

**Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.**

Zeroing in on the Expected Returns of Anomalies

Andrew Y. Chen and Mihail Velikov

2020-039

Please cite this paper as:

Chen, Andrew Y., and Mihail Velikov (2020). "Zeroing in on the Expected Returns of Anomalies," Finance and Economics Discussion Series 2020-039. Washington: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2020.039>.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

Zeroing in on the Expected Returns of Anomalies

Andrew Y. Chen

Mihail Velikov

Federal Reserve Board

Pennsylvania State University

andrew.y.chen@frb.gov

velikov@psu.edu

May 2020*

Abstract

We zero in on the expected returns of long-short portfolios based on 120 stock market anomalies by accounting for (1) effective bid-ask spreads, (2) post-publication effects, and (3) the modern era of trading technology that began in the early 2000s. Net of these effects, the average anomaly's expected return is a measly 8 bps per month. The strongest anomalies return only 10-20 bps after accounting for data-mining with either out-of-sample tests or empirical Bayesian methods. Expected returns are negligible despite cost optimizations that produce impressive net returns in-sample and the omission of additional trading costs like price impact.

*First posted to SSRN: November 2017. This paper originated from a conversation with Svetlana Bryzgalova. We thank Marie Briere (HFPE discussant), Victor DeMiguel, Yesol Huh, Nina Karnaukh, Alberto Martin-Utrera (FDU discussant), Andy Neuhierl, Steve Sharpe, Nitish Sinha, Ingrid Tierens (Jacobs Levy discussant), Tugkan Tuzun, Michael Weber, Haoxiang Zhu, and seminar participants at the Federal Reserve Board, Penn State University, University of Georgia, the 11th Annual Hedge Fund and Private Equity Research Conference, 2019 Finance Down Under Meetings, 2019 Eastern Finance Association Meetings, and 2019 Jacobs Levy Frontiers in Quantitative Finance conference for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the position of the Board of Governors of the Federal Reserve or the Federal Reserve System.

1. Introduction

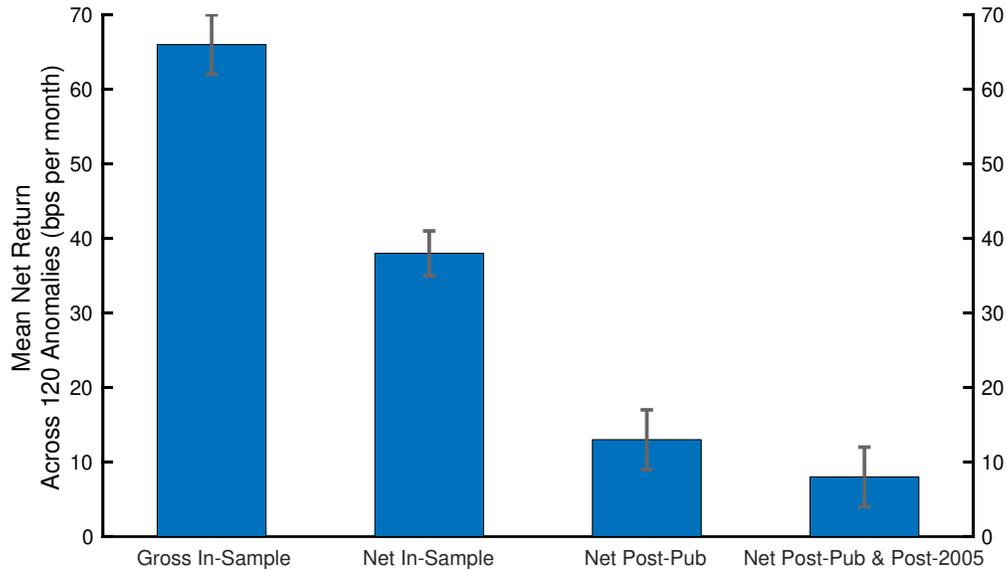
The literature on stock market anomalies has documented more than one hundred predictors of the cross-section of stock returns. Using historical data, these papers demonstrate market-neutral returns that average around 8% per year. These anomalies range from those based on past return patterns, to those based purely on accounting variables, and still others based on institutional stock holdings. Few economic risk factors or behavioral theories are so broad that they can make a dent in this wide variety of return predictors.

Anomalies' expected returns, however, may be much lower than the mean returns found in the literature. With only a couple exceptions, the literature ignores trading costs, which can significantly reduce expected payoffs, and thus expected returns. Moreover, the historical data used in these papers are stale. The literature uses data going back to the 1920s, leading to questions about whether returns from so long ago are still relevant. Indeed, data-mining bias and investor learning imply that returns in recent years have been much smaller (McLean and Pontiff 2016). And the early 2000s saw a revolution in information and trading technologies, implying that data from earlier decades may not be informative about the future (Chordia, Subrahmanyam, and Tong 2014).

In this paper, we zero in on the expected returns of anomalies by accounting for both trading costs and the staleness of historical data. Our main result is that, net of these effects, expected returns are effectively zero.

Figure 1 illustrates how we “zero in.” To generate this figure, we replicate 120 anomaly signals, construct long-short portfolios using state-of-the-art cost mitigation techniques, and reduce portfolio payoffs by half of the effective bid-ask spread whenever a portfolio weight is adjusted. Each bar, moving from left to

Figure 1: Anomaly Mean Long-Short Returns. Error bars show one standard error.



right, provides a more refined estimate of the average anomaly's expected return.

The first bar is the mean return before trading costs (gross return) within the original papers' sample periods (in-sample). In our dataset we find an impressive 66 bps per month. Accounting for trading costs reduces the expected return to 38 bps, which is still a notable 4.6% per year. Adding post-publication effects, however, results in a measly 13 bps per month. Additionally restricting the sample to the modern era of trading technology (post-2005), we should expect a negligible 8 bps per month.¹ These results omit additional trading costs such as price impact and short-sale fees. Indeed, short-sale costs average 10-20 basis points per month (Cohen, Diether, and Malloy 2007; Drechsler and Drechsler 2016), and would wipe out the remaining profits.

Though the average anomaly is unprofitable, perhaps the strongest anomalies still offer large expected returns. Indeed, seven anomalies have mean net returns in excess of 60 bps per month in the the data that is both post-publication

¹We thank Marie Briere for suggesting this analysis. Post-2003 and post-2004 samples lead to similar results.

and post-2005. This performance should be viewed with suspicion, however, as some portion of it must be due to luck. Indeed, reporting only anomalies with the largest mean returns is the very definition of data-mining.

We find, however, that even the strongest anomalies offer negligible expected returns. To come to this conclusion, we use two data-mining adjustments that have distinct statistical motivations. Despite their different origins, but both adjustments lead to the same quantitative result.

The first data-mining adjustment is a simple out-of-sample test. We sort anomalies based on information available in their in-sample periods, and then average post-publication and post-2005 net returns within quantiles. This exercise avoids using the same data to select and make inference on anomalies, thus eliminating data-mining bias. We examine four in-sample predictors: the net return, the net Sharpe ratio, the return reduction due to trading costs, and turnover.

The best expected returns come from sorting anomalies on their in-sample net Sharpe ratios. The top quartile of has a mean net return of 21 bps per month in post-publication and post-2005 data. But net returns are not monotonic, and the second strongest predictor (turnover) produces at most only 14 bps per month. Moreover, these results require using equal-weighted implementations. Restricting our implementations to value-weighting implies expected returns of 11 bps per month, at best.

The second data-mining adjustment uses an empirical Bayes estimator. These estimators have been shown to effectively adjust for data-mining in a wide variety of settings (Efron 2012; Azevedo et al. 2019; Liu, Moon, and Schorfheide 2020; Chen and Zimmermann 2019). Such estimators compare the cross-anomaly dispersion of mean returns to their standard errors to determine how much dispersion is due to luck. The overall contribution of luck is estimated

from empirical data using frequentist methods, and then adjustments for individual anomalies are derived using Bayesian formulas, hence the name “empirical Bayes.”

This estimation finds that most of the dispersion in mean net returns in post-publication and post-2005 data is due to luck. As a result, even the 90th percentile anomaly has an expected return of 20 bps per month after adjusting for data-mining. Even worse, implementations that use only value-weighting produce only 6 bps in the 90th percentile. These results are remarkably consistent with our first data-mining adjustment, despite their very different methodologies.

Our data-mining results are intuitive given the distribution of mean net returns in recent data. The distribution closely resembles a standard normal distribution. Only 9% of t-stats exceed 2.0 in absolute value, not far from the 5% implied by a standard normal. Thus, the data can be largely explained by the null of no predictability, and returns in the right tail of the distribution are mostly due to luck.

These results may be surprising, as other papers find size, B/M, and momentum survive trading costs (Novy-Marx and Velikov 2016; Frazzini, Israel, and Moskowitz 2015; Briere et al. 2019). Individual anomalies, however, have noisy mean returns that are very sensitive to the sample period. Size, B/M, and momentum have positive net returns of 30 to 70 bps in the 1998-2013 sample studied by Frazzini, Israel, and Moskowitz (2015), but their net returns drop to between -30 and +25 bps post-2005. This change in performance is consistent with the fact that standard errors on mean returns are around 40 bps per month for samples of 15 years. This fragility demonstrates the importance of aggregating across many anomalies, as we do in our paper.

A limitation of our study is that we do not allow for combining multiple

anomalies. Combining anomalies can improve portfolio performance, particularly when accounting for trading costs (Novy-Marx and Velikov 2016, for example). Indeed, DeMiguel et al. (Forthcoming) apply state-of-the-art optimization techniques to 50 anomalies, and find that combining anomalies has very powerful effects on trading costs in the long historical sample. Combining anomalies, however, does not allow for sharp inferences about more recent data. It is only by averaging over 120 anomalies that we can obtain the small standard errors in Figure 1. Indeed, our large dataset allows us to make sharp inferences about the recent performance of the best anomalies. Both data-mining adjustments produce standard errors on mean net returns of around 5-10 bps.

Another limitation is that we measure trading costs with effective spreads. Spreads are a lower bound trading cost because they correspond to the smallest market orders, but one might argue that even lower costs can be obtained with the strategic use of limit orders. Indeed, Frazzini, Israel, and Moskowitz (2018) argue that traders can act as market makers and pay negative spreads, receiving rather than paying trading costs. Acting as a market maker, however, results in adverse selection costs (Glosten and Milgrom 1985) and execution risk (Cont and Kukanov 2017). Moreover, theory suggests that there are fundamental trading costs that cannot be avoided regardless of the implementation (Kyle and Obizhaeva 2016), and empirical studies find that effective bid-ask spreads are closely related to these fundamental costs (Fong, Holden, and Tobek 2017).

In the remainder of the Introduction, we relate our study to existing literature. Section 2 describes our methods. Section 3 presents results for the average anomaly. Section 4 examines the strongest anomalies. We examine size, B/M, and momentum in Section 4.3. Section 5 concludes.

Relation to the Literature In a closely-related study, Novy-Marx and Velikov (2016) (NV) find that trading costs have a large effect on the mean returns of twenty-three anomalies. However, there are several reasons why NV’s results do not allow for inference about expected returns on the “anomaly zoo” (McLean and Pontiff 2016; Freyberger, Neuhierl, and Weber 2017; Feng, Giglio, and Xiu 2017; etc)

First, and foremost, the anomalies studied in NV are not representative of the anomaly zoo. NV’s anomalies are “twenty-three of the best known, and strongest performing, anomaly strategies.” In contrast, anomaly zoo papers like McLean and Pontiff (2016) are drawn from a more-or-less exhaustive literature search, and include dozens of anomalies that are not popularly known.²

Unlike NV, our anomalies include all 68 of MP’s anomalies that allow for cost optimization and 52 additional anomalies from Green, Hand, and Zhang (2017) and Hou, Xue, and Zhang (2017). This large set of anomalies also differentiates our paper from other trading cost studies, all of which examine small sets of well-known anomalies (Frazzini, Israel, and Moskowitz 2015, and Briere et al. 2019, for example).³ Moreover, we reconcile our results with studies of selected anomalies by using data-mining adjustments. These adjustments allow us to study the strongest anomalies using objective statistics, unlike previous papers which use judgment to determine notable anomalies.

The second reason NV’s results cannot be used to study expected returns is

²McLean and Pontiff’s anomalies “were mostly identified with search engines such as Econlit by searching for articles in finance and accounting journals using words such as ‘cross-section.’ ” David McLean informed us that they also surveyed asset pricing experts to make sure they were not missing anything.

³For other trading cost studies, see Stoll and Whaley (1983), Schultz (1983), Ball, Kothari, and Shanken (1995), Knez and Ready (1996), Pontiff and Schill (2001), Korajczyk and Sadka (2004), Lesmond, Schill, and Zhou (2004), and Hanna and Ready (2005), McLean (2010), Hou, Kim, and Werner (2016), Patton and Weller (2017), Frazzini, Israel, and Moskowitz (2015), and Briere et al. (2019). For other papers on the decay of predictability over time, see Schwert (2003), Marquering, Nisser, and Valla 2006, Huang and Huang 2013, Chordia, Subrahmanyam, and Tong (2014), Jacobs and Müller (2017), Chu, Hirshleifer, and Ma (2017), and Chen and Zimmermann (2019).

that NV's trading cost exhibits a large upward bias in recent years. NV measure trading costs using Hasbrouck's (2009) low-frequency spreads, and as we show in Section 2.2, low-frequency spreads are upward biased by 25-50 bps after 2003. This bias is consistent with changes in the trading environment since decimalization (Jahan-Parvar and Zikes 2019).

We account for this bias by using high-frequency spreads from NYSE's Trade and Quote (TAQ) database. Spreads from TAQ serve as benchmark measures of liquidity in the microstructure literature (Goyenko, Holden, and Trzcinka 2009; Fong, Holden, and Trzcinka 2017), and indeed all low frequency (LF) spreads demonstrate their validity by examining their correlations with HF spreads (Corwin and Schultz 2012, for example).

Finally, it is not obvious how to combine NV's trading cost effects with the performance decay found in other papers. While performance decay tends to reduce net returns in recent data, trading costs have plummeted too, with opposite effects. Moreover, theories of limited arbitrage predict that performance decay and trading costs interact cross-sectionally, implying that the measurement of expected returns must be done at the anomaly level. To accommodate these interactions, we provide the first joint study of trading costs and performance decay that uses empirical trading cost data.⁴

2. Anomalies Data, Trading Cost Measurement, and Portfolio Implementations

Here we describe our methods. We begin with the anomalies data (Section 2.1), then describe trading cost measurement (Section 2.2), and then describe

⁴Huang and Huang (2013) also examine trading costs and post-publication returns for many anomalies and find that expected returns are positive, but they impute trading costs based on statistics reported in the literature and study only 14 anomalies.

our portfolio implementations (Section 2.3).

2.1. Anomalies Data

Our anomalies dataset is created from Chen and Zimmermann's (2019) (CZ's) set of 156 cross-sectional return predictors from 115 publications in accounting, economics, and finance journals. This dataset contains all 97 from McLean and Pontiff (2016) and adds 59 predictors from Green, Hand, and Zhang (2017), Hou, Xue, and Zhang (2017), and Harvey, Liu, and Zhu (2016).

Chen and Zimmermann show that their replicated predictors perform quite well. The average in-sample (original publication's sample) return is 0.72% per month, with an average t-stat of 4.3. Moreover, their in-sample returns are very similar to hand collected statistics from the original publications, differing by only a handful of basis points on average.

We exclude 34 predictors that have difficult-to-evaluate trading costs. Many of these predictors are created from event studies (such as Ritter's (1991) study of long-run IPO performance) that are difficult to compare with predictors that change on a regular basis. In particular, the optimal rebalancing of event study-based portfolios is difficult to determine, and rebalancing has a large effect when examining trading costs. We also exclude predictors that are too discrete to be used in our trading cost mitigation techniques such as Hong and Kacperczyk's (2009) sin stock classification. Continuity is important, because our most reliable cost mitigation, the buy-hold spread, relies on the continuity of the predictor for more efficient rebalancing.

We also exclude the Fama and MacBeth (1973) CAPM beta and Kelly and Jiang's (2014) tail risk factor because some academics may object that these are not anomalies. Nevertheless, including them has almost no effect on our results.

The anomalies are constructed from the usual data sources. More than half of the predictors focus on Compustat data, and about 30% use purely price data. Most of the remainder use analyst forecasts, though several focus on institutional ownership data, trading volume, or specialized data (such as Gompers, Ishii, and Metrick's (2003) governance index). Appendix A.1 provides a list of the anomalies. For further details, please see Chen and Zimmermann (2019).

