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

Publication Bias and the Cross-Section of Stock Returns

Andrew Y. Chen and Tom Zimmermann

2018-033

Please cite this paper as:
Chen, Andrew Y., and Tom Zimmermann (2018). "Publication Bias and the Cross-Section of Stock Returns," Finance and Economics Discussion Series 2018-033. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2018.033.

Publication Bias and the Cross-Section of Stock Returns

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

Tom Zimmermann
University of Cologne
tom.zimmermann@uni-koeln.de

May 2018*

Abstract

We develop an estimator for publication bias and apply it to 156 hedge portfolios based on published cross-sectional return predictors. Publication bias adjusted returns are only 12% smaller than in-sample returns. The small bias comes from the dispersion of returns across predictors, which is too large to be accounted for by data-mined noise. Among predictors that can survive journal review, a low t-stat hurdle of 1.8 controls for multiple testing using statistics recommended by Harvey, Liu, and Zhu (2015). The estimated bias is too small to account for the deterioration in returns after publication, suggesting an important role for mispricing.

*We thank Rebecca Wasyk for excellent scientific programing and Chris Hollrah for excellent research assistance. This paper originated from a conversation with Frank Schorfheide. We also thank Jack Bao, Svetlana Bryzgalova, Stefano Giglio, Yan Liu (AFA discussant), David McLean, Steve Sharpe, Nitish Sinha, Gustavo Suarez, and seminar participants at the Federal Reserve Board and Georgetown for helpful comments. Earlier versions of this paper relied on data provided by Robert Novy-Marx and Mihail Velikov, and we are grateful that they made their data available. The views expressed herein are those of the authors and do not necessarily reflect the position of the Board of Governors of the Federal Reserve or the Federal Reserve System.

1. Introduction

The nature of academia leads to an extremely thorough investigation of stock return data. Some argue that, subject to this much questioning, the data will tell you whatever you want to hear. Indeed, the data have informed us of more than one hundred portfolios with high returns and low market risk, leading many to be suspicious of information obtained in this manner (for example, Lo and MacKinlay (1990), Sullivan, Timmermann, and White (1999), Harvey, Liu, and Zhu (2015), Linnainmaa and Roberts (2016), Chordia, Goyal, and Saretto (2017)).[1]

Our interrogation of the data is subject to controls, however. Though a skilled investigator may be able to coerce her desired answer, the confession is only published if editors and referees deem it trustworthy. Indeed, in reflecting on his years as the editor of the Journal of Finance, Harvey (2014) recommends that authors should "convince the reader that there has been minimal data mining[2]." The effectiveness of the journal review process finds support in recent empirical studies that suggest that stock market anomalies are real (McLean and Pontiff (2016), Jacobs and Müller (2016), Yan and Zheng (2017)).

Publication bias is the net result of data mining and the journal review process. These effects oppose each other, and the net result remains an open question. In this paper, we propose an estimate of the net result, and apply it to a dataset of 156 published cross-sectional return predictors.

We find that the controlled interrogation of the CRSP tapes is surprisingly effective at uncovering true cross-sectional variation in returns. We estimate that a modest 12% of the typical predictor's in-sample return is due to publication bias—that is, while the typical equal-weighted quintile long-short return is about 8% per year, the bias-adjusted return is $(1-0.12)8\% \approx 7\%$.

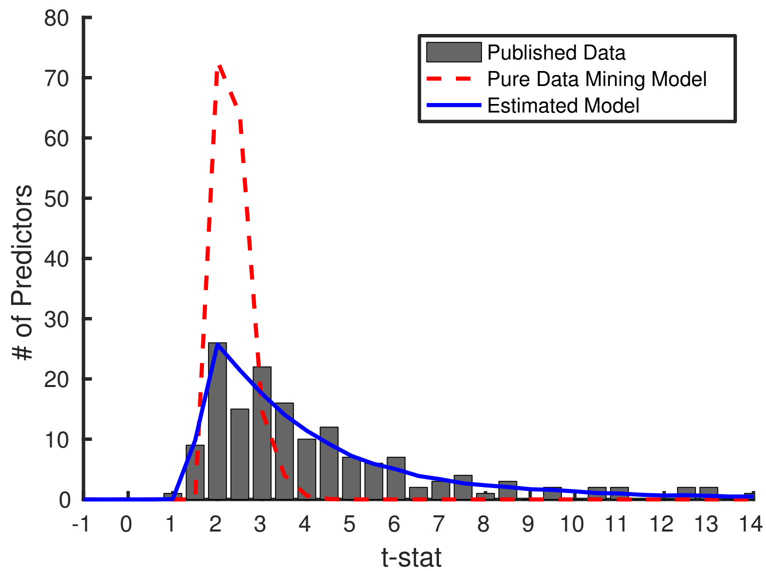
This modest bias adjustment comes from the shape of the distribution of published t-stats, seen in Figure 1. The left side of this distribution displays clear evidence of data mining, as predictors with t-stats less than 2.0 are conveniently missing.

The right side of this distribution, however, cannot be accounted for by data

[1]Throughout this paper, "return" refers to "mean return." We also omit the word "mean" in "in-sample mean return," "true mean return," etc.

[2]Data-mining is also known as "data-snooping," "p-hacking," "the file-drawer problem," "researcher degrees of freedom," and "the Garden of Forking Paths."

Figure 1: Published t-stats vs Pure Data Mining.



mining alone. Under pure data mining, t-stats should bunch up at the t-stat hurdle, and so we should see not only a sharp shoulder, but also a quick decay. To see this, suppose that there is no predictability anywhere, and published predictors are simply those which happened to have t-stats larger than 2. Then the t-stats would follow, well, a t-distribution, with many degrees of freedom, truncated around 2. This pure data mining distribution (dashed line) fits the left shoulder of the published data, but it decays far too quickly to account for the right tail.

In contrast, our model allows for both data mining and the possibility that the journal review process is effective. The power of journal review is embodied in the dispersion of true returns, that is, the amount of true variation in expected returns. This dispersion can be extracted by fitting the data, and the estimator finds that a large dispersion produces a tight fit (solid line). Intuitively, if true returns are dispersed, then t-stats pick up some of this dispersion, leading to the slow decay of the solid line in Figure 1.

The estimated model also implies that a predictor's in-sample return is informative about its underlying true return. This result is formalized in a Bayesian expression related to James and Stein (1961) shrinkage, and averaging across predictors produces our headline 12% number.

We find this modest bias using two unique data sets. Our primary dataset consists of 156 replicated long-short portfolios drawn from 115 publications in finance, accounting, and economics journals. To our knowledge, this is the most comprehensive dataset of cross-sectional predictors to date,[3] and is large enough to produce a precise estimate of the bias adjustment. The bootstrapped standard error for our 12% adjustment is just 1.8 percentage points.

Moreover, our replications perform quite well: the average in-sample return is 0.72% per month, with an average t-stat of 4.3. Indeed, we were able to replicate almost every predictor we attempted. Only four predictors that we made serious attempts to replicate produced t-stats less than 1.5. We make our data available online at http://sites.google.com/site/chenandrewy/code-and-data/.

We focus on replications instead of the original published statistics because performance measures are only partially standardized across publications. Only about half of the original papers report portfolio returns, with the other half reporting only regression results. By replicating portfolios, we ensure that we are feeding comparable performance measures into our estimator.

Nevertheless, we supplement our replications with hand-collected data. We search the publications of our 156 portfolios and write down any average portfolio returns and t-stats that are reasonably comparable. The resulting 77 portfolios produce a slightly smaller bias adjustment of 10%, though the difference is not statistically significant.

It's important to note that our small bias adjustments apply only to a select group of predictors. As our data consists of predictors published in peer-reviewed journals, our estimates are only relevant to predictors that have the possibility of being published. More specifically, our estimated model can be considered a formal description of portfolios with narratives and supplementary results that can survive the journal review process. Thus, our small adjustments do not apply to portfolios generated by uncontrolled data-mining experiments, which tend to be dominated by data-mining bias (Chordia, Goyal, and Saretto (2017)).

Our model estimates do apply, however, to predictors in the cross-sectional asset pricing literature. Indeed, our estimates lead to two broad implications for the zoo of anomalies.

The first implication is that the vast majority of published predictors are *not* statistical figments. We follow the multiple-testing literature and construct the

[3] Hou, Xue, and Zhang (2017)'s dataset of 447 anomalies contains many alternative lagging choices and variables which were not demonstrated to produce predictability in the original papers. Excluding these, their dataset contains 149 anomalies.

false discovery rate (FDR) implied by our model. We find that the FDR among published predictors is a tiny 1.5%.[4] Thus, we find that the traditional t-stat hurdle of 1.96 can actually be lowered, and even a t-stat hurdle of 1.79 leads to an FDR of 1.0%.

This surprising result may appear to contradict multiple-testing logic. If one runs 156 traditional hypothesis tests, the null of no predictability will likely be rejected by pure chance. Doesn't this logic imply that t-stat hurdles must be raised?

The problem with this logic is that, while running many tests raises concerns about lucky rejections, the many tests also provide information unavailable in a single test. Critically, examining many published predictors tells us about the nature of the publication process. We find that this process leads to highly dispersed true returns, that each t-statistic is informative about the underlying true return, and thus a high t-stat hurdle is not required. This logic contrasts with that of less structured multiple testing controls (such as the Bonferroni adjustment or Benjamini and Hochberg (1995)), which do not estimate the distribution of true returns. Instead, they use the same no predictability assumption from the single test setting, leading to strictly higher t-stat hurdles.

The second implication of our bias adjustment is that the deterioration in returns after publication cannot be attributed to publication bias. We replicate McLean and Pontiff (2016)'s result that post-publication returns are about 50% smaller than the returns in the original samples. We go beyond McLean and Pontiff, however, in that we produce a precise measurement of the amount of deterioration that is due to publication bias. We find that post-publication deterioration is 25 basis points per month larger than the publication bias adjustment, and we can reject with extreme confidence that there is no non-statistical deterioration (p-value < 0.0001).

This second implication is important because it suggests that mispricing plays a large role in the typical stock return anomaly. With statistical effects accounted for, the deterioration in returns post-publication must be due to either a decline in risk or a reduction in mispricing. The mispricing story has a compelling economic explanation: traders act on the published mispricing, bidding up underpriced assets and avoiding overpriced ones. Risk-based stories, on the other hand, do not provide a clear prediction.

[4]This FDR omits three published predictors that we failed to replicate. Even assuming these predictors are false discoveries, however, results in a low FDR of 4%.

Our results, combined with a couple other recent papers, provide a complete accounting for the returns of the anomaly zoo. We find that the typical anomaly return of 8% per year is 12% publication bias. McLean and Pontiff (2016) show that another 35% is mispricing that can be traded away. Chen and Velikov (2017) complete the story, showing that much of the remaining 53% can be accounted for by trading costs.

Related Literature Our paper is closely related to Harvey, Liu, and Zhu (2015) (HLZ), who also examine publication bias in cross-sectional asset pricing using a structured approach. They find that a t-stat hurdle in excess of 2.88 is required to obtain an FDR of 1%, far above our estimate of 1.79.

HLZ's data is substantially different than ours, however. While our dataset contains only variables that predict returns cross-sectionally, HLZ's dataset is comprised of asset pricing factors, broadly defined. Thus, the two sets of results suggest that there is much more publication bias in factor models and aggregate return predictors than in cross-sectional return predictors. There are other difference in methodology, however, which may be responsible for our different results, and unfortunately, we cannot provide a definitive reconciliation. In our view, such a reconciliation is an important question for future research.

Concerns about data mining bias in stock market predictors go back at least to Jensen and Bennington (1970) (see also Merton (1987), Lo and MacKinlay (1990), Black (1993)). Formal evaluations of data mining include Sullivan, Timmermann, and White (1999), Sullivan, Timmermann, and White (2001), and Chordia, Goyal, and Saretto (2017), who find strong evidence that data mining leads to spurious inference about predictability.

Publishing, however, involves both data mining (at least, collective data mining) and the journal review process. To measure the effects of journal review, one needs a body of evidence on the review process, something which was not available until the recent proliferation of published predictors.

Studies that take advantage of this proliferation have yet to come to a consensus. Harvey, Liu, and Zhu (2015), Linnainmaa and Roberts (2016), and Hou, Xue, and Zhang (2017) find that most published results are false, while Green, Hand, and Zhang (2014), McLean and Pontiff (2016), Jacobs and Müller (2016) come to the opposite conclusion.

Our paper brings to the debate a more structured model. This structure allows us to examine both bias-adjusted returns (à la McLean and Pontiff (2016)) and bias-adjusted statistical significance (à la Harvey, Liu, and Zhu (2015)) in the same framework. Additionally, our paper brings to bear the most comprehensive set of cross-sectional predictors to date, and we make this data publically available at http://sites.google.com/site/chenandrewy/code-and-data/.

Outside of finance, the literature on publication bias is large (see Christensen and Miguel (2016) for a review). Our approach is similar to Hedges (1992) and Andrews and Kasy (2017), who also explicitly model publication bias. Elements of our bias adjustment are also found in Efron (2011) and Liu, Moon, and Schorfheide (2016).

Our model complements Liu, Lu, Sun, and Yan (2015)'s model of anomaly discovery. While their model focuses on trading effects and abstracts from publication bias, we do exactly the converse. Thus, the two models capture two distinct components of the decay in returns post-sample. Other papers that study the long-term dynamics of anomaly returns include Alti and Titman (2017) and Penasse (2017).

The next section describes a quick 2-page version of our bias adjustment. Section 3 describes the full bias adjustment's methodology. The main results are in Section 4, which presents bias adjusted returns for 156 published predictors. Section 5 explains why the bias adjustment is small, and Section 6 examines the implications of our estimates for hypothesis testing and mispricing.

2. A Quick and Dirty Bias Adjustment

This section presents a quick and dirty bias adjustment that captures the intuition and magnitudes of our more rigorous estimation. This quick and dirty adjustment requires only two inputs: (1) the typical standard error on the long-short return and (2) the cross-sectional dispersion of long-short returns.

Suppose that in-sample returns r_i are noisy signals of true returns μ_i, and that μ_i are on average zero but have some dispersion:

$$r_i \sim N(\mu_i, \sigma) \tag{1}$$

$$\mu_i \sim N(0, \sigma_\mu), \tag{2}$$

where σ is the standard error of r_i (the same for all i), and σ_μ is the dispersion of true returns. Only statistically significant portfolios are published and observed. Thus we observe portfolio i only if

$$\frac{r_i}{\sigma} > 2. \tag{3}$$

To adjust for publication bias, we compute the expected true return conditional on publication using Bayes rule:

$$\hat{\mu}_i = (1-s)r_i \tag{4}$$

$$s = \frac{\sigma^2}{\sigma_\mu^2 + \sigma^2}. \tag{5}$$

The bias-adjusted return is simply the in-sample return r_i, shrunk at a rate s, where s is a transformed signal-to-noise ratio. Intuitively, if the publication process involves pure noise, the standard error σ is much larger than the dispersion in true returns σ_μ, and shrinkage is 100%. Alternatively, a large σ_μ relative to σ implies little shrinkage.

To calculate the bias adjustment, one needs σ and σ_μ. For now, let's assume that the published average standard error is a good estimate of σ. Using our dataset of 156 predictors, we have $\sigma \approx 0.18$, corresponding to a portfolio volatility of around 3% per month and a sample of around 20 years.

σ_μ, the dispersion of true returns, is not observed. However, σ_μ is observed indirectly through the dispersion of in-sample returns. To see this, note Equations (1) - (3) imply that that published returns follow a truncated normal distribution. The standard deviation of a truncated normal is

$$\text{Std}[r_i | r_i > 2\sigma] = f(\sigma, \sigma_\mu)\sqrt{\sigma^2 + \sigma_\mu^2} \tag{6}$$

where $f(\sigma, \sigma_\mu)$ is an adjustment due to truncation.[5] The LHS of equation (6) is directly observed, and everything on the RHS is observed except for σ_μ. Thus, Equation (6) can be used to estimate σ_μ by method of moments.

Figure 2 illustrates the estimation of σ_μ using Equation (6). The figure plots the model-implied standard deviation of published in-sample returns as a func-

[5]The adjustment is

$$f(\sigma, \sigma_\mu) = 1 + \alpha\phi(\alpha)/(1-\Phi(\alpha)) - [\phi(\alpha)/(1-\Phi(\alpha))]^2$$

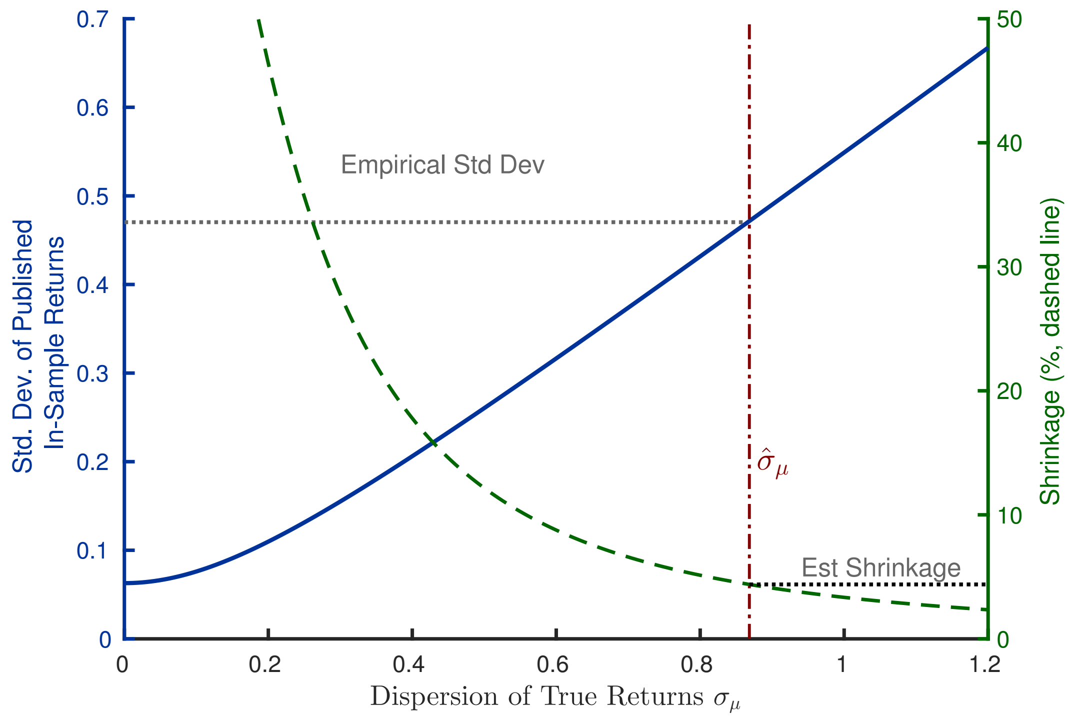
where $\alpha = 2\sigma/\sqrt{\sigma^2 + \sigma_\mu^2}$, $\phi()$ is the standard normal pdf, and $\Phi()$ is the standard normal cdf.

tion of σ_μ (Equation (6)). A very high $\sigma_\mu \approx 0.90$ is required to fit the empirical standard deviation of 0.47% found in our dataset. This high $\hat{\sigma}_\mu$ implies a very small shrinkage (dotted line) of about 5%. McLean and Pontiff (2016)'s dataset of 97 predictors leads to a standard deviation of 0.40%, and thus a lower $\hat{\sigma}_\mu$, but shrinkage is still below 10%.

Figure 2: Quick and Dirty Bias Adjustment. The solid line (left axis) plots the theoretical standard deviation of published in-sample returns (Equation (6)). "Empirical std dev" is the cross-sectional standard deviation of published in-sample returns in our dataset. $\hat{\sigma}_\mu$ (vertical dashed-dot line) is the dispersion of true returns implied by the data and Equation (6). The dashed line (right axis) is the theoretical shrinkage, defined by

$$\text{Bias-adjusted return} = (1 - [\text{Shrinkage}]) \times [\text{In-Sample Return}]$$

where shrinkage is calculated using Equation (5). "Est shrinkage" is the shrinkage implied by the data. Implied shrinkage is small, less than 10%. Returns are % per month. The model assumes $\sigma = 0.19$, the average published standard error in our dataset of 156 predictors.



The quick-and-dirty estimate overlooks a number of issues. It assumes homoskedasticity, normality, and no publication bias in standard errors. Moreover, we have not shown that this simple model provides a good fit for other moments in the data. These and other issues are addressed in the full estimation that follows.

3. Methods: Model, Bias Adjustment, and Data

This section describes our methodology. We describe our model (Section 3.1), bias adjustment estimation (Section 3.2), and data (Sections 3.3-3.4). Readers eager for results may wish to skip to Section 4.

3.1. A Simple Model of Predictor Publication

The model is summarized in Table 1. It is a statistical description of the portfolio publication process. We introduce characters like "academics" and "journals" to clarify the interpretation. There are no dynamics, trading, or strategic behavior. Thus the "true return" in the model is best understood as the publication-bias adjusted return, or the return in a world in which the predictor remains untouched by traders.

Table 1: Model Summary

This table summarizes the model (Section 3.1). All portfolios that have a remote possibility of being published are submitted. Only portfolios with narratives have a chance of publication, and the probability of publication $p(r_i/\sigma_i|t_{\text{cut}}, t_{\text{slope}})$ is increasing in the t-stat. All distributions are independent. The model has 7 parameters: μ_μ, σ_μ, ν_μ, μ_σ, σ_σ, t_{slope}, and t_{cut}.

