Publication Bias and the Cross-Section of Stock Returns

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

Tom Zimmermann
Quantco, Inc
zimmermann.tom@gmail.com

December 2017†

Abstract

Due to data mining, the expected returns of stock market predictors may be biased. This bias, however, may be mitigated by the journal review process. We develop an estimator for the net bias and apply it to replications of 172 cross-sectional stock return predictors. Bias-adjusted long-short returns are only 13% smaller than in-sample long-short returns. This small bias comes from the dispersion of t-stats across predictors, which is too large to be accounted for by noise, indicating that many predictors have positive true returns. The bias is too small to account for the deterioration in average returns after publication (p-value = 0.0002), suggesting an important role for mispricing. Among predictors that can survive journal review, a low t-stat hurdle of 1.8 controls for multiple testing using statistics recommended by Harvey, Y. Liu, and Zhu (2015).

*Corresponding author.

†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, 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, Y. 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), X. S. 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 effect, and apply it to a dataset of 172 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 13% 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 $(1-0.13)8\% \approx 7\%$.

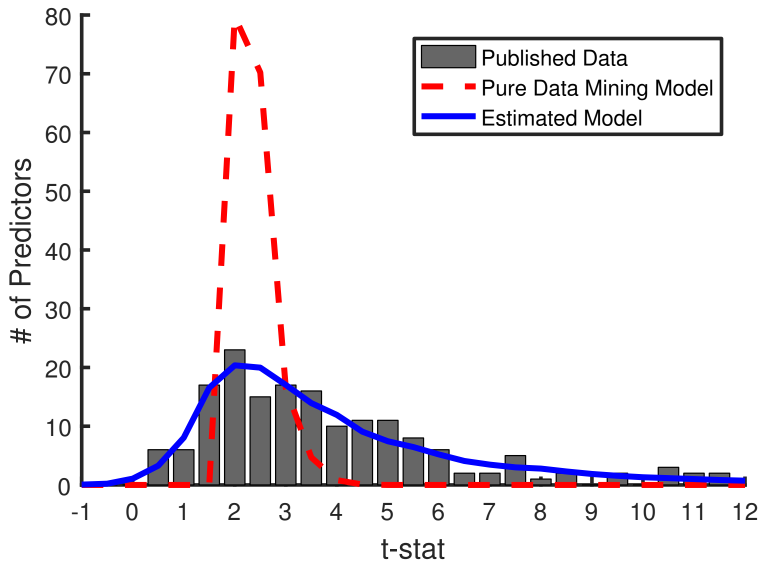
This modest bias adjustment comes from the shape of the distribution of published t-stats, seen in Figure 1. The left side of this distribution displays clear evidence of data mining, as predictors with t-stats less than 2.0 are conveniently missing.

Data mining, however, cannot account for the right side of the distribution. Under pure data mining, t-stats should bunch up at the t-stat hurdle, and so we

[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.



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 by pure luck 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. The dashed line in Figure 1 plots this distribution. This pure data mining distribution fits the left shoulder of the published data, but it decays far too quickly to account for the right tail.

In this paper, we estimate a model of data mining that allows for 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 13% number.

Our data consists of 172 long-short portfolio returns from replications of 120 publications in accounting, economics, and finance journals. To our knowledge, this is the most comprehensive dataset of cross-sectional predictors to date.[3] Our data includes all 97 of McLean and Pontiff (2016)'s predictors, as well as 75 additional statistically significant predictors that are published in "top tier" journals.[4] We make our data available online at http://sites.google.com/site/chenandrewy/code-and-data/.

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 highly selective 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 false discovery rate (FDR) implied by our model. We find that, among published predictors with t-stats > 1.96, the FDR is a tiny 0.6%—that is, nearly 100% of published anomalies that meet this threshold are true. Thus, we find that the traditional t-stat hurdle of 1.96 can actually be lowered, and even a t-stat hurdle of 1.8 leads to an FDR of 1.0%.

This surprising result may appear to contradict multiple-testing logic. If one runs 172 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

[3] Hou, Xue, and L. 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 148 anomalies.

[4] "Top tier" journals are basically the top 5 journals in each of the disciplines: finance, accounting, and economics.

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 increased 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 24 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 = 0.0002).

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.

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 13% 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 the remaining 52% can be accounted for by trading costs.

Related Literature Our paper is closely related to Harvey, Y. 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 above 2.88 is required to obtain an FDR of 1%, far above our estimate of 1.8.

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, Y. Liu, and Zhu (2015), Linnainmaa and Roberts (2016), and Hou, Xue, and L. Zhang (2017) find that most published results are false, while Green, Hand, and F. 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, Y. 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 L. Liu, Moon, and Schorfheide (2016).

Our model complements Q. Liu, Lu, Sun, and H. 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 172 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.

Suppose that in-sample returns r_i are noisy signals of the 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 172 predictors, we have $\sigma \approx 0.19$, corresponding to an average portfolio volatility of 3% per month and a typical sample of 324 months.

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 function of σ_μ (Equation (6)). A very high $\sigma_\mu \approx 0.90$ is required to fit the empirical standard deviation of 0.51% found in our dataset. This high $\hat{\sigma}_\mu$ implies a very small shrinkage (dotted line) of less than 10%. 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 still the shrinkage is quite small and below 10%.

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. Readers eager for results may wish to skip to Section 4.

[5] The adjustment is

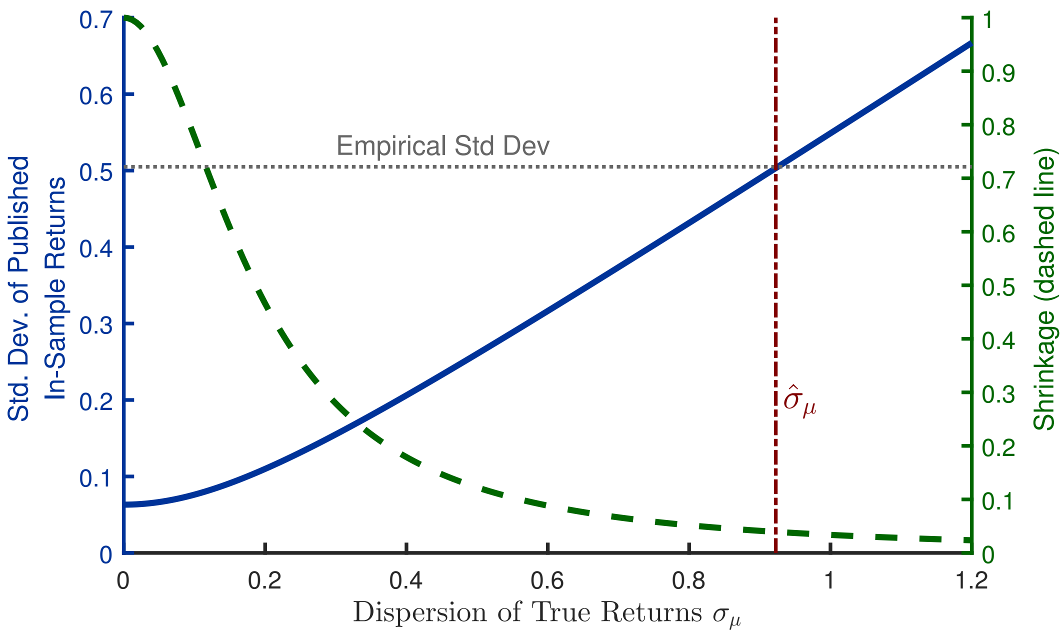
$$f(\sigma, \sigma_\mu) = 1 + \alpha\phi(\alpha)/(1 - \Phi(\alpha)) - [\phi(\alpha)/(1 - \Phi(\alpha))]^2$$

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

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" (horizontal dotted line) is the cross-sectional standard deviation of published in-sample returns in our dataset. $\hat{\sigma}_\mu$ (vertical dashed-dot line) is the implied estimate for the dispersion of true returns. 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). $\hat{\sigma}_\mu$ implies a small shrinkage of less than 10%. Returns are % per month. The model assumes $\sigma = 0.19$, the average published standard error in our dataset of 172 predictors.



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.

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 re-

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	$p(r_i/\sigma_i \mid t_{\text{cut}}, t_{\text{slope}}) = \dfrac{1}{1 + \exp(t_{\text{slope}}(r_i/\sigma_i - t_{\text{cut}}))}$

quirement 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 more than 170 predictors, the median pairwise correlation is 0.026, and 80% of correlations are between -0.37 and 0.41. The full distribution of correlations can be found in Appendix A.2.

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

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

[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 $\sqrt{\frac{1}{2(600-1)}} \approx 0.04$.

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 equilibrium theories based on risk. 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

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

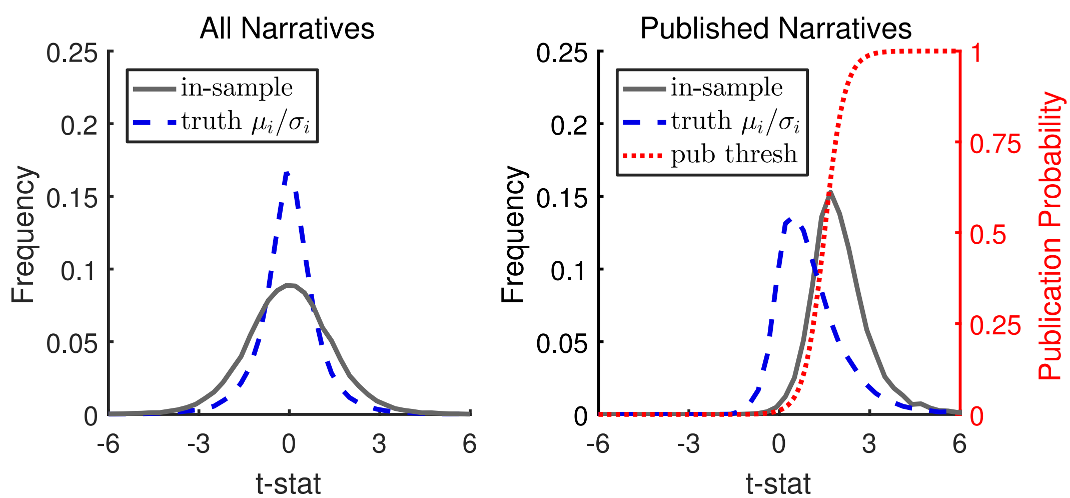
This function implies that t_{cut} is the midpoint of the soft threshold and t_{slope} is the slope. The slope captures that fact that journals make editorial decisions that can soften a strict statistical requirement. For example, many published papers report only Fama-Macbeth regression coefficients. While the regression coefficients may be statistically significant, the analogous long-short portfolio may not be.

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.