2.2. Trading Cost Measurement

We measure returns before trading costs using the ubiquitous monthly CRSP data. Then to adjust for trading costs, we track portfolio weights, and each time a position is entered or exited, we assume the effective half spread is paid. This notion of trading costs is also studied in Hanna and Ready (2005), Korajczyk and Sadka (2004), and Novy-Marx and Velikov (2016).

To understand this trading cost measure, it helps to know that prices in CRSP are predominately determined by closing auctions.⁵ The hypothetical anomaly portfolios studied by academics would have added additional demand or supply to these auctions, increasing the prices for buys and decreasing the prices for sells. These price deviations, then, would reduce returns compared to the CRSP benchmark. Our trading cost aims to measure the minimal amount by which these prices would have been moved.⁶ An alternative method for measuring trading costs is to exclusively use intraday data as in Knez and Ready (1996), but this would deviate significantly from the anomalies literature which is based on closing auction prices.

Our measure of the minimal price deviation is the effective half bid-ask

⁵The NYSE and NASDAQ closing auctions are described at <https://www.nyse.com/article/nyse-closing-auction-insiders-guide> and <https://www.nasdaqtrader.com/content/productsservices/Trading//ClosingCrossfaq.pdf>.

⁶We are grateful to Haoxiang Zhu for suggesting this interpretation.

spread—that is, the absolute difference between the trade price and the prevailing quoted midpoint. Supposing that the prevailing midpoint is an unbiased estimated of the frictionless price, a buy trade “overpays” by the effective half spread and a sell trade receives too little by the same amount. Effective spreads use trades that are actually executed, and typically imply smaller spreads than quoted prices due to price improvement (Stoll 2003).

We use high-frequency HF data to compute spreads whenever it is available. Our HF data combines the Daily TAQ, Monthly TAQ, and ISSM datasets. Computation of spreads follows Holden and Jacobsen (2014) (HJ) closely.⁷ To match the monthly data frequencies used in the anomalies literature, we first aggregate to a daily level by taking a share-weighted average of intra-day spreads, and then aggregate across days within each month by taking a simple average following Hanna and Ready (2005) and others. Anomaly returns are measured using end-of-month closing prices and thus one may argue that end-of-month spreads are a better match. However, averaging across the month ensures that our spreads are not sensitive to outliers. For additional details see Appendix A.2.

Our HF data provide a mostly continuous history of transactions on the NYSE and AMEX from 1983-2016.⁸ These datasets are sufficient for estimating trading costs of anomalies post-publication, as 97% of anomalies are published after 1983. However, we also wish to study the effects of cost optimization, and to avoid data-mining bias we run our optimizations on pre-publication data.

Thus, we compute effective spreads pre-1983 (and whenever HF data is missing) using low frequency (LF) proxies based on daily CRSP data. Rather than

⁷We are grateful to Craig Holden for providing SAS code on his website.

⁸Data for NASDAQ stocks is somewhat shorter (1987-2016), as ISSM is missing NASDAQ data before 1987. The older ISSM data also features several gaps in data. NASDAQ data is missing in April and May 1987, April and July 1988, November and December 1989. In addition, there are 46 trading days with no data for NASDAQ stocks between 1987 and 1991, and 146 trading days with no data for NYSE/AMEX. These data gaps are also found by Barber, Odean, and Zhu (2008).

choose any particular LF proxy, we compute four different LF proxies and use the simple average as our spread. The four LF proxies we use are Hasbrouck's (2009) Gibbs estimate (Gibbs), Corwin and Schultz's (2012) high-low spread (HL), Abdi and Rinaldo's (2017) close-high-low spread (CHL), and Fong, Holden, and Tobek's (2017) implementation of Kyle and Obizhaeva (2016) invariance-based volume-over-volatility measure (VoV).

This approach is motivated by the idea that the LF proxies are a forecast (or backcast) of the unobserved high frequency effective spread. The literature on economic forecasting has shown that a simple average of forecasts (a.k.a. combination forecasts) significantly outperforms individual forecasts in a wide variety of settings (Bates and Granger 1969; Timmermann 2006). This improvement can be understood from a simple diversification argument: the predictive power of a particular forecast varies across observations, and combining multiple forecasts averages out these errors. The averaging of multiple LF illiquidity proxies is also used in Karnaukh, Rinaldo, and Soderlind (2015), who find that averaging improves on using the constituent proxies alone. Indeed, we find that our LF average outperforms any individual LF proxy in terms of its ability to match HF data. For further details see Appendix A.3.

[Table 1 about here.]

Table 1 illustrates the performance of our LF average proxy. Panel A begins by showing that our four LF proxies, while highly correlated, still contain distinct information. The typical correlation is around 75%, but can be as low as 0.59 (between HL and VoV). These results suggest that the logic of combination forecasts applies here: by combining proxies we can average out their errors.

Panels B and C shows that this logic works. These panels compare our LF average with HF spreads when they are available. The LF average has the high-

est correlation with TAQ spreads, at 90%. In comparison, the best individual LF proxies are Gibbs and VoV, which both have 84% correlations with TAQ. Panel C shows a similar result with ISSM. The LF average has an even higher 94% correlation with ISSM spreads, compared to 90% for the best individual LF proxy, CHL.

Though LF spreads are highly correlated with HF spreads, they exhibit a strong bias, especially in recent data. This problem is shown in Figure 2, which plots the median difference between LF and HF spreads over time. Post-2003, spreads are biased upward by 25-50 basis points. This bias indicates that it is important to use HF data to examine trading costs in recent years, and that the LF trading costs used by Novy-Marx and Velikov (2016) overestimate expected costs going forward.

[Figure 2 about here.]

Figure 3 illustrates how our combined effective spread measure have evolved over time. Trading costs rise sharply in the early 1970s as NASDAQ stocks enter the CRSP universe. Costs rise further in the late 1980's, a phenomenon which is seen in other papers (Corwin and Schultz 2012; Abdi and Rinaldo 2017). Trading costs plummet in the 2000's as electronic trading and decimalization have improved liquidity. Overall, our combined effective spread is consistent with key features of stock market history.

[Figure 3 about here.]

2.3. Portfolio Implementations

We examine three different implementations for each anomaly: (1) academic implementations, (2) constrained cost optimizations that allow for equal-weighting, and (3) constrained cost optimizations that enforce value-weighting.

Implementation is important because the more general notion of trading costs includes not only the direct costs of trades (e.g. effective spreads), but also the lost returns that come from avoiding the direct costs (Perold 1988). Thus, a full accounting of trading costs requires the study of cost optimization. Moreover, the relevant implementation depends on the investor in question, so we study two versions of our constrained optimized implementation.

2.3.1. Academic Implementations

Our academic implementations are simply equal-weighted long-short quintiles. We sort stocks into quintiles based on the signal, equally-weight stocks, and re-calculate portfolio weights when the signal updates.⁹

This implementation represents the modal approach in the literature. Almost all anomaly papers report either equal-weighted portfolios or equal-weighted regressions, but only a minority report value-weighted portfolios (Green, Hand, and Zhang 2013). Similarly, though the decile and quintile sorts are both frequently reported, many papers that study decile sorts also combine the 9th and 10th deciles in the long leg of their hedge portfolios, suggesting that the original authors would similarly advocate the use of quintile sorts.

2.3.2. Constrained Optimized Implementations

Optimal implementation with many assets and proportional trading costs is an extremely difficult problem. Theoretical solutions have been found only by imposing stark approximations such as uncorrelated returns (Liu 2004) or an exogenous and constant target portfolio (Leland 2000). For tractability, empirical studies often optimize within a restricted set of linear portfolio rules (Brandt, Santa-Clara, and Valkanov 2009; DeMiguel et al. Forthcoming; Moallemi and

⁹For a detailed list of signal updating frequencies see Appendix A.1

Saglam 2017), despite the fact that theory tends to imply non-linear policies.

We optimize within a set of simple non-linear rules that capture the intuition from optimal theory. This set of rules is called the “buy/hold spread” (also known as “banding”), and is best described with an example: a 20/40 buy/hold spread goes long stocks with signals that are in the top 20th percentile, but only exits stocks that have signals below the top 40th percentile (and similarly for the short end). Between the 20th and 40th percentiles is an inaction region where no trading occurs.

Inaction regions are a key feature of optimal trading under trading costs (Magill and Constantinides 1976). Intuitively, while frictionless trading implies that one could always benefit from trading to improve the expected return, with frictions there are states in which the cost of trading outweighs this benefit. Empirical evidence supports this intuition. Novy-Marx and Velikov (2016, 2019) show that the buy/hold spread outperforms other rules commonly used in industry.

Buy/hold spreads also have the advantage that they nest the standard academic implementation: a 20/20 buy/hold rule is equivalent to the standard quintile sort. This feature makes it easy to interpret how our optimization improves on the academic benchmark.

The buy/hold spread rules only prescribe which stocks to long or short—it does not prescribe the weights of each position. For stocks that are prescribed to go long or short, we consider both equal-weighting and value-weighting stocks in our optimization. This choice allows our constrained optimal implementation to tilt toward more larger and more liquid stocks if the lower cost of trading outweighs the gain in expected gross returns. One could consider a more complex weighting function, but we consider only equal- and value- weighting to avoid overfitting.

Optimization proceeds in two steps. In the first step, choose the buy/hold spread exit parameter to maximize the average in-sample net return of anomalies within turnover quartiles. We apply this optimization twice: first assuming equal-weighting and a buy/hold enter threshold of 20% using all stocks, and second assuming value-weighting and a buy/hold enter threshold of 10% using NYSE stocks only. In the second step, choose equal- or value-weighting to maximize in-sample net returns at the anomaly level. For further details see Appendix A.4.

When we examine cost optimizations that enforce value-weighting, we simply enforce value-weighting in the second step of the optimization.

Our optimization is clearly constrained. We take as given a buy/hold decision rule, the enter thresholds of these rules, and only allow for equal or value weighting. Optimizing over additional choices would, by construction, improve performance in-sample, but would lead to more overfitting. Indeed, the fact that our optimization dramatically improves net returns in-sample suggests that cost of more overfitting outweighs benefits. These costs tend to be large in portfolio choice (DeMiguel, Garlappi, and Uppal 2009, for example).

We optimize using only in-sample information for similar reasons. Our main object of interest is the mean net return in samples that are both post-publication and post-2005. Optimizing using only in-sample information ensures that our main object of interest is not affected by data-mining bias coming from optimization.

3. Zeroing in on the Average Anomaly

Having described our methods, we can now zero in on expected returns. We begin with academic implementations because they are widely understood and

thus are helpful for understanding how the anomaly zoo interacts with trading costs. We then present our first main result which examines cost-optimized implementations (Section 3.2).

3.1. The Average Academic Implementation

Table 2 shows that academic implementations offer no expected returns at all. Though the historical gross return (in-sample) was 66 bps per month, one should expect closer to a net return of -3 bps going forward (net of costs and post-publication). Notably, our large set of anomalies produces a standard error on the post-publication net return of just 5 bps.

[Table 2 about here.]

Table 2 offers a few decompositions for understanding this lack of expected returns. The post-publication row shows that roughly half of the in-sample gross returns are eliminated by data-mining bias and changes in the investing environment, consistent with McLean and Pontiff (2016). Though this decay is large, post-publication data still imply a notable 30 bps per month of expected returns (4% per year) before trading costs.

Trading costs wipe out the remaining expected returns, however. A second decomposition shows that this return reduction (column d) is roughly equal to the product of 2-sided turnover (column c) and the average spread paid (column d). As the typical anomaly turns over 15% of its long portfolio and 15% of its short portfolio each month, the total 2-sided turnover is 30%. Multiplying this turnover by the average paid post-publication spread of 111 bps (column d) leads to the return reduction of 32 bps.

The large impact of trading costs may be surprising, since decimalization implies that the quoted spread on many stocks is just one penny. Dividing \$0.01 by

the typical share price of \$20 leads to a tiny spread of 5 bps, far from the 111 bps post-publication spread paid in Table 2.

Trading costs are extremely right-skewed, however, and anomaly strategies require trading stocks from all over the liquidity spectrum. Thus, the typical spread paid by an anomaly strategy is more similar to the mean spread, and much larger than the modal spread one typically sees at a brokerage.

This skewness is seen in Figure 4, which compares distributions of spreads in 2014. NYSE spreads (dotted line) display a mode at around 5 basis points, consistent with the tiny spread implied by decimalization. The NYSE contains many stocks with much larger spreads, however, as seen in the long right tail of the distribution. Indeed, about 20% of NYSE stocks have effective spreads in excess of 20 bps.

[Figure 4 about here.]

Anomaly portfolios load up on this right tail. The distribution of spreads paid by academic implementations in 2014 (solid line) shares the same mode as the NYSE distribution, but the peak is only half as tall, and the missing mass is shifted into the right tail. As a result, the mean spread paid by anomaly strategies in 2014 is 67 bps, more than 4 times the average NYSE spread of 16 bps.

While academic portfolios tend to trade stocks that are more illiquid than the NYSE, their trading costs are similar to that of the broad universe of stocks. Indeed, the anomaly paid spread distribution (solid line) lines up closely with the distribution for all stocks (dash-dotted line), and is significantly shifted to the left compared with the distribution for the Russell 2000 (dashed line).

Returning to Table 2, the “in-sample” row shows that academic implementations are not even profitable in-sample. Compared to post-publication results, turnover is about the same in-sample, but the average spread paid is more than

twice as large, and thus the return reduction doubles to 61 bps per month. This return reduction effectively wipes out the in-sample gross return.

These results suggest academic strategies naively trade stocks that are too illiquid. But simply avoiding illiquid stocks may not be wise, as predictability is much stronger in the more illiquid stocks. Indeed, Novy-Marx and Velikov (2019) find that simply avoiding illiquid stocks also naive, as the reduction in gross returns is as large or larger than the improvement in trading costs.

3.2. The Average Cost-Optimized Anomaly

This section presents our first main result. Here we zero in on the expected returns of cost-optimized implementations.

We begin by showing that our constrained optimization is very effective. Panel B of Table 2 shows that, relative to the academic implementation, optimization improves in-sample net returns by 33 bps per month, leading to a noteworthy 38 bps net return. This improvement comes from a 35% decrease in turnover and a 38% decrease in the spreads paid, while the lost returns are just 7 bps (66 - 59 bps).

Post-publication, however, the mean net return is just 13 bps per month. This negligible return comes from the fact that the gross return drops to just 20 bps post-publication. Thus, even with a miniscule return reduction of 8 bps, the net return is tiny.

Figure 5 provides a more graphic view of this decline in performance. This figure shows the details of our estimates as an event study: we average net returns across 120 anomalies within each month relative to publication (light line). The extreme volatility of the light line is a reminder that anomalies portfolios are not at all sure bets.

[Figure 5 about here.]

The dark line shows the trailing 5-year moving average net return, once again averaging across 120 anomalies. This moving average shows a sharp decline in performance dropping from about 40 bps before publication to around 12 bps afterwards.

Returning to Table 2, the “Post-Pub & Post-2005” row further isolates expected returns by accounting for the change in trading technologies that happened during the early 2000s. This change saw an explosion in trading volume and institutional activity, which implies that the data pre-2005 is unlikely to be representative of the future (Chordia, Subrahmanyam, and Tong 2014). We account for this change by limiting the data to anomaly-months that are both post-publication and post-2005.¹⁰ In this more refined isolation, the typical anomaly is expected to return only 8 bps per month, with a standard error of just 4 bps.

Even this tiny 8 bps per month may be unachievable on larger scales, as panel B allows for equal-weighting for ease of comparison with the broader anomalies literature. Panel C limits our cost-optimized strategies to value-weighting. There we find 4 bps per month of expected returns. Despite the small standard error of 3 bps per month, these expected returns are statistically indistinguishable from zero.

4. Zeroing in the Strongest Anomalies

We’ve seen that the average anomaly’s expected return is effectively zero. But what should we expect from the strongest anomalies? This section presents our second main result: the strongest anomalies’ expected returns are only 10-20 bps per month.

¹⁰Using only post-2003 or post-2004 data leads to very similar results.

To come to this result, we need to account for data-mining bias. To understand this, it helps to examine the heterogeneity in post-publication and post-2005 (“post-pub05”) mean net returns, shown in Figure 6. Some anomalies have notable net returns. Cash flow to price (CF2Price), tangibility (Tangibili), and momentum for young firms (MomYoung) all produce net returns in excess of 80 bps per month in this recent sample.

[Figure 6 about here.]