Properties of the Portfolio Based on Narrative i	
True return	$\mu_i = \mu_\mu + \sigma_\mu \tau_{\nu_\mu}$
	$\tau_{\nu_\mu} \sim$ student's t with ν_μ d.o.f
In-sample return	$r_i = \mu_i + \epsilon_i$
	$\epsilon_{r,i} \sim N(0, \sigma_i)$
Log standard error	$\log \sigma_i \sim N(\mu_\sigma, \sigma_\sigma)$
Publication probability	$p(r_i/\sigma_i|t_{\text{cut}}, t_{\text{slope}}) = \dfrac{1}{1 + \exp(t_{\text{slope}}(r_i/\sigma_i - t_{\text{cut}}))}$

In search of tenure or other glory, academics search financial market data for publishable material. As a collective, the academics submit every portfolio that has a remote possibility of being published.

Journals only publish portfolios that meet two requirements. The first requirement is that the portfolio must contain a "narrative," or display a set of soft characteristics that meets the journals' standards. For example, a narrative for momentum is that investors overreact to the past year's returns. Thus, a narrative implicitly places a sign on the portfolio (long past winners and short past losers). Additionally, this narrative implies that returns are generally increasing in the past year's returns, and, perhaps, its returns are robust to various subsamples and portfolio construction methods.

We do not measure these soft characteristics directly. Instead, we model the quality of narrative i using its unobservable true return μ_i. The quality of all narratives is described by a scaled t-distribution:

$$\mu_i = \mu_\mu + \sigma_\mu \tau_{\nu_\mu} \tag{7}$$

$$\tau_{\nu_\mu} \sim \text{student's t with } \nu_\mu \text{ d.o.f,} \quad \text{i.i.d..} \tag{8}$$

where μ_μ, σ_μ, and ν_μ are parameters that govern the quality of narratives. Large ν_μ implies that μ_i is very close to a normal distribution with mean μ_μ and standard deviation σ_μ. We allow for small ν_μ in order to capture the idea that there may be rare portfolios with extremely good returns.

The scaled t-distribution of (7), with its single peak, is somewhat restrictive. In particular, it implies that there are many distinct signals in the data. For example, an alternative model might have three peaks (one for value, size, and momentum), and then the other predictors are just related variants around those peaks. We will see that single peak model is a good description of the data.

μ_μ, σ_μ, and τ_{ν_μ} are the net result of authors' data mining and the journals' narrative screening, and they ultimately describe whether the journals are publishing true returns. Clearly, if $\mu_\mu \gg 0$ the net result is that the narrative screen is effective at eliminating spurious portfolios. However, if $\mu_\mu = 0$ but $\sigma_\mu \gg 0$, the narrative screen is still effective. In this case, though the average narrative produces no returns, *some* narratives will have truly high expected returns. A low degrees of freedom ν_μ has similar effects.

The narrative screen is not perfect. Equation (7) means that some (perhaps many) narratives have $\mu_i < 0$, and the journals can't observe μ_i. Instead, they observe the in-sample return r_i, which is a noisy signal of μ_i. For a randomly selected narrative i, the in-sample return follows:

$$r_i = \mu_i + \epsilon_i \tag{9}$$

$$\epsilon_i \sim N(0, \sigma_i), \text{ i.i.d.} \tag{10}$$

where we assume that the standard error of the return σ_i is observed without error. This assumption can be justified by the fact that the standard error of a typical portfolio's standard error is two orders of magnitude smaller than the standard error itself.[6]

The above assumptions imply that the in-sample mean returns are uncorrelated across accepted portfolios. Theoretically this can be justified because journals are unlikely to accept a new predictor unless it is distinct from previously published ones. Moreover, the empirical pairwise time-series correlation between in-sample monthly returns is typically small. In our sample of 156 predictors, the median pairwise correlation is 0.036, and 80% of correlations are between -0.36 and 0.43. The full distribution of correlations can be found in Appendix A.3.

Narrative portfolios are heterogeneous in standard errors, and standard errors are lognormal

$$\log \sigma_i \sim N(\mu_\sigma, \sigma_\sigma) \text{ i.i.d.} \tag{11}$$

This assumption implies that standard errors are independent of true returns. One might think that the volatility component of the standard error should be correlated with the true return, as in risk-based theories. The cross-sectional asset pricing literature, however, is focused on portfolios that survive risk adjustment. Indeed, the literature tends to find a wide variety of in-sample returns with similar volatilities.

This setting leads to the journals' second requirement for portfolio i's publication: its in-sample return and t-stat must meet a soft threshold. The threshold is soft in that it is not a strict cutoff, but a probabilistic rule:

$$\text{decision}_i = \begin{cases} \text{pub}_i & \text{with prob } p(r_i/\sigma_i | t_{\text{cut}}, t_{\text{slope}}) \\ \text{reject}_i, & \text{otherwise} \end{cases} \tag{12}$$

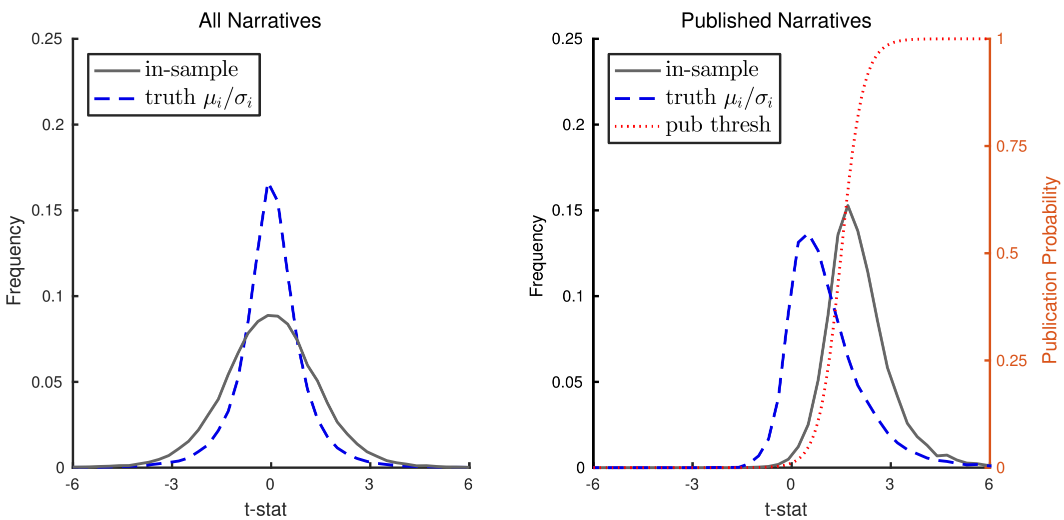
where the probability of publication is given by a logistic function

$$p(r_i/\sigma_i | t_{\text{cut}}, t_{\text{slope}}) = \frac{1}{1 + \exp(t_{\text{slope}}(r_i/\sigma_i - t_{\text{cut}}))}. \tag{13}$$

This function implies that t_{cut} is the midpoint of the soft threshold and t_{slope} is

[6] If the monthly return is normally distributed, then the standard error of the sample volatility is about $\sqrt{\frac{1}{2(T-1)}}$ times the true volatility. A sample size of 30 years leads to a factor of $\sqrt{\frac{1}{2(600-1)}} \approx$ 0.04.

Figure 3: Model Illustration. We simulate the model of biased publication (Table 1), and plot the distributions of t-stats. The left panel shows all "narratives," that is, all portfolios which have soft characteristics that satisfy the journals' requirements. In-sample t-stats (solid line) are noisy measures of the true t-stat (dashed line). The true t-stat is defined as the true return μ_i divided by the standard error. The right panel shows narratives which pass the publication threshold (dotted line). Publication bias is evident in the fact that the published in-sample t-stats are further from zero than the true t-stats. The simulation assumes $\mu_\mu = 0$, $\sigma_\mu = 0.30$, $\nu_\mu = 7$, $\mu_\sigma = -1.31$, $\sigma_\sigma = 0.45$, $t_{\text{cut}} = 1.5$, $t_{\text{slope}} = 3$.



the slope. The slope captures that fact that journals make editorial decisions that can soften a strict statistical requirement.

The statistical requirement improves the quality of published portfolios, as r_i and r_i/σ_i are a noisy signals about μ_i, and thus $\mathbb{E}(\mu_i|\text{pub}_i) > \mathbb{E}(\mu_i)$. Unfortunately, the cost of this quality control is a bias: $\mathbb{E}(\epsilon_i|\text{pub}_i) \neq 0$.

Figure 3 illustrates the model by plotting simulation results. The left panel shows the distribution of all narrative t-stats from a model simulation. The in-sample t-stats r_i/σ_i are more dispersed than their true counterparts μ_i/σ_i, as a result of measurement error ϵ_i. Despite the noise, the average of all narrative in-sample t-stats is an unbiased measure of the average true t-stat.

This unbiasedness is missing, however, from published narratives in the right panel. Only narratives that meet the publication threshold (dotted line) are published. Since narratives must have large t-stats to be published, this publication bias leads to a bias in the in-sample t-stats. In this particular simulation, the true t-stats are much closer to zero than their in-sample counterparts.

In the formalism that follows, it's helpful to gather all parameters into a vector θ:

$$\theta = [\mu_\mu, \sigma_\mu, \nu_\mu, \mu_\sigma, \sigma_\sigma, \nu_\sigma, t_{\text{cut}}, t_{\text{slope}}]. \tag{14}$$

3.2. Bias Adjustment and Estimation

Given parameters θ, Bayesian logic implies a bias adjustment formula. Then estimation of θ by maximum likelihood gives the empirical bias adjustment. Code for both estimation and bias adjustment can be found at http://sites.google.com/site/chenandrewy/code-and-data/.

To understand the bias adjustment formula, suppose we observe a published return r_i and standard error σ_i and want to estimate the true return μ_i. The naive rule assumes in-sample = true:

$$\hat{\mu}_i^{\text{naive}} = r_i. \tag{15}$$

The above expression is biased because it fails to condition on the fact that r_i is published—that is,

$$\mathbb{E}(\hat{\mu}_i^{\text{naive}}|\text{pub}_i) = \mathbb{E}(\mu_i|\text{pub}_i) + \underbrace{\mathbb{E}(\epsilon_i|\text{pub}_i)}_{\neq 0}. \tag{16}$$

In fact, typically, $\mathbb{E}(\epsilon_i|\text{pub}_i) \gg 0$, since the publication process selects for portfolios with large r_i (and thus large ϵ_i).

To correct for publication bias, we need to condition on the fact that the portfolio is published, as well as all other information about μ_i that is contained in the model. Thus, we define our estimator as follows:

$$\hat{\mu}_i \equiv \mathbb{E}(\mu_i|\text{pub}_i, r_i, \sigma_i; \theta). \tag{17}$$

As we have a complete model of publication, we can compute this expectation. The simplest way to compute this is to simulate the model, and then plot the average published μ_i as a function of (r_i, σ_i). This brute force approach, however, results in a bit of a black box.

Instead, we compute Equation (17) by applying Bayesian reasoning in two steps. The first and key step is to realize that, within the model, the fact that narrative i is published provides no information over and above the model parameters θ. This result comes from Equations (7)-(10), which provide a complete description of μ_i given θ. Intuitively, publication means that r_i is probably high, but the information set on the RHS implies that we already know r_i anyway. Formally, this reasoning means we can simplify our estimator:

$$\hat{\mu}(r_i, \sigma_i, \theta) \equiv \mathbb{E}(\mu_i | \text{pub}_i, r_i, \sigma_i; \theta) = \mathbb{E}(\mu_i | r_i, \sigma_i; \theta). \tag{18}$$

This result may be surprising, and indeed, has sometimes been called a paradox (Dawid (1994), Senn (2008)).[7]

Finally, we can write down an expression for our publication bias-adjusted return. To do this, rewrite the RHS of equation (18), using the definition of expectation and Bayes formula:

$$\hat{\mu}(r_i, \sigma_i, \theta) = \int_{-\infty}^{\infty} \mu' f_{\mu|r,\sigma}(\mu' | r_i, \sigma_i, \theta) \tag{19}$$

$$f_{\mu|r,\sigma}(\mu | r_i, \sigma_i, \theta) = \frac{f_N(r_i | \mu, \sigma_i) f_\tau(\mu | \mu_\mu, \sigma_\mu, \nu_\mu)}{\int d\tilde{\mu} f_N(r_i | \tilde{\mu}, \sigma_i) f_\tau(\tilde{\mu} | \mu_\mu, \sigma_\mu, \nu_\mu)} \tag{20}$$

where $f_N(r_i|\mu_i, \sigma_i)$ is just a normal pdf and $f_\tau(\mu_i|\mu_\mu, \sigma_\mu, \tau_\mu)$ is a scaled student's t pdf (Equations (7) and (9)). The above bias adjustment lacks closed form solutions, so we use numerical integration to compute both integrals.

To gain some intuition, consider the special case that $\nu_\mu \longrightarrow \infty$—that is, the true returns μ_i are normally distributed. In this case, $\hat{\mu}_i$ can be calculated using textbook normal-normal updating:

$$\hat{\mu}_i = (1 - s_j) r_j + s_j \mu_\mu \tag{21}$$

where the "shrinkage" s_j is

$$s_j = \frac{\sigma_i^2}{\sigma_\mu^2 + \sigma_i^2}. \tag{22}$$

Equations (21) and (22) capture intuitive aspects of publication bias. The negative effect of publication bias comes down to mistaking luck (high ϵ_i) for true

[7] The reasoning is quite straightforward, however, in a simpler problem. Suppose $x \sim N(y, 1)$, x is not observed, and we only observe y if $y > 0$. Then the density of x conditional on y being observed is still $N(y, 1)$. The same result holds if y is only observed with probability $p(y)$. The shape of $p(y)$ may imply that y is large, but we already know y anyway.

returns (high μ_i). Lucky portfolios have higher in-sample returns, and thus the adjustment in (21) increases in r_i. Lucky portfolios still contain signal, however. Thus, the adjustment contains a signal-to-noise ratio (22), where the noise is the standard error, and the dispersion of true returns σ_μ measures the signal. This signal measures the positive effect of publication bias, that is, the quality of the narrative controls.

Equations (21) and (22) show that our bias adjustment is an extension of the celebrated James and Stein (1961) estimator for a vector of means. These, like other empirical Bayes estimators, improve on the in-sample mean of a given observation by incorporating information from other observations (Efron (2012)).

The bias adjustment (Equations (19) and (20)) assumes that the model parameters θ are known. For our empirical application, we estimate θ using a large cross section of predictors via maximum likelihood. The invariance property of maximum likelihood implies that plugging our estimate $\hat{\theta}$ into Equations (19) and (20) gives the maximum likelihood bias-adjusted return.

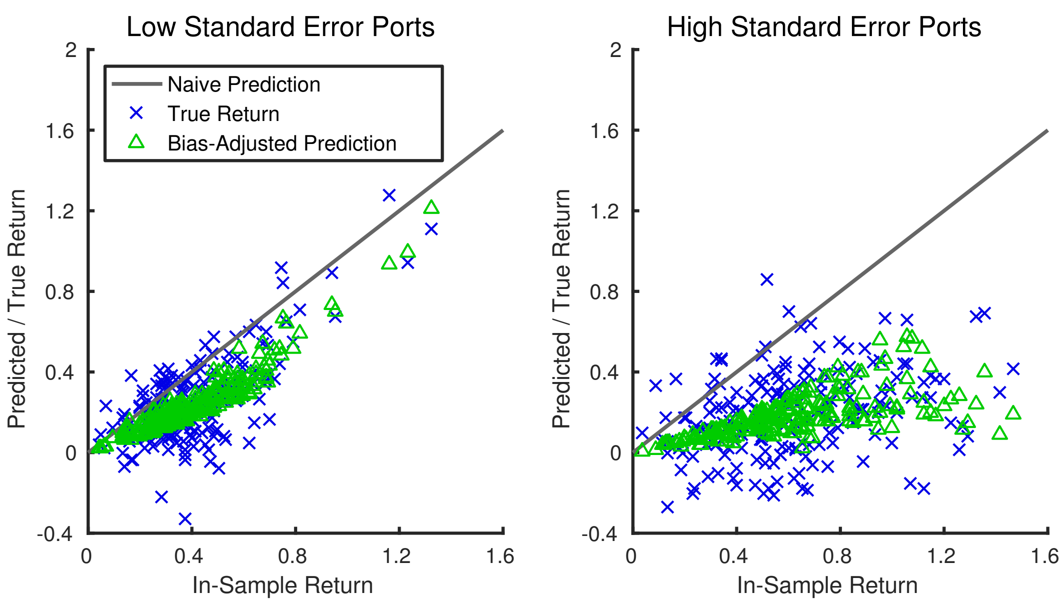
Maximum likelihood runs into a few technical difficulties. The most important is that estimation of μ_μ can lead to bad behavior for numerical optimizers, as the likelihood tends to become very flat for negative values of μ_μ. This problem is intuitive: Since only positive returns are published, the published returns have very little information about the true mean if the true mean is negative.

Thus, we do not estimate μ_μ and instead set it to the most conservative value of 0. Nevertheless, we find that setting $\mu_\mu = 0$ appears to result in a local maximum (Figure A.2), and, moreover, assuming other values for μ_μ has little effect on our estimated shrinkage (Table A.2). For details on how the likelihood is calculated and optimization performed, see Appendix A.2.

The effectiveness of our estimator is illustrated in Figure 4, which plots bias adjustments computed from applying maximum likelihood to simulated data. As the data is simulated, we can directly observe publication bias. This bias is manifested by the fact that the naive prediction that assumes in-sample = true (solid line) is typically higher than the true returns (blue x's). Moreover, this bias increases in the in-sample return, as well as the standard error (left vs right panels).

The figure shows that our estimator effectively removes publication bias from simulated data. The bias adjusted predictions run right through the center of the clouds of true returns, and adjust effectively regardless of the in-sample return

Figure 4: Bias Adjustment Illustration. We simulate 400 published portfolios, run a maximum likelihood estimation on the simulated data, and apply the bias adjustment (Equations (19)-(20)). Portfolios are separated into those below the median standard error (left panel) and those above the median (right panel). The naive prediction (solid line) assumes the in-sample return is equal to the expected true return. This prediction is biased upward compared to the true returns (x's) due to publication bias, and this bias increases in standard error and the in-sample return. The bias adjusted predictions (triangles) effectively adjust for publication bias. Parameters are in the Figure 3 caption.



and standard error of the portfolio.

3.3. Replications of 156 Cross-Sectional Return Predictors

Ideally, our data should (1) cover a comprehensive set of predictors, (2) use standardized performance measures, and (3) use statistics reported in the original publications. Achieving all three goals is impossible, however, as performance measures are only partially standardized across publications. Only about half of the predictors we examine report portfolio returns, with the other half reporting regression results. Moreover, the published portfolios use a variety of constructions and often only report returns adjusted for factor exposures or characteristics.

Thus, we construct standardized performance measures for 156 long-short portfolios by replicating 115 publications in accounting, economics, and finance journals. To our knowledge, this data is the most comprehensive set of cross-sectional predictors to date. The Hou, Xue, and Zhang (2017) dataset of 447 anomalies consists of only 149 anomalies if one excludes alternative lagging choices and anomalies that were not demonstrated to produce predictability in the original papers. The complete predictor list and detailed definitions can be found in Table A.1. We also make our dataset available at http://sites.google.com/site/chenandrewy/code-and-data/.

We cannot perfectly replicate all 156 predictors. Our replications have the more modest goals of (1) capturing the spirit of the paper and (2) producing t-statistics above 1.5. We use the cutoff of 1.5 instead of the traditional cutoff of 1.96 to allow for differences in t-stat calculations and updates to data sources. For example, Ritter (1991)'s study of long-term IPO underperformance finds t-stats in excess of 5.0 using event time returns. To compare across predictors, however, we must use calendar time, and our calendar time version of the IPO predictor produces a more modest t-stat of 1.6.

We achieved both goals for almost every predictor that we made serious attempts to replicate. In only four cases we failed to generate t-stats above 1.5. For three of these cases we assume the error was ours and omit the predictor. The remaining case is cash flow yield variance, which is shown to be statistically significant in Haugen and Baker (1996)'s regression results. Our portfolio sorts using cash flow yield variance produce a low t-stat of 1.20, but we keep it because the returns are monotonic and the long-short return is a respectable 0.39% per month.

Panel A of Table 2 describes the types of predictors in our dataset. Our predictors come from top tier academic journals and are quite diverse in the data sources used.

About 90% of our predictors are published in the “top 3” finance journals, “top 3” accounting journals, or the “top 5” general interest economics journals. The remaining 22 are also published in reputable journals, and include important predictors like Titman, Wei, and Xie (2004)’s investment anomaly and Amihud (2002)’s illiquidity measure.

The predictors use a wide variety of data sources. As one might expect, the majority use only accounting data (62) or use market prices in some fashion (48). But an equally large number (46) use diverse data that include analyst forecasts, trading-related measures, and corporate events. Many predictors are valuation measures (13), but they represent only 8% of our 156 predictors. Similarly, 9 of our 156 predictors are related to momentum. We include multiple valuation and momentum measures as they were considered distinct enough to be published.

The “top 3” finance journals appear to take interest in diversity. All six of our predictor categories are well-represented in the “top 3” finance row of Panel A. In contrast, the accounting journals are focused on accounting only predictors. This stark change in focus between accounting and finance journals does not affect our main results, however. We will see that our main results hold if we only use predictors from the “top 3” finance journals.

