Accounting for the Anomaly Zoo: a Trading Cost Perspective

Andrew Y. Chen
Federal Reserve Board
andrew.y.chen@frb.gov

Mihail Velikov
Pennsylvania State University
velikov@psu.edu

September 2019*

Abstract

We study the post-publication trading costs of 120 stock market anomalies. Trading costs use effective bid-ask spreads from high-frequency ISSM and TAQ data when available and average four low-frequency proxies otherwise. The average equal-weighted long-short portfolio nets -3 bps per month post-publication after costs. Optimized cost mitigation using value-weighting and buy/hold spreads dramatically improves net returns in-sample but nets only 4 to 12 bps post-publication on average. The strongest cost-optimized anomalies in-sample net just 10-20 bps post-publication. These results show that the average investor should expect tiny profits (alternatively, a tiny risk premium) from investing in any individual anomaly.

*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, Nitish Sinha, Tugkan Tuzun, Michael Weber, 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, and 2019 Eastern Finance Association Meetings 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

This paper adjusts the post-publication returns of 120 stock market anomalies for trading costs. We find that the average equal-weighted long-short quintile portfolio has a negative return of -3 bps per month post-publication and net of costs. Using portfolios that optimally mitigate costs with value-weighting and/or a buy/hold spread rule following Novy-Marx and Velikov (2016), the average post-publication net return is a tiny 4 to 13 basis points per month (0.48% to 1.56% per year), depending on how the average is calculated.

These conclusions come from adjusting the returns of anomalies for effective bid-ask spreads. In general, trading costs are a high-dimensional object that depends on the trader under consideration. We focus on effective spreads because they lead to a simple interpretation: our costs are the lower bound for the average trader who uses market orders. Thus, even the tiny net returns we compute may be unachievable for many traders. Indeed, short sale costs average 10-20 basis points per month (Cohen, Diether, and Malloy 2007; Drechsler and Drechsler 2016), and would likely eliminate the remaining profits.

Post-publication net returns are important because they tell us how much profit we should expect from anomalies in the future. In a practical sense, these are the only returns that matter. Gross returns (before trading costs) are simply not profits. In-sample net returns are largely not available to investors, as anomalies are typically discovered at the end of their in-sample periods. Only by measuring post-publication net returns can we get a sense of future anomaly profits, free of trading costs and publication effects like data-mining bias.

More generally, post-publication net returns isolate the permanent and risk-based component of predictability. Roughly 50% of the average anomaly return is transient and disappears after publication, suggesting that mispricing and data-mining bias play a large role (McLean and Pontiff 2016). The remaining 50%, then, is due to more permanent effects like risk and market frictions. As described by Cochrane (1999), "[i]f a high average return comes from exposure to risk, ... Even if the opportunity is widely publicized, investors will not change their portfolio decisions, and the relatively high average return will remain." Isolating this risk component, however, requires first removing trading costs.

We measure trading costs using high frequency (HF) data whenever it is available. Our HF data combines data from the Institute for the Study of Security

Markets (ISSM) and NYSE's Trade and Quote (TAQ) database. Nearly all of our post-publication trading costs use HF data, as 97% of the anomalies we study are published after 1983, when the HF data begins. When HF data is unavailable, we use the average of four low-frequency (LF) proxies: Hasbrouck's (2009) Gibbs estimate; Corwin and Schultz's (2012) high-low spread; Fong, Holden, and Tobek's (2017) implementation of Kyle and Obizhaeva's (2016) invariance hypothesis; and Abdi and Ranaldo's (2017) close-high-low measure.

The effect of trading costs can be understood in the following back-of-the-envelope calculation. The typical anomaly hedge portfolio requires turning over 15% of its positions on each leg (long and short) each month, and each of these transactions requires paying a bid-ask spread. The average spread paid in the post-publication sample is about 100 bps. These costs are weighed against the average post-publication gross return of 30 bps per month. Putting these numbers together, we get the following back-of-the envelope calculation:

$$\begin{aligned}
\text{[Net Return]} &\approx \text{[Gross Return]} - 2 \times \text{[Each Leg's Turnover]} \times \text{[Bid-Ask Spread]} \\
&= 30 \text{ bps} - 2 \times 0.15 \times 100 \text{ bps} \\
&= 0 \text{ bps per month.}
\end{aligned}$$

Thus, trading costs wipe out the remaining post-publication predictability.

One may be surprised that bid-ask spreads are so large post-publication, given the tiny spreads of the post-decimalization era. The distribution of spreads has a very long right tail, however, and published anomaly strategies require trading all over the spread distribution. As a result, the typical spread paid by equal-weighted anomaly strategies in 2014 is 67 bps, an order of magnitude larger than the modal spread of 5 bps.

Perhaps published strategies can be made profitable if they are modified to avoid trading costs. To examine this question, we mitigate transaction costs by applying value-weighting and/or a buy/hold spread following Novy-Marx and Velikov (2016). Value-weighting reduces costs by reducing the average spread paid. Buy/hold spreads reduce costs by reducing turnover, mimicking the optimal portfolio rule in the presence of transaction costs (Magill and Constantinides 1976; Brandt, Santa-Clara, and Valkanov 2009). We choose stock weighting and buy/hold spread parameters to maximize net returns in-sample. This optimization increases the average net return in-sample from 5 bps to 38 bps per month, consistent with Novy-Marx and Velikov (2016), who find that buy/hold spreads outperform a variety of other cost mitigation techniques.

Post-publication, however, the average cost-mitigated anomaly net return is only 13 bps. Moreover, even these modest profits are fragile. Net returns tend to decay as time-since-publication increases. As a result, the average net return is sensitive to the details of the averaging. The 13 bps mentioned previously focuses on the relatively high net returns in the years soon after publication, as it comes from first averaging across time and then averaging across anomalies. Averaging across anomalies and then averaging across time leads to a tiny 4 bps net return.

Net returns are also concentrated among anomalies that perform better using equal-weighting. Indeed, restricting our cost-mitigated strategies to using value-weighting leads to an average post-publication net return of 4 to 7 bps, once again depending on how the average is taken.

So far, we've discussed results that average across all anomalies. Averaging keeps the analysis simple, but ignores heterogeneity. Indeed, we find that size, B/M, and momentum performed well post-publication net of trading costs (consistent with Frazzini, Israel, and Moskowitz 2015 and Briere, Lehalle, Nefedova, and Raboun 2019). Could there be a subset of exceptional anomalies that survives trading costs post-publication?

Answering this question is not straightforward, as focusing on the right tail of the net return distribution introduces selection bias. This bias is especially pronounced in the relatively short post-publication samples. To control for selection bias, we examine the best-performing anomalies using two approaches: (1) we attempt to forecast post-publication net returns using in-sample information and (2) we estimate a simple model that explicitly accounts for selection bias following the empirical Bayes literature (Efron 2012; Azevedo, Deng, Montiel Olea, and Weyl 2019; Liu, Moon, and Schorfheide Forthcoming).

Both methods lead to a similar result: the strongest anomalies deliver only 10-20 bps of net returns post-publication. Overall, the average investor should expect little profit in the future from published anomalies, even before any risk adjustment is made. Alternatively, there is little room for anomaly risk premiums for the average investor, once limits to arbitrage are accounted for.

Our results paint a picture of a dynamic equilibrium process, but one more in line with Lo's (2004) adaptive market hypothesis than standard dynamic equilibrium models (Campbell and Cochrane 1999; Bansal and Yaron 2004). Every month, academic researchers and market participants find imperfections in the existing market equilibrium. After discovery, the net returns of these imperfections are traded away, leading to a new equilibrium.

We do not show that all anomalies are unprofitable to all investors after publication. A perfectly efficient market is impossible, as shown by Grossman and Stiglitz (1980). Rather, our results measure the magnitude of anomaly profits available to the average trader. These profits are not only relevant for the majority of market participants, but also informative about the overall efficiency of financial markets, imperfect as they must be. Indeed, our finding that the average trader should expect to earn negligible anomaly profits is closely in-line with the predictions of Gârleanu and Pedersen's (2018) refinement of the Grossman and Stiglitz (1980) model: in an "efficiently inefficient" market, the average trader is best served by investing passively.

Related Literature Our paper is closely related to Novy-Marx and Velikov (2016), who also study effective spread costs for a broad array of anomalies. We differ from Novy-Marx and Velikov in that (1) we have a much more accurate measure of trading costs that combines several HF and LF measures, and (2) we use a much larger set of anomalies. These two improvements allow us to make strong conclusions about post-publication net returns, the profits available to average investors in the future, and the risk premium average investors should expect. HF data is important for accuracy. LF spreads are highly correlated with HF spreads, but produce large root-mean-squared errors (Fong, Holden, and Trzcinka 2017). Indeed, we find that low frequency spreads are biased upward by 25-50 bps in the post-decimalization sample (Appendix A.4).

Another closely related paper is DeMiguel, Martin-Utrera, Nogales, and Uppal (2017), who study trading costs and 50 anomaly signals from a portfolio choice perspective. They achieve out-of-sample annual Sharpe ratios in excess of 1.0 net of trading costs using state-of-the-art portfolio optimization techniques. In contrast, our study's long-short portfolios are more accessible to the average trader. Moreover, while DeMiguel et al focus on Brandt, Santa-Clara, and Valkanov's (2009) reduced form trading costs, our study combines several data sources to create a more accurate stock-level measure. Nevertheless, our studies complement each other, as they both emphasize the importance of thinking of anomalies in combination, as each anomaly on its own is not especially powerful.

