.18	-0.26	-0.10
(16)	Regression	Pooled	Out of sample	Economic sign	Z, capped	0.18	-0.26	-0.10
(17)	Regression	Pooled	Full sample	None	Z, capped	0.00	-0.40	-0.27
(18)	PCA	Individual asset	Out of sample	Economic sign	Raw	-0.04	-0.08	0.00
(19)	PCA	Individual asset	Full sample	Economic sign	Raw	0.14	0.21	0.24

Table IA8: Value and Momentum Factor Timing Using Different Methodologies

The table reports in and out of sample value and momentum factor timing for the multi-factor, all asset class implementation of these timing strategies using various timing methodologies that include: a z-score to time the strategy, a fitted regression of the strategy's returns on value spreads and past 12-month returns that uses the coefficients from those regressions to time the factor out of sample, and the PCA timing strategy of Haddad, Kozak, and Santosh (2018). The first two rows report the results for the z-score timing methodology both out of sample and in-sample, where the first row corresponds to the last row of Table 5. The next 15 rows report results from timing strategies based off of regressions which are run both over the full sample as well as an expanding window that relies only on ex ante information (out of sample) as indicated below. Regressions are run allowing the coefficient on the value spread and momentum to vary across factors and assets ("individual asset") as well as forcing the coefficients to be the same for a given factor ("factor") and forcing the coefficients to be the same for every asset and factor ("pooled"). In some cases, an economic sign constraint is imposed on the coefficient (e.g., Campbell and Thompson (2007)), where value spreads should positively predict returns, so that if the regression coefficient is negative it is set to zero. In some cases, the standardized measures, whether z-scores or regression weights, are capped at +2, -2 (identified in the "Weight" column). We report annualized Sharpe ratios of the raw returns, the information ratio of the alpha of each timing strategy with respect to the static single factor being timed, and the information ratio of the alpha of each timing strategy with respect to all static factors (value, momentum, carry, and defensive) applied across all asset classes.

	Timing signal	Variation by	Sample	Constraints	Weight	Raw Sharpe ratio	Univariate alpha info. ratio	Multivariate alpha info. ratio
(1)	Z-score	Individual asset	Out of sample	Economic sign	Capped	0.45	0.18	0.11
(2)	Z-score	Individual asset	Full sample	Economic sign	Capped	0.63	0.52	0.41
(3)	Regression	Individual asset	Out of sample	None	Raw	1.35	0.34	0.46
(4)	Regression	Individual asset	Out of sample	None	Z-score	0.44	0.40	0.37
(5)	Regression	Individual asset	Out of sample	None	Z, capped	0.45	0.49	0.46
(6)	Regression	Individual asset	Out of sample	Economic sign	Z, capped	0.59	0.38	0.32
(7)	Regression	Individual asset	Full sample	None	Z, capped	0.63	0.52	0.41
(8)	Regression	Factor	Out of sample	None	Raw	1.26	0.01	0.18
(9)	Regression	Factor	Out of sample	None	Z-score	0.42	0.00	0.01
(10)	Regression	Factor	Out of sample	None	Z, capped	0.47	0.04	0.04
(11)	Regression	Factor	Out of sample	Economic sign	Z, capped	0.46	0.00	0.03
(12)	Regression	Factor	Full sample	None	Z, capped	0.50	0.17	0.10
(13)	Regression	Pooled	Out of sample	None	Raw	1.41	-0.16	0.02
(14)	Regression	Pooled	Out of sample	None	Z-score	0.26	-0.23	-0.15
(15)	Regression	Pooled	Out of sample	None	Z, capped	0.21	-0.30	-0.20
(16)	Regression	Pooled	Out of sample	Economic sign	Z, capped	0.21	-0.30	-0.20
(17)	Regression	Pooled	Full sample	None	Z, capped	0.12	-0.33	-0.23
(18)	PCA	Individual asset	Out of sample	Economic sign	Raw	-0.05	0.23	0.13
(19)	PCA	Individual asset	Full sample	Economic sign	Raw	0.28	0.74	0.59

Figure IA3: Other Factor Timing Variables Using PCA. Reported are results for the PCA methodology of Haddad, Kozak, and Santosh (2018) applied to all of the other timing signals, using both out of sample PCA from the expanding historical window and the in-sample PCA method over the full sample.