To evaluate predictor performance, we create long-short portfolios and measure monthly returns. For most (130) predictors, we use long-short extreme quintiles of the predictor. Both quintiles and deciles are commonly used in the literature, but using quintiles reduces the noise in mean returns and makes for easy comparison with McLean and Pontiff (2016).

The remaining 26 portfolios go long a particular indicator and short another set of stocks. These portfolios are mostly constructed from discrete predictors, such as Hong and Kacperczyk (2009)’s sin stock (stocks involved with alcohol, tobacco, or gaming) predictor or Cusatis, Miles, and Woolridge (1993) spinoff event predictor. We will see that these indicator variables perform similarly to our quintile sorts.

All but two portfolios are equal-weighted, as most of the original publications focus on either equal-weighted portfolios or Fama-Macbeth regressions.

The two exceptions are idiosyncratic volatility (Ang, Hodrick, Xing, and Zhang (2006)) and the Gompers, Ishii, and Metrick (2003) governance index. We value-weight these two predictors, as they perform far better using value-weighting in both the original papers and our replications.

Panel B of Table 2 shows the mean returns and t-stats from these portfolios using the papers' original in-sample periods. The returns in the original samples are around 0.70% per month on average, somewhat higher than McLean and Pontiff (2016)'s average in-sample return of 0.58%. To put these averages in perspective, our equal-weighted versions of value and momentum are significantly more anomalous than the average predictor, earning 1.01% and 1.24% in their original samples, respectively.

Value and momentum are not the only predictors with especially high returns. The cross-predictor standard deviation of about 0.50% is nearly as large as the mean, and similar to the 0.40% standard deviation in McLean and Pontiff (2016)'s dataset. Most of this dispersion comes from a long right tail in returns, which will be important for identification of our bias adjustment (Section 5.3).

The data show that journal categories produce fairly similar returns. The average top 3 finance return of 0.75% is 10 basis points higher than the average top 3 accounting return, but the two average returns are statistically similar given the large dispersion within each category. The average top 3 finance return is also similar to the average "other journal" return. T-stats do differ across journals categories, however. In particular, the top 3 accounting predictors have significantly larger t-stats, perhaps due to the smaller noise in accounting data compared to market data.

In contrast to journal categories, data categories produce significant dispersion in returns. The analyst forecast-based predictors have an average return of 1.00% per month, more than twice that of the corporate event predictors. The variation in returns across categories is fairly smooth, however, as for all categories there is another category with a similar return. This smooth variation is consistent with our assumption that all predictors are drawn from the same distribution.

Along the same lines, our quintile-sorts and indicator predictors produce very similar mean returns of 0.73% per month and 0.69%, respectively. Thus, we view predictors from both constructions as drawn from the same distribution in our estimation.

3.4. Alternative Data: Hand-Collected Returns for 77 Predictors

As our main results use replicated data, they do not measure the bias in the original reported numbers. Instead, they capture a more general concept of publication bias. In this concept, each publication suggests a trading strategy and provides a volume of statistics in support. It is then up to the reader to turn the statistics into precise portfolio constructions, as we do in our replications. These replicated expected returns contain bias due to the publication process, which our estimator then measures.

Our replications, however, may already remove some publication bias. Intuitively, the many authors of the original publications examined more specifications than our two-person co-authorship team did. Our less exhaustive specification search, then, may produce less data-mining bias.[8]

To examine this issue, we supplement our replications with hand-collected statistics. We search through the original publications of our 156 predictors and write down any average portfolio returns and t-stats that are reasonably comparable to our replications. We collect raw return quintile sorts when available, but also collect returns that adjust for factor models and characteristics, as well as portfolios that use alternative portfolio breakpoints, when necessary.

Table 3 compares this hand collected data with our replications. Panel A shows that our replications are on average faithful to the original papers. The hand-collected portfolio returns average 0.83% per month, just 2 basis points higher than our replications of the same portfolios. Similarly, the average hand-collected t-stat is 4.57, very close to our average replicated t-stat of 4.60. This similar performance is seen in Panels B and C, which examine subsets of the data that depend on the availability of statistics in the original papers. In the panel that examines hand-collected raw return quintile sorts (the same construction we use in our replications), the two datasets deviate in mean returns by just two basis points. Panel C shows that the hand-collected raw returns are slightly higher than our replications, perhaps because that subsample of hand-collected data contains many decile sorts.

Overall, the statistics for hand-collected and the matched replicated portfolios are similar. This close match suggests that our replications and the original reported statistics contain a similar amount of publication bias.

[8] We thank Stefano Giglio for making this point.

Ultimately, the amount of publication bias in both datasets can be measured using our estimator. We will see in Section 4.1 that both lead to similar bias adjustments.

Table 2: Summary Statistics for Replications of 156 Cross-Sectional Return Predictors

Table describes our replicated data. *Top 3 Finance* includes the Journal of Finance, the Journal of Financial Economics and the Review of Financial Studies. *Top 3 Accounting* includes the Accounting Review, the Journal of Accounting Research and the Journal of Accounting and Economics. *Top 5 Econ* includes the Quarterly Journal of Economics and the Journal of Political Economy (other econ journals did not have predictors that we replicated). Appendix A.1 provides a complete list of predictors. Monthly portfolio returns can be found at http://sites.google.com/site/chenandrewy/code-and-data/.

Panel A: Predictor Counts by Journal and Data Category

	Acct Only	Mkt Price	Analyst	Trading	Event	Other	Total
Top 3 Finance	21	37	11	7	10	8	94
Top 3 Accounting	30	4	0	2	0	0	36
Top 5 Econ	1	2	0	0	0	1	4
Other	10	5	2	3	2	0	22
Total	62	48	13	12	12	9	156

Panel B: Statistics for Long-Short Returns in Original Sample Periods

		Mean Return (% per month)		t-statistic	
	N	Mean	Std	Mean	Std
Journal					
Top 3 Finance	94	0.75	0.47	3.84	2.21
Top 3 Accounting	36	0.65	0.54	5.20	3.97
Top 5 Econ	4	0.55	0.16	2.82	2.02
Other	22	0.79	0.39	5.30	3.39
Data Category					
Accounting Only	62	0.65	0.42	4.97	3.32
Market Price	48	0.82	0.49	3.81	2.13
Corporate Event	13	0.41	0.18	2.73	0.92
Analyst Forecast	12	1.00	0.52	6.88	4.13
Trading	12	0.65	0.18	2.84	1.03
Other	9	0.92	0.77	3.70	2.85
Portfolio Construction					
Quintiles	130	0.73	0.47	4.41	2.97
Indicator	26	0.69	0.48	3.98	2.81

Table 3: Comparison of Hand-Collected and Replicated Portfolios

We hand-collect long-short returns and t-stats from original papers. We collect raw return quintiles when available, but include adjustments for factor models and characteristics, as well as alternative sorting methods when necessary. Our replications are all raw returns and quintile sorts or indicator variables. Panel A shows all 77 hand-collected portfolios, as well as the our replications of the same 77 portfolios. Panels B keeps only portfolios where the original paper provided raw quintile returns. Panel C keeps only portfolios where the original paper provided raw returns. Our replicated statistics are similar to the originals.

	Panel A: All Hand Collected Data				
	N	Mean Return		t-stat	
		Mean	Std	Mean	Std
Hand Collected	77	0.83	0.53	4.57	2.58
Replicated	77	0.81	0.54	4.61	3.11

	Panel B: Raw Returns and Quintiles Only				
	N	Mean Return		t-stat	
		Mean	Std	Mean	Std
Hand Collected	16	0.78	0.69	3.95	1.96
Replicated	16	0.76	0.56	3.69	1.51

	Panel C: Raw Returns Only				
	N	Mean Return		t-stat	
		Mean	Std	Mean	Std
Hand Collected	47	0.90	0.54	4.34	2.25
Replicated	47	0.81	0.54	4.27	2.97

4. Main Result: Estimated Publication Bias Adjustments

Having described our model, estimation, and data, we are finally in a position to show the main results. This section focuses on describing the estimated parameters, bias adjustments, and robustness. Section 5 explains why the bias adjustments are so small.

4.1. Estimated parameters and bias adjustments

Table 4 shows the main result: estimated model parameters and implied bias adjustments. We estimate a model of biased publication (Table 1) on our database of cross-sectional predictors (Table 2) by maximum likelihood (Section 3.2). The table shows our baseline specification "all," as well as four alternative specifications for robustness. We begin by discussing the baseline, but all specifications lead to similar results.

The first estimated parameter shows that many predictors are real. In the baseline estimation, the dispersion of true returns is estimated to be quite large at 0.45. Combined with the estimated fat tail parameter of 3.89, this implies that the cross-sectional standard deviation of all narrative true returns is $\sigma_\mu \sqrt{\frac{\nu_\mu}{\nu_\mu - 2}} =$ 0.64% per month. In other words, it's quite common to find narrative portfolios with true, bias-adjusted returns of 0.64% per month.

Moreover, the dispersion of true returns and the fat tail parameter are precisely estimated. Indeed, the dispersion of true returns is 9 standard errors from zero, showing that we can, with little doubt, reject the hypothesis that all predictors are false (equivalently, we reject that all shrinkage is 100%).

The remainder of the parameters demonstrate that the estimator works properly. The standard error parameters imply that the mean of all narrative standard errors is $\exp(\mu_\sigma + 0.5\sigma_\sigma^2) = 0.22\%$ per month. This is somewhat higher than the mean standard error for published data (0.19%) per month, indicating that there is a bit of downward publication bias in standard errors. The dispersion of log standard errors is similar to its naive counterpart, the standard deviation of published log standard errors. Similarly, the midpoint of the t-stat cutoff is close to the 1.50 cutoff imposed in our portfolio replications (Section 3.4). The threshold slope of 11 indicates that the publication threshold is very sharp, which can

Table 4: Estimation Results

Bootstrap standard errors are in parentheses. We estimate a model of biased publication (Table 1) on many cross-sectional stock return predictor portfolios by maximum likelihood assuming. Shrinkage for a portfolio is defined by

$$\text{Bias-adjusted return} = (1 - [\text{Shrinkage}]) \times [\text{In-Sample Return}]$$

where the bias-adjusted return is calculated using Equations (19)-(20). "All" uses our replications of 156 published predictors (Table 2). "t-stat > 2.0 only" uses only portfolios with t-stats > 2.0, "top 3 finance only" uses only portfolios from the Journal of Finance, Journal of Financial Economics, and Review of Financial Studies, and "Normal true returns" assumes $\nu_\mu = 100$. "Hand-collected data" uses 77 hand-collected portfolio statistics (Table 3). $\mu_\mu = 0$ throughout (see Section 3.2). The bias adjustment is small, at 12% of the in-sample return. This small shrinkage is well-estimated, and robust across all specifications.

		Replicated Data				
Estimated Parameters		All	t-stat > 2.0 Only	Top 3 Finance Only	Normal True Returns	Hand Collected Data
σ_μ	dispersion of	0.45	0.46	0.43	0.66	0.58
	true returns	(0.05)	(0.05)	(0.07)	(0.05)	(0.10)
ν_μ	fat tail (d.o.f.) of	3.89	3.89	3.85	100	6.56
	true returns	(1.38)	(1.40)	(2.55)	-	(4.63)
μ_σ	mean of log	-1.67	-1.68	-1.54	-1.69	-1.51
	standard error	(0.05)	(0.05)	(0.05)	(0.04)	(0.09)
σ_σ	std of log	0.51	0.50	0.49	0.52	0.68
	standard error	(0.02)	(0.03)	(0.03)	(0.03)	(0.10)
t_{cut}	midpoint of	1.61	2.00	1.64	1.56	1.88
	t-stat threshold	(0.08)	(0.02)	(0.14)	(0.05)	(0.18)
t_{slope}	slope of	10.97	100	8.31	11.51	6.91
	t-stat threshold	(41.78)	-	(42.62)	(42.35)	(37.49)
Estimated Bias Adjustments						
	Mean Shrinkage (%)	12.5	10.7	15.1	8.3	9.9
		(1.8)	(1.5)	(3.0)	(1.3)	(2.5)
	Median Shrinkage (%)	9.6	8.5	12.5	6.0	7.1
		(1.8)	(1.6)	(3.0)	(1.0)	(2.6)
	Std Shrinkage (%)	9.5	8.0	10.0	7.0	7.4
		(1.1)	(1.0)	(1.6)	(1.0)	(1.6)
Number of Predictors		156	132	94	156	77

be seen directly in the shape of the published t-stat distribution (Section 5.3). This slope comes with a very large standard error, as slopes larger than 20 are essentially vertical and distinguishing them is impossible. Nevertheless, we can say with certainty that the slope is fairly steep, as the 5th percentile of the bootstrapped distribution is 6.2.

The bottom of Table 4 provides the headline bias adjustment number from the abstract. It shows summary statistics for bias adjusted returns, where bias adjusted returns are calculated by applying Bayes rule within the context of the model (Equations (19)-(20)). To ease interpretation, we express bias adjusted returns in terms of "shrinkage," which is defined as

$$\text{Bias-adjusted return} = (1 - [\text{Shrinkage}]) \times [\text{In-Sample Return}]. \tag{23}$$

The baseline specifications finds that the mean shrinkage is modest, at 12.46%. In other words, for the typical in-sample return of 0.70% per month, the bias adjusted return is 0.61% = 0.70 × (1 − 0.1246)%). This small shrinkage is well-identified, with the a standard error of just 1.8 percentage points.

These mean shrinkage numbers summarize our main result. We can say with confidence that the net effect of publication bias on the cross-sectional return prediction literature is small.

4.2. Alternative Specifications

Our headline result is robust. Table 4 shows that the mean shrinkage is around 12% regardless of the specification. Further robustness is seen in even more specifications in Appendix A.3.

The first alternative specification in Table 4 limits the data to predictors with t-stats > 2.0. This specification addresses the concern that the predictors with small t-stats are not comparable to the other predictors. This concern does not affect our main result, as estimating on the 132 portfolios with t-stats > 2.0 results in a slightly smaller shrinkage of 10.7%.

One can argue that our measure of publication bias is muddled as it mixes together journals from different fields and of various reputations. The "top 3 finance only" specification shows that our main results hold if we restrict our predictors to the 94 published in the Journal of Finance, Journal of Financial Economics, and the Review of Financial Studies. According to our estimator, the top 3 finance journals have a publication bias of 15.1%, slightly larger than our baseline estimate, but within one standard error. Indeed, the standard error is noticeably larger for the top3 finance estimate, as the number of predictors used in the estimation drops by roughly 50%.

Some readers may be concerned that our small shrinkage is reliant on the estimated fat tail in true returns, and that fat tailed models are tricky to estimate. The "normal true returns" column should alleviate this concern. It assumes that the degrees of freedom parameter is 100 and omits it from the estimation. The resulting shrinkage of 8.3% is somewhat smaller than the baseline, but leads to the same conclusion: publication bias is modest.

4.3. Estimation on Hand-Collected Data

The final column ("hand-collected data") of Table 4 examines our hand-collected dataset (Section 3.4). This alternative data should alleviate the concern that our replications do not fully capture the publication bias in the original papers.

The hand-collected data column shows that similar bias adjustments result from using the original reported statistics. The mean shrinkage is 9.9%, economically and statistically close to the baseline estimate of 12.5%.

Indeed, the parameters are overall similar to the baseline. The hand-collected data shows a weaker fat tail with $\nu_\mu = 6.56$ compared to the baseline estimate of 3.89, however both estimates are within one standard error of each other. This weaker fat tail comes with a larger σ_μ, so the overall dispersion is similar: the hand-collected data leads to a cross-predictor standard deviation of 0.70, not far from the baseline estimate of 0.66.

The t-stat threshold t_{cut} is estimated to be 1.88, just a touch lower than the 1.96 threshold used in a 2-sided 5% significance test. This lower threshold comes from the fact that 4 of our hand-collected portfolios have t-stats < 1.96. These 4 predictors come with additional tests, however, that produce statistical significance. For example, while Ikenberry, Lakonishok, and Vermaelen (1995) report that their repurchase announcement portfolios produce a modest t-stat of 1.5, their event studies show t-stats in excess of 6.

The 77 hand-collected portfolios include both quintile and decile sorts, as well as portfolio returns that are adjusted for factor exposure and characteristics.

Estimations on the 47 portfolios that do not make adjustments led to similar results.

4.4. Heterogeneity in bias adjustments

The mean shrinkage is just 12%, but the cross-predictor standard deviation is somewhat large at 10 percentage points (Table 4). Figure 5 shows, however, that modest shrinkage is a good description of the estimates overall.

Figure 5 shows a histogram of the shrinkage distribution, as well as the identities of the predictors. Each name represents one predictor, and thus each stack of names represents the number of predictors within a particular bin of shrinkage.

The shrinkage distribution is right-skewed, with a large mass at low shrinkage. Thus, 78% of predictors have shrinkage less than 20%. Moreover, even the high shrinkage predictors have only a moderate amount of publication bias. The maximum shrinkage among 156 predictors is 38.6%.

Figure 5 also illustrates the determinants of the predictor-level shrinkage. Portfolios with high return volatility (red or purple text) dominate the distribution above 20% shrinkage. This result is intuitive: portfolios with a lot of noise are more likely to have had lucky in-sample returns, and thus exhibit more publication bias (on average).

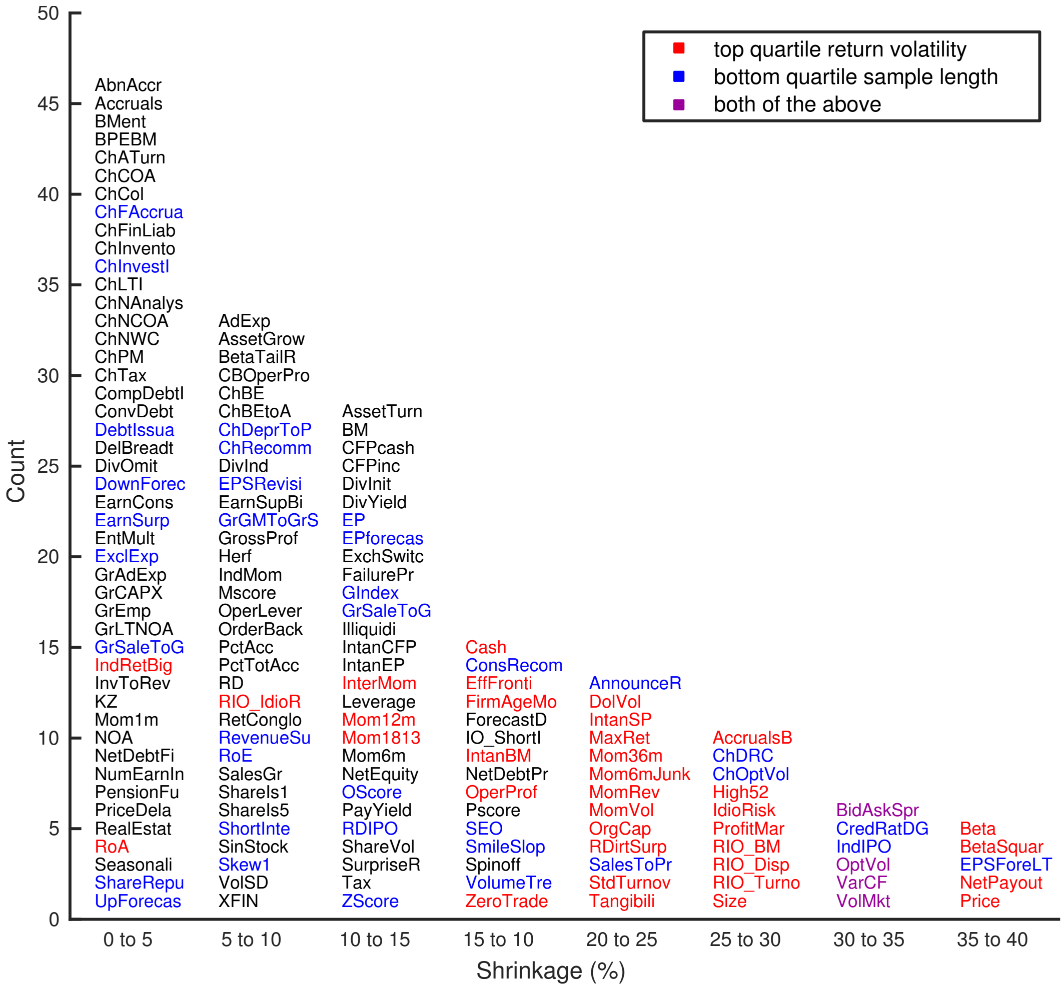
Theoretically, the sample length should also play a key role in the amount of noise, and thus the magnitude of shrinkage (Equation (22)). However, we find that the empirical correlation between the sample lengths and return standard errors is only mildly negative, at -0.14.

It's worth noting that higher shrinkage does not imply poor bias-adjusted returns. Higher shrinkage portfolios have larger standard errors, and these portfolios need to have higher in-sample returns in order to meet the publication t-stat threshold. The higher in-sample returns, then, compensates for the large shrinkage. Indeed, the highest shrinkage portfolio, La Porta (1996)'s long term earnings growth predictor (EPSForeLT), earns a respectable bias-adjusted return of 0.47% per month.

Figure 5: Distribution of publication bias adjustments. We estimate a model of biased publication (Table 1) on 156 long-short portfolios (Table 4). Shrinkage is defined by

$$\text{Bias-adjusted return} = (1 - [\text{Shrinkage}]) \times [\text{In-Sample Return}]$$

where the bias-adjusted return is calculated using Equations (19)-(20). Each name represents one portfolio. The full references are in Table A.1. Publication bias is heterogeneous and right skewed, but modest overall.



