The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXV, NO. 5 • OCTOBER 2020

The Causal Effect of Limits to Arbitrage on Asset Pricing Anomalies

YONGQIANG CHU, DAVID HIRSHLEIFER, and LIANG MA*

ABSTRACT

We examine the causal effect of limits to arbitrage on 11 well-known asset pricing anomalies using the pilot program of Regulation SHO, which relaxed short-sale constraints for a quasi-random set of pilot stocks, as a natural experiment. We find that the anomalies became weaker on portfolios constructed with pilot stocks during the pilot period. The pilot program reduced the combined anomaly long–short portfolio returns by 72 basis points per month, a difference that survives risk adjustment with standard factor models. The effect comes only from the short legs of the anomaly portfolios.

OVER THE LAST SEVERAL DECADES, finance researchers have discovered many cross-sectional asset pricing anomalies, whereby predetermined security characteristics predict future stock returns.[1] Such patterns can derive from either rational risk premia or market mispricing. The mispricing explanation is consistent with the idea that limits to arbitrage delay the flow of wealth from irrational to sophisticated investors (Shleifer and Vishny (1997)). In contrast, if return predictability is the result of rational risk premia that compensate investors for bearing factor risk, limits to arbitrage should not affect expected returns.

An interesting question that arises is whether, to the extent that return anomalies reflect mispricing, such anomalies are persistent because limits to arbitrage prevent sophisticated investors from trading profitably against

*Yongqiang Chu is with the Belk College of Business, University of North Carolina, Charlotte. David Hirshleifer is with the Merage School of Business, University of California at Irvine and NBER. Liang Ma is with the Darla Moore School of Business, University of South Carolina. For helpful comments and suggestions, we thank Wei Xiong (the editor); an anonymous associate editor; two anonymous referees; Karl Diether; Lukasz Pomorski; Jeffrey Pontiff; Lin Sun; and participants at the 2016 Rodney L. White Center for Financial Research Conference on Financial Decisions and Asset Markets at Wharton and the 2017 American Finance Association Meetings. This research has received financial support from the Moore School Research Grant Program. The authors have read *The Journal of Finance* disclosure policy and have no conflicts of interest to disclose.

Correspondence: Liang Ma, Department of Finance, Darla Moore School of Business, University of South Carolina, 1014 Greene Street, Columbia, SC 29208; e-mail: liang.ma@moore.sc.edu.

[1] Harvey, Liu, and Zhu (2016) provide a comprehensive list of variables that can predict cross-sectional stock returns.

DOI: 10.1111/jofi.12947

them. However, it is empirically challenging to capture pure variations in limits to arbitrage that exclude variations in other economic forces that might affect risk premia or mispricing. In this paper, we study the *causal* effect of limits to arbitrage on 11 well-known asset pricing anomalies, namely, the momentum, gross profitability, asset growth, investment to assets, return on assets, net operating assets, accruals, net stock issuance, composite equity issuance, failure probability, and O-score anomalies. These 11 anomalies, which were the focus of Stambaugh, Yu, and Yuan (2012) in their study of sentiment and anomalies, survive after adjusting for the Fama–French three factors. Examining the causal effect of limits to arbitrage on these anomalies provides insight into the extent to which well-known return anomalies derive from risk versus mispricing.

It is difficult to identify the causal effect of limits to arbitrage as we seldom observe them, or pure variations in them, directly. Accordingly, existing literature often relies on firm characteristics, such as idiosyncratic volatility, size, and stock liquidity, as proxies for limits to arbitrage. However, these proxies are likely to be correlated with risk. For example, size has been offered as the basis for a risk factor in the three-factor model of Fama and French (1993), and volatility can be a risk measure in models with limited diversification such as settings with costs of trading or asymmetric information. This raises the possibility that effects attributed to limits to arbitrage may actually be due to rational risk premia.

Here we offer a pure test of the causal effect of limits to arbitrage on asset pricing anomalies. Short-sale constraints are one of the most important limits of arbitrage (e.g., Jones and Lamont (2002), Lamont and Thaler (2003), Nagel (2005), Gromb and Vayanos (2010)). Research on the effect of short-sale constraints on asset prices relies mainly on indirect proxies such as breadth of ownership (Chen, Hong, and Stein (2002)), institutional ownership (Asquith, Pathak, and Ritter (2005), Nagel (2005), Hirshleifer, Teoh, and Yu (2011)), firm size (Ali and Trombley (2006), Israel and Moskowitz (2013)), short interest (Asquith, Pathak, and Ritter (2005)), and shorting cost estimated from stock borrowing and lending behavior (Geczy, Musto, and Reed (2002), Jones and Lamont (2002), Drechsler and Drechsler (2014)). Several of these proxies may be correlated across stocks or over time with variations in factor risk.

We exploit a natural experiment—the Rule 202T pilot program of Regulation SHO (hereafter the pilot program)—to identify the *causal* effect of limits to arbitrage, and in particular short-sale constraints, on asset pricing anomalies. Regulation SHO was adopted by the Securities and Exchange Commission (SEC) in July 2004. Among stocks in the Russell 3000 index as of June 2004, the pilot program designated every third stock ranked by average daily trading volume (in the prior year) on each of NYSE, Amex, and NASDAQ as pilot stocks. The pilot program then removed short-sale price tests on this quasi-randomly selected group of pilot stocks. Prior to Regulation SHO, the specific form of short-sale price tests differed across stock markets. NYSE/Amex imposed the uptick rule, which only allowed a short sale to be placed on a plus tick or a zero-plus tick, whereas NASDAQ imposed the bid price test, which did not allow short sales at or below the (inside) bid when the inside bid was at or below the previous inside bid. From May 2, 2005 to August 6, 2007, the pilot stocks on NYSE/Amex were exempted from the uptick rule and those on NASDAQ were exempted from the bid price test. The pilot program therefore made it easier to short sell pilot stocks relative to nonpilot stocks. Because the assignment of pilot and nonpilot firms is quasi-random, the program provides an ideal setting to examine the causal effect of short-sale constraints on asset pricing anomalies. The bid price test for NASDAQ stocks is not very restrictive (see Diether, Lee, and Werner (2009) and discussion in Section I), and a significant fraction of trading volume in NASDAQ-listed stocks is executed on ArcaEx and INET, which do not enforce the bid price test. We therefore exclude NASDAQ stocks and only include pilot and nonpilot stocks traded on NYSE/Amex in our main analysis.

We examine two main hypotheses regarding the differential performance of pilot versus nonpilot anomaly portfolios *during* the pilot period of Regulation SHO. The first is that the anomalies become weaker for pilot firms relative to nonpilot firms during the pilot period. During the pilot period, arbitrageurs could more easily short pilot stocks to construct arbitrage portfolios, which should reduce mispricing. It follows that the return spread of arbitrage portfolios should decline for pilot stocks relative to nonpilot stocks.

To test the first hypothesis, for each asset pricing anomaly we construct long–short portfolios with pilot and nonpilot stocks separately. Specifically, we first sort all pilot stocks into deciles according to the return-predicting characteristic and calculate the anomaly returns as the return differences between the highest performing decile based on existing anomaly evidence (the long leg) and the lowest performing decile (the short leg). We then do the same with all nonpilot stocks. In a difference-in-differences framework, we find that the anomalies were much weaker in long–short portfolios constructed using pilot stocks during the pilot period. The effect is statistically significant in five of the 11 anomalies. When the 11 anomalies are combined in a joint test, the effect is both statistically and economically significant. The pilot program reduced the anomaly returns by 72 basis points per month, or 8.64% per year.

The second hypothesis is that the decrease in anomaly returns for pilot stocks during the pilot period comes mostly from the short-leg portfolios. In general, anomaly returns can come from either overpriced short legs or underpriced long legs. A loosening of short-sale constraints should reduce profitability of short-leg arbitrage portfolios. Using the same difference-in-differences framework, we find that the returns of short-leg portfolios constructed with pilot stocks were significantly and substantially higher during the pilot period, that is, short strategies became less profitable. In contrast, there is no significant effect of the pilot program on long-leg portfolios.

We next consider two additional hypotheses. First, the difference in anomaly returns between pilot and nonpilot stocks should vanish after the ending of the pilot program, with the disappearance of the difference in short-sale restrictions between pilot and nonpilot stocks.[2] We find empirical evidence consistent with this hypothesis. Furthermore, we expect to observe return dynamics of short-leg portfolios at the beginning and the end of the pilot program. Specifically, pilot short legs should underperform nonpilot short legs at the beginning of the pilot program, right after the uptick rule was lifted for pilot stocks. Similarly, pilot short legs should outperform nonpilot short legs at the end of the pilot program, right after the uptick rule was lifted for nonpilot stocks. We find evidence supporting this hypothesis as well.

We conduct a battery of robustness checks on our main results, which correspond to our two main hypotheses. We first rerun our tests on different sample periods. We find that our main results continue to hold. Next, we carry out two falsification tests. In a placebo test, we maintain the assignment of pilot and nonpilot firms but change the timing of the pilot period fictitiously to 2001 to 2003 and test whether this fictitious pilot program also affected the asset pricing anomalies during the 1980 to 2003 period.[3] We find that the fictitious pilot program had no effect on the asset pricing anomalies, suggesting that the main results are indeed driven by the pilot program. In another falsification test, we show that the difference in anomaly strength between pilot and nonpilot stocks during the pilot period was small and insignificant for NASDAQ stocks, which again confirms that our main results come from the relaxation of short-sale constraints.

Finally, we perform subsample analyses of our main results. As argued and documented by Diether, Lee, and Werner (2009), small and less liquid stocks were more affected by the suspension of the uptick rule (see the discussion in Section IV.D). Consistent with this observation, we find that the main effect (i.e., the effect of easier short selling on anomalies during the pilot period) is more pronounced among small and less liquid stocks. Moreover, in a more direct subsample analysis that splits stocks based on the extent to which they were restricted by the uptick rule before the pilot program, we find that our main effect is stronger among stocks that were more restricted by the uptick rule before the pilot program, which again corroborates our mechanism.

Taken together, the results above show that limits to arbitrage, and in particular, short-sale constraints, play an important role in generating the 11 well-known anomalies. These findings therefore suggest that, at a minimum, these anomalies are driven in large part by mispricing.

A potential alternative explanation for our main results is that the pilot program made pilot stocks more salient. Even though the pilot program is a quasi-experiment, it is not a double-blind study—market participants were aware of the change. It is possible that the sheer fact that the list of pilot stocks was publicly known drew attention to these stocks, and that higher investor attention to pilot stocks weakened anomalies, driving our main results. The results on NASDAQ stocks indicate that this mechanism is less plausible than

[2] After the pilot program ended, the uptick rule was lifted for nonpilot stocks as well.

[3] We end the placebo test sample in 2003 so that the actual pilot program does not affect the placebo test results.

our proposed mechanism of change in limits to arbitrage, as under this story NASDAQ pilot stocks that experienced an increase in investor attention should also have weakened anomalies, which is not the case. In contrast, our limits to short arbitrage hypothesis explains why the pilot effects are not present on NASDAQ. Furthermore, our results on shorting activity and the subsample results based on uptick rule restrictiveness probe further into whether short-sale constraints are the source of the effects of the pilot program and lend further support to our proposed mechanism. However, we do not assert that our tests rule out the alternative mechanism completely.

The behavioral finance literature has long argued that limits to arbitrage help explain the persistence of asset pricing anomalies despite the incentives of sophisticated investors to trade profitably against such anomalies (Shleifer and Vishny (1997), Hirshleifer (2001), Barberis and Thaler (2003), Gromb and Vayanos (2010)). Empirical tests examine the association between various proxies for limits to arbitrage and asset returns. These proxies for limits to arbitrage include stock price (Pontiff (1996), Mashruwala, Rajgopal, and Shevlin (2006)), size (Pontiff (1996), Ali, Hwang, and Trombley (2003), Israel and Moskowitz (2013)), idiosyncratic volatility (Ali, Hwang, and Trombley (2003), Mashruwala, Rajgopal, and Shevlin (2006)), transaction costs (Ali, Hwang, and Trombley (2003)), investor sophistication (Ali, Hwang, and Trombley (2003)), dollar trading volume (Mashruwala, Rajgopal, and Shevlin (2006)), and capital constraints of merger arbitrageurs in the context of merger arbitrage (Baker and Savaşoglu (2002)).

Many of the proxies for limits to arbitrage used in existing literature may actually capture risk, which makes it hard to distinguish between risk-based and mispricing-based explanations of anomalies. As documented by Lam and Wei (2011), proxies for limits to arbitrage are often highly correlated with proxies for investment frictions (risk).[4]

Our paper is more closely related to the empirical literature on how short-sale constraints or short-sale costs affect asset prices and asset pricing anomalies.[5] One strand of this literature employs indirect proxies for short-sale constraints (Chen, Hong, and Stein (2002), Asquith, Pathak, and Ritter (2005), Nagel (2005), Ali and Trombley (2006), Hirshleifer, Teoh, and Yu (2011), Israel and Moskowitz (2013)). However, the indirect proxies of short-sale constraints used in this strand of literature may also capture variation in risk.

Another strand of this literature uses more direct proxies for short-sale constraints or short-sale costs, measured using data on stock borrowing and lending (D'Avolio (2002), Geczy, Musto, and Reed (2002), Jones and Lamont (2002),

[4] Lam and Wei (2011) attempt to distinguish between mispricing-based and risk-based (q-theory with investment frictions) explanations of the asset growth anomaly. They examine a comprehensive list of proxies for limits to arbitrage: idiosyncratic volatility, the number of institutional shareholders, three measures of information uncertainty (analyst coverage, dispersion in analysts' earnings forecasts, and cash flow volatility), and five measures of transaction costs (stock price, effective bid-ask spread, institutional ownership, Amihud illiquidity, and dollar trading volume).

[5] See Reed (2013) and the references therein for more discussion on the role of short selling in financial markets.

Cao et al. (2007), Cohen, Diether, and Malloy (2007), Saffi and Sigurdsson (2010), Drechsler and Drechsler (2014), Beneish, Lee, and Nichols (2015), Engelberg, Reed, and Ringgenberg (2018)). These proxies are more direct in the sense that they are associated with the equity lending process, for example, with stock loan fees and stock lending supply. Nevertheless, these proxies can still be correlated across stocks or over time with shifts in factor risk, so that the return effects can still be driven by risk premia. In contrast, the natural experiment in our paper focuses on a regulatory shift that only alters permitted short-selling behavior and therefore is unlikely to be correlated with shifts in factor risk.