More broadly, our paper builds on a large literature that finds that microstructure frictions have a large effect on anomaly returns (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). Recent papers in this literature use specialized datasets to study the implementation shortfall of size, B/M, and momentum (Frazzini, Israel, and Moskowitz 2015; Briere, Lehalle, Nefedova, and Raboun 2019). Our paper differs in the size of our anomaly dataset. This huge amount of data allows us to draw conclusions about post-publication net returns, and thus the profits that investors should expect in the future.

Several other papers study the deterioration of anomaly performance over time (Schwert 2003; Marquering, Nisser, and Valla 2006; Huang and Huang 2013; McLean and Pontiff 2016; Jacobs and Müller 2017; Chen and Zimmermann 2018). Of these papers, ours is most closely related to Chordia, Subrahmanyam, and Tong (2014) and Chu, Hirshleifer, and Ma (2017), who also demonstrate that improvements in aggregate liquidity have reduced anomaly returns, and to Huang and Huang (2013), who show that a long-only strategy that optimally picks the best performing published anomaly beats the market after accounting for trading costs. We differ from these papers in that we quantitatively examine returns net of transaction costs for a large set of anomalies. Thus, we show that not only have gross returns declined as liquidity has improved, but that trading costs almost completely eliminate gross returns post-publication for the average anomaly.

2. Anomalies Data and Trading Cost Estimates

Our anomalies are 120 published anomalies from the Chen and Zimmermann (2018) (CZ) dataset that reliably produce decile sorts. This requirement omits many event studies and other discrete variables but ensures that our cost-mitigation techniques can be applied.

Our trading cost measure is the effective bid-ask spread. We measure spreads using high frequency ISSM and TAQ data when available (1983-2016 for NYSE/AMEX and 1987-2016 for NASDAQ). For the earlier sample, we use a simple average of four low frequency proxies: Hasbrouck's (2009) Gibbs estimate (Gibbs); Corwin and Schultz's (2012) high-low spread (HL); Fong, Holden, and Tobek's (2017) implementation of Kyle and Obizhaeva's (2016) volume-over-volatility (VoV); and Abdi and Ranaldo's (2017) close-high-low measure (CHL). These LF proxies build on daily CRSP data and allow us to estimate spreads back to 1926. We post data for our low-frequency average spread and returns gross and net of spreads at http://sites.google.com/site/chenandrewy.

2.1. Anomalies Data

Our anomalies dataset is created from Chen and Zimmermann's (2018) (CZ's) set of 156 cross-sectional return predictors from 115 publications in accounting, economics, and finance journals. From this set, we remove 34 predictors that have difficult-to-evaluate trading costs and 2 predictors that are clearly based on risk to arrive at our baseline set of 120 anomalies.

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.

The CZ dataset is a collection of return predictors, and as such is not formally a set of anomalies. Nevertheless, most return predictors in the CZ dataset were tested against the CAPM. We choose to be open in our interpretation of what is an anomaly and include everything except for the Fama and MacBeth (1973) CAPM beta and Kelly and Jiang's (2014) tail risk factor.

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 (2018).

2.2. Trading Cost Estimates

For each portfolio we examine, we construct trading costs by tracking portfolio weights and applying the effective bid-ask spread whenever trading occurs. Specifically, each time a position is entered or exited, we assume half of the effective spread is paid.

This measure of trading costs aims for relevance and simplicity. Indeed, it leads to a simple interpretation: our costs are the lower bound on the cost to an average trader who uses market orders. Using the effective spread means that we capture the cost of completed trades, which can be smaller than those implied by quoted spreads (Stoll 2003).

Effective spreads omit shorting costs and price impact. We omit these costs because including them would require taking a strong stand on the trader under consideration. Similarly, traders who use limit orders instead of market orders incur execution risk and adverse selection costs, which would require a dramatically more complicated analysis (see Cont and Kukanov 2017, for example).

We now discuss details of our effective spread measurement.

2.2.1. High Frequency Spread Measurement

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

$$[\text{Effective Spread}]_k = 2|\log(P_k) - \log(M_k)|, \tag{1}$$

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

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.

Combining ISSM, MTAQ, and DTAQ, our HF data provide a mostly continuous history of transactions on the NYSE and AMEX from 1983-2016. Data 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). We discuss how we fill in the gaps in Section 2.2.2.

Construction of the matched BBO quotes and data cleaning follows Holden and Jacobsen (2014) (HJ) closely.[1] In addition to the screens used in HJ, we also delete spreads > 40% at the trade level. DTAQ spreads use HJ's DTAQ code. MTAQ spreads for 1999-2003 use HJ's MTAQ code. For pre-1999 data, we add a 2 second delay to the HJ interpolation-matching algorithm. ISSM spreads use an adapted version of HJ's MTAQ code following the quote screens used in Lou and Shu (2014). For additional details see Appendix A.2.

2.2.2. Low-Frequency Spread Measurement

When HF data is not available, we use LF proxies based on daily CRSP data. Rather than focus on a particular LF proxy, we compute four different LF proxies and use the simple average as our spread. 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

[1] We are grateful to Craig Holden for providing SAS code on his website.

averaging of multiple LF illiquidity proxies is also used in Karnaukh, Ranaldo, and Söderlind (2015), who find that averaging improves on using the constituent proxies alone.

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, Martin-Utrera, Nogales, and Uppal (2017).

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 Söderlind (2015); McLean and Pontiff (2016); Koch, Ruenzi, and Starks (2016); and Chen and Zimmermann (2018).

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:

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

   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 (2). 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.

We compute a LF average if we have at least one LF proxy with data. Each proxy can produce missing values, as each proxy requires multiple firm-day observations within a given firm-month. Firm-day observations may be missing if, for example, the stock did not trade.

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. Regardless, to fill in missing spreads, we match the firm with missing data to the nearest firm with data in the same month using market equity rank and idiosyncratic volatility rank. This data filling procedure follows Novy-Marx and Velikov (2016). For further details, please see Appendix A.3.

2.2.3. Performance and Summary Statistics of Effective Spread Measures

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 the logic works. These panels compare our LF average with HF spreads when they are available. Panel B shows that the LF average has the highest 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.

Figure 1 illustrates how summary statistics for our 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 Ranaldo 2017). Trading costs plummet in the 2000's as electronic trading and decimalization have improved liquidity.

[Figure 1 about here.]

3. How Profitable is the Average Anomaly Publication After Trading Costs?

Having described our trading cost measure and anomaly data, we are now in a position to address the main question of the paper. Do anomaly returns survive trading costs post-publication?

In particular, this section examines the returns net of trading costs for the average anomaly. We examine both published strategies and strategies that mitigate trading costs following Novy-Marx and Velikov (2016). We will see that post-publication net returns are negative before cost optimization, and tiny after cost optimization.

Here, we focus on the average anomaly, as examining unusually strong anomalies introduces selection bias. Section 4 takes a closer look at the best-performing anomalies using a variety of selection bias adjustments.

3.1. Returns of Published Strategies Net of Trading Costs

We begin by examining the net returns of published strategies. Specifically, we examine equal-weighted long-short quintile portfolios. This approach follows the modal portfolio construction in the anomalies literature (McLean and Pontiff 2016). While one can find papers that use value-weighting, the vast majority of papers use equal-weights. Similarly, though the decile sorts are frequently used, many of these papers 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.

We rebalance each anomaly portfolio following the original publication's suggestions when possible. If the original paper uses only regressions or is unclear about rebalancing, we choose the frequency that matches the frequency of the signal updates (for example, annual for annual Compustat variables). For a detailed list of rebalancing frequencies see Appendix A.1.

Table 2 summarizes our main findings. The table shows the average return across 120 anomalies, gross and net of trading costs. Panel A examines the modal published portfoio construction: equal-weighted long-short quintiles.

Trading costs have a massive effect. While the average gross return is an impressive 66 bps per month (8 percent per year) in-sample, net of trading costs the average return is a measly 5 bps per month (0.60 percent per year). The post-publication net return is even worse, at -3 bps per month.

[Table 2 about here.]

To understand why trading costs have such a massive effect, it helps to look at a decomposition provided in Table 2. The net return (column e) is approximately the gross return (column a), minus the product of 2-sided turnover (column c) and the average spread paid (column d).

The average anomaly turns over 15% of its long portfolio and 15% of its short portfolio each month. Adding both legs leads to the 31% turnover seen in column (b). Multiplying this turnover by the average spread paid in-sample of 219 bps (column d) leads to a return reduction of about 61 bps per month, nearly eliminating the gross return of 66 bps.

The destruction of profits is echoed in the post-publication results of Table 2. Liquidity has improved over time, and thus, anomaly trading costs are lower in the post-publication samples. Accordingly, average spreads paid post-publication are roughly 50% lower compared to the in-sample periods (column c). Gross returns have declined by more than 50%, however (column a). As a result, net returns are actually negative post-publication, despite the improved liquidity.

Thus, the average trader would have lost money trading on academic publications. Moreover, our large set of anomalies means that we can be confident that these results are not due to sampling error and will likely persist into the future. The standard error on the -3 bps post-publication net return is just 5 basis points. Estimation error coming from spreads is also small, as post-publication net returns use almost exclusively HF data. 97% of the anomalies we study are published after 1983 (Figure A.2), when the ISSM data begins.

Table 2 also examines the performance of cost-mitigated portfolios (Panels B and C). We will return to these results after we explain our cost mitigation (Section 3.3). There we will see that cost-mitigation significantly improves net returns in-sample, but post-publication net returns are tiny.

3.2. Why Are Trading Costs so Large?

The large impact of trading costs may be surprising, particularly in the post-publication samples. Since decimalization, 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 2, 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 2 about here.]

Anomaly portfolios load up on this right tail. The distribution of spreads paid by our 120 anomaly strategies 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 anomaly strategies 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).

