Estimating The Anomaly Base Rate[*]

Alex Chinco[†], Andreas Neuhierl[‡], and Michael Weber[§]

February 10, 2020

Abstract

The anomaly zoo has caused many to question whether researchers are using the right tests of statistical significance. We point out that even if researchers are using the right tests, they will still draw the wrong conclusions from their econometric analyses if they start out with the wrong priors—i.e., if they start out with incorrect beliefs about the ex ante probability of encountering a tradable anomaly. We call this ex ante probability the *'anomaly base rate'*. We propose a way to estimate this quantity by combining two key insights: #1) Empirical-Bayes methods capture the implicit process by which researchers form priors about the likelihood that a new variable is a tradable anomaly based on their past experience with other variables in the anomaly zoo. #2) Under certain conditions, a one-to-one mapping exists between these prior beliefs and the best-fit tuning parameter in a penalized regression. The anomaly base rate varies substantially over time, and we study trading-strategy performance to verify our estimation results. If you trade on two variables with similar one-month-ahead return forecasts in different base-rate regimes (low vs. high), the variable in the low base-rate regime consistently underperforms the otherwise identical variable in the high base-rate regime.

JEL Classification: C12, C52, G11

Keywords: Return Predictability, Data Mining, Empirical Bayes, Penalized Regressions

1 Introduction

Imagine you are a financial economist sitting down at your weekly research seminar. Today's speaker has regressed each stock's excess return, R_n, on lagged values of a new variable, X_n:

$$R_n = \hat{\mu} + \hat{\beta} \cdot X_n + \hat{\varepsilon}_n \qquad \text{for stocks } n = 1, \ldots, (N+1). \tag{1}$$

$\hat{\mu}$ is the mean excess return in the current month, $\hat{\beta}$ is the estimated slope coefficient, and $\hat{\varepsilon}_n$ is the residual for the nth stock. Before running this regression, the speaker standardized X_n to have zero mean and unit variance in the cross-section each period. And, on the current slide, he is reporting a positive estimated slope coefficient, $\hat{\beta} > 0$, that is statistically significant at the 1% level.

Predictive regressions and trading-strategy returns are two sides of the same coin (Fama, 1976). So, the $\hat{\beta}$ in Equation (1) can be viewed as the realized return to a zero-cost portfolio that is long stocks with high X_n values last month and short stocks with low X_n values last month:

$$\hat{\beta} \overset{\text{def}}{=} \frac{\widehat{\text{Cov}}[R_n, X_n]}{\widehat{\text{Var}}[X_n]} = \tfrac{1}{N} \cdot \sum_n (R_n - \hat{\mu}) \cdot (X_n - 0). \tag{2}$$

Thus, an estimated $\hat{\beta} > 0$ implies not only that high-X_n stocks last month, $(X_n - 0) > 0$, tended to have high excess returns this month, $(R_n - \hat{\mu}) > 0$, but also that it was profitable to trade on X_n in the past. If you had bought high-X_n stocks last month, then you would have earned high returns this month. Since investors should exploit and thereby eliminate such an arbitrage opportunity, this finding might suggest a gap in our economic understanding of how financial markets work.

However, the speaker's estimated slope coefficient is just that... an estimate. An estimated $\hat{\beta} > 0$ implies that it *was* profitable to trade on X_n, but you want to know whether it *will be* profitable to trade on X_n going forward. And, this change in verb tense makes a world of difference since "many of the results being published fail to hold up in the future" (Harvey, 2017). Predictability often turns out to be a decidedly in-sample phenomenon (McLean and Pontiff, 2016; Linnainmaa and Roberts, 2018). X_n still might not represent a tradable anomaly even if the seminar speaker estimates a large, positive, and statistically significant slope coefficient. His results might just be a fluke, a chance event not warranting any change in our economic understanding.

Your task as an audience member is to figure out $\Pr[\,anom\,|\,signif\,]$, the probability that X_n is a tradable anomaly given the speaker's statistically significant results. Bayes' rule tells you how:

$$\Pr[\,anom\,|\,signif\,] = \left(\frac{\Pr[\,signif\,|\,anom\,]}{\Pr[\,signif\,]} \right) \times \Pr[\,anom\,]. \tag{3}$$

Multiply the ex ante probability that X_n is a tradable anomaly, $\Pr[\,anom\,]$, times the relative increase in the base rate due to the speaker's statistically significant results, $\left(\frac{\Pr[\,signif\,|\,anom\,]}{\Pr[\,signif\,]} \right)$.

The current literature on cross-sectional predictability is primarily concerned with how data mining distorts the first term on the right-hand side of Equation (3), $\left(\frac{\Pr[\,signif \,|\, anom\,]}{\Pr[\,signif\,]} \right)$. Computing this 'Bayes factor' requires knowing the correct unconditional probability of a significant result, $\Pr[\,signif\,]$. And, knowing the correct unconditional probability requires using the right *p-value*:

$$\Pr[\,signif\,] = \Pr[\,signif \,|\, anom\,] \times \Pr[\,anom\,] + \underbrace{\Pr[\,signif \,|\, \neg anom\,]}_{=p\text{-}value} \times \Pr[\,\neg anom\,], \qquad (4)$$

where, $\neg$ denotes the logical not symbol. Earlier we said that the seminar speaker reported an estimated slope coefficient, $\hat{\beta} > 0$, that was statistically significant at the 1% level—i.e., a *p-value* = 0.01 and a *t-statistic* = 2.54. A *p-value* = 0.01 means that there is only a 1/100 chance of a single variable having such an estimate by pure chance under the null hypothesis of no predictability. But, if the seminar speaker also ran regressions involving 99 other variables, then on average you would expect one of these 100 regressions to produce a *p-value* = 0.01. Thus, by ignoring the effects of data mining, financial economists tend to think that significant results are less common than they actually are. And, as a result, they tend to over-react when they see one.

Using a higher *t-statistic* cutoff (or equivalently, a lower *p-value* cutoff) when assessing statistical significance is one way of correcting this error (Harvey et al., 2016). Financial economists can and should be doing this. But, it is also worth remembering that the Bayes factor is not the only term on the right-hand side of Equation (3). Someone who enters the seminar room with wildly inaccurate priors, $\Pr[\,anom\,]$, is going to draw the wrong conclusions from the speaker's results even if they correct their Bayes factor for data mining. And, it seems like some of your colleagues must have entered the seminar room today with wildly inaccurate priors. Before the talk even started, you already knew who would leave the room skeptical. You could also point to a group of your colleagues known to be much more receptive to evidence of cross-sectional predictability. Regardless of where you stand, it is clear that both priors cannot be right. We are not the first to point out the importance of prior beliefs in asset pricing (Shanken, 1987; Harvey and Zhou, 1990). But, the current financial-economics literature gives no guidance about which priors to use.

So, how did you estimate an anomaly base rate for use in today's seminar? We recognize that there is a whiff of paradox about the idea of estimating one's priors. But, the notion is not completely absurd. Penalized-regression procedures such as Ridge (Hoerl and Kennard, 1970) can be viewed as something like this when they solve the optimization problem below:

$$\min_{\beta} \left\{ \tfrac{1}{N} \cdot \sum_n \left(R_n - \hat{\mu} - \beta \cdot X_n \right)^2 + \lambda \cdot \beta^2 \right\}. \qquad (5)$$

$\lambda \geq 0$ is called the 'tuning parameter'. And, it is well-known that this optimization problem has a clear Bayesian interpretation when the true slope coefficients are drawn from the appropriate

distribution. For Ridge, the appropriate distribution is $\beta_i^\star \sim \text{Normal}[0, \sigma^2]$. Under these conditions, the Ridge's success can be viewed as the result of incorporating one's prior beliefs about σ^2, the variance of the market conditions leading to predictability (Pástor, 2000; DeMiguel et al., 2009; Kozak et al., 2018b).

In this simple setup, the best-fit tuning parameter (λ) is inversely related to the variance of one's prior beliefs about the market conditions for predictability (σ^2). This mathematical connection between the best-fit tuning parameter and prior beliefs is a promising avenue, which suggests it might be possible to estimate this prior using a penalized-regression framework. Unfortunately, the economic meaning of this connection is not entirely clear. Suppose λ is inversely related to the variance of market conditions for return predictability. There is no general theory of these market conditions. The variables that seem to predict the cross-section of expected returns are constructed using a wide variety of different data sets and typically have little to do with one another economically (Lewellen et al., 2010). So, how should we interpret the associated prior variance, σ^2? What exactly are the corresponding prior beliefs beliefs about?