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 $\mu_\mu = 0$, $\sigma_\mu = 0.30$, $\nu_\mu = 7$, $\mu_\sigma = -1.31$, $\sigma_\sigma = 0.45$, $t_{\text{cut}} = 1.5$, and $t_{\text{slope}} = 3$.



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. Publication Bias Adjustment

The model implies a formula for adjusting in-sample returns for publication bias. We derive this formula assuming that we know θ, the parameters that govern the model. In the next subsection we explain how we extract parameters from data.

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 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)).

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.

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 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 estimator, improve on the in-sample mean of a given observation by incorporating information from other observations (Efron (2012)).

3.3. Estimation of Publication Bias

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.

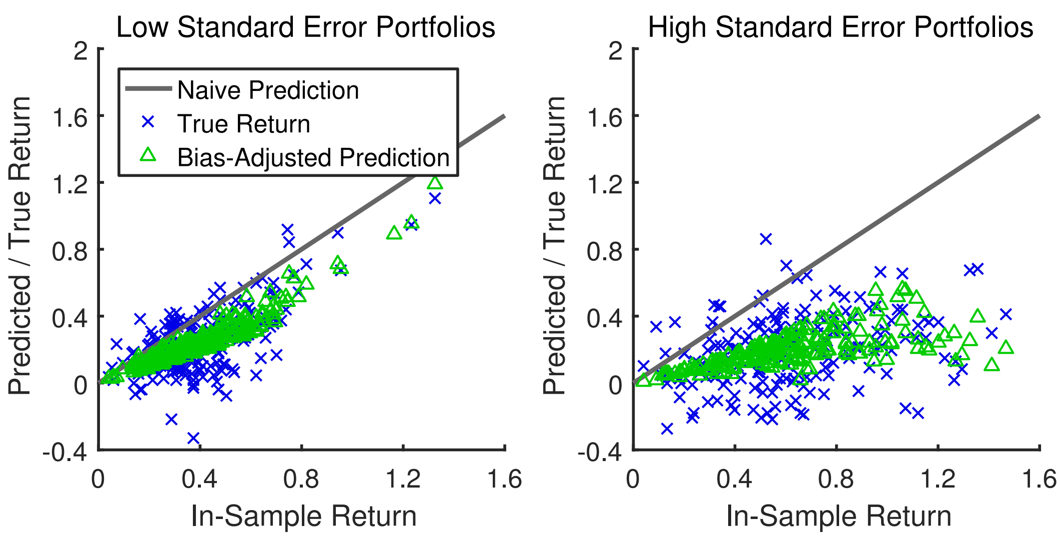
The result of this process is illustrated in Figure 4, which plots bias adjustments computed from applying maximum likelihood to simulated data. As the data is simulated, we're able to observe the true returns, and can directly observe publication bias. This bias is manifested by the fact that the naive prediction (assuming in sample returns equal the true return) is typically higher than the true returns. Moreover, this bias increases in the in-sample return, as well as the standard error.

The figure shows that the bias adjustments effectively remove publication bias. The bias adjusted predictions run right through the center of the clouds of true returns, and adjust effectively regardless of the in-sample return and standard error of the portfolio.

Maximum likelihood estimation of our model is somewhat tricky. We discuss how we handle the difficult aspects of the estimation in the remainder of this section.

In our baseline estimation, we do not estimate μ_μ and instead simply set it to the most conservative value of 0. We do this since the likelihood can be quite

Figure 4: Bias Adjustment Illustration. We simulate 400 published portfolios (parameters are in Figure 3), 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) is that the in-sample return is equal to the true return. This prediction is biased upward compared to the true returns (x's) due to the publication bias, and this bias increases in standard error and the in-sample return. The bias adjusted predictions (triangles) effectively adjust for publication bias.



flat for negative μ_μ, which then leads to bad behavior of the numerical optimizers. This problem is intuitive: Since only positive returns are published, the published returns have very little information about the mean of all returns if the mean is negative. Nevertheless, we find that setting $\mu_\mu = 0$ results in at least a local maximum (Figure A.2), and, moreover, assuming other values for μ_μ has little effect on our main results.

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:

$$f_{r,\sigma|\text{pub}}(r_i, \sigma_i|\text{pub}_i, \theta) = \frac{p(r_i/\sigma_i|\theta) f_{r|\sigma}(r_i|\sigma_i, \theta) f_N(\log\sigma_i|\mu_\sigma, \sigma_\sigma)}{\int d\tilde{\sigma} \left[\int d\tilde{r}\, p(\tilde{r}/\tilde{\sigma}|\theta) f_{r|\sigma}(\tilde{r}|\tilde{\sigma}, \theta_\mu)\right] f_N(\log\tilde{\sigma}|\mu_\sigma, \sigma_\sigma)}, \quad (23)$$

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

$$f_{r|\sigma}(r|\sigma, \theta_\mu) = \int d\tilde{\mu}\, f_\tau(\tilde{\mu}|\mu_\mu, \sigma_\mu, \nu_\mu) f_N(r|\tilde{\mu}, \sigma). \quad (24)$$

The numerator of Equation (23) 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.

3.4. Data on 172 Published Cross-Sectional Return Predictors

In principle, our model can be estimated by hand-collecting statistics from the original publications. The original statistics use various portfolio constructions and regression specifications, however, and the assumption that true returns are drawn from a single-peaked distribution (Equation (7)) means that we should standardize the returns as much as possible.

Thus, we construct 172 long-short portfolios by replicating 120 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 L. Zhang (2017) dataset of 447 anomalies consists of only 148 anomalies if one excludes alternative lagging choices and anomalies that were not demonstrated to produce predictability in the original papers. The full list and predictor definitions can be found in Table A.1. We also make our dataset available at http://sites.google.com/site/chenandrewy/code-and-data/.

To construct our data, we begin with McLean and Pontiff (2016)'s set of 97 predictors and add predictors listed in Green, Hand, and X. F. Zhang (2013) (GHZ), Harvey, Y. Liu, and Zhu (2015) (HLZ), and Hou, Xue, and L. Zhang (2017) (HXZ). In adding predictors, we require that (1) the predictor is published in a "top tier" journal,[7] (2) the portfolio is shown to generate statistically significant return predictability in the original paper, and (3) the portfolio uses publicly-available data that are regularly updated. This process adds 23 variables from GHZ, 26 variables from HLZ, 47 variables from HXZ, and 20 variables which overlap between these papers.

We do not aim to perfectly replicate the original papers, but rather try to (1) capture the spirit of the paper and (2) produce t-statistics above 1.5. In our experience, simply following the instructions in the original paper does not guarantee a good replication, and including the t-statistic minimum helps avoid errors which invariably occur when replicating so many studies. Aiming for high

[7]We define "top tier" journals as: the Journal of Finance, the Journal of Financial Economics, the Review of Financial Studies, and the Journal of Financial and Quantitative Analysis, Management Science, the Accounting Review, the Journal of Accounting Research, the Journal of Accounting and Economics, the Review of Accounting Studies, Contemporary Accounting Research, the Quarterly Journal of Economics, the Journal of Political Economy, the American Economic Review, Econometrica, and the Review of Economic Studies.

t-stats is also important for avoiding "reverse p-hacking," that is, suffering from the publishing process's bias for negative replication results (Galiani, Gertler, and Romero (2017)).

We set the t-statistic goal to be less than 2 to allow for differences in t-stat calculations and updates to data sources. We include 12 portfolios with t-stats < 1.5 that were in McLean and Pontiff (2016)'s dataset. 4 additional portfolios also had t-stats < 1.5 but reasonable in-sample returns. We include these portfolios to avoid adding selection beyond what the publication process produces. Nevertheless, excluding these portfolios has little effect on our results.

The resulting dataset covers 90% of the 121 factors in HLZ that meet our screening criteria. Similarly, it covers 91% of the 149 predictors in HXZ that meet our screening criteria and are not duplicates. Most of the remaining predictors require either specialized or proprietary data (TAQ, BEA input-output tables, conference call transcripts). In a handful of cases, we had difficulty replicating the original papers. In these cases, we assume that the error was ours and exclude these portfolios.

Table 2 gives an overview of our dataset. The top panel describes the types of journals as well as the categories of predictors we cover. More than half of the predictors (104) come from the "top 3" finance journals. The bulk of the remainder (44) come from the "top 3" accounting journals.

We categorize predictors primarily by the data source used for constructing the signals. More than half of the predictors focus on Compustat data, and indeed, nearly all of the accounting journal predictors have this focus. The second largest group of predictors relies on only market prices (equity prices or equity option prices). The remainder of the predictors use either analyst forecasts, institutional ownership data, or focus on events (such as mergers or IPOs).

In our baseline results, we focus on equal-weighted long short portfolios.[8] Most variables are long-short quintiles, though our dataset also includes a numer of indicator variables (like recent IPOs). Panel B of Table 2 shows the mean returns and t-stats from these portfolios. The returns in the original samples are around 0.70% per month on average. For comparison, the mean return of equal-

[8]Nearly all of the original papers use equal-weighted portfolios or Fama-Macbeth regressions. An important exception is the idiosyncratic volatility (Ang, Hodrick, Xing, and X. Zhang (2006)) which focuses on value-weighted portfolios. We equal-weight our idiosyncratic volatility portfolios for ease of communication, but our results are not sensitive to using value-weights for this particular predictor.

weighted long short quintiles based on B/M is around 1.00% in Fama and French (1992)'s sample. The standard deviation of mean returns (across predictors) is around 0.50%, indicating that there is substantial heterogeneity in returns. These values are close to those from McLean and Pontiff (2016), who find a mean return of 0.58% and a standard deviation of 0.40%.

Interestingly, the data show that all journal categories and predictor data sources lead to similar mean returns. Accounting journals and analyst forecast data appear to produce higher t-stats, but considering the large dispersion in t-statistics it's hard to say that any category is special. Indeed, we find that including only predictors from top 3 finance journals does not affect our results.