3.3. A Simple and Effective Cost Mitigation

We've seen that the average academic anomaly is wiped out by trading costs. Academic anomaly strategies, however, are not designed to account for trading costs. Can cost mitigation rescue the average anomaly?

In this section, we use value-weighting and/or the buy/hold spread to mitigate trading costs. Stock weights and buy/hold spread parameters are chosen to maximize net returns using in-sample data. Here, we describe the optimized cost mitigation technique. Section 3.4 examines performance post-publication.

Value-weighting mitigates costs by reducing the average effective spread paid. Intuitively, trading micro-cap stocks is costly, and value-weighting ensures that these micro-caps make up a tiny portion of the portfolio.

The buy/hold spread (also known as "banding" or an "(s,S)" rule) mitigates costs by reducing turnover. The buy/hold spread is a trading rule that enters positions when a stock's signal is strong, but only exits the position when the signal is weak enough to justify incurring a transaction cost. For example, a 20/40 buy/hold spread goes long stocks that enter the top 20th percentile of the anomaly signal, but only exits the position if the stock drops below the 40th percentile. This trading rule mimics the optimal portfolio rule in the presence of transaction costs (Magill and Constantinides 1976; Brandt, Santa-Clara, and Valkanov 2009, for example) and similar decision rules are often seen in dynamic models with frictions (Arrow, Harris, and Marschak 1951). Novy-Marx and Velikov (2016) and Novy-Marx and Velikov (Forthcoming) show that the buy/hold spread outperforms other simple cost mitigation strategies such as limiting trading to low-cost stocks and reducing rebalancing frequency.

We optimize over stock weighting and buy/hold spreads in 4 steps:

1. We consider either (a) equal-weighted quintiles, or (b) value-weighted NYSE deciles.

2. Using each selection in step 1, we group anomalies into turnover quartiles using in-sample data.

3. For each turnover group in step 2, we try different lower bounds for the buy/hold spread and measure net returns. For each anomaly, we choose the buy/hold spread lower bound that maximizes the average net return of its turnover group.

4. Using the optimized buy/hold spreads from step 3, for each anomaly we choose the stock weighting that maximizes net returns in-sample.

This approach to cost mitigation is somewhat restricted. We only consider equal-weighted all-stock quintiles and value-weighted NYSE deciles, rather than try all combinations of weighting and sorting parameters. We also restrict the buy/hold spread choice to the turnover group level rather than optimize for each anomaly. This restricted optimization keeps our portfolio constructions transparent and limits overfitting.

Table 3 illustrates steps 1-3 of the 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.

[Table 3 about here.]

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 3. This last step of the optimization is done at the anomaly level, and is not shown in the table.

3.4. Cost Mitigated Net Returns

Figures 3 and 4 show that our cost mitigation is effective in-sample. The figures show the distribution of in-sample net returns before (Figure 3) and after (4) 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 3 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.

[Figure 3 about here.]

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 3, value-weighting combined with a buy/hold spread produces positive net returns even among anomalies in the highest turnover quartile.

Indeed, Figure 4 shows that our cost-mitigation is quite effective in-sample. The long left tail of net returns from Figure 3 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 4. 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 3).

[Figure 4 about here.]

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 4 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 4 likely cannot earn the in-sample profits, as the strategies were developed toward the end of the in-sample period.

Instead, the profits available to investors are largely post-publication net returns, which are examined in Figure 5. The figure examines the net returns of cost-mitigated anomalies by month since publication. The light line takes the simple average across all anomalies within each month. The extreme volatility of the light line is a reminder that anomalies portfolios are not at all sure bets, even after averaging across 120 anomalies.

[Figure 5 about here.]

The dark line shows the trailing 5-year moving average net return, once again averaging across 120 anomalies. The 5-year moving average shows a sharp decay in performance that occurs around publication time. Net returns drop from about 40 bps per month 5 years before publication to around 10 bps per month within a couple years after publication. Net returns improve slightly between 5 and 10 years after publication, as if traders on average forget about the anomaly, but they trend downward afterwards, reaching close to zero within 20 years.

Averaging across months first and then averaging across anomalies leads to a meager net return of 13 bps per month, as seen Panel B of Table 2. This pooled average focuses on the first 10 years after publication, however, as most of our anomalies were published relatively recently (Appendix A.4). Averaging across anomalies within each month after publication and then averaging across months results in a tiny, statistically insignificant net return of 4 bps per month.

Table 2 provides a more detailed look at how cost-mitigation works in-sample but largely fails post-publication. Panel B shows that cost-mitigation reduces both turnover by about 30% and spreads paid by almost 50% in the in-sample period, while reducing gross-returns by only about 10%. Combining these effects leads to the 33 bps improvement in-sample.

Post-publication cost-mitigation is similarly effective, with about a 30% improvement in turnover and a more than 50% improvement in spreads paid. But gross-returns decline precipitously for cost-mitigated strategies post-publication, dropping 66%, compared to 55% for unmitigated strategies.

Using post-publication samples accounts explicitly for information effects, but does not directly account for the modern era of high liquidity. As the high liquidity era is likely to remain in the future, traders may be interested in examining net returns both post-publication and post-2005, when the high liquidity era began. The bottom row of Panel B examines this subsample.[2] Controlling for both information effects and liquidity, the mean cost-mitigated net return is a tiny 8 bps per month, even smaller than the 13 bps per month seen in the post-publication sample without excluding pre-2005 data.

Overall, these results are consistent with the idea that arbitrageurs act on academic publications. Indeed, they appear to concentrate their efforts on cost-mitigated strategies.

3.5. Cost-Mitigated, Only Value-Weighting

Even the small post-publication net returns implied by our cost-mitigated portfolios may be unachievable, as they require equal-weighting for 40% of anomalies (Figure 4). Our trading costs omit price impact and short sale fees, and these additional costs are likely to be larger in small stocks. Indeed, Cohen, Diether, and Malloy (2007) finds that short sale fees for stocks with below NYSE-median market value have short sale fees of 33 bps per month, compared to just 3 bps per month for above-median stocks.

These additional costs are more limited in cost-mitigated value-weighted portfolios, which we examine in Figure 6. These portfolios apply buy/hold spreads following Table 3, but use only value-weighting.

Figure 5 shows that post-publication decay is even more pronounced in value-weighted portfolios. Net returns drop from around 40 bps per month in the 5 years before publication to 0 within 3 years of publication. Once again, net returns rise about 5-10 years after publication, but trend down afterwards, reaching zero at around 13 years post-publications.

[Figure 6 about here.]

Averaging across time and then across anomalies, post-publication net returns are a tiny 7 bps per month (Table 2, Panel C). This pitiful post-publication

[2] We thank Marie Briere for this suggestion.

profit comes despite the fact that net returns in-sample are a respectable 30 bps per month. As in Section 3.4, post-publication profits are tiny because gross returns drop precipitously. For these cost mitigated portfolios that use only value-weighting, this decay is 74%, suggesting once again that arbitrageurs are attracted to variations of academic strategies that have reduced trading costs. Indeed, post-publication and post-2005 (bottom row of Panel C), the mean net return is close to one standard error from zero.

These results are robust to the exclusion of the 9 anomalies that have negative net returns in-sample. Excluding these 9 anomalies increases the average post-publication net return by only 3 bps per month weighing anomalies equally, and has little effect on the average weighting months since publication equally. Section 4.1 takes a more detailed look at predicting post-publication net returns using in-sample data.

Overall, the average anomaly publication offers economically insignificant profits net of trading costs for the average trader. Even though cost mitigation can rescue the average anomaly in-sample, the average post publication net return is at most 13 bps per month, and is even smaller in value-weighted portfolios. Most of the remaining profits are likely claimed by short sale fees, which are omitted from our simple estimate of trading costs.

4. How Profitable are the Best Anomaly Publications after Trading Costs?

Up till now, we have focused on the average anomaly's post-publication net return. Averaging across all anomalies keeps the analysis simple, but ignores heterogeneity across anomalies.

In this section, we examine whether anomalies are truly heterogeneous in post-publication net returns. To answer this question, one cannot simply average the right tail of net returns in the post-publication distribution. Such an average would be polluted by selection bias, and this bias would be especially large in the short post-publication samples.

To understand this selection bias, it helps to plot the distribution of post-publication net returns. Figure 7 plots this distribution, and shows that net returns form a nice bell shape, centered around the average anomaly return of 13 bps per month.

[Figure 7 about here.]

Figure 7 shows that some anomalies have large net returns post-publication. Size, B/M, and momentum are among the better performers, consistent with Frazzini, Israel, and Moskowitz's (2015) and Briere, Lehalle, Nefedova, and Raboun's (2019) finding that size, B/M, and momentum perform well after trading costs in recent years. Other anomalies have performed well too: asset tangibility (Tangibili), Gross Profitability (ProfGross), and net external financing (ExtFinNet) all produce net returns in excess of 60 bps per month.

It's not clear that these large net returns are due to true predictability, however. We've been using the term "net returns" to refer to *sample* mean net returns. Sample means include sampling variation, and this variation is not random when we focus on the right tail of the distribution. Indeed, the right tail will tend to include not only anomalies with large true returns, but also anomalies with unusually positive sampling error.

This selection bias is especially pronounced for anomalies with short post-publication samples. For example, cash to assets (Cash) and gross profitability (ProfGross) were published in 2012 and 2013, respectively, and we have only 3-4 years of post-publication returns as our data ends in 2016. Thus, it's hard to say whether the large net returns are due to pure chance. As these anomalies are in the right tail, it is likely that sampling variation has a positive effect on the sample mean return.