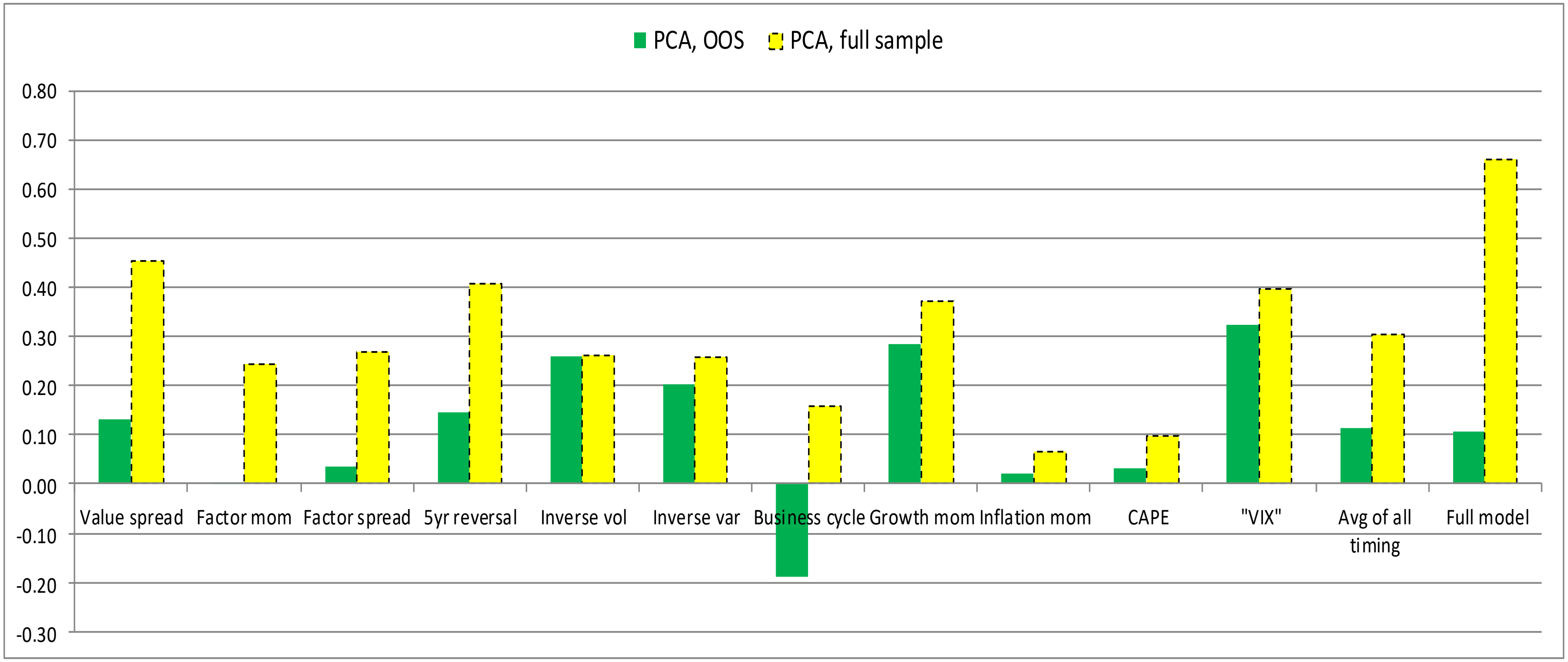
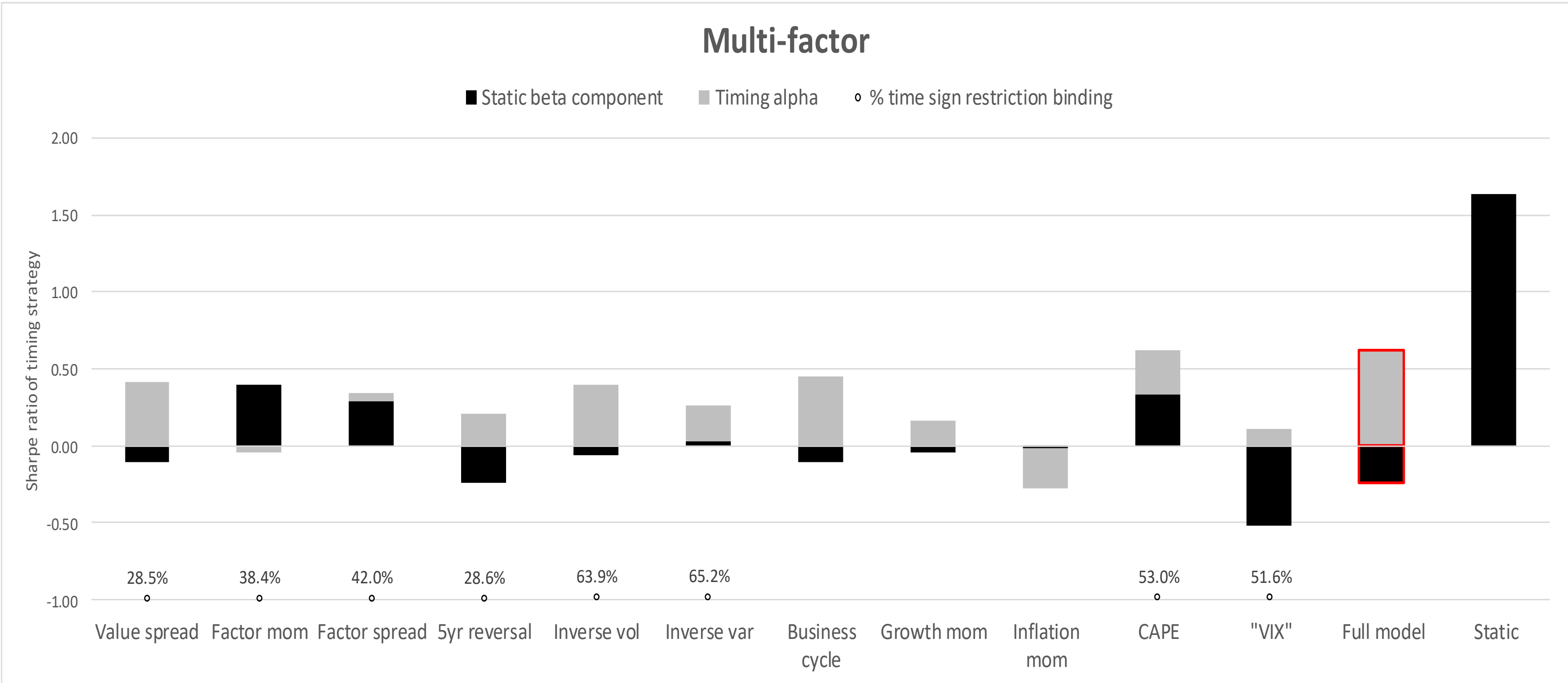
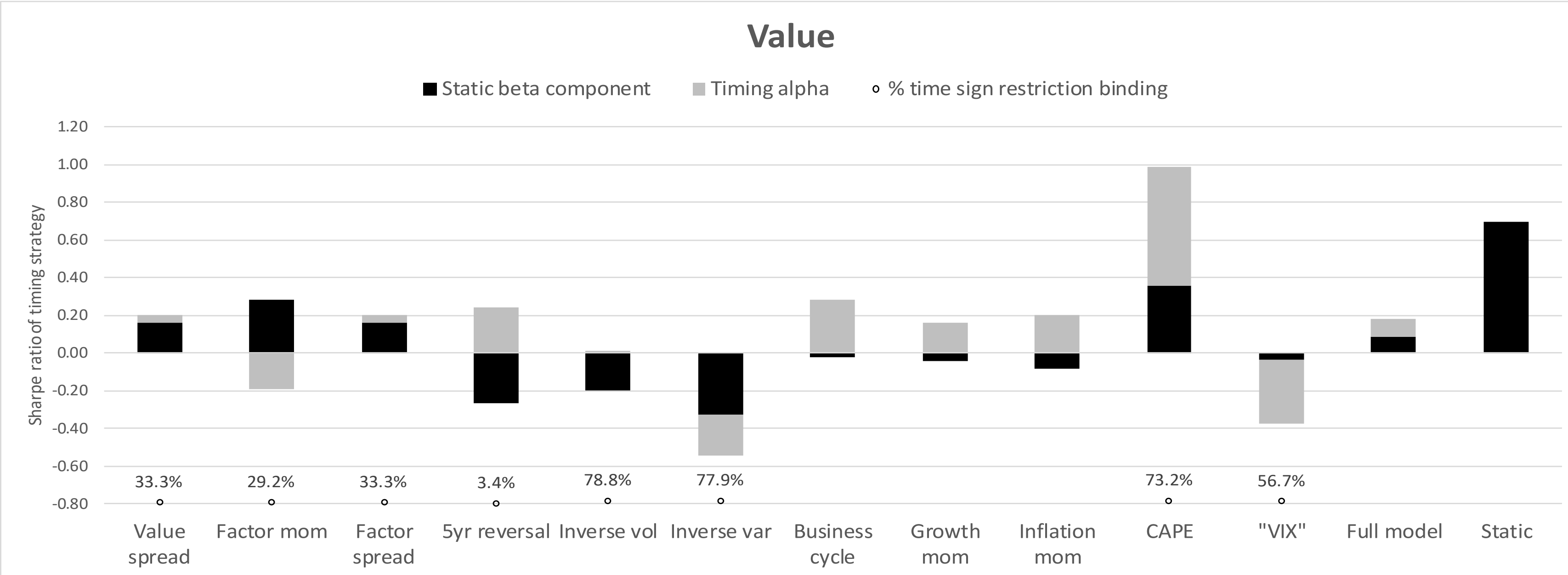