With these difficulties in mind, this paper proposes a way to estimate the prevailing anomaly base rate. Here is the logic. The existence of the anomaly zoo tells us that research practices matter. So, maybe your prior beliefs walking into the seminar room today had more to do with your first-hand knowledge of current research practices than with the market conditions leading to predictability. After all, as a well-trained financial economist, you know the econometric tips and tricks that other well-trained financial economists use. You yourself frequently employ these same techniques. So, your past experience with other variables in the anomaly zoo contains information about the rate at which these techniques produce tradable anomalies rather than spurious predictors.

When reframed this way, the idea of estimating one's priors no longer seems quite so paradoxical. Financial economists are constantly evaluating their current research practices based on how well published results hold up out of sample. This is an integral part of doing good research, and it is a process that is captured by empirical-Bayes methods (Robbins, 1956; Efron and Morris, 1972, 1973, 1975). "The essential empirical-Bayes task [is] learning an appropriate prior distribution from ongoing statistical experience rather than knowing it by assumption (Efron, 2013)." Or, in the words of Efron and Hastie (2016, p. 77), "large data sets of parallel situations carry within them their own Bayesian information" about the appropriate prior. And, what is the anomaly zoo if not a large data set of many parallel situations, one for each variable in the published literature $i = 1, \ldots, I$ that seems to predict the cross-section of expected returns?

To exploit this reframing, we model the true predictive strength of each of variable in the anomaly zoo as being drawn from a common normal distribution:

$$\beta_i^\star \overset{\text{iid}}{\sim} \text{Normal}[0, \sigma^2] \qquad \text{for each variable } i = 1, \ldots, I. \tag{6}$$
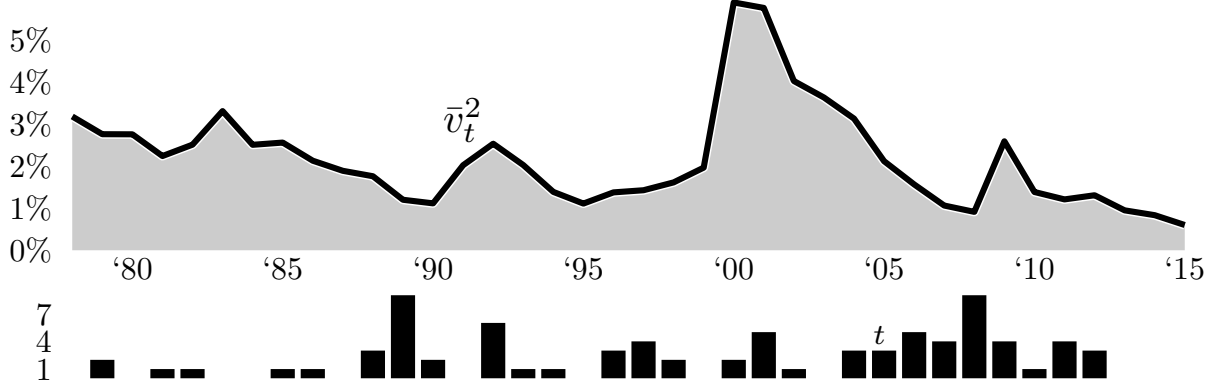
Figure 1. Forecasted $\bar{v}_t^2$. Top Panel: Average forecasted prior variance each year. For each variable $i = 1, \ldots, I_t$ first introduced prior to month t, we estimate the in-sample parameter $\hat{v}_{i,t'}^2$ for all months $t' < t$ via Proposition 2.2. We then make a one-month-ahead forecast for month t, $\bar{v}_{i,t}^2$, by fitting an AR(3) model in months $\{t - 60, \ldots, t - 1\}$. Finally, we average the forecasts for all past variables to compute $\bar{v}_t^2$. **Bottom Panel:** Number of new variables published in the academic literature each year, ΔI_t. Sample: June 1978 to June 2015.

The idea here is to parameterize the efficacy of current research practices by the variance of this distribution. If $\sigma^2 = 0$, current research practices only produce spurious predictors. As σ^2 increases, it becomes more likely that a randomly selected variable in the anomaly zoo is a tradable anomaly. The zero-mean assumption reflects the idea that some variables positively predict future returns while others negatively predict future returns—e.g., medium-term momentum vs. short-term reversal. We provide a more detailed discussion of the assumptions involved in Equation (6) below.

Once we view σ^2 as related to research methods and not market conditions, this quantity has a clear economic meaning. It then makes sense to invert the best-fit Ridge penalty, λ, to estimate the anomaly base rate. We can use empirical-Bayes logic to capture the way in which researchers combine their past experiences with other variables to inform their priors. We apply this approach to monthly data on 85 different variables in the anomaly zoo. Let $I_t \leq 85$ denote the number of variables published prior to month t. Each month, we run separate univariate Ridge regressions of excess returns over the risk-free rate on lagged values of each variable $i = 1, \ldots, I_t$. Then, we invert the best-fit tuning parameter from each of these regressions to produce variable-specific estimates for σ^2 in each month t, which we call $\hat{v}_{i,t}^2$. We use these variable-specific estimates to construct a market-wide base rate by making a one-month-ahead forecast of each variable-specific estimate and then averaging these out-of-sample forecasts to find the aggregate prior variance, $\bar{v}_t^2$. This quantity represents our estimate for the ex ante probability that the next predictor will be a tradable anomaly, that is, σ^2.

The top panel of Figure 1 shows how this prior variance has evolved over time. Note that $\bar{v}_t^2$ is not just counting the number of variables recently added to the academic literature. The bottom

panel of Figure 1 shows that the anomaly base rate can be low at times when lots of new variables are being 'discovered' if these new variables tend to have poor subsequent performance—i.e., if they tend to be spurious predictors. $\bar{v}_t^2$ is also not just a stand-in for return volatility or related variables as we show below. $\bar{v}_t^2$ is measuring the strength of predictors in the anomaly zoo as we argue in equation (6). Finally, recall that our statistical approach is agnostic about the market conditions that lead to cross-sectional return predictability. So, the variables we are calling 'tradable anomalies' do not necessarily have to be pricing errors. We are using this term as a convenient short-hand since it reflects the original motivation for looking for such variables as described in Equation (2).

A natural intuition is that the anomaly base rate should be steadily declining over time. And, if there were a fixed number of tradable anomalies to discover and financial economists were repeatedly using the exact same empirical methods to look for them in the exact same data sets, then this is exactly what we would find. In that world, a researcher who entered the game early would be more likely to discover and publish a tradable anomaly (high base rate). Whereas, later researchers would face a lower anomaly base rate since most of the low-hanging fruit would have already been published. However, this intuition does not take into account the fact that empirical methods and data sets evolve over time. Groups of new candidate predictors become available; growth in computing power displays jumps; new statistical methods get deployed. These sorts of changes lead to time-series variation in the anomaly base rate.[1]

This paper offers a practical means of estimating the anomaly base rate and its evolution over time. The anomaly zoo contains a few tradable anomalies and many more spurious predictors. Our core economic insight is that a researcher's past experience with these variables must contain information about the efficacy of his research practices. So, in the absence of an overarching theory for cross-sectional predictability, a researcher likely uses this information to inform the prior beliefs he uses when evaluating new variables. We have deliberately used a highly stylized statistical setup. But, numerical simulations show our approach is readily extensible. By introducing a clear and simple way of estimating the anomaly base rate, our hope is that future research can shed more light on the drivers of this important economic object.

We show via a purely out-of-sample trading-strategy performance shows that, in spite of its transparent simplicity, our statistical approach still produces actionable estimates of the anomaly base rate. We consider trading on two variables in different base-rate regimes (one low, the other high) with the same one-month-ahead return forecast. Theory suggests that the variable in the low base-rate regime should under-perform the otherwise identical variable in the high base-rate regime. And, this is exactly what we find in the data. Let $direction_i$ be the direction of the ith variable. e.g., $direction_i = +1$ for medium-term momentum while $direction_i = -1$ for short-term reversals. And, let $\bar{\beta}_{i,t}$ be the one-month-ahead return forecast for the ith variable.

---

[1] We would like to thank an anonymous referee for helping to develop this intuition more clearly.