5. Why is Publication Bias So Small?

Our results may be surprising, especially to those who work in the cross-sectional returns literature. Some might feel certain that, we *must* be mining the data. At least, as a collective we must be.

But there are controls on the publication process that are designed to limit the negative effects of data-mining. And a priori, it's hard to know which force dominates.

Our estimator takes an empirical approach, and lets the data speak about which force is stronger. The estimator belongs to the empirical Bayes family, and as such, it learns about a given predictor by studying the larger family of predictors (Efron (2012)). This family displays considerable dispersion, much more dispersion than would be implied by pure noise. Using this information, the estimator concludes that there is a lot of signal in each predictor.

This section explains the intuition behind our estimated bias adjustments. Section 5.1 shows that the mean bias adjustment is determined by the dispersion of true returns. Section 5.2 shows how the dispersion of true returns is determined by the dispersion of in-sample returns. Section 5.3 finishes up by comparing our bias adjusted returns with McLean and Pontiff (2016)'s lower bound.

5.1. Mean Shrinkage is Determined by the Dispersion of True Returns

Our bias adjusted return comes from a complicated expression (Equations (19) and (20)), but plotting the bias adjustments reveals some intuition for how the adjustment works.

Figure 6 plots the bias adjustments against the standard errors of the portfolio's in-sample return. The scatter shows a clear pattern: the larger the standard error, the more shrinkage is recommended. This result is intuitive: more volatile portfolios or publications with shorter samples are more likely to have lucky in-sample returns. Thus, these lucky portfolios require a larger adjustment.

Indeed, the relationship between the standard error and shrinkage can be expressed in closed form for the normal approximation of our model. In this approximation, the shrinkage is a sort of noise-to-signal ratio, where the noise is the portfolio-specific standard error (Section 3.2). The normal approximation

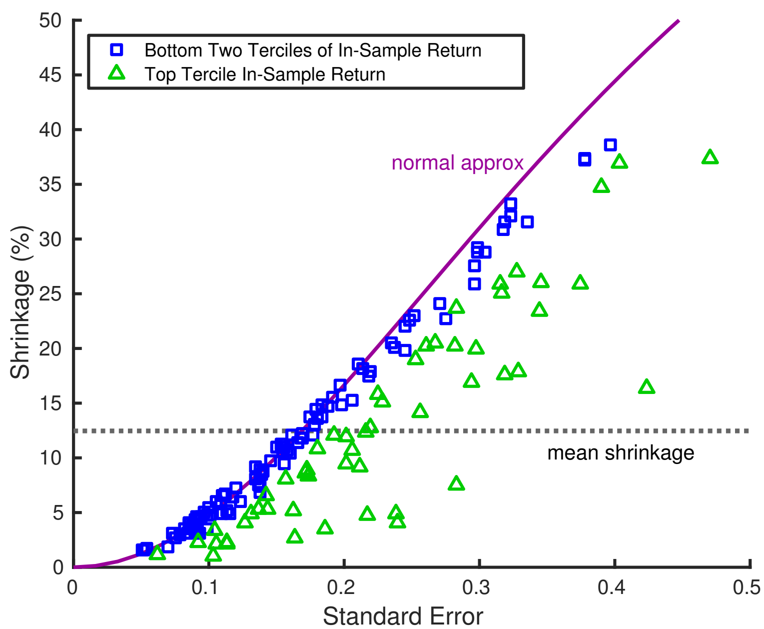
Figure 6: Determinants of the bias adjustments. Each marker represents one portfolio from our database of 156 predictors. Shrinkage is defined by

$$\text{Bias-adjusted return} = (1 - [\text{Shrinkage}]) \times [\text{In-Sample Return}]$$

where the bias-adjusted return is Eqns. (19)-(20). The normal approximation is

$$[\text{Shrinkage}]_i = \frac{\sigma_i^2}{\hat{\sigma}_\mu^2 + \sigma_i^2}.$$

where σ_i is the standard error and $\hat{\sigma}_\mu^2$ is the estimated dispersion of true returns. The normal approximation works well for most portfolios. The primary determinant of the mean shrinkage is $\hat{\sigma}_\mu^2$.



(solid line) describes the full shrinkage formula well for most of the portfolios, though it misses the portfolios with very high in-sample returns (triangles). This deviation occurs because the full model has a fat tail in true returns, and these high return portfolios are more likely to belong in the tail.

But overall, the normal approximation does a good job of capturing shrinkage. Indeed, our headline 12% shrinkage can be derived using this approximation. Plugging in the mean standard error of 0.19% and our estimated $\hat{\sigma}_\mu = 0.45$, the typical shrinkage is approximately

$$\frac{\sigma_i^2}{\hat{\sigma}_\mu^2 + \sigma_i^2} = \frac{0.19^2}{0.45^2 + 0.19^2} = 15\% \tag{24}$$

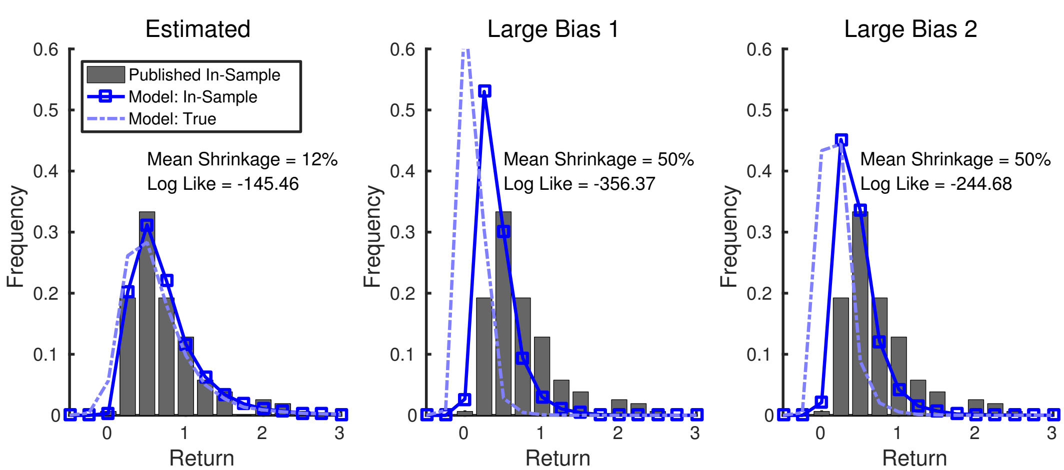
This analysis begs the question: where does our estimate of $\hat{\sigma}_\mu = 0.42$ come from?

5.2. The Estimated Dispersion of True Returns is Determined by the Dispersion of In-Sample Returns.

We've seen that the mean shrinkage is determined by the estimated dispersion of true returns $\hat{\sigma}_\mu$. Here, we show that $\hat{\sigma}_\mu$ is identified by the dispersion of in-sample returns.

This identification is illustrated in Figure 7, which plots the distribution of in-sample returns in the data (bars) and model-implied distributions (squares). The left panel shows our estimated model. The other panels show models that display large publication bias.

Figure 7: Identification of the Dispersion of True Returns σ_μ. Each panel illustrates the fit of a different model. The left panel compares the distribution of published returns in the data (bars) with the estimated model (squares, Table 4). The distribution of true returns implied by the model is plotted for comparison (dash dotted line). "Large bias 1" uses $\sigma_\mu = 0.075$, but all other parameters remain the same. "Large bias 2" has $\sigma_\mu = 0.075$, $t_{\text{slope}} = 3$, and $t_{\text{cut}} = 3.0$. Both large bias models are poor fits for right tail of the distribution.



The estimated model (left panel) is a tight fit for the published data. The model histogram counts are close to the data throughout the distribution. This tight fit comes despite the fact that the model has only six estimated parameters, and that the model also must fit the distribution of standard errors (not shown). The left panel also shows the distribution of true returns implied by the model (dash-dot line). True returns are quite close to the in-sample returns, leading to the small mean shrinkage of just 12%.

The middle panel plots the distribution of in-sample returns implied by a model with large bias. This model deviates from the estimated model only in that $\sigma_\mu = 0.075$, compared to the estimated $\sigma_\mu = 0.421$. $\sigma_\mu = 0.075$ is chosen in order to achieve a mean shrinkage of 0.50. This shrinkage is important because McLean and Pontiff (2016) find that post-publication returns are lower than in-sample returns by 58%. Thus, $\sigma_\mu = 0.075$ is required to assign the bulk of this decline to publication bias.

The middle panel shows that this large bias model is a poor fit for the data. This model fails to capture the dispersion of in-sample returns. Our estimator sees much more than just the dispersion however. As we use maximum likelihood, the estimator sees the fit of every in-sample return bin, and the excessively high counts for the low return bins as well as the excessively low counts for the high returns bins are all penalized by the estimator. Indeed, the log-likelihood of this model is more than 200 log points lower than our maximum likelihood estimate.

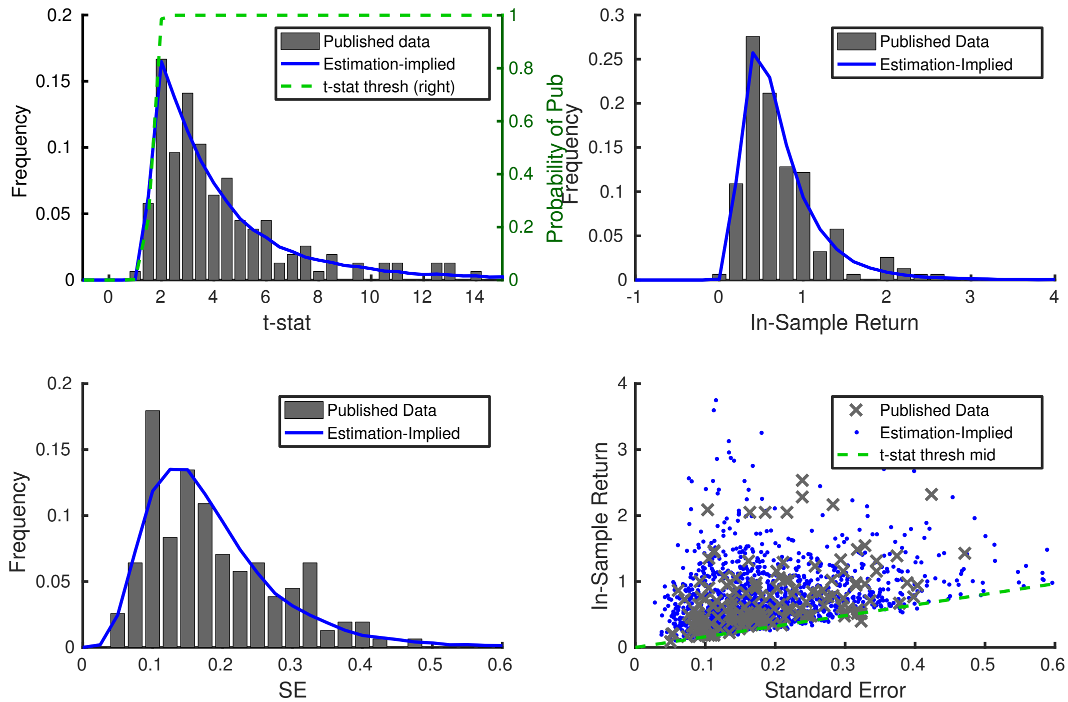
One might argue that the large bias model needs to have other parameters adjusted to fit the data. One adjustment consistent with the idea that the data exhibits a large bias is a large increase in t_{cut} and a large decrease in t_{slope}—that is, the journals exhibit a strong preference for large t-stats.

The right panel of Figure 7 shows that increasing t_{cut} and decreasing t_{slope} does not fit the data either. The panel assumes a high t threshold midpoint of $t_{\text{cut}} = 3.00$ and a relatively shallow slope of $t_{\text{slope}} = 3$, in addition to $\sigma_\mu = 0.075$. This strong preference for large t-stats improves the fit on the left side of the distribution, but overall the fit is still poor, with the log-likelihood is still about 100 log points below our maximum likelihood estimate. Experiments with other strong preferences for high t-stats led to similar results.

This identification discussion begs the question: does the estimated model fit the other dimensions of the data? Figure 8 shows that the answer is yes.

The figure's 4 panels plot the distribution of t-stats, in-sample returns, standard errors, as well as a plot that illustrates the correlation between in-sample returns and standard errors. All 4 panels of show that the estimated model captures the data very well.

Figure 8: Model fit. We simulate the model using estimated parameter values (Table 4) and compare the distribution of observables with those from our database of 156 predictors (Table 2). The t-stat thresh uses estimated parameter values. The model fits all observable distributions very well, including the correlation between in-sample returns and standard error (bottom right).



5.3. External Verification: McLean and Pontiff (2016)'s Lower Bound

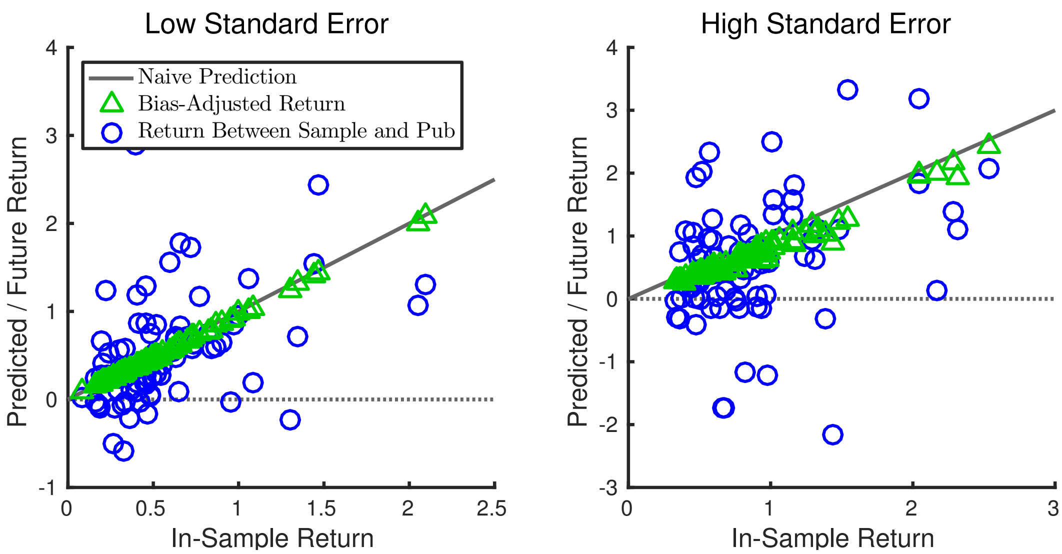
We've shown that the model fits data that it algorithmically targets. But is there a way to bring to bear data from outside the estimation?

The natural external test would be to compare our estimated shrinkage to out-of-sample returns. Out-of-sample returns, however, are polluted by trading effects, since investors that read the publications may decide to change their portfolio allocations.

McLean and Pontiff (2016) (MP) develop a clever way to isolate trading effects. Assuming that papers are less widely known before publication, the return between the end of the original sample and the publication date should exhibit a limited amount of trading effects. Thus, the mean return in this "between" sample serves as a lower bound on the publication bias adjusted return.

Figure 9 examines MP's lower bound. The figure shows scatterplots of bias-adjusted returns against the in-sample return, as well as the mean returns between the end of the original sample and publication. For comparison, the figure also plots the naive prediction: in-sample return = true return.

Figure 9: Bias adjusted returns and returns after the original in-sample period. Each marker represents one long-short portfolio. The naive prediction (solid line) assumes that the true return is equal to the in-sample return. "Return between sample and pub" (circles) are the mean returns between the end of the original paper's sample and the publication date. The bias-adjusted prediction (triangles) use Equations (19)-(20) and estimated parameters (Table 4). Low standard error portfolios are those with standard error below the median. The bias-adjusted returns are consistent with the lower bound implied by the mean return between sample and publication.



The plot shows that bias-adjusted returns are very similar to naive predictions. This, essentially, is the main message of our paper: publication bias is modest. This modest bias is particularly evident in low-standard error portfolios (left panel).

More importantly, the figure provides external validation of our bias adjustment. The circles represent the returns between the end of the sample and publication. High publication bias implies that these circles would be symmetrically spread across 0. Instead, the circles are more or less symmetrically spread around the naive prediction line.

Moreover, our bias adjusted returns are slightly above the middle of the cloud of circles. Averaging across the blue circles we find that our mean bias-adjusted return is consistent with MP's lower bound. The average return in the between sample is 0.56% per month, slightly below our mean bias-adjusted return of 0.63%.

6. Implications for the Anomaly Zoo

The asset pricing literature has uncovered hundreds of patterns in the cross-section of stock returns. Recent research has aimed to place some order on this zoo of anomalies (Cochrane (2011), Harvey, Liu, and Zhu (2015), Kozak, Nagel, and Santosh (2017), Feng, Giglio, and Xiu (2017)).

Our bias adjusted returns imply that (1) correcting for data mining does not reduce the size of the anomaly zoo, and (2) much of the predictability at the time of publication was due to mispricing. Sections 6.1 and 6.2 discuss these implications, respectively.

6.1. Hypothesis Tests Adjusted for Publication Bias

The small bias adjustments suggest that the zoo of anomalies cannot be simply attributed to publication bias. Here, we look more closely at the question and show that nearly 100% of published anomalies are true using multiple testing statistics.

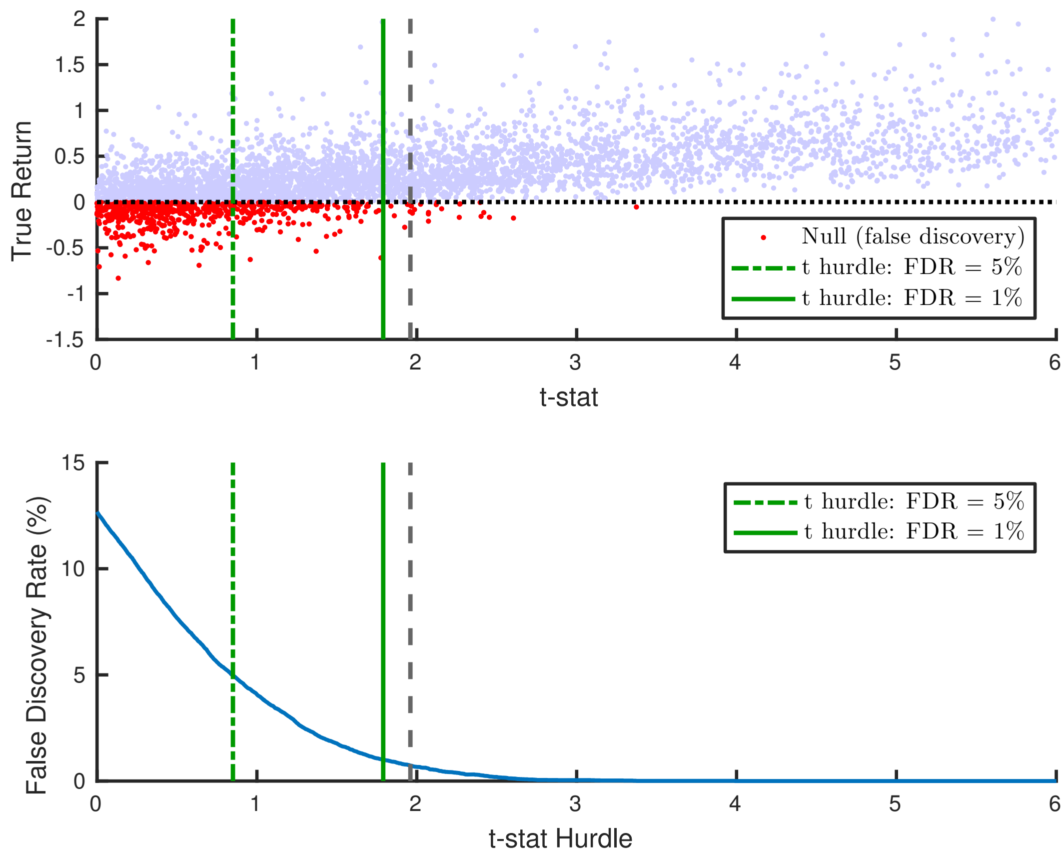
To demonstrate this, we use our estimated model to calculate the false discovery rate (FDR), one of the multiple testing statistics recommended by Harvey, Liu, and Zhu (2015) (HLZ). We focus on the FDR instead of the family wise error rate because of its simple interpretation: the FDR is the share of anomalies that are statistical figments.

We define a null predictor as one with non-positive true returns ($\mu_i \leq 0$). This definition stays close to the classical definition and also to HLZ, both of which define the null as $\mu_i = 0$. This null is also important as it is used in popular multiple testing adjustments (for example, Bonferroni and Benjamini and Hochberg (1995)). In contrast to these approaches, our model considers portfolios with worse-than-zero true returns, and thus we must label $\mu_i < 0$ as null in addition to $\mu_i = 0$.[9]

[9] An alternative to using $\mu_i \leq 0$ as the null is to estimate a model in which μ_i is drawn from a strictly positive distribution and a point mass at zero. The point mass at zero serves a similar function as the distribution of negative μ_i in our model.

Using this definition of a null predictor, we can calculate false discovery rates by simulating the estimated model. Figure 10 illustrates this calculation. The top panel shows the distribution of narrative portfolios' true returns against t-stats. Null portfolios, that is, portfolios with non-positive true returns are highlighted in red.