Our paper contributes to existing literature by providing a clean and powerful test of the causal effect of limits to arbitrage in general and short-sale constraints in particular on asset pricing anomalies. In contrast to existing literature that mainly relies on proxies for limits to arbitrage and short-sale constraints (which may capture risk or be correlated with risk as discussed above), we use exogenous shocks to short-sale constraints generated by a natural experiment, namely, the pilot program of Regulation SHO. The quasi-randomness of the assignment of pilot and nonpilot stocks makes a stock's assignment unlikely to be correlated with the loadings of stocks on risk factors. We are therefore able to conclude from our analysis whether limits of arbitrage and in turn mispricing actually affect asset pricing anomalies.

Our paper also adds to the literature that studies the impact of the pilot program of Regulation SHO. A few recent papers examine its effect on aspects related to stock prices. Diether, Lee, and Werner (2009) examine whether the suspension of short-sale price tests by the pilot program affects market quality. They find that short-selling activity increases for NYSE and NASDAQ pilot stocks and that NYSE pilot stocks experience a slight increase in spreads and intraday volatility, while the effect on market quality for NASDAQ stocks is smaller. Grullon, Michenaud, and Weston (2015) find that the pilot program leads to an increase in short-selling activity and a decline in prices for pilot stocks and that this effect is stronger for small firms, and these firms react by reducing equity issues and investment. Li and Zhang (2015) show that the pilot program increases price sensitivity to bad news and as a result makes managers more likely to reduce the precision of bad news forecasts. Fang, Huang, and Karpoff (2016) show that the pressure of short-selling on stock prices due to the pilot program curbs managers' willingness to manipulate earnings. To the best of our knowledge, our paper is the first to show that the pilot program affects the strength of well-known anomalies and reduces overpricing on short legs of these anomalies. Moreover, the implications of our results go beyond the effect of the pilot program per se and provide insight into a broad question in the asset pricing literature—whether these anomalies reflect mispricing or compensation for risk.

The rest of the paper is organized as follows. Section I provides background on Regulation SHO. Section II describes the data and anomalies. Section III presents results of our main empirical analysis. Section IV provides results of robustness checks and subsample analyses. Section V concludes.

I. Background on Regulation SHO

Following the stock market crash of 1929, short-selling restrictions (price tests) were introduced in the 1930s on stocks traded in the United States. Prior to Regulation SHO, the specific form of short-selling price tests differed across stock markets. NYSE/Amex imposed the uptick rule, which only allowed a short sale to be placed on a plus tick or a zero-plus tick, where a zero-plus tick is a zero tick such that the most recent price change preceding the trade was a plus tick. NASDAQ imposed the bid price test, which did not allow short sales at or below the (inside) bid when the inside bid was at or below the previous inside bid.

Regulation SHO was designed by the SEC to investigate whether the uptick rule imposed by NYSE (and Amex) and the bid price test imposed by NASDAQ affect market quality, and to develop potential uniform price tests if these rules are necessary. The regulation was announced by the SEC on July 28, 2004. For stocks in the Russell 3000 index as of June 2004, the pilot program of Regulation SHO designated every third stock ranked by average daily trading volume (in the prior year) on each of NYSE, Amex, and NASDAQ as pilot stocks. The SEC focused on the Russell 3000 stocks as they represent a broad cross section of U.S. stocks. The volume-ranking design was adopted to provide a quasi-random assignment of pilot versus nonpilot stocks. The pilot program removed the uptick rule and the bid price test on this quasi-randomly selected group of pilot stocks. From May 2, 2005 to August 6, 2007, the pilot stocks on NYSE/Amex were exempted from the uptick rule and those on NASDAQ were exempted from the bid price test.

The pilot program ended on August 6, 2007. Slightly before the pilot program ended, on July 6, 2007, the SEC eliminated short-sale price tests for all exchange-listed stocks. Therefore, the pilot program effectively ran from May 2, 2005 to July 6, 2007. As discussed in Fang, Huang, and Karpoff (2016), the elimination of the short-sale price tests for all exchange-listed stocks received extensive criticism from managers and politicians. This led the SEC to partially restore a modified uptick rule on February 24, 2010. Under the modified rule, short-sale price tests were imposed when a stock's price declined by 10% or more from its closing price on the previous trading day.

As discussed in detail in Diether, Lee, and Werner (2009), the bid price test is not very restrictive (compared to the uptick rule).[6] Furthermore, a significant fraction of trading volume in NASDAQ-listed stocks is executed on ArcaEx and INET, which do not enforce the bid price test. As a result, the effect of the pilot program on NASDAQ-listed stocks in terms of relaxing short-sale arbitrage

[6] Here, we quote the example given on page 44 of Diether, Lee, and Werner (2009) to illustrate why the bid price test is less restrictive than the uptick rule. Suppose for a stock the last sale price is \$28.05 on a plus tick, and the quotes are \$28.00 to \$28.05. To comply with the bid price test, a short sale can be placed on NASDAQ at a marketable limit sell order at \$28.00, as long as the most recent bid was \$28.00 or below. A NYSE short seller, however, has to place the short-sale order at \$28.05, which is 4 cents higher, in order to comply with the uptick rule.

constraints should be minimal. We therefore exclude NASDAQ stocks and only include pilot and nonpilot stocks traded on NYSE/Amex in our main analysis.

II. Data and Anomalies

A. Sample

Starting with the June 2004 Russell 3000 index, we follow the procedure described in the SEC's first pilot order of Regulation SHO (Securities Exchange Act Release No. 50104) to build our sample of pilot and nonpilot stocks. We exclude stocks that were not listed on the NYSE, Amex, or NASDAQ NM as well as stocks that went public or had spin-offs after April 30, 2004. The initial sample consists of 986 pilot stocks (based on the list published in the SEC's pilot order[7]) and 1,966 nonpilot stocks. We then merge this initial sample with data from the Center for Research in Security Prices (CRSP) and Compustat (both annual and quarterly data) to form portfolios and analyze portfolio returns of the 11 anomalies. As discussed in the introduction and Section I, the bid price test for NASDAQ-listed stocks is likely to have minimal effect. Our final sample therefore consists of pilot and nonpilot stocks in the pilot program that are listed on NYSE or Amex at portfolio formation. From the initial sample of pilot and nonpilot stocks of the pilot program, 1,025 nonpilot stocks and 515 pilot stocks are included in our final sample, of which 1,477 stocks are traded on NYSE and 63 stocks are traded on Amex. The ratio of nonpilot stocks to pilot stocks is roughly 2:1.[8] The sample period for our main empirical analysis is from January 1980 to June 2007, after which the pilot program of Regulation SHO ended.[9]

Our sample of pilot and nonpilot stocks is selected at the end of June 2004. For most of the prepilot period in our difference-in-differences analysis, the sample is selected (in terms of selecting pilot versus nonpilot stocks) using information not yet available. However, this is not an issue for our analysis. The reason is in the prepilot period, whether a stock is part of the pilot is only used to classify it into different groups of comparison. Furthermore, Table IV shows that our main results can also be identified using the pilot period per se, when information as to whether a stock is part of the pilot is available.

B. Anomalies

We focus on the 11 anomalies studied by Stambaugh, Yu, and Yuan (2012), which they select based on survival after adjustment for the Fama–French three factors.

[7] See https://www.sec.gov/rules/other/34-50104.htm.

[8] In Table IA.I in the Internet Appendix, we examine the robustness of our results to setting the number of pilot and nonpilot firms to be equal by randomly removing half of the nonpilot firms with simulation. Our main results continue to go through. The Internet Appendix is available in the online version of this article on *The Journal of Finance* website.

[9] The pilot program officially ended on August 6, 2007. However, on July 6, 2007, all exchange-listed stocks were exempted from short-sale price tests, which effectively ended the pilot program.

Table I
Characteristics of Stocks for the 11 Anomalies

This table summarizes the characteristics of stocks in the long leg (the highest performing group) and those in the short leg (the lowest performing group) for the 11 anomalies.

	Stocks in the Long Leg	Stocks in the Short Leg
Momentum	High past return	Low past return
Gross profitability	High gross profitability	Low gross profitability
Asset growth	Low asset growth	High asset growth
Investment to assets	Low investment to assets	High investment to assets
Return on assets	High return on assets	Low return on assets
Net operating assets	Low net operating assets	High net operating assets
Accruals	Low accruals	High accruals
Net stock issues	Low equity issuance	High equity issuance
Composite equity issues	Low equity issuance	High equity issuance
Failure probability	Low failure probability	High failure probability
O-score	Low O-score	High O-score

Below we briefly describe each anomaly, relegating details on variable construction to Appendix A. For each anomaly, there is a corresponding long–short trading strategy that goes long in the stocks that earn high returns (the long leg) and short in those that earn low returns (the short leg). The relationship between the subsequent stock performance and the ranking variable is positive for some anomalies and negative for others. For example, stocks with high past returns outperform those with low past returns for the momentum anomaly, whereas stocks with a low asset growth rate outperform those with a high asset growth rate for the asset growth anomaly. Table I summarizes the characteristics of stocks in the long and short legs for each anomaly.

Anomaly 1: Momentum. The momentum effect in stock returns, first documented by Jegadeesh and Titman (1993), is one of the most prominent anomalies in asset pricing. It refers to the phenomenon whereby stocks with higher past recent returns continue to outperform stocks with lower past recent returns. We employ the conventional 11-1-1 momentum strategy to construct our momentum portfolios. The ranking period is the 11 months from $t-12$ to $t-2$. The holding period is month t. Month $t-1$ is skipped to eliminate the short-run reversal effect.

Anomaly 2: Gross profitability. As documented by Novy-Marx (2013), stocks with high gross profitability earn on average higher returns than stocks with low gross profitability. He further shows that the profitability premium becomes more pronounced after controlling for the value premium. Following Novy-Marx (2013), we measure gross profitability as total revenue minus cost of goods sold, scaled by total assets.

Anomaly 3: Asset growth. Cooper, Gulen, and Schill (2008) find that stocks with high-growth total assets earn low subsequent returns. A possible explanation for this phenomenon is that investors tend to overreact to the growth rate of total assets. We measure asset growth as the change in total assets, scaled by lagged total assets.

Anomaly 4: Investment to assets. Titman, Wei, and Xie (2004) find that firms that increase capital investments earn negative benchmark-adjusted returns subsequently. They propose that this phenomenon is consistent with investors underreacting to the empire-building implications of increased investment expenditures. We measure investment to assets as the annual change in gross property, plant, and equipment plus the annual change in inventories, scaled by lagged total assets.

Anomaly 5: Return on assets. Fama and French (2006) document that in Fama–MacBeth cross-sectional regressions, earnings can predict stock returns. Chen, Novy-Marx, and Zhang (2011) and Stambaugh, Yu, and Yuan (2012) find that return on assets, measured on a quarterly basis, can predict subsequent stock returns: a higher past return on assets leads to higher subsequent stock returns. We measure return on assets as quarterly earnings scaled by quarterly total assets.

Anomaly 6: Net operating assets. Hirshleifer et al. (2004) find that firms with higher net operating assets earn lower subsequent returns. They attribute this phenomenon to investor limited attention. Net operating assets capture cumulative differences between operating income and free cash flow. Investors with limited attention may not process all available information and therefore may focus on accounting profitability without taking into account cash profitability information, leading to overvaluation of firms with higher net operating assets. We measure net operating assets as the difference between all operating assets and all operating liabilities on the balance sheet, scaled by lagged total assets.

Anomaly 7: Accruals. As documented by Sloan (1996), firms with higher accruals earn on average lower subsequent returns. This result suggests that stock prices fail to fully reflect information contained in the accruals and cash flow components of current earnings, which is consistent with investors having limited attention. We measure operating accruals as changes in noncash working capital minus depreciation expense, scaled by lagged total assets.

Anomaly 8: Net stock issues. As documented by Loughran and Ritter (1995) and Pontiff and Woodgate (2008), net share issues negatively predict stock returns in the cross section. One explanation for this phenomenon proposed in the literature is that firms issue stocks when they are overvalued and retire stocks when they are undervalued. We measure net stock issues on an annual basis as the change in the natural logarithm of a firm's adjusted shares over the last year.

Anomaly 9: Composite equity issues. Daniel and Titman (2006) find that an alternative measure of equity issuance, composite equity issuance, is also a negative predictor of stock returns in the cross section. They propose that this measure is related to the “intangible” component of past returns. Measured as the part of market equity growth not attributable to stock returns, composite equity issuance captures the amount of equity a firm issues (or retires) in exchange for cash or services. This measure therefore increases with seasoned equity issuance, employee stock option plans, and share-based acquisitions, whereas it decreases with share repurchases, dividends, and other actions that take cash out of the firm.

Anomaly 10: Failure probability. We use the failure probability proposed by Campbell, Hilscher, and Szilagyi (2008) to measure financial distress, which is estimated from a dynamic logit model to match empirically observed default events, with both market and accounting information taken into account. Campbell, Hilscher, and Szilagyi (2008) show that with this measure, more distressed firms earn lower subsequent returns on average than less distressed firms, especially after 1981.

Anomaly 11: O-score. We also use an alternative measure of financial distress, the O-score proposed by Ohlson (1980). Dichev (1998) shows that with this measure, more distressed firms earn lower subsequent returns on average than less distressed firms.

C. Summary of Anomaly Returns in Our Sample

Before proceeding to the main empirical analysis, we first verify the existence of the 11 anomalies in our sample of pilot and nonpilot firms. For each anomaly, we sort stocks in our sample into deciles based on the corresponding ranking variables and calculate the gross-return-weighted anomaly returns as the return differences between the highest performing decile (the long leg) and the lowest performing decile (the short leg). In other words, the portfolio break points we use are the decile break points in our pilot and nonpilot samples (that contain only NYSE/Amex stocks), respectively.

Equal-weighted portfolio returns can lead to various statistical and microstructure biases (Asparouhova, Bessembinder, and Kalcheva (2013)). However, it is useful for testing purposes to employ the information in small-firm returns, because small firms are especially informative for understanding the effects of limits to arbitrage. As discussed by Diether, Lee, and Werner (2009), the suspension of short-sale price tests is likely to affect smaller stocks more. So a test that makes use of small-firm returns maximizes our power to test the relevant hypothesis, namely, whether limits to short arbitrage (in the form of short-sale price tests) affect anomalies. Our main tests therefore use gross-return-weighted portfolio returns unless otherwise noted. The gross-return weight for stock i in each month t is its gross return $R_{i,t-1}$ in the preceding month $t-1$. As discussed in Asparouhova, Bessembinder, and Kalcheva (2010, 2013), gross-return weighting places similar weight in drawing inferences on the information provided by each stock in the sample while mitigating the statistical and microstructure biases associated with equal-weighted portfolio returns.

We use data from CRSP to construct portfolios for Anomalies 1 and 9, Compustat annual data to construct portfolios for Anomalies 2, 3, 4, 6, 7, and 8, Compustat quarterly data to construct portfolios for Anomaly 5, and CRSP and Compustat quarterly data to construct portfolios for Anomalies 10 and 11. For anomalies that use annual Compustat data, we follow Fama and French (1992) to match the accounting data for all fiscal years ending in calendar year $t-1$ with the stock returns from July of year t to June of $t+1$. For anomalies that use quarterly Compustat data, we use accounting information lagged by one quarter to match with stock returns.