We start with a benchmark strategy that holds all variables in the anomaly zoo that have $\bar{\beta}_{i,t} \cdot direction_i > 1\%$ in month t. The 1%-per-month threshold captures the idea, to be a tradable anomaly, a variable must generate enough predictability to cover implementation costs. The specific threshold level is not important, and our results are robust to using a wide range of thresholds. Rather, the key thing is that, when the prior variance implied by all other variables $i' \neq i$ is small, $\bar{v}^2_{\neg i,t} \stackrel{\text{def}}{=} \frac{1}{I_t - 1} \cdot \sum_{i' \neq i} \bar{v}^2_{i',t} \approx 0$, you should expect the ith variable to have a small amount of predictability too, $\beta^\star_{i,t} \approx 0$. As a result, you should be less willing to trade on the ith variable if $\bar{\beta}_{i,t} \cdot direction_i \gg threshold$. Following this logic, our base-rate-adjusted strategy only holds variables $i = 1, \ldots, I_t$ that still have return forecasts in excess of 1% after adjusting for $\bar{v}^2_{\neg i,t}$.

This base-rate-adjusted strategy delivers net excess returns of 0.74% per month and has an annualized out-of-sample Sharpe ratio of 0.57. The strategy works by discarding variables in the benchmark strategy that are implausibly strong predictors given the prevailing anomaly base rate. And, we find that an alternative strategy which only invests in these discarded variables has net returns of just 0.18% per month and an annualized out-of-sample Sharpe ratio of 0.11. This drop in performance persists even when we control for the magnitude and standard error of each variable's forecasted $\bar{\beta}_{i,t}$ to control for estimating uncertainty. In other words, we find that trading on the same one-month-ahead return forecast is less profitable in months when the anomaly base rate is low just as predicted by the theory. We also verify that the base-rate-adjusted strategy is also no more likely to hold high-turnover predictors.

1.1 Related Literature

This paper connects to three strands of the literature on cross-sectional predictability.

Data Mining. There is a large financial-econometrics literature on data mining (Lo and MacKinlay, 1990; Ferson et al., 1999; Sullivan et al., 1999; White, 2000; Barras et al., 2010; Bajgrowicz and Scaillet, 2012; McLean and Pontiff, 2016; Harvey et al., 2016; Yan and Zheng, 2017; Harvey and Liu, 2018a,b,c; Linnainmaa and Roberts, 2018). We want to emphasize that, while clearly related to the topic, this is not a paper about data mining. We are doing something different. We are proposing a method of estimating the anomaly base rate, $\Pr[\,anom\,]$. In fact, our analysis specifically takes the Bayes factor, $\left(\frac{\Pr[\,signif\,|\,anom\,]}{\Pr[\,signif\,]}\right)$, as given. You should use the many excellent papers listed above to adjust the Bayes factor for data mining when interpreting a seminar speaker's results. Then, you should apply this data-mining adjusted Bayes factor to the prevailing anomaly base rate, which we show how to estimate in this paper.

Factor Structure. Even if there are many statistically significant predictors, it still might be possible to summarize the information in all these variables using a few well-chosen factors. Papers on the factor structure of predictors try to simplify investors' lives by collapsing the anomaly zoo down into a few manageable variables. Ideally, once you condition on these variables, there would be

no incremental value in considering any other variables when forecasting returns. Researchers typically try to accomplish this goal using some form of principal-component analysis (Kelly et al., 2017; Green et al., 2017; Kelly et al., 2018; Lettau and Pelger, 2018).

Penalized Regressions. Many recent papers have used penalized-regression procedures to solve asset-pricing problems (Pástor, 2000; DeMiguel et al., 2009; Bryzgalova, 2017; Feng et al., 2017; Freyberger et al., 2017; Ledoit and Wolf, 2017; Chinco et al., 2018; Kozak et al., 2018b). We are using these same techniques with an entirely different goal in mind: estimating the anomaly base rate. Here is the key distinction. If there were only a handful of variables to choose from, then these earlier papers would not need to use a penalized regression. By contrast, estimating the correct anomaly base rate is important regardless of the size of the anomaly zoo. To our knowledge, this is the first instance where insights from machine learning are being used to shape our understanding of a more fundamental asset-pricing problem, a problem that exists even in a low-dimensional setting.

2   Statistical Approach

Let us return to our motivating example. Imagine you are a financial economist sitting at your weekly research seminar. The speaker is showing evidence that it was profitable to trade on a new variable, X_n. You want to figure out the probability of X_n being a tradable anomaly going forward. To do this, you need to start out with the anomaly base rate. Then, you need to update this prior belief based on the speaker's results. Your financial-econometrics training tells you how to do the updating. This section describes a statistical approach for estimating the anomaly base rate.

2.1   Inference Problem

We begin by defining the inference problem you face in greater detail. What precisely is the anomaly base rate? And, what does it mean to learn about this quantity from past experience?

Data-Generating Process. Suppose there are $(N+1)$ stocks indexed by $n = 1, \ldots, (N+1)$. Let R_n denote the excess return of the nth stock in the current month. Each stock's excess return this month is related to the lagged value of the ith variable in the previous month, $X_{n,i}$:

$$R_n = \mu^\star + \beta_i^\star \cdot X_{n,i} + \varepsilon_{n,i}^\star \qquad \text{for each variable } i = 1, \ldots, I. \tag{7}$$

$\mu^\star$ is the average excess return in the current month, $\beta_i^\star$ is the true slope coefficient associated with the ith variable, and $\varepsilon_{n,i}^\star \overset{\text{iid}}{\sim} \text{Normal}[0, N \cdot se_i^2]$ is the idiosyncratic shock for the nth stock for the ith variable. In other words, $\varepsilon_{n,i}^\star$ represents the portion of the nth stock's excess return that is not explained by the ith variable. This residual might come from idiosyncratic shocks that are fundamentally random or represent the effects of other variables $i' \neq i$.

We normalize the lagged values of each variable to have zero mean and unit variance in the

cross-section each month: $\frac{1}{N+1} \cdot \sum_n X_{n,i} = 0$ and $\frac{1}{N} \cdot \sum_n (X_{n,i} - 0)^2 = 1$. This normalization is done period-by-period for each variable, does not involve any future information, and hence it does not affect investability. All it does is ensure that in-sample estimates of $\hat{\beta}_i$ are comparable across variables. Consider the following example to see why this is important. Suppose variable 1 corresponds to a portfolio that is long/short the top/bottom deciles while variable 2 corresponds to a portfolio that is long/short the top/bottom quintiles. Variable 1 will have variance $\widehat{\mathrm{Var}}[X_{n,1}] = \frac{1}{10} \cdot (+1)^2 + \frac{1}{10} \cdot (-1)^2 = 1/5$; whereas, variable 2 will have variance $\widehat{\mathrm{Var}}[X_{n,2}] = \frac{1}{5} \cdot (+1)^2 + \frac{1}{5} \cdot (-1)^2 = 2/5$. Thus, since $\hat{\beta}_i \stackrel{\text{def}}{=} \frac{\widehat{\mathrm{Cov}}[R_n, X_{n,i}]}{\widehat{\mathrm{Var}}[X_{n,i}]}$, we would expect that $\hat{\beta}_1 > \hat{\beta}_2$ even if both variables have equal predictive power, $\widehat{\mathrm{Var}}[X_{n,1}] = 1/5 < 2/5 = \widehat{\mathrm{Var}}[X_{n,2}]$.

Tradable Anomalies. We say that the ith variable is a tradable anomaly if the magnitude of its true predictive power exceeds some minimum performance threshold:

$$anom_i \stackrel{\text{def}}{=} \mathbf{1}\Big[|\beta_i^\star| > threshold\Big] \qquad \text{for } threshold \geq 0. \tag{8}$$

You can think about this minimum performance threshold as coming from trading costs (Novy-Marx and Velikov, 2015). Or, you can think about it more broadly as any form of implementation costs. If you are running a trading desk, then how strong does a predictor need to be before you redeploy scarce resources to trade on it? In our empirical analysis, we will typically set $threshold = 1\%$ per month for a trading strategy that is long/short the top/bottom deciles. However, we show in the online appendix that using a $threshold \in [0.5\%, 1.5\%]$ yields qualitatively similar results.

We take the absolute value of the true slope coefficient in Equation (8) because the sign of the $\beta_i^\star$ associated with a long-short strategy is arbitrary. Fama and French (1993) could just as easily have defined their HML factor using market-to-book rather than book-to-market ratios. This alternative choice would have flipped the sign of the associated slope coefficient but left everything else about the authors' analysis unchanged. Contrarian strategies, such as the long-term reversals (De Bondt and Thaler, 1985), will have $\beta_i^\star < 0$. What really matters is the magnitude of $\beta_i^\star$.