A portion of these large net returns is due to data-mining bias, however. This bias is clearly seen if we break down the mean post-pub05 net return of predictor i into two components

$$\bar{r}_i = \mu_i + \epsilon_i \tag{1}$$

where \bar{r}_i is the observed mean, μ_i is the true expected return, and ϵ_i is a zero mean noise term due to sampling variability. And suppose we define large net returns as those where \bar{r}_i is larger than the 80th percentile \bar{r}_{80} . Then the conditional expectation for large net returns is

$$\mathbb{E}(\bar{r}_i | \bar{r}_i > \bar{r}_{80}) = \mathbb{E}(\mu_i | \bar{r}_i > \bar{r}_{80}) + \underbrace{\mathbb{E}(\epsilon_i | \bar{r}_i > \bar{r}_{80})}_{>0}. \tag{2}$$

The noise term $\mathbb{E}(\epsilon_i | \bar{r}_i > \bar{r}_{80})$ is positive because mining for large mean returns also selects for large realizations of noise. As a result, the mean returns in the right tail $\mathbb{E}(\bar{r}_i | \bar{r}_i > \bar{r}_{80})$ are upward biased compared to their true returns $\mathbb{E}(\mu_i | \bar{r}_i > \bar{r}_{80})$.

We examine two approaches to removing the bias $\mathbb{E}(\epsilon_i | \bar{r}_i > \bar{r}_{80})$. Section 4.1 uses an out-of-sample test, and Section 4.2 uses an empirical Bayesian adjustment. Though the methods are very different, they lead to very similar results.

4.1. Data-Mining Adjustments Using Out-of-Sample Tests

A simple way to remove the bias in Equation (2) is with an out-of-sample test. Specifically, we sort anomalies based on in-sample predictors, and then measure post-pub05 net returns within quantiles to measure conditional expectations. This exercise ensures that the data used to select anomalies is not the same as that used to evaluate them, thus eliminating data-mining bias.

Formally, suppose we use net returns as the in-sample predictor, and focus on anomalies above the 80th percentile $\bar{r}_{IS,80}$. Then the conditional expectation of the post-pub05 net return is

$$\mathbb{E}(\bar{r}_i | \bar{r}_i > \bar{r}_{IS,80}) = \mathbb{E}(\mu_i | \bar{r}_i > \bar{r}_{IS,80}) + \mathbb{E}(\epsilon_i | \bar{r}_i > \bar{r}_{IS,80}) \quad (3)$$

$$= \mathbb{E}(\mu_i | \bar{r}_i > \bar{r}_{IS,80}), \quad (4)$$

where $\mathbb{E}(\epsilon_i | \bar{r}_i > \bar{r}_{IS,80}) = 0$ because monthly stock returns are nearly i.i.d. and thus sampling error in the mean ϵ_i is uncorrelated across the two, non-overlapping samples. The sample analogue of $\mathbb{E}(\bar{r}_i | \bar{r}_i > \bar{r}_{IS,80})$, then, provides an unbiased estimate of the true expected return $\mathbb{E}(\mu_i | \bar{r}_i > \bar{r}_{IS,80})$.

We consider the following in-sample predictors: the mean net return, net Sharpe ratio, return reduction from trading costs, and turnover. In-sample net returns would predict post-pub05 net returns if μ_i is persistent across samples, and Sharpe ratios would predict for similar reasons. Trading costs should predict because net returns are the difference between gross returns and trading costs, and once again trading costs may be persistent. Turnover, finally, may predict as it is one of the components of trading costs.

Table 3 shows the results. The table shows the mean post-pub05 net return of anomalies grouped by predictor quartiles. Predictability is weak and fragile.

In implementations that allow for equal-weighting (panel A), the best net returns come from using the net Sharpe ratio, with the top quartile producing expected returns of 21.2 bps per month. But the net returns from this sort are not monotonically increasing, and indeed, three out of four predictors fail to produce monotonicity. Moreover, the second strongest predictor (Turnover) produces only 14.3 bps per month in its top quartile.

[Table 3 about here]

Predictability is even essentially gone when using only value-weighting (Panel B). The net Sharpe ratio sort is very fragile, with the second quartile performing much better than the first and third, suggesting that the 11.4 bps in its top quartile cannot be trusted. Indeed, only turnover seems to produce a reliable improvement in mean returns, and it only leads to a statistically insignificant 9.7 bps per month in its top quartile.

Overall, post-pub05 mean net returns show little predictability in out-of-sample tests. Taken together, these results lead us to conclude that the strongest anomalies offer at most 10-20 bps per month, once data-mining bias is accounted for.

4.2. Data-Mining Adjustments Using Empirical Bayes

As an alternative data-mining adjustment, we study an “empirical Bayesian” estimator. This method can be motivated by Equation (2). Bias comes from the noise term $\mathbb{E}(\epsilon_i | \bar{r}_i > \bar{r}_{80})$. Thus, one can remove bias by directly estimating $\mathbb{E}(\mu_i | \bar{r}_i > \bar{r}_{80})$. In other words, what the econometrician really wishes to know is μ_i for the strongest anomalies, and thus our goal is not the conditional sample mean $\mathbb{E}(\bar{r}_i | \bar{r}_i > \bar{r}_{80})$, but the conditional expectation of true returns $\mathbb{E}(\mu_i | \bar{r}_i > \bar{r}_{80})$.

Given an estimated model, Bayes rule provides the logic for computing this expectation. And to generate an estimated model, we specify a DGP and fit it to empirical data using frequentist methods. This combination of empirical frequentist methods and Bayesian logic gives the name “empirical Bayes.” Empirical Bayes has been shown to effectively remove data-mining bias in a variety of settings (Efron 2011; Azevedo et al. 2019; Liu, Moon, and Schorfheide 2020).

We first develop the adjustment and then examine adjusted expected returns. Throughout this section, we refer to mean returns that are post-publication, post-2005, and net of trading costs. For ease of reading, we drop all of the qualifiers in what follows (“Sharpe ratio” refers to the post-publication, post-2005, net Sharpe ratio).

4.2.1. Empirical Bayes Methodology

The Sharpe ratio for predictor i is normally distributed around the true Sharpe ratio

$$\frac{\bar{r}_i}{\sigma_i} \sim N\left(\frac{\mu_i}{\sigma_i}, \text{SE}(\text{SR}_i)\right), \quad (5)$$

where σ_i is the volatility of net returns and $\text{SE}(\text{SR}_i)$ is the standard error for Sharpe ratio i . The normal distribution is justified by the central limit theorem and the fact that the sample sizes are in the order of hundreds. We assume Sharpe ratios are uncorrelated across predictors, consistent with the near-zero median correlation in returns across anomalies (McLean and Pontiff 2016; Green, Hand, and Zhang 2014; Chen and Zimmermann 2019).

Modeling Sharpe ratios rather than mean returns effectively rescales portfolios to have the same volatility. We find that modeling mean returns leads to even smaller expected returns, consistent with the strong performance of net Sharpe

ratios as in Table 3.

We assume σ_i is observed. This assumption can be justified by the small standard error in sample volatility for samples of 360 months.¹¹ Under this assumption and the standard assumption of zero autocorrelation in monthly returns, $SE(SR_i) = SE(\bar{r}_i) / \sigma_i = 1/\sqrt{T_i}$.

True Sharpe ratios are location-scale t-distributed

$$\frac{\mu_i}{\sigma_i} \sim t(\mu_{SR}, \sigma_{SR}, \nu_{SR}) \quad (6)$$

where μ_{SR} is the location (mean), σ_{SR} is the scale (dispersion), ν_{SR} is the degrees of freedom parameter. This bell-shaped distribution is consistent with the data (Figure 6). Using a t-distribution allows for fat tails and thus the idea that there may be a few predictors that are truly exceptional.

Equations (5) and (6) summarize the model. The model has just three parameters: μ_{SR} , σ_{SR} , and ν_{SR} . For simplicity, we fix ν_{SR} at different values to examine how our results change.

Given ν_{SR} , method of moments implies a simple estimate (Xie, Kou, and Brown 2012)¹²

$$\hat{\mu}_{SR} \equiv \frac{1}{N} \sum_{i=1}^N \frac{\bar{r}_i}{\sigma_i} \quad (8)$$

$$\hat{\sigma}_{SR}^2 \equiv \max \left\{ \left(\frac{\nu_{SR} - 2}{\nu_{SR}} \right) \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{\bar{r}_i}{\sigma_i} - \hat{\mu}_{SR} \right)^2 - \frac{1}{N} \sum_{i=1}^N \frac{1}{T_i} \right], 0 \right\}. \quad (9)$$

¹¹If the monthly return is normally distributed, then sample volatility is $\hat{\sigma}_i = \frac{\sigma_i}{\sqrt{T-1}} \chi_{T-1}$. Then the standard error of $\hat{\sigma}_i = 0.037s$ for a sample size of 30 years.

¹²To see this, note

$$\mathbb{E}[(\bar{r}_i/\sigma_i - \mu_{SR})^2] = \mathbb{E}[(\mu_i/\sigma_i - \mu_{SR})^2 + (\mu_i/\sigma_i - \mu_{SR})\delta_i + \delta_i^2], \quad (7)$$

where δ_i is a noise term. The cross term drops out, and then population moments are replaced by sample moments to arrive at (7). Restricting the parameter set to positive σ_{SR}^2 results in the max operation.

Intuitively, the grand mean is estimated using the average of all Sharpe ratios, and the scale parameter is estimated as the dispersion in Sharpe ratios $\frac{1}{N} \sum_{i=1}^N \left(\frac{\bar{r}_i}{\sigma_i} - \hat{\mu}_{\text{SR}} \right)^2$ that cannot be accounted for by noise $\frac{1}{N} \sum_{i=1}^N \frac{1}{T_i}$. Finally, the factor $\left(\frac{\nu_{\text{SR}} - 2}{\nu_{\text{SR}}} \right)$ adjusts for the assumed fat tail parameter ν_{SR} .

Given estimated parameters, we calculate the bias-adjusted expected return for predictor i with

$$\hat{\mu}_i \equiv \mathbb{E} \left[\frac{\hat{\mu}_i}{\sigma_i} \middle| \bar{r}_i, \sigma_i, \hat{\mu}_{\text{SR}}, \hat{\sigma}_{\text{SR}}, \nu_{\text{SR}} \right] \sigma_i. \quad (10)$$

That is, the bias adjusted return is the conditional expectation of the true Sharpe ratio given all available information, rescaled by volatility. We rescale by volatility for ease of comparison with our other results.

Equation (10) is free of data-mining bias, even for predictors with large \bar{r}_i . This feature comes from the fact Equation (10) already conditions on all available information. This property is sometimes considered a paradox (Dawid 1994), but Senn (2008) demonstrates that it is entirely logical. Indeed, the removal of data-mining bias using estimations analogous to Equation (10) has been demonstrated in numerous settings (Efron 2011; Azevedo et al. 2019; Liu, Moon, and Schorfheide 2020; Chen and Zimmermann 2019).

The mechanics of the adjustment can be seen in the special case $\nu_{\text{SR}} \rightarrow \infty$. In this case, normal-normal updating formulas imply

$$\hat{\mu}_i = \hat{s}_i \hat{\mu}_{\text{SR}} \sigma_i + (1 - \hat{s}_i) \bar{r}_i \quad (11)$$

where the “shrinkage” \hat{s}_i is given by

$$\hat{s}_i \equiv \frac{1/T_i}{\hat{\sigma}_{\text{SR}}^2 + 1/T_i}. \quad (12)$$

Intuitively, we shrink large \bar{r}_i toward the grand mean $\hat{\mu}_{\text{SR}}\sigma_i$. Predictors with smaller samples are shrunk more, as they are more vulnerable to data-mining bias. The overall shrinkage is determined by $\hat{\sigma}_{\text{SR}}$, where in the extreme case that there is no dispersion in true Sharpe ratios, shrinkage is 100%. Equation (11) shows our estimator is closely related to the celebrated James and Stein (1961) estimator. Thus, similar estimators can also be derived from quadratic loss arguments, as well as Galtonian reverse regression (Stigler 1990).

4.2.2. Empirical Bayes Results

Table 4 describes the estimation results and bias adjusted returns. Panel A shows our baseline cost optimizations, which allow for equal-weighting. There we find that assuming that true Sharpe ratios are approximately normal, ($\nu_{\text{SR}} = 100$), the standard deviation of true Sharpe ratios is 0.20 (annualized). Considering that the mean standard error on the observed net Sharpe ratio is 0.35, this implies that the adjustment is very large (Equation (11)). Indeed, 80th and 90th percentile adjusted net post-pub05 returns are only about 20 bps per month. Assuming that true Sharpe ratios are fat tailed ($\nu_{\text{SR}} = 4$) has almost no effect on the results. These results are quantitatively very similar to those from our predictability-based adjustment (Table 3).

[Table 4 about here.]

Bias adjustments for implementations that only use value-weighting (Panel B) are even stronger. Indeed, our estimates imply that there is no dispersion of true Sharpe ratios at all. This result comes from the fact that the dispersion in observed Sharpe ratios is smaller than the average standard error, and thus method of moments hits the positivity constraint on σ_{SR} . In other words, all value-weighted anomalies have the same true Sharpe ratios, and the strongest

expected returns come only from taking on more volatility. As a result, the 90th percentile of adjusted net post-pub returns is just 6.4 bps per month. This result is also consistent with our predictability results, where we saw that no in-sample information is a reliable predictor of post-pub05 net returns.

The intuition for these results can be seen in Figure 6. Only 11 out of 120 anomalies produce t-stats > 2.0 in absolute value, not far from the 6 implied by a model in which there is no predictability ($\sigma_{SR} = \mu_{SR} = 0$). As a result, noise can account for most of the heterogeneity in post-pub05 performance, Bayesian logic implies that bias adjustments are large, leading to our finding that even the strongest anomalies offer only 10-20 bps of expected returns.

4.3. Performance of Size, B/M, and Momentum

Objective statistics show that the strongest anomalies provide little expected returns. But these results group famous anomalies like size, B/M, and momentum with lesser known ones from the broader anomaly zoo. This section examines the performance of size, B/M, and momentum and compares our results with the literature, which tends to focus on these well-known anomalies.

We find that size, B/M, and momentum have unremarkable performance, in-line with the broader anomaly zoo in post-pub05 samples. This can be seen in Figure 6, in which B/M is represented by “BM,” and momentum is represented by “Mom12m.” These famous anomalies lie in the middle of the distribution, centered around zero.

Table 5 takes a closer look at these anomalies. Our baseline results emphasize the post-pub05 sample, in which size, B/M, and momentum net -26 bps, 33 bps, and 16 bps, respectively. This sample corresponds to 2006-2016, as size, B/M, and momentum are all published before our post-2005 period begins.

This poor performance appears to conflict with those of other papers, which often find that these select anomalies perform well net of costs. In particular, Frazzini, Israel, and Moskowitz (2015) (FIM) conclude that “size, value, and momentum — are robust, implementable, and sizeable in the face of transactions costs.”

[Table 5 about here.]

FIM’s conclusions, however, come from examining long historical samples going back to 1926, and mean net returns are highly sensitive to the sample period. Indeed, FIM’s results for the 1998-2013 sample show more of a mixed result. We reprint these results in the “FIM (2015)” column of Table 5. There we see that in the more recent data, size and B/M have notable net returns, but momentum has a slightly negative net return.

This sensitivity is also seen in using our methodology. While size, B/M, and momentum are unremarkable post 2006, they seem to have above-average performance 1998-2013, as seen in the 3rd column of Table 5. Indeed, this earlier-sample performance is as good or better than those reported by FIM for the same sample period.

Consistent with these fragile results, Table 5 shows that individual anomalies produce huge standard errors of 30-60 bps per month. This sampling noise makes it impossible to tell if the any individual anomaly has strong performance in the modern era trading technology. Indeed, our post-2005 net returns are not statistically different than any of the other results shown in the table.

Overall, Table 5 highlights the importance of studying a large set of anomalies for making inference about expected returns. The performance of individual anomalies is very noisy in the post-2005 period. It is only by aggregating information over many anomalies that we can make precise measurements of what

we should expect after excluding stale data.

5. Conclusion

We zero in on the expected returns of anomalies by accounting for trading costs and the staleness of historical data. Net of these effects, the expected return on even the best anomalies is effectively zero.

This conclusion comes from applying data-mining adjustments to data that includes high-frequency trading costs and a large set of anomalies. High-frequency data is necessary as low-frequency spreads are biased upward in recent years. A large set of anomalies is required as individual anomaly returns are very noisy after excluding stale data. Finally, data-mining adjustments are required to control for the bias that comes from selecting the best anomalies. Our study is unique in combining these datasets and methods.