We examine the average of raw anomaly returns and benchmark-adjusted anomaly returns controlling for the Capital Asset Pricing Model (CAPM) and the Fama–French three-factor model over the January 1980 to December 2004 sample period. We end the sample in December 2004 to avoid overlap with the pilot program. The average of benchmark-adjusted returns is the alpha from regressing the time series of excess returns onto the time series of the appropriate factors (the market excess return for the CAPM, and the market excess return plus two additional factors, the small minus big (SMB) and high minus low (HML) factors, for the Fama–French three-factor model). Table II reports these average returns.

Table II shows that the long–short portfolio returns for all 11 anomalies survive risk-adjustment with the Fama–French three-factor model. The average Fama–French three-factor-adjusted anomaly returns are presented in the last column of Table II and are positive and statistically significant for all 11 anomalies. These results are consistent with the evidence in Stambaugh, Yu, and Yuan (2012). We therefore confirm that these anomalies exist in our more restricted sample of stocks.

III. Empirical Analysis

As stated in the introduction, our two main hypotheses are as follows.

Hypothesis 1: The relaxation of short-sale constraints caused by the pilot program of Regulation SHO reduces anomaly returns for pilot stocks relative to nonpilot stocks during the pilot period.

Hypothesis 2: This decrease in anomaly returns comes primarily from the short-leg anomaly portfolios. Short legs of pilot stocks outperform those of nonpilot stocks during the pilot period.

We also test two additional hypotheses.

Hypothesis 3: The difference in anomaly returns between pilot and nonpilot stocks disappears after the pilot program ends.

Hypothesis 4: At the beginning of the pilot program, short legs of pilot stocks underperform those of nonpilot stocks. At the end of pilot program, short legs of pilot stocks outperform those of nonpilot stocks.

Table II

Summary of Anomaly Returns in Our Sample

This table reports the mean monthly raw return, the CAPM α, and the Fama–French three-factor α for the 11 anomalies individually, constructed using stocks in our sample. The sample consists of nonpilot and pilot stocks from the pilot program that are traded on NYSE/Amex. The sample period is January 1980 to December 2004. For each anomaly, stocks are sorted into deciles based on the corresponding ranking variable and the raw anomaly return is obtained as the portfolio return of buying the highest performing decile and shorting the lowest performing decile. All returns and alphas are in percentage terms. Robust t-statistics are presented in parentheses below the estimates. ** and *** denote significance at the 5% and 1% levels, respectively.

	Raw	CAPM α	Fama–French α
Momentum	0.94***	0.94***	1.13***
	(2.87)	(2.92)	(3.11)
Gross profitability	0.15	0.08	0.40**
	(0.84)	(0.44)	(2.24)
Asset growth	0.52***	0.57***	0.38**
	(3.10)	(3.44)	(2.56)
Investment to assets	0.45***	0.51***	0.42**
	(2.84)	(3.27)	(2.58)
Return on assets	0.22	0.28	0.60***
	(1.01)	(1.30)	(3.02)
Net operating assets	0.52***	0.49***	0.51***
	(3.64)	(3.50)	(3.48)
Accruals	0.34**	0.39**	0.36**
	(2.12)	(2.51)	(2.35)
Net stock issues	0.42***	0.45***	0.43***
	(3.21)	(3.51)	(3.29)
Composite equity issues	0.27	0.43***	0.40***
	(1.65)	(2.76)	(2.60)
Failure probability	−0.02	0.21	0.45**
	(−0.10)	(1.01)	(1.97)
O-score	0.06	0.09	0.48**
	(0.26)	(0.38)	(2.42)

A. Verifying the Quasi-Randomness of Pilot Stock Assignment

As discussed in Section I, the pilot firms were assigned in a quasi-random experiment (every third firm in a sorting of firms by trading volume on NYSE and, separately, on Amex).

In our context, we further confirm that the pilot firms were in fact quasi-randomly assigned with respect to firm characteristics associated with the 11 anomalies. To do so, we compare these firm characteristics between pilot and nonpilot firms at the end of 2003, before the announcement of the pilot program (July 2004). We calculate the mean of these anomaly variables for pilot and nonpilot firms at the end of 2003, and calculate both their differences and robust t-statistics of the differences. All variables are winsorized at the 1^{st} and 99^{th} percentiles of all firm-month observations to limit the effect of outliers. The results are reported in Panel A of Table III. Except for the measure of gross profitability, for which the difference is significant only at the 10% level, the

Table III

Comparing Nonpilot and Pilot Firms: Anomaly Variables

This table compares pilot and nonpilot firms in terms of the 11 ranking variables corresponding to the 11 asset pricing anomalies at the end of 2003. The sample consists of nonpilot and pilot firms from the pilot program that are traded on NYSE/Amex. Panel A reports the means of these variables for nonpilot and pilot firms, as well as their difference, over the full sample. Panel B (Panel C) reports the means of these variables for nonpilot and pilot firms, as well as their difference, for stocks that are in the long-leg (short-leg) portfolios of these anomalies. Variable definitions are in Appendix A. All variables are winsorized at the 1st and 99th percentiles of all firm-month observations to remove the effect of outliers. Robust t-statistics are presented in parentheses below the coefficient estimates.

		Panel A: Whole Sample		Panel B: Long Leg		Panel C: Short Leg	
	Variable	Nonpilot Pilot	Diff.	Nonpilot Pilot	Diff.	Nonpilot Pilot	Diff.
Momentum	*PRET*	0.452	−0.010	1.873	0.031	−0.210	0.037
		0.442	(−0.452)	1.904	(0.573)	−0.173	(2.463)
Gross profitability	*GP*/*A*	0.269	0.024	0.788	0.016	0.026	0.002
		0.293	(1.867)	0.804	(0.629)	0.029	(0.874)
Asset growth	*AG*	0.166	−0.032	−0.256	−0.042	1.418	−0.136
		0.134	(−0.772)	−0.297	(−1.506)	1.282	(−0.386)
Investment to assets	*IVA*	0.066	−0.015	−0.127	−0.039	0.492	−0.070
		0.052	(−1.077)	−0.166	(−1.775)	0.422	(−0.668)
Return on assets	*ROA*	0.010	0.001	0.042	0.001	−0.027	0.000
		0.010	(0.775)	0.043	(0.567)	−0.027	(0.033)
Net operating assets	*NOA*	0.593	0.012	−0.000	0.012	1.679	−0.173
		0.605	(0.474)	0.012	(1.077)	1.506	(−1.011)
Accruals	*AC*	−0.049	0.001	−0.170	0.006	0.098	−0.028
		−0.048	(0.182)	−0.164	(0.704)	0.070	(−1.517)
Net stock issues	*NSI*	0.052	−0.007	−0.074	−0.020	0.452	0.008
		0.045	(−0.603)	−0.094	(−1.629)	0.460	(0.096)
Composite equity issues	*CEI*	0.048	−0.020	−0.531	0.078	0.987	−0.120
		0.028	(−0.847)	−0.453	(2.042)	0.867	(−1.506)
Failure probability	*FP*	−8.143	−0.029	−9.111	−0.024	−6.957	0.057
		−8.172	(−0.830)	−9.135	(−0.788)	−6.899	(0.691)
O-score	*OS*	−3.160	−0.019	−5.065	0.018	−1.054	−0.006
		−3.179	(−0.264)	−5.047	(0.299)	−1.060	(−0.072)

other anomaly predictors show no statistically significant differences between pilot and nonpilot firms. Furthermore, the difference in gross profitability between pilot and nonpilot firms is small in magnitude compared with the two sample means.[10] As our empirical analysis focuses on the long legs and short legs of anomalies, in Panels B and C of Table III we also compare the mean of these anomaly variables for pilot and nonpilot stocks that fall into these two legs. The differences are again mostly small and statistically insignificant. Taken together, these results suggest that there were no significant differences

[10] In Table IA.II in the Internet Appendix, we confirm that the difference in size and book-to-market between pilot and nonpilot firms is also small and statistically insignificant.

between pilot and nonpilot firms prior to the announcement of the pilot program.

B. Main Difference-in-Differences Results

We now test the two main hypotheses, Hypotheses 1 and 2. We explore whether the pilot program led to differences in anomaly returns for the pilot stock sample relative to the nonpilot stock sample during the pilot period. We first construct anomaly portfolios based on pilot and nonpilot firms separately. Specifically, we sort all pilot stocks into deciles according to the predictors of the anomalies and calculate the returns of the highest performing decile (long-leg returns), the returns of the lowest performing decile (short-leg returns), and the differences between the two (long-short returns). We then do the same on all nonpilot firms. We next examine whether returns of pilot portfolios are different from returns of nonpilot portfolios during the pilot period (relative to the prepilot period), using a difference-in-differences approach. Throughout the paper, we use the terms difference-in-differences and DiD interchangeably.

The main difference-in-differences test employs the specification

$$r_{it} = \gamma_t + \beta Pilot_i \times During_t + \beta_1 Pilot_i + \epsilon_{it}, \tag{1}$$

where r_{it} is the gross-return-weighted monthly return of portfolio i (which can be the long leg, the short leg, or the long–short portfolio of an anomaly) in month t, γ_t denotes time fixed effects, $Pilot_i$ is a dummy variable that is equal to 1 if portfolio i is formed on pilot firms, and zero otherwise, and $During_t$ is a dummy variable that is equal to 1 if month t is between July 2005 and June 2007 (i.e., during the pilot period of Regulation SHO). As $During_t$ is subsumed by the time fixed effects, it is dropped from the regression. The time fixed effects γ_t capture the common factors and/or common macroeconomic variables that drive the portfolio returns for both pilot and nonpilot portfolios. In all analysis involving equation (1), we drop the two months at the beginning of the pilot program (May and June 2005), to avoid capturing price movement over this short time window. In these regressions, the unit of analysis is a portfolio-month observation. We estimate equation (1) for the long leg, the short leg, and the long–short portfolios separately.[11] The difference-in-differences coefficient β in equation (1) is the main coefficient of interest. It captures the difference

[11] We also explore an alternative empirical design to capture the effect of the pilot program on anomalies. Specifically, for each portfolio (the long/short leg or the long–short portfolio) of an anomaly, we take the return difference between the pilot portfolio and the nonpilot portfolio and denote the time series of this difference as r^d_{it}. By doing so, we isolate the cross-sectional difference between pilot and nonpilot portfolios. We expect that this difference will only predict returns when $During_t = 1$. Therefore, we run the regression $r^d_{it} = \gamma + \beta During_t + \epsilon_{it}$, where the coefficient β captures the difference in portfolio returns between pilot and nonpilot stocks during the pilot period (relative to that in the prepilot period). It can be shown that this specification and the specification in equation (1) are mathematically equivalent, and in Table IA.III in the Internet Appendix we confirm that they produce identical estimates of β. In the rest of the paper, we focus our discussion on results from the specification in equation (1).

in anomaly portfolio returns between pilot stocks and nonpilot stocks during the pilot period, relative to that in the prepilot period. We run the regression in equation (1) for each individual anomaly and also for all 11 anomalies combined. In the aggregate analysis, we replace the time fixed effects by the anomaly-time fixed effects (i.e., the fixed effects associated with each anomaly-time pair). The aggregate analysis enhances the power of our test and produces the average effect of the pilot program across all 11 anomalies. The results are reported in Table IV.

Hypothesis 1 predicts that β is negative for anomaly long–short returns. Hypothesis 2 predicts that β is positive for short-leg returns and close to zero for long-leg returns. The results support these two hypotheses. For the long–short returns (the last column of Table IV, Panel A), β is consistently negative for all 11 anomalies and statistically significant for five of them. When the 11 anomalies are combined, that is, in the aggregate analysis where equation (1) is estimated for all 11 anomalies together with the time fixed effects replaced by the anomaly-time fixed effects, β is -0.72% with a t-statistic of -4.37. In other words, the pilot program reduced the monthly anomaly returns by 72 basis points per month, or 8.64% per year, on average. These results are consistent with Hypothesis 1.

In addition to the difference-in-differences coefficient β, we report the mean anomaly returns for nonpilot and pilot stocks in the prepilot and during-pilot periods. These results are presented in the first six columns of Table IV, Panel A together with the differences in anomaly returns between pilot and nonpilot stocks in the prepilot and during-pilot periods. The results show that the difference between pilot and nonpilot anomaly returns does not exist in the prepilot period—it is small and insignificant for all 11 anomalies. Instead, the effect of the pilot program on anomaly returns comes mainly from the during-pilot period. The average difference in anomaly returns between pilot and nonpilot stocks in the during-pilot period is -0.65% with a t-statistic of -4.36. This evidence again suggests that our results are driven by the pilot program.

The results in Table IV, Panels B and C show that the decreases in anomaly returns come almost entirely from the short legs. For short-leg portfolios (Panel C), β is consistently positive for all 11 anomalies and statistically significant for five of them. When the 11 anomalies are combined, β is 0.63% with a t-statistic of 5.31. The first six columns of Panel C again show that the effect comes mainly from the during-pilot period. In contrast, for long-leg portfolios (Panel B), β is close to zero and statistically insignificant for most anomalies. When the 11 anomalies are combined, β is still close to zero and statistically insignificant. These results support Hypothesis 2.

In Appendix B, we show that using benchmark-adjusted (CAPM- and Fama–French three-factor-adjusted) returns as the dependent variable in equation (1) delivers almost identical results as in Table IV. This is expected, because the loadings on the benchmark factors of pilot versus nonpilot firms should also be similar, as the selection of pilot firms is quasi-random.