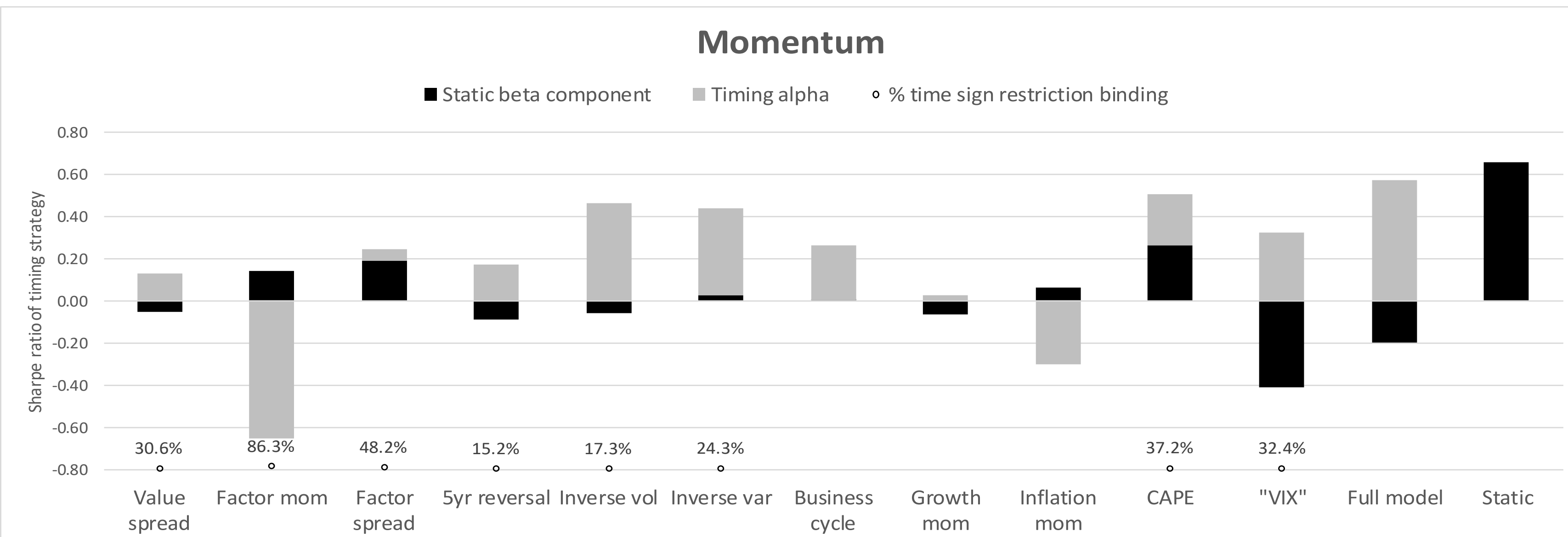
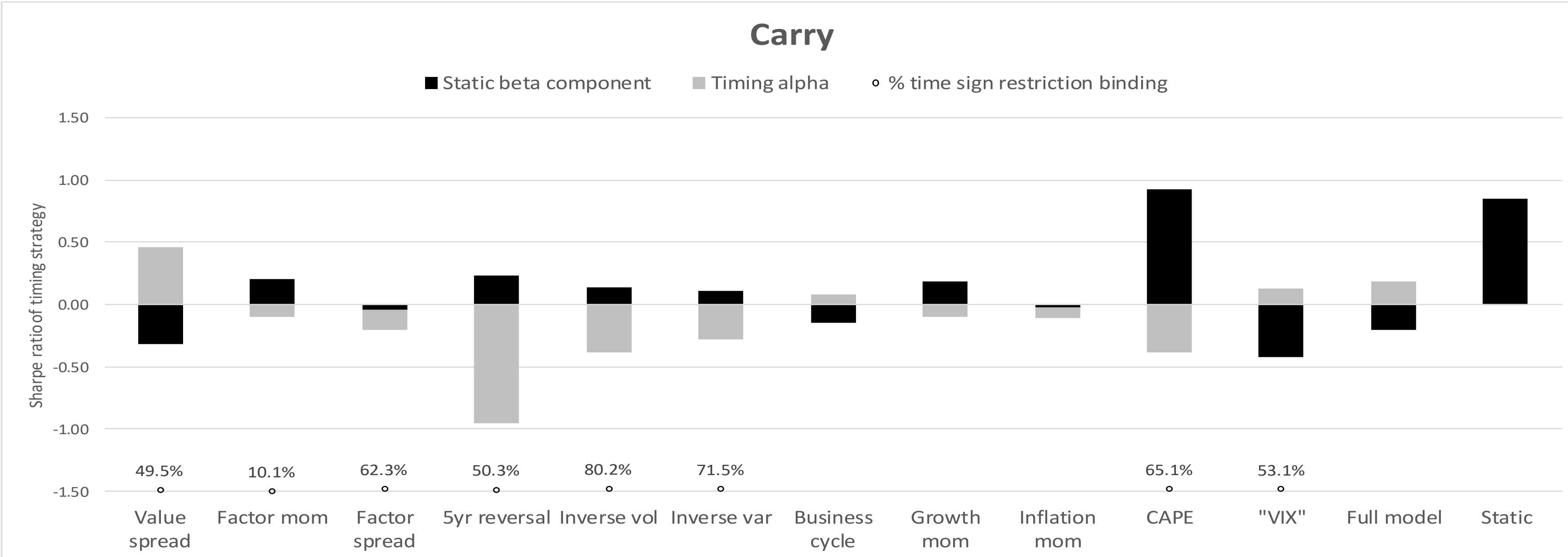