Indeed, Figure 7 resembles what would be generated by the null hypothesis with no predictability. Only a 12.5% of anomalies have t-stats that exceed 2.0 (bold), compared to 4.6% under the null. Similarly, 73% have t-stats less than 1.5, compared to 87% under the null. In other words, much of the tail of this distribution can be attributed to noise.

We use two methods to separate true net returns from noise. Section 4.1 examines whether post-publication net returns can be predicted using in-sample data. The amount of predictability, then, determines the amount of heterogeneity in true net returns. Section 4.2 explicitly separates true returns from noise by estimating a simple model that accounts for selection bias.

Though the methods are different, they lead to very similar results: allowing for equal-weighted strategies, one can hope to find at most 15-20 bps per month of true net returns post-publication. Value-weighted strategies provide at most 7-15 bps per month.

4.1. Predicting Post-Publication Net Returns with In-Sample Net Returns

We use in-sample turnover and in-sample net returns to predict post-publication net returns. Both of these anomaly properties should theoretically predict net returns. Sections 3.3 and 3.4 showed that turnover has a massive impact on net returns, and that some anomalies have negative in-sample net returns even after cost mitigation. High turnover anomalies may also have high net returns, however. For example, momentum anomalies and idiosyncratic volatility both have high turnover but large net returns once trading costs are mitigated.

Table 4 shows the predictability results. The table shows the mean post-publication net return of anomalies grouped by predictor quartiles. Panel A uses cost-mitigated anomaly strategies that include equal weighting, while panel B uses cost-mitigated strategies that use only value-weighting.

[Table 4 about here]

The table shows that predictability is relatively minor. Turnover is the stronger predictor. Including equal-weighted strategies (Panel A), the best (lowest) turnover quartiles net about 20 bps per month post-publication. These net returns are highly statistically significant, but economically modest.

Moreover, net returns are notably smaller using value-weighting (Panel B). Restricting the portfolios to value-weighting, the best two turnover quartiles net 15 and 8 bps per month. It may be hard to determine whether an anomaly is in the best or second best turnover quartile in real time, however. Thus, the net return that can be actually obtained is likely between 8 and 15 bps per month.

Table 4 shows that in-sample net returns are a fairly poor predictor of post-publication net returns. The relationship is non-monotonic in strategies that include equal-weighting and strategies that use value-weighting only. Indeed, predictability is essentially non-existent using value-weighting, with the best and worst quartiles differing by just a single basis point per month.

Overall, post-publication returns are not very predictable. Using in-sample data, one can obtain at most 20 bps per month of net returns post-publication, and about 10-15 bps per month if one is restricted to value-weighting. These results suggest that the right tail in post-publication net returns (Figure 7) is largely due to noise, and net returns are modest, even among the best anomalies and after cost mitigation.

4.2. Large Net Returns Adjusted for Selection Bias

Another way to examine whether any anomalies survive is to explicitly adjust for selection bias in the estimation of mean returns. This section develops an "empirical Bayesian" adjustment and applies it to anomalies with especially large post-publication net returns.

4.2.1. An Empirical Bayesian Selection Bias Adjustment

Our bias adjustment is derived from a statistical model of post-publication net returns. Predictor i's post-publication sample mean return $\bar{r}_i$ is a noisy signal of its true return μ_i

$$\bar{r}_i = \mu_i + \epsilon_i \tag{3}$$

$$\epsilon_i \sim N(0, \mathrm{SE}_i), \tag{4}$$

where ϵ_i is noise and SE_i is the standard error of $\bar{r}_i$. The idea that the noise is uncorrelated is consistent with the near-zero mean and median time-series correlation between pairs of anomaly returns (Green, Hand, and Zhang 2013, McLean and Pontiff 2016, Chen and Zimmermann 2018). True returns are also normally distributed

$$\mu_i \sim N(\mu_\mu, \sigma_\mu) \tag{5}$$

where μ_μ is the mean of true returns and σ_μ is the standard deviation of true returns.

This simple model helps illustrate the hazards of focusing on large post-publication returns. Suppose we want to look at a portfolio i which has a post-publication sample mean return that is equal to the 90th percentile post-publication sample mean return $\bar{r}_{90\text{th}}$. This selection process implies that $\bar{r}_i$ is upward biased, as it leads to large ϵ_i:

$$\mathbb{E}(\bar{r}_i|\bar{r}_i = \bar{r}_{90\text{th}}) = \mathbb{E}(\mu_i|\bar{r}_i = \bar{r}_{90\text{th}}) + \underbrace{\mathbb{E}(\epsilon_i|\bar{r}_i = \bar{r}_{90\text{th}})}_{>0}. \tag{6}$$

To control for luck, we want to explicitly condition on selection and keep only the $\mathbb{E}(\mu_i|\bar{r}_i > \bar{r}_{90\text{th}})$ term:

$$\hat{\mu}_{90} = \mathbb{E}(\mu_i|\bar{r}_i = \bar{r}_{90\text{th}}) \tag{7}$$

We calculate similar conditional expectations by first estimating μ_μ and σ_μ, and then applying standard probability theory formulas.[3]

We estimate μ_μ and σ_μ using method of moments. μ_μ is estimated by averaging across all $\bar{r}_i$.

$$\hat{\mu}_\mu = \frac{1}{N}\sum_i \bar{r}_i \tag{10}$$

σ_μ can be estimated by comparing the variance across $\bar{r}_i$ to the average standard error SE_i:[4]

$$\hat{\sigma}^2_\mu = \max\left\{\left[\frac{1}{N}\sum_i(\bar{r}_i - \hat{\mu}_\mu)^2 - \frac{1}{N}\sum_i \text{SE}_i^2\right], \quad 0 \right\}, \tag{11}$$

where the max operator ensures non-negativity of $\hat{\sigma}_\mu$, and can be justified using expected loss function arguments (Efron and Morris 1973).

This approach to adjusting for selection bias is often described as "empirical Bayesian." We calculate $\mathbb{E}(\mu_i|\bar{r}_i > \bar{r}_{90\text{th}})$ using Bayesian formulas, but then use frequentist methods to estimate the "priors" μ_μ and σ_μ. Similar methods

[3]Normal-normal updating formulas give

$$\mathbb{E}(\mu_i|\bar{r}_i, \text{SE}_i, \mu_\mu, \sigma_\mu) = s_i\mu_\mu + (1-s)\bar{r}_i \tag{8}$$

where the shrinkage s_i is

$$s_i = \frac{\text{SE}_i^2}{\sigma_\mu^2 + \text{SE}_i^2}. \tag{9}$$

And then the properties of the normal distribution give the 90th percentile.

[4]To see this, note that

$$(\bar{r}_i - \mu_\mu)^2 = (\mu_i - \mu_\mu)^2 + (\mu_i - \mu_\mu)\epsilon_i + \epsilon_i^2.$$

Then taking expectations removes the cross term $(\mu_i - \mu_\mu)\epsilon_i$.

are often used to control for luck in statistics (Efron 2012). These methods have recently been applied in economics to large scale A/B testing (Azevedo, Deng, Montiel Olea, and Weyl 2019; Azevedo et al. 2019) and forecasting revenues of a large panel of banks (Liu, Moon, and Schorfheide Forthcoming).

4.2.2. Bias Adjusted Post-Publication Net Returns

Table 5 describes the estimation results and bias adjusted returns for the best-performing anomalies. The top panel illustrates the estimation of the model of post-publication returns (Equations (3)-(5)).

[Table 5 about here.]

Using cost-mitigated portfolios that allow for equal-weighting (top row), the cross-predictor variance of returns (var $\bar{r}_i$) is very close to the mean squared standard error (mean SE_i^2). This implies that the variance of true returns $\hat{\sigma}_\mu^2$ is very small (Equation (11)). In other words, noise can account for the vast majority of the dispersion in post-publication returns across predictors. This result holds regardless of whether one allows equal-weighting (top row) or uses only value-weighted portfolios. Indeed, using value-weighting implies that there is no dispersion at all in true returns.

These results are intuitive. The net returns in Figure 7 are mostly within 35 bps of the center of the distribution. The typical post-publication sample of 13 years and typical monthly volatility of around 400 bps implies a standard error of about $400/\sqrt{13 \times 12} \approx 32$ bps. These two measurements of dispersion are very close to another, allowing little room for variation in true returns. Informally, Figure 7 has just about as many large t-stats as one would expect by pure chance.

As there is little variation in true returns, the adjusted returns are close to the mean across anomalies. This is seen in the bottom panel of Table 5, which shows the top percentiles of adjusted returns $\hat{\mu}$. The 90th and 95th percentile anomalies have bias-adjusted net returns of only about 20 bps per month. Limiting strategies to value weighting, the estimation implies no variation in true returns $\hat{\sigma}_\mu = 0$ and thus even the 95th percentile true return is equal to the mean across anomalies. Both of these results are quantitatively consistent with those that come from predicting post-publication net returns (Table 4).

Taken with Section 4.1, we can be quite confident that there is little true dispersion in performance post-publication. Anomaly publications offer little in the way of returns net of trading costs for the average trader, especially if one focuses on value-weighted portfolios.

5. Conclusion

We show that anomalies largely provide the illusion of profits. For the average trader, publications reveal only profits that were already taken out of the market. Post-publication, the average anomaly return is negligible net of trading costs. Even the strongest anomaly's net returns are modest at best. Moreover, the modest net returns we manage to uncover are fragile, relying on equal-weighting and decaying toward zero as time passes.

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 (2018). 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 squred	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	Jegedeesh	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

We use Holden and Jacobsen's (2014) (HJ's) code to calculate effective spreads. DTAQ spreads use 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, F, 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 ≤ 0 or missing
- BID ≤ 0 or missing
- ofrsize ≤ 0 or missing
- bidsize ≤ 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 ≤ 0 or missing
- BID ≤ 0 or missing
- ofrsiz ≤ 0 or missing
- bidsiz ≤ 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 bidsize 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.