Our main tests for Hypotheses 1 and 2 based on equation (1) use a relatively short (two-year) pilot period to estimate the effect of relaxing short-sale

Table IV

Main Difference-in-Differences Results

This table presents the main DiD analysis results. The DiD coefficient β from the regression $r_{it} = \gamma_t + \beta Pilot_i \times During_t + \beta_1 Pilot_i + \epsilon_{it}$ is reported for the 11 anomalies individually and for all of them in aggregate. The variable r_{it} is the return of portfolio i in month t, γ_t denotes time fixed effects, $Pilot_i$ equals 1 if portfolio i is formed on pilot firms and zero otherwise, and $During_t$ equals 1 if month t is between July 2005 and June 2007 and zero otherwise. The portfolio returns are gross-return weighted. The mean portfolio returns for nonpilot and pilot stocks in the prepilot period and the during-pilot period, as well as their difference (pilot minus nonpilot) in these two periods, are reported. The DiD coefficient β is the difference in these two differences (one in the prepilot period and the other in the during-pilot period). Panels A, B, and C present results for the long-short anomaly portfolios, the long-leg portfolios, and the short-leg portfolios, respectively. The sample consists of nonpilot and pilot stocks from the pilot program that are traded on NYSE/Amex. The sample period is January 1980 to June 2007. We drop the two months at the beginning of the pilot program (May and June 2005) from the sample. Portfolio returns, difference in portfolio returns, and the DiD coefficient β are all in percentage terms. Robust t-statistics are presented in parentheses below the coefficient estimates and are only presented for differences in mean returns and the DiD coefficient β for brevity. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Long–Short anomaly returns

	Nonpilot Pre	Pilot Pre	Diff.	Nonpilot During	Pilot During	Diff.	DiD (β)
Momentum	0.96	1.04	0.08	1.02	0.47	−0.54	−0.62
			(0.31)			(−0.86)	(−0.93)
Gross profitability	0.06	0.39	0.34	0.47	−0.03	−0.50	−0.83*
			(1.60)			(−1.18)	(−1.79)
Asset growth	0.47	0.55	0.08	0.50	−0.03	−0.53	−0.61
			(0.35)			(−1.44)	(−1.42)
Investment to assets	0.45	0.36	−0.09	0.31	0.04	−0.28	−0.19
			(−0.40)			(−0.50)	(−0.31)
Return on assets	0.27	0.21	−0.06	0.83	0.07	−0.76	−0.70
			(−0.28)			(−1.22)	(−1.09)
Net operating assets	0.45	0.60	0.15	0.63	−0.21	−0.84*	−0.99**
			(0.70)			(−1.97)	(−2.11)
Accruals	0.29	0.40	0.12	−0.11	−1.29	−1.18**	−1.30***
			(0.47)			(−2.72)	(−2.62)
Net stock issues	0.39	0.50	0.11	0.42	−0.31	−0.73**	−0.84**
			(0.54)			(−2.25)	(−2.22)
Composite equity issues	0.16	0.41	0.25	0.17	−0.68	−0.85**	−1.10***
			(1.14)			(−2.43)	(−2.70)
Failure probability	0.10	0.05	−0.04	1.02	0.69	−0.33	−0.29
			(−0.20)			(−0.79)	(−0.61)
O-score	0.04	−0.16	−0.19	0.39	−0.24	−0.63	−0.44
			(−0.70)			(−0.82)	(−0.54)
Combination	0.33	0.40	0.07	0.51	−0.14	−0.65***	−0.72***
			(0.97)			(−4.36)	(−4.37)

(*Continued*)

Table IV—*Continued*

Panel B: Long-Leg Portfolio Returns

	Nonpilot Pre	Pilot Pre	Diff.	Nonpilot During	Pilot During	Diff.	DiD (β)
Momentum	2.21	2.29	0.08	1.57	1.67	0.10	0.03
			(0.47)			(0.28)	(0.07)
Gross profitability	1.75	1.81	0.06	1.10	1.12	0.02	−0.04
			(0.42)			(0.06)	(−0.14)
Asset growth	1.82	1.86	0.04	1.62	1.48	−0.15	−0.18
			(0.22)			(−0.42)	(−0.48)
Investment to assets	1.78	1.94	0.15	1.56	1.69	0.12	−0.03
			(0.86)			(0.36)	(−0.07)
Return on assets	1.73	1.75	0.02	1.52	1.58	0.06	0.04
			(0.17)			(0.35)	(0.17)
Net operating assets	1.77	1.86	0.09	1.31	1.28	−0.03	−0.12
			(0.55)			(−0.09)	(−0.33)
Accruals	1.76	1.84	0.08	1.41	0.97	−0.44	−0.51
			(0.41)			(−1.37)	(−1.41)
Net stock issues	1.72	1.80	0.08	1.25	1.26	0.00	−0.08
			(0.60)			(0.02)	(−0.27)
Composite equity issues	1.60	1.84	0.23*	1.37	1.06	−0.31	−0.55*
			(1.66)			(−1.11)	(−1.75)
Failure probability	1.58	1.53	−0.06	1.58	1.68	0.10	0.16
			(−0.45)			(0.39)	(0.55)
O-score	1.54	1.36	−0.18	1.30	1.48	0.18	0.36
			(−1.20)			(0.38)	(0.75)
Combination	1.75	1.81	0.05	1.42	1.39	−0.03	−0.09
			(1.15)			(−0.33)	(−0.81)

Panel C: Short-Leg Portfolio Returns

	Nonpilot Pre	Pilot Pre	Diff.	Nonpilot During	Pilot During	Diff.	DiD (β)
Momentum	1.25	1.25	−0.00	0.56	1.20	0.64	0.65
			(−0.02)			(1.46)	(1.37)
Gross profitability	1.69	1.42	−0.27*	0.63	1.14	0.51	0.79**
			(−1.74)			(1.55)	(2.17)
Asset growth	1.35	1.31	−0.04	1.12	1.50	0.38*	0.42
			(−0.27)			(1.79)	(1.64)
Investment to assets	1.33	1.58	0.25	1.25	1.65	0.40	0.16
			(1.57)			(1.22)	(0.44)
Return on assets	1.46	1.54	0.08	0.69	1.51	0.82	0.74
			(0.47)			(1.43)	(1.26)
Net operating assets	1.32	1.26	−0.06	0.68	1.49	0.81***	0.87***
			(−0.42)			(3.86)	(3.47)
Accruals	1.48	1.44	−0.04	1.52	2.26	0.74**	0.79**
			(−0.24)			(2.47)	(2.29)
Net stock issues	1.32	1.29	−0.03	0.83	1.56	0.73***	0.76***
			(−0.19)			(3.96)	(3.21)
Composite equity issues	1.44	1.43	−0.01	1.20	1.73	0.53**	0.55*
			(−0.08)			(2.09)	(1.80)
Failure probability	1.49	1.47	−0.01	0.55	0.99	0.43	0.45
			(−0.07)			(1.18)	(1.11)
O-score	1.51	1.52	0.01	0.91	1.72	0.81	0.80
			(0.04)			(1.54)	(1.39)
Combination	1.42	1.41	−0.01	0.90	1.52	0.62***	0.63***
			(−0.24)			(5.79)	(5.31)

constraints on anomaly returns. This immediately raises the question of whether the sample size generates enough power to distinguish hypotheses. In the results presented in Table IV, we do indeed obtain statistically significant effects on some individual anomalies, especially on the short legs, as well as strong significance for the results that aggregate across the 11 anomalies. This is reassuring. We discuss why our testing approach has sufficient power to distinguish hypotheses in Appendix C.

C. Postpilot-Program Results

After the pilot program ended, the difference in short-sale restrictions between pilot and nonpilot stocks disappeared. If our main results in Table IV are indeed driven by the pilot program, the difference in anomaly returns between pilot and nonpilot firms should also vanish after the end of the pilot program. This is formally stated in Hypothesis 3. To test this hypothesis, we estimate the revised difference-in-differences specification

$$r_{it} = \gamma_t + \beta Pilot_i \times During_t + \beta_1 Pilot_i + \beta_2 Pilot_i \times Post_t + \epsilon_{it}, \tag{2}$$

where $Post_t$ is a dummy variable that equals 1 if month t is after August 2007 and zero otherwise, that is, that represents the postpilot period, and the other variables are defined as in equation (1). The difference-in-differences coefficient β_2 in equation (2) is the coefficient of interest in this analysis, which we expect to be close to zero.

We run the regression in equation (2) for the entire January 1980 to December 2016 sample period. In this analysis, in addition to dropping the two months at the beginning of the pilot program (May and June 2005) as in the main analysis, we also drop the two months at the end of the pilot program (July and August 2007) to avoid capturing price movement over this short time window. The β estimates (representing the effect of the pilot program during the pilot period) are identical to those reported in Table IV (which can be shown mathematically) and therefore are not shown. Table V reports the β_2 estimates for the 11 anomalies individually and in aggregate. The coefficient β_2 is statistically insignificant for all 11 anomalies and close to zero when the 11 anomalies are combined. These results confirm that as the difference in short-sale restrictions between pilot and nonpilot firms disappeared, the difference in anomaly returns between them also vanished, consistent with our main conclusion.

We now consider what may have happened after the pilot program ended. On the one hand, as we hypothesize, the nonpilot stocks may have become like the pilot stocks because they also experienced a relaxation of short-sale constraints, which is also our main proposed mechanism. On the other hand, the pilot stocks may have reverted back and once again become like nonpilot stocks, if they were subject to some unknown temporary influence other than the relaxation of short-sale constraints. We can have an insignificant β_2 in

Table V

Postpilot-Program Results

This table reports the coefficient β_2 from the regression $r_{it} = \gamma_t + \beta Pilot_i \times During_t + \beta_1 Pilot_i + \beta_2 Pilot_i \times Post_t + \epsilon_{it}$ for the 11 anomalies individually and for all of them in aggregate. The variable r_{it} is the return of portfolio i in month t, γ_t denotes time fixed effects, $Pilot_i$ equals 1 if portfolio i is formed on pilot firms and zero otherwise, $During_t$ equals 1 if month t is between July 2005 and June 2007 and zero otherwise, and $Post_t$ equals 1 if month t is after August 2007 and zero otherwise. The sample consists of nonpilot and pilot stocks from the pilot program that are traded on NYSE/Amex. The sample period is January 1980 to December 2016. We drop the two months at the beginning of the pilot program (May and June 2005) and the two months at the end of the pilot program (July 2007 and August 2007) from the sample. The β_2 estimates are in percentage terms. Robust t-statistics are presented in parentheses below the coefficient estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Long Leg	Short Leg	Long–Short
Momentum	−0.09	0.39	−0.48
	(−0.31)	(0.72)	(−0.78)
Gross profitability	0.03	0.11	−0.08
	(0.12)	(0.31)	(−0.19)
Asset growth	0.16	−0.07	0.23
	(0.45)	(−0.24)	(0.56)
Investment to assets	−0.21	0.06	−0.27
	(−0.53)	(0.16)	(−0.58)
Return on assets	0.26	0.49	−0.23
	(1.27)	(1.16)	(−0.49)
Net operating assets	−0.25	−0.06	−0.19
	(−0.76)	(−0.21)	(−0.49)
Accruals	0.12	−0.17	0.29
	(0.26)	(−0.60)	(0.56)
Net stock issues	−0.16	0.18	−0.34
	(−0.63)	(0.59)	(−0.83)
Composite equity issues	−0.11	0.12	−0.23
	(−0.49)	(0.32)	(−0.52)
Failure probability	0.12	−0.10	0.21
	(0.58)	(−0.27)	(0.51)
O-score	0.19	−0.01	0.20
	(0.71)	(−0.03)	(0.37)
Combination	0.01	0.09	−0.08
	(0.07)	(0.75)	(−0.57)

both cases. We now perform further tests to distinguish between these two possible scenarios.

To this end, we compare the strength of anomaly returns in the during-pilot and postpilot periods for pilot versus nonpilot stocks. The average anomaly raw return across the 11 anomalies in the during-pilot period is 0.51% ($t = 2.82$) for nonpilot stocks (see also Table C.I) and −0.14% ($t = -0.85$) for pilot stocks.

In Scenario 1 (our paper's main proposed mechanism), we should see anomaly returns become *weaker* for nonpilot stocks right after the end of the pilot program. In Scenario 2, we should see anomaly returns become *stronger* for pilot stocks right after the end of the pilot program.

We find that in September 2007 to August 2009, which spans a period of equal length to the pilot period right after the pilot program, the average anomaly raw return across the 11 anomalies during the pilot period is -0.10% ($t = -0.12$) for nonpilot stocks and 0.04% ($t = 0.04$) for pilot stocks. This evidence is consistent with Scenario 1, not Scenario 2. This finding further supports our paper's main proposed mechanism and shows that easier arbitrage also reduced anomaly returns for nonpilot stocks right after the end of the pilot program.

D. Return Dynamics of Short Legs

Our main results indicate that short-leg portfolios of pilot stocks outperformed those of nonpilot stocks *during* the pilot period. As stated in Hypothesis 4, our mechanism has additional implications regarding the price movement of short legs at the beginning and end of the pilot program. Specifically, if the relaxation of short-sale constraints is indeed the underlying force driving our main results, we should see a large price decrease (negative returns) at the beginning of the pilot program for pilot short legs (relative to nonpilot short legs), right after the uptick rule was lifted for pilot stocks. Similarly, at the end of the pilot program, when the uptick rule was lifted for nonpilot stocks as well, the relaxation of short-sale constraints should reduce overpricing for nonpilot short legs relative to pilot short legs. We therefore expect to see a price increase in the short legs of pilot stocks relative to those of nonpilot stocks.

We next test Hypothesis 4. We also provide a more complete view of the return dynamics of pilot short-leg portfolios (relative to nonpilot short-leg portfolios), over the prepilot, during-pilot, and postpilot periods.

Specifically, we construct consecutive rolling time windows that begin in May 1999 and cover the three periods. Most of the time windows are two years long, in line with the length of the during-pilot period. At the beginning and the end of the pilot program, we use a short window of two months to capture potential large price movements around these two dates. The exact timing of these windows is May 1999 to April 2001, May 2001 to April 2003, May 2003 to April 2005, May 2005 to June 2005, July 2005 to June 2007, July 2007 to August 2007, September 2007 to August 2009, September 2009 to August 2011, and September 2011 to December 2013. The first three windows cover the prepilot period. The fourth (two-month) window covers the beginning of the pilot program. The fifth window covers the during-pilot period and corresponds to the dummy variable *During* in equation (1). The sixth (two-month) window covers the end of the pilot program. The last three windows cover the postpilot period. The last rolling window is longer than two years to cover the full sample period ending in December 2013; our results are similar if we instead use the two-year period from September 2011 to August 2013 for the last window.