Figure 10: Multiple Tests of the Null of Non-Positive True Returns. We simulate narrative portfolios using our estimated model (Table 4). The top panel shows a scatter plot of 10,000 true returns against t-stats. The false discovery rate (FDR) for a given t hurdle is the fraction of predictors exceeding the hurdle that have non-positive true returns (red dots). Incorporating information from multiple tests leads to the t-hurdles given by the green lines, which are more lenient t-stat hurdle than the traditional 1.96 (grey dashed line).



For portfolios with t-stats of 0.5, the probability of being null is about 50%, as indicated by the even split between red and light blue dots near the left side of the panel. The cloud of dots, however, is upward sloping, and thus, higher t-stat portfolios are more likely to be non-null.

This pattern is more precisely described in the bottom panel. The panel plots the FDR as a function of the t-stat hurdle. Even the extremely generous hurdle of 0 leads to a low FDR of 12.7%. Increasing the t-stat hurdle decreases the false discoveries sharply. At a t-stat hurdle of 0.85 we already have an FDR of 5%, one of the FDR values in recommended by HLZ. Raising the t-stat hurdle to 1.79 reduces the FDR to 1%, HLZ's alternative recommendation.

Thus, our results suggest that the traditional t-stat hurdle of 1.96 could actually be *loosened*. Even a t-stat hurdle of 0.85 effectively controls for false discoveries, given that the portfolio has a top-tier quality narrative. This surprising result comes from the fact that we estimate the dispersion of narrative true returns to be very large. This large dispersion implies that the t-stat is a strong signal about the underlying true return, the cloud of dots in the top panel of Figure 10 is upward sloping, and thus a large t-stat is not required for concluding the true return is positive.

In contrast, single hypothesis tests do not allow for any inferences about the dispersion of true returns. With a single predictor, the only reasonable approach is to assume that the predictor is useless, leading to a high t-stat hurdle. Less structured multiple-testing adjustments such as the Bonferroni and Benjamini-Hochberg adjustments also do not estimate the distribution of true returns and instead assume the worst case, as we explain in Appendix A.4.

An important caveat is that our results apply only to predictors that are judged to have quality narratives, that is, soft characteristics that satisfy the journal review process. Thus, our results do *not* imply that a randomly data-mined portfolio with a t-stat of 0.85 is 95% likely to have positive true returns, and our low FDR estimates are consistent with Chordia, Goyal, and Saretto (2017)'s results regarding randomly generated signals. Similarly, our results do not imply that journals should consider loosening their t-stat restrictions without carefully maintaining their narrative controls.

Nevertheless, our results do apply to predictors that are published in peer-reviewed journals. Indeed, as peer-reviewed journals only allow narratives that meet a threshold centered around 1.5 (Table 4), almost all published predictors are true. According to our estimates, the FDR among published predictors is a tiny 1.5%. This FDR would increase if we regard the three predictors that we omitted due to replication failures as false discoveries. However, even assuming these are false discoveries leads to a low FDR of 4%.

One interpretation of the low estimated t-stat hurdles is that the traditional null hypothesis of $\mu_i = 0$ is inadequate. This null describes only a tiny portion of narrative predictors. As a result, non-null predictors are not unusual, and the null does not help separate interesting cases from typical ones. In this setting, one may want to use an "empirical null" that is designed to generate unusual cases (Efron (2012). We discuss one such empirical null in Appendix A.5.

Our results contrast with HLZ, who find that t-stat hurdles close to 3.0 are required to reduce the FDR below 1%. HLZ's data is substantially different than ours, however. While our dataset includes only predictors that demonstrate return predictability, HLZ's dataset is comprised of asset pricing factors, broadly defined. Perhaps as a result, the dispersion of t-statistics is larger in our data. The 90th percentile t-statistic in our sample is 8.0, compared to the 90th percentile of 6.3 in HLZ.

There are other differences in methodology which may contribute to the deviation in results, however. Our model uses both mean returns and standard errors, while HLZ consider only the t-stat. HLZ's model assumes a mixture distribution for true returns, while ours assumes a single fat-tailed distribution. A clear reconciliation of our low t-stat hurdle and HLZ's t-stat hurdles above 3.0 is, in our view, an important question for future research.

6.2. Implied Mispricing

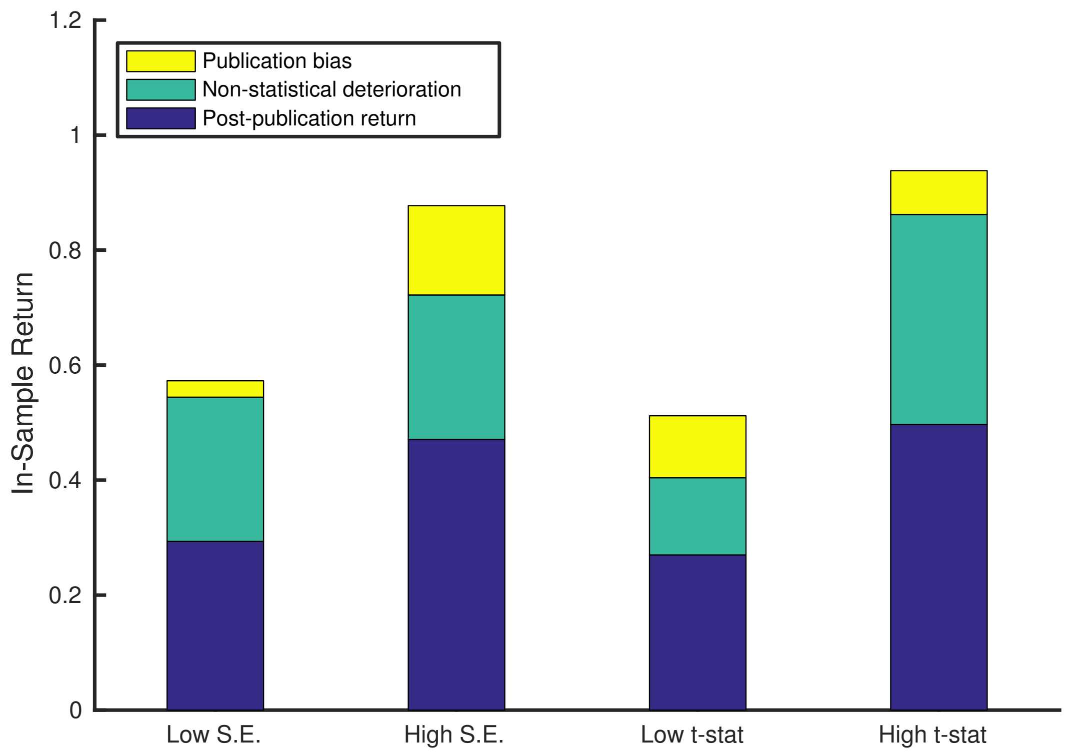
Our estimation results thus far are negative. We find that publication bias cannot account for the zoo of published stock return anomalies.

In this section, we present evidence in favor of a more positive conclusion. Combining our estimation with the empirical methodology of McLean and Pontiff (2016) and Marquering, Nisser, and Valla (2006), we find evidence suggesting that mispricing plays an important role throughout the anomaly zoo.

McLean and Pontiff (2016) (MP) and Marquering, Nisser, and Valla (2006) build on the insight that the average return post publication is informative about the nature of the cross-sectional predictability. If predictability is due to mispricing, post publication returns should be poor, as traders bid up underpriced assets and avoid overpriced ones. Similarly, if predictability is due to publication bias, post-publication returns should be poor as the pre-publication predictability was a statistical figment. On the other hand, risk-based stories do not provide a clear prediction.

This logic leads to the decompositions of in-sample returns seen in Figure 11. The figure decomposes the average in-sample return across predictors into (1) publication bias (2) non-statistical deterioration, and (3) the post-publication return.

Figure 11: Implied mispricing. This chart decomposes the average in-sample return across predictors into publication bias, non-statistical deterioration, and the post-publication return. Publication bias is the average in-sample return minus the average bias-adjusted return (Equations (19)-(20)). Post-publication return is the mean return in the sample after publication. Non-statistical deterioration is the difference between the average bias adjusted return and the average post-publication return. Each bar computes averages within a subset of the predictors. "Low S.E." consists of portfolios with below the median standard error, and similarly for "low t-stat." A significant portion of in-sample returns is due to non-statistical deterioration, suggesting that mispricing is important.



The decomposition comes from computing average returns of different types and taking differences. Publication bias is the difference between the average in-sample return and the average bias adjusted return (Equations (19) and (20)). Post-publication returns are the average returns after the publication date. The non-statistical deterioration is the difference between the average bias-adjusted return and the average post-publication return.

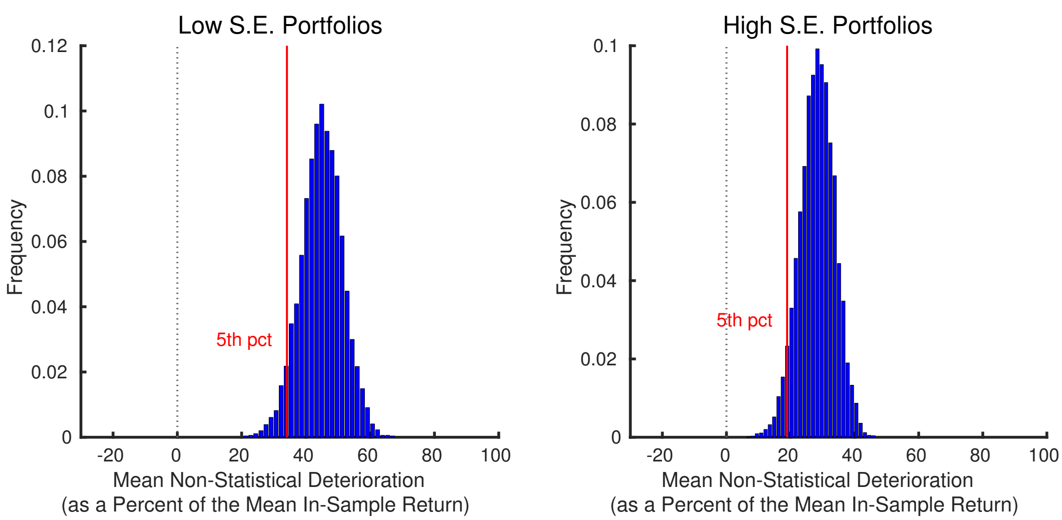
The figure shows that a significant portion of in-sample returns is due to non-statistical deterioration. The average post-publication return is 25 basis points per month lower than the average bias-adjusted return. This non-statistical deterioration accounts for a significant 34% of the average in-sample return. Non-statistical deterioration is largest for high t-stat, low standard error, and high in-sample return portfolios, consistent with MP and the hypothesis that mispricing is the underlying driver of predictability.

Our results go beyond MP, however, in several ways. While MP place an upper bound on publication bias, and thus a lower bound on non-statistical deterioration, our bias adjustments provide direct estimates of both. MP's upper bound also runs into a couple theoretical concerns: namely it assumes that there is no selection happening between the end of the in-sample period and publication. Our estimator avoids these concerns by explicitly modeling and estimating the selection process. Finally, our dataset is nearly twice the size of MP's, which is important considering how volatile stock returns are and how short the post-publication period can be.

These refinements mean that we can make inferences on subsamples of predictors with confidence. This increased precision is highlighted in Figure 12, which plots the bootstrapped distribution of non-statistical deterioration.

The figure splits the data into portfolios with standard errors below the median (left panel) and those above (right panel). In both panels, the 5th percentile of the bootstrapped standard errors are very far from zero. Thus we can be confident that a significant share of in-sample returns is due to non-statistical deterioration. Indeed, the hypothesis that publication bias can account for all of the deterioration is soundly rejected for both low and high standard error portfolios (p-values < 0.0001).

Figure 12: Implied mispricing: bootstrapped distribution. We resample the data 10,000 times and run our estimator on each resampling. Low S.E. portfolios have standard errors below the median. Mean non-statistical deterioration is the average bias-adjusted in-sample return minus the average post-publication return, all divided by the average in-sample return. The hypothesis that publication bias accounts for all deterioration in returns post-publication is soundly rejected, suggesting that mispricing is important.



7. Conclusion

We find that the net effect of publication bias on cross-sectional stock predictors is modest. These results suggest that editors and referees provide an important control on our collective mining of the data, leading to the discovery of a multitude of portfolios with high returns and low market risk. These high returns, however, are short-lived, as traders quickly act on the publication of return predictability and eliminate mispricing.

Our results, combined with a couple other recent papers, provide a complete accounting for the returns of the anomaly zoo. We find that the typical anomaly return of 8% per year is 12% publication bias. McLean and Pontiff (2016) show that another 35% is mispricing that can be traded away. Chen and Velikov (2017) complete the story, showing that much of the remaining 53% can be accounted for by trading costs.

A. Appendix

A.1. Additional Details on the Replicated Data

Table A.1: Description of Anomaly Construction. This table provides details of the construction of return predictors used in the paper. Data come from the CRSP stock return database, Compustat North America Annual and Quarterly databases, IBES earnings estimates database, OptionMetrics, Thomson SDC and a number of additional databases noted in the descriptions of specific anomalies. Our final database is set up at monthly frequency. We lag annual Compustat data by five months and quarterly Compustat data by 3 months to assure availability of relevant data at the time of trading.

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
AbnAccr	Abnormal Accruals	Xie	2001	1971	1992	Define Accruals as net income (ib) minus operating cash flow (oancf), divided by average total assets (at) for years t-1 and t. If oancf is missing, replace operating cash flow with funds from operations (fopt) minus the annual change in total current assets (act) plus the annual change in cash and short-term investments (che) plus the annual change in current liabilities (lct) minus the annual change in debt in current liabilities (dlc). For each year t, regress Accruals on: the inverse of average total assets for years t-1 and t, the change in revenue (sale) from year t-1 to t divided by average total assets, propery plant and equipment (ppegt) divided by average total assets, industry dummies for Fama-French's 48 industry classification. AbnormalAccrual is the residual from this cross-sectional regression.
Accruals	Accruals	Sloan	1996	1962	1991	Annual change in current total assets (act) minus annual change in cash and short-term investements (che) minus annual change in current liabilities (lct) minus annual change in debt in current liabilities (dlc) minus change in income taxes (txp). All divided by average total assets (at) over this year and last year. Exclude if abs(prc) < 5.
AccrualsBM	Book-to-market and accruals	Bartov and Kim	2004	1980	1998	Binary variable equal to 1 if stock is in the highest Accrual quintile and the lowest BM quintile, and equal to 0 if stock is in the lowest Accrual quintile and the highest BM quintile. Exclude if book equity (ceq) is negative.
AdExp	Advertising Expense	Chan et al	2001	1975	1996	Advertising expense (xad) over market value of equity (shrout*abs(prc))
AnnounceR	Earnings announcement return	Chan et al	1996	1977	1992	Get announcement date for quarterly earnings from IBES (fpi = 6). AnnouncementReturn is the sum of (ret - mktrf + rf) from one day before an earnings announcement to 2 days after the announcement.
AssetGrowth	Asset Growth	Cooper et al	2008	1968	2003	Annual growth rate of total assets (at)

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
AssetTurnover	Asset Turnover	Soliman	2008	1984	2002	Sales (sale) divided by two year average of net operating assets. Net operating assets is the sum of receivables (rect), inventories (invt), current assets other (aco), net property, plants and equipment (ppent) and intangibles (intan), minus accounts payable (ap), other current liabilities (lco) and other liabilities (lo). Exclude if abs(prc) < 5 or AssetTurnover < 0.
Beta	CAPM beta	Fama and MacBeth	1973	1926	1968	Coefficient of a 60-month rolling window regression of monthly stock returns minus the riskfree rate on market return minus the risk free rate (ewretd - rf). Exclude if estimate based on less than 20 months of returns.
BetaSquared	CAPM beta squred	Fama and MacBeth	1973	1926	1968	Square of Beta (defined above).
BetaTailRisk	Tail risk beta	Kelly and Jiang	2014	1963	2010	Each month, compute the 5th percentile over daily returns over all firms. For all daily return observations with return below that 5th percentile, compute the average of (log(ret/5th percentile of cross-sectional return distribution). Call that average tailEX. BetaTailRisk is the coefficient of a 120-month rolling regression of a firm's stock return on tailEX. Exclude if price less than 5 or share code greater than 11.
BidAskSpread	Bid-ask spread	Amihud and Mendelsohn	1986	1961	1980	Spread estimates from Shane Corwin's website (https://www3.nd.edu/ scorwin/) divided by price (abs(prc)).
BM	Book to market	Fama and French	1992	1963	1990	Log of annual book equity (ceq) over market equity (see above).
BMent	Enterprise component of BM	Penman Richardson Tuna	2007	1961	2001	(ceq + che - dltt - dlc - dc - dvpa+ tstkp) / (mve_c + che - dltt - dlc - dc - dvpa+ tstkp). Exclude if price less than 5.
BPEBM	Leverage component of BM	Penman Richardson Tuna	2007	1961	2002	BP - EBM, where BP = (ceq + tstkp - dvpa)/(shrout*abs(prc)), and EBM is defined above. Exclude if price less than 5.
Cash	Cash to assets	Palazzo	2012	1972	2009	Ratio of quarterly cash and short-term investments (cheq) and total assets (atq).
CBOperProf	Cash-based operating profitability	Ball et al	2016	1963	2014	Revenue (revt) minus cost (cogs) - (administrative expenses (xsga) - R&D expenses (xrd)) minus annual change in receivables (rect), annual change in investment (invt) and annual change in prepaid expenses, plus annual change in current deferred revenue (drc), long-term deferred revenue (drlt), accounts payable (ap) and accrued expenses (xacc), all divided by total assets (at) in year t-1. Replace all variables in the numerator with 0 if they are missing. Exclude if share code is greater 11, market value of equity, BM or total assets are missing, or if SIC code between 6000 and 6999.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
CFPcash	Operating Cash flows to price	Desai, Rajgopal, and Venkatachalam	2004	1973	1997	Operating cash-flow (oancf) divided by market value of equity. If operating cash-flow is missing, replace by difference betwee net income (ib) and level of accruals, where the latter is the annual change in current assets (act) minus the annual change in cash and short-term investments (che), minus the annual change in current liabilities (lct) plus the annual change in debt in current liabilities (dlc) plus the annual change in payable income taxes (txp) plus depreciation (dp).
CFPinc	Cash flow to market	Lakonishok et al	1994	1968	1990	Net income (ib) plus depreciation (dp) divided by market equity. Exclude NASDAQ stocks.
ChATurn	Change in Asset Turnover	Soliman	2008	1984	2002	Annual change in AssetTurnover (defined above). Exclude if price less than 5.
ChCOA	Change in current operating assets	Richardson et al	2005	1962	2001	Difference in current operating assets (total current assets (act) minus cash and short-term investments (che)) between years t-1 and t, scaled by average total assets (at) in years t-1 and t.
ChCol	Change in current operating liabilities	Richardson et al	2005	1962	2001	Difference in current operating liabilities (total current liabilities (lct) minus debt in current liabilities (dlc)) between years t-1 and t, scaled by average total assets (at) in years t-1 and t.
ChDeprToPPE	Change in depreciation to gross PPE	Holthausen and Larcker	1992	1978	1988	Annual percentage change in the ratio of depreciation (dp) to property, plant and equipment (ppent).
ChDRC	Deferred Revenue	Prakash and Sinha	2012	2002	2007	Annual change in deferred revenue (drc) scaled by average total assets (at) in t-1 and t. Exclude if negative book equity (ceq), deferred revenue equal to 0 in both years, revenue less than 5m, or SIC code between 6000 and 6999.
ChEQ	Sustainable Growth	Lockwood and Prombutr	2010	1964	2007	Ratio of book equity (ceq) to book equity in the previous year. Include only if book equity is positive this year and last year.
ChEqu	Change in equity	Richardson et al	2005	1962	2001	Difference in book equity (ceq) between years t-1 and t, scaled by average total assets (at) in years t-1 and t.
ChFAccrual	Change in Forecast and Accrual	Barth and Hutton	2004	1981	1996	Within upper half of Accruals distribution, equal to 1 if mean earnings estimate increased relative to the previous month. 0 if it decreased.
ChFinLiab	Change in financial liabilities	Richardson et al	2005	1962	2001	Difference in financial liabilities (sum of long-term debt (dltt), current liabilitites (dlc) and preferred stock (pstk)) between years t-1 and t, scaled by average total assets (at) in years t-1 and t.
ChInventory	Inventory Growth	Thomas and Zhang	2002	1970	1997	12 month change in inventory (invt) divided by average total assets.
ChInvestInd	Change in capital inv (ind adj)	Abarbanell and Bushee	1998	1974	1988	Growth in capital expenditure (capx) minus average growth in capital expenditure in the same industry (two-digit SIC). If capx is missing, capital expenditure is defined as the annual change in property, plant and equipment (ppent). Capital expenditure growth is defined as the percentage growth of capx today relative to the average capx over the previous two years (.5*(capx$_t-1$ + capx$_t-2$), or as percentage growth relative to the previous year only if t-2 is missing.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
ChLTI	Change in long-term investment	Richardson et al	2005	1962	2001	Difference in investment and advances (ivao) between years t-1 and t, scaled by average total assets (at) in years t-1 and t.
ChNAnalyst	Decline in Analyst Coverage	Scherbina	2008	1982	2005	Binary variable equal to 1 if the number of analysts (numest) for next quarter's EPS estimate decreased relative to three months ago, and 0 if it increased.
ChNCOA	Change in Noncurrent Operating Assets	Soliman	2008	1984	2002	Twelve-month change in noncurrent operating assets. Noncurrent operating assets is ((at - act - ivao) - (lt - dlc - dltt))/at.
ChNWC	Change in Net Working Capital	Soliman	2008	1984	2002	Twelve-month change in net working capital. Net working capital is ((act - che) - (lct - dlc))/at
ChOptVol	Option Volume relative to recent average	Johnson and So	2012	1996	2010	Based off of OptionVolume1. OptionVolume2 = OptionVolume1 / average of OptionVolume1 from months t-6 to t-1.
ChPM	Change in Profit Margin	Soliman	2008	1984	2002	Annual change in profit margin PM (profit margin defined below). Exclude if price less than 5.
ChRecomm	Change in recommendation	Jegadeesh Kim Krische Lee	2004	1985	1998	(As in MP). If an analyst issues a new strong buy recommendation (ireccd == 1), we assign a value of 1 to that event, if an analyst issues any other change in recommendation, we assign a value of -1; we assign 0 if the recommendation is unchanged. The final variable is the average over the constructed variable over all analysts each month.
ChTax	Change in Taxes	Thomas and Zhang	2011	1977	2006	4-quarter change in quarterly total taxes (txtq), scaled by lagged total assets (at).
CompDebtI	Composite debt issuance	Lyandres Sun Zhang	2008	1970	2005	Log of long-term debt (dltt) plus debt in current liabilties (dlc) minus log of the same variable 5 years ago.
ConsRecomm	Consensus Recommendation	Barber Lehavy MicNichols Trueman	2001	1985	1997	Binary variable if the monthly mean of recommendations (ireccd) over analysts is greater than 3, and 0 if it is less or equal than 1.5.
ConvDebt	Convertible debt indicator	Valta	2016	1985	2012	Binary variable equal to 1 if deferred charges (dc) greater than 0 or common shares reserved for convertible debt (cshrc) greater than 0.
CredRatDG	Credit Rating Downgrade	Dichev and Piotroski	2001	1970	1997	A downgrade happens if credit rating (splticrm) decreased by at least one notch relative to the previous month. CredRatDG = 1 if a downgrade happened over the past 3 months.
DebtIssuance	Debt Issuance	Spiess and Affleck-Graves	1999	1975	1989	Equal to 1 if debt issuance (dltis) greater 0 and 0 otherwise. Exclude if share code > 11 or missing book-to-market.
DelBreadth	Breadth of ownership	Chen, Hong and Stein	2002	1979	1998	Quarterly change in the number of institutional owners (numinstowners) from 13F data. Exclude if in the lowest quintile of stocks by market value of equity (based on NYSE stocks only).
DivInd	Dividends	Hartzmark and Salomon	2013	1927	2011	Binary variable equal to 1 if return with dividends (ret) is greater than return without dividends (retx) 11 months ago or 2 months ago, and 0 otherwise or if price less than 5.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
DivInit	Dividend Initiation	Michaely Thaler Womack	1995	1964	1988	Define dividend initiation as having paid a dividend in month t (divamt > 0) and not having paid a dividend in the 24 preceding months. DivInit is equal to 1 if a dividend was initiated in the past 12 months and 0 otherwise. Exclude if share code greater 11 and use NYSE stocks only.
DivOmit	Dividend Omission	Michaely Thaler Womack	1995	1964	1988	Define dividend omission as not having paid a dividend in the current month or the two preceding months, but having paid dividends in the 3, 6, 9, 12, 15, 18 months before. DivOmit is equal to 1 if a dividend was omitted in the previous 12 months and 0 otherwise.
DivYield	Dividend Yield	Naranjo et al	1998	1963	1994	4 times latest dividend (divamt) divided by price (prc). Include only if dividend has been paid in all of the past 4 quarters.
DolVol	Past trading volume	Brennan Chordia Subrahmanyam	1998	1966	1995	Log of two-month lagged trading volume (vol) times two-month lagged price (prc).
DownForecast	Down forecast EPS	Barber Lehavy MicNichols Trueman	2001	1985	1997	Binary variable equal to 1 if mean earnings forecast (meanest) decreased over the past month.
EarnCons	Earnings Consistency	Alwathainani	2009	1971	2002	Average earnings growth over previous 48 months. Earnings growth is defined as EPS (epspx) minus EPS 12 months ago divided by average EPS 12 and 24 months ago. Exclude if price less than 5, absolute value of 12 month earnings growth greater 600%, or earnings growth and earnings growth 12 months ago have different signs.
EarnSupBig	Earnings surprise of big firms	Hou	2007	1972	2001	Average monthly value of EarningsSurprise (defined above) of the 30% largest companies by market value of equity in the same Fama-French 48 industry. Exclude the largest 30% of companies for EarnSupBig (not to compute the anomaly)
EarnSurp	Earnings Surprise	Foster et al	1984	1974	1981	EPS (epspxq) minus EPS twelve months ago - Drift, scaled by standard deviation of that expression. Drift is the average earnings growth (EPS - EPS twelve months ago) over the past two years. Exclude if price less than 5
EffFrontier	Efficient frontier index	Nguyen and Swanson	2009	1980	2003	Frontier is the residual of a regression of log(BM) on log(book equity (ceq)), long-term debt (dltt) to assets (at), capital expenditures (capx) to revenue (sale), R&D expense (xrd) to revenue, advertising expense (xad) to revenue, property plant and equipment (ppent) to assets, EBIT (ebitda) to assets, and dummies for Fama-French's 48 industry definitions. Regression is updated each month with a rolling window of 60 months.
EntMult	Enterprise Multiple	Loughran and Wellman	2011	1963	2009	Market value of equity + long-term debt (dltt) + debt in current liabilities (dlc) + deferred charges (dc) - cash and short-term investments (che) , divided by operating income (oibdp). Exclude if missing book equity or negative operating income.
EP	Earnings-to-Price Ratio	Basu	1977	1957	1971	ib / lag(market value of equity, 6 months). NYSE stocks only. Exclude if EP < 0. Lag simulates the Dec 31 market equity used in original paper