Table 2: Summary Statistics for 172 Cross-Sectional Return Predictors

Most portfolios are quintile sorts. A minority of portfolios use indicator variables. *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 (as these two journals are the only ones among top 5 economics journals with predictors that we replicated). *Accounting* includes predictors based on Compustat information, *Analyst forecasts* includes predictors based on IBES, *Event* includes company events such as mergers or initial public offerings, *Price Only* includes predictors based on stock or options market information based on CRSP or OptionMetrics, and *Owner* which includes predictors based on 13F data. 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 Categories

	Accounting	Price Only	Analyst Forecasts	Event	Owner	Total
Top 3 Finance	42	40	8	8	5	103
Top 3 Accounting	41	0	3	0	0	44
Top 5 Econ	2	2	0	0	0	4
Other	11	5	4	1	0	21
Total	96	47	15	9	5	172

Panel B: In-Sample Statistics for Equal-Weighted Long-Short Portfolios

		Mean Return (% per month)		t-statistic	
	#Portfolios	Mean	Std	Mean	Std
Journal					
Top 3 Finance	103	0.68	0.49	3.57	2.43
Top 3 Accounting	44	0.69	0.60	5.23	4.05
Top 5 Econ	4	0.55	0.18	2.79	1.95
Other	21	0.76	0.41	4.82	3.02
Predictor Type					
Accounting	96	0.60	0.45	4.39	3.19
Analyst Forecast	15	0.86	0.53	5.81	4.15
Market Price	47	0.76	0.55	3.26	2.27
Event	9	0.56	0.29	2.55	0.69
Ownership	5	1.42	0.71	5.13	3.27

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 3 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 3.3). The table shows our baseline specification, as well as four alternative specifications for robustness.

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 3.53, this implies that the cross-sectional standard deviation of all narrative true returns is $\sigma_\mu\sqrt{\frac{\nu_\mu}{\nu_\mu-2}} =$ 0.68% per month. In other words, it's quite common to find narrative portfolios with true, bias-adjusted returns of 0.68% per month.

Moreover, the dispersion of true returns and the fat tail parameter are precisely estimated. Indeed, the dispersion of true returns is more than 5.5 standard errors from zero, showing that we can, with little doubt, reject the hypothesis that all predictors are false (equivalently, we reject that shrinkage is 100%).

The remainder of the parameters demonstrate that the estimator works properly, but they are otherwise not very interesting. The standard error parameters imply that the mean of all narrative standard errors is $\exp(\mu_\sigma + 0.5\sigma_\sigma^2) = 0.21\%$ per month. This is slightly 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, as is the midpoint of the t-statistic cutoff. The t-stat threshold slope of 2.47 indicates that the publication threshold is soft, which can be seen directly in the shape of the published t-stat distribution (Section 5.3).

The bottom of Table 3 provides the key numbers from the estimation: sum-

Table 3: Estimation Results

We estimate a model of biased publication (Table 1) on our library of cross-sectional return predictors (Table 2) by maximum likelihood. Bootstrap standard errors are in parentheses. 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). The bias adjustment is small, at 13% of the in-sample return. This small shrinkage is well-estimated, and robust across several alternative specifications. "t-stat > 2.0 only" uses 132 portfolios with t-stats > 2.0, "top 3 finance only" uses 104 portfolios from the Journal of Finance, Journal of Financial Economics, and Review of Financial Studies, "Normally distributed true returns" assumes that $\nu_\mu = 100$, and "negative mean of true returns" assumes that μ_μ is equal to the negative of the average in-sample return.

Parameter		Baseline	Alternative Specifications			
		172 EW long-short portfolios	t-stat > 2.0 Only	Top 3 Finance Only	Normally distributed true returns	Negative mean of true returns
Assumed Parameters						
μ_μ	mean true return	0	0	0	0	-0.69
Estimated Parameters						
σ_μ	dispersion of	0.45	0.41	0.43	0.70	0.75
	true returns	(0.08)	(0.05)	(0.11)	(0.06)	(0.08)
ν_μ	fat tail (d.o.f.) of	3.53	3.56	3.64	100.00	12.58
	true returns	(0.99)	(0.74)	(2.33)	-	(3.11)
μ_σ	mean of log	-1.70	-1.71	-1.59	-1.74	-1.72
	standard error	(0.06)	(0.05)	(0.07)	(0.04)	(0.05)
σ_σ^2	std of log	0.51	0.51	0.51	0.51	0.51
	standard error	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
t_{cut}	midpoint of	1.49	1.99	1.45	1.09	1.40
	t-stat threshold	(0.47)	(0.02)	(0.67)	(0.20)	(0.20)
t_{slope}	slope of	2.47	100.00	2.37	2.98	2.75
	t-stat threshold	(0.50)	-	(1.98)	(3.01)	(0.44)
Estimated Bias Adjustments						
	Mean Shrinkage	0.13	0.11	0.16	0.08	0.14
		(0.05)	(0.02)	(0.08)	(0.01)	(0.03)
	Median Shrinkage	0.09	0.09	0.12	0.05	0.09
		(0.05)	(0.01)	(0.09)	(0.01)	(0.02)
	Std Shrinkage	0.10	0.09	0.12	0.07	0.15
		0.03	0.01	0.04	0.01	0.04

mary statistics for bias adjusted returns. The 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

$$\text{Bias-adjusted return} = (1 - [\text{Shrinkage}]) \times [\text{In-Sample Return}]. \tag{25}$$

The baseline specifications finds that the mean shrinkage is modest, at just 13%. In other words, for the typical in-sample return of 0.70% per month, the bias adjusted return is $0.61\% = 0.70 \times (1 - 0.13)\%$). This small shrinkage is well-identified, with the standard error for the mean shrinkage at just 5%.

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

We can be even more confident in our headline result because it is robust across model specifications. The alternative specifications columns of Table 3 show that the mean shrinkage around 13% regardless of whether we keep only replications with t-stats > 2.0, top 3 finance journal papers, assume normally distributed true returns, or assume a negative mean true return.

The "t-stat > 2.0 only" specification uses only the 132 portfolios with t-stats > 2.0. It's arguable that the predictors with t-stats < 2.0 were due to errors in our replications, and this specification shows that our main results are robust to excluding these portfolios. In this estimation, we do not estimate the slope of the t-stat adjustment since we know by construction that the slope is nearly infinite.

The "top 3 finance only" specification uses 104 portfolios published in the Journal of Finance, Journal of Financial Economics, and the Review of Financial Studies. This specification should alleviate the concern that our results are driven by the 21 predictors in the "other" journal category (Table 2). Moreover it shows that the small bias exists within the finance journals alone.

The last two alternative specifications concern model assumptions. "Normally distributed true returns" assumes that the degrees of freedom parameter is very large and omits it from the estimation. This specification should alleviate the concern that our results are due to an artifact of the fat tail assumption in our model. Indeed, the normal assumption results in even smaller mean shrinkage of 8%.

"Negative mean of true returns" explores the assumption that the mean of true returns is very negative. In the baseline estimation, we assumed that the mean of true returns is 0 because the likelihood becomes very flat if this parameter is negative (Section 3.3). Interestingly, assuming a very negative mean of true returns has little effect on the mean shrinkage. This invariance occurs because the dispersion of true returns increases in order to compensate for the lower mean. These two parameters have opposing effects on the bias adjustment (Equation (22)), and in the end cancel out. Estimations with other assumptions for the mean of true returns lead to a similar cancellation effect.

4.3. Heterogeneity in bias adjustments

The mean shrinkage is just 13%, but the cross-sectional standard deviation is somewhat large at 10 percentage points (Table 3). Nevertheless, modest shrinkage is a good description of the estimates overall.

This heterogeneity is explored in detail in Figure 5, which shows a histogram of the shrinkage distribution, as well as the identities of the predictors. The distribution is very right-skewed, with 80% of predictors having shrinkage estimates less than 20%. Even the high shrinkage predictors have only a moderate amount of publication bias, however. 169 out of 172 predictors have shrinkage estimates less than 40%, and the maximum shrinkage is 53%.

Figure 5 also illustrates the determinants of the predictor-level shrinkage. Portfolios with high return volatility (red text) occupy nearly all of 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).

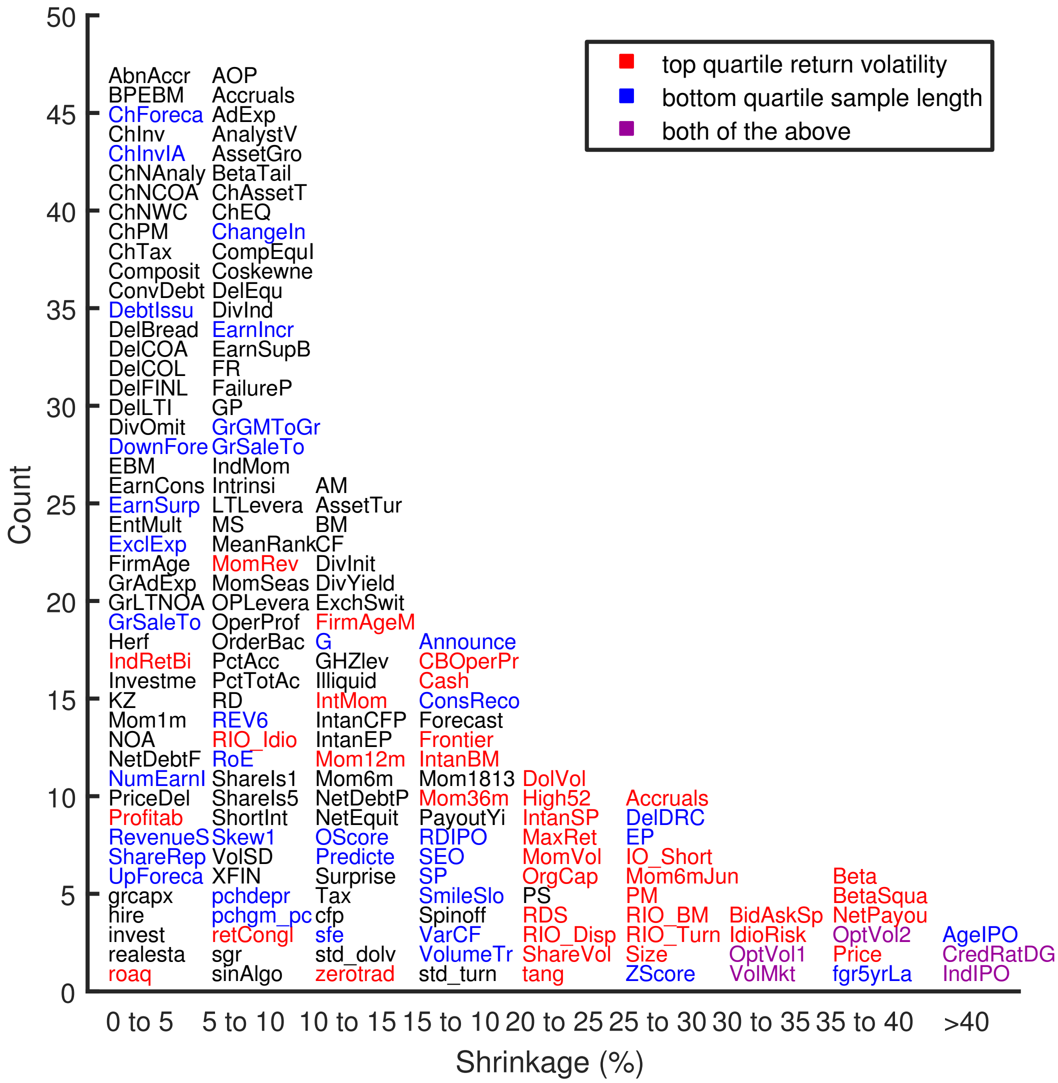
Theoretically, the sample length should also play a key role in the amount of noise, and thus the magnitude of shrinkage (Equation (22)). However, we find that the empirical correlation between the sample lengths and return standard errors is only mildly negative, at -0.20.