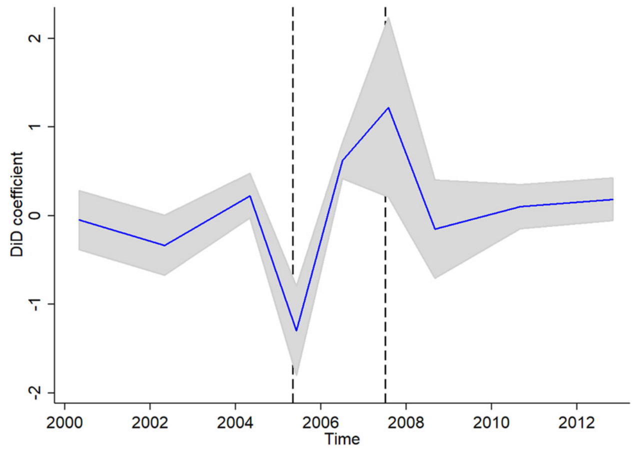

Figure 1. Return dynamics of short-leg portfolios. This figure illustrates the return dynamics of pilot short-leg portfolios relative to nonpilot short-leg portfolios over nine rolling windows that cover the prepilot, during-pilot, and postpilot periods as well as the beginning and end of the pilot program. The regression $r_{it} = \gamma_t + \sum_{j=1}^{9} \beta_{wj} Pilot_i \times Window_{jt} + \beta_1 Pilot_i + \epsilon_{it}$ is conducted for the 11 anomalies in aggregate, where r_{it} is the monthly return of short-leg portfolio i in month t, γ_t denotes time fixed effects, $Pilot_i$ is a dummy variable that equals 1 if portfolio i is formed on pilot firms and zero otherwise, and $Window_{jt}$ is a dummy variable that equals 1 if month t is in $Window_j$. The timing of the nine windows is May 1999 to April 2001, May 2001 to April 2003, May 2003 to April 2005, May 2005 to June 2005, July 2005 to June 2007, July 2007 to August 2007, September 2007 to August 2009, September 2009 to August 2011, and September 2011 to December 2013. The sample period for this analysis is January 1980 to December 2013. The blue solid line plots the nine DiD coefficients β_{wj} in percentage terms. The shaded area shows their 90% confidence intervals. The two dashed vertical lines denote the beginning and end of the pilot program. (Color figure can be viewed at wileyonlinelibrary.com)

We define a dummy variable for each of these rolling windows, $Window_j$, $j = 1, 2, \ldots, 9$, and then estimate the difference-in-differences specification

$$r_{it} = \gamma_t + \sum_{j=1}^{9} \beta_{wj} Pilot_i \times Window_{jt} + \beta_1 Pilot_i + \epsilon_{it}, \tag{3}$$

where r_{it} is the monthly return of short-leg portfolio i in month t, γ_t denotes time fixed effects, $Pilot_i$ is a dummy variable that equals 1 if portfolio i is formed on pilot firms and zero otherwise, and $Window_{jt}$ is a dummy variable that equals 1 if month t is in $Window_j$. We perform this analysis for all 11 anomalies in aggregate, to provide an overall view of the return dynamics on the short-leg portfolios and to enhance the power of the test.[12] The sample period for this analysis is January 1980 to December 2013.

We obtain nine DiD coefficients β_{wj}, one for each window. Figure 1 plots these coefficients over time, together with their 90% confidence intervals.

[12] This aggregation is especially important for the short two-month windows at the beginning and end of the pilot program (i.e., to test Hypothesis 4).

In the prepilot period, the three DiD coefficients are close to zero and statistically insignificant, which indicates that there was no notable difference between pilot and nonpilot short-leg returns before the pilot program. In the during-pilot period, the DiD coefficient is positive and significant, which is consistent with our main DiD results in Table IV. We also define a dummy variable for the transition period between the announcement and introduction of the pilot program (i.e., August 2004 to April 2005) and find that the DiD coefficient for this transition period, 0.06% ($t = 0.25$), is close to zero and statistically insignificant. This result suggests that the actual lifting of the uptick rule introduced by the pilot program was needed to reduce overpricing of short legs.

In the postpilot period, the three DiD coefficients are close to zero and statistically insignificant. These results suggest that the difference between pilot and nonpilot short-leg returns vanished in the postpilot period, as all stocks were exempted from the uptick rule, consistent with the results in Table V.

As for Hypothesis 4, at the beginning of the pilot program, we find that the DiD coefficient is large and negative. It is −1.30% ($t = -4.27$) per month over the two-month window from May to June 2005. This result suggests that there was a large initial price decrease on pilot short legs once the uptick rule was lifted, consistent with Hypothesis 4. This finding provides direct evidence that overpricing is reduced once short-sale constraints are relaxed.

At the end of the pilot program, the uptick rule was lifted for nonpilot stocks as well. This relaxation of short-sale constraints should reduce overpricing for nonpilot short legs relative to pilot short legs. Hypothesis 4 predicts a price increase in the short legs of pilot stocks relative to those of nonpilot stocks (i.e., a positive DiD coefficient). Consistent with this conjecture, the DiD coefficient, 1.22% ($t = 1.99$), is positive, statistically significant, and large in magnitude over the two-month window from July to August 2007.

Overall, the results presented in Figure 1 support the view that the relaxation of short-sale constraints reduced the mispricing associated with the asset pricing anomalies.

E. Shorting Activity

We hypothesize that the introduction of the pilot program will result in an increase in shorting for pilot short-leg portfolios relative to nonpilot short-leg portfolios, and that this increase will be sustained throughout the pilot period. At the end of the pilot program, the difference in shorting activity between pilot and nonpilot short-leg portfolios should decrease, as the uptick rule was also lifted for nonpilot stocks.[13]

[13] Rational arbitrageurs might potentially short preemptively, resulting in a relative increase in shorting on pilot short legs after the announcement but before the introduction of the pilot program. However, this effect may not be strong enough to leave clear tracks in the data, for two reasons. First, the expected benefit of preemptive shorting is not realized until the pilot program starts; during the interim, it is costly/risky to hold a short position from pilot program announcement to the start of the pilot program without commensurate incremental compensation. Second,

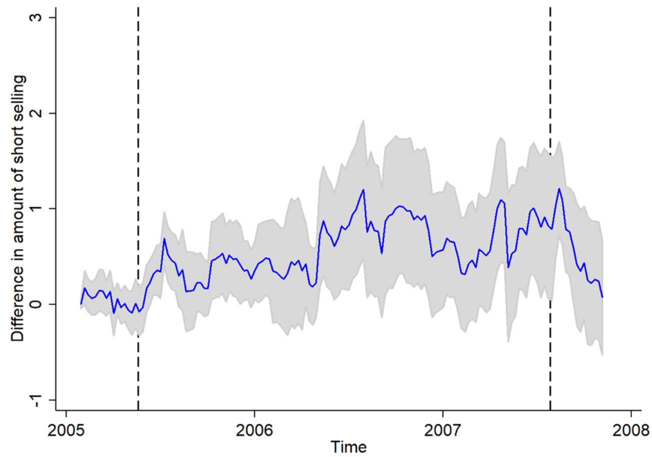

Figure 2. Difference in shorting activity on short-leg portfolios between pilot and nonpilot stocks. This figure shows the difference in the amount of short selling between pilot and nonpilot short-leg portfolios from January 2005 to October 2007. The difference is normalized to zero at the beginning. We use weekly Markit Securities Finance data. For each stock-week, we measure the amount of shorting selling as the ratio between the number of shares on loan and the number of shares outstanding. We calculate the portfolio-level amount of short selling as the average amount of short selling across all stocks in the short-leg portfolios. We then take the difference in the portfolio-level amount of shorting selling between pilot and nonpilot short legs and average this difference across the 11 anomalies. The solid line plots the average difference in percentage terms. The shaded area shows the 90% confidence intervals. The two dashed vertical lines denote the beginning and end of the pilot program. (Color figure can be viewed at wileyonlinelibrary.com)

We focus our analysis on the January 2005 to October 2007 window. Over this window, we compare the difference in the amount of short selling between pilot and nonpilot short-leg portfolios using weekly Markit Securities Finance data. For each stock and week, we measure the amount of short selling as the ratio between the number of shares on loan and the number of shares outstanding. We calculate the portfolio-level amount of short selling as the average amount of short selling across all stocks in the short-leg portfolios. We take the difference in the amount of short selling between pilot and nonpilot short legs, and average this difference across the 11 anomalies. We normalize the difference to be zero at the beginning of the window.

Figure 2 plots the difference (together with its 90% confidence interval). In line with the hypotheses above, there is no apparent increase in shorting on pilot short legs (relative to nonpilot short legs) until the beginning of the pilot program. The increase is then sustained throughout the pilot period. There is a decrease in the difference around the end of the pilot program, when the uptick rule was also lifted for nonpilot stocks. In Figure IA.1 in the Internet

as anomaly characteristics tend to revert to the mean over time, the shorting opportunity presented by an anomaly characteristic before the pilot period will be weaker on average by the time the pilot program starts, which reduces the expected amount of postinitiation shorting.

Appendix, we find that the results using short interest from Compustat as a measure of shorting activity are qualitatively similar.

Overall, this set of results suggests that pilot short-leg portfolios do indeed experience more short selling once short-sale constraints are relaxed. This finding is consistent with our argument that the relaxation of short-sale constraints reduces mispricing associated with the asset pricing anomalies.

IV. Robustness Checks and Subsample Analyses

In this section, we present a battery of robustness checks and subsample analyses corresponding to the main results in Table IV.

A. Different Sample Periods

We first test the robustness of our main results with respect to the sample period. Doing so addresses the concern that the during-pilot period is shorter than the prepilot period as well as sheds light on whether our choice of sample period is critical. Specifically, we explore two shorter sample periods, January 1990 to June 2007 and January 2000 to June 2007, and estimate equation (1) over these two sample periods. The results are presented in Panels A and B of Table VI, respectively. The β estimates are qualitatively similar to those in Table IV. In aggregate, the β estimates are statistically and economically significant for both the short leg and the long–short portfolios using the two different sample periods. With all 11 anomalies combined, the β for the long-short portfolio is -0.53% with a t-statistic of -3.06 from 1990 to 2007 and -0.72% with a t-statistic of -3.46 from 2000 to 2007. The β for the short leg is 0.50% with a t-statistic of 3.98 from 1990 to 2007 and 0.74% with a t-statistic of 4.98 from 2000 to 2007.

B. Placebo Tests

In general, a potential problem with the difference-in-differences method is that the results can be driven by unobservable shocks that affect pilot and nonpilot firms differently, which may undermine the causal inference of the main difference-in-differences results. The volume-ranking design used to choose pilot firms (in which every third firm in a sorting of firms by trading volume on NYSE and, separately, on Amex was chosen as a pilot firm) makes the assignment of pilot and nonpilot firms quasi-random and unlikely to be highly correlated with unobserved shocks. Nonetheless, as a precaution, we conduct a set of falsification tests.

As a first placebo test, we create a pseudo-event in 2000 and perform a test as if the pseudo-event relaxed short-sale constraints for pilot firms.[14] To mimic

[14] We choose 2000 for a pseudo-event as it is prior to the real event and the resulting pseudo-pilot program does not overlap with the real pilot program. In Table IA.IV in the Internet Appendix, we find similar results when we create a pseudo-event in 1998 or 1999.

Table VI

Different Sample Periods

This table reports the coefficient β from the regression $r_{it} = \gamma_t + \beta Pilot_i \times During_t + \beta_1 Pilot_i + \epsilon_{it}$ for the 11 anomalies individually and for all of them in aggregate. The variable r_{it} is the return of portfolio i in month t, γ_t denotes time fixed effects, $Pilot_i$ equals 1 if portfolio i is formed on pilot firms and zero otherwise, and $During_t$ equals 1 if month t is between July 2005 and June 2007 and zero otherwise. The sample consists of nonpilot and pilot stocks from the pilot program that are traded on NYSE/Amex. The sample period is January 1990 to June 2007 for Panel A and January 2000 to June 2007 for Panel B. We drop the two months at the beginning of the pilot program (May and June 2005) from the sample. The β estimates are in percentage terms. Robust t-statistics are presented in parentheses below the coefficient estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Panel A: 1990 to 2007			Panel B: 2000 to 2007		
	Long Leg	Short Leg	Long-Short	Long Leg	Short Leg	Long-Short
Momentum	0.06	0.36	−0.30	−0.10	0.58	−0.68
	(0.15)	(0.73)	(−0.43)	(−0.19)	(1.05)	(−0.82)
Gross profitability	0.05	0.56	−0.51	0.18	0.44	−0.26
	(0.14)	(1.46)	(−1.04)	(0.46)	(1.05)	(−0.48)
Asset growth	−0.19	0.48*	−0.68	0.09	0.83**	−0.74
	(−0.48)	(1.79)	(−1.50)	(0.16)	(2.35)	(−1.18)
Investment to assets	−0.03	−0.04	0.01	−0.23	0.07	−0.30
	(−0.08)	(−0.11)	(0.02)	(−0.47)	(0.15)	(−0.40)
Return on assets	0.19	0.66	−0.47	0.35	1.03	−0.68
	(0.78)	(1.09)	(−0.71)	(1.00)	(1.55)	(−0.91)
Net operating assets	0.09	0.80***	−0.71	−0.08	1.11***	−1.19*
	(0.24)	(2.99)	(−1.44)	(−0.17)	(3.07)	(−1.92)
Accruals	−0.66	0.72*	−1.39**	−0.26	1.06**	−1.31*
	(−1.63)	(1.95)	(−2.55)	(−0.51)	(2.23)	(−1.82)
Net stock issues	0.03	0.63**	−0.61	0.04	0.85**	−0.81
	(0.09)	(2.32)	(−1.47)	(0.13)	(2.09)	(−1.39)
Composite equity issues	−0.56*	0.31	−0.87**	−0.52	0.74*	−1.26**
	(−1.72)	(0.93)	(−2.00)	(−1.33)	(1.77)	(−2.24)
Failure probability	0.26	0.34	−0.08	0.16	0.64	−0.48
	(0.87)	(0.82)	(−0.16)	(0.44)	(1.22)	(−0.83)
O-score	0.47	0.66	−0.19	0.65	0.83	−0.17
	(0.95)	(1.10)	(−0.23)	(1.16)	(1.21)	(−0.19)
Combination	−0.03	0.50***	−0.53***	0.03	0.74***	−0.72***
	(−0.26)	(3.98)	(−3.06)	(0.19)	(4.98)	(−3.46)

the actual pilot program closely, we assume that this pseudo pilot program was effective from May 2001 to July 2003. We then run the difference-in-differences regression

$$r_{it} = \gamma_t + \beta Pilot_i \times PseudoDuring_t + \beta_1 Pilot_i + \epsilon_{it}, \tag{4}$$

where $PseudoDuring_t$ is a dummy variable that equals 1 if month t is between July 2001 and June 2003 (i.e., when the pseudo-event was effective) and the other variables are defined as in equation (1). Following closely the main analysis in Table IV, the sample period is from January 1980 to June 2003, and we

Table VII
Results From a Placebo Test

This table reports results from a placebo test. We create a pseudo-event in 2000 and assume that the pseudo-event also relaxed short-sale constraints for the pilot firms. To mimic the actual pilot program closely, we assume that this pseudo-event was effective from May 2001 to July 2003. We then run the difference-in-differences regression $r_{it} = \gamma_t + \beta Pilot_i \times PseudoDuring_t + \beta_1 Pilot_i + \epsilon_{it}$ for the 11 anomalies individually and for all of them in aggregate. The variable r_{it} is the return of portfolio i in month t, γ_t denotes time fixed effects, $Pilot_i$ equals 1 if portfolio i is formed on pilot firms and zero otherwise, and $PseudoDuring_t$ equals 1 if month t is between July 2001 and June 2003 and zero otherwise. The sample consists of nonpilot and pilot stocks from the pilot program that are traded on NYSE/Amex. The sample period is from January 1980 to June 2003. We drop the two months at the beginning of the pseudo-event (May and June 2001) from the sample. The β estimates are in percentage terms. Robust t-statistics are presented in parentheses below the coefficient estimates. * denotes significance at the 10% level.