In combination with recent findings, our results provide a complete accounting for the average return on the anomaly zoo. Previous papers show that the gross return is about 15% publication bias (McLean and Pontiff 2016; Chen and Zimmermann 2019). We find that trading costs account for another 40%, and that the remaining net returns (45%) are traded away over time, consistent with the idea that mispricing is removed as information proliferates and technology improves (Chordia, Subrahmanyam, and Tong 2014; McLean and Pontiff 2016).

This decomposition paints a picture of a dynamic equilibrium process, but one more in line with Lo's (2004) adaptive market hypothesis or "efficiently inefficient" markets (Grossman and Stiglitz 1980; Gârleanu and Pedersen 2018) than standard dynamic equilibrium models (Campbell and Cochrane 1999). Every month, researchers find imperfections in the existing market equilibrium. As information about predictability diffuses and trading technology improves, the

net returns of these imperfections are traded away, leading to a new equilibrium.

A. Appendix

A.1. Description of the Anomaly Dataset

Table A.1: List of Cross-Sectional Return Predictors Part 1/3

This table lists the anomalies in our dataset. For further details, please see the Appendix of Chen and Zimmermann (2019). Freq lists the rebalancing frequencies we assume.

| Acronym | Description | Freq | Publication |
|--------------|---------------------------------------|------|--------------------------------|
| AccrAbn | Abnormal Accruals | A | Xie 2001 AR |
| AccrOper | Percent Operating Accruals | A | Hafzalla et al 2011 AR |
| AccrPct | Percent Total Accruals | A | Hafzalla et al 2011 AR |
| Accruals | Accruals | A | Sloan 1996 AR |
| AdExpGr | Growth in advertising expenses | A | Lou 2014 RFS |
| AnnounRet | Earnings announcement return | Q | Chan et al 1996 JF |
| AssetCGr | Change in current operating assets | A | Richardson et al 2005 JAE |
| InvestAG | Asset Growth | A | Cooper et al 2008 JF |
| ATurn | Asset Turnover | A | Soliman 2008 AR |
| BEgrowth | Sustainable Growth | A | Lockwood Prombutr 2010 JFR |
| BetaSquared | CAPM beta squared | M | Fama MacBeth 1973 JPE |
| BidAskSpread | Bid-ask spread | M | Amihud Mendelsohn 1986 JFE |
| BM | Book to market | A | Fama French 1992 JF |
| BMent | Enterprise component of BM | A | Penman et al 2007 JAR |
| BMlev | Leverage component of BM | A | Penman et al 2007 JAR |
| CAPXgr | Change in capex (two years) | A | Anderson Garcia-Feijoo 2006 JF |
| Cash | Cash to assets | Q | Palazzo 2012 JFE |
| CF2Price | Cash flow to market | A | Lakonishok et al 1994 JF |
| CFOper2Price | Operating Cash flows to price | A | Desai et al 2004 AR |
| DebtFinC | Composite debt issuance | A | Lyandres Sun Zhang 2008 RFS |
| DeferRev | Deferred Revenue | A | Prakash Sinha 2012 CAR |
| DepGr | Change in depreciation to gross PPE | A | Holthausen Larcker 1992 JAE |
| EarnCons | Earnings Consistency | Q | Alwathainani 2009 BAR |
| EarnSupBig | Earnings surprise of big firms | M | Hou 2007 RFS |
| EarnSurp | Earnings Surprise | Q | Foster et al 1984 AR |
| EffFrontier | Efficient frontier index | A | Nguyen Swanson 2009 JFQA |
| EntMult | Enterprise Multiple | A | Loughran Wellman 2011 JFQA |
| EP | Earnings-to-Price Ratio | A | Basu 1977 JF |
| EPforecast | Earnings Forecast | M | Elgers Lo Pfeiffer 2001 AR |
| EPSDisp | EPS Forecast Dispersion | M | Diether et al 2002 JF |
| EPSForeLT | Long-term EPS forecast | M | La Porta 1996 JF |
| EPSrevise | Earnings forecast revisions | M | Chan et al 1996 JF |
| Eq2AGr | Change in equity to assets | A | Richardson et al 2005 JAE |
| ExcludExp | Excluded Expenses | M | Doyle et al 2003 RAS |
| ExtFinNet | Net external financing | A | Bradshaw et al 2006 JAE |
| FailurePr | Failure probability | Q | Campbell et al 2008 JF |
| FinLiabGr | Change in financial liabilities | A | Richardson et al 2005 JAE |
| GIndex | Governance Index | A | Gompers et al 2003 QJE |
| GM2SaleGr | Gross Margin growth over sales growth | A | Abarbanell Bushee 1998 AR |
| Herf | Industry concentration (Herfindahl) | A | Hou Robinson 2006 JF |
| High52 | 52 week high | M | George Hwang 2004 JF |
| IdioVol | Idiosyncratic risk | M | Ang et al 2006 JF |
| Illiquid | Amihud's illiquidity | M | Amihud 2002 JFM |
| IndMom | Industry Momentum | M | Grinblatt Moskowitz 1999 JFE |
| IndRetBig | Industry return of big firms | M | Hou 2007 RFS |

Table A.2: List of Cross-Sectional Return Predictors Part 2/3

| Acronym | Description | Freq | Publication |
|----------------|--|------|--------------------------------|
| InstOwnSI | Inst own among high short interest | Q | Asquith Pathak Ritter 2005 JFE |
| IntanBM | Intangible return using BM | A | Daniel Titman 2006 JF |
| IntanCFP | Intangible return using CFtoP | A | Daniel Titman 2006 JF |
| IntanEP | Intangible return using EP | A | Daniel Titman 2006 JF |
| IntanSP | Intangible return using Sale2P | A | Daniel Titman 2006 JF |
| InvestGr | Change in capital inv (ind adj) | A | Abarbanell Bushee 1998 AR |
| Invntory | Inventory Growth | A | Thomas Zhang 2002 RAS |
| InvToRev | Investment to revenue | A | Titman et al 2004 JFQA |
| KZ | Kaplan Zingales index | A | Lamont et al 2001 RFS |
| LaborGr | Employment growth | A | Bazdresch Belo Lin 2014 JPE |
| Leverage | Market leverage | A | Bhandari 1988 JFE |
| LiabCGr | Change in current operating liabilities | A | Richardson et al 2005 JAE |
| LTAssetGr | Change in Noncurrent Operating Assets | A | Soliman 2008 AR |
| LTNOAgr | Growth in Long term net operating assets | A | Fairfield et al 2003 AR |
| MaxRet | Maximum return over month | M | Bali et al 2010 JF |
| Mom12m | Momentum (12 month) | M | Jegadeesh Titman 1993 JF |
| Mom12to7 | Intermediate Momentum | M | Novy-Marx 2012 JFE |
| Mom1813 | Momentum-Reversal | M | De Bondt Thaler 1985 JF |
| Mom1m | Short term reversal | M | Jegadeesh 1989 JF |
| Mom36m | Long-run reversal | A | De Bondt Thaler 1985 JF |
| Mom6Jnk | Junk Stock Momentum | M | Avramov et al 2007 JF |
| Mom6m | Momentum (6 month) | M | Jegadeesh Titman 1993 JF |
| MomVol | Momentum and Volume | M | Lee Swaminathan 2000 JF |
| MomYoung | Firm Age - Momentum | M | Zhang 2004 JF |
| NDebtFin | Net debt financing | A | Bradshaw et al 2006 JAE |
| NDebtPrice | Net debt to price | A | Penman et al 2007 JAR |
| NEqFin | Net equity financing | A | Bradshaw et al 2006 JAE |
| NOA | Net Operating Assets | A | Hirshleifer et al 2004 JAE |
| NPayYield | Net Payout Yield | A | Boudoukh et al 2007 JF |
| NWCgr | Change in Net Working Capital | A | Soliman 2008 AR |
| OperLeverage | Operating Leverage | A | Novy-Marx 2010 ROF |
| OptVol | Option Volume to Stock Volume | M | Johnson So 2012 JFE |
| OptVolGr | Option Volume relative to recent average | M | Johnson So 2012 JFE |
| OrderBacklog | Order backlog | A | Rajgopal et al 2003 RAS |
| OrgCap | Organizational Capital | A | Eisfeldt Papanikolaou 2013 JF |
| OScore | O Score | A | Dichev 1998 JFE |
| PayYield | Payout Yield | A | Boudoukh et al 2007 JF |
| PensionFunding | Pension Funding Status | A | Franzoni Marin 2006 JF |
| PMGrowth | Change in Profit Margin | A | Soliman 2008 AR |
| Price | Price | M | Blume Husic 1972 JF |
| PriceDelay | Price delay | M | Hou Moskowitz 2005 RFS |
| ProfCash | Cash-based operating profitability | A | Ball et al 2016 JFE |
| ProfGross | gross profits / total assets | A | Novy-Marx 2013 JFE |
| ProfitMargin | Profit Margin | A | Soliman 2008 AR |
| ProfOper | operating profits / book equity | A | Fama French 2006 JFE |

Table A.3: List of Cross-Sectional Return Predictors Part 3/3

| Acronym | Description | Freq | Publication | |
|-----------------|--------------------------------------|------|----------------------|-----------|
| RDirtSurp | Real dirty surplus | A | Landsman et al | 2011 AR |
| RealEstate | Real estate holdings | A | Tuzel | 2010 RFS |
| RetConglomerate | Conglomerate return | M | Cohen Lou | 2012 JFE |
| Rev2Price | Sales-to-price | A | Barbee et al | 1996 FAJ |
| RevG2InvG | Sales growth over inventory growth | A | Abarbanell Bushee | 1998 AR |
| RevG2OHG | Sales growth over overhead growth | A | Abarbanell Bushee | 1998 AR |
| RevGrowth | Revenue Growth Rank | A | Lakonishok et al | 1994 JF |
| RevSurprise | Revenue Surprise | Q | Jegadeesh Livnat | 2006 JFE |
| RoA | earnings / assets | Q | Balakrishnan et al | 2010 JAE |
| RoE | net income / book equity | A | Haugen Baker | 1996 JFE |
| Seasonality | Return Seasonality | M | Heston Sadka | 2008 JFE |
| ShareIs1 | Share issuance (5 year) | A | Daniel Titman | 2006 JF |
| ShareIs5 | Share issuance (1 year) | A | Pontiff Woodgate | 2008 JF |
| VolumeShare | Share Volume | Q | Datar Naik Radcliffe | 1998 JFM |
| ShortInterest | Short Interest | Q | Dechow et al | 2001 JFE |
| Size | Size | A | Banz | 1981 JFE |
| OSmirkNTM | Volatility smirk near the money | M | Xing Zhang Zhao | 2010 JFQA |
| OSmirkCP | Put volatility minus call volatility | M | Yan | 2011 JFE |
| Tangibility | Tangibility | A | Hahn Lee | 2009 JF |
| Tax2E | Taxable income to income | A | Lev Nissim | 2004 AR |
| TaxGr | Change in Taxes | Q | Thomas Zhang | 2011 JAR |
| ATurnGr | Change in Asset Turnover | A | Soliman | 2008 AR |
| TurnovVol | Share turnover volatility | M | Chordia et al | 2001 JFE |
| CF2Pvar | Cash-flow to price variance | A | Haugen Baker | 1996 JFE |
| Volume2Mkt | Volume to market equity | M | Haugen Baker | 1996 JFE |
| VolumeDol | Past trading volume | M | Brennan et al | 1998 JFE |
| VolumeSD | Volume Variance | M | Chordia et al | 2001 JFE |
| VolumeTrend | Volume Trend | M | Haugen Baker | 1996 JFE |
| ZeroTrade | Days with zero trades | M | Liu | 2006 JFE |
| ZScore | Altman Z-Score | A | Dichev | 1998 JFE |

A.2. Details of High Frequency Data

The HF effective spread for the k th trade of a given stock is

$$[\text{Effective Spread}]_k = 2|\log(P_k) - \log(M_k)|, \quad (13)$$

where P_k is the price of the k th trade and M_k is the midpoint of the matched consolidated best bid and offer (BBO) quote.

We use Daily TAQ (DTAQ) data with its milli-nanosecond time-stamps whenever it is available (October 2003 to December 2016). Holden and Jacobsen (2014) find that DTAQ leads to a more accurate and precise measurement of effective spreads in the modern market environment relative to the Monthly TAQ (MTAQ) data with its second-level time stamps.

DTAQ spreads use Holden and Jacobsen's (2014) (HJ's) HJ's DTAQ code. ISSM and MTAQ spreads use HJ's monthly code. For pre-1999 data, we add a 2 second delay to the HJ interpolation-matching algorithm. For data in 1999-2002 we use the 1 millisecond delay following HJ's MTAQ code.

In addition to the data screens used by HJ, we also discard any spreads $> 40\%$ at the trade level (before averaging), following Abdi and Ranaldo (2017). We also adapt the mode screens to ISSM data following Lou and Shu (2014).

The details of the data cleaning are described below.

A.2.1. ISSM Data Details

We adapt HJ's MTAQ code to calculate ISSM spreads.

One of HJ's screens deletes quotes in which the offer or bid size are ≤ 0 or missing. These depth fields are missing or appear to have errors in some subsamples of the data, and we choose not to apply this screen on these subsamples.

NASDAQ stocks in ISSM from 1987-1989 are all missing depth data. Roughly half of the stocks in MTAQ from January 1, 1993 to April 5, 1993 (inclusive) are have zero for all observations of depth, while close to 0% of stocks are have zeros beginning April 6. HJ use the depth screen in order to avoid withdrawn quotes. We choose to not use the depth screen on these subsamples, as the noise in LF spreads is likely to be much larger than the errors introduced by withdrawn quotes.

Quotes are excluded if any of the following hold:

- Time is before 9:00 am or after 4:00 pm
- if mode in (C, D, E, G, I, L, N, P, S, V, X, Z)
- $BID > OFR$ and $BID > 0$ and $OFR > 0$
- $BID > 0$ and $OFR = 0$
- $OFR - BID > 5$ and $BID > 0$ and $OFR > 0$
- $OFR \leq 0$ or missing
- $BID \leq 0$ or missing
- $ofrsize \leq 0$ or missing
- $bidsize \leq 0$ or missing.

NASDAQ listed stocks from 1987-1989 and NYSE listed stocks in 1986 are not subject to the size filters as they are all missing $ofrsize$ and $bidsize$.

Trades are kept if all of the following hold

- Time is after 9:30 am and before 4:00 pm
- $Price > 0$
- $Type = T$
- Cond not in (C, L, N, R, O, Z) and $Size > 0$
- From TAQ and correction field is zero

We add a 2-second interpolated delay using Holden and Jacobsen's (2014) interpolation code.

A.2.2. MTAQ Data Details

We follow HJ's MTAQ code to calculate MTAQ spreads. MTAQ data spans Jan 1, 1993 to Dec 31, 2014 with trades and quotes timestamped to the second.

Quotes are excluded if any of the following hold:

- Time is before 9:00 am or after 4:00 pm
- if mode in (4,7,9,11,13,14,15,19,20,27,28)
- $BID > OFR$ and $BID > 0$ and $OFR > 0$
- $BID > 0$ and $OFR = 0$
- $OFR - BID > 5$ and $BID > 0$ and $OFR > 0$
- $OFR \leq 0$ or missing
- $BID \leq 0$ or missing
- $ofrsiz \leq 0$ or missing
- $bidsiz \leq 0$ or missing.

Data from January 1, 1993 to April 5, 1993 are not subject to the size filters because about 50% of stocks have zero for all observations of ofrsize and bidsiz during this period. In contrast, close to 0% have zeros beginning April 6, 1993, suggesting there are errors for bid and offer sizes at the beginning of the MTAQ data.

Trades are kept if all of the following hold

- Time is after 9:30 am and before 4:00 pm
- Price > 0
- Type = T
- Corr = 0

Following Holden and Jacobsen (2014), we delay quotes as follows:

- Add 2 second interpolated delay pre-1999
- Add 1 millisecond interpolated delay based on HJ for 1999-2002

A.2.3. DTAQ data details

We exactly follow HJ's DTAQ code to calculate DTAQ spreads. DTAQ spans Sep 10, 2003 to the present with trades, quotes, and NBBOs originally timestamped to the millisecond. On Aug 25, 2015 the Daily TAQ timestamps were switched to the microsecond and on Oct 24, 2016 the Daily TAQ timestamps were switched to the nanosecond. Our DTAQ code uses nanosecond timestamps throughout even though some of the trailing digits will be zeros during the millisecond and microsecond eras.