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 three highest shrinkage portfolios (the credit rating downgrade predictor of Dichev and Piotroski (2001), the IPO-age predictor of Ritter (1991), and the recent IPO predictor (also of Ritter (1991)) have reasonably good bias-adjusted returns of 0.37%, 0.48%, and 0.54%, respectively.

Figure 5: Distribution of publication bias adjustments for 172 cross-sectional return predictors. We estimate a model of biased publication (Table 1) on 172 long-short returns and t-stats (Table 3). 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 cross-sectional predictor. The full references are in Table A.1. Publication bias is heterogeneous and right skewed, but modest overall: 80% of the portfolios have shrinkage below 0.20. High shrinkage portfolios are mostly those with high return volatility.



5. Why is Publication Bias So Small?

Our results may be surprising, especially to those who work in the cross-sectional returns literature. Those in the field 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 our empirical results in detail. Section 5.1 begins by showing how the dispersion of true returns is critical for determining the mean shrinkage. Section 5.2 goes on to show how the dispersion of true returns is determined by the dispersion of in-sample returns. Section 5.3 finishes up by comparing our estimated bias adjustments with McLean and Pontiff (2016)'s upper 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 for all 172 of our predictors. The bias adjustment is plotted in terms of shrinkage, and is shown as a scatter against the standard error of the portfolio's in-sample return. Clearly, the portfolio's standard error is an important determinant of shrinkage. The higher 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

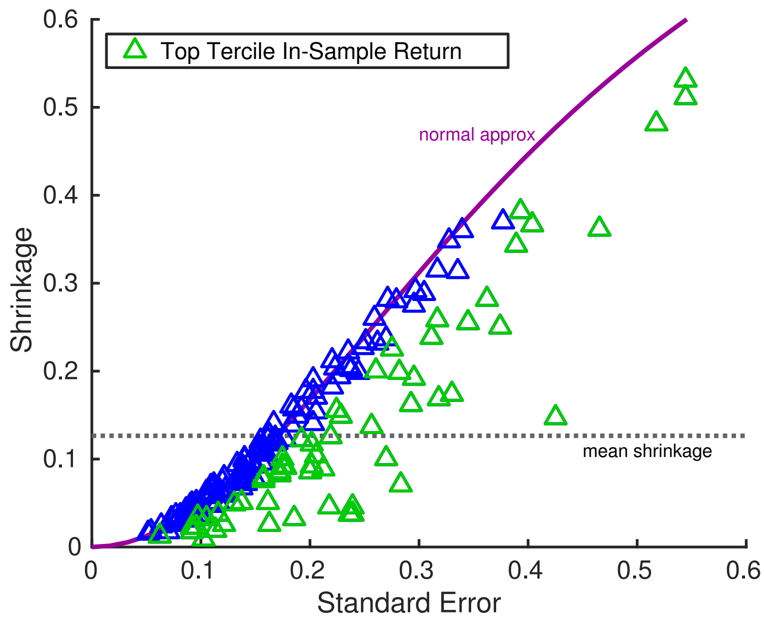
Figure 6: Determinants of the bias adjustments. Each marker represents one portfolio from our database of 172 predictors. Shrinkage is defined by

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

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 $\hat{\sigma}_\mu^2$.



expressed in closed form for the normal approximation of our model (Section 3.2). In this approximation, the shrinkage is a sort of noise-to-signal ratio, where the noise is the portfolio-specific standard error (Equation (22)).

Figure 6 also plots the normal approximation. The normal approximation works well for most of the portfolios, though it misses the portfolios with very high in-sample returns. 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 13% shrinkage can be derived using this approximation. Plugging in the mean standard error of 0.20% and our estimated $\hat{\sigma}_\mu = 0.45$, the typical shrinkage is approximately

$$\frac{\sigma_i^2}{\hat{\sigma}_\mu^2+\sigma_i^2}=\frac{0.20^2}{0.45^2+0.20^2}=0.16. \tag{26}$$

This analysis begs the question: where does our estimate of $\hat{\sigma}_\mu = 0.45$ 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.

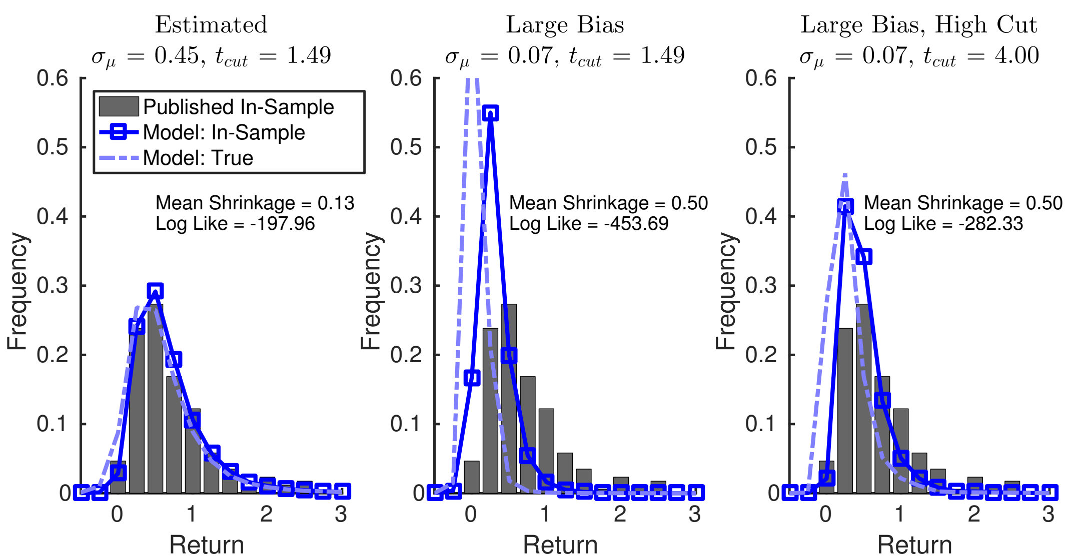
This identification is illustrated in Figure 7, which plots the distribution of in-sample returns in the data (bars) and model-implied distributions (lines with square markers). The left panel shows our estimated model. The other panels show models that display large publication bias.

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 0.13.

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 $\sigma_\mu = 0.07$, compared to the estimated $\sigma_\mu = 0.45$. $\sigma_\mu = 0.07$ 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, $\sigma_\mu = 0.07$ is required to assign the majority of this decline to publication bias.

The middle panel shows that this large bias model is a terrible fit for the data. This model utterly 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 ex-

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 in-sample returns in the data (grey bars) with the distribution implied by the estimated model of Table 3 (blue squares). The distribution of true returns implied by the model is plotted for comparison (dash dotted line). The middle panel shows a model in which σ_μ is decreased to 0.07, but all other parameters remain the same. The right panel shows a model in which σ_μ is decreased to 0.07 and t_{cut} is increased to 4.00. Both of the alternative models imply large mean shrinkage, and both are poor fits for the data.



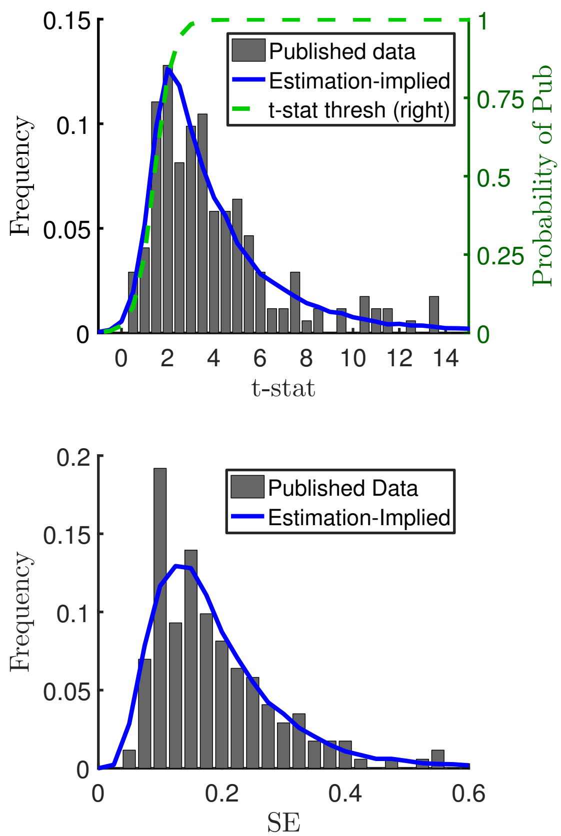
cessively high counts for the low return bins as well as the excessively low counts for the in-sample returns around 1 are all penalized by the estimator. Indeed, the log-likelihood of this model is -454, more than 200 log points away from our maximum likelihood estimate of -198.

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}—that is, the journals exhibit a large bias for statistical significance.

The right panel of Figure 7 shows that increasing t_{cut} does not fit the data either. The panel assumes a high t threshold midpoint of $t_{\text{cut}} = 4.00$, which, we found fits the data the best holding other parameters constant. This high cutoff still leads to too little dispersion in in-sample returns, and the log likelihood of -282 is still very far from our baseline estimate. Moreover, this high t-stat cutoff results in an implausibly strong preference for high t-stats. $t_{\text{cut}} = 4.00$ implies that a quality narrative portfolio with a t-stat of 2.5 gets only a 2.4% chance of publication, while a t-stat of 5.5 implies a 97.6% chance (Equation (13)).

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.

Figure 8: Model fit. We simulate the model using estimated parameter values (Table 3) and compare the distribution of observables with those from our database of 172 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).



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.

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

We've shown that the model fits data that it algorithmically targets. 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 learn about predictability may decide to change their portfolio allocations.

McLean and Pontiff (2016) (MP) develop a clever way to try 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.