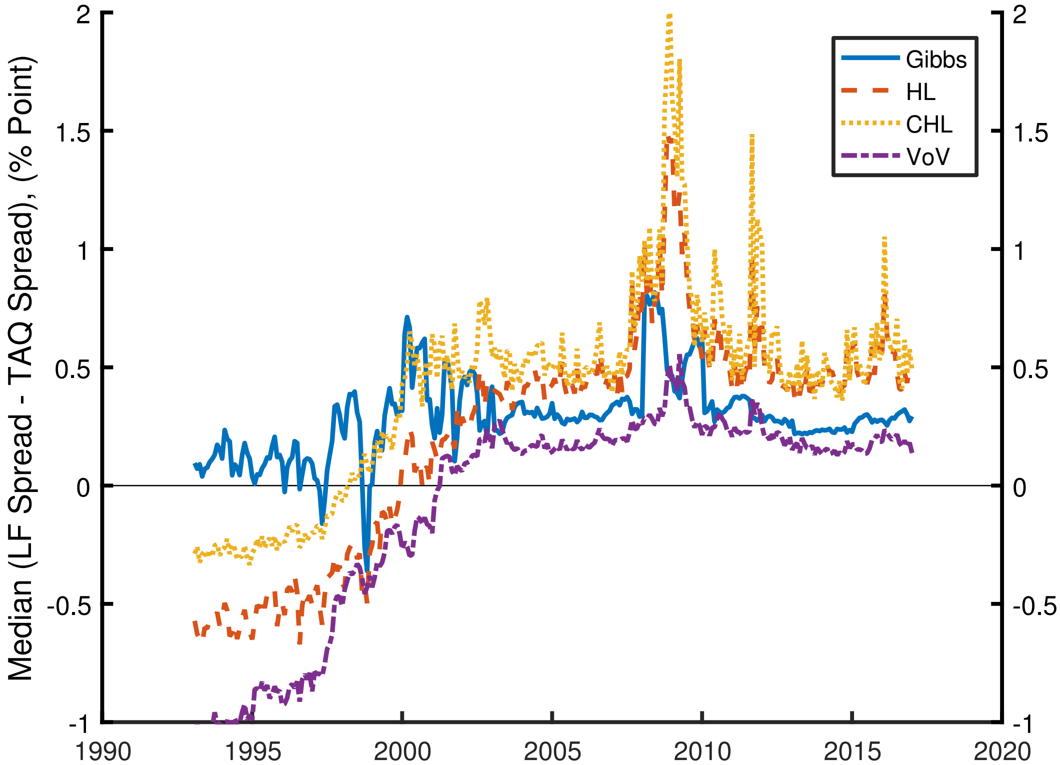
A.3. Details of Low Frequency Measures

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.

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.

Figure A.1: Low-Frequency Spread Proxy Errors Over Time. We subtract low-frequency (LF) effective spread proxies from TAQ effective spreads at the firm-month level to calculate an error. We then compute the median error across firms within each month. Definitions of the LF proxies are found in Section 2.2.2. Post-decimalization, LF proxies are upward biased by roughly 25-50 bps.



A.4. Additional Results

Figure A.2: Distribution of Publication Years.

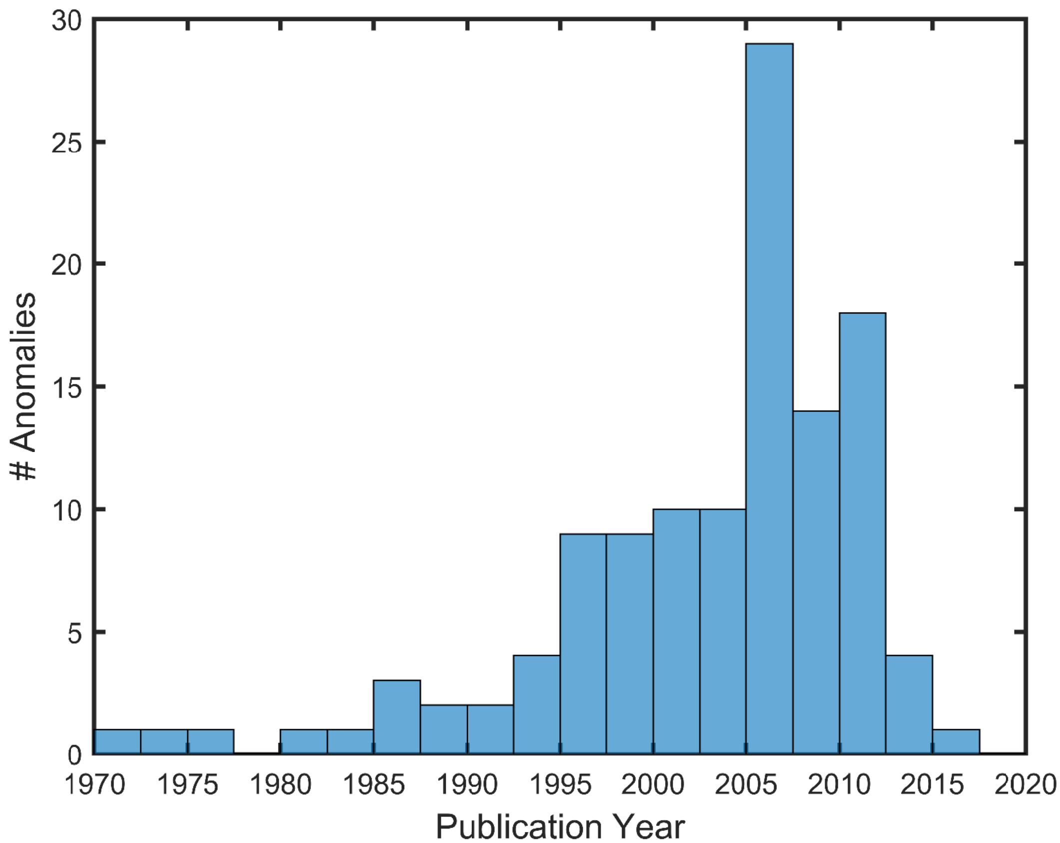


Figure A.3: Distribution of Post-Publication Sample Lengths.

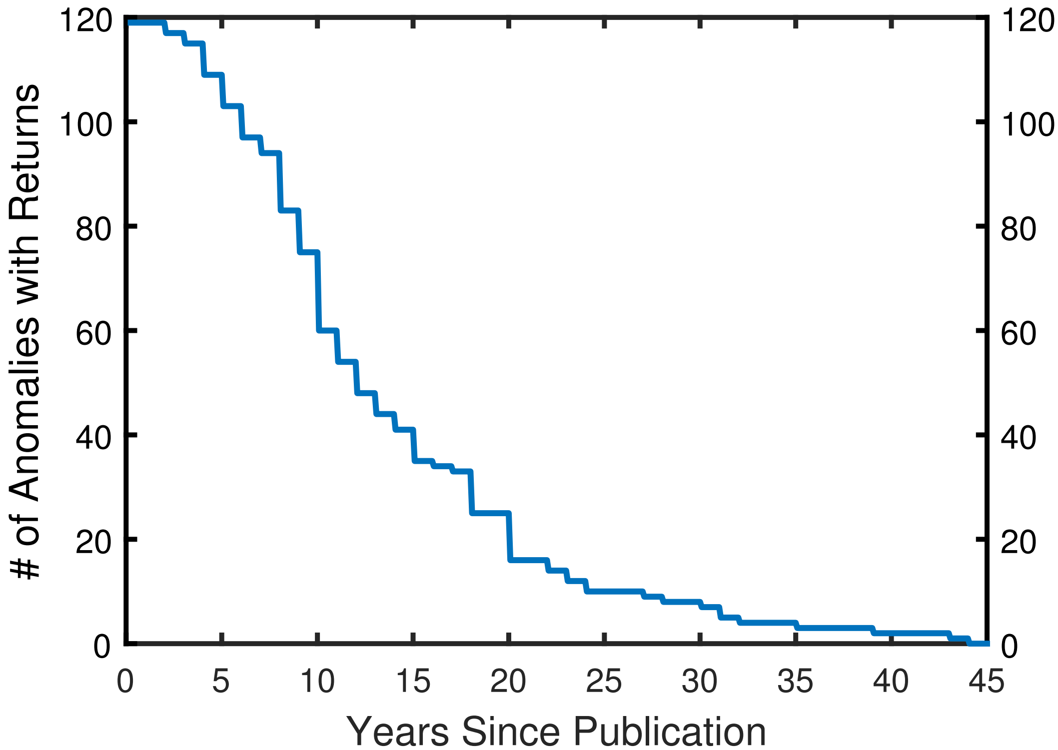


Table A.4: Returns Gross and Net of Trading Costs: Post-Pub and Post-2005

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) ≈ (b) × (c) Return Reduction	(e) = (a) - (d) Net Return
Panel A: Equal-Weighted Long-Short Quintiles					
Post-Pub & Post-2005	0.30	0.30	1.11	0.32	-0.03
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Panel B: Cost-Mitigated using Value-Weighting and Buy/Hold Spreads					
Post-Pub & Post-2005	0.20	0.20	0.60	0.08	0.13
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Panel C: Cost-Mitigated using Buy/Hold Spreads, Value-Weighted only					
Post-Pub & Post-2005	0.12	0.19	0.31	0.05	0.07
	(0.10)	(0.10)	(0.10)	(0.10)	(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.