The statistical approach we describe in this section is agnostic about the market conditions leading to cross-sectional return predictability. So, although we will refer to strong predictors as tradable anomalies, their predictability does not necessarily have to be the result of a pricing error (Kozak et al., 2018a). Nevertheless, we use the term 'tradable anomaly' rather than 'sufficiently strong predictor' because we feel it captures the motivation behind researchers' interest in cross-sectional predictability as we outline in the introduction.

Research Practices. There is no single unified explanation for the origins of cross-sectional predictability. The variables that seem to predict the cross-section of expected returns are all quite different. Cross-sectional predictability might be the result of any number of different limits-to-arbitrage models (Barberis and Thaler, 2003; Gromb and Vayanos, 2010). There is no

shortage of risk-based explanations to choose from either (Fama and French, 1996). And, different variables often use entirely different data sources. Some variables only involve past market data (e.g., medium-term momentum; Jegadeesh and Titman, 1993). Others use only accounting data (e.g., investment growth; Titman et al., 2004). Why would there be a single economic explanation for the predictability associated with both medium-term momentum and investment growth? The only common theme connecting these two variables is Sheridan Titman.

In the absence of a general theory for the origins of cross-sectional predictability, your prior beliefs about the likelihood of a tradable anomaly likely have more to do with your first-hand knowledge of current research practices. Even if the economic rationales for medium-term momentum and investment growth are completely different, you know that both variables were discovered by financial economists using a shared set of research practices. We have all attended the same PhD programs. We have taken the same courses. We have all been trained by the same advisers. We have all been given access to the same data sources. While computing power has grown over time, at any one point in time all researchers have access to similar amounts of processing power. Thus, while some researchers are better at searching for new variables than others, everyone's discovery process is constrained by the same inputs.

We use a statistical model for the anomaly-discovery process to capture this commonality in research practices. Specifically, we assume that the strength of each variable is drawn from a common normal distribution as described in Equation (6). The key assumption embedded in this equation is that the efficacy of these research practices can be encapsulated by a single parameter σ^2, which governs the range of predictability that financial economists discover. A larger value of σ^2 means that researchers are more likely to discover strong predictors. It is not essential to assume normality or independence across variables as we show in Section 2.3 below. The mean-zero assumption is also not crucial as we verify using numerical analysis in Section 2.3 as well. We make these simplifying assumptions for ease of exposition. Our main goals are highlighting the importance of the anomaly base rate and proposing a simple way to estimate this economic object.

Inference Problem. If the strength of the ith variable is drawn as in Equation (6), then the parameter σ^2 controls the typical size of cross-sectional predictability. This assumption implies that calibrating your expectations for what today's seminar speaker will say during the next hour is tantamount to learning about the prevailing value for σ^2.

Proposition 2.1 (Inference Problem). *Suppose the true slope coefficients are drawn as in Equation (6), excess returns are generated by the process in Equation (7), and there exists some minimum performance threshold as in Equation (8), then the anomaly base rate is given by*

$$\Pr[\, anom_i \,] = 2 \cdot \Phi[\, -\, threshold/\sigma\,], \tag{9}$$

where $\Phi[\cdot]$ *represents the standard normal CDF.*

Suppose today's seminar speaker is presenting evidence that the ith variable is a very strong predictor, $|\hat{\beta}_i| \gg$ *threshold*. If your past experience with $(I-1)$ other variables tells you that $\sigma^2 \gg 0$, this result should strike you as quite plausible. $\sigma^2 \gg 0$ means that it is quite common for a researcher to draw a $\beta_i^\star$ far from zero. By contrast, if your past experience suggests that $\sigma^2 \approx 0$, today's result will strike you as highly implausible. You will rationally discount the seminar speaker's evidence, not because he did anything wrong, but because you think his result is very unlikely to begin with. $\sigma^2 \approx 0$ means that tradable anomalies are quite rare. But, your past experience could not have dictated that $\sigma^2 = 0$ exactly. If it did, then you would not have even bothered to show up to the seminar in the first place. There would have been nothing that the seminar speaking could say to convince you that $\beta_i^\star \neq 0$, that is, you had 'dogmatic priors'.

If researchers have been applying the same empirical techniques to the same collection of data sets for a long time already, then σ^2 will be close to zero. It is unlikely that these same techniques will suddenly uncover a new variable after years and years of trying. But, these exact same techniques can be given new life by the introduction of an important new data set, such as the Trade and Quote Database (TAQ). That being said, we do not have a full explanation for variation in σ^2 over time. Our approach is similar in spirit to the idea of estimating stock-market volatility (Andersen et al., 2003). We certainly do not think that there is a single unified explanation for all stock-market fluctuations. And, almost by definition, uncertainty can rise for any number of reasons. Better understanding the drivers of the anomaly-discovery process is an important and understudied topic.

We are proposing a simple model of the anomaly-discovery process. It is possible to extend this simple model in various ways to make it more realistic.[2] These extensions are promising avenues for future research, and we use numerical simulations to explore many of these possibilities in Section 2.3. We consider our paper only as a first step in explicitly studying the anomaly-discovery process. We hope our paper can spark interest in this topic.

Publication Bias. At first glance, the assumption of a normally distributed $\beta_i^\star$ seems to be at odds with the existence of publication bias. If researchers are engaging in data mining, why are there not peaks in the $\beta_i^\star$ distribution right above/below the positive/negative 5% p-value cutoffs? After all, this is the level of statistical significance that is needed to get a paper published.

Using a cross-sectional OLS regression like in Equation (1) to estimate $\beta_i^\star$ based on pre-publication data does produce an empirical distribution with two large masses immediately above/below the positive/negative 5% p-value cutoffs as shown by the red shaded region in Figure 2.

---

[2]We would like to thank several discussants for pointing out possible ways to do this. For example, you could model priors over Sharpe ratios rather than the level of predictability as suggested by Svetlana Bryzgalova and we plan to pursue these extensions in future work.
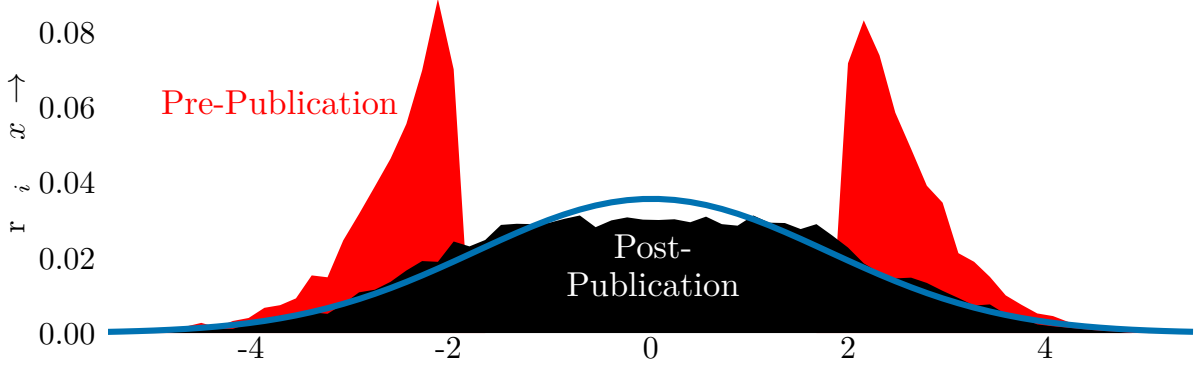
Figure 2. Publication Bias. x-axis: value of β_i. y-axis: empirical density computed using 10,000 simulations. Each simulation repeatedly samples from the data-generating process governed by Equations (6) and (7) with $(N + 1) = 1{,}000$ stocks and parameter values $\mu^\star = 0\%$, $\sigma = 1\%$, and $se_i = 1\%$ until the realized OLS estimate, $\hat{\beta}_i$, has $p < 0.05$. Red-shaded region reports empirical density of the resulting $\hat{\beta}_i$ estimates, which correspond to pre-publication empirical results. Black-shaded region reports empirical density of $\beta_i^\star$ values associated with each $\hat{\beta}_i$ estimate. These values correspond to long-run average post-publications empirical results. Blue line is a normal distribution with same mean and variance as the empirical $\beta_i^\star$ distribution.