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
EPforecast	Earnings Forecast	Elgers, Lo and Pfeiffer	2001	1982	1998	Mean earnings estimate (meanest) for next quarter's earnings divided by stock price (prc). Exclude if price less than 1.
EPSForeLT	Long-term EPS forecast	La Porta	1996	1983	1990	Long-term earnings forecast (fgr5yr) lagged by twelve months. Exclude if book equity (ceq), net income (ib), deferred taxes (txdi), dividends (dvp), revenue (sale) or depreciation (dp) is missing.
EPSRevisions	Earnings forecast revisions	Chan et al	1996	1977	1992	Define revisions as the change in the mean earnings estimate (meanest) for the next quarter from month t-1 to t, scaled by stock price in month t-1. REV6 is the sum of that variable from months t-6 to t.
ExchSwitch	Exchange Switch	Dharan and Ikenberry	1995	1962	1990	Binary variable equal to 1 if a firm switched from AMEX or NASDAQ to NYSE within the past year, or from NASDAQ to AMEX within the past year.
ExclExp	Excluded Expenses	Doyle, Lundholm and Soliman	2003	1988	1999	Difference between unadjusted earnings (EPSActualUnadj) from IBES and quarterly earnings per share (epspiq). Exclude the highest and lowest 1% of values.
FailureProbability	Failure probability	Campbell et al	2008	1981	2003	Failure probability is -9.16 -.058*PRICE + .075*MB - 2.13*CASHMTA -.045*RSIZE + 1.41*IdioRisk - 7.13*EXRETAVG + 1.42*TLMTA - 20.26*NIMTAAVG. PRICE is log(min(abs(prc), 15)); MB is shrout*abs(prc)/ceqq; CASHMTA is cheq/(shrout*abs(prc) + ltq); RSIZE is log(shrout*abs(prc)/ sum of shrout*abs(prc) for the largest 500 companies each month); IdioRisk is defined above, EXRETAVG is the weighted average excess return (log(1 + ret) - log(1 + mktrf)) over the previous 12 months, with weight on month t-j being $\phiĵ$ and the sum scaled by $\frac1-\phi1-\phiÎ2$; TLMTA is total liabilities (ltq/(shrout*abs(prc)); NIMTAAVG is a weighted average of net income over total assets (ibq/(shrout*abs(prc) + ltq)) over four quarters, with weight $\phiq̂$ on quarter $t-q$ and the sum scaled by $\frac1-\phi3̂1-\phiÎ2$. $\phi = 2≏\frac13$. All input variables are winsorized at the 5th and 95th percentile. Exclude if price less than 1.
FirmAgeMom	Firm Age - Momentum	Zhang	2004	1983	2001	6 month return, restricted to the bottom quintile of the cross-sectional firm age distribution. Exclude if price less than 5 or firm younger than 12 months.
ForecastDispersion	EPS Forecast Dispersion	Diether, Malloy and Scherbina	2002	1976	2000	Standard deviation of earnings estimates (stdev_est) scaled by mean earnings estimate.
GIndex	Governance Index	Gompers et al	2003	1990	1999	Index available from http://fac ulty.som.yale.edu/andrewmetrick/data.html . The index is only available every 2-3 years for each firm, we replace intermediate missing values with the latest available one. Value-weighted.
GrAdExp	Growth in advertising expenses	Lou	2014	1974	2010	Log of advertising expense (xad) minus log of advertising expense last year. Exclude if price less than 5, xad less than .1 or stock in the lowest decile of market value of equity.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
GrCAPX	Change in capex (two years)	Anderson and Garcia-Feijoo	2006	1976	1999	Growth rate of capital expenditures (capx) relative to two years ago. If capx is missing, replace with annual change in property, plant and equipment (ppent).
GrEmp	Employment growth	Bazdresch, Belo and Lin	2014	1965	2010	Change in number of employees (emp) between t -1 and t, scaled by average number of employees in t-1 and t. Replace hire with 0 if emp or lagged emp is missing.
GrGMToGrSales	Gross Margin growth over sales growth	Abarbanell and Bushee	1998	1974	1988	Define gross margin GM as revenue (sale) minus cost of goods sold (cogs). GrGMToGrSales is the percentage growth of GM relative to average GM in years t-1 and t-2, minus the percentage growth of revenue relative to average revenue in years t-1 and t-2. Replace growth rates with growth relative to the previous year only if data for t-2 are not available.
GrLTNOA	Growth in Long term net operating assets	Fairfield et al	2003	1964	1993	Annual growth in net operating assets, minus accruals. Net operating assets are (rect + invt + ppent + aco + intan + ao- ap- lco- lo) / at. Accruals are (rect-l12.rect + invt - l12.invt + aco - l12.aco - (ap - l12.ap + lco - l12.lco) - dp) / ((at + l12.at)/2)
GrossProf	gross profits / total assets	Novy-Marx	2013	1963	2010	Revenue (sale) - cost of goods solds (cogs), divided by 12 months lagged total assets.
GrSaleToGrInv	Sales growth over inventory growth	Abarbanell and Bushee	1998	1974	1988	Percentage growth in sales (sale) relative to average sales of t-1 and t-2, minus percentage growth in inventory (invt) relative to average inventory of t-1 and t-2. Both growth terms are calculated relative to t-1 only if t-2 is missing.
GrSaleToGrOverhead	Sales growth over overhead growth	Abarbanell and Bushee	1998	1974	1988	GrSaleToGrOverHead = Percentage growth in sales (sale) relative to average sales of t-1 and t-2, minus percentage growth in administrative expenses (xsga) relative to average administrative expenses of t-1 and t-2. Both growth terms are calculated relative to t-1 only if t-2 is missing. Remove if in the highest quintile of GrSaleToGrOverHead. Returns are nicely monotonic until the highest quintile, consistent with original paper's rank regressions.
Herf	Industry concentration (Herfindahl)	Hou and Robinson	2006	1963	2001	Three-year rolling average of the three digit industry Herfindahl index based on firm revenue (sale). Exclude regulated industries (4011, 4210, 4213 & year $\leq$ 1980; 4512 & year $\leq$ 1978, 4812, 4813 & year $\leq$ 1982, 4900-4999 in any year)
High52	52 week high	George and Hwang	2004	1963	2001	Let temphigh = price / by the maximum daily price over the past twelve months. High52 is the rolling 6 month average of temphigh to simulate the original paper's 6-month holding periods
IdioRisk	Idiosyncratic risk	Ang et al	2006	1963	2000	Standard deviation of residuals from CAPM regressions using the past month of daily data. Value weighted
Illiquidity	Amihud's illiquidity	Amihud	2002	1964	1997	Past twelve month average of: daily return (abs(ret)) divided by turnover((abs(prc)*vol)

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
IndIPO	Initial Public Offerings	Ritter	1991	1975	1984	1 if IPO in the past 6-36 months. 0 otherwise. IPO dates are taken from Jay Ritter's IPO data available at: http://bear.warrington.ufl.edu/ritter/ipodata.htm. Missing IPO dates imply IndIPO = 0
IndMom	Industry Momentum	Grinblatt and Moskowitz	1999	1963	1995	Weighted average of firm-level 6 month buy-and-hold return. Average is taken over two digit industries each month and weights are based on market value of equity.
IndRetBig	Industry return of big firms	Hou	2007	1972	2001	Average monthly return (ret) of the 30% largest companies by market value of equity in the same Fama-French 48 industry. Exclude the largest 30% of companies for IndRetBig (not to compute the anomaly!)
IntanBM	Intangible return	Daniel and Titman	2006	1968	2003	In each month, run a cross-sectional regression of a firm's five-year stock return on 5 year lagged BM (defined above) and a constructed regressor that is the change in BM from 5 years ago to today plus the five-year stock return. The residual from that regression is IntanBM.
IntanCFP	Intangible return	Daniel and Titman	2006	1968	2003	In each month, run a cross-sectional regression of a firm's five-year stock return on the 5 year lagged CFP = (net income (ni) plus depreciation (dp))/market value of equity and a constructed regressor that is the change in CFP from 5 years ago to today plus the five-year stock return. The residual from that regression is IntanCFP.
IntanEP	Intangible return	Daniel and Titman	2006	1968	2003	In each month, run a cross-sectional regression of a firm's five-year stock return on the 5 year lagged EP = net income (ni)/market value of equity and a constructed regressor that is the change in EP from 5 years ago to today plus the five-year stock return. The residual from that regression is IntanEP.
IntanSP	Intangible return	Daniel and Titman	2006	1968	2003	In each month, run a cross-sectional regression of a firm's five-year stock return on 5 year lagged SP (defined above) and a constructed regressor that is the change in SP from 5 years ago to today plus the five-year stock return. The residual from that regression is IntanSP.
IntMom	Intermediate Momentum	Novy-Marx	2012	1926	2010	Stock return between months t-12 and t-6
Investment	Investment	Titman et al	2004	1973	1996	Ratio of capital investment (capx) to revenue (revt) divided by the firm-specific 36-month rolling mean of that ratio. Exclude if revenue less than \$10m.
IO_ShortInterest	Institutional Ownership for stocks with high short interest	Asquith, Pathak and Ritter	2005	1980	2002	Exclude all stocks with short interest (ShortInterest) below .025. IO_ShortInterest is institutional ownership (instown_perc). Keep NYSE Only.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
KZ	Kaplan Zingales index	Lamont et al	2001	1968	1997	-1.002* (net income (ni) + depreciation (dp))/total assets (at) + .283*(total assets (at) + market value of equity - book value of equity (ceq) - deferred taxes (txdi))/total assets (at) + 3.319*(debt in current liabilities (dlc) + long-term debt (dltt))/(debt in current liabilities + long-term debt + book value of equity) - 39.368*(Dividends (divamt)/total assets) - 1.315*(cash and short-term investments (che)/total assets). Replace txdi and divamt with 0 if missing.
Leverage	Market leverage	Bhandari	1988	1946	1981	Total liabilities (lt) divided by market value of equity.
MaxRet	Maximum return over month	Bali et al	2010	1962	2005	Maximum of daily returns (ret) over the previous month
Mom12m	Momentum (12 month)	Jegadeesh and Titman	1993	1964	1989	Stock return between months t-12 and t-1.
Mom1813	Momentum-Reversal	De Bondt and Thaler	1985	1933	1980	Stock return between months t-18 and t-13.
Mom1m	Short term reversal	Jegedeesh	1989	1934	1987	Stock return (ret) over the previous month.
Mom36m	Long-run reversal	De Bondt and Thaler	1985	1926	1982	Stock return between months t-36 and t-13.
Mom6m	Momentum (6 month)	Jegadeesh and Titman	1993	1964	1989	Stock return between months t-6 and t-1. Exclude if price less than 5.
Mom6mJunk	Junk Stock Momentum	Avramov et al	2007	1985	2003	Mom6m. Include only stocks with a credit rating (splticrm) of BBB or lower
MomRev	Momentum and LT Reversal	Chan and Kot	2006	1965	2001	Binary variable equal to 1 if firm is in the highest Mom6m quintile and the lowest Mom36m quintile, and equal to 0 if firm is in the lowest Mom6m quintile and the highest Mom36m quintile. Exclude if price less than 5.
MomVol	Momentum and Volume	Lee and Swaminathan	2000	1965	1995	Mom6m. Include only stocks in the highest quintile of average trading volume (vol) over the previous 6 months. Exclude NASDAQ stocks, if price less than 1 or if stock has been trading for less than 24 months.
Mscore	Mohanram G-score	Mohanram	2005	1978	2001	Examine only stocks in lowest BM quintile. Binary variable based on sum of eight indicator variables which are: 1 if return on assets (ni/average assets) above the two digit industry median; 1 i net cash flow to assets (oancf/average assets) above the two digit indstry median; 1 if net cash flow greater than net income; 1 if R&D expense to assets (xrd/average assets) greater than two digit industry median; 1 if capital expenditure (capx/average assets) greater than two digit industry median; 1 if advertising expenses (xad/average assets) greater than two digit industry median; 1 if the volatility of net income over the past 3 years is below the two digit industry median, 1 if the volatility of revenue (revt) over the past 3 years is below the two digit industry median. The final variable is equal to 1 if the sum of the above 8 indicators is greater than 5 and 0 if the sum is less than 2.
NetDebtFinance	Net debt financing	Bradshaw et al	2006	1971	2000	Long-term debt issuance (dltis) minus long-term debt reduction (dltr) minus current debt changes (dlcch), scaled by average total assets (at) in years t-1 and t. Replace missing values of dlcch with 0. Exclude if ratio is greater than 1.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
NetDebtPrice	Net debt to price	Penman Richardson Tuna	2007	1961	2001	Long-term debt (dltt) plus debt in current liabilities (dlc) plus preferred stock (pstk) plus preferred dividends in arrears (dvpa) minus treasury stock (tstkp) minus cash and short-term investments (che), scaled by market value of equity. Exclude if SIC between 6000 and 6999, or if missing value for total assets (at), net income (ib), common shares outstanding (csho), book value of equity (ceq) or price close fiscal year (prcc_f). Keep only 3rd B/M Quintile, following Table 4 (and in contrast to Table 1).
NetEquityFinance	Net equity financing	Bradshaw et al	2006	1971	2000	Sale of common stock (sstk) minus purchase of common stock (prstkc), scaled by average total assets (at) from years t and t-1. Exclude if absolute value of ratio is greater than 1.
NetPayoutYield	Net Payout Yield	Boudoukh et al	2007	1984	2003	Dividends (dvc) plus purchase of common and preferred stock (prstkc) minus sale of common and preferred stock (sstk), divided by market value of equity.
NOA	Net Operating Assets	Hirshleifer et al	2004	1964	2002	Difference between operating assets and operating liabilities, scaled by lagged total assets. Operating assets are total assets (at) minus cash- and short-term investments (che), operating liabilities are total assets minus long-term debt (dltt), minority interest (mib), deferred charges (dc) and book equity (ceq).
NumEarnIncrease	Number of consecutive earnings increases	Loh and Warachka	2012	1987	2009	Number of 4-quarter net income (ibq) increases over the previous 2 years.
OperLeverage	Operating Leverage	Novy-Marx	2010	1963	2008	Sum of administrative expenses (xsga) and cost of goods sold (cogs), scaled by total assets (at). Use xsga = 0 if xsga is missing.
OperProf	operating profits / book equity	Fama and French	2006	1977	2003	Revenue (revt) minus cost (cogs) - administrative expenses (xsga) - interest expenses (xint), scaled by book value of equity (ceq). Exclude smallest size tercile.
OptVol	Option Volume to Stock Volume	Johnson and So	2012	1996	2010	Total monthly option volume (volume) over all puts and calls, divided by monthly stock trading volume (vol). Exclude if price less than 1 or share code greater 11 or option volume or stock volume data are missing for the previous month.
OrderBacklog	Order backlog	Rajgopal et al	2003	1981	1999	Order backlog (ob) divided by average total assets (at) in years t-1 and t. Exclude if order backlog is 0.
OrgCap	Organizational Capital	Eisfeldt and Papanikolaou	2013	1970	2008	Defined recursively. Initialize with OrgCap = 4*general expenses (xsga) in the first year, and calculate as .85*OrgCap previous year + xsga current year thereafter. Scale by total assets (at).