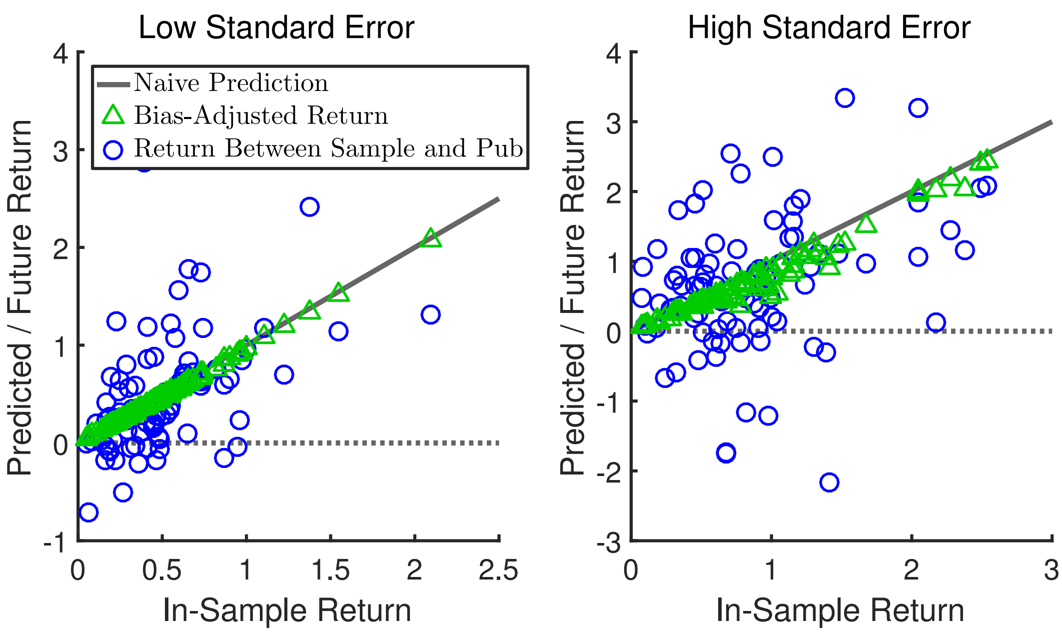
Figure 9 examines MP's lower bound. The figure shows scatterplots of bias-adjusted returns against the in-sample return, as well as the mean returns between the end of the original sample and publication. For comparison, the figure also plots the naive prediction: in-sample return = true return.

The first thing that jumps out from the plot is 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 a simple external validation of our bias adjustment. The circles represent the returns between the end of the sample and publication. High publication bias would imply 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 0.57% per month, slightly below our mean bias-adjusted return of 0.60%.

Figure 9: Bias adjusted returns and returns after the original in-sample period. Each marker represents one long-short portfolio. The naive prediction assumes that the true return is equal to the in-sample return. "Return between sample and pub" is the mean return between the end of the original paper's sample and publication date. Low standard error portfolios are those with standard error below the median. The bias-adjusted prediction is calculated by using Equations (19)-(20) and the maximum likelihood estimates of our model (Table 3). The bias-adjusted predictions are slightly above the mean return between sample and pub.



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, Y. 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 help reduce the multitude of cross-sectional predictors, and (2) much of the predictability throughout the zoo of anomalies was due to mispricing at the time of publication. Sections 6.1 and 6.2 discuss these implications, respectively.

6.1. Hypothesis Tests Adjusted for Publication Bias

The small bias adjustments for expected returns suggests that the zoo of anomalies cannot be simply attributed to publication bias. Here, we look more closely at the question, and show that 96% 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, Y. 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]

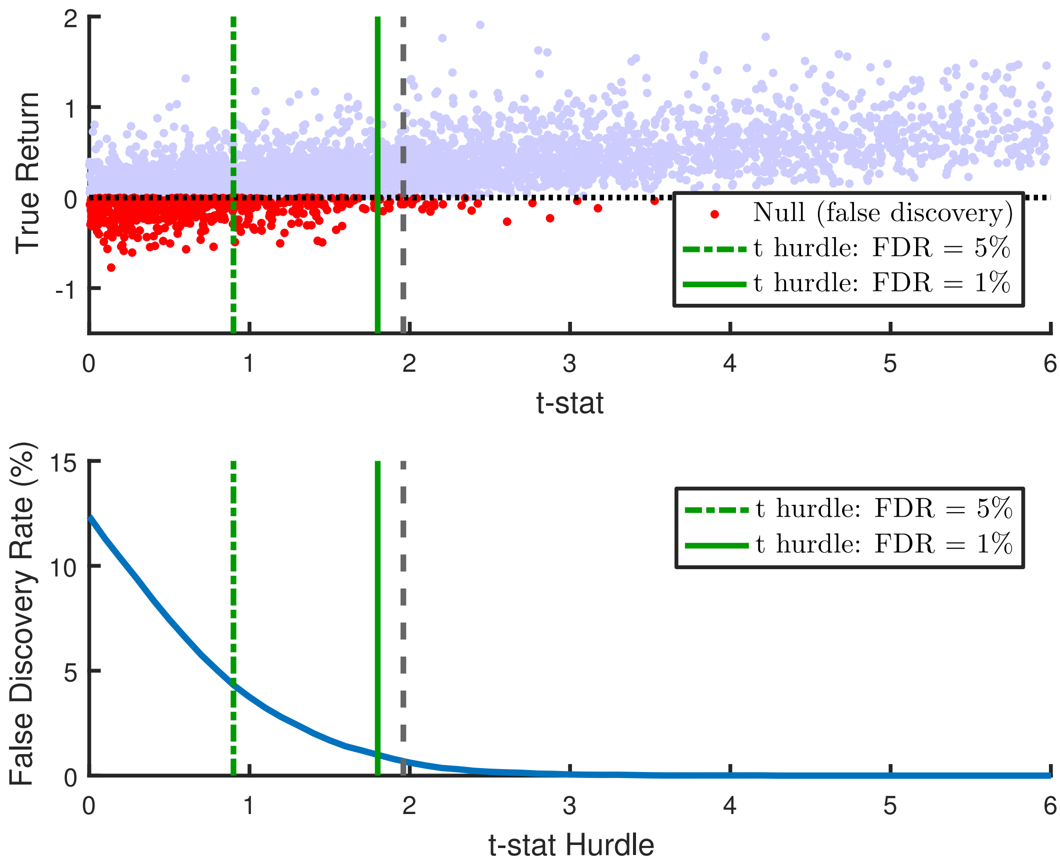
Using this definition of a null predictor, we can calculate false discovery rates by simulating the estimated model. Figure 10 illustrates this calculation. The top panel shows the distribution of narrative portfolios' true returns against t-stats. Null portfolios, that is, portfolios with negative true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

For portfolios with t-stats less than 0.5, the probability of being null is about 0.5, 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%. Increasing the t-stat hurdle decreases the false discoveries sharply. At a t-stat hurdle of 0.9 we already have an FDR of 5%, one of the FDR values in recommended by HLZ. Raising the t-stat hurdle to 1.8 reduces the FDR to 1%, HLZ's alternative recommendation.

[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.

Figure 10: Multiple Tests of the Null of Non-Positive True Returns. We simulate narrative portfolios using our estimated model (Table 3). 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 which exceed 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).



Thus, our results suggest that the traditional t-stat hurdle of 1.96 could actually be *loosened*. Even a t-stat hurdle of 0.9 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.3.

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 1.0 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 top tier journals. Indeed, as top tier journals only allow narratives that meet a soft threshold centered around 1.5 (Table 3), published predictors are almost all true. According to our model estimates, the FDR among published predictors is only 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 case, 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.4.

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 which 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 7.8, 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 point estimates 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.

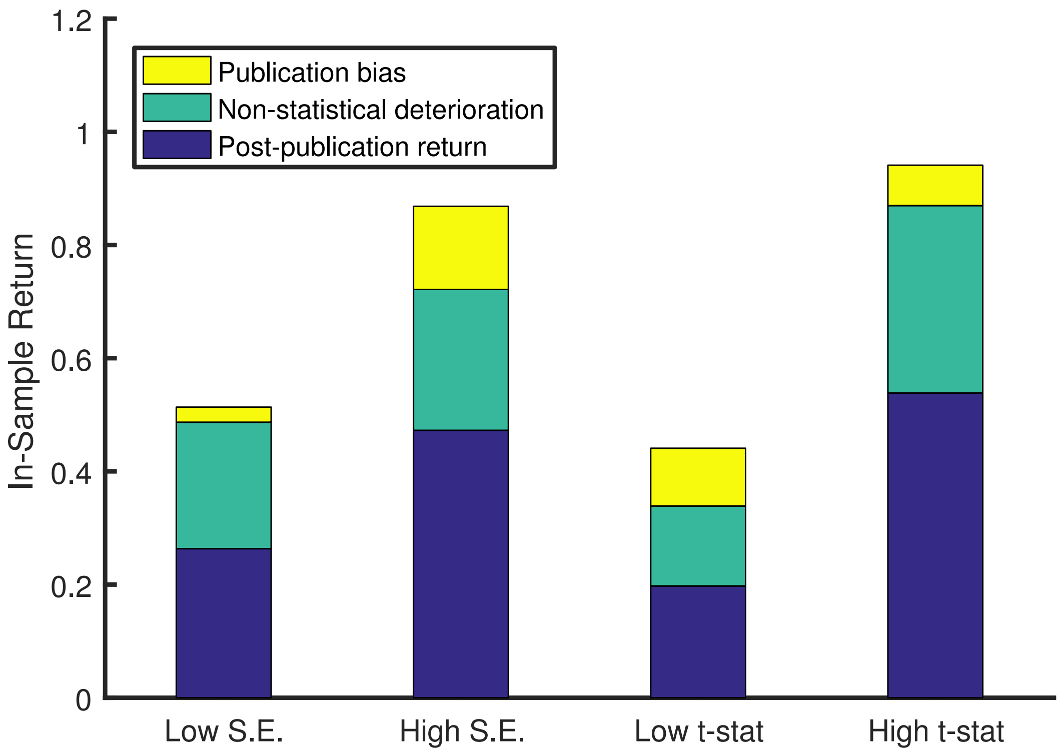
McLean and Pontiff (2016) (MP) and Marquering, Nisser, and Valla (2006) build on the insight that the average return post publication is informative about the nature of the cross-sectional predictability. If predictability is due to mispricing, post publication returns should be poor, as traders bid 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 deterioration (2) non-statistical deterioration, and (3) the post-publication return.

The decomposition comes from computing average returns of different types and taking differences. The publication bias deterioration is the average difference between in-sample returns and bias adjusted returns calculated according to Equations (19) and (20). Post-publication returns are the average returns in the sample after the publication date. The non-statistical deterioration is the average difference between bias-adjusted returns and post-publication returns. These averages are computed within subsets of predictors. Low standard error and low t-stat predictors are those below the median, while high standard errors and high t-stat are above the median.

The figure shows that a significant portion of in-sample returns is due to non-statistical deterioration. On average, post-publication returns are 0.25 percentage points per month lower than bias-adjusted returns, and this non-statistical deterioration accounts for 35% 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

Figure 11: Implied mispricing. This chart decomposes the average in-sample return 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 average mean return in the sample after publication. Non-statistical deterioration is the average difference between bias adjusted returns and post-sample returns. Each bar computes averages within 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 plays a role across many anomalies.