Arrow, Kenneth J, Theodore Harris, and Jacob Marschak. “Optimal inventory policy”. *Econometrica: Journal of the Econometric Society* (1951), pp. 250–272.

Azevedo, Eduardo M, Deng Alex, Jose Montiel Olea, Justin M Rao, and E Glen Weyl. “A/B Testing with Fat Tails”. *Available at SSRN 3171224* (2019).

Azevedo, Eduardo M, Alex Deng, José L Montiel Olea, and E Glen Weyl. “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.

Bansal, Ravi and Amir Yaron. “Risks for the long run: A potential resolution of asset pricing puzzles”. *The Journal of Finance* 59.4 (2004), pp. 1481–1509.

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, Charles-Albert Lehalle, Tamara Nefedova, and Amine Raboun. “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”. *Available at SSRN: https://ssrn.com/abstract=2802357* (2018).

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.

Cochrane, John H. "Portfolio Advice for a Multi-Factor World". *Economic Perspectives, Federal Reserve Bank of Chicago.* Vol. 23. Citeseer. 1999, pp. 59–78.

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.

DeMiguel, Victor, Alberto Martin-Utrera, Francisco J Nogales, and Raman Uppal. "A portfolio perspective on the multitude of firm characteristics" (2017).

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 and Carl Morris. "Stein's estimation rule and its competitors-an empirical Bayes approach". *Journal of the American Statistical Association* 68.341 (1973), pp. 117–130.

Fama, Eugene F and James D MacBeth. "Risk, return, and equilibrium: Empirical tests". *The Journal of Political Economy* (1973), pp. 607–636.

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 of asset pricing anomalies" (2015).

Gârleanu, Nicolae and Lasse Heje Pedersen. "Efficiently inefficient markets for assets and asset management". *The Journal of Finance* 73.4 (2018), pp. 1663–1712.

Gompers, Paul, Joy Ishii, and Andrew Metrick. "Corporate governance and equity prices". *The quarterly journal of economics* 118.1 (2003), pp. 107–156.

Green, Jeremiah, John RM Hand, and X Frank Zhang. "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.

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).

Huang, Jing-Zhi and Zhijian James Huang. "Real-Time Profitability of Published Anomalies: An Out-of-Sample Test". *Quarterly Journal of Finance* 3.3n4 (2013).

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

Karnaukh, Nina, Angelo Ranaldo, and Paul Söderlind. "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.

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, Laura, Hyungsik Roger Moon, and Frank Schorfheide. “Forecasting with Dynamic Panel Data Models”. *Econometrica: Journal of the Econometric Society* (Forthcoming).

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.

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* (Forthcoming).

Patton, Andrew J and Brian M Weller. “What you see is not what you get: The costs of trading market anomalies” (2017).

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.

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.

Tables and Figures

Table 1: Correlations Between Effective Spread Measures

Correlations are pooled. We examine four low frequency measures that use daily CRSP data: 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 Ranaldo'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 measures. 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.

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: Returns Gross and Net of Trading Costs

We adjust 120 anomaly portfolio returns for effective bid-ask spreads. Figures average across months and then across anomalies, with standard errors in parentheses. Anomalies come from the Chen and Zimmermann (2018) dataset (Table A.1-A.3). Spreads combine data from ISSM, TAQ, and four low-frequency proxies (Section 2.2.1-2.2.2). Columns (a)-(d) report an approximate net return decomposition. Panel A uses the equal-weighted long-short quintiles (the modal published portfolio construction, Section 3.1). Panel B examines cost-mitigation which optimizes over equal- vs value-weighting and the buy-hold spread lower bound (Section 3.4). Panel C examines cost-mitigation using buy-hold spreads and permitting only value-weighting (Section 3.5). All figures are in percent per month except for turnover, which is a ratio per month.

	(a) Gross Return	(b) Turnover (2-sided)	(c) Ave Spread Paid	(d) ≈ (b) × (c) Return Reduction	(e) = (a) - (d) Net Return
Panel A: Equal-Weighted Long-Short Quintiles					
In-Sample	0.66	0.31	2.19	0.61	0.05
	(0.04)	(0.04)	(0.06)	(0.07)	(0.06)
Post-Publication	0.30	0.30	1.11	0.32	-0.03
	(0.04)	(0.04)	(0.06)	(0.05)	(0.05)
Panel B: Cost-Mitigated					
In-Sample	0.59	0.20	1.36	0.21	0.38
	(0.04)	(0.02)	(0.07)	(0.02)	(0.03)
Post-Publication	0.20	0.20	0.60	0.08	0.13
	(0.04)	(0.02)	(0.06)	(0.01)	(0.04)
Post-Pub & Post-2005	0.14	0.20	0.46	0.06	0.08
	(0.04)	(0.02)	(0.04)	(0.01)	(0.04)
Panel C: Cost-Mitigated, Value-Weighted only					
In-Sample	0.46	0.20	0.86	0.16	0.30
	(0.04)	(0.02)	(0.05)	(0.02)	(0.03)
Post-Publication	0.12	0.19	0.31	0.05	0.07
	(0.03)	(0.02)	(0.05)	(0.01)	(0.03)
Post-Pub & Post-2005	0.07	0.19	0.21	0.03	0.04
	(0.03)	(0.02)	(0.03)	(0.00)	(0.03)

Table 3: 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
	Q1	**0.39**	0.39	0.38	0.37	0.36	0.34	0.33
Turnover	Q2	0.31	**0.32**	0.31	0.31	0.30	0.29	0.28
Quartile	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
	Q1	**0.33**	0.28	0.26	0.24	0.23
Turnover	Q2	**0.34**	0.32	0.30	0.26	0.22
Quartile	Q3	0.16	**0.23**	0.22	0.19	0.19
	Q4	0.07	0.23	0.28	0.31	**0.32**

Table 4: Predicting Post-Publication Net Returns with In-Sample Data

The table shows the average post-publication net returns (% per month) of anomalies within quartiles of in-sample turnover and in-sample net return. Quartiles are numbered from worst expected net return a priori—that is, quartile 1has the highest turnover and lowest expected net return, and quartile 4 has the lowest turnover and highest expected net return. All portfolios use cost-mitigation following Table 3. Panel A includes equal-weighted portfolios if they perform better in-sample. Panel B is restricted to value-weighting. The best anomalies return only about 10-20 bps net of trading costs.

Panel A: Including Equal-Weighting

In-Sample Predictor	Predictor Quartile 1 (Worst)	2	3	4 (Best)
Turnover	-0.18	-0.01	0.18	0.21
	(0.06)	(0.05)	(0.05)	(0.05)
Net Return	0.00	0.10	0.06	0.13
	(0.04)	(0.05)	(0.06)	(0.05)

Panel B: Value-Weighted Only

In-Sample Predictor	Predictor Quartile 1 (Worst)	2	3	4 (Best)
Turnover	0.06	0.00	0.08	0.15
	(0.08)	(0.08)	(0.07)	(0.07)
Net Return	0.10	0.10	0.04	0.11
	(0.07)	(0.08)	(0.07)	(0.08)

Table 5: Large Net Returns Post Publication Adjusted for Selection Bias

We use empirical Bayesian methods to adjust large post-publication sample mean net returns for luck. $\bar{r}_i$ is the post-publication sample mean net return of predictor i. SE_i is the standard error of $\bar{r}_i$. $\hat{\sigma}^2_\mu$ is the variance of unobserved true returns. $\hat{\mu}$ is the selection bias adjusted return using Equation (7). All portfolios use cost-mitigation following Table 3. "Including EW" includes equal-weighted portfolios if it improves net returns in-sample. "VW only" is restricted to value-weighting.

Parameter Estimation				
	mean $\bar{r}_i$	var $\bar{r}_i$	mean σ^2_i	$\hat{\sigma}^2_\mu$
Including EW	0.127	0.132	0.120	0.013
VW only	0.069	0.139	0.160	0.000

Bias-Adjusted Net Returns Post Publication				
	$\hat{\mu}$ (%, monthly)			
	50 pct	75 pct	90 pct	95 pct
Including EW	0.13	0.16	0.19	0.21
VW only	0.07	0.07	0.07	0.07

Figure 1: Summary Statistics for Effective Spreads Over Time. Spreads use high-frequency data from Daily TAQ (DTAQ), Monthly TAQ (MTAQ), and ISSM when available. When high-frequency data is not available, we use the average of four low frequency (LF) proxies: Gibbs (Hasbrouck 2009), HL (Corwin and Schultz 2012), CHL (Abdi and Ranaldo 2017), and VoV (Kyle and Obizhaeva 2016).