	Long Leg	Short Leg	Long–Short
Momentum	0.88*	−0.43	1.31
	(1.77)	(−0.72)	(1.52)
Gross profitability	0.16	0.45	−0.29
	(0.37)	(1.16)	(−0.50)
Asset growth	−0.02	−0.60	0.58
	(−0.03)	(−1.31)	(0.70)
Investment to assets	−0.43	−0.50	0.08
	(−0.94)	(−0.84)	(0.10)
Return on assets	0.06	−0.01	0.07
	(0.17)	(−0.02)	(0.11)
Net operating assets	0.14	−0.51	0.66
	(0.32)	(−0.95)	(0.97)
Accruals	0.32	−0.13	0.45
	(0.41)	(−0.19)	(0.39)
Net stock issues	−0.23	−0.40	0.17
	(−0.59)	(−0.77)	(0.24)
Composite equity issues	−0.25	0.34	−0.59
	(−0.51)	(0.62)	(−0.86)
Failure probability	0.28	−0.62	0.90*
	(0.85)	(−1.18)	(1.74)
O-score	−0.48	0.21	−0.69
	(−0.83)	(0.29)	(−0.77)
Combination	0.04	−0.20	0.24
	(0.26)	(−1.18)	(1.02)

drop the two months at the beginning of the pseudo pilot program (May and June 2001) from the sample.

The results of the placebo test are presented in Table VII. The coefficients on $Pilot_i \times PseudoDuring_t$ are mostly statistically insignificant and have mixed signs. When all 11 anomalies are combined, the coefficients are much smaller in magnitude than those in Table IV for the short leg and the long–short portfolio and are statistically insignificant. The placebo test results therefore suggest that our main results are unlikely to be driven by unobserved shocks that affect pilot and nonpilot firms differently.

We next conduct more systematic falsification tests in which we recreate the stratified trading volume design the SEC used in the pilot program for rolling synthetic samples and examine how likely our main results are to arise in a random two-year period.

Specifically, at the end of June of each year τ (from 1991 to 2000), we create the stratified sample following closely the trading volume design the SEC used for the pilot program. We select the largest (in terms of market capitalization at the end of June of year τ) 3,000 stocks. We then rank stocks on each of NYSE, Amex, and NASDAQ based on their average daily trading volume during the prior year (June of year $\tau - 1$ through May of year τ) and select every third stock (beginning with the second one) as a pilot stock and the remainder as nonpilot stocks. We then rerun our main DiD analysis for these rolling stratified samples, keeping only NYSE and Amex stocks to be consistent with the main analysis. For each sample created at the end of June of year τ, we define a pseudo-pilot program that was effective from May of year $\tau + 1$ to July of year $\tau + 3$. We then run the regression

$$r_{it} = \gamma_t + \beta Pilot_i \times PseudoDuring_{\tau,t} + \beta_1 Pilot_i + \epsilon_{it}, \tag{5}$$

where $PseudoDuring_{\tau,t}$ is a dummy variable that equals 1 if month t is between July of year $\tau + 1$ and June of year $\tau + 3$, and the other variables are defined as in equations (1) and (4). The sample period is from January 1980 to June of year $\tau + 3$. We drop the two months at the beginning of the pseudo pilot program (May and June of year $\tau + 1$) from the sample.

None of these falsification tests generates results that are similar to our main results in Table IV. The most positive β for the short-leg portfolios (all 11 anomalies in aggregate) is 0.18% ($t = 0.54$), which is much weaker than our main result, 0.63% ($t = 5.31$). The most negative β for the long-short portfolios (all 11 anomalies in aggregate) is −0.10% ($t = -0.38$), which is much weaker than our main result, −0.72% ($t = -4.37$). This evidence further suggests that our main results are unlikely to arise by chance.

C. Results from NASDAQ Stocks

Our main empirical analysis is carried out on the sample of pilot and nonpilot stocks traded on NYSE/Amex. In this subsection, we conduct a falsification test using the sample of pilot and nonpilot stocks traded on NASDAQ. As stated in the introduction and Section I, the pilot program also removed the bid price test for pilot stocks traded on NASDAQ. However, the bid price test for NASDAQ stocks is not very restrictive, and a significant fraction of trading volume in NASDAQ-listed stocks is executed on ArcaEx and INET, which do not enforce the bid price test (see, e.g., Diether, Lee, and Werner (2009)). We therefore expect at most a minimal effect of the pilot program on anomaly returns of NASDAQ-listed stocks.

This falsification test helps rule out a potential alternative explanation for our main results. Specifically, one may argue that the pilot program made pilot stocks more salient to investors, with the increase in investor attention to pilot stocks weakening anomalies and driving our main results. Since both NASDAQ pilot stocks and NYSE/Amex pilot stocks were included in the pilot program, under this explanation the same shift in investor attention would also occur for NASDAQ pilot stocks during the same pilot period. It follows that if this salience mechanism were driving our main results, we should observe an effect of the pilot program on anomaly returns of NASDAQ stocks similar to that on anomaly returns of NYSE/Amex stocks. On the other hand, if the relaxation of short-sale constraints drives our main results, we would expect to see a minimal effect on anomaly returns of NASDAQ stocks.

We repeat our main DiD analysis (equation (1)) on the sample of pilot and nonpilot stocks traded on NASDAQ National Market. The results are reported in Table VIII. The DiD coefficient β is mostly statistically insignificant and has mixed signs across the 11 anomalies for the long-leg, short-leg, and long–short returns. In aggregate, the coefficient is also small and insignificant for the long-leg, short-leg, and long–short returns. Overall, these results suggest that the pilot program had little effect on anomaly returns for NASDAQ stocks and thus confirm that our main results derive from the relaxation of short-sale constraints generated by the pilot program.

D. The Effect of Short-Sale Constraints on Mispricing of Different Kinds of Stocks

We next explore whether our main results documented in Table IV differ among different classes of stocks. Diether, Lee, and Werner (2009) argue that small and less liquid stocks are likely to be more affected by the suspension of the uptick rule, that is, that the uptick rule impeded short selling more for small and less liquid stocks. The reason is that these stocks have wider spreads and therefore short sellers have to become liquidity providers to ensure compliance with the uptick rule, which makes short-sale orders more passive in the presence of the uptick rule. In addition, for small stocks, a penny tick may be a more significant impediment to shorting them. Consistent with this argument, Diether, Lee, and Werner (2009) find that the suspension of the uptick rule has a more pronounced effect on spreads and some intraday volatility measures for small and less liquid stocks. In our context, we test whether the effect of the pilot program on asset pricing anomalies is more pronounced for small and less liquid stocks. Furthermore, we directly construct a stock-level measure that captures the restrictiveness of the uptick rule using TAQ data and test whether the effect of the pilot program on asset pricing anomalies is more pronounced for stocks that are more restricted by the uptick rule.

We first explore the difference in the effects of the pilot program on small versus large stocks. We split our main sample into small and large stock subsamples based on stock market capitalization at the end of April 2005, before the beginning of the pilot program. Large stocks are those with market capitalization above the median while small stocks are those with market capitalization below the median. We then form anomaly decile portfolios using pilot/nonpilot

Table VIII

Results from NASDAQ Stocks

This table reports the coefficient β from the regression $r_{it} = \gamma_t + \beta Pilot_i \times During_t + \beta_1 Pilot_i + \epsilon_{it}$ for the 11 anomalies individually and for all of them in aggregate. The variable r_{it} is the return of portfolio i in month t, γ_t denotes time fixed effects, $Pilot_i$ equals 1 if portfolio i is formed on pilot firms and zero otherwise, and $During_t$ equals 1 if month t is between July 2005 and June 2007 and zero otherwise. The sample consists of pilot and nonpilot stocks traded on the NASDAQ National Market and the sample period is January 1980 to June 2007. We drop the two months at the beginning of the pilot program (May and June 2005) from the sample. The β estimates are in percentage terms. Robust t-statistics are presented in parentheses below the coefficient estimates. * and ** denote significance at the 10% and 5% levels, respectively.

	Long Leg	Short Leg	Long–Short
Momentum	0.40	−0.68	1.08
	(0.72)	(−0.90)	(1.26)
Gross profitability	−0.41	−0.62	0.21
	(−0.97)	(−0.78)	(0.23)
Asset growth	−0.57	−0.41	−0.16
	(−0.74)	(−0.79)	(−0.20)
Investment to assets	−0.48	−0.87	0.39
	(−0.66)	(−1.41)	(0.43)
Return on assets	−0.19	−1.01	0.82
	(−0.32)	(−1.09)	(0.76)
Net operating assets	−0.71	0.04	−0.75
	(−0.97)	(0.07)	(−0.89)
Accruals	0.30	0.49	−0.19
	(0.47)	(1.10)	(−0.22)
Net stock issues	−0.36	−1.05**	0.69
	(−0.85)	(−2.06)	(0.97)
Composite equity issues	0.60*	0.56	0.05
	(1.76)	(0.83)	(0.07)
Failure probability	−1.01**	−0.35	−0.67
	(−2.10)	(−0.41)	(−0.63)
O-score	−0.86*	−0.30	−0.56
	(−1.76)	(−0.35)	(−0.51)
Combination	−0.30	−0.38	0.08
	(−0.96)	(−1.03)	(0.17)

stocks in these subsamples and repeat the main difference-in-differences analysis (equation (1)) for each subsample. The results are presented in Table IX.

Comparing Panels A and B of Table IX, the effect of the pilot program on asset pricing anomalies is indeed more pronounced among small stocks. When we aggregate over all 11 anomalies, the pilot program reduced the long–short portfolio return by 95 basis points for small stocks compared with 47 basis points for large stocks. The DiD estimate on the short legs is 88 basis points for small stocks compared with 37 basis points for large stocks. The results from subsample analysis split by liquidity are similar to those in Table IX and are reported in Table IA.V in the Internet Appendix.

Table IX

Different Effects on Small and Large Stocks

This table reports the coefficient β from the regression $r_{it} = \gamma_t + \beta Pilot_i \times During_t + \beta_1 Pilot_i + \epsilon_{it}$ for the 11 anomalies individually and for all of them in aggregate. The variable r_{it} is the return of portfolio i in month t, γ_t denotes time fixed effects, $Pilot_i$ equals 1 if portfolio i is formed on pilot firms and zero otherwise, and $During_t$ equals 1 if month t is between July 2005 and June 2007 and zero otherwise. The sample consists of nonpilot and pilot stocks from the pilot program that are traded on NYSE/Amex. Panel A displays results for the subsample of small stocks and Panel B reports results for the subsample of large stocks. The subsample is split based on market capitalization at the end of April 2005. The sample period is January 1980 to June 2007. We drop the two months at the beginning of the pilot program (May and June 2005) from the sample. The β estimates are in percentage terms. Robust t-statistics are presented in parentheses below the coefficient estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Panel A: Small Stocks			Panel B: Large Stocks		
	Long Leg	Short Leg	Long–Short	Long Leg	Short Leg	Long–Short
Momentum	−0.08	0.56	−0.64	0.19	0.74	−0.55
	(−0.15)	(0.61)	(−0.57)	(0.31)	(1.38)	(−0.61)
Gross profitability	−0.30	1.23**	−1.53**	0.35	0.39	−0.05
	(−0.62)	(2.32)	(−2.05)	(0.95)	(0.97)	(−0.09)
Asset growth	0.25	0.69	−0.45	−0.66	0.08	−0.74
	(0.41)	(1.59)	(−0.66)	(−1.35)	(0.24)	(−1.28)
Investment to assets	0.08	0.41	−0.33	−0.12	−0.11	−0.02
	(0.14)	(0.60)	(−0.32)	(−0.26)	(−0.30)	(−0.03)
Return on assets	−0.17	0.97	−1.15	0.16	0.51	−0.35
	(−0.41)	(1.11)	(−1.15)	(0.47)	(0.71)	(−0.45)
Net operating assets	0.03	1.13***	−1.10	−0.20	0.59*	−0.80
	(0.05)	(2.92)	(−1.48)	(−0.48)	(1.68)	(−1.51)
Accruals	−0.25	0.66	−0.92	−0.74	0.82*	−1.55***
	(−0.31)	(1.16)	(−0.92)	(−1.47)	(1.93)	(−2.96)
Net stock issues	−0.33	1.73***	−2.06***	0.12	−0.24	0.36
	(−0.65)	(5.15)	(−3.88)	(0.35)	(−0.70)	(0.70)
Composite equity issues	−0.30	1.27**	−1.57**	−0.79**	−0.17	−0.62
	(−0.58)	(2.39)	(−2.45)	(−2.01)	(−0.46)	(−1.23)
Failure probability	−0.43	0.12	−0.55	0.74*	0.76	−0.02
	(−0.97)	(0.18)	(−0.65)	(1.94)	(1.32)	(−0.03)
O-score	0.72	0.89	−0.17	−0.13	0.70	−0.83
	(1.10)	(1.05)	(−0.13)	(−0.25)	(1.18)	(−1.14)
Combination	−0.07	0.88***	−0.95***	−0.10	0.37***	−0.47**
	(−0.42)	(4.51)	(−3.50)	(−0.73)	(2.59)	(−2.41)

Next, we conduct subsample analysis split by the uptick-rule restrictiveness. We measure the uptick-rule restrictiveness in April 2005 (right before the beginning of the pilot program) using TAQ data as follows.