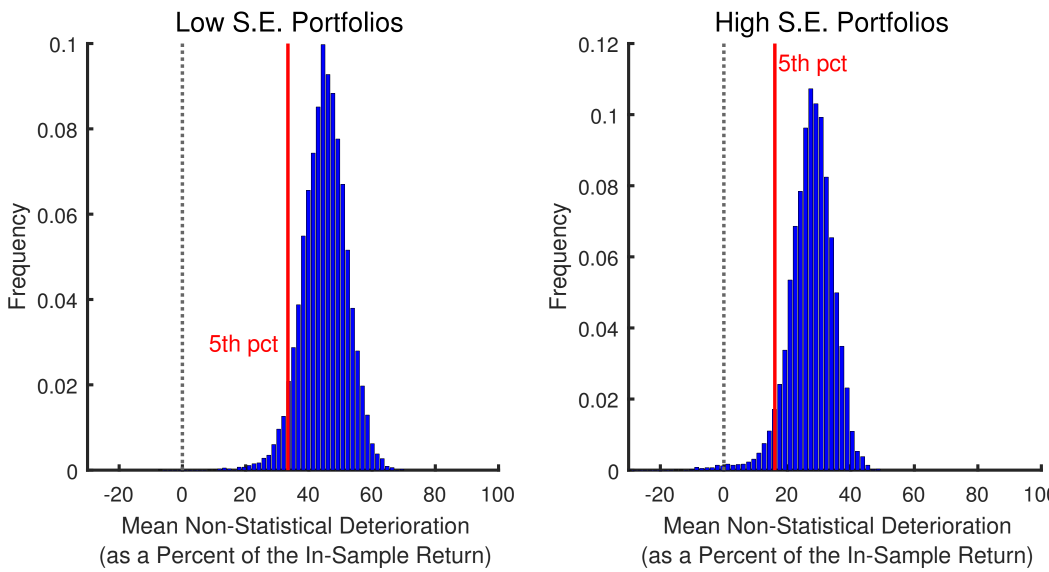


underlying driver of predictability.

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.

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 in-sample return 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.



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 far from zero, indicating that 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 of 0.0002 and 0.0056, respectively).

7. Conclusion

We find that the net effect of publication bias on cross-sectional stock predictors is small. 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 13% 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 the remaining 52% can be accounted for by trading costs.

A. Appendix

A.1. Additional Exhibits on the Data

Table A.1: Description of Anomaly Construction. This table provides details of the construction of 174 anomalies 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
AbnormalAccruals	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, Lakonishok and Sougiannis	2001	1975	1996	Advertising expense (xad) over market value of equity (shrout*abs(prc))
AgeIPO	IPO and age	Ritter	1991	1975	1984	Age (current year - founding year from Jay Ritter's dataset). Exclude if IndIPO == 0 or if there are fewer than 150 firms with IndIPO equal to 1 in a month.
AM	Total assets to market	Fama and French	1992	1963	1990	Total assets (at) divided by market value of equity.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
AnalystValue	Analyst Value	Frankel and Lee	1998	1975	1993	Analyst value is (1 + (FROE - .1)/1.1 + (FROE - .1)/(.1*1.1))*BMAve. FROE is the most recent mean analyst EPS forecast (meanest) times shares outstanding (shrout) divided by book value of common equity (ceq). BMAve is average book to market equity (ceq/(shrout*abs(prc)) over the past two years. Exclude if FROE > 1 or book equity negative or abs(prc) < 1 (Frankel and Lee, page 291).
AnnouncementReturn	Earnings announcement return	Chan, Jegadeesh and Lakonishok	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.
AOP	Analyst Optimism	Frankel and Lee	1998	1975	1993	AnalystValue (defined above) minus IntrinsicValue (defined above), divided by abs(IntrinsicValue).
AssetGrowth	Asset Growth	Cooper, Gulen and Schill	2008	1968	2003	Annual growth rate of total assets (at)
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 use the Corwin-Schultz 2012 estimate 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).
BPEBM	Leverage component of BM	Penman, Richardson and 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).

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
CBOperProf	Cash-based operating profitability	Ball, Gerakos, Linnainmaa and Nikolaev	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.
CF	Cash flow to market	Lakonishok, Shleifer and Vishny	1994	1968	1990	Net income (ib) plus depreciation (dp) divided by market equity. Exclude NASDAQ stocks.
cfp	Cash flow to price	Desai et al	2005	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).
ChangeInRecommendation	Change in recommendation	Jegadeesh, Kim, Krische and 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.
ChAssetTurnover	Change in Asset Turnover	Soliman	2008	1984	2002	Annual change in AssetTurnover (defined above). Exclude if price less than 5.
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.
ChForecastAccrual	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.
ChInv	Inventory Growth	Thomas and Zhang	2002	1970	1997	12 month change in inventory (invt) divided by average total assets.
ChInvIA	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.
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.

Continued on next page

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
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
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.
ChTax	Change in Taxes	Thomas and Zhang	2011	1977	2006	4-quarter change in quarterly total taxes (txtq), scaled by lagged total assets (at).
CompEquIss	Composite equity issuance	Daniel and Titman	2006	1968	2003	5 year growth rate of market value of equity minus 5 year stock return.
CompositeDebtIssuance	Composite debt issuance	Lyandres, Sun and 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, McNichols and 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.
Coskewness	Coskewness	Harvey and Siddique	2000	1963	1993	
CredRatDG	Credit Rating Downgrade	Dichev and Piotroski	2001	1970	1997	Equal to 1 if credit rating (splticrm) decreased by at least one notch relative to the previous month and 0 otherwise. Exclude if price less than 5.
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).
DelCOA	Change in current operating assets	Richardson, Sloan, Soliman and Tuna	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.
DelCOL	Change in current operating liabilities	Richardson, Sloan, Soliman and Tuna	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.
DelDRC	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.
DelEqu	Change in equity	Richardson, Sloan, Soliman and Tuna	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.
DelFINL	Change in financial liabilities	Richardson, Sloan, Soliman and Tuna	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.
DelLTI	Change in long-term investment	Richardson, Sloan, Soliman and Tuna	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.

Continued on next page

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
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.
DivInit	Dividend Initiation	Michaely, Thaler and 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 and 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, Nimalendran and Ryngaert	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	Chordia, Roll and Subrahmanyam	2001	1966	1995	Log of the product of two-month lagged trading volume (vol) and two-month lagged price (prc).
DownForecast	Down forecast EPS	Barber, Lehavy, McNichols and Trueman	2001	1985	1997	Binary variable equal to 1 if mean earnings forecast (meanest) decreased over the past month.
EarnIncrease	Consistent earnings increases	Barth, Elliott and Finn	1999	1982	1992	Binary variable equal to 1 if the change in quarterly net income (ibq) from t-1 to 1 was positive in quarters t, t-1, t-2, t-3 and t-4, and 0 otherwise.
EarningsConsistency	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.
EarningsSurprise	Earnings Surprise	Foster, Olsen and Shevlin	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
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!)
EBM	Enterprise component of BM	Penman, Richardson and Tuna	2007	1961	2001	(book equity (ceq) + cash and short-term investments (che) - long-term debt (dltt) - debt in current liabilities (dlc) - deferred charges (dc) - preferred dividends in arrears (dvpa) + treasury stock (tstkp)) / (market value of equity (shrout*abs(prc) + cash and short-term investments (che) - long-term debt (dltt) - debt in current liabilities (dlc) - preferred dividends (dvpa) + treasury stock (tstkp)). Exclude if price less than 5.

Continued on next page

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
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	1964	1971	net income (ni) divided by market value of equity. NYSE stocks only.
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, Hilscher and Szilagyi	2008	1963	2003	Using the specification for 12-month ahead failure probability in column 3 of table 4, 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 ϕ^j and the sum scaled by $\frac{1-\phi}{1-\phi^{12}}$; 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 ϕ^q on quarter $t-q$ and the sum scaled by $\frac{1-\phi^3}{1-\phi^{12}}$. $\phi = 2^{-\frac{1}{3}}$. All input variables are winsorized at the 5th and 95th percentile. Exclude if price less than 1.
fgr5yrLag	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.
FirmAge	Firm Age	Barry and Brown	1984	1931	1980	Months since start of CRSP coverage. Exclude if price less than 5.
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.
FR	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.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
Frontier	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.
G	Governance Index	Gompers, Ishii and Metrick	2003	1990	1999	Index available from http://faculty.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.
GHZlev	Leverage component of BM	Bhandari	1988	1946	1981	Total liabilities (lt) divided by market value of equity.
GP	Gross Profitability	Novy-Marx	2013	1962	2010	Revenue (sale) - cost of goods solds (cogs), divided by 12 months lagged total assets.
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.
grcapx	Change in capex (two years)	Anderson and Garcia-Feijoo	2006	1976	1999	Growth rate of capital expenditures (capx) relative to t-2. If capx is missing, replace with annual change in property, plant and equipment (ppent).
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, divided by 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, Whisenant and Yohn	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)
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	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. Exclude if price less than 5.

Continued on next page

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
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 <= 1980; 4512 and year <=1978, 4812, 4813 and year <= 1982, 49)
High52	52 week high	George and Hwang	2004	1963	2001	Divide the absolute value of prc by the maximum value of prc over the past twelve months.
hire	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.
IdioRisk	Idiosyncratic risk	Ang, Hodrick, Xing and Zhang	2006	1963	2000	Standard deviation of residuals from monthly CAPM regressions.
Illiquidity	Amihud's illiquidity	Amihud	2002	1964	1997	Past twelve month average of: daily return (abs(ret)) divided by turnover((abs(prc)*vol)
IndIPO	Initial Public Offerings	Ritter	1991	1975	1984	Binary variable equal to 1 if IPO in the past 36 months. IPO dates are taken from Jay Ritter's IPO data available at: http://bear.warrington.ufl.edu/ritter/ipodata.htm
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.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
IntMom	Intermediate Momentum	Novy-Marx	2012	1926	2010	Stock return between months t-12 and t-6
IntrinsicValue	Intrinsic value	Frankel and Lee	1998	1975	1993	Define FROE as net income (ib) divided by book equity (ceq), and drop if abs(FROE) > 1. DefineAveBM as average book equity to market value of equity for years t-1 and t (in the first year of coverage, use book-to-market equity directly). IntrinsicValue is $\left(1+\frac{FROE-.1}{1.1}+\frac{FROE-.1}{.1\times 1.1}\right)AveBM$. Exclude if price less than 1 or book equity less than 0.
invest	Capex and Inventory Change	Chen and Zhang	2010	1972	2006	Annual change in property, plant and equipment (ppegt) plus annual change in inventory (invt), scaled by lagged total assets (at). Use ppent if ppegt is missing.
Investment	Investment	Titman, Wei and Xie	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 USD 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. Among the remaining stocks, IO_ShortInterest is equal to 1 if a stock is in the highest institutional ownership (instown_perc) tercile and 0 if it is in the lowest tercile.
KZ	Kaplan Zingales index	Lamont, Polk and Saa-Requejo	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.
LTLeverage	Long-term leverage	Bhandari	1988	1946	1981	Log of long-term debt (dltt) to market value of equity (shrout*abs(prc))
MaxRet	Maximum return over month	Bali et al	2010	1962	2005	Maximum of daily returns (ret) over the previous month
MeanRankRevGrowth	Revenue Growth Rank	Lakonishok, Shleifer and Vishny	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.
Merger	Mergers	Langetieg	1978	1929	1969	Binary variable equal to 1 if involved in a merger in previous 12 months, and 0 otherwise. Merger data are from SDC.
Mom12m	Momentum (12 month)	Jegadeesh and Titman	1993	1964	1989	Stock return between months t-12 and t-1.
Mom18m13m	Momentum-Reversal	Jegadeesh and Titman	1993	1964	1989	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	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, Chordia, Jostova and Philipov	2007	1985	2003	Mom6m. Include only stocks with a credit rating (splticrm) of BBB or lower