Observations in the DATQ NBBO and quote file are excluded if any of the following hold:

- Qu_Cond not in (A, B, H, O, R, W)
- Ask ≤ 0 or missing
- Ask size ≤ 0 or missing
- Bid ≤ 0 or missing
- Bid size ≤ 0 or missing

Observations in the DTAQ NBBO are also excluded if Qu_Cancel = B. Observations in the quote file are also excluded if Bid > Ask or Bid - Ask > 5.

We also keep only quotes that meet the following additional restrictions:

- (Qu_Source = C and NatBBO_Ind=1) or (Qu_Source = N and NatBBO_Ind=4)
- sym_suffix = "
- Time is between 9:00 am and 4:00 pm

Trades are kept if the all of the following hold:

- Tr_Corr = 00
- price > 0
- sym_suffix = "
- Time is between 9:30 am and 4:00 pm

Following Holden and Jacobsen (2014), we delay quotes as follows:

- Add 1 nanosecond (one-billionth of a second) delay post Oct 24, 2016
- Add 1 microsecond (one-millionth of a second) delay post Jul 24, 2015
- Add 1 millisecond (one-thousand of a second) delay post Sep 9, 2003

Explicitly, the Holden and Jacobsen (2014) DTAQ code adds a nanosecond delay, but due to the data variable data availability in DTAQ the delays are as listed above.

A.3. Details of Low Frequency Spreads

Three of our four proxies build off of Roll's (1984) classic microstructure model. The Roll model assumes that the true value of a stock follows a random walk, and that the observed trade prices deviate from the true value by the effective spread. The fourth proxy uses a completely different framework: the Kyle and Obizhaeva (2016) microstructure invariance hypothesis. All 4 proxies have been shown to be highly correlated with HF spreads.

The LF proxies we use are as follows:

1. Hasbrouck's (2009) Gibbs sampler estimate of the Roll model (Gibbs)

Hasbrouck (2009) estimates the Roll model using Bayesian methods (Gibbs sampler) and daily closing prices. Identification comes from the "bid-ask bounce"—the phenomenon in which buyer initiated trades tend to occur at higher prices than seller initiated trades. Bid-ask bounce induces a negative serial correlation in transaction prices, that is stronger for stocks that are more expensive to trade. The Bayesian approach ensures that the measured serial correlation is negative, and thus the estimated spread is well

defined. Our Gibbs proxy is estimated using annual samples, following the approach recommended in Hasbrouck (2009).

Gibbs forms the basis for transaction costs in several other studies of portfolio returns, including Brandt, Santa-Clara, and Valkanov (2009); Hand and Green (2011); Novy-Marx and Velikov (2016); and DeMiguel et al. (Forthcoming).

2. **Corwin and Schultz's (2012) High-Low Spread (HL).**

Corwin and Schultz (2012) estimate the Roll model from daily high and low prices (hence, HL) that are available in CRSP. Identification comes from the fact that the daily high-low ratio reflects both spreads and return volatility, but these two components decay at different rates. Thus, the comparison of 1-day and 2-day price ranges provides information about the effective spread.

HL is used in many studies including Karnaukh, Ranaldo, and Soderlind (2015); McLean and Pontiff (2016); Koch, Ruenzi, and Starks (2016); and Chen and Zimmermann (2019).

3. **Abdi and Ranaldo's (2017) Close-High-Low (CHL)**

Abdi and Ranaldo's (2017) CHL proxy estimates the Roll model using daily closing prices as well as the daily high and low (hence, CHL). Abdi and Ranaldo's identification builds off the insight that the average of the daily high and low prices (the midpoint) contains important information about the true price. Abdi and Ranaldo (2017) show that CHL outperforms both Gibbs and HL using a number of empirical tests.

4. **Volume-over-Volatility (VoV), based on Kyle and Obizhaeva's (2016) microstructure invariance hypothesis.**

Our last LF proxy takes a rather different approach. Rather than build off of Roll (1984), VoV is based on the Kyle and Obizhaeva’s (2016) microstructure invariance hypothesis. In particular, we use Fong, Holden, and Tobek’s (2017) (FHT’s) implementation:

$$[\text{VoV}]_{i,t} = \frac{8.0 [\text{Std Dev of Daily Returns}]^{\frac{2}{3}}}{[\text{Mean Real Daily Dollar Volume}]^{\frac{1}{3}}} \quad (14)$$

where $[\text{VoV}]_{i,t}$ is the proxy for effective spread for stock i in month t , the $\frac{2}{3}$ and $\frac{1}{3}$ exponents are predictions of Kyle and Obizhaeva’s (2016) invariance hypothesis, and the 8.0 coefficient was chosen by FHT to fit the average monthly TAQ effective spread in their U.S. sample. Nominal dollar volume is converted to real dollar volume using the CPI.

The invariance hypothesis is that the distribution of transaction costs is the same across assets and time periods when expressed in terms of “business time,” that is, the speed with which “bets” arrive at the market. This hypothesis leads to the prediction that the constant term in trading costs (alternatively, the bid-ask spread) is proportional to the RHS of Equation (14). Fong, Holden, and Tobek (2017) find that VoV is the best performing LF proxy among many proxies in terms of correlations and RMSE with respect to TAQ spreads.

HL and CHL both use daily high and low prices. For days in which stocks do not trade, we use the most recent observation of high and low prices. As noted in Abdi and Ranaldo (2017) and Corwin and Schultz (2012), on days in which stocks do not trade CRSP provides closing quoted spreads, and closing quoted spreads are very highly correlated with effective HF spreads in the recent sample. In these cases, we do not use the closing quoted spread in order to make interpretation of our LF proxy average simple.

The LF proxies require multiple firm-day observations to compute a spread for a given firm-month. We follow the original papers and do not compute the proxy if the data is insufficient. Specifically, HL requires 12 daily observations, CHL requires 12 eligible days following the definition in Abdi and Ranaldo (2017), VoV requires 5 positive volume and 11 non-zero return observations, and Gibbs requires the sampler to converge.

We compute a LF average if we have at least one LF proxy with data. In 12.24% of observations, all LF and HF spreads are missing data. These missing observations have little effect on our main results, however, as only 0.27% of post-1993 observations are missing, and 90% of our anomalies are published after 1993. If ISSM, TAQ, and the LF spreads are all missing, we match the firm to the nearest firm with available data in terms of Euclidean distance of market equity rank and idiosyncratic volatility rank. If idiosyncratic volatility is missing, we use just the market equity rank. This data filling procedure follows Novy-Marx and Velikov (2016).

A.4. Details of Cost Optimization

Table A.4 illustrates the first step of our optimization. Panel A shows the net returns of equal-weighted quintile strategies within turnover quantiles, after implementing a variety of buy-hold spreads. The panel shows that buy/hold spreads improve the net returns of high turnover anomalies, but do not help much among anomalies with low turnover. Anomalies in the 3rd turnover quartile perform best on average using a 20/35 buy-hold spread—that is, long positions should only be exited when they drop below the top 35th percentile of the anomaly signal. 4th turnover quartile anomalies benefit significantly from a 20/50 buy/hold spread, but they do not produce positive net returns on average.

Panel B shows that buy/hold spreads are reliably effective for value-weighted NYSE decile strategies. As with equal-weighted quintiles, buy/hold spreads do not significantly improve the net returns of anomalies with below-median turnover. Buy/hold spreads produce significantly positive net returns for 3rd turnover quartile anomalies and even the 4th turnover quartile anomalies, however. These results are consistent with Novy-Marx and Velikov (2016), who also find that the trading costs for low turnover anomalies are too small to justify implementing a buy/hold spread.

Bold numbers indicate the best-performing buy/hold spreads for each stock weighting and turnover quartile combination. In the last step of our cost mitigation, we choose the stock weighting and breakpoint choice that maximizes the net return in-sample, given the bold buy/hold spreads in Table A.4. This last step of the optimization is done at the anomaly level, and is not shown in the table.

Figures A.2 and A.3 show that our cost mitigation is effective in-sample. The figures show the distribution of in-sample net returns before (Figure A.2) and after (A.3) cost mitigation. Rather than use bars to indicate the histogram counts, we list acronyms, with each acronym identifying a different anomaly. Full references for each acronym are found in Appendix A.1.

Figure A.2 shows that net returns before cost mitigation feature a long left tail. While most anomalies have positive net returns ranging between 0 and 60 bps per month, many anomalies have very negative net returns of -50 to -300 bps. Averaging across all anomalies leads to the tiny net return of 6 bps per month in Table 2.

Anomalies with above-median turnover are shown in bold. These high turnover anomalies occupy the vast majority of the left tail of net returns. These high turnover anomalies include many momentum anomalies like 12-month

momentum (Mom12m) and momentum among junk-rated firms (Mom6Jnk), but also includes a variety of unrelated anomalies like idiosyncratic volatility (IdioVol), earnings forecast dispersion (EPSDisp), and detrended trading volume (VolumeTre). Persistent anomaly signals like B/M (BM) and size (Size) are little affected by bid-ask spreads and occupy the right tail of this distribution.

Cost-mitigation should be very helpful with this left tail of net returns. As seen in Table A.4, value-weighting combined with a buy/hold spread produces positive net returns even among anomalies in the highest turnover quartile.

Indeed, Figure A.3 shows that our cost-mitigation is quite effective in-sample. The long left tail of net returns from Figure A.2 is gone. As a result, the average anomaly net return increases to a notable 38 bps per month.

Cost mitigation techniques used on each anomaly are also shown in Figure A.3. Anomalies that use value-weighting are shown in italics. Strategies that use buy/hold spreads larger than 5 percentage points are underlined. We do not underline equal-weighted 20/25 buy/hold spreads as the improvement in net returns is very small (Table A.4).

60% of anomalies perform best using value-weighting once trading costs are accounted for. A large fraction of these anomalies work best with a combination of value-weighting and a buy/hold spread. Indeed, most of the anomalies with negative net returns before optimization (**bold**) become profitable once both of these techniques are applied.

The anomalies that are rescued by cost-mitigation include the momentum anomalies (Mom6m, Mom12m, Mom6Jnk, etc). Indeed, momentum anomalies move from among the worst performers using the academic strategies to among the best performers once value-weighting and buy/hold spreads are applied. Other anomalies that have significantly improved by cost mitigation include idiosyncratic volatility (IdioVol), the distress anomaly (FailurePr), and the

forecasted earnings-price ratio (EPforecas).

Still, there are a few anomalies that cost mitigation cannot resuscitate. Many of these are related to information diffusion, such as price delay (PriceDela) or the earnings surprise of matched large firms (EarnSupBig). Intuitively, profiting on slow information diffusion may require trading neglected and illiquid stocks, as well as frequent trading.

The net returns in Figure A.3 are largely not available to the public, however. Many readers may not be able to trade on the anomalies until after they are published. Even the academics who developed the original strategies in Figure A.3 likely cannot earn the in-sample profits, as the strategies were developed toward the end of the in-sample period.

Table A.4: Optimizing Buy-Hold Spreads: Mean Net Returns In-Sample by Turnover Quartile

The table shows mean net returns in-sample for various buy/hold spread trading rules within turnover quartiles. Bold numbers indicate the best-performing buy/hold spread for each turnover quartile. Turnover quartiles are calculated using the EW quintile benchmark (panel A) and the VW NYSE deciles (panel B). For buy/hold spreads in panel A, we enter a long position for stocks that enter the top 20th percentile of the anomaly signal, but only exit the long position when the stock drops below the percentile indicated by the buy/hold lower bound in the table. Similarly, we enter short positions when stocks enter the bottom 20th percentile, but only exit when stocks rise above the indicated buy/hold lower bound. Panel B enters long positions when stocks enter the 10th NYSE percentile and exits when the stock drops below the NYSE percentile indicated by the buy/hold lower bound.

| | | Panel A: EW Quintiles | | | | | | |
|----------------------|----|-----------------------|-------------|-------|-------------|-------|-------|--------------|
| | | Buy/Hold Lower Bound | | | | | | |
| | | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
| Turnover Quartile | Q1 | 0.39 | 0.39 | 0.38 | 0.37 | 0.36 | 0.34 | 0.33 |
| | Q2 | 0.31 | 0.32 | 0.31 | 0.31 | 0.30 | 0.29 | 0.28 |
| | Q3 | 0.12 | 0.16 | 0.17 | 0.18 | 0.17 | 0.17 | 0.17 |
| | Q4 | -0.65 | -0.51 | -0.41 | -0.34 | -0.29 | -0.24 | -0.21 |

| | | Panel B: VW NYSE Deciles | | | | |
|----------------------|----|--------------------------|-------------|------|------|-------------|
| | | Buy/Hold Lower Bound | | | | |
| | | 10 | 20 | 30 | 40 | 50 |
| Turnover Quartile | Q1 | 0.33 | 0.28 | 0.26 | 0.24 | 0.23 |
| | Q2 | 0.34 | 0.32 | 0.30 | 0.26 | 0.22 |
| | Q3 | 0.16 | 0.23 | 0.22 | 0.19 | 0.19 |
| | Q4 | 0.07 | 0.23 | 0.28 | 0.31 | 0.32 |

Figure A.2: Distribution of Net Returns: In-Sample, Before Cost Optimization We adjust anomaly returns for effective bid-ask spreads (Figure 3). All portfolios use equal-weighted quintile sorts, following the modal approach in the literature. Anomalies with above median turnover (15% per month, two-sided) are shown in bold. Hash marks indicate larger bins. Published anomaly strategies have a long left tail in net returns, and produce an average net return of only 5 bps per month.