For each stock on each trading day of April 2005 and at a given time of day, we calculate the minimum shortable price that complies with the uptick rule. We then compare the minimum shortable price with the current bid. If the minimum shortable price is lower than or equal to the bid, then a short seller can potentially execute a short-sale transaction at this price successfully. For each stock on each day of April 2005, we calculate the frequency

Table X

Different Effects on Stocks Split by Uptick-Rule Restrictiveness

This table reports the coefficient β from the regression $r_{it} = \gamma_t + \beta Pilot_i \times During_t + \beta_1 Pilot_i + \epsilon_{it}$ for the 11 anomalies individually and for all of them in aggregate. See the legend of Table IV for the definitions of r_{it}, γ_t, $Pilot_i$, and $During_t$. The sample consists of nonpilot and pilot stocks from the pilot program that are traded on NYSE/Amex. We split the sample into two subsamples based on the uptick-rule restrictiveness measured over April 2005, right before the beginning of the pilot program. For each stock on each trading day of April 2005 and at a given time of day, we calculate the minimum shortable price that complies with the uptick rule. We then compare the minimum shortable price with the current bid. If the minimum shortable price is lower than or equal to the bid, then a short seller can potentially execute a short-sale transaction at this price successfully. The uptick-rule restrictiveness is measured as the average frequency of these potential short-sale transactions under the restriction of the uptick rule over April 2005. Panel A displays results for the subsample of stocks that are more restricted by the uptick rule (low average frequency) and Panel B reports results for the subsample of stocks that are less restricted by the uptick rule (high average frequency). The sample period is January 1980 to June 2007. We drop the two months at the beginning of the pilot program (May and June 2005) from the sample. The β estimates are in percentage terms. Robust t-statistics are presented in parentheses below the coefficient estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Panel A: More restricted stocks			Panel B: Less restricted stocks		
	Long leg	Short leg	Long–short	Long leg	Short leg	Long–short
Momentum	−0.02	1.04	−1.06	0.09	0.26	−0.17
	(−0.04)	(1.22)	(−1.00)	(0.17)	(0.50)	(−0.21)
Gross profitability	−0.58	0.79	−1.37*	0.50	0.80*	−0.30
	(−1.25)	(1.51)	(−1.86)	(1.15)	(1.91)	(−0.51)
Asset growth	0.48	0.80*	−0.33	−0.84*	−0.00	−0.83
	(0.75)	(1.66)	(−0.41)	(−1.76)	(−0.01)	(−1.43)
Investment to assets	−0.19	0.70	−0.88	0.13	−0.38	0.52
	(−0.32)	(1.09)	(−0.87)	(0.33)	(−1.00)	(0.94)
Return on assets	−0.10	1.32	−1.42	0.18	0.14	0.04
	(−0.29)	(1.43)	(−1.38)	(0.58)	(0.23)	(0.06)
Net operating assets	0.44	1.07***	−0.63	−0.68*	0.67	−1.34**
	(0.72)	(2.60)	(−0.81)	(−1.74)	(1.55)	(−2.21)
Accruals	−0.68	1.04*	−1.73**	−0.33	0.52	−0.85*
	(−1.02)	(1.95)	(−2.18)	(−0.92)	(1.23)	(−1.74)
Net stock issues	−0.27	1.50***	−1.76***	0.12	0.02	0.10
	(−0.55)	(3.62)	(−3.40)	(0.30)	(0.05)	(0.18)
Composite equity issues	−0.76	0.96*	−1.72**	−0.32	0.13	−0.45
	(−1.61)	(1.76)	(−2.55)	(−0.81)	(0.36)	(−0.90)
Failure probability	−0.08	0.37	−0.45	0.40	0.52	−0.12
	(−0.19)	(0.60)	(−0.53)	(0.96)	(0.92)	(−0.16)
O-score	0.32	0.98	−0.66	0.38	0.60	−0.23
	(0.44)	(1.15)	(−0.53)	(0.85)	(0.92)	(−0.31)
Combination	−0.13	0.96***	−1.09***	−0.03	0.30**	−0.33*
	(−0.79)	(4.96)	(−4.08)	(−0.26)	(2.05)	(−1.72)

of these potential short-sale transactions under the restriction of the uptick rule. Taking the average of this frequency over all trading days in April 2005, we obtain a stock-level measure of uptick-rule restrictiveness. A stock for which short sales can be carried out more frequently has a lower degree of

Table XI

Results Using Value-Weighted Portfolio Returns

This table reports the coefficient β from the regression $r_{it} = \gamma_t + \beta Pilot_i \times During_t + \beta_1 Pilot_i + \epsilon_{it}$ for the 11 anomalies individually and for all of them in aggregate. The variable r_{it} is the return of portfolio i in month t, γ_t denotes time fixed effects, $Pilot_i$ equals 1 if portfolio i is formed on pilot firms and zero otherwise, and $During_t$ equals 1 if month t is between July 2005 and June 2007 and zero otherwise. The portfolio returns are value weighted. The sample consists of nonpilot and pilot stocks from the pilot program that are traded on NYSE/Amex. The sample period is January 1980 to June 2007. We drop the two months at the beginning of the pilot program (May and June 2005) from the sample. The β estimates are in percentage terms. Robust t-statistics are presented in parentheses below the coefficient estimates. * denotes significance at the 10% level.

	Long Leg	Short Leg	Long–Short
Momentum	0.29	−0.33	0.62
	(0.56)	(−0.61)	(0.75)
Gross profitability	−0.17	0.70	−0.87
	(−0.44)	(1.42)	(−1.39)
Asset growth	−0.33	0.26	−0.59
	(−0.85)	(0.58)	(−1.11)
Investment to assets	0.10	0.34	−0.24
	(0.24)	(0.78)	(−0.40)
Return on assets	0.50	−0.22	0.72
	(1.42)	(−0.39)	(1.23)
Net operating assets	−0.34	0.60	−0.94
	(−0.87)	(1.30)	(−1.36)
Accruals	−0.14	0.54	−0.68
	(−0.26)	(1.20)	(−1.31)
Net stock issues	−0.12	0.02	−0.14
	(−0.37)	(0.04)	(−0.30)
Composite equity issues	−0.52	0.01	−0.53
	(−1.32)	(0.02)	(−0.80)
Failure probability	0.41	0.33	0.07
	(1.14)	(0.54)	(0.11)
O-score	−0.45	0.52	−0.97
	(−1.07)	(1.13)	(−1.60)
Combination	−0.07	0.25*	−0.32*
	(−0.56)	(1.66)	(−1.70)

uptick-rule restrictiveness. We then split our main sample into two subsamples based on their uptick-rule restrictiveness in April 2005. More restricted stocks have above-median restrictiveness while less restricted stocks have below-median restrictiveness. For more restricted stocks, the average waiting time for a short seller before she can short is 1.62 days, which seems a meaningful restriction. The average magnitude of price change (return) during this waiting time is 1.90%. We form anomaly decile portfolios using pilot/nonpilot stocks in these subsamples and repeat the main difference-in-differences analysis (equation (1)) for each subsample. The results are presented in Table X.

Comparing Panels A and B of Table X, the effect of the pilot program on asset pricing anomalies is indeed more pronounced among stocks that are more restricted by the uptick rule. It is also useful to note that the difference between the two subsamples is slightly larger than the difference between small and large stocks in Table IX. When we aggregate over all 11 anomalies, the pilot program reduced the long–short portfolio return by 109 basis points for more restricted stocks compared with 33 basis points for less restricted stocks. The DiD estimate on the short legs is 96 basis points for more restricted stocks compared with 30 basis points for less restricted stocks. The results from this subsample analysis further suggest that the suspension of the uptick rule is the underlying force driving our main results.

As another way to evaluate differences in the effects of the uptick rule on small versus large stocks, we also report the main difference-in-differences results using value-weighted portfolio returns in Table XI and find a small effect. Comparing with the results using gross-return-weighted portfolio returns in Table IV indicates that the uptick rule is indeed more important for small stocks.

V. Conclusion

Using the pilot program of Regulation SHO, which relaxed short-sale constraints for a quasi-random set of stocks, we examine the causal effect of limits to arbitrage, and in particular short-sale constraints, on 11 well-known asset pricing anomalies. We find that the long–short strategies for the 11 anomalies produced much smaller abnormal returns on portfolios constructed with pilot stocks during the pilot period. This result suggests that these anomalies reflect mispricing, and that making arbitrage easier reduces such mispricing. The effect of the pilot program is significant only for the short legs of the anomaly long–short portfolios, which is consistent with the prediction that easy short arbitrage weakens the short side of the anomalies.

We further show that the difference in anomaly returns between pilot and nonpilot stocks vanished after the pilot program ended, as the difference in short-sale constraints between pilot and nonpilot stocks disappeared. In addition, we show that pilot short legs underperformed (outperformed) nonpilot short legs at the beginning (end) of the pilot program. Finally, we show that the difference in anomaly portfolio returns between pilot and nonpilot stocks during the pilot period was more pronounced among small stocks and stocks that were more restricted by the uptick rule before the pilot program. Taken together, these findings provide strong and clear-cut confirmation that limits to arbitrage have a *causal* effect on the strength of well-known asset pricing anomalies and that these anomalies reflect mispricing to a large extent.

Initial submission: July 17, 2016; Accepted: January 8, 2019
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

Appendix A: Definition of Anomaly Variables

The data to construct anomaly variables come from CRSP and annual and quarterly Compustat.

Anomaly 1: Momentum (PRET). A stock's past return $PRET_t$ is calculated as the compounded return over the 11-month ranking period $t-12$ to $t-2$.

Anomaly 2: Gross profitability (GP/A). A firm's gross profitability GP/A_t is calculated as the difference between total revenue $REVT_t$ and cost of goods sold $COGS_t$, scaled by total assets AT_t.

Anomaly 3: Asset growth (AG). A firm's asset growth AG_t is calculated as the change in total assets $AT_t - AT_{t-1}$, scaled by lagged total assets AT_{t-1}.

Anomaly 4: Investment to assets (IVA). Investment to assets IVA_t is defined as the annual change in gross property, plant, and equipment plus the annual change in inventories, $PPEGT_t - PPEGT_{t-1} + INVT_t - INVT_{t-1}$, scaled by lagged total assets AT_{t-1}.

Anomaly 5: Return on assets (ROA). Return on assets ROA_t is measured as quarterly earnings, or income before extraordinary items, IBQ_t, scaled by quarterly total assets ATQ_t.

Anomaly 6: Net operating assets (NOA). Net operating assets is calculated as the difference between operating assets and operating liabilities, scaled by lagged total assets: $NOA_t = (Operating\ Assets_t - Operating\ Liabilities_t)/AT_{t-1}$, where $Operating\ Assets = Total\ Assets\ (AT) - Cash\ and\ Short\text{-}Term\ Investment\ (CHE)$, and $Operating\ Liabilities = Total\ Assets\ (AT) - Short\text{-}Term\ Debt\ (DLC) - Long\text{-}Term\ Debt\ (DLTT) - Minority\ Interests\ (MIB) - Preferred\ Stocks\ (PSTK) - Common\ Equity\ (CEQ)$.

Anomaly 7: Accruals (AC). Operating accruals is measured as the change in noncash current assets, less the change in current liabilities (exclusive of short-term debt and taxes payable), less depreciation and amortization, all divided by lagged total assets: $Accruals_t = [(\Delta Current\ Assets - \Delta Cash) - (\Delta Current\ Liabilities - \Delta Short\text{-}Term\ Debt - \Delta\ Taxes\ Payable) - Depreciation\ and\ Amortization\ Expense)]/AT_{t-1}$. In terms of Compustat item notation, *Current Assets* is ACT, *Cash* is CHE, *Current Liabilities* is LCT, *Short-Term Debt* is DLC, *Taxes Payable* is TXP, and *Depreciation and Amortization* is DP.

Anomaly 8: Net stock issues (NSI). Net stock issues on an annual basis are measured as the change in the natural logarithm of a firm's split-adjusted shares over the last year, $NSI_t = Ln(Adjusted\ Shares_t) - Ln(Adjusted\ Shares_{t-1})$, where $Adjusted\ Shares_t$ is the product of common shares outstanding ($CSHO_t$) and the adjustment factor ($AJEX_t$).

Anomaly 9: Composite equity issues (CEI). Composite equity issues are measured over the past five-year window and are defined as the part of the growth in market equity not attributable to stock returns, $CEI_t = Ln(ME_t/ME_{t-5}) - r(t-5, t)$. In June of year t, for example, ME_t is the market equity at the end of June of year t and ME_{t-5} is the market equity at the end of June of year $t-5$, while $r(t-5, t)$ is the cumulative log return of the stock from the end of June of year $t-5$ to the end of June of year t.

Anomaly 10: Failure probability (FP). Following Campbell, Hilscher, and Szilagyi (2008), we use the coefficients in column 4 of their table IV to construct FP_t, a monotonic transformation of a firm's failure probability. Specifically, FP_t is given as

$$\begin{aligned} FP_t = & -9.16 - 20.26NIMTAAVG_t + 1.42TLMTA_t - 7.13EXRETAVG_t \\ & + 1.41SIGMA_t - 0.045RSIZE_t - 2.13CASHMTA_t \\ & + 0.075MB_t - 0.058PRICE_t, \end{aligned}$$

where details on the variables can be found in Campbell, Hilscher, and Szilagyi (2008).

Anomaly 11: O-score (OS). Following Ohlson (1980) and Chen, Novy-Marx, and Zhang (2011), we construct the O-score as

$$\begin{aligned} OS_t = & -1.32 - 0.407log(ADJASSET_t) + 6.03TLTA_t - 1.43WCTA_t \\ & + 0.076CLCA_t - 1.72OENEG_t - 2.37NITA_t - 1.83FUTL_t \\ & + 0.285INTWO_t - 0.521CHIN_t, \end{aligned}$$

where $ADJASSET$ is adjusted total assets and equals total assets (ATQ)+0.1× (market equity-book equity), $TLTA$ is the leverage ratio and equals the book value of debt (DLCQ plus DLTTQ) divided by $ADJASSET$, $WCTA$ is working capital (ACTQ minus LCTQ) divided by $ADJASSET$, $CLCA$ is current liabilities (LCTQ) divided by current assets (ACTQ), $OENEG$ is one if total liabilities (LTQ) exceeds total assets (ATQ) and is zero otherwise, $NITA$ is net income (NIQ) divided by $ADJASSET$, $FUTL$ is funds provided by operations (PIQ) divided by liabilities (LTQ), $INTWO$ is equal to 1 if net income (NIQ) is negative for the last two quarters and zero otherwise, and $CHIN$ is $(NI_t - NI_{t-1})/(|NI_t| + |NI_{t-1}|)$, where NI is net income (NIQ).[15]

[15] Further details on these variables can be found in the Appendix of Chen, Novy-Marx, and Zhang (2011). They have $-0.407\log(ADJASSET_t/CPI_t)$, where CPI is the consumer price index, as the second term of O-score. As scaling $ADJASSET$ by CPI does not affect the cross-sectional sorting of stocks based on O-score, we drop CPI in our calculation of O-score.