This figure is based on 10,000 simulations of the data-generating process from Equations (6) and (7) with $(N + 1) = 1{,}000$ stocks and parameter values $\mu^\star = 0\%$, $\sigma = 1\%$, and $se_i = 1\%$. Each simulation starts by repeatedly sampling from this data-generating process until the realized in-sample OLS estimate, $\hat{\beta}_i$, has a $p < 0.05$. The red-shaded region reports the empirical density of the resulting $\hat{\beta}_i$ realizations conditional on statistical significance.

But, we are not modeling the $\hat{\beta}_i$ distribution in Equation (6); we are modeling the $\beta_i^\star$ distribution. The $\hat{\beta}_i$ distribution, which is depicted by the red-shaded region in Figure 2, is a function of both the $\beta_i^\star$ distribution and publication bias. If you are concerned about data mining, p-hacking, publication bias, etc. . . , then you must believe that these research practices have a big effect on the $\hat{\beta}_i$ distribution. So, there is no reason to expect the $\hat{\beta}_i$ distribution to look anything like the $\beta_i^\star$ distribution. This is precisely what the black-shaded region in Figure 2 shows.[3] This region reports the empirical density of the realized $\beta_i^\star$ values in those 10,000 simulations which lead to a significant $\hat{\beta}_i$ estimate at the 5% level. In spite of the existence of an extreme form of publication bias, the black-shaded region is not far from normal as depicted by the blue curve representing a normal density with the same mean and variance.[4]

Empirical Bayes. The existence of the anomaly zoo is a sign that financial economists have adopted bad research practices. But, if these research practices matter, then they must be affecting

---

[3]We would like to thank our discussant, Yan Liu, for raising these points.

[4]The black distribution is slightly flatter around zero. But, we verify in Figure 4 that our econometric estimator is robust to sampling $\beta_i^\star$ from this flattened post-publication distribution rather than the assumed normal distribution.

your prior beliefs about the likelihood of encountering a tradable anomaly in today's seminar. Your past experience with other published results in the anomaly zoo—some tradable anomalies and many more spurious predictors—must contain valuable information about how often the variables produced by prevailing research practices turn out to be tradable anomalies going forward.

Thus, the cross-sectional asset-pricing literature represents a large data set of parallel past situations for you to draw on, one for each variable you have encountered over the course of your career $i = 1, \ldots, I$. We use empirical-Bayes methods to formalize the process by which researchers convert past experience with such variables into a prior for next time around. As Efron and Hastie (2016, p. 88) writes, empirical-Bayes methods "remove the usual Bayesian scaffolding. In place of a reassuring prior, the statistician [puts] his or her faith in the relevance of the 'other' cases in a large data set to the direct case of interest." After being introduced by Robbins (1956), the first major work on empirical-Bayes methods was a series of papers by Efron and Morris (1972, 1973, 1975).[5]

Empirical-Bayes applications all have the same basic structure (cf. Maritz, 1970, p. 13). They start with a problem involving repeated sampling from an unknown prior distribution, $g[\sigma^2]$. This repeated sampling yields unseen realizations, $\{\beta_1^\star, \ldots, \beta_I^\star\}$. Each of these unseen realizations, $\beta_i^\star$, provides an observation, $\hat{\beta}_i \sim f_{\beta_i^\star}[\cdot]$, drawn from a known probability family. On the basis of these observations, $\{\hat{\beta}_1, \ldots, \hat{\beta}_I\}$, a researcher wishes to solve some inference problem that would be easy if only he knew the true prior distribution, $g[\sigma^2]$, e.g., in today's seminar, you want to estimate the true slope coefficient associated with the ith variable, $\beta_i^\star$, after seeing the speaker's in-sample OLS estimate, $\hat{\beta}_i$—a problem that would be easy if only you knew the correct choice of priors.

Here is how empirical Bayes solves this problem. First, convert the information about each other variable $i' \neq i$ into an estimated value for σ^2, which we will denote $\hat{v}_{i'}^2$. Then, since each variable in the anomaly zoo represents a parallel situation with some noise, take the average of these $(I - 1)$ variable-specific estimates, $\hat{v}_{\neg i}^2 \overset{\text{def}}{=} \frac{1}{I-1} \cdot \sum_{i' \neq i} \hat{v}_{i'}^2$. Finally, plug this average value into your prior distribution, $g[\hat{v}_{\neg i}^2]$. These steps allow you to "[learn the] appropriate prior distribution from ongoing statistical experience rather than knowing it by assumption (Efron, 2013)."

The existing Bayesian asset-pricing research interprets priors as having something to do with the market conditions for predictability. But, there is no general theory of these conditions, which makes it hard to make sense of such priors. As a result, this literature offers no guidance about which priors to use. We resolve this paradox by pointing out that, in the absence of a general theory,

---

[5]There is some ambiguity about naming conventions. As Efron (2012, p. 14) writes, "Robbins reserved the name 'empirical Bayes' for situations where a genuine prior distribution was being estimated, using 'compound Bayes' for more general parallel estimation and testing situations, but Efron and Morris (1973) hijacked 'empirical Bayes' for James and Stein-type estimators." "One might refer to Robbins's formulation as nonparametric empirical Bayes; whereas, the formulation discussed here can be referred to as parametric empirical Bayes. (Casella, 1985)" Parametric empirical Bayes sometimes goes by the name of 'objective Bayes' (Berger, 2006) or 'hierarchical Bayes' (Gelman et al., 2013) in the statistics literature, too. There are a couple of recent financial-econometrics papers that use empirical Bayes methods to correct for the effects of data mining (see Liu et al., 2018; Chen and Zimmermann, 2020).

researchers' priors are likely determined by their first-hand knowledge of the anomaly-discovery process. We may not know much about the underlying model underpinning both medium-term momentum (Jegadeesh and Titman, 1993) and investment growth (Titman et al., 2004), but we know a lot about the empirical toolkit Sheridan Titman used to discover both variables.

2.2 Econometric Estimator

We are using σ^2, the variance of $\beta_i^\star$, to parameterize the efficacy of current research practices. We now show how to apply univariate Ridge regressions to estimate this parameter based on the subsequent performance of every other variable in the anomaly zoo. Since each of these variables is just one of many parallel situations, we can then average these variable-specific estimates for σ^2 to compute the prevailing anomaly base rate. This averaging reflects the implicit process by which a researcher uses his past experience with other variables to inform his prior for use next time around.

Ridge Regression. We study a penalized-regression procedure known as the Ridge regression (Hoerl and Kennard, 1970). A Ridge regression combines a standard OLS regression with a quadratic penalty that shrinks OLS-regression estimates towards zero. Using the Ridge to compute the slope coefficient, $\hat{\beta}_i[\lambda]$, for the ith predictor means solving the following optimization problem:

$$\hat{\beta}_i[\lambda] \stackrel{\text{def}}{=} \arg\min_\beta \left\{ \tfrac{1}{N} \cdot \sum_n \left(R_n - \hat{\mu} - \beta \cdot X_{n,i} \right)^2 + \lambda \cdot \beta^2 \right\}. \tag{10}$$

$\lambda \cdot \beta^2$ represents the quadratic penalty, and $\lambda \geq 0$ is known as the 'tuning parameter'. While there are a total of $(N+1)$ stocks, the sum-of-squares term is divided by N in Equation (10) because it already includes the estimated mean, $\hat{\mu}$, leading to a degrees-of-freedom correction.

It is possible to analytically solve for $\hat{\beta}[\lambda]$ in a univariate setting:

$$\hat{\beta}_i[\lambda] = \left(\tfrac{1}{1+\lambda} \right) \cdot \hat{\beta}_i. \tag{11}$$

When $\lambda = 0$, the penalty function disappears, $0 \cdot \beta^2 = 0$, and the results of the Ridge coincide with those of an OLS regression, $\hat{\beta}_i = \hat{\beta}_i[0]$. But, for all $\lambda > 0$, the quadratic penalty in Equation (10) shrinks OLS-regression coefficients toward zero, with larger λs resulting in more shrinkage.[6]

Bayesian Interpretation. There is a Bayesian interpretation for this shrinkage. You can view it as the effect of incorporating prior beliefs about the volatility of the true slope coefficients,

---

[6]The first-order condition (FOC) of the optimization problem in Equation (10) is

$$\tfrac{1}{N} \cdot \sum_n \left(R_n - \hat{\mu} \right) \cdot X_{n,i} = \beta \cdot \left(\tfrac{1}{N} \cdot \sum_n X_{n,i}^2 + \lambda \right)$$

We have normalized each predictor in the cross-section to have zero mean and unit variance, $\tfrac{1}{N+1} \cdot \sum_n X_{n,i} = 0$ and $\tfrac{1}{N} \cdot \sum_n X_{n,i}^2 = 1$. Thus, the left-hand side is just the standard OLS slope coefficient, $\hat{\beta}_i = \hat{\beta}_i[0]$; and, the right-hand side simplifies to $\beta \cdot (1 + \lambda)$. Isolating β on the left-hand side then results in formula given by Equation (11).