Figure IA4. Breaking Down Timing Strategies By Factor. The figure reports a breakdown of the Sharpe ratio of the out-of-sample regression timing strategy that imposes economic sign restrictions on the coefficients for every timing variable applied across all asset classes. For each timing variable, we report the contribution to returns coming from increased beta to the static factor from timing, plus the pure timing alpha (component of timing returns orthogonal to the static factors). The graphs plot the results for the multifactor timing strategy, and each individual factor timing strategy, applied across all factors, separately for each factor. On the first plot, we highlight in red the multifactor timing strategy using the full model of timing variables applied across all asset classes, which is also highlighted in Figure 6. Also plotted on each graph is the Sharpe ratio of the static factor portfolio applied across all asset classes for each factor for comparison (the no-timing factor portfolio). The percentage of times the economic sign restriction binds is also reported, except for the business cycle and macro variables, where there is no ex ante sign prediction.




Value
Static beta component
Timing alpha
% time sign restriction binding
Sharpe ratio of timing strategy
33.3%
29.2%
33.3%
3.4%
78.8%
77.9%
73.2%
56.7%
Value spread
Factor mom
Factor spread
5yr reversal
Inverse vol
Inverse var
Business cycle
Growth mom
Inflation mom
CAPE
"VIX"
Full model
Static


Momentum
Static beta component
Timing alpha
% time sign restriction binding
Sharpe ratio of timing strategy
30.6%
86.3%
48.2%
15.2%
17.3%
24.3%
37.2%
32.4%
Value spread
Factor mom
Factor spread
5yr reversal
Inverse vol
Inverse var
Business cycle
Growth mom
Inflation mom
CAPE
"VIX"
Full model
Static


Carry
Static beta component
Timing alpha
% time sign restriction binding
Sharpe ratio of timing strategy
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
49.5%
10.1%
62.3%
50.3%
80.2%
71.5%
65.1%
53.1%
Value spread
Factor mom
Factor spread
5yr reversal
Inverse vol
Inverse var
Business cycle
Growth mom
Inflation mom
CAPE
"VIX"
Full model
Static