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
OScore	O Score	Dichev	1998	1981	1995	OScore = -1.32 - .407*log(at/GNP deflator) + 6.03*(lt/at) - 1.43*((act - lct)/at) + .076*(lct/act) - 1.72*I(lt > at) - 2.37*(ib/at) - 1.83*(fopt/lt) + .285*(ib + ib$_t-12$ + ib$_t-24$ < 0) - .521*((ib - ib$_t-12$)/(abs(ib) + .abs(ib$_t-12$))). fopt = oancf if fopt is missing. Exclude Exclude if SIC code between 3999 and 4999, or greater than 5999. Exclude if price less than 5. Then exclude if OScore is in bottom quintile of OScore (original paper shows non-monotonic returns, as does our replication)
PayYield	Payout Yield	Boudoukh et al	2007	1984	2003	Sum of dividends (dvc), purchase of common and preferred stock (prstkc) and max(preferred stock redemption value (pstkrv), 0), divided by lag(market value of equity, 6 months). Exclude if PayoutYield $\leq$ 0.
PctAcc	Percent Operating Accruals	Hafzalla et al	2011	1989	2008	Income before extraordinary items (ib) minus net cash flow (oancf) divided by absolute value of ib. If oancf is missing, PctAcc is defined as ((act - act$_t-12$) - (che - che$_t-12$) - ((lct - lct$_t-12$) - (dlc - dlc$_t-12$) - (txp - txp$_t-12$) - dp))/abs(ib). In either case, if ib is equal to 0, divide by .01 instead. Exclude if price less than 5.
PctTotAcc	Percent Total Accruals	Hafzalla et al	2011	1989	2008	Net income (ni) minus (purchase of common and preferred stock (prstkcc) minus sale of common and preferred stock (sstk) plus dividends (dvt), cash flow from operations (oancf), from financing (fincf) and investment (ivncf)). Scaled by absolute value of net income.
PensionFunding	Pension Funding Status	Franzoni and Marin	2006	1980	2002	FR = (FVPA - PBO), scaled by market value of equity. FVPA is pbnaa from 1980 to 1986, pplao + pplao from 1987 to 1997, and pplao after 1997. PBO is pbnvv from 1980 to 1986, pbpro + pbpru from 1987 to 1997, and pbpro after 1997. Exclude if price less than 5 or shrcd > 11.
Price	Price	Blume and Husic	1972	1932	1971	Log of absolute value of price (prc).
PriceDelay	Price delay	Hou and Moskowitz	2005	1964	2001	Regress daily stock return (ret) on market return (mktrf) in $t, t-1, \ldots, t-4$ with observations over the previous year. Trim the highest and lowest 1% of estimated coefficients. Define PriceDelay as the ratio of 1*beta on mktrf$_t-1$ + 2*beta on mktrf$_t-2$ + 3*beta on mktrf$_t-3$ + 4*beta on mktrf$_t-4$, and beta on mktrf$_t$ + beta on mktrf$_t-1$ + beta on mktrf$_t-2$ + beta on mktrf$_t-3$ + beta on mktrf$_t-4$. The final variable is the average of that ratio over the previous month.
Profitability	earnings / assets	Balakrishnan, Bartov and Faurel	2010	1976	2005	Quarterly earnings per share (epspxq) times quarterly shares outstanding used to calculate EPS (cshprq) divided by total assets (at). Exclude if price less than 1.
ProfitMargin	Profit Margin	Soliman	2008	1984	2002	Net income (ni) over revenue (revt). Exclude if price less than 5.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
Pscore	Piotroski F-score	Piotroski	2000	1976	1996	Sum of nine indicator variables which are: 1 if net income (ib) greater 0; 1 if net cash flow (oancf) greater 0; 1 if return on assets (ib/at) increased relative to previous year; 1 if net cash flow greater net income; 1 if long-term debt to assets (dltt/at) declined over the previous year; if current assets to current liabilities (act/lct) increased over the previous year; 1 if ebit/sale (ebit = ib + txt + xint) increased over the previous year; 1 if revenue to assets increased over the previous year; 1 if shrout $\leq$ shrout last year. Include highest quintile of book-to-market only. Exclude if missing any of the input variables.
RD	R&D over market cap	Chan et al	2001	1975	1995	R&D expense (xrd) over market value of equity.
RDIPO	IPO and no R&D spending	Gou et al	2006	1980	1995	Binary variable equal to 1 if R&D expense (xrd) = 0 and IndIPO = 1. 0 otherwise.
RDirtSurp	Real dirty surplus	Landsman, Miller, Peasnell and Yeh	2011	1976	2003	Define Dirty Surplus as annual change in marketable securities adjustment msa plus annual change in retained earnings adjustment (recta) + .65 times the annual change in min(Unrecognized prior service cost (pcupsu) - Pension additional minimum liability (paddml),0). Real dirty surplus is the annual change in book equity (ceq) minus dirty surplus minus (net income (ni) minus dividends preferred (dvp)) + dividends (divamt) - end-of-fiscal-year-stock-price (prcc_f)*annual change in common shares outstanding (csho).
RealEstate	Real estate holdings	Tuzel	2010	1971	2005	Industry-adjusted value of real estate holdings. Real estate holdings are calculated as: PPE Buildings at cost (fatb) plus PPE Leases at cost (fatl), divided by PPE (ppegt). Use ppent if ppegt is missing. Subtract monthly industry-mean at the 2 digit SIC level.
RetConglomerate	Conglomerate return	Cohen and Lou	2012	1977	2009	Identify conglomerate firms as those with multiple OPSEG or BUSSEG entries in the Compustat segment data (and require that at least 80% of firm's total assets are covered by segment data). Compute monthly stock return at the 2-digit SIC level for stand-alone (non-conglomerate) firms only, and match those returns to conglomerates' segments. Compute weighted conglomerate return as the industry return of stand-alone companies, weighted with a conglomerate's total sales in each industry.
RevenueSurprise	Revenue Surprise	Jegadeesh and Livnat	2006	1987	2003	Define revenue per share as quarterly revenue (revtq) divided by quarterly common shares outstanding (cshprq). RevenueSurprise is the 4-quarter change in revenue per share minus the average 4-quarter change in revenue per share over the previous 2 years. RevenueSurprise is scaled by its standard deviation over the previous 2 years. Exclude if price less than 5.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
RIO_BM	Inst Own and BM	Nagel	2005	1980	2003	Residual institutional ownership (RIO) is defined as log(institutional ownership (instown_perc)/(1-institutional ownership)) + 23.6 - 2.89*log(market value of equity) +.08*log(market value of equity)$2̂$. Replace instown_perc with 0 if it is missing, with .9999 if it's above .9999, and with .0001 if it's below .0001. RIO_BM is a binary variable equal to 1 if a firm is in the highest quintile of the monthly RIO distribution and has BM below the cross-sectional median, and 0 if a firm is in the lowest quintile of RIO and has BM below the median.
RIO_Disp	Inst Own and Forecast Dispersion	Nagel	2005	1980	2003	Binary variable equal to 1 if RIO (defined above) is in the highest quintile and ForecastDispersion (defined above) is above the median, 0 if RIO is in the lowest quintile and ForecastDispersion is above the median.
RIO_IdioRisk	Inst Own and Idio Vol	Nagel	2005	1980	2003	Binary variable equal to 1 if RIO (defined above) is in the highest quintile and monthly IdioRisk (defined above) is above the median, 0 if RIO is in the lowest quintile and IdioRisk is above the median.
RIO_Turnover	Inst Own and Turnover	Nagel	2005	1980	2003	Binary variable equal to 1 if RIO (defined above) is in the highest quintile and monthly turnover (vol/shrout) is above the median, 0 if RIO is in the lowest quintile and turnover is above the median.
RoE	net income / book equity	Haugen and Baker	1996	1979	1993	Net income (ni) over book value of equity (ceq). Exclude if price less than 5.
SalesGr	Revenue Growth Rank	Lakonishok et al	1994	1968	1990	Rank firms by their annual revenue growth each year over the past 5 years. MeanRankRevGrowth is the weighted average of ranks over the past 5 years, that is, MeanRankRevGrowth = (5*Rank$_t-1$ + 4*Rank$_t-2$ + 3*Rank$_t-3$ + 2*Rank$_t-4$ + 1*Rank$_t-5$)/15. Exclude NASDAQ stocks.
SalesToPrice	Sales-to-price	Barbee et al	1996	1979	1991	Ratio of annual sales (sale) to market value of equity.
Seasonality	Return Seasonality	Heston and Sadka	2008	1965	2002	Average return in the same month over the preceding 5 years. Exclude NASDAQ stocks.
SEO	Public Seasoned Equity Offerings	Loughran and Ritter	1995	1975	1984	Binary variable equal to 1 if seasoned equity offering within the previous 12 months. SEO data are from SDC.
ShareIs1	Share issuance (5 year)	Daniel and Titman	2006	1968	2003	5-year growth in number of shares. Number of shares is calculated as shrout/cfacshr to adjust for splits.
ShareIs5	Share issuance (1 year)	Pontiff and Woodgate	2008	1970	2003	Growth in number of shares between t-18 and t-6. Number of shares is calculated as shrout/cfacshr to adjust for splits.
ShareRepurchase	Share repurchases	Ikenberry, Lakonishok and Vermaelen	1995	1980	1990	Binary variable equal to 1 if stock repurchase indicated in cash flow statement (prstkc > 0), and 0 if prstkc = 0.
ShareVol	Share Volume	Datar Naik Radcliffe	1998	1962	1991	Sum of monthly share trading volume (vol) over the previous three months, scaled by 3 times common shares outstanding (shrout). Drop if ShareVol is below its median

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
ShortInterest	Short Interest	Dechow et al	2001	1976	1993	Short-interest from Compustat (shortint) scaled by shares outstanding (shrout). Short-interest data are available bi-weekly with a four day lag. We use the mid-month observation to make sure data would be available in real time.
SinStock	Sin Stock (selection criteria)	Hong and Kacperczyk	2009	1926	2006	Using Compustat Segment data, sinAlgo is defined as a binary variable equal to 1 if at least one segment of a firm is listed as being in at least one of the following industries: sic $\geq$ 2100 & sic $\leq$ 2199, sic $\geq$2080 & sic $\leq$ 2085, NAICS in \7132, 71312, 713210, 71329, 713290, 72112, 721120\. As in the original paper, we assume that the sin stock indicator applies to the entire history and future of the identified firm. sinAlgo is equal to 0 if the firm is not identified in the CS Segment data as a sin stock and if the firm is in one of the following industries: (sic $\geq$ 2000 & sic $\leq$ 2046) OR (sic $\geq$ 2050 & sic $\leq$ 2063) OR (sic $\geq$ 2070 & sic $\leq$ 2079) OR (sic $\geq$ 2090 & sic $\leq$ 2092) OR (sic $\geq$ 2095 & sic $\leq$ 2099) OR (sic $\geq$ 2064 & sic $\leq$ 2068) OR (sic $\geq$ 2086 & sic $\leq$ 2087) OR (sic $\geq$ 920 & sic $\leq$ 999) OR (sic $\geq$ 3650 & sic $\leq$ 3652) OR sic == 3732 OR (sic $\geq$ 3931 & sic $\leq$ 3932) OR (sic $\geq$ 3940 & sic $\leq$ 3949) OR (sic $\geq$ 7800 & sic $\leq$ 7833) OR (sic $\geq$ 7840 & sic $\leq$ 7841) OR (sic $\geq$ 7900 & sic $\leq$ 7911) OR (sic $\geq$ 7920 & sic $\leq$ 7933) OR (sic $\geq$ 7940 & sic $\leq$ 7949) OR sic $==$ 7980 OR (sic $\geq$ 7990 & sic $\leq$ 7999)
Size	Size	Banz	1981	1926	1975	Log of monthly market value of equity (abs(prc)*shrout)).
Skew1	Volatility smirk	Xing, Zhang and Zhao	2010	1996	2005	Using OptionMetrics data, among options with duration between 10 and 60 days, implied volatility of put option with moneyness closest to but above 1 minus implied volatility of call option with moneyness closest to but below 1.
SmileSlope	Slope of smile	Yan	2011	1996	2005	Using OptionMetrics data, average implied volatility of put options with duration between 15 and 30 days and rounded delta of -.5 minus average implied volatility of call options with duration between 15 and 30 days and rounded delta of .5.
Spinoff	Spinoffs	Cusatis et al	1993	1965	1988	Spinoffs are identified as all observations in the CRSP acquisition file with valid acperm entry. Spinoff is a binary variable equal to 1 if a firm is identified in the CRSP Acquisition data and if it has at most one year of history in the CRSP stock return data. Spinoff is equal to 0 otherwise.
StdTurnover	Turnover volatility	Chordia Roll Subrahmanyam	2001	1966	1995	Standard deviation of turnover (vol/shrout) over the past 36 months.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
SurpriseRD	Unexpected R&D increase	Eberhart et al	2004	1974	2001	Binary variable equal to 1 if: R&D (xrd) scaled by revenue (revt) is positive, R&D scaled by total assets (at) is positive, annual R&D growth is greater than 5%, annual growth in R&D over total assets is greater than 5%. SurpriseRD is 0 otherwise.
Tangibility	Tangibility	Hahn and Lee	2009	1973	2001	Cash and short-term investments (che) plus .715*receivables (rect) + .547*inventory (invt) + .535* property, plant and equipment (ppent), scaled by total assets (at). Only defined for manufacturing firms (SIC $\geq$ 2000 and SIC <4000). Exclude the lowest tercile of manufacturing firms by total assets.
Tax	Taxable income to income	Lev and Nissim	2004	1973	2000	Ratio of Taxes paid and tax share of net income. Numerator is defined as the sum of foreign (txfo) and federal (txfed) income taxes. If either one is missing, numerator is defined as total taxes (txt) minus deferred taxes (txdi). Denominator is the product of the prevailing tax rate and net income (ib). Tax rate is .48 before 1979, .46 from 1979 to 1986, .4 in 1987, .34 between 1988 and 1992 and .35 from 1993 onwards. If net income is negative, and the numerator is positive, tax is defined as 1. Exclude if price less than 5.
UpForecast	Up Forecast	Barber Lehavy MicNichols Trueman	2001	1985	1997	Binary variable equal to 1 if mean analyst earnings forecast for the next quarter (meanest) has improved over the previous month, and 0 otherwise.
VarCF	Cash-flow variance	Haugen and Baker	1996	1979	1993	Rolling variance of (ib+dp)/mve_c over the past 60 months (minimum 24 months data required).
VolMkt	Volume to market equity	Haugen and Baker	1996	1979	1993	Average monthly dollar trading volume (vol*abs(prc)) over the previous 12 months, scaled by market value of equity. Exclude if price less than 5.
VolSD	Volume Variance	Chordia Roll Subrahmanyam	2001	1966	1995	Rolling standard deviation of monthly trading volume (vol) over the past 36 months (require at least 24 observations). Include only NYSE stocks.
VolumeTrend	Volume Trend	Haugen and Baker	1996	1979	1993	Rolling coefficient from regressing monthly trading volume on a linear time trend over a window of 60 months (require that at least 30 exist). Scale coefficient by 60-month average of trading volume.
XFIN	Net external financing	Bradshaw et al	2006	1971	2000	Sale of common stock (sstk) minus dividends (dv) minus purchase of common stock (prstkc) plus long-term debt issuance (dltis) minus long-term debt reductions (dltr). Scaled by total assets (at).
ZeroTrade	Days with zero trades	Liu	2006	1960	2003	In each month, count the number of days with no trades. Define zerotrade as the number of days without trades plus (the sum of monthly turnover (vol/shrout) divided by 48*10$5̂$), multiplied by 21/number of trading days per month. Zerotrade is the 6-month average of that variable.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
ZScore	Altman Z-Score	Dichev	1998	1981	1995	1.2*(current assets (act) - current liabilities (lct))/total assets (at) + 1.4*(Retained earnings (re)/total assets (at)) + 3.3*(net income (ni) + interest expense (xint) + total taxes (txt))/total assets (at) + .6*(market value of equity/Total liabilities (lt)) + revenue (revt)/ total assets (at). Include only NYSE stocks. Exclude if SIC code between 4000 and 4999, or above 5999. Exclude if ZScore is in bottom quintile of ZScore (original paper shows non-monotonic returns, as does our replication)

A.2. Details of the Maximum Likelihood Estimation

The likelihood is a bit tricky to write down as a result of publication bias. The likelihood of a observing a pair (r_i, σ_i) needs to be conditioned on publication:

$$f_{r,\sigma|\text{pub}}(r_i, \sigma_i|\text{pub}_i, \theta) = \frac{p(r_i/\sigma_i|\theta) f_{r|\sigma}(r_i|\sigma_i, \theta) f_N(\log\sigma_i|\mu_\sigma, \sigma_\sigma)}{\int d\tilde{\sigma} \left[\int d\tilde{r}\, p(\tilde{r}/\tilde{\sigma}|\theta) f_{r|\sigma}(\tilde{r}|\tilde{\sigma}, \theta_\mu)\right] f_N(\log\tilde{\sigma}|\mu_\sigma, \sigma_\sigma)}, \quad (25)$$

where, as before, f_N and f_τ are the normal and scaled student's t densities, and $f_{r|\sigma}$ is the conditional density of $r|\sigma$. $f_{r|\sigma}$ is found by convolution

$$f_{r|\sigma}(r|\sigma, \theta_\mu) = \int d\tilde{\mu} f_\tau(\tilde{\mu}|\mu_\mu, \sigma_\mu, \nu_\mu) f_N(r|\tilde{\mu}, \sigma). \quad (26)$$

The numerator of Equation (25) is intuitive: Due to publication bias, the likelihood of observing a pair r_i, σ_i includes not only the densities of σ_i and $r_i|\sigma_i$, but also the probability of passing the statistical requirements for publication $p(r_i/\sigma_i|\theta)$. The denominator comes from the fact that, since some portfolios are not published, we need to renormalize the density and make sure it integrates to 1.

We evaluate the convolution in the numerator by standard numerical quadrature. The denominator of the likelihood involves three integrals, which is tricky to do using traditional methods. Thus, we compute the denominator by monte carlo.

Another issue in the estimation is that the fat tail parameter ν_μ has non-smooth derivatives, which tends to make standard optimizers fail. To overcome this problem, we optimize by iterating between a quasi-newton method for all parameters besides ν_μ, and using a more robust golden section search based algorithm for ν_μ. The iteration stops when the likelihood stops updating. We find this algorithm to be quite robust, and far outperforms starting simplex optimizers at various points, for example.

The last issue is that derivative-based standard errors may not work well with the fat tail parameter. Indeed, we find that the Hessian standard error underestimates uncertainty in ν_μ in simulated data. To overcome this issue, we calculate standard errors by bootstrap.

A.3. Additional Estimation Figures

Figure A.1: Pairwise Correlations. This histogram shows the distribution of pairwise correlations in our database of monthly long-short portfolio returns.

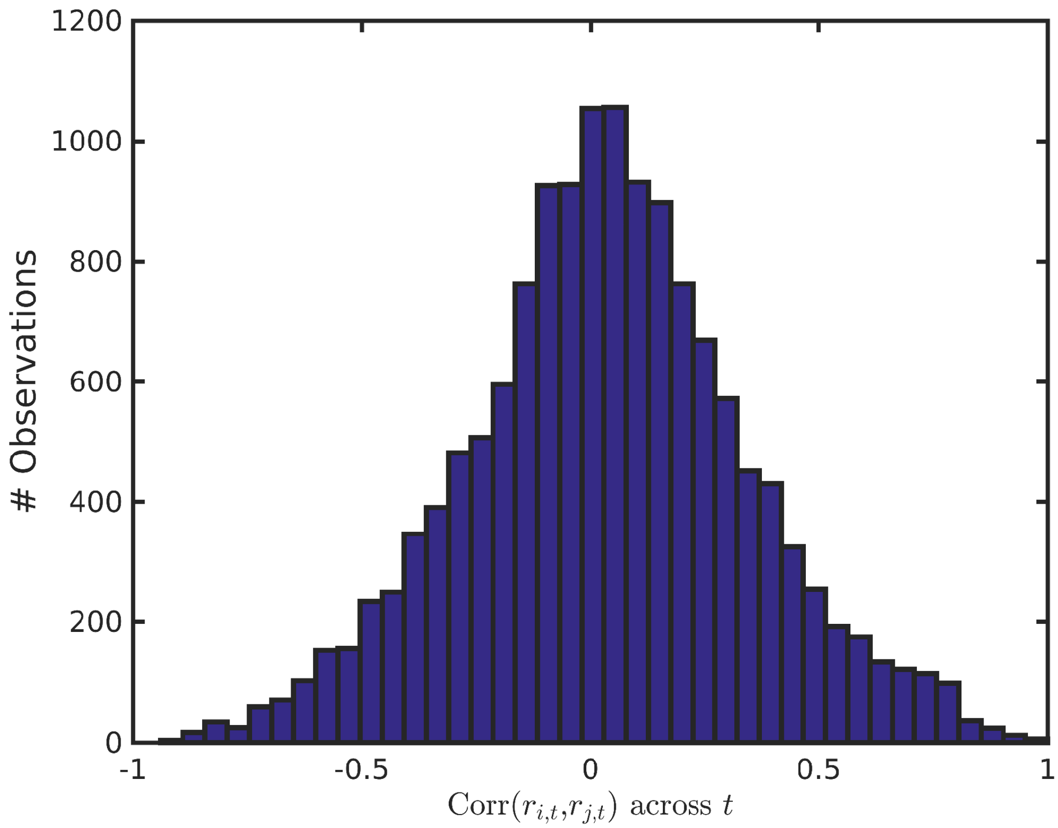


Figure A.2: Likelihood function. This figure plots the likelihood function near our maximum likelihood estimate (Table 4).