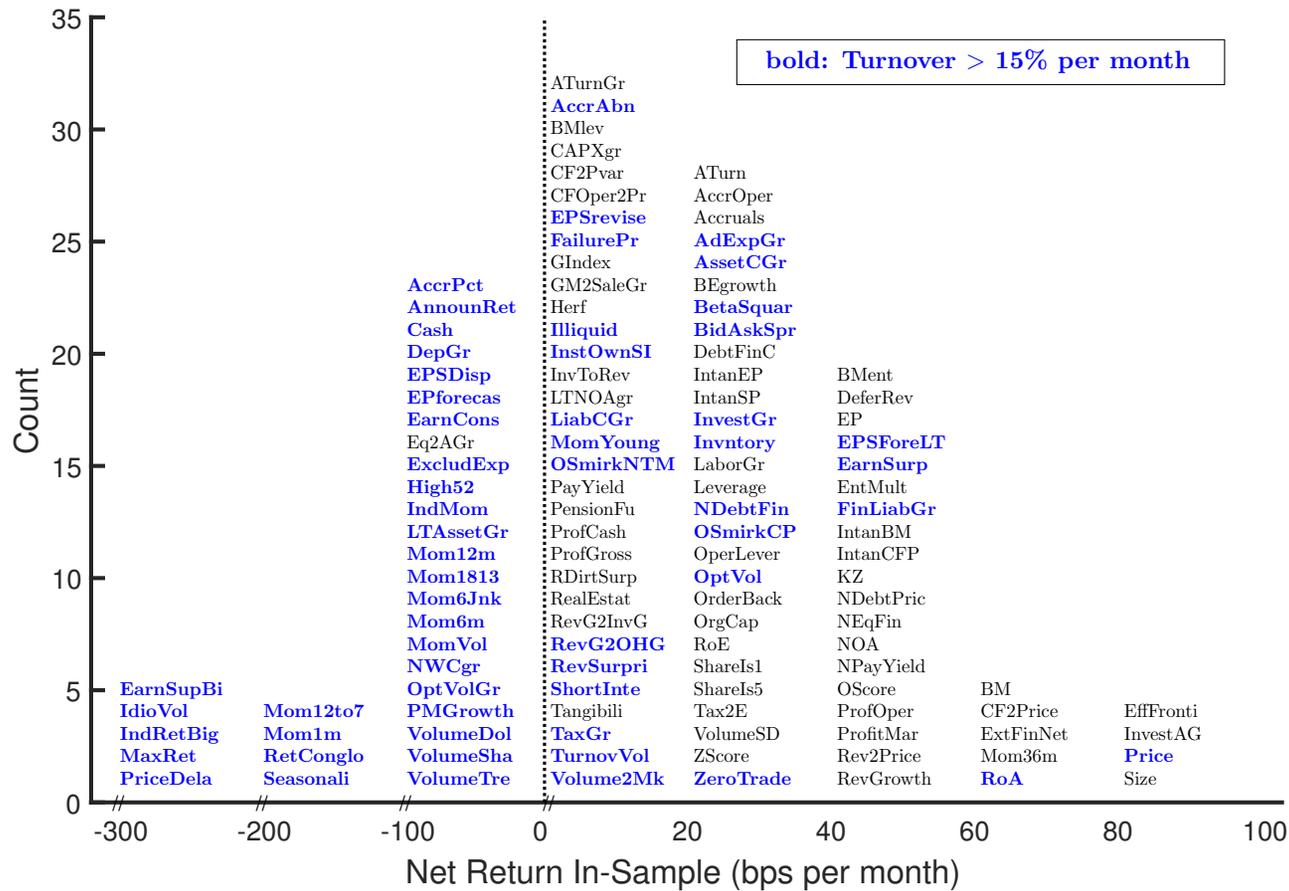
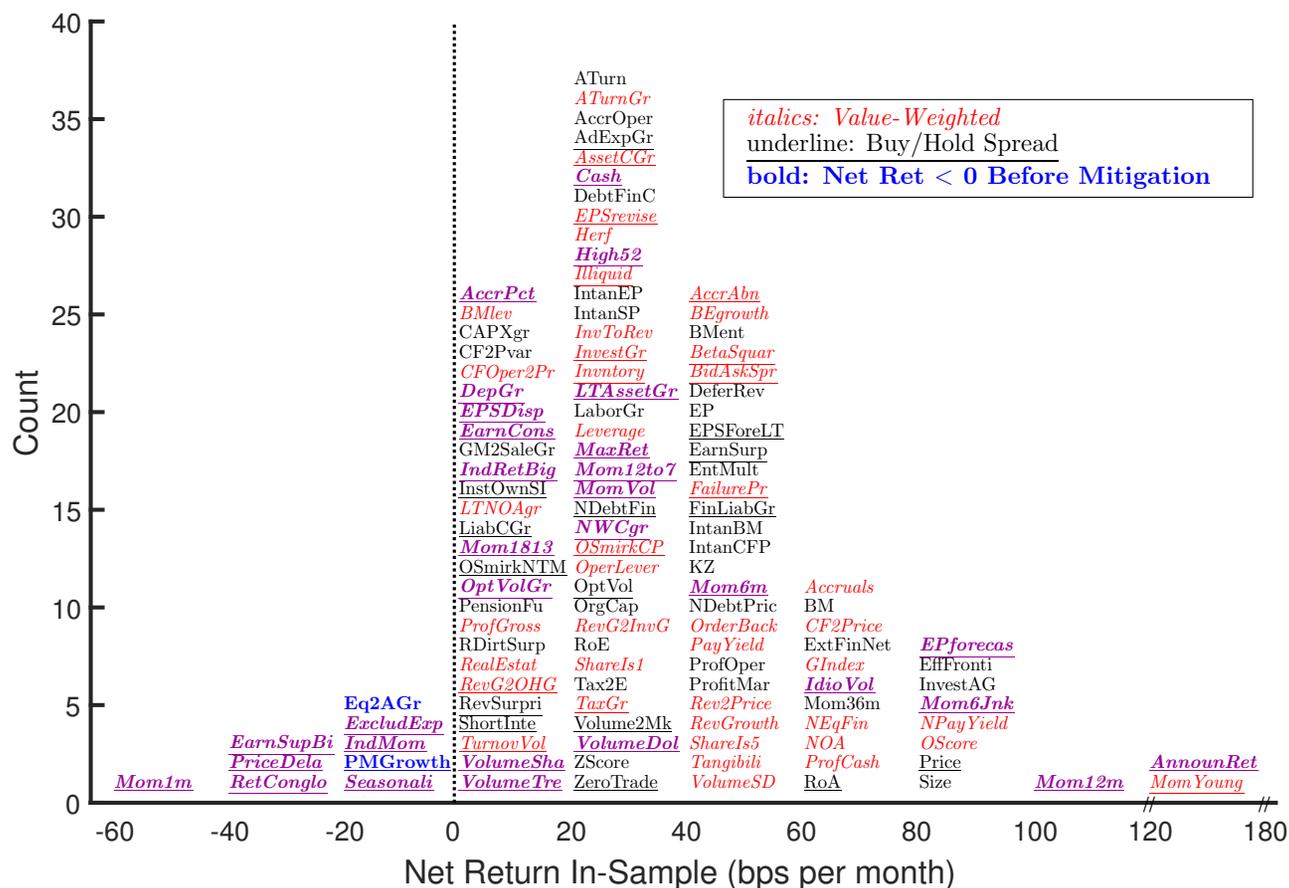


Figure A.3: Cost Optimization Results: Distribution of Net Returns: In-Sample. We mitigate transaction costs by applying value-weighting and/or buy/hold spreads to 120 anomaly portfolios. Buy/hold spreads are chosen to maximize net returns in-sample following Table A.4. Stock weighting is chosen to maximize the in-sample net return given the optimized buy/hold spread. Italicized anomalies benefit from value-weighting. Underlined anomalies benefit from buy/hold spreads. Bold indicates anomalies with negative net returns before cost mitigation. Hash marks indicate larger bins. Cost mitigation leads to positive net returns for the vast majority of anomalies, and raise the average net return to 38 bps per month.



A.5. Additional Results

Figure A.4: Distribution of Publication Years.

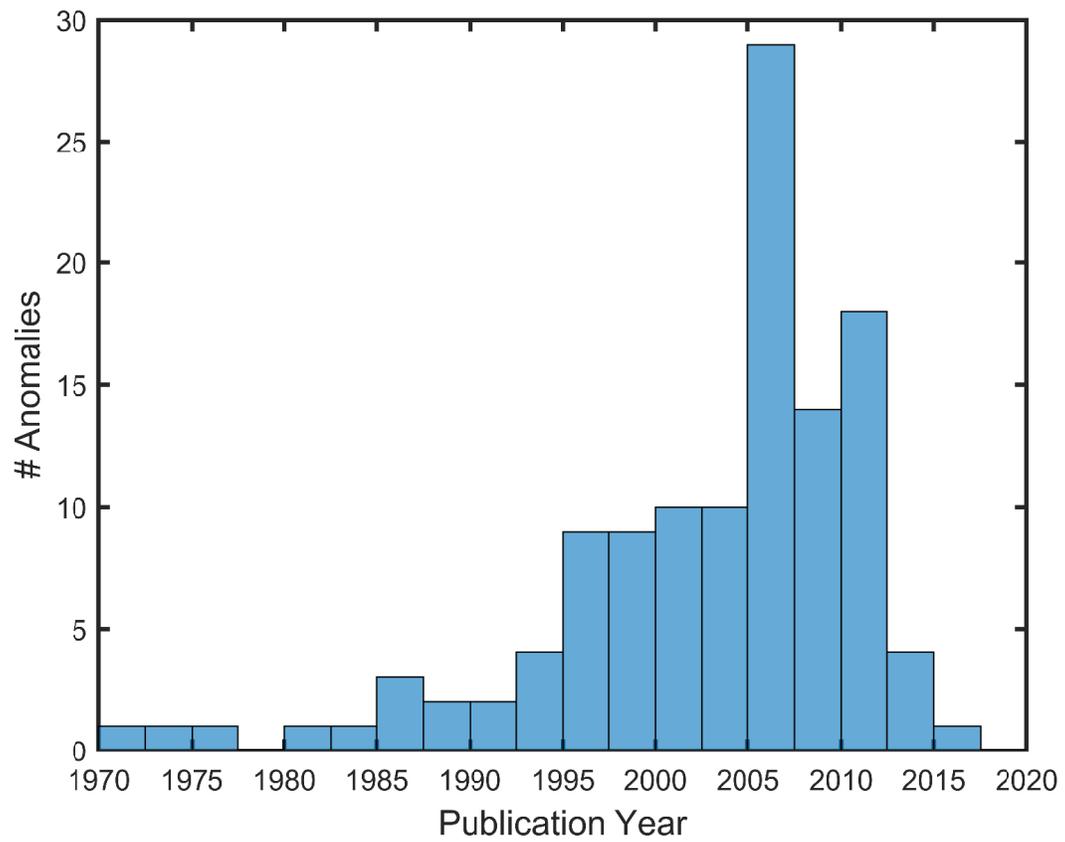


Figure A.5: Distribution of Post-Publication Sample Lengths.

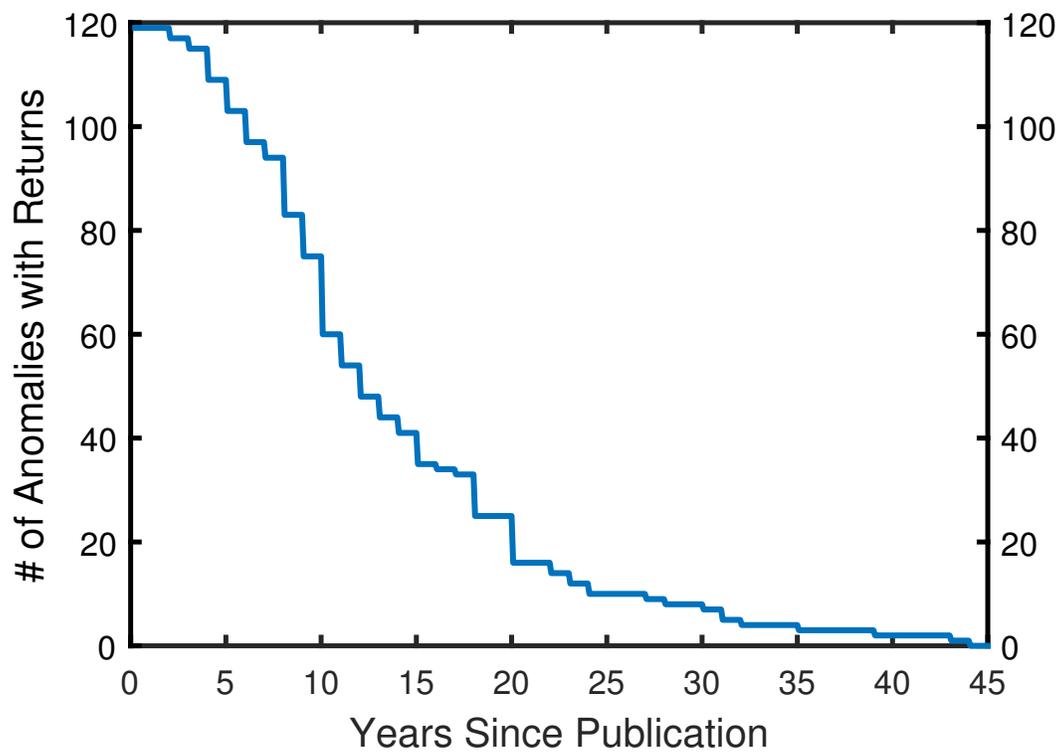


Table A.5: Returns Gross and Net of Trading Costs: Post-Pub and Post-2005

This table shows the same calculations as Table 2 but uses post-2005 data only.

| | (a) Gross Return | (b) Turnover (2-sided) | (c) Ave Spread Paid | (d) \approx (b) \times (c) Return Reduction | (e) = (a) - (d) Net Return |
|--|------------------------|------------------------------|---------------------------|---|----------------------------------|
| Panel A: Equal-Weighted Long-Short Quintiles | | | | | |
| Post-Pub & Post-2005 | 0.30 (0.10) | 0.30 (0.10) | 1.11 (0.10) | 0.32 (0.10) | -0.03 (0.10) |
| Panel B: Cost-Mitigated using Value-Weighting and Buy/Hold Spreads | | | | | |
| Post-Pub & Post-2005 | 0.20 (0.10) | 0.20 (0.10) | 0.60 (0.10) | 0.08 (0.10) | 0.13 (0.10) |
| Panel C: Cost-Mitigated using Buy/Hold Spreads, Value-Weighted only | | | | | |
| Post-Pub & Post-2005 | 0.12 (0.10) | 0.19 (0.10) | 0.31 (0.10) | 0.05 (0.10) | 0.07 (0.10) |

References

- Abdi, Farshid and Angelo Ranaldo. "A Simple Estimation of Bid-Ask Spreads from Daily Close, High, and Low Prices". *The Review of Financial Studies* 30.12 (2017), pp. 4437–4480.
- Azevedo, Eduardo M et al. "Empirical Bayes Estimation of Treatment Effects with Many A/B Tests: An Overview". *AEA Papers and Proceedings*. Vol. 109. 2019, pp. 43–47.
- Ball, Ray, SP Kothari, and Jay Shanken. "Problems in measuring portfolio performance An application to contrarian investment strategies". *Journal of Financial Economics* 38.1 (1995), pp. 79–107.
- Barber, Brad M, Terrance Odean, and Ning Zhu. "Do retail trades move markets?" *The Review of Financial Studies* 22.1 (2008), pp. 151–186.
- Bates, John M and Clive WJ Granger. "The combination of forecasts". *Journal of the Operational Research Society* 20.4 (1969), pp. 451–468.
- Brandt, Michael W, Pedro Santa-Clara, and Rossen Valkanov. "Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns". *The Review of Financial Studies* 22.9 (2009), pp. 3411–3447.
- Briere, Marie et al. "Stock Market Liquidity and the Trading Costs of Asset Pricing Anomalies". *Available at SSRN 3380239* (2019).
- Campbell, John Y and John H Cochrane. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior". *Journal of Political Economy* 107.2 (1999), pp. 205–251.
- Chen, Andrew Y and Tom Zimmermann. "Publication Bias and the Cross-Section of Stock Returns". *The Review of Asset Pricing Studies* (Nov. 2019). raz011. ISSN: 2045-9920. eprint: <https://academic.oup.com/raps/advance->

- article-pdf/doi/10.1093/rapstu/raz011/31618483/raz011.pdf.
URL: <https://doi.org/10.1093/rapstu/raz011>.
- Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong. "Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?" *Journal of Accounting and Economics* 58.1 (2014), pp. 41–58.
- Chu, Yongqiang, David Hirshleifer, and Liang Ma. *The causal effect of limits to arbitrage on asset pricing anomalies*. Tech. rep. National Bureau of Economic Research, 2017.
- Cohen, Lauren, Karl B Diether, and Christopher J Malloy. "Supply and demand shifts in the shorting market". *The Journal of Finance* 62.5 (2007), pp. 2061–2096.
- Cont, Rama and Arseniy Kukanov. "Optimal order placement in limit order markets". *Quantitative Finance* 17.1 (2017), pp. 21–39.
- Corwin, Shane A and Paul Schultz. "A simple way to estimate bid-ask spreads from daily high and low prices". *The Journal of Finance* 67.2 (2012), pp. 719–760.
- Dawid, AP. "Selection paradoxes of Bayesian inference". *Lecture Notes-Monograph Series* (1994), pp. 211–220.
- DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" *The review of Financial studies* 22.5 (2009), pp. 1915–1953.
- DeMiguel, Victor et al. "A portfolio perspective on the multitude of firm characteristics" (Forthcoming).
- Drechsler, Itamar and Qingyi Freda Drechsler. "The Shorting Premium and Asset Pricing Anomalies" (2016).
- Efron, Bradley. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Vol. 1. Cambridge University Press, 2012.

- Efron, Bradley. “Tweedie’s formula and selection bias”. *Journal of the American Statistical Association* 106.496 (2011), pp. 1602–1614.
- Fama, Eugene F and James D MacBeth. “Risk, return, and equilibrium: Empirical tests”. *The Journal of Political Economy* (1973), pp. 607–636.
- Feng, Guan hao, Stefano Giglio, and Dacheng Xiu. “Taming the Factor Zoo” (2017).
- Fong, Kingsley YL, Craig W Holden, and Charles A Trzcinka. “What are the best liquidity proxies for global research?” *Review of Finance* 21.4 (2017), pp. 1355–1401.
- Fong, Kingsley, Craig Holden, and Ondrej Tobek. “Are Volatility Over Volume Liquidity Proxies Useful For Global Or US Research?” (2017).
- Frazzini, Andrea, Ronen Israel, and Tobias J Moskowitz. “Trading costs”. *Available at SSRN 3229719* (2018).
- “Trading costs of asset pricing anomalies” (2015).
- Freyberger, Joachim, Andreas Neuhierl, and Michael Weber. *Dissecting characteristics nonparametrically*. Tech. rep. National Bureau of Economic Research, 2017.
- Gârleanu, Nicolae and Lasse Heje Pedersen. “Efficiently inefficient markets for assets and asset management”. *The Journal of Finance* 73.4 (2018), pp. 1663–1712.
- Glosten, Lawrence R and Paul R Milgrom. “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders”. *Journal of financial economics* 14.1 (1985), pp. 71–100.
- Gompers, Paul, Joy Ishii, and Andrew Metrick. “Corporate governance and equity prices”. *The quarterly journal of economics* 118.1 (2003), pp. 107–156.