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
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.
MomSeas	Return Seasonality	Heston and Sadka	2008	1965	2002	Average return in the same month over the preceding 20 years. Exclude NASDAQ stocks.
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.
MS	Mohanram G-score	Mohanram	2005	1978	2001	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, Richardson and Sloan	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.
NetDebtPrice	Net debt to price	Penman, Richardson and 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).
NetEquityFinance	Net equity financing	Bradshaw, Richardson and Sloan	2006	1971	2000	Sale of common stock (sstk) minus 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, Michaely, Richardson and Roberts	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.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
NOA	Net Operating Assets	Hirshleifer, Hou, Teoh and Zhang	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	Barth, Elliott and Finn	1999	1982	1992	Number of 4-quarter net income (ibq) increases over the previous 2 years.
OperProf	Operating Profitability	Fama and French	2015	1963	2013	Revenue (revt) minus cost (cogs) - administrative expenses (xsga) - interest expenses (xint), scaled by book value of equity (ceq). Exclude smallest size tercile.
OPLeverage	Operating Leverage	Novy-Marx	2010	1963	2008	Sum of administrative expenses (xsga) and cost of goods sold (cogs), scaled by total assets (at).
OptionVolume1	Option 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.
OptionVolume2	Option Volume	Johnson and So	2012	1996	2010	Total monthly option volume (volume) over all puts and calls, relative to the average of that same variable from months t-6 to t-1. 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, Shevlin and Venkatachalam	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).
OScore	O Score	Dichev	1998	1981	1995	-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}))). Funds from operations (fopt) is the sum of net income (ni), total taxes (txt) and depreciation (dp). NYSE stocks only. Exclude if SIC code between 3999 and 4999, or greater than 5999. Exclude if price less than 5.
PayoutYield	Payout Yield	Boudoukh, Michaely, Richardson and Roberts	2007	1984	2003	Sum of dividends (dvc), purchase of common and preferred stock (prstkc) and max(preferred stock redemption value (pstkrv), 0), divided by market value of equity.
pchdepr	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).
pchgm_pchsale	Change in gross margin vs sales	Abarbanell and Bushee	1998	1974	1988	Annual percentage change in revenue (sale) minus cost (cogs), minus annual percentage change in revenue.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
PctAcc	Percent Operating Accruals	Hafzalla, Lundholm and van Winkle	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, Lundholm and van Winkle	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.
PM	Profit Margin	Soliman	2008	1984	2002	Net income (ni) over revenue (revt). Exclude if price less than 5.
PredictedFE	Predicted Analyst forecast error	Frankel and Lee	1998	1975	1993	Define FROE as mean earnings estimate (meanest) times shares outstanding (shrout), divided by book equity (ceq). Define the prediction error as net income (ib) over book equity (ceq), minus FROE. In each month t, regress the prediction error on 3 year lagged values of a firm's relative ranks in the cross-sectional revenue (sale), BM (defined above), AOP (defined above) and FROE distributions. PredictedFE is the fitted value from that regression. Update monthly.
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	Profitability	Karthik, 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.
PS	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 gross margin ((sale - cogs)/sale) increased over the previous year; 1 if revenue to assets increased over the previous year; 1 if no issuance of common shares. Include highest quintile of book-to-market only.
RD	R&D over market cap	Chan, Lakonishok and Sougiannis	2001	1975	1995	R&D expense (xrd) over market value of equity.
RDIPO	IPO and no R&D spending	Gou, Lev and Shi	2006	1980	1995	Binary variable equal to 1 if positive R&D expense (xrd) and 0 otherwise. Only defined for firms with IndIPO equal to 1.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
RDS	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.
REV6	Earnings forecast revisions	Chan, Jegadeesh and Lakonishok	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.
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.
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.

Continued on next page

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)^2. 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.
roaq	Return on assets	Balakrishnan, Bartov and Faurel	2010	1976	2005	Quarterly net income (ibq) divided by lagged total assets (atq). Exclude if price less than 1.
RoE	Return on Equity	Haugen and Baker	1996	1979	1993	Net income (ni) over book value of equity (ceq). Exclude if price less than 5.
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.
sfe	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.
sgr	Annual sales growth	Lakonishok et al	1994	1968	1990	Sales (sale) relative to t-1.
ShareIss1Y	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.
ShareIss5Y	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.
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	Datair et al	1998	1962	1991	Sum of monthl share trading volume (vol) over the previous three months, scaled by 3 times common shares outstanding (shrout). Exclude if common shares outstanding changed over the previous three months, or if SIC code between 6000 and 6999. Trim the highest and lowest 1% of observations.
ShortInterest	Short Interest	Dechow, Pathak and Ritter	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.

Continued on next page

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
sinAlgo	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 >= 2100 & sic <= 2199, sic >=2080 & sic <= 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 >= 2000 & sic <= 2046) OR (sic >= 2050 & sic <= 2063) OR (sic >= 2070 & sic <= 2079) OR (sic >= 2090 & sic <= 2092) OR (sic >= 2095 & sic <= 2099) OR (sic >= 2064 & sic <= 2068) OR (sic >= 2086 & sic <= 2087) OR (sic >= 920 & sic <= 999) OR (sic >= 3650 & sic <= 3652) OR sic == 3732 OR (sic >= 3931 & sic <= 3932) OR (sic >= 3940 & sic <= 3949) OR (sic >= 7800 & sic <= 7833) OR (sic >= 7840 & sic <= 7841) OR (sic >= 7900 & sic <= 7911) OR (sic >= 7920 & sic <= 7933) OR (sic >= 7940 & sic <= 7949) OR sic == 7980 OR (sic >= 7990 & sic <= 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	Smile of slope	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.
SP	Sales-to-price	Barbee, Mukherji and Raines	1996	1979	1991	Ratio of annual sales (sale) to market value of equity.
Spinoff	Spinoffs	Cusatis, Miles and Woolridge	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.
std_dolvol	Dollar volume volatility	Chordia, Roll and Subrahmanyam	2001	1966	1995	Standard deviation of log daily dollar trading volume (abs(prc*vol)). Exclude if NASDAQ stock.
std_turn	Turnover volatility	Chordia, Roll and Subrahmanyam	2001	1966	1995	Standard deviation of daily turnover (vol/shrout). Exclude if NASDAQ stock.
SurpriseRD	Unexpected R&D increase	Eberhart, Maxwell and Siddique	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.

Continued on next page

Table A.1: (continued)

Acronym	Description	Author(s)	Pub year	Sample Start	Sample End	Description
tang	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 >= 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, Tax is defined as 1 if the numerator of the ratio is positive. Exclude if price less than 5.
UpForecast	Up Forecast	Barber, Lehavy, McNichols and 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 CF over the past 60 months. Exclude if less than 24 months of data available, or NASDAQ stock, price less than 5 or shrcd > 11.
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 and 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, Richardson and Sloan	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 $48*10^5$), multiplied by 21/number of trading days per month. Zerotrade is the 6-month average of that variable.
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.

A.2. Additional Estimation Figures

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

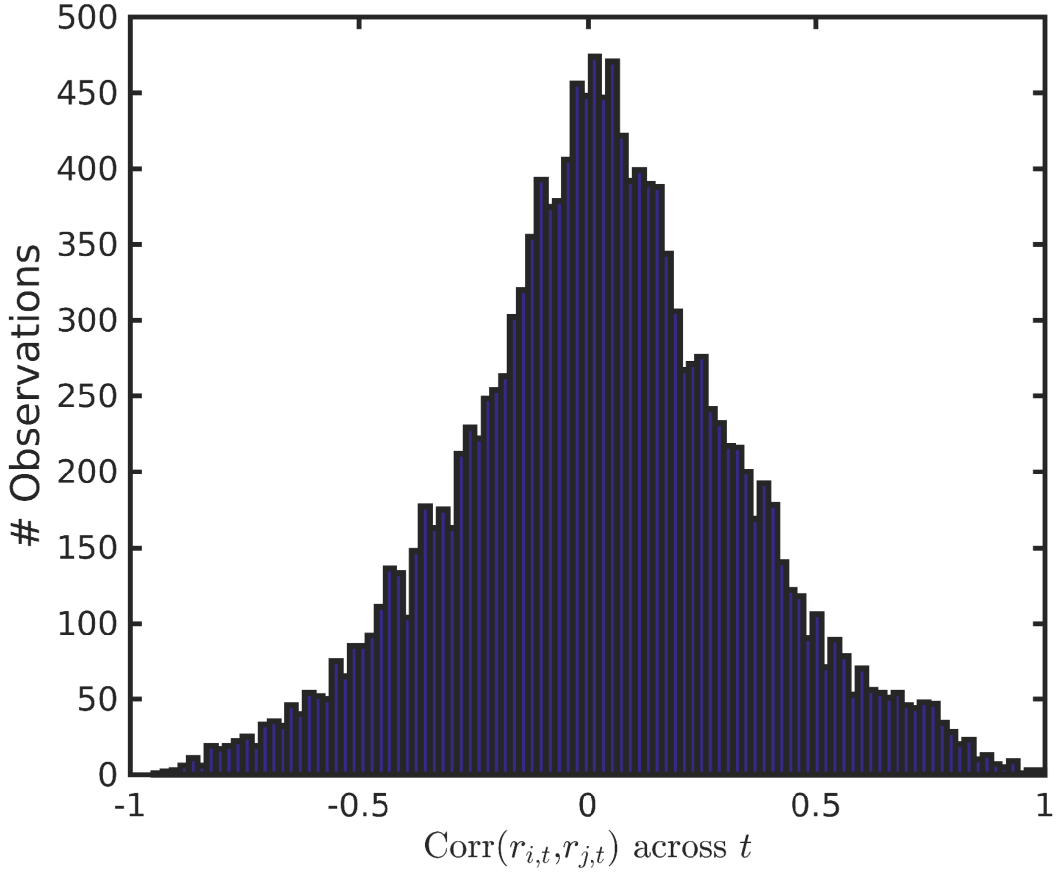


Figure A.2: Likelihood function. This figure plots the likelihood function near our maximum likelihood estimate (Table 3).