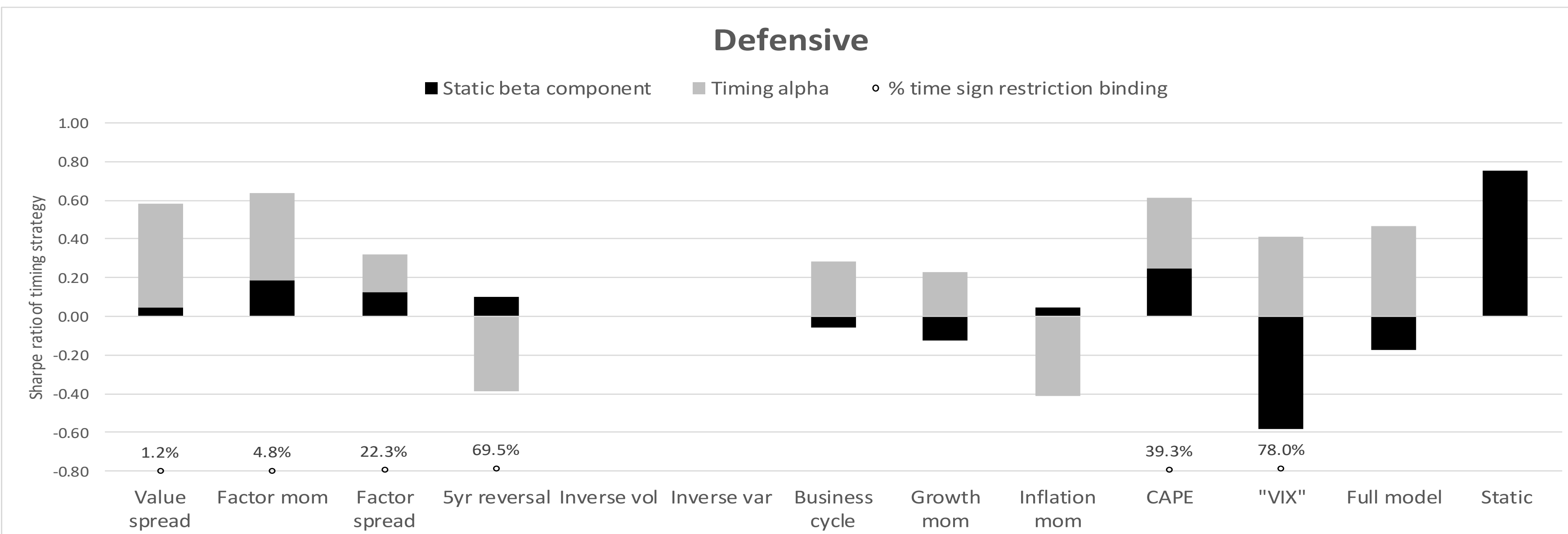
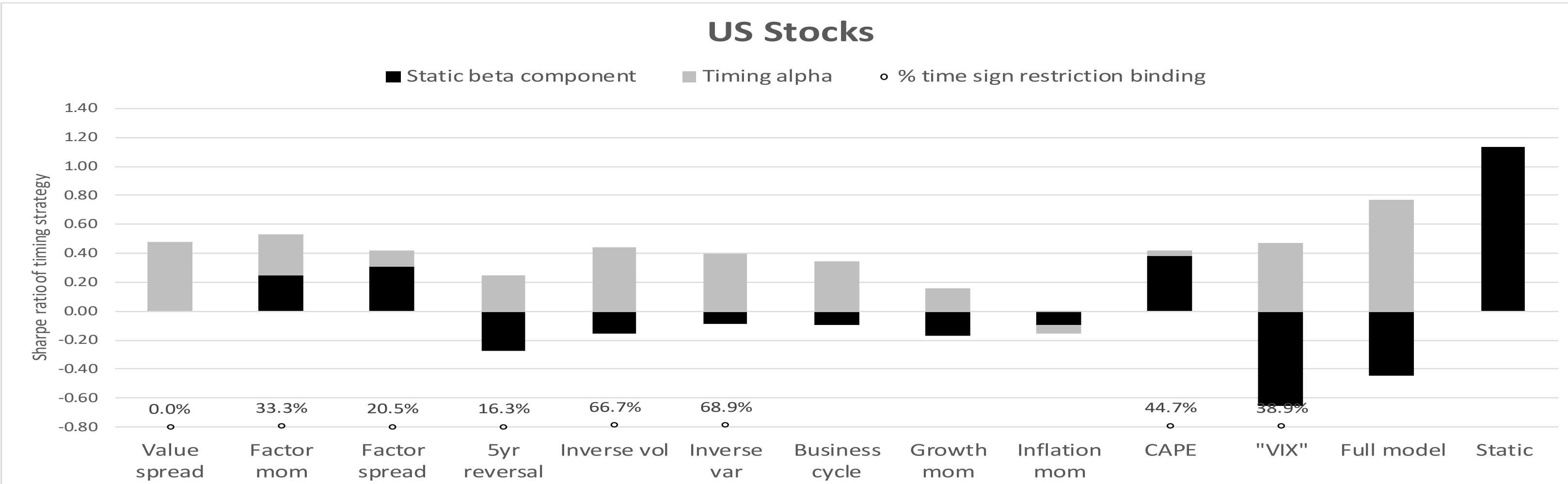
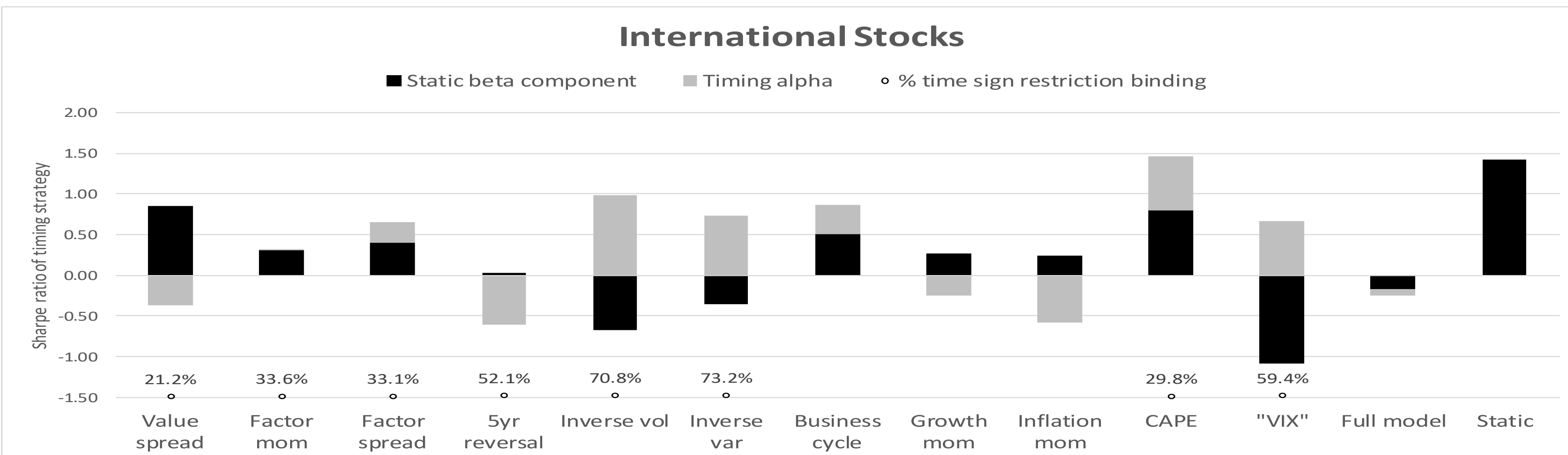