- Goyenko, Ruslan Y, Craig W Holden, and Charles A Trzcinka. “Do liquidity measures measure liquidity?” *Journal of Financial Economics* 92.2 (2009), pp. 153–181.
- Green, Jeremiah, John RM Hand, and Frank Zhang. “The remarkable multidimensionality in the cross-section of expected US stock returns”. *Available at SSRN 2262374* (2014).
- Green, Jeremiah, John RM Hand, and X Frank Zhang. “The characteristics that provide independent information about average us monthly stock returns”. *The Review of Financial Studies* (2017), hhx019.
- “The supraview of return predictive signals”. *Review of Accounting Studies* 18.3 (2013), pp. 692–730.
- Grossman, Sanford J and Joseph E Stiglitz. “On the impossibility of informationally efficient markets”. *The American economic review* 70.3 (1980), pp. 393–408.
- Hand, John RM and Jeremiah Green. “The importance of accounting information in portfolio optimization”. *Journal of Accounting, Auditing & Finance* 26.1 (2011), pp. 1–34.
- Hanna, J Douglas and Mark J Ready. “Profitable predictability in the cross section of stock returns”. *Journal of Financial Economics* 78.3 (2005), pp. 463–505.
- Harvey, Campbell R, Yan Liu, and Heqing Zhu. “... and the cross-section of expected returns”. *The Review of Financial Studies* 29.1 (2016), pp. 5–68.
- Hasbrouck, Joel. “Trading costs and returns for US equities: Estimating effective costs from daily data”. *The Journal of Finance* 64.3 (2009), pp. 1445–1477.
- Holden, Craig W and Stacey Jacobsen. “Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions”. *The Journal of Finance* 69.4 (2014), pp. 1747–1785.

Hong, Harrison and Marcin Kacperczyk. “The price of sin: The effects of social norms on markets”. *Journal of Financial Economics* 93.1 (2009), pp. 15–36.

Hou, Kewei, Sehoon Kim, and Ingrid M Werner. “(Priced) Frictions” (2016).

Hou, Kewei, Chen Xue, and Lu Zhang. *Replicating Anomalies*. Tech. rep. National Bureau of Economic Research, 2017.

Huang, Jing-Zhi and Zhijian James Huang. “Real-Time Profitability of Published Anomalies: An Out-of-Sample Test”. *Quarterly Journal of Finance* 3 (2013).

Institute for the Study of Security Markets. *NYSE/AMEX and NASDAQ Historical Tick Data*. Wharton Research Data Services, <http://www.whartonwrds.com/datasets/crsp/>.

Jacobs, Heiko and Sebastian Müller. “Anomalies across the globe: Once public, no longer existent?” (2017).

Jahan-Parvar, Mohammad and Filip Zikes. “When do low-frequency measures really measure transaction costs?” (2019).

James, William and Charles Stein. “Estimation with quadratic loss”. *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 1961. 1961, pp. 361–379.

Karnaukh, Nina, Angelo Ranaldo, and Paul Soderlind. “Understanding FX liquidity”. *The Review of Financial Studies* 28.11 (2015), pp. 3073–3108.

Kelly, Bryan and Hao Jiang. “Tail risk and asset prices”. *Review of Financial Studies* (2014), hhu039.

Knez, Peter J and Mark J Ready. “Estimating the profits from trading strategies”. *The Review of Financial Studies* 9.4 (1996), pp. 1121–1163.

Koch, Andrew, Stefan Ruenzi, and Laura Starks. “Commonality in liquidity: a demand-side explanation”. *The Review of Financial Studies* 29.8 (2016), pp. 1943–1974.

- Korajczyk, Robert A and Ronnie Sadka. "Are momentum profits robust to trading costs?" *The Journal of Finance* 59.3 (2004), pp. 1039–1082.
- Kyle, Albert S and Anna A Obizhaeva. "Market microstructure invariance: Empirical hypotheses". *Econometrica* 84.4 (2016), pp. 1345–1404.
- Leland, Hayne. "Optimal portfolio implementation with transactions costs and capital gains taxes". *Haas School of Business Technical Report* (2000).
- Lesmond, David A, Michael J Schill, and Chunsheng Zhou. "The illusory nature of momentum profits". *Journal of financial economics* 71.2 (2004), pp. 349–380.
- Liu, Hong. "Optimal consumption and investment with transaction costs and multiple risky assets". *The Journal of Finance* 59.1 (2004), pp. 289–338.
- Liu, Laura, Hyungsik Roger Moon, and Frank Schorfheide. "Forecasting with dynamic panel data models". *Econometrica* 88.1 (2020), pp. 171–201.
- Lo, Andrew W. "The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective". *Journal of Portfolio Management* 30 (2004), pp. 15–29.
- Lou, Xiaoxia and Tao Shu. "Price impact or trading volume: Why is the Amihud (2002) illiquidity measure priced". *Available at SSRN* 2291942 (2014).
- Magill, Michael JP and George M Constantinides. "Portfolio selection with transactions costs". *Journal of Economic Theory* 13.2 (1976), pp. 245–263.
- Marquering, Wessel, Johan Nisser, and Toni Valla. "Disappearing anomalies: a dynamic analysis of the persistence of anomalies". *Applied Financial Economics* 16.4 (2006), pp. 291–302.
- McLean, R David. "Idiosyncratic risk, long-term reversal, and momentum". *Journal of Financial and Quantitative Analysis* 45.4 (2010), pp. 883–906.
- McLean, R David and Jeffrey Pontiff. "Does academic research destroy stock return predictability?" *The Journal of Finance* 71.1 (2016), pp. 5–32.

- Moallemi, Ciamac C. and Mehmet Saglam. “Dynamic Portfolio Choice with Linear Rebalancing Rules”. *Journal of Financial and Quantitative Analysis* 52.3 (2017), 1247fffd1278.
- New York Stock Exchange. *Daily TAQ (Historical Trades Quotes)*. Wharton Research Data Services, <http://www.whartonwrds.com/datasets/crsp/>.
- *Monthly TAQ (Historical Trades Quotes)*. Wharton Research Data Services, <http://www.whartonwrds.com/datasets/crsp/>.
- Novy-Marx, Robert and Mihail Velikov. “A taxonomy of anomalies and their trading costs”. *Review of Financial Studies* 29.1 (2016), pp. 104–147.
- “Comparing Cost-Mitigation Techniques”. *Financial Analysts Journal* 75.1 (2019), pp. 85–102.
- Patton, Andrew J and Brian M Weller. “What you see is not what you get: The costs of trading market anomalies” (2017).
- Perold, Andre F. “The implementation shortfall: Paper versus reality”. *Journal of Portfolio Management* 14.3 (1988), p. 4.
- Pontiff, Jeffrey and Michael Schill. “Long-run seasoned equity offering returns: Data snooping, model misspecification, or mispricing? A costly arbitrage approach” (2001).
- Ritter, Jay R. “The long-run performance of initial public offerings”. *The journal of finance* 46.1 (1991), pp. 3–27.
- Roll, Richard. “A simple implicit measure of the effective bid-ask spread in an efficient market”. *The Journal of finance* 39.4 (1984), pp. 1127–1139.
- Schultz, Paul. “Transaction costs and the small firm effect: A comment”. *Journal of Financial Economics* 12.1 (1983), pp. 81–88.
- Schwert, G William. “Anomalies and market efficiency”. *Handbook of the Economics of Finance* 1 (2003), pp. 939–974.

- Senn, Stephen. "A note concerning a selection "paradox" of Dawid's". *The American Statistician* 62.3 (2008), pp. 206–210.
- Stigler, Stephen M. "The 1988 Neyman memorial lecture: a Galtonian perspective on shrinkage estimators". *Statistical Science* 5.1 (1990), pp. 147–155.
- Stoll, Hans R. "Market microstructure". *Handbook of the Economics of Finance*. Vol. 1. Elsevier, 2003, pp. 553–604.
- Stoll, Hans R and Robert E Whaley. "Transaction costs and the small firm effect". *Journal of Financial Economics* 12.1 (1983), pp. 57–79.
- Timmermann, Allan. "Forecast combinations". *Handbook of economic forecasting* 1 (2006), pp. 135–196.
- WRDS: Center for Research in Security Prices. *CRSP/Compustat Merged Database*. Wharton Research Data Services, <http://www.whartonwrds.com/datasets/crsp/>.
- Xie, Xianchao, SC Kou, and Lawrence D Brown. "SURE estimates for a heteroscedastic hierarchical model". *Journal of the American Statistical Association* 107.500 (2012), pp. 1465–1479.

Tables and Figures

Table 1: Correlations Between Low-Frequency Proxies and High-Frequency Effective Bid-Ask Spreads

Correlations are pooled. We examine four low frequency proxies for spreads: Gibbs is Hasbrouck's (2009) Gibbs estimate of the Roll model, HL is Corwin and Schultz's (2012) high-low spread, CHL is Abdi and Rinaldo's (2017) close-high-low, and VoV (volume-over-volatility) is Fong, Holden, and Tobek's (2017) implementation of Kyle and Obizhaeva (2016) microstructure invariance hypothesis. LF_ave is the equal weighted average of the four low frequency proxies. TAQ and ISSM are computed from high-frequency data. The low frequency measures are imperfectly correlated, suggesting that they contain distinct information. LF_ave has the highest correlation with high-frequency spreads. LF spread data are available at <http://sites.google.com/site/chenandrewy/code-and-data/>.

| Panel A: LF spread correlations (1926-2017; 2,114,436 obs.) | | | | | | |
|---|-------|------|------|------|--|--|
| | Gibbs | HL | CHL | VoV | | |
| Gibbs | 1.00 | | | | | |
| HL | 0.68 | 1.00 | | | | |
| CHL | 0.76 | 0.88 | 1.00 | | | |
| VoV | 0.75 | 0.59 | 0.74 | 1.00 | | |

| Panel B: Correlations with TAQ (1993-2014; 1,183,068 obs.) | | | | | | |
|--|------|-------|------|------|------|--------|
| | TAQ | Gibbs | HL | CHL | VoV | LF_Ave |
| TAQ | 1.00 | | | | | |
| Gibbs | 0.84 | 1.00 | | | | |
| HL | 0.71 | 0.67 | 1.00 | | | |
| CHL | 0.80 | 0.74 | 0.88 | 1.00 | | |
| VoV | 0.84 | 0.73 | 0.60 | 0.75 | 1.00 | |
| LF_Ave | 0.90 | 0.90 | 0.86 | 0.93 | 0.87 | 1.00 |

| Panel C: Correlations with ISSM (1983-1992; 262,381 obs.) | | | | | | |
|---|------|-------|------|------|------|--------|
| | ISSM | Gibbs | HL | CHL | VoV | LF_Ave |
| ISSM | 1.00 | | | | | |
| Gibbs | 0.88 | 1.00 | | | | |
| HL | 0.84 | 0.79 | 1.00 | | | |
| CHL | 0.90 | 0.84 | 0.92 | 1.00 | | |
| VoV | 0.86 | 0.82 | 0.66 | 0.78 | 1.00 | |
| LF_Ave | 0.94 | 0.95 | 0.90 | 0.95 | 0.88 | 1.00 |

Table 2: Zeroing in on the Average Anomaly's Expected Return

We estimate the average net return (e) of 120 anomaly long-short portfolios after accounting for effective bid-ask spreads and stale data. All figures are in bps per month except for turnover, which is a ratio per month. Figures average across months and then across anomalies, with standard errors in parentheses. Panel A examines the typical academic implementation (Section 2.3.1). Panels B and C examine cost-optimized implementations (Section 2.3.2). Columns (a)-(d) report an approximate net return decomposition. Anomalies are drawn from McLean and Pontiff (2016), Green, Hand, and Zhang (2017), and Hou, Xue, and Zhang (2017) (Section 2.1, Tables A.1-A.3). After accounting for trading costs and stale data, the expected return is approximately zero. Source: Center for Research in Security Prices, New York Stock Exchange, and Institute for the Study of Security Markets.

| | (a) Gross Return | (b) Turnover (2-sided) | (c) Ave Spread Paid | (d) \approx (b) \times (c) Return Reduction | (e) = (a) - (d) Net Return |
|--|------------------------|------------------------------|---------------------------|---|----------------------------------|
| Panel A: Equal-Weighted Long-Short Quintiles | | | | | |
| In-Sample | 66 (4) | 0.31 (0.04) | 219 (6) | 61 (7) | 5 (6) |
| Post-Publication | 30 (4) | 0.30 (0.04) | 111 (6) | 32 (5) | -3 (5) |
| Panel B: Cost-Optimized | | | | | |
| In-Sample | 59 (4) | 0.20 (0.02) | 136 (7) | 21 (2) | 38 (3) |
| Post-Publication | 20 (4) | 0.20 (0.02) | 60 (6) | 8 (1) | 13 (4) |
| Post-Pub & Post-2005 | 14 (4) | 0.20 (0.02) | 46 (4) | 6 (1) | 8 (4) |
| Panel C: Cost-Optimized, Value-Weighted only | | | | | |
| In-Sample | 46 (4) | 0.20 (0.02) | 86 (5) | 16 (2) | 30 (3) |
| Post-Publication | 12 (3) | 0.19 (0.02) | 31 (5) | 5 (1) | 7 (3) |
| Post-Pub & Post-2005 | 7 (3) | 0.19 (0.02) | 21 (3) | 3 (0) | 4 (3) |

Table 3: The Best Expected Returns Using Out-of-Sample Tests

To avoid data-mining bias, we sort anomaly portfolios based on in-sample data and average net returns post-publication and post-2005. Quartiles are numbered from worst expected net return a priori—for example, quartile 1 has the highest turnover, and quartile 4 has the lowest turnover. All portfolio implementations use cost-optimization following Section 2.3.2. Panel B restricts implementations to value-weighting. Even the strongest anomalies have expected returns of only 10-20 bps per month. Source: Center for Research in Security Prices, New York Stock Exchange, and Institute for the Study of Security Markets.

| Panel A: Including Equal-Weighting | | | | |
|------------------------------------|---|--------------|---------------|---------------|
| In-Sample Predictor | Post-Pub Post-05 Net Return (bps monthly) | | | |
| | Predictor Quartile | | | |
| | 1 (Worst) | 2 | 3 | 4 (Best) |
| Net Return | 4.8 (5.8) | 6.0 (6.2) | 12.5 (6.8) | 10.6 (7.4) |
| Net Sharpe | 4.9 (6.9) | 3.5 (7.0) | 3.9 (6.4) | 21.2 (6.2) |
| Return Reduction | 14.0 (7.1) | 7.0 (6.4) | 8.5 (5.7) | 4.3 (7.1) |
| Turnover | 1.4 (8.1) | 7.2 (6.2) | 11.0 (5.5) | 14.3 (6.5) |

| Panel B: Value-Weighted Only | | | | |
|------------------------------|---|---------------|---------------|---------------|
| In-Sample Predictor | Post-Pub Post-05 Net Return (bps monthly) | | | |
| | Predictor Quartile | | | |
| | 1 (Worst) | 2 | 3 | 4 (Best) |
| Net Return | 1.4 (6.6) | 9.3 (7.8) | -4.6 (7.4) | 10.6 (8.3) |
| Net Sharpe | -0.7 (7.0) | 10.2 (8.7) | -4.2 (7.7) | 11.4 (7.0) |
| Return Reduction | 4.1 (8.4) | 1.9 (7.1) | 9.3 (7.4) | 0.2 (7.2) |
| Turnover | 2.4 (8.1) | -0.5 (7.0) | 4.3 (7.2) | 9.7 (7.9) |

Table 4: Empirical Bayes Estimates of the Best Expected Returns

We adjust large mean net returns in post-publication and post-2005 (post-pub05) samples for data-mining using empirical Bayes. Bootstrapped standard errors are in parentheses. Adjustments assume Sharpe ratios are the sum of the true Sharpe ratio and an error term, and true Sharpe ratios are t-distributed with d.o.f. ν_{SR} , scale σ_{SR} , and mean μ_{SR} . Given ν_{SR} , we estimate σ_{SR} and μ_{SR} by method of moments (Equation (7)). Adjusted expected returns are computed from the conditional expectation of true Sharpe ratios (Equation (10)). Value-weighted implementations imply that method of moments hits a positivity constraint, and thus $\hat{\sigma}_{SR} = 0$. Even the strongest anomalies have expected returns of only 5-20 bps per month, consistent with Table 3. Source: Center for Research in Security Prices, New York Stock Exchange, and Institute for the Study of Security Markets.