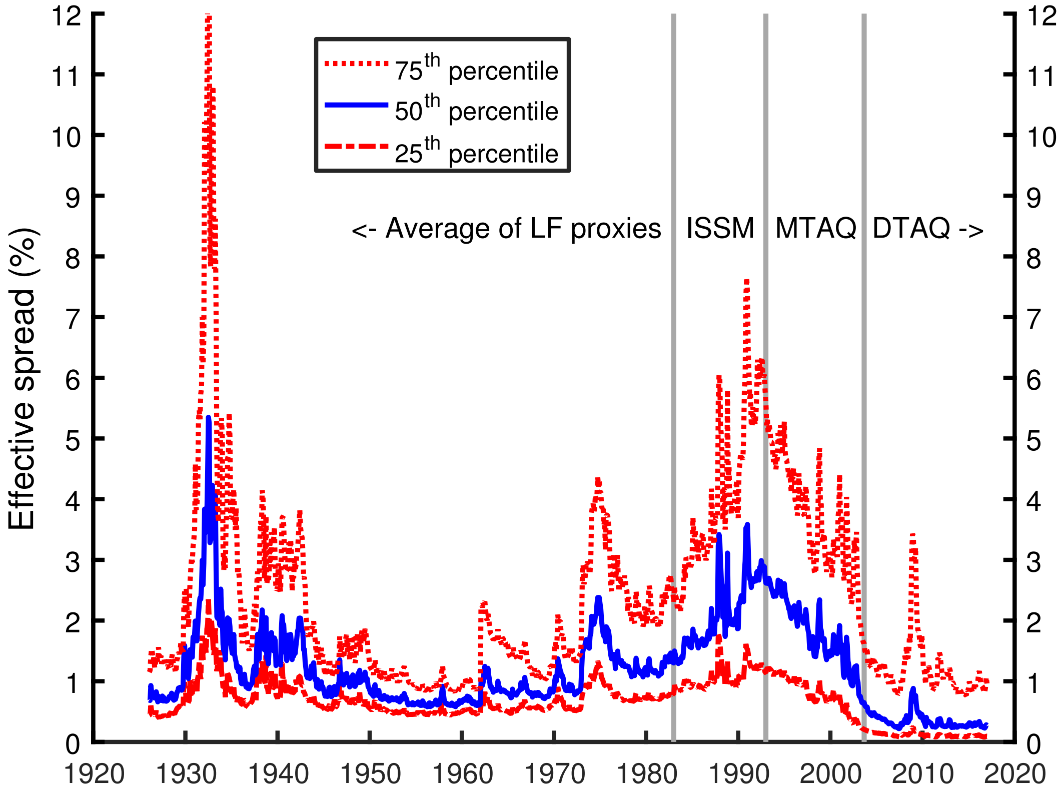


Figure 2: Distribution of Effective Spreads in 2014. This figure compares the distribution of effective spreads that are paid by anomaly portfolios with those across all stocks, NYSE stocks, and Russell 2000 stocks. "Paid by anomaly portfolios" spreads are pooled across all trades implied by 120 anomaly portfolios in 2014. The stock distributions are pooled across all stock-months in 2014. Anomaly portfolios trade stocks across the entire liquidity spectrum, including the long right tail.

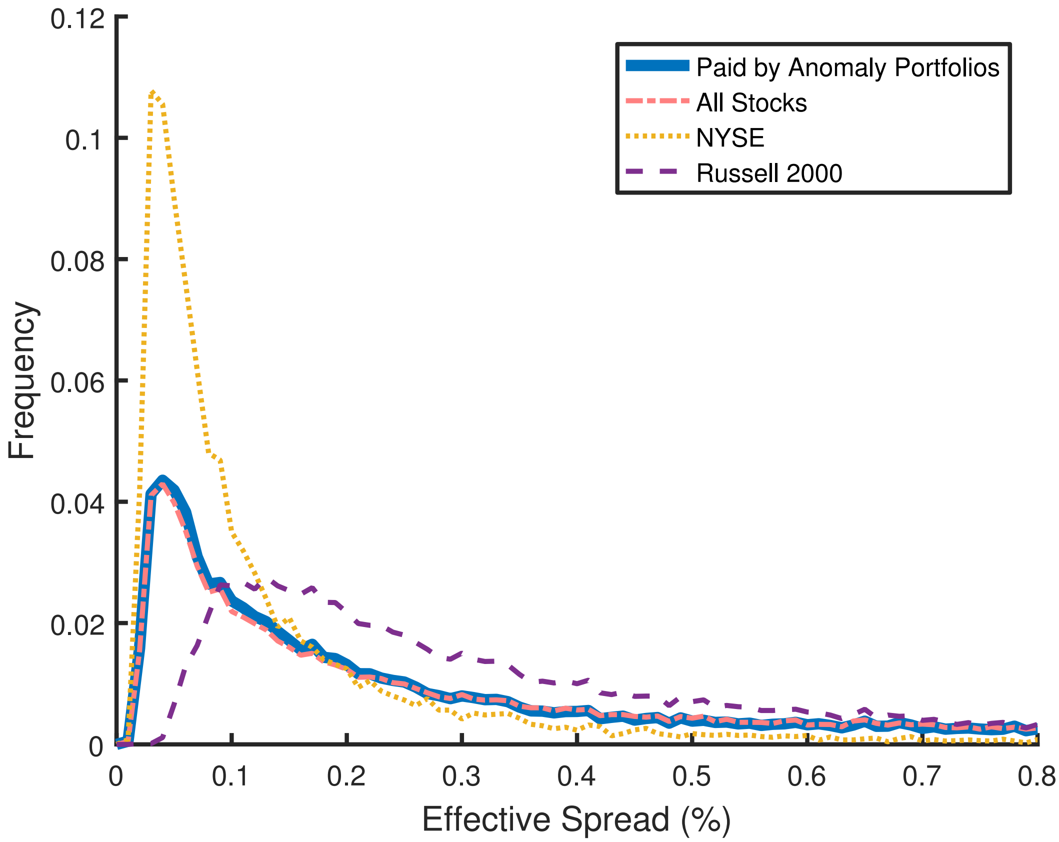


Figure 3: Distribution of Net Returns: In-Sample, Before Cost-Mitigation We adjust anomaly returns for effective bid-ask spreads (Figure 1). All portfolios use equal-weighted quintile sorts, following the modal approach in the literature. Anomalies with above median turnover (15% per month, two-sided) 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.

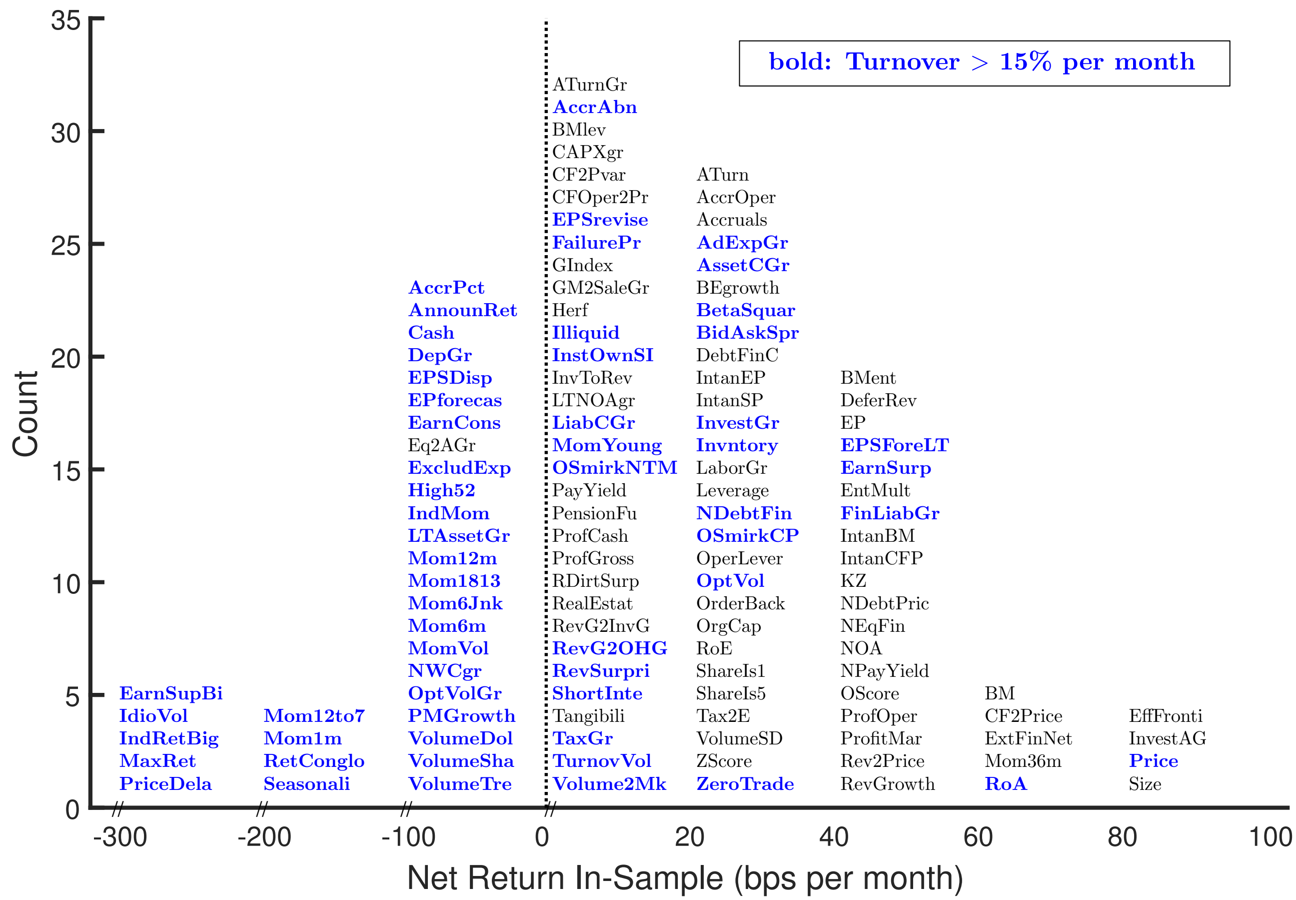


Figure 4: Cost-Mitigation Results: Distribution of Net Returns: In-Sample. We mitigate transaction costs by applying value-weighting and/or buy/hold spreads to 120 anomaly portfolios. Buy/hold spreads are chosen to maximize net returns in-sample following Table 3. Stock weighting is chosen to maximize 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.