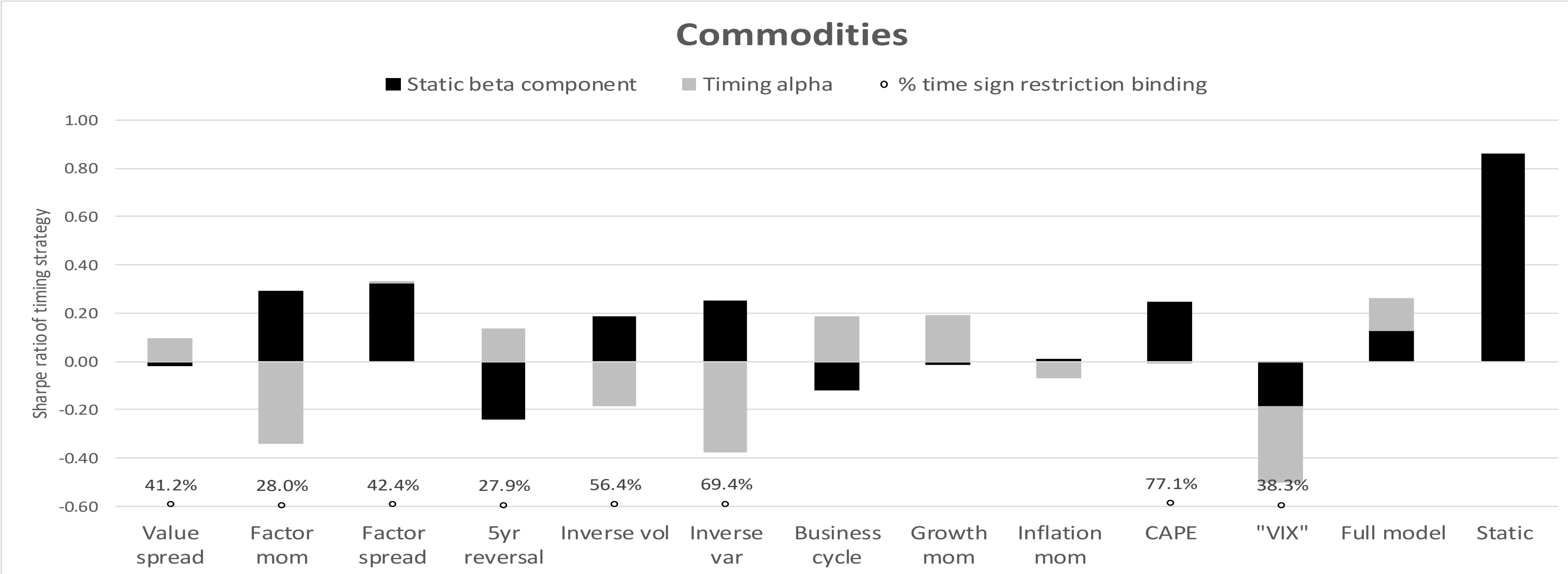
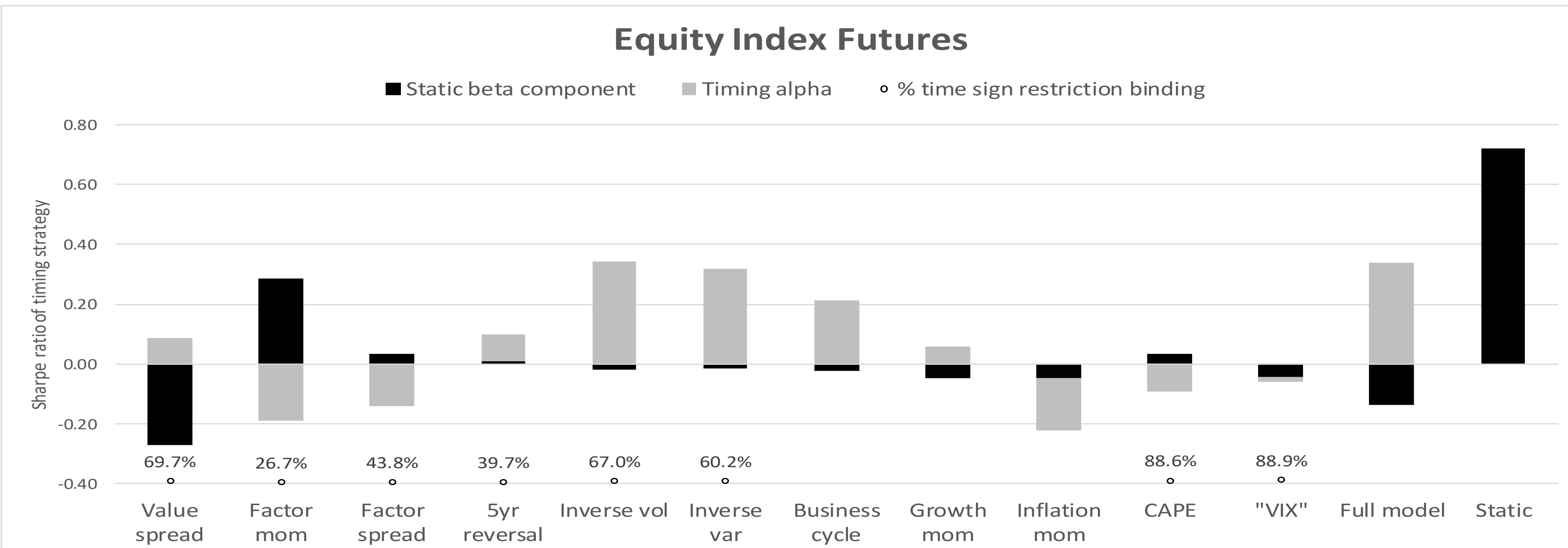