| Panel A: Including Equal-Weighting | | | | | | |
|------------------------------------|---------------------|------------------|-------------------------------------|-------|-------|-------|
| Parameters (annualized) | | | Post-Pub05 Net Return (bps monthly) | | | |
| Assumed | Estimated | | Percentile | | | |
| ν_{SR} | $\hat{\sigma}_{SR}$ | $\hat{\mu}_{SR}$ | 50 | 70 | 80 | 90 |
| 100 | 0.20 | 0.11 | 10.2 | 14.3 | 18.9 | 21.3 |
| | (0.06) | (0.03) | (3.2) | (4.1) | (4.2) | (5.0) |
| 4 | 0.15 | 0.11 | 10.0 | 14.3 | 18.1 | 20.2 |
| | (0.05) | (0.03) | (3.1) | (3.8) | (4.0) | (4.3) |

| Panel B: Value-Weighted Only | | | | | | |
|------------------------------|---------------------|------------------|-------------------------------------|-------|-------|-------|
| Parameters (annualized) | | | Post-Pub05 Net Return (bps monthly) | | | |
| Assumed | Estimated | | Percentile | | | |
| ν_{SR} | $\hat{\sigma}_{SR}$ | $\hat{\mu}_{SR}$ | 50 | 70 | 80 | 90 |
| 100 | 0.00 | 0.04 | 4.1 | 4.9 | 5.6 | 6.4 |
| | (0.06) | (0.03) | (3.5) | (4.4) | (4.7) | (5.3) |
| 4 | 0.00 | 0.04 | 4.1 | 4.9 | 5.6 | 6.4 |
| | (0.04) | (0.03) | (3.6) | (4.3) | (4.6) | (5.2) |

Table 5: Performance of Size, B/M, and Momentum

Returns are in bps per month. Post-Pub05 is our baseline post-publication and post-2005 sample, and is equivalent to 2006-2016 for these three anomalies. FIM (2015) is taken from Table IV of Frazzini, Israel, and Moskowitz (2015). Size, B/M, and momentum perform well in earlier data, consistent with FIM. The performance of individual anomalies is highly sensitive to the sample period, and thus we need many anomalies to estimate expected returns post-2005. Source: Center for Research in Security Prices, New York Stock Exchange, and Institute for the Study of Security Markets.

| Panel A: Size | | | |
|---------------|-------------------------|----------------|-------------------------|
| Return | Post-Pub05 2006-2016 | 1998-2013 | FIM (2015) 1998-2013 |
| Gross | -25.8 (33.9) | 60.0 (39.2) | 66.5 (22.1) |
| Net | -33.1 (33.8) | 48.5 (39.2) | 54.3 (21.9) |

| Panel B: B/M | | | |
|--------------|-------------------------|----------------|-------------------------|
| Return | Post-Pub05 2006-2016 | 1998-2013 | FIM (2015) 1998-2013 |
| Gross | 32.9 (28.7) | 79.9 (31.3) | 40.5 (36.2) |
| Net | 24.5 (29.1) | 66.4 (31.8) | 29.3 (36.6) |

| Panel C: Momentum | | | |
|-------------------|-------------------------|----------------|-------------------------|
| Return | Post-Pub05 2006-2016 | 1998-2013 | FIM (2015) 1998-2013 |
| Gross | 16.4 (60.6) | 36.2 (59.9) | 18.8 (47.1) |
| Net | 12.6 (60.5) | 28.8 (59.9) | -6.4 (45.8) |

Figure 2: The Bias in Low-Frequency Effective Spread Proxies. We take the difference between low-frequency proxies and TAQ spreads at the firm-month level, and then take the median across firms to calculate an error in each month. Low-frequency spreads are from Hasbrouck (2009) (Gibbs), Corwin and Schultz (2012) (HL), Abdi and Ranaldo (2017) (CHL), and Kyle and Obizhaeva (2016) (VoV). Post-decimalization, low-frequency proxies are biased upward by roughly 25-50 bps.

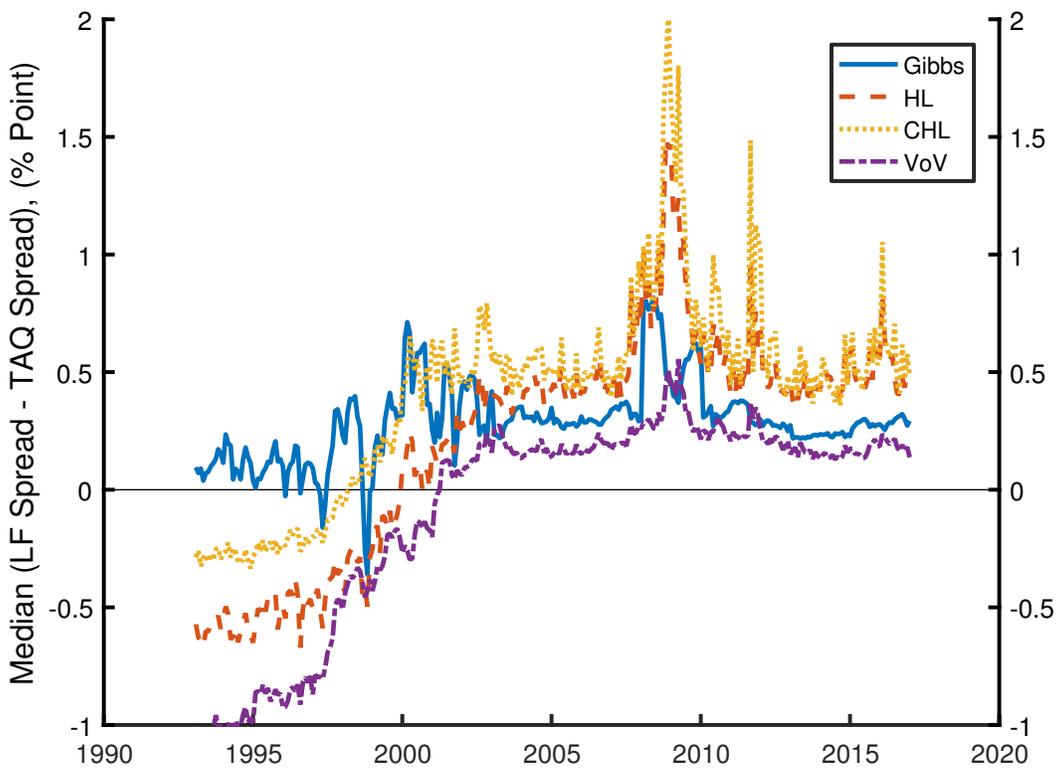


Figure 3: Combined Effective Spreads Over Time. Spreads combine high-frequency and low-frequency data. We use high-frequency Daily TAQ (DTAQ), Monthly TAQ (MTAQ), and ISSM when available. Otherwise, we use the average of four low frequency proxies: Gibbs (Hasbrouck 2009), HL (Corwin and Schultz 2012), CHL (Abdi and Rinaldo 2017), and VoV (Kyle and Obizhaeva 2016). The combined spread tracks well-known structural changes like the entry of NASDAQ (early 1970s) and decimalization (early 2000s). LF spread data are available at <http://sites.google.com/site/chenandrewy/code-and-data/>. Source: Center for Research in Security Prices, New York Stock Exchange, and Institute for the Study of Security Markets.

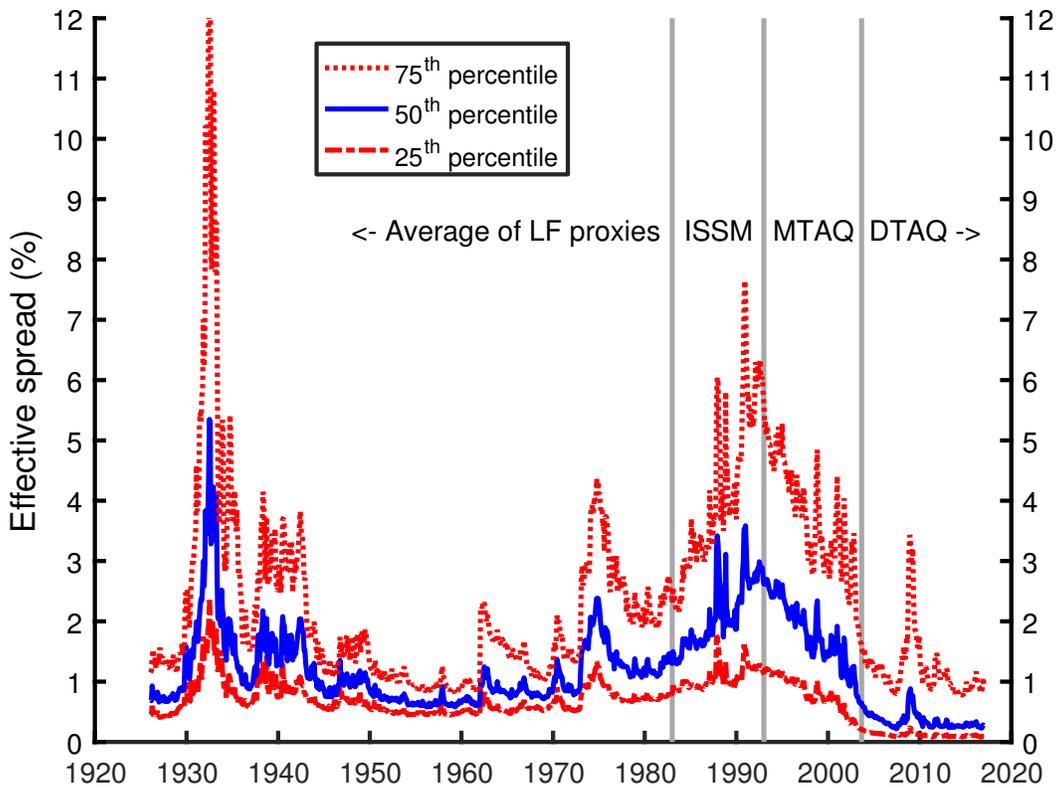


Figure 4: Distribution of Spreads Paid by Academic Implementations in 2014.

We compare the effective spreads paid by academic implementations with those of all stocks, NYSE stocks, and Russell 2000 stocks. “Paid by anomaly portfolios” pools across all trades implied by 120 academic implementations in 2014. Other distributions are pooled across all stock-months in 2014. Academic implementations trade stocks across the entire liquidity spectrum, resulting in large trading costs despite the near-zero modal spreads of recent years. Source: Center for Research in Security Prices, New York Stock Exchange, and Institute for the Study of Security Markets.

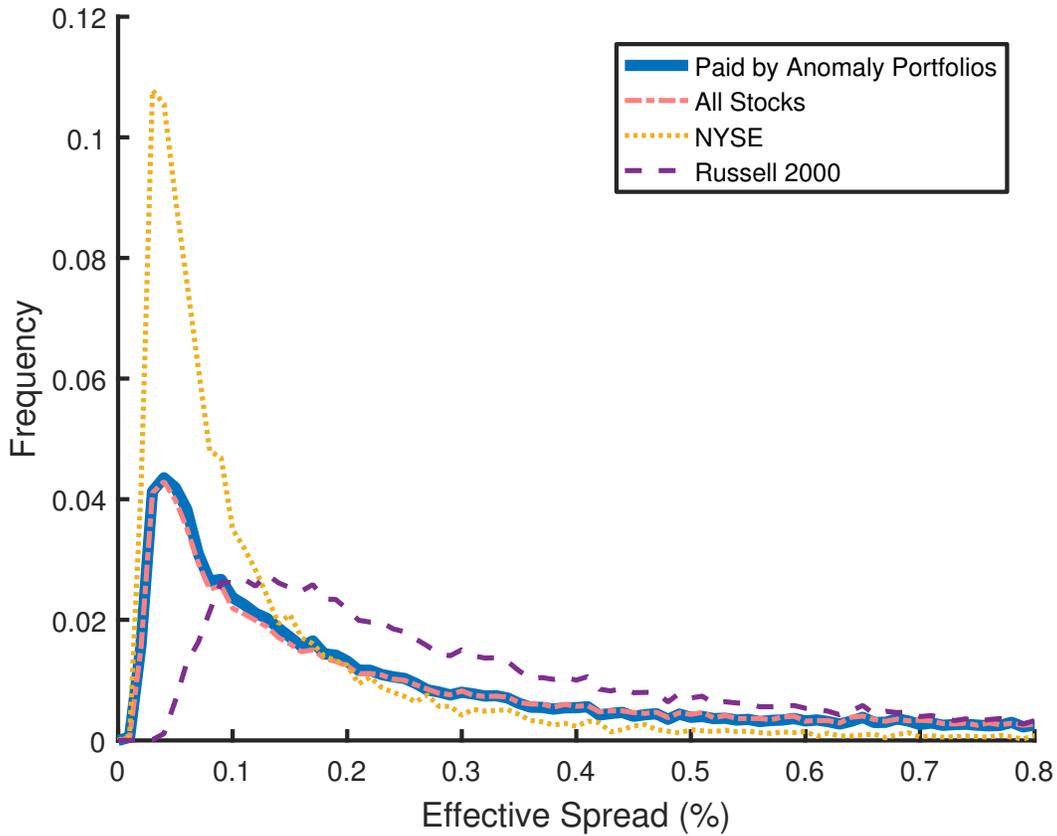


Figure 5: Event-Time Net Returns for Cost-Optimized Implementations. For a given month relative to publication, light lines plot the mean net return across all anomalies. Dark lines show the trailing 5-year moving average of mean returns, and dashed lines show 2 standard error confidence bounds. Cost optimization is effective before publication, but net returns become tiny afterwards. Source: Center for Research in Security Prices, New York Stock Exchange, and Institute for the Study of Security Markets.

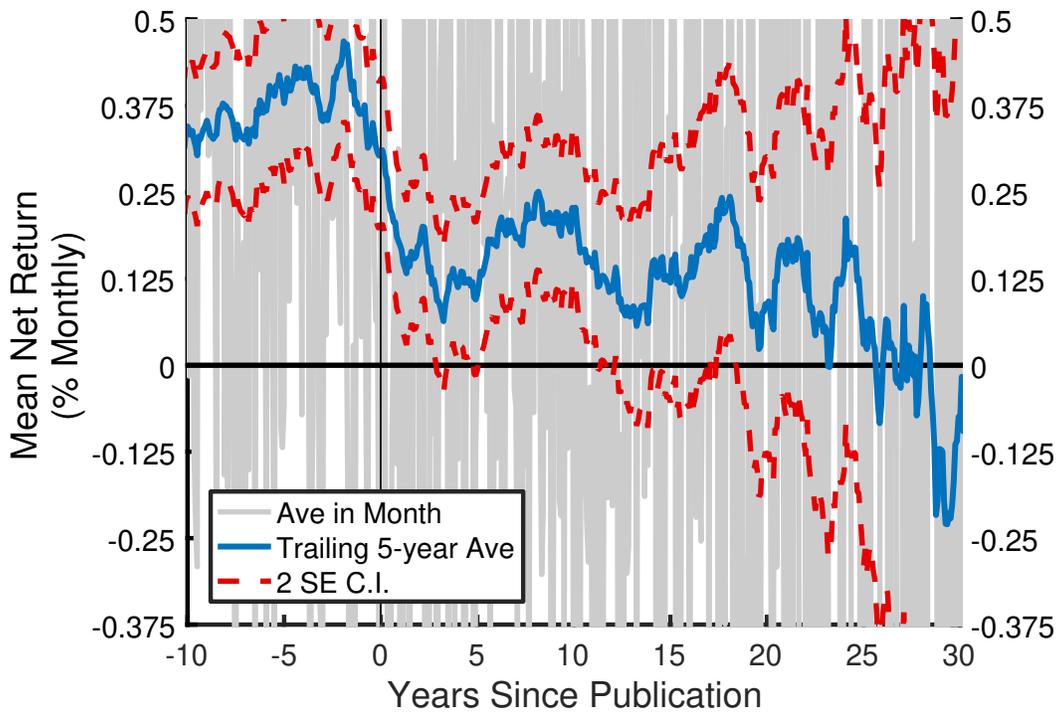


Figure 6: Heterogeneity in Cost-Optimized Mean Net Returns Post-Publication and Post-2005. Many anomalies have notable net returns, but the distribution closely resembles the null of no predictability, consistent with the idea notable net returns are largely due to luck. Source: Center for Research in Security Prices, New York Stock Exchange, and Institute for the Study of Security Markets.