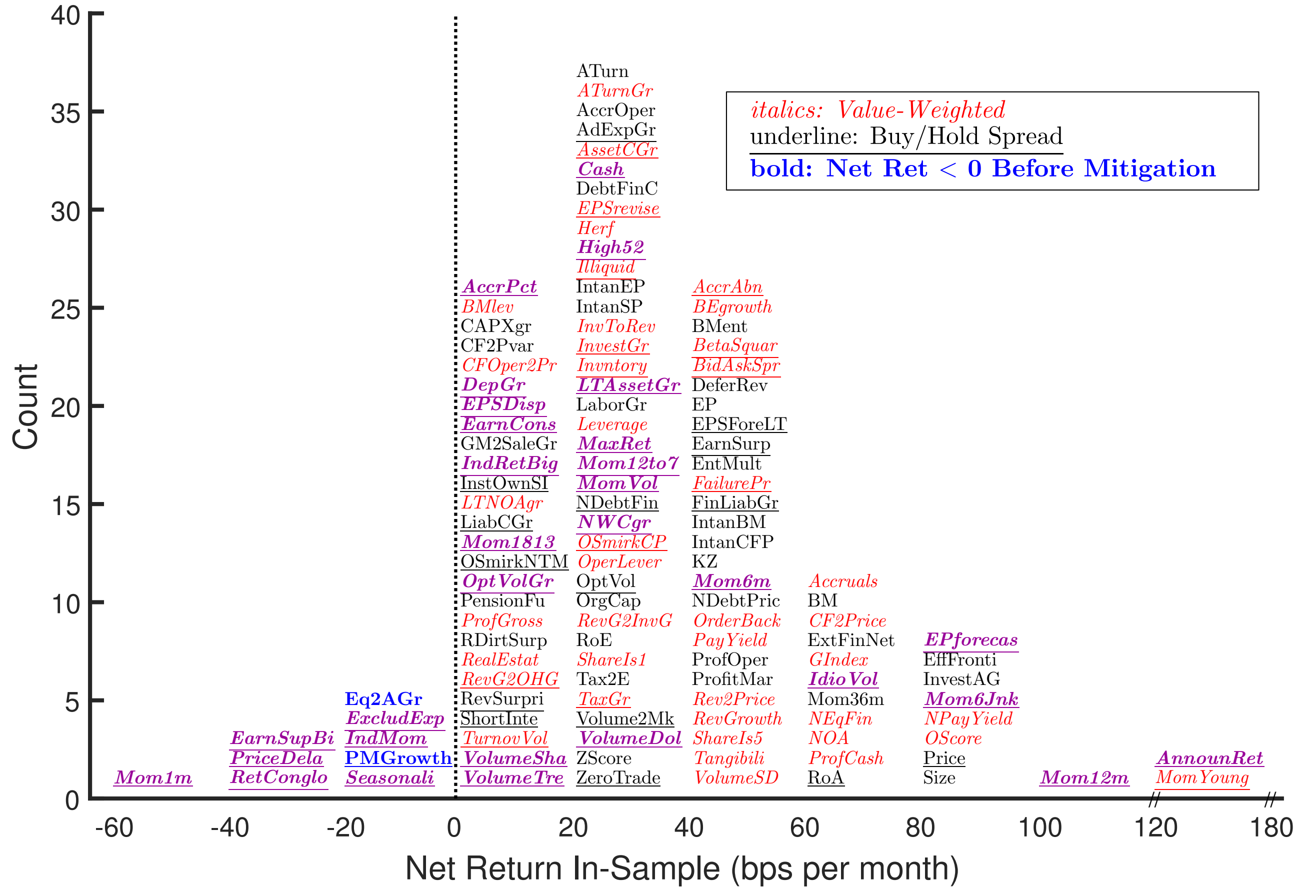


Figure 5: Performance Decay of Cost-Mitigated Anomaly Portfolios. This figure plots the net return by month since publication for portfolios that use cost-mitigation following Table 3. For a given month relative to publication, light lines plot the mean return across all anomalies. Dark lines show the trailing 5-year moving average of mean returns, and dashed lines show 2 standard error confidence bounds. Positive net returns in-sample become small soon after publication, and trend toward zero afterwards.

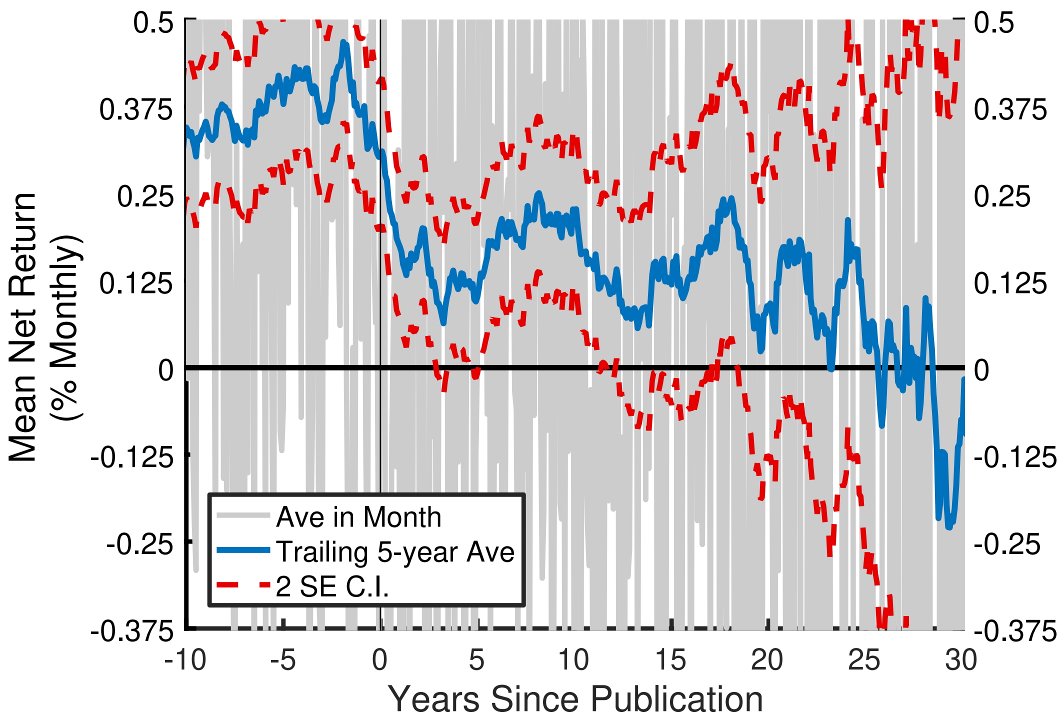


Figure 6: Performance Decay of Cost-Mitigated Anomaly Portfolios: Value-Weighted Only. This figure plots the net return by month since publication for portfolios that use cost-mitigation following Table 3 using only value-weighting. For a given month relative to publication, light lines plot the mean 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. Post-publication performance is even worse than if one allows for equal-weighting (Figure 5).

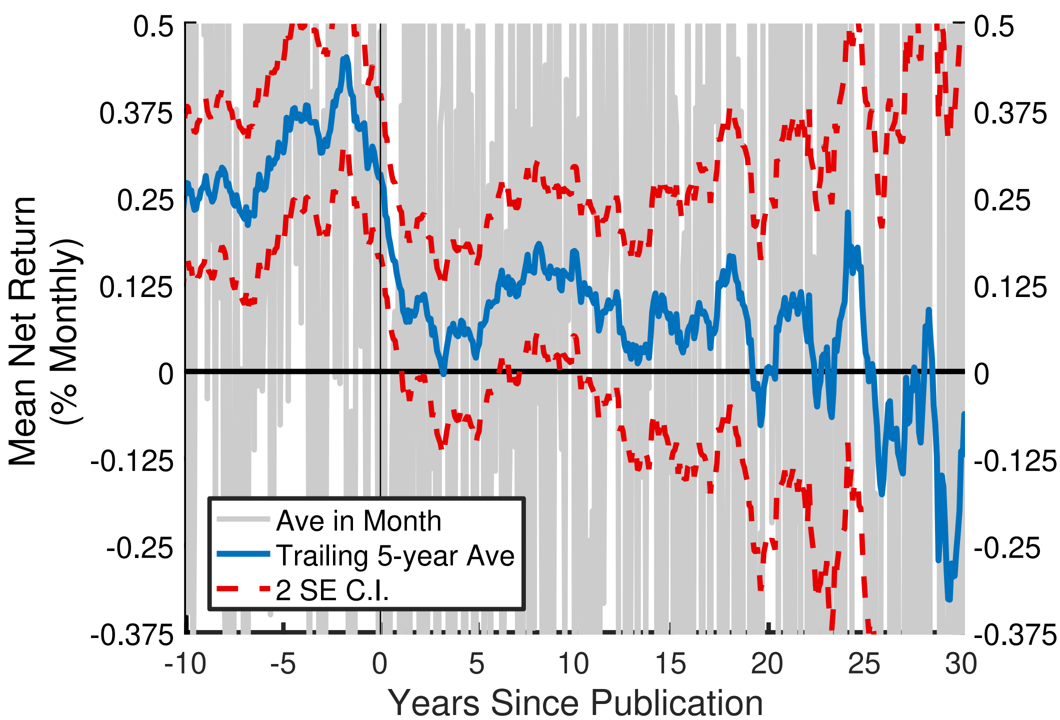


Figure 7: Cost-Mitigation Results: Distribution of Net Returns: Post-Publication. We mitigate transaction costs by optimally applying value-weighting and/or buy/hold spreads to 120 anomaly portfolios using in-sample data (Table 3 and Figure 4). We then measure the net returns to these mitigated strategies post-publication. Italics indicates anomalies with post-publication net return t-stats < 1.5. Bold indicates t-stats > 2.0. Only a handful of anomalies have t-stats > 2.0, suggesting that many of the large net returns are due to pure chance.