Defensive
Static beta component
Timing alpha
% time sign restriction binding
Sharpe ratio of timing strategy
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
-0.80
1.2%
4.8%
22.3%
69.5%
39.3%
78.0%
Value spread
Factor mom
Factor spread
5yr reversal
Inverse vol
Inverse var
Business cycle
Growth mom
Inflation mom
CAPE
"VIX"
Full model
Static

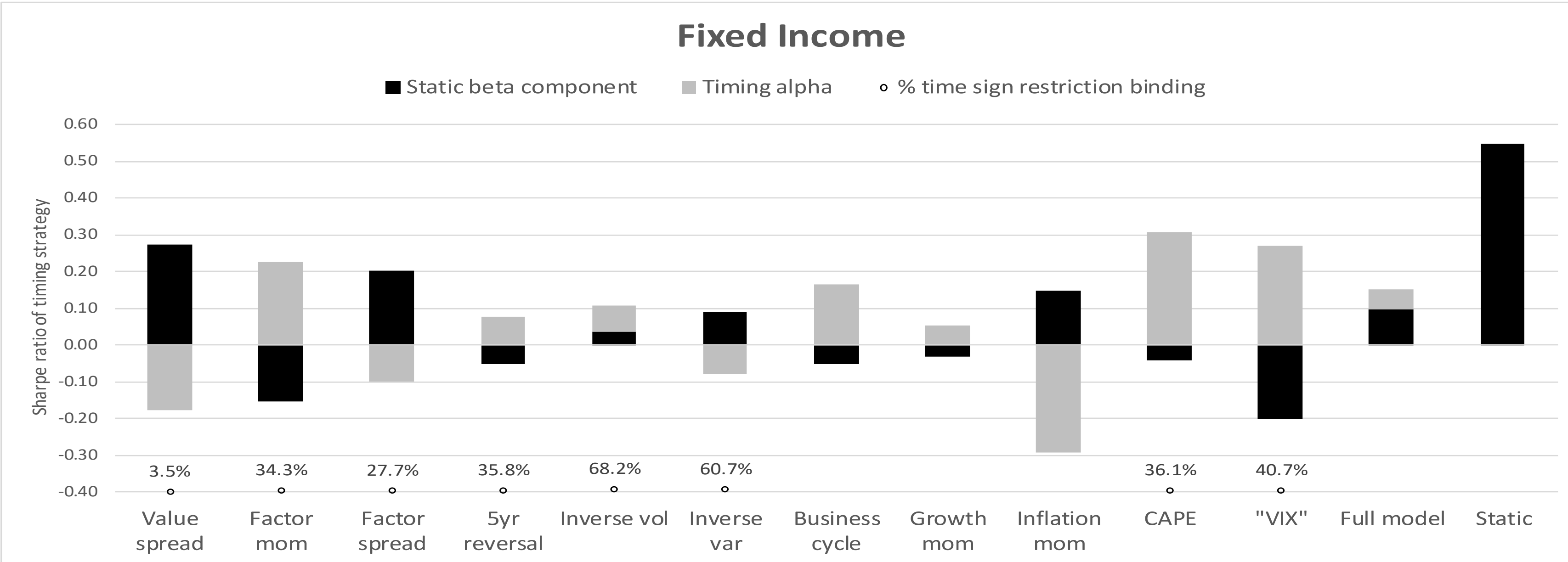
Figure 8. Breaking Down Timing Strategies By Asset Class. The figure reports a breakdown of the Sharpe ratio of the out-of-sample regression timing strategy that imposes economic sign restrictions on the coefficients for every timing variable applied across all asset classes. For each timing variable, we report the contribution to returns coming from increased beta to the static factor from timing, plus the pure timing alpha (component of timing returns orthogonal to the static factors). The graphs plot the results for the multifactor timing strategy separately for each asset class. Also plotted on each graph is the Sharpe ratio of the static multifactor portfolio within each asset class for comparison (the no-timing multifactor portfolio). The percentage of times the economic sign restriction binds is also reported, except for the business cycle and macro variables, where there is no ex ante sign prediction.