Appendix B: Benchmark-Adjusted Returns

In the main empirical analysis, we use raw portfolio returns as the dependent variable in equation (1). In this subsection, we test the validity of our main results when we use benchmark-adjusted portfolio returns as the dependent variable in equation (1). If the mean return premia of the Fama–French three factors represent rational risk premia (an issue on which we do not take a

Table BI

Benchmark-Adjusted Returns

This table reports the coefficient β from the regression $r_{it} = \gamma_t + \beta Pilot_i \times During_t + \beta_1 Pilot_i + \epsilon_{it}$ for the 11 anomalies individually and for all of them in aggregate, with benchmark-adjusted returns used as the dependent variable. The variable r_{it} is the benchmark-adjusted return of portfolio i in month t, γ_t denotes time fixed effects, $Pilot_i$ equals 1 if portfolio i is formed on pilot firms and zero otherwise, and $During_t$ equals 1 if month t is between July 2005 and June 2007 and zero otherwise. Panel A displays results for CAPM-adjusted returns while Panel B displays results for Fama–French three-factor-adjusted (FF-adjusted) returns. The sample consists of nonpilot and pilot stocks from the pilot program that are traded on NYSE/Amex. The sample period is January 1980 to June 2007. We drop the two months at the beginning of the pilot program (May and June 2005) from the sample. The β estimates are in percentage terms. Robust t-statistics are presented in parentheses below the coefficient estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Panel A: CAPM-Adjusted Returns			Panel B: FF-Adjusted Returns		
	Long Leg	Short Leg	Long–Short	Long Leg	Short Leg	Long–Short
Momentum	0.03	0.65	−0.62	0.01	0.66	−0.65
	(0.07)	(1.37)	(−0.92)	(0.03)	(1.39)	(−0.93)
Gross profitability	−0.04	0.81**	−0.85*	−0.05	0.80**	−0.86*
	(−0.13)	(2.24)	(−1.81)	(−0.17)	(2.23)	(−1.85)
Asset growth	−0.18	0.42	−0.60	−0.19	0.43*	−0.62
	(−0.46)	(1.64)	(−1.40)	(−0.48)	(1.67)	(−1.41)
Investment to assets	−0.01	0.15	−0.16	−0.02	0.16	−0.19
	(−0.01)	(0.43)	(−0.27)	(−0.06)	(0.45)	(−0.30)
Return on assets	0.05	0.74	−0.69	0.03	0.73	−0.70
	(0.23)	(1.26)	(−1.08)	(0.14)	(1.25)	(−1.09)
Net operating assets	−0.11	0.88***	−0.99**	−0.12	0.87***	−1.00**
	(−0.32)	(3.50)	(−2.11)	(−0.34)	(3.46)	(−2.10)
Accruals	−0.51	0.80**	−1.31***	−0.51	0.80**	−1.31***
	(−1.41)	(2.29)	(−2.61)	(−1.45)	(2.24)	(−2.73)
Net stock issues	−0.08	0.77***	−0.85**	−0.08	0.79***	−0.86**
	(−0.26)	(3.24)	(−2.25)	(−0.26)	(3.22)	(−2.26)
Composite equity issues	−0.56*	0.54*	−1.10***	−0.55*	0.54*	−1.09***
	(−1.80)	(1.78)	(−2.73)	(−1.79)	(1.78)	(−2.75)
Failure probability	0.16	0.45	−0.29	0.14	0.45	−0.30
	(0.56)	(1.12)	(−0.62)	(0.48)	(1.12)	(−0.64)
O-score	0.36	0.80	−0.44	0.35	0.79	−0.43
	(0.76)	(1.39)	(−0.54)	(0.73)	(1.40)	(−0.54)
Combination	−0.08	0.64***	−0.72***	−0.09	0.64***	−0.73***
	(−0.76)	(5.34)	(−4.36)	(−0.83)	(5.36)	(−4.39)

stand here), then this analysis would indicate whether relaxation of short-sale constraints reduces mispricing measured against this benchmark.

To obtain benchmark-adjusted returns, we first regress the time series of excess returns onto the time series of appropriate factors (the market excess return for the CAPM, and the market excess return plus two additional factors, the SMB and HML factors, for the Fama–French three-factor model). We then obtain the time series of benchmark-adjusted returns as the constant plus the residuals from the regression.

We would not expect this factor adjustment to affect the results much, as the selection of pilot firms is quasi-random, implying similar loadings on the benchmark factors of pilot versus nonpilot firms. The results for benchmark-adjusted returns are presented in Table BI. Consistent with this intuition, all of the β estimates are similar to those in Table IV of main results.

Appendix C: Power of the Main Tests

Our main tests for Hypotheses 1 and 2 based on equation (1) use a relatively short (two-year) pilot period to estimate the effect of relaxing short-sale constraints on anomaly returns. Here, we address the issue of power explicitly. Intuitively, our main tests using equation (1) gain power via two means. First is the aggregation across 11 anomalies. Second, even for a single anomaly during the pilot period, what is relevant for our test is not the raw strength of that anomaly, but rather the *difference* in the strength of an anomaly between pilot versus nonpilot firms over the same time period.[16] This differencing effectively hedges away much of the factor volatility of returns, greatly increasing the precision of the test. To see this in a very simple way, suppose that the momentum return of Portfolio A were equal to the return of Portfolio B plus a constant. Then even if both portfolios were highly volatile, the difference in returns would be a constant, implying that the difference would be significant with an infinite t-statistic. Of course a constant difference is unrealistic, but this example illustrates that testing for a difference filters out a large amount of variation from the test.

Consistent with this point, in Table IA.VI in the Internet Appendix, we show that taking differences between portfolios constructed with pilot and nonpilot stocks substantially reduces return volatility. Monthly standard deviations of return differences between long-leg/short-leg portfolios constructed with pilot and nonpilot stocks are much smaller than those of returns on long-leg/short-leg portfolios themselves. For example, averaged across the 11 anomalies, the monthly standard deviation of return differences between pilot and nonpilot stocks for the short leg is 1.71%, while the monthly standard deviation of short-leg returns is 3.64% for nonpilot stocks and 3.70% for pilot stocks.

This contrasts with conventional tests for estimating average anomaly returns (rather than differences in returns), in which sampling noise derived

[16] Econometrically, this is achieved by including anomaly-time fixed effects in our regression specification. Also, our tests actually examine the difference in this difference between the pilot and nonpilot periods, but this is not crucial for our argument.

from factor realizations reduces power. In such tests, much longer time periods are often needed to confirm an anomaly reliably. It is of course sometimes possible to identify anomaly returns using a sample period measured in years rather than decades. For example, in an out-of-sample test of their 1993 paper (Jegadeesh and Titman (1993)), Jegadeesh and Titman (2001) find significant momentum in the 1990 to 1998 sample period (nine years), with a t-statistic of 4.71.

A further consideration that enhances the power of our tests is that the pilot period is one in which the 11 anomalies are relatively strong. If we estimate the mean anomaly returns (long-minus-short returns) on *nonpilot stocks* during the pilot period, the anomalies tend to be stronger, both economically and statistically, than might ordinarily be expected for a two-year period.

Specifically, we calculate the mean monthly anomaly returns and CAPM/Fama–French three-factor alphas for the 11 anomalies individually and

Table CI

Asset Pricing Anomalies During the Pilot Period

This table presents the mean monthly raw return, the CAPM α, and the Fama–French three-factor α of the 11 asset pricing anomalies individually and in aggregate during the pilot period from July 2005 to June 2007. The sample consists of nonpilot and pilot stocks from the pilot program that are traded on NYSE/Amex. The anomaly portfolios are constructed with nonpilot stocks. Robust t-statistics are presented in parentheses below the estimates. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Raw	CAPM α	FF α
Momentum	1.02	1.11	0.84
	(1.47)	(1.24)	(0.92)
Gross profitability	0.47	0.24	0.40
	(1.24)	(0.64)	(1.06)
Asset growth	0.50	0.71**	0.82**
	(1.33)	(2.26)	(2.27)
Investment to assets	0.31	0.66	1.02*
	(0.62)	(1.35)	(1.85)
Return on assets	0.83	0.82	0.79
	(1.54)	(1.27)	(1.09)
Net operating assets	0.63	0.86**	1.16***
	(1.71)	(2.09)	(2.99)
Accruals	−0.11	0.17	−0.27
	(−0.24)	(0.31)	(−0.53)
Net stock issues	0.42	0.53	0.74**
	(1.30)	(1.63)	(2.24)
Composite equity issues	0.17	0.42	0.43
	(0.54)	(1.27)	(1.33)
Failure probability	1.02*	1.20*	1.31**
	(1.85)	(1.75)	(2.12)
O-score	0.39	0.37	0.38
	(0.88)	(0.93)	(0.96)
Combination	0.51***	0.64***	0.69***
	(2.82)	(3.12)	(3.22)

in aggregate over the pilot period from July 2005 to June 2007. Table CI presents the results for nonpilot stocks. As can be seen, the anomaly returns and alphas of the 11 anomalies for nonpilot stocks are mostly positive (31 out of 33). When we combine the 11 anomalies together, both the mean return and the alphas are positive and statistically significant. The magnitudes are also large—the mean monthly return and alphas are about 51bps to 69 bps when the 11 anomalies are combined.

REFERENCES

Ali, Ashiq, Lee-Seok Hwang, and Mark A. Trombley, 2003, Arbitrage risk and the book-to-market anomaly, *Journal of Financial Economics* 69, 355–373.

Ali, Ashiq, and Mark A. Trombley, 2006, Short sales constraints and momentum in stock returns, *Journal of Business Finance & Accounting* 33, 587–615.

Asparouhova, Elena, Hendrik Bessembinder, and Ivalina Kalcheva, 2010, Liquidity biases in asset pricing tests, *Journal of Financial Economics* 96, 215–237.

Asparouhova, Elena, Hendrik Bessembinder, and Ivalina Kalcheva, 2013, Noisy prices and inference regarding returns, *Journal of Finance* 68, 665–714.

Asquith, Paul, Parag A. Pathak, and Jay R. Ritter, 2005, Short interest, institutional ownership, and stock returns, *Journal of Financial Economics* 78, 243–276.

Baker, Malcolm, and Serkan Savaşoglu, 2002, Limited arbitrage in mergers and acquisitions, *Journal of Financial Economics* 64, 91–115.

Barberis, Nicholas, and Richard Thaler, 2003, A survey of behavioral finance, in George M. Constantinides, Milton Harris, and René M. Stulz, eds.: Handbook of the Economics of Finance (North-Holland, Amsterdam).

Beneish, Messod D., Charles MC Lee, and D. Craig Nichols, 2015, In short supply: Short-sellers and stock returns, *Journal of Accounting and Economics* 60, 33–57.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *Journal of Finance* 63, 2899–2939.

Cao, Bing, Dan S. Dhaliwal, Adam C. Kolasinski, and Adam V. Reed, 2007, Bears and numbers: Investigating how short sellers exploit and affect earnings-based pricing anomalies, Working paper, University of Arizona, University of Washington, and University of North Carolina.

Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* 66, 171–205.

Chen, Long, Robert Novy-Marx, and Lu Zhang, 2011, An alternative three-factor model, Working paper, Washington University in St. Louis, University of Rochester, and Ohio State University.

Cohen, Lauren, Karl B. Diether, and Christopher J. Malloy, 2007, Supply and demand shifts in the shorting market, *Journal of Finance* 62, 2061–2096.

Cooper, Michael J., Huseyin Gulen, and Michael J. Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance* 63, 1609–1651.

Daniel, Kent, and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *Journal of Finance* 61, 1605–1643.

D'Avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.

Dichev, Ilia D., 1998, Is the risk of bankruptcy a systematic risk? *Journal of Finance* 53, 1131–1147.

Diether, Karl B., Kuan-Hui Lee, and Ingrid M. Werner, 2009, It's SHO time! Short-sale price tests and market quality, *Journal of Finance* 64, 37–73.

Drechsler, Itamar, and Qingyi Freda Drechsler, 2014, The shorting premium and asset pricing anomalies, Working paper, New York University and Wharton Research Data Services.

Engelberg, Joseph E., Adam V. Reed, and Matthew C. Ringgenberg, 2018, Short-selling risk, *Journal of Finance* 73, 755–786.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2006, Profitability, investment and average returns, *Journal of Financial Economics* 82, 491–518.

Fang, Vivian W., Allen H. Huang, and Jonathan M. Karpoff, 2016, Short selling and earnings management: A controlled experiment, *Journal of Finance* 71, 1251–1294.

Geczy, Christopher C., David K. Musto, and Adam V. Reed, 2002, Stocks are special too: An analysis of the equity lending market, *Journal of Financial Economics* 66, 241–269.

Gromb, Denis, and Dimitri Vayanos, 2010, Limits of arbitrage, *Annual Review of Financial Economics* 2, 251–275.

Grullon, Gustavo, Sébastien Michenaud, and James P. Weston, 2015, The real effects of short-selling constraints, *Review of Financial Studies* 28, 1737–1767.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, … and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Hirshleifer, David, 2001, Investor psychology and asset pricing, *Journal of Finance* 56, 1533–1597.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics* 38, 297–331.

Hirshleifer, David, Siew Hong Teoh, and Jeff Jiewei Yu, 2011, Short arbitrage, return asymmetry and the accrual anomaly, *Review of Financial Studies* 24, 2429–2461.

Israel, Ronen, and Tobias J. Moskowitz, 2013, The role of shorting, firm size, and time on market anomalies, *Journal of Financial Economics* 108, 275–301.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jegadeesh, Narasimhan, and Sheridan Titman, 2001, Profitability of momentum strategies: An evaluation of alternative explanations, *Journal of Finance* 56, 699–720.

Jones, Charles M., and Owen A. Lamont, 2002, Short-sale constraints and stock returns, *Journal of Financial Economics* 66, 207–239.

Lam, F.Y. Eric C., and K.C. John Wei, 2011, Limits-to-arbitrage, investment frictions, and the asset growth anomaly, *Journal of Financial Economics* 102, 127–149.

Lamont, Owen A., and Richard H. Thaler, 2003, Can the market add and subtract? Mispricing in tech stock carve-outs, *Journal of Political Economy* 111, 227–268.

Li, Yinghua, and Liandong Zhang, 2015, Short selling pressure, stock price behavior, and management forecast precision: Evidence from a natural experiment, *Journal of Accounting Research* 53, 79–117.

Loughran, Tim, and Jay R. Ritter, 1995, The new issues puzzle, *Journal of Finance* 50, 23–51.

Mashruwala, Christina, Shivaram Rajgopal, and Terry Shevlin, 2006, Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs, *Journal of Accounting and Economics* 42, 3–33.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Ohlson, James A., 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 18, 109–131.

Pontiff, Jeffrey, 1996, Costly arbitrage: Evidence from closed-end funds, *Quarterly Journal of Economics* 111, 1135–1151.

Pontiff, Jeffrey, and Artemiza Woodgate, 2008, Share issuance and cross-sectional returns, *Journal of Finance* 63, 921–945.

Reed, Adam V., 2013, Short selling, *Annual Review of Financial Economics* 5, 245–258.

Saffi, Pedro A.C., and Kari Sigurdsson, 2010, Price efficiency and short selling, *Review of Financial Studies* 24, 821–852.

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.

Sloan, Richard, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71, 289–315.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.

Titman, Sheridan, Kuo-Chiang Wei, and Feixue Xie, 2004, Capital investments and stock returns, *Journal of Financial and Quantitative Analysis* 39, 677–700.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.
Replication code.