$\beta_i^\star$, when these coefficients are drawn from a common normal distribution as in Equation (6). This is because the negative posterior log likelihood of $\beta_i^\star = \beta$ given the realized data, $\boldsymbol{R} \stackrel{\text{def}}{=} \{R_1, \ldots, R_{N+1}\}$ and $\boldsymbol{X}_i \stackrel{\text{def}}{=} \{X_{1,i}, \ldots, X_{N+1,i}\}$, can be written as follows

$$
\begin{aligned}
-\log \Pr[\beta \mid \boldsymbol{R}, \boldsymbol{X}_i] &= \tfrac{1}{2 \cdot (N \cdot \widehat{se}_i^2)} \cdot \textstyle\sum_n \left(R_n - \hat{\mu} - \beta \cdot X_{n,i} \right)^2 + \tfrac{1}{2 \cdot \sigma^2} \cdot (\beta - 0)^2 + \cdots \\
&= \tfrac{1}{2 \cdot \widehat{se}_i^2} \cdot \left\{ \tfrac{1}{N} \cdot \textstyle\sum_n \left(R_n - \hat{\mu} - \beta \cdot X_{n,i} \right)^2 + \tfrac{\widehat{se}_i^2}{\sigma^2} \cdot \beta^2 \right\} + \cdots
\end{aligned}
\tag{12}
$$

where the "$\cdots$" terms represent constants that do not depend on the choice of β.

Thus, when using the appropriate tuning parameter,

$$
\lambda_i = \widehat{se}_i^2 / \sigma^2,
\tag{13}
$$

estimating a univariate Ridge regression in Equation (10) is equivalent to finding the most likely estimate for $\beta_i^\star$ given both your prior beliefs and the observed data which, that is, to minimizing the negative posterior log likelihood in Equation (12). Since this fully Bayesian estimate will have the lowest out-of-sample prediction error on average, we can use this connection to turn your past experience with $(I - 1)$ other variables into $(I - 1)$ different variable-specific estimates for σ^2 by inverting the best-fit λ. In our empirical implementation, we estimate λ period-by-period and make out-of-sample forecasts using a constant sample window estimate to study the performance of newly-discovered characteristics.[7]

Standard Error. Notice that λ_i might be large for either of two reasons in Equation (13). First, the anomaly base rate might be really low. If it is very unlikely that a researcher discovers a new cross-sectional predictor, then σ^2 must be tiny and you should heavily discount any empirical evidence that suggests otherwise for the ith variable. Second, you might not be very confident in your estimated slope coefficient for the ith variable. If the variance of your estimated slope coefficient is quite large after seeing data on the cross-section of $(N + 1)$ excess returns, $\widehat{se}_i^2 \gg 0$, then you should also place less weight on any estimated $|\hat{\beta}_i| \gg 0$.

So, in order to learn about σ^2 by inverting the best-fit tuning parameter for the ith variable, we have to be able to estimate $\widehat{se}_i^2$ separately for each variable. We can do this using OLS:

$$
N \cdot \widehat{se}_i^2 = \tfrac{1}{N-1} \cdot \textstyle\sum_n \left(R_n - \hat{\mu} - \hat{\beta}_i \cdot X_{n,i} \right)^2.
\tag{14}
$$

We use this OLS-based approach to estimating $\widehat{se}_i^2$ when searching for the best-fit tuning parameter

---

[7]We use Ridge regressions because we are assuming that $\beta_i^\star$ is normally distributed. A different choice of prior distributions would suggest using a different kind of penalized regression. e.g., if $\beta_i^\star$ were drawn from a Laplace distribution, then the most efficient way to estimate σ^2 would be to use the LASSO (Park and Casella, 2008). Such an approach would result in a slightly different functional relationship between the best-fit λ_i and σ^2. However, the same core economic intuition would still apply. We discuss this extension in more detail in Appendix B.

associated with each variable $i = 1, \dots, I$.

In-Sample Overfitting. What exactly does it mean to choose the 'best-fit tuning parameter' though? One approach you could take would be to choose the tuning parameter $\lambda_i > 0$ that best fits the data you have observed for each variable in-sample. Let $\text{Err}_i[\lambda]$ denote a Ridge regression's expected in-sample prediction error when using a particular value of λ given the realized cross-section of excess returns in a particular month and lagged values of the ith variable:

$$\text{Err}[\lambda | \boldsymbol{R}, \boldsymbol{X}_i] \overset{\text{def}}{=} \text{E}\left[\left(R_n - \hat{\mu} - \hat{\beta}_i[\lambda] \cdot X_{n,i} \right)^2 \right]. \tag{15}$$

This is called the 'training error' (Hastie et al., 2001). We will typically write $\text{Err}[\lambda | \boldsymbol{R}, \boldsymbol{X}_i] = \text{Err}_i[\lambda]$.

Unfortunately, this simple approach is too simple. No matter what the true value of σ^2 is, the training error will always be minimized by setting $\lambda_i = 0$. In other words, an OLS regression will always outperform a Ridge regression in sample according to this metric. If $\tilde{v}_i^2$ denotes the parameter estimate with the minimum in-sample prediction error for the ith variable,

$$\tilde{v}_i^2 \overset{\text{def}}{=} \arg\min_{v^2 > 0} \left\{ \text{Err}_i[\widehat{se}_i^2 / v^2] \right\}, \tag{16}$$

then we have that $\text{E}[\tilde{v}_i^2] = \infty$ regardless of which $\sigma^2 > 0$ was used to generate the data.

Minimizing the training error requires fine-tuning the slope coefficient to explain variation in excess returns coming from in-sample noise. And, this is easiest when there is no penalty for doing so—i.e., when $\lambda_i = 0$; or equivalently, when $\tilde{v}_i^2 = \infty$. Put another way, the ratio $\lambda_i = \widehat{se}_i^2 / \sigma^2$ governs the relative likelihood that a statistically significant estimate for the slope coefficient, $|\hat{\beta}_i| \gg 0$, is due to in-sample overfitting rather than the existence of an honest-to-goodness tradable anomaly, $\beta_i^\star \neq 0$. But, the estimator $\tilde{v}_i^2$ does not reflect this comparison. It only reveals whether the in-sample fit is good; it does not tell you anything about the origins of this good performance.

Econometric Estimator. If we want an unbiased estimator for σ^2, then we need to adjust the metric used above so that it does not reward in-sample overfitting. Large values of σ^2 will result in lots of in-sample overfitting, so these choices should receive a large penalty; whereas, small values of σ^2 will result in little in-sample overfitting, so these choices should receive a small penalty.

Proposition 2.2 (Econometric Estimator). *Let* $\text{E}[\cdot]$ *denote an expectations operator with respect to realizations of* $\beta_i^\star$, *and let* $\text{Err}_i[\lambda]$ *denote the training error as specified in Equation* (15). *If* $\hat{v}_i^2$ *denotes the parameter estimate with the minimum penalized training error for the ith variable,*

$$\hat{v}_i^2 \overset{\text{def}}{=} \arg\min_{v^2 > 0} \left\{ \text{Err}_i[\widehat{se}_i^2 / v^2] + 2 \cdot \left(\frac{1}{1 + \widehat{se}_i^2 / v^2} \right) \cdot \widehat{se}_i^2 \right\}, \tag{17}$$

then for any $\sigma^2 > 0$ *we have that* $\text{E}[\hat{v}_i^2] = \sigma^2$.

Each variable implicated in the anomaly-zoo literature represents one of many parallel situations. So, after estimating the $\hat{v}_{i'}^2$ associated with each variable $i' \neq i$, we can average these $(I-1)$ variable-specific estimates to form a prior for use when evaluating the ith variable.