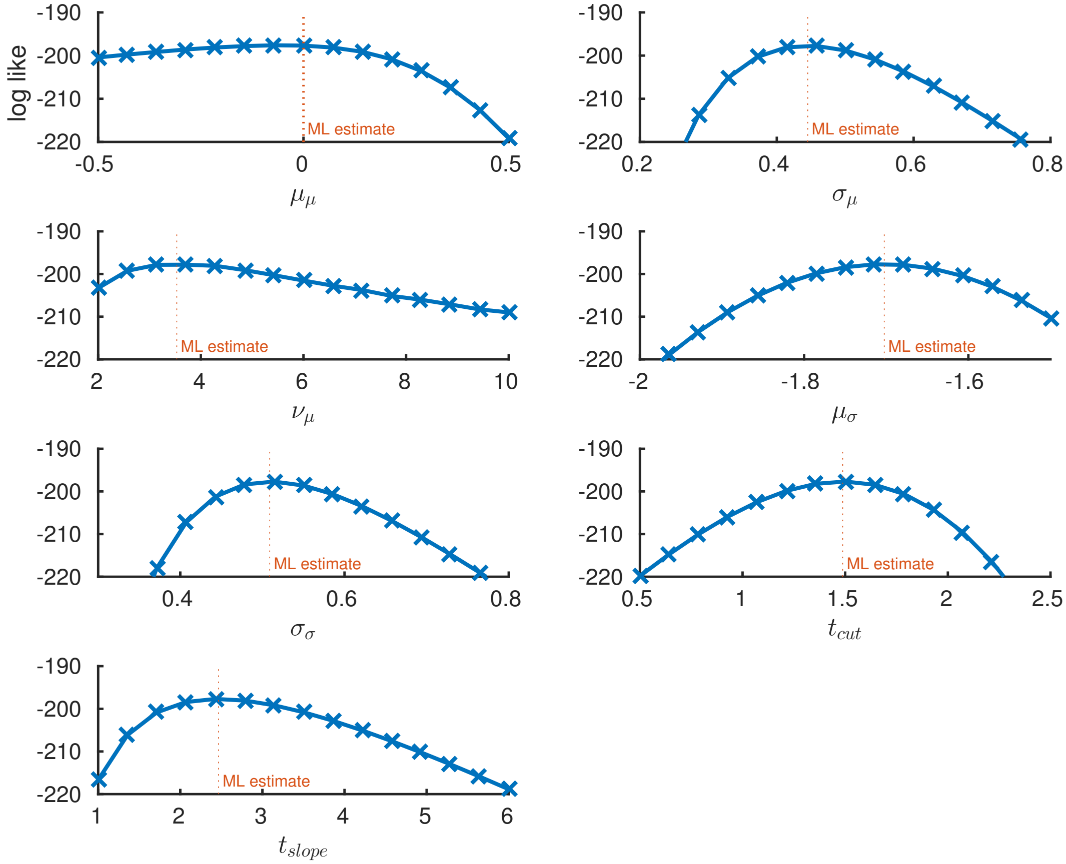
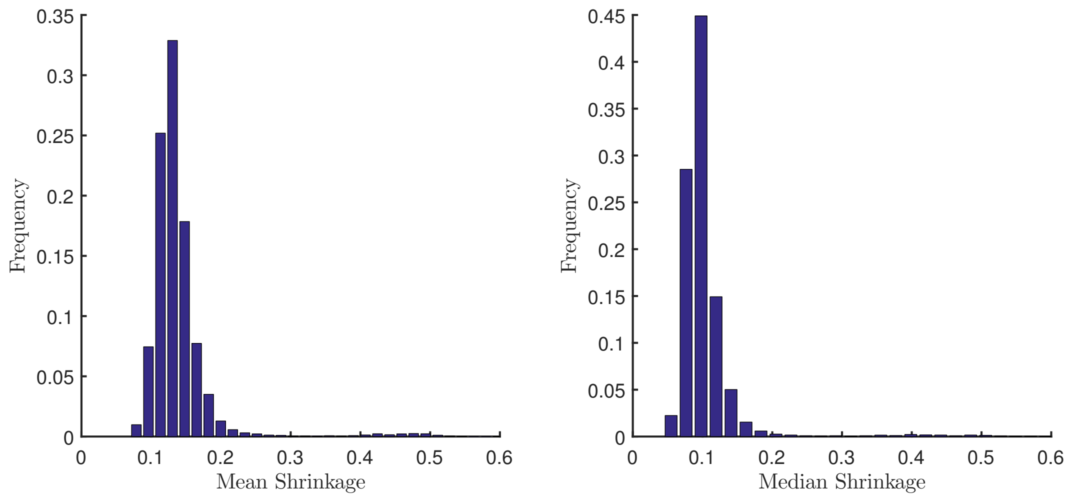


Figure A.3: Bootstrapped distribution of mean and median shrinkage. This figure plots details of the mean shrinkage standard errors in Table 3.



A.3. 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, Y. 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

$$Pr(\text{null}_i|t_i > t_h) = \frac{Pr(t_i > t_h|\mu_i \leq 0)Pr(\mu_i \leq 0)}{Pr(t_i > t_h)} \tag{31}$$

$$= \int_{-\infty}^{0} d\mu f_\mu(\mu|\theta)\left[1-\Phi\left(t_i - \frac{\mu_i}{\sigma_i}\right)\right]\frac{Pr(\mu_i \leq 0)}{Pr(t_i > t_h)}. \tag{32}$$

where $f_\mu(\mu|\theta)$ is the distribution of true means that we estimate in Section 3.3.

A.4. 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

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

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.

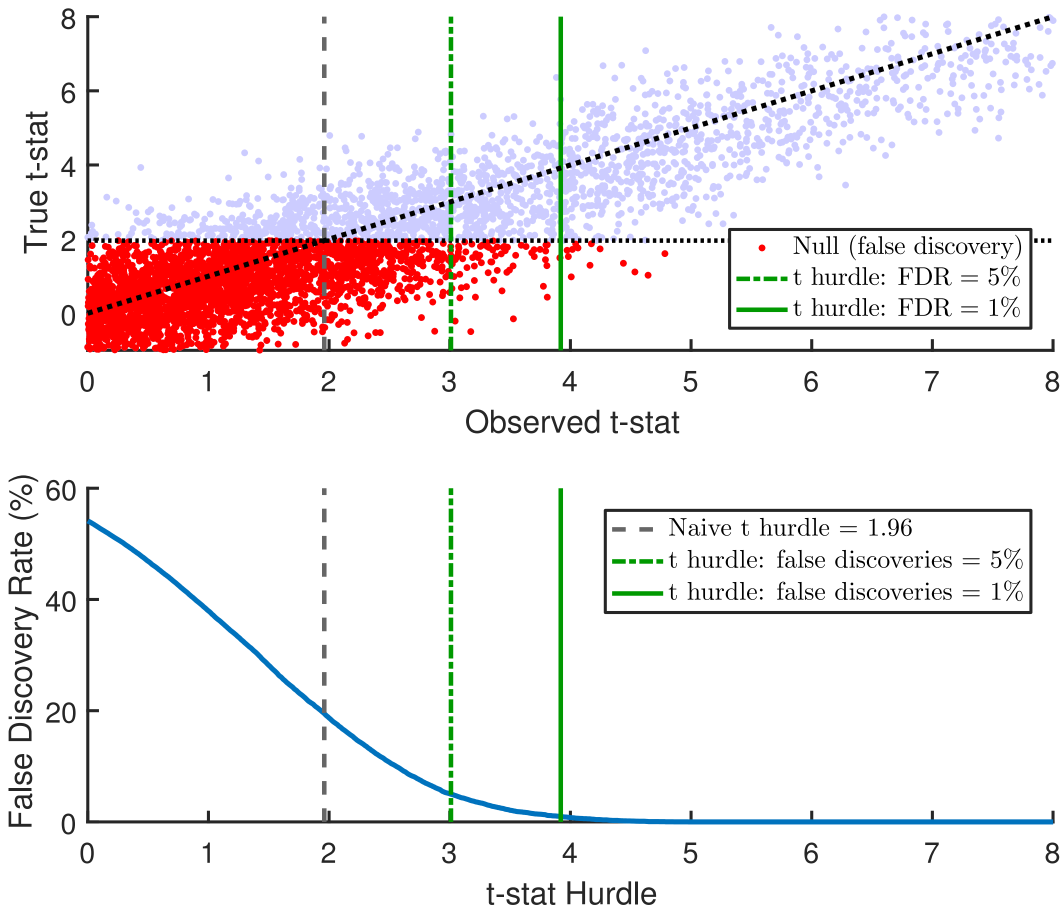
Figure A.4 illustrates the FDR implied by the null (33). The top panel shows a scatterplot of published true t-stats against observed t-stats from simulating the estimated model. If there was no publication bias, observed t-stats would be an unbiased estimate of the true t-stat, and the scatterplots 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.

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.

Figure A.4: Multiple Tests of the Null: True t-stat < 1.96. We simulate narrative portfolios using our estimated model (Table 3). 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.



References

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

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.

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

Dichev, Ilia D and Joseph D Piotroski. "The Long-Run Stock Returns Following Bond Ratings Changes". *The Journal of Finance* 56.1 (2001), pp. 173–203.

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.

Fama, E.F. and K.R. French. "The cross-section of expected stock returns". *Journal of finance* (1992), pp. 427–465.

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

Galiani, Sebastian, Paul Gertler, and Mauricio Romero. *Incentives for Replication in Economics*. Working Paper 23576. National Bureau of Economic Research, July 2017. DOI: 10.3386/w23576. URL: http://www.nber.org/papers/w23576.

Green, Jeremiah, John RM Hand, and Frank Zhang. "The remarkable multidimensionality in the cross-section of expected US stock returns". *Available at SSRN 2262374* (2014).

Green, Jeremiah, John RM Hand, and X Frank Zhang. "The supraview of return predictive signals". *Review of Accounting Studies* 18.3 (2013), pp. 692–730.

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.

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

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

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).

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.

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.