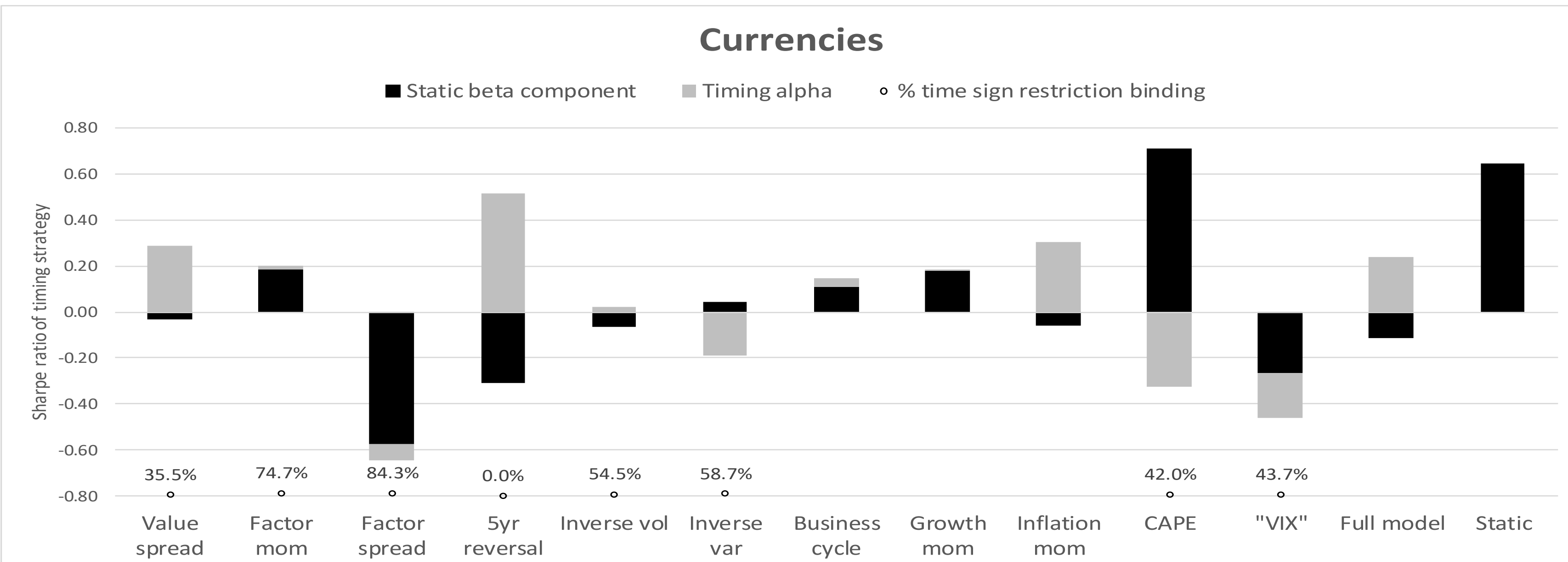





Commodities
Static beta component
Timing alpha
% time sign restriction binding
Sharpe ratio of timing strategy
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
41.2%
28.0%
42.4%
27.9%
56.4%
69.4%
77.1%
38.3%
Value spread
Factor mom
Factor spread
5yr reversal
Inverse vol
Inverse var
Business cycle
Growth mom
Inflation mom
CAPE
"VIX"
Full model
Static


Equity Index Futures
Static beta component
Timing alpha
% time sign restriction binding
Sharpe ratio of timing strategy
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
69.7%
26.7%
43.8%
39.7%
67.0%
60.2%
88.6%
88.9%
Value spread
Factor mom
Factor spread
5yr reversal
Inverse vol
Inverse var
Business cycle
Growth mom
Inflation mom
CAPE
"VIX"
Full model
Static


Fixed Income
Static beta component
Timing alpha
% time sign restriction binding
Sharpe ratio of timing strategy
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
-0.40
3.5%
34.3%
27.7%
35.8%
68.2%
60.7%
36.1%
40.7%
Value spread
Factor mom
Factor spread
5yr reversal
Inverse vol
Inverse var
Business cycle
Growth mom
Inflation mom
CAPE
"VIX"
Full model
Static


Currencies
Static beta component
Timing alpha
% time sign restriction binding
Sharpe ratio of timing strategy
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
-0.80
35.5%
74.7%
84.3%
0.0%
54.5%
58.7%
42.0%
43.7%
Value spread
Factor mom
Factor spread
5yr reversal
Inverse vol
Inverse var
Business cycle
Growth mom
Inflation mom
CAPE
"VIX"
Full model
Static