Proposition 2.2 follows the same basic logic as many other information-theoretic selection criteria, such as the Akaike information criterion (AIC; Akaike, 1974). Such procedures minimize a mean squared-error loss function plus an additional penalty proportional to the number of degrees of freedom, df, times the noise variance, $\widehat{\text{Var}}[\varepsilon_{n,i}]$. In fact, an alternative way to derive the penalty function in Equation (17) is to note that the effective degrees of freedom in a univariate Ridge regression is given by $1/(1 + \lambda)$. Thus, since $\widehat{\text{Var}}[\varepsilon_{n,i}] = N \cdot \widehat{se}_i^2$, the Akaike penalty function, $2 \cdot (df/N) \times \widehat{\text{Var}}[\varepsilon_{n,i}]$, reduces to the one in Equation (17) when $\lambda_i = \widehat{se}_i^2/\sigma^2$. Similarly, Stone (1977) shows that using Proposition 2.2 to estimate σ^2 will deliver estimates that are asymptotically equivalent to those of cross-validation. And, we illustrate this point in the simulation analysis below.

Forecasting vs. Learning. We have just outlined a univariate approach to estimating σ^2. In essence, we are asking: 'How can you separately use the information in each of the other $(I-1)$ variables to learn about σ^2?' We are not looking for the best combination of variables. But, that is not to say looking for the best combination is wrong. It just depends on what you are after. Taming the factor zoo (Feng et al., 2017) will mean different things to different people.

If all you care about is making good forecasts, taming the factor zoo will mean looking for the best combination of variables. It will mean using something like principle-component analysis to collapse the information in all I variables down into a single forecasting variable. But, if you want to use these same variables to learn how to make actionable predictions, you do not want to combine variables. This would throw away information and make it harder for you to learn.

Consider an example. Driving a car is a complicated activity; there are lots of different factors involved. It has taken quite a while for engineers to teach a computer how to do it.[8] But, it is nevertheless easy to forecast who will be a good driver. You can collapse all the information in the zoo of driving-related factors into a single variable: `IsTeenager` $\in \{\textit{True}, \textit{False}\}$ (Jonah, 1986). If you are an insurance underwriter trying to forecast future claims, this single variable effectively tames the factor zoo. But, if you are a Google engineer, it does not. You cannot create a self-driving car by *not installing* a 17-year-old operating system.

2.3   Simulation Analysis

A researcher's past experience with other variables in the anomaly zoo contains information about the efficacy of financial economists' research practices. So, it is possible to turn these many parallel signals about methodological efficacy into an base rate for use when evaluating the next candidate predictor. We apply empirical-Bayes methods to formalize this reasoning. We then encode the

---

[8]Wired. 12/13/2018. *The Wired Guide to Self-Driving Cars.*

efficacy of our shared research practices as a variance, σ^2, and use the Bayesian interpretation of the Ridge regression to extract information about this variance from other variables.

We are not developing empirical-Bayes methods from scratch or pointing out the Bayesian interpretation of univariate Ridge regressions for the first time. Instead, our core insight has to do with using these ideas to choose the right priors. We have simplified our statistical approach as much as possible to highlight the economic intuition guiding our analysis. Modeling assumptions, like having the true slope coefficients being drawn iid normal, are there for ease of exposition. They are not central to our analysis. And, we now run a variety of numerical simulations to demonstrate the robustness of our simple statistical approach.

Simulation Setup. We simulate data from the following generalization of Equation (6):

$$\beta_i^\star \overset{\text{iid}}{\sim} \text{Dist}[\mu, \sigma^2]. \tag{18}$$

In the analysis below, we vary both the distribution function—i.e., the functional form of Dist[·]—as well as the mean and variance of the $\beta_i^\star$ distribution—i.e., the parameter values μ and σ^2. In addition, we also verify that our results using a regularized maximum-likelihood approach as outlined in Proposition 2.2 closely match the results when using a cross-validation approach.

Unless otherwise stated, all simulations will use the data-generating process in Equation (7) for a cross-section of $(N+1) = 1{,}000$ excess returns where $\mu^\star = 0\%$ per month and $se_i = 1\%$ per month. Everything will be the same as in our main statistical framework except for the fact that the true slope coefficients will be drawn as in Equation (18). When studying the effects of changing either μ or Dist[·], we will compare the average estimated $\hat{v}_i^2$ in 10,000 simulations to the true σ^2 used to generate the data. When studying the effects of changing σ^2, we will compute the average fit of the econometric estimator outlined in Proposition 2.2 across the entire range of choices for v^2:

$$\text{Fit}_i[v^2] \overset{\text{def}}{=} \text{Err}_i[\widehat{se}_i^2/v^2] + 2 \cdot \left(\tfrac{1}{1+\widehat{se}_i^2/v^2}\right) \cdot \widehat{se}_i^2. \tag{19}$$

Note that in each simulation we will be separately estimating $\widehat{se}_i^2$ every time we estimate $\hat{v}_i^2$, just like we will have to when studying real-world data.

Benchmark Simulation. We begin with a benchmark simulation to confirm that our statistical approach recovers the correct value of σ^2 when we simulate data drawn from our hypothesized anomaly-discovery process in Equation (6), Dist[·] = Normal[·] and $\mu = 0\%$ per month. We study three different regimes, $\sigma^2 \in \{0.25\%, 1.00\%, 4.00\%\}$. $\sigma^2 = 4.00\%$ denotes a regime where tradable anomalies are likely; $\sigma^2 = 0.25\%$ denotes a regime where tradable anomalies are unlikely; and, $\sigma^2 = 1.00\%$ denotes a regime somewhere in between. Each panel in Figure 3 reports the average fit at a range of input values $v^2 \in (0, 16]$ when the data are generated using a

17
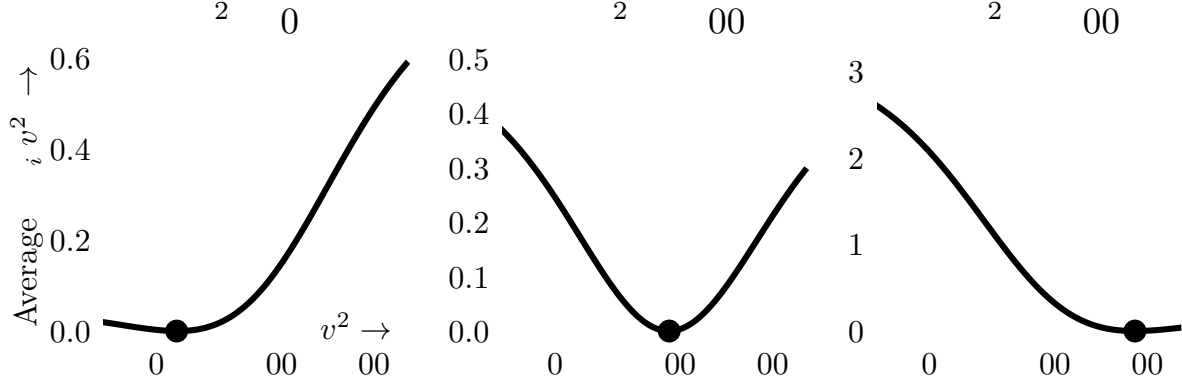
Figure 3. Benchmark Simulation. Each panel reports the results of 10,000 simulations of a market with $(N+1) = 1{,}000$ stocks using a different prior variance, $\sigma^2 \in \{0.25\%, 1.00\%, 4.00\%\}$. Each simulation starts by sampling $\beta_i^\star \overset{\text{iid}}{\sim} \text{Normal}[0, \sigma^2]$. Then, using the realized value of $\beta_i^\star$, we generate a cross-section of excess returns as in Equation (7) using the parameters $\mu^\star = 0$ and $\varepsilon_{n,i}^\star \overset{\text{iid}}{\sim} \text{Normal}[0, N \cdot 1\%^2]$. Finally, we compute the penalized fit characterized by Proposition 2.2 at values $v^2 \in (0, 16]$. x-axis: $v^2 \in (0, 16]$ on a log scale. y-axis: average $\text{Fit}_i[v^2]$ at each value of v^2 across all 10,000 simulations. Large dot: v^2 value that minimize each curve, which corresponds to estimate defined by Proposition 2.2 and should match σ^2 for each panel.

particular prior variance, σ^2. The figure confirms that, when we simulate data using a particular value of σ^2, the objective function in Proposition 2.2 is minimized at that value.