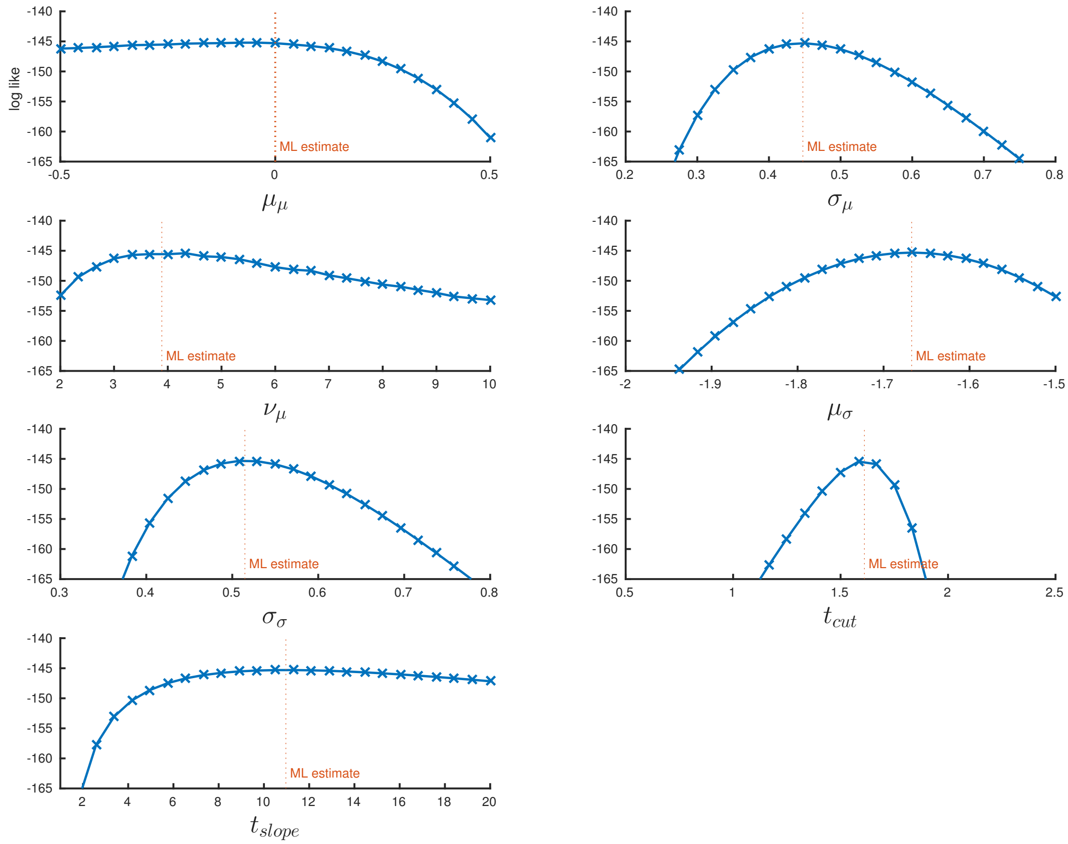


Figure A.3: Bootstrapped distribution of mean and median shrinkage. This figure plots details of the mean shrinkage standard errors in Table 4.

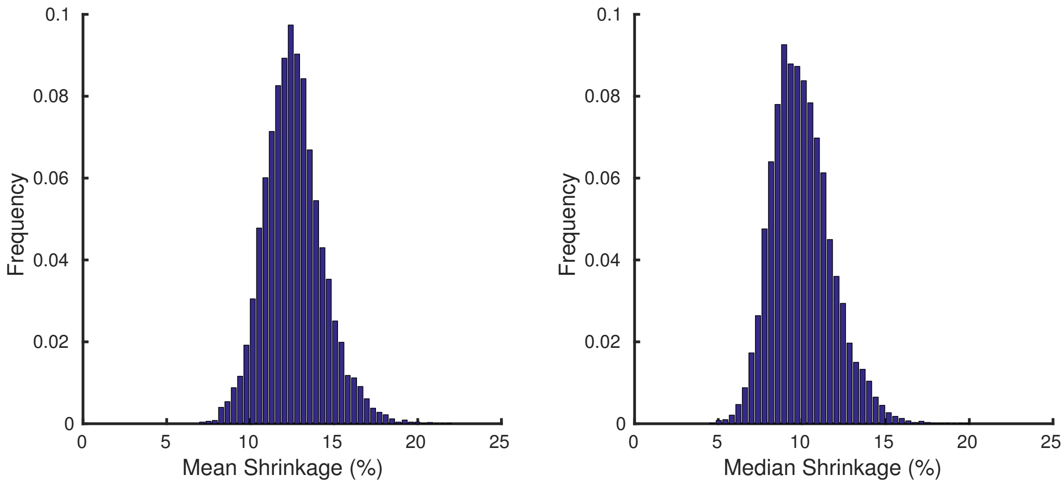


Table A.2: Estimations under Alternative Assumptions for μ_μ

This table shows that assuming alternative values for the mean of true returns μ_μ does not affect our headline shrinkage. We run the same estimation as in Table 4 column "all" but assume various values of μ_μ. "Baseline" reprints the "all" column of Table 4 for comparison. The mean shrinkage is small regardless of the assumption for μ_μ. An alternative definition for mean shrinkage that takes averages before taking ratios (mean $\hat{\mu}_i$ / mean r_i) finds similar shrinkage for all μ_μ assumptions. The similar shrinkage happens because the estimated σ_μ decreases as one increases μ_μ. Intuitively, if the average true return is high, then one needs little dispersion to match the positive published true returns. This robustness is consistent with the generality of James and Stein (1961) shrinkage, which improves on the sample mean regardless of the choice of μ_μ. The log-likelihood suggests that μ_μ close to zero is ideal, however.

		Baseline	Alternative μ_μ			
Assumed Parameters						
μ_μ	mean of true returns	0.00	-0.40	-0.20	0.20	0.40
Estimated Parameters						
σ_μ	dispersion of	0.45	0.56	0.55	0.36	0.31
	true returns	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)
ν_μ	fat tail (d.o.f.) of	3.89	5.96	6.57	2.70	2.37
	true returns	(1.38)	(2.14)	(1.92)	(0.70)	(0.39)
μ_σ	mean of log	-1.67	-1.67	-1.67	-1.66	-1.68
	standard error	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)
σ_σ	std of log	0.51	0.51	0.51	0.52	0.52
	standard error	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
t_{cut}	midpoint of	1.61	1.62	1.61	1.60	1.56
	t-stat threshold	(0.08)	(0.09)	(0.07)	(0.07)	(0.06)
t_{slope}	slope of	10.97	10.90	11.09	11.02	11.24
	t-stat threshold	(41.78)	(42.06)	(41.96)	(41.99)	(42.77)
Estimated Bias Adjustments						
	Mean $\hat{\mu}_i$ / Mean r_i	12.68	13.24	12.52	12.03	9.58
		(5.82)	(5.63)	(5.62)	(5.48)	(5.13)
	Mean Shrinkage (%)	12.46	13.63	12.31	11.01	6.49
		(1.75)	(1.84)	(1.85)	(1.33)	(0.84)
	Std Shrinkage (%)	9.47	10.48	9.39	9.18	10.08
		(1.13)	(1.40)	(1.31)	(1.00)	(1.19)
Log likelilhood		-145.46	-145.56	-145.85	-146.22	-149.39

A.4. The Benjamini-Hochberg Adjustment in Our Model

The Benjamini-Hochberg (BH) adjustment requires very few assumptions. It merely assumes that a certain, unspecified proportion of t-statistics are close to the null $N(0,1)$ distribution.

This generality comes at a cost, however. Without specifying the proportion of null t-statistics, the adjustments can only provide an upper bound on the false discovery rate. Indeed, in many cases the BH adjustment will be excessively conservative, as we illustrate in this section.

To illustrate the mechanics of the BH adjustment, it helps to derive the adjustment within the context of our model. Suppose there is a small number Δ, such that any portfolio with $\mu_i \in [-\Delta, \Delta] \approx 0$. Let's label these portfolios as null_i. These are portfolios with zero true returns, so their in-sample returns follow the traditional null distribution $r_i|\text{null}_i \sim N(0, \sigma_i)$. This leads to a binary transformation of the model of Section 3.1:

$$t_i \sim \begin{cases} \epsilon_i & \text{w/ prob} \quad Pr(\text{null}_i) \\ \frac{\mu_i}{\sigma_i} + \epsilon_i & \text{otherwise} \end{cases} \tag{27}$$

Consider the t-stat hurdle t_h. For portfolios which meet this hurdle, the false discovery rate is

$$Pr(\text{null}_i | t_i > t_h) = \frac{Pr(t_i > t_h | \text{null}_i) Pr(\text{null}_i)}{Pr(t_i > t_h)} = \frac{(1 - \Phi(t_i)) Pr(\text{null}_i)}{Pr(t_i > t_h)}. \tag{28}$$

Where $\Phi()$ is the standard normal CDF. Note that $(1 - \Phi(t_i)) = p_i$, the p-value corresponding to t_i. Also, the denominator can be estimated using its sample counterpart (assuming all narrative portfolios are observed). These facts lead to the BH adjustment

$$Pr(\text{null}_i | t_i > t_h) = \frac{Pr(\text{null}_i) p_h}{\text{Proportion of portfolios with } t_i > t_h} \tag{29}$$

$$\leq \frac{p_h}{\text{Proportion of portfolios with } t_i > t_h}. \tag{30}$$

Thus, BH is an upper bound, rather than a direct estimate of the false discovery rate. Moreover, the BH adjustment is excessively conservative if $Pr(\text{null}_i)$ is far from 1. For example, if null portfolios comprise roughly half the data (as in our estimation and in Harvey, Liu, and Zhu (2015)), then the BH FDR bound exceeds the actual FDR by a factor of 2.

The null hypothesis discussed in Section 6.1 $\mu_i \leq 0$ cannot be examined using BH's algorithm without the additional estimation of the distribution of μ_i. To see this, note that the false discovery rate for $\mu_i \leq 0$ is

$$
\begin{aligned}
Pr(\text{null}_i | t_i > t_h) &= \frac{Pr(t_i > t_h | \mu_i \leq 0) Pr(\mu_i \leq 0)}{Pr(t_i > t_h)} && (31) \\
&= \int_{-\infty}^{0} d\mu f_\mu(\mu|\theta) \left[1 - \Phi\left(t_i - \frac{\mu_i}{\sigma_i} \right) \right] \frac{Pr(\mu_i \leq 0)}{Pr(t_i > t_h)}. && (32)
\end{aligned}
$$

where $f_\mu(\mu|\theta)$ is the distribution of true means that we estimate in Section **??**.

A.5. Multiple Tests of the Null: Bias-Adjusted t-stat < 1.96

The low t-stat hurdles in Section 6.1 are due to the inadequacy of the traditional null hypothesis of $\mu_i = 0$. This null describes only a tiny portion of narrative predictors. As a result, the null is ineffective for isolating cases worthy of further study.

When the traditional null is a poor fit, one may want to use an empirical null, that is, a null which is designed to generate unusual and interesting cases. This notion of estimating a null distribution is not possible in classical single test statistics, but is common in large-scale studies (Efron (2012)).

In this section, we examine a null hypothesis which effectively generates interesting predictors. Specifically, we define a null predictor as one that satisfies

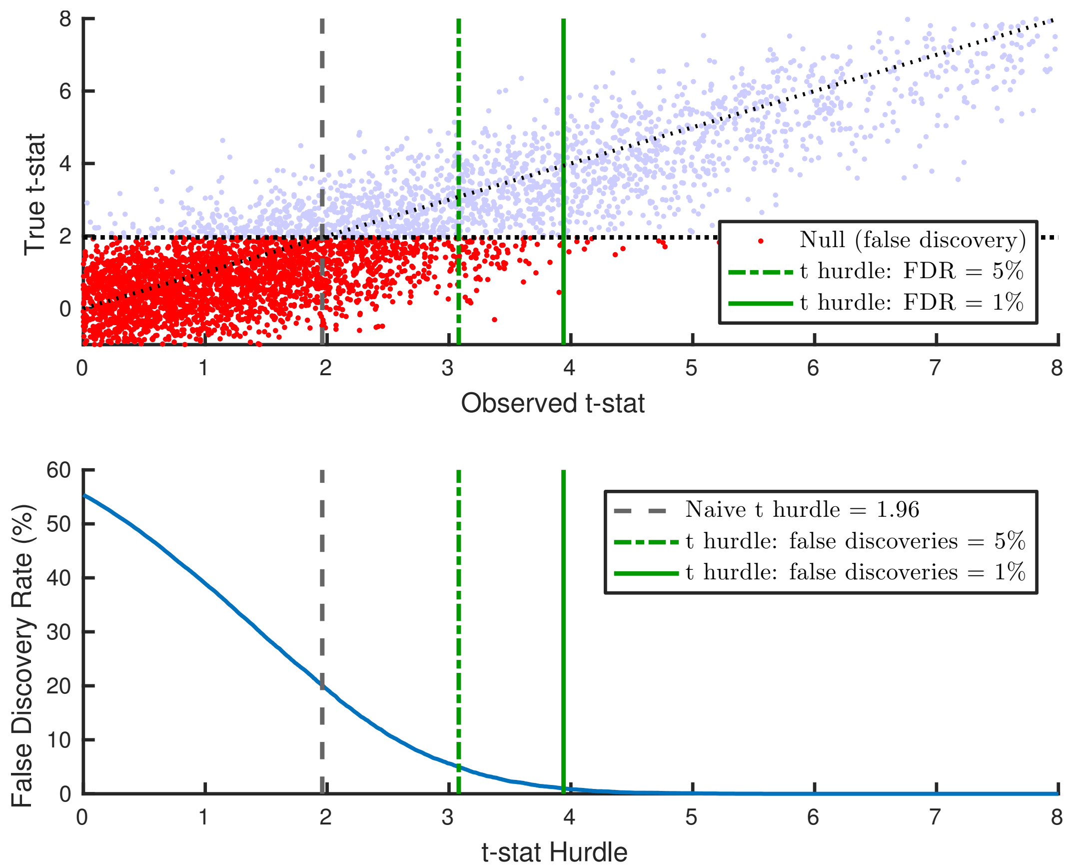
$$
\text{true t-stat} \equiv \frac{\text{true return}}{\text{standard error}} < 1.96. \qquad (33)
$$

This null is motivated by both theory and data. From a theoretical perspective, Equation (33) is a natural extension of the traditional t-stat < 1.96 hurdle. As the observed t-stat is a noisy estimate of the true t-stat, roughly half of the true t-stats will be below the observed one. Using the null in Equation (33) limits this uncertainty, and provides a higher order assurance that the true t-stat exceeds 1.96.

From an empirical perspective, the data show that we need a rather strict definition of a null in order to isolate unusual cases. As we will see, relatively few narrative portfolios satisfy equation (33), and those that do are likely to be portfolios worthy of further research.

Figure A.4 illustrates the FDR implied by the null (33). The top panel shows a scatterplot of published true t-stats against observed t-stats from simulating the estimated model. If there was no publication bias, observed t-stats would be an unbiased estimate of the true t-stat, and the scatterplots would be evenly spread across the 45 degree line (dotted line). There is a bit of publication bias, and thus there are more markers below the 45 degree line than above it.

Figure A.4: Multiple Tests of the Null: True t-stat < 1.96. We simulate narrative portfolios using our estimated model (Table 4). The top panel shows a scatter plot of true t-stats against observed t-stats, where true t-stat = [true return]/[standard error]. Non-null predictors are those with true t-stats > 1.96 (light dots). The false discovery rate for a given t hurdle is the fraction of predictors which exceed the hurdle that are null.



Despite the fact that the bias adjustments are small, many predictors are null (red dots). The presence of many null predictors is due to the stringency of our null definition. By design, only about half of the predictors with observed t-stats around 2 are "significant."

The bottom panel shows the FDR as a function of the t-stat hurdle. Using a hurdle of 0, 54% of predictors are null, and roughly 20% of predictors are null using the traditional hurdle of 1.96. It's not until t-stat hurdles above 3.0 that one achieves an FDR recommended by HLZ. Indeed, a high t-stat of 3.92 is required to achieve an FDR of 1%.

The t-stat hurdle of 3.92 effectively generates interesting academic case studies. Predictors that meet this hurdle are very likely to be notable in the traditional academic sense. As the number of predictors available for study has become unwieldy, this higher hurdle may be helpful for focusing the literature.

References

Alti, Aydoğan and Sheridan Titman. "A Dynamic Model of Characteristic-Based Return Predictability" (2017).

Amihud, Yakov. "Illiquidity and stock returns: cross-section and time-series effects". *Journal of financial markets* 5.1 (2002), pp. 31–56.

Andrews, Isaiah and Maximilian Kasy. *Identification of and correction for publication bias*. Tech. rep. National Bureau of Economic Research, 2017.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. "The cross-section of volatility and expected returns". *The Journal of Finance* 61.1 (2006), pp. 259–299.

Benjamini, Yoav and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". *Journal of the royal statistical society. Series B (Methodological)* (1995), pp. 289–300.

Black, Fischer. "Beta and return". *The Journal of Portfolio Management* 20.1 (1993), pp. 8–18.

Chen, Andrew Y and Mihail Velikov. "Accounting for the Anomaly Zoo: A Trading Cost Perspective". *Available at SSRN: https://papers.ssrn.com/abstract=3073681* (2017).

Chordia, Tarun, Amit Goyal, and Alessio Saretto. "p-hacking: Evidence from two mmillion trading strategies" (2017).

Christensen, Garret S and Edward Miguel. *Transparency, Reproducibility, and the Credibility of Economics Research*. Tech. rep. National Bureau of Economic Research, 2016.

Cochrane, John H. "Presidential address: Discount rates". *The Journal of Finance* 66.4 (2011), pp. 1047–1108.

Cusatis, Patrick J, James A Miles, and J Randall Woolridge. "Restructuring through spinoffs: The stock market evidence". *Journal of Financial Economics* 33.3 (1993), pp. 293–311.

Dawid, AP. "Selection paradoxes of Bayesian inference". *Lecture Notes-Monograph Series* (1994), pp. 211–220.

Efron, Bradley. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Vol. 1. Cambridge University Press, 2012.

— "Tweedie's formula and selection bias". *Journal of the American Statistical Association* 106.496 (2011), pp. 1602–1614.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu. “Taming the Factor Zoo” (2017).

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 Frank Zhang. “The remarkable multidimensionality in the cross-section of expected US stock returns”. *Available at SSRN 2262374* (2014).

Harvey, Campbell R. “Reflections on Editing the Journal of Finance, 2006 to 2012”. *Secrets of Economics Editors* (2014), p. 67.

Harvey, Campbell R, Yan Liu, and Heqing Zhu. “... And the cross-section of expected returns”. *Review of Financial Studies* 29.1 (2015), pp. 5–68.

Haugen, Robert A and Nardin L Baker. “Commonality in the determinants of expected stock returns”. *Journal of Financial Economics* 41.3 (1996), pp. 401–439.

Hedges, Larry V. “Modeling publication selection effects in meta-analysis”. *Statistical Science* (1992), pp. 246–255.

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

Ikenberry, David, Josef Lakonishok, and Theo Vermaelen. “Market underreaction to open market share repurchases”. *Journal of financial economics* 39.2-3 (1995), pp. 181–208.

Jacobs, Heiko and Sebastian Müller. “Anomalies Across the Globe: Once Public, No Longer Existent?” (2016).

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

Jensen, Michael C and George A Bennington. “Random walks and technical theories: Some additional evidence”. *Journal of Finance* 25 (1970), pp. 469–482.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. “Shrinking the Cross Section” (2017).

La Porta, Rafael. “Expectations and the cross-section of stock returns”. *The Journal of Finance* 51.5 (1996), pp. 1715–1742.

Linnainmaa, Juhani T and Michael R Roberts. *The history of the cross section of stock returns.* Tech. rep. National Bureau of Economic Research, 2016.

Liu, Laura, Hyungsik Roger Moon, and Frank Schorfheide. *Forecasting with Dynamic Panel Data Models*. Tech. rep. Working Paper, 2016.

Liu, Qi, Lei Lu, Bo Sun, and Hongjun Yan. "A model of anomaly discovery". *Available at SSRN 2431498* (2015).

Lo, Andrew W and A Craig MacKinlay. "Data-snooping biases in tests of financial asset pricing models". *Review of financial studies* 3.3 (1990), pp. 431–467.

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 and Jeffrey Pontiff. "Does academic research destroy stock return predictability?" *The Journal of Finance* 71.1 (2016), pp. 5–32.

Merton, Robert C. *On the current state of the stock market rationality hypothesis*. 1987.

Penasse, Julien. "Understanding Anomaly Decay" (2017).

Ritter, Jay R. "The long-run performance of initial public offerings". *The journal of finance* 46.1 (1991), pp. 3–27.

Senn, Stephen. "A note concerning a selection "paradox" of dawid's". *The American Statistician* 62.3 (2008), pp. 206–210.

Sullivan, Ryan, Allan Timmermann, and Halbert White. "Dangers of data mining: The case of calendar effects in stock returns". *Journal of Econometrics* 105.1 (2001), pp. 249–286.

— "Data-snooping, technical trading rule performance, and the bootstrap". *The Journal of Finance* 54.5 (1999), pp. 1647–1691.

Titman, Sheridan, KC John Wei, and Feixue Xie. "Capital investments and stock returns". *Journal of financial and Quantitative Analysis* 39.4 (2004), pp. 677–700.

Yan, Xuemin Sterling and Lingling Zheng. "Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach". *The Review of Financial Studies* 30.4 (2017), pp. 1382–1423.