Non-Zero Mean. Next, we explore the sensitivity of our statistical approach to the assumption that $\mu = 0\%$ per month. We do this by running simulations where $\text{Dist}[\cdot] = \text{Normal}[\cdot]$, $\sigma^2 = 1\%^2$, and $\mu \in [-0.3\%, 0.3\%]$. The top panel of Figure 4 shows the results. Each black dot reports the average value of $\hat{v}_i^2$ produced by our statistical approach in 10,000 simulations using a particular choice for μ. We look at values of μ ranging from -0.3% per month to 0.3% per month. The true prior variance in all these simulations is $\sigma^2 = 1\%$ per month, which is depicted by the dashed red line in Figure 4. And, for any choice of $\mu \in [-0.3\%, 0.3\%]$, our results are quite close to this theoretical target.

The inset figure shows that, if we extend the range of μ to $\pm 1\%$ per month, then the econometric estimator in Proposition 2.2 will start to over-estimate σ^2. The black dots are the same as in the main figure. The white dots represent analogous calculations made using $\mu \in [-1.0\%, -0.3\%) \cup (0.3\%, 1.0\%]$. The tendency of our econometric estimator to over-estimate σ^2 when $\mu \neq 0$ makes intuitive sense. The econometric estimator outlined in Proposition 2.2 assumes that $\mu = 0$. So, it reads any deviation of $\beta_i^\star$ from zero as evidence of high prior variance, even if the deviations are due to a non-zero mean.

Using an econometric estimator that is based on a misguided assumption about the true data-generating process is not innocuous; however, the top panel in Figure 4 suggests that the anomaly-discovery process would have to be implausibly efficient to materially affect our estimation

18

results. If $\mu = 0.30\%$ per month, then the long-short portfolio associated with a randomly selected variable $i = 1, \ldots, I$ would have excess returns of 30 bps per month. This is too large an amount of out-of-sample cross-sectional return predictability for the average variable (Harvey et al., 2016; McLean and Pontiff, 2016; Linnainmaa and Roberts, 2018), especially since the sign of $\beta_i^\star$ can be both positive and negative.

Non-Gaussian Prior. Section 2.2 shows how it is possible to use a univariate Ridge regression to learn about σ^2 when the true slope coefficients are drawn from a normal distribution. We just saw that allowing this normal distribution to have a non-zero mean is unlikely to affect our statistical approach in real-world settings. But, what if the true slope coefficients are not even being drawn from a normal distribution to begin with? If the vast anomaly zoo only contains a small handful of tradable anomalies, then you might expect the $\beta_i^\star$ distribution to have fat tails such that extreme observations are more likely. We also saw in Figure 2 that publication bias results in a $\beta_i^\star$ that is not exactly normal, but we demonstrate in the simulation that the normal approximation still yields good estimates in this case.

We address these concerns in the bottom panel of Figure 4 by showing that our statistical approach is still able to recover the correct value of σ^2 when this parameter is the variance of a non-Gaussian prior distribution. In other words, this simulation exercise varies the $\beta_i^\star$ distribution used in Equation (18). Each panel reports results for a different prior distribution, $\mathrm{Dist}[\cdot] \in \{\mathrm{Normal}, \mathrm{Student's}\ t, \mathrm{Post}\text{-}\mathrm{Publication}\}$.

The left chart shows results when data are generated using $\mathrm{Dist}[\cdot] = \mathrm{Normal}[\cdot]$ just like in our primary analysis. Each black dot represents an average taken over 10,000 simulations using $(N + 1) = 1{,}000$, $\mu = 0\%$ per month, and $\sigma^2 \in [0.25, 4.00]$. The x-axis represents the prior variance used to simulate the data, σ^2, while the y-axis reports the average estimated $\hat{v}_i^2$. Thus, if our statistical approach is successful, all the black dots should be along the 45° dashed red line. And, this is exactly what happens, providing further confirmation of the earlier benchmark-simulation results in Figure 3 at a wider range of input values.

The middle chart is different. Rather than drawing $\beta_i^\star$ from a normal distribution, this chart shows analogous results when $\beta_i^\star$ is sampled from a Student's-t distribution:

$$\beta_i^\star = \sigma \cdot \xi_i$$
$$\xi_i \overset{\mathrm{iid}}{\sim} \mathrm{StudentT}[3]. \tag{20}$$

Using a Student's-t distribution for $\beta_i^\star$ is one way to capture the idea that the distribution of true slope coefficients is fat tailed. A Student's-t distribution with 3 degrees of freedom has unit variance. So, to make sure that $\beta_i^\star$ has the appropriate variance when simulating the data, we first sample $\xi_i \overset{\mathrm{iid}}{\sim} \mathrm{StudentT}[3]$ and then multiply the result by the desired prior volatility, $\beta_i^\star = \sigma \cdot \xi_i$. Just like
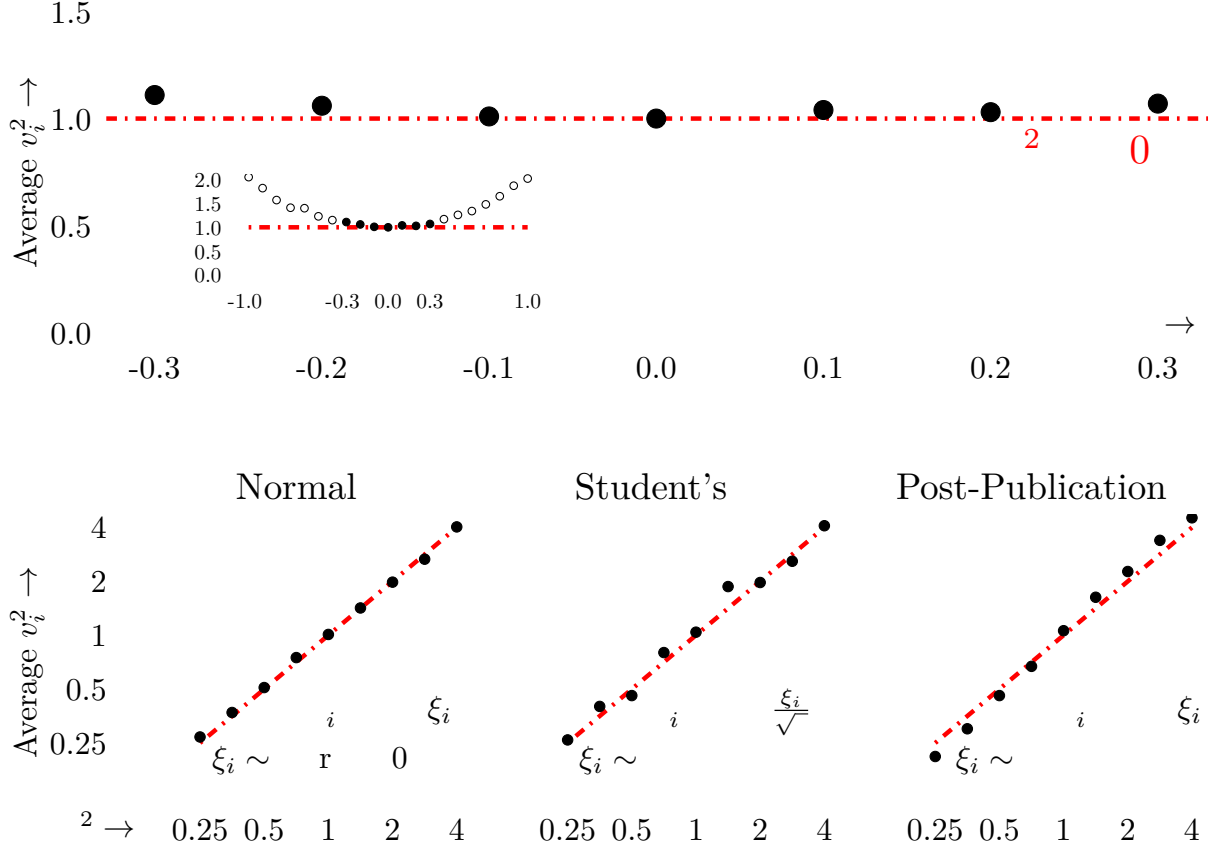
Figure 4. Estimator Robustness. Top Panel: Simulation showing robustness to zero-mean assumption. Each dot represents the average estimated $\hat{v}_i^2$ across 10,000 simulations of a market with $(N + 1) = 1{,}000$ stocks simulated using a given value of $\mu \in [-1\%, 1\%]$. Each simulation starts by sampling $\beta_i^\star \overset{\text{iid}}{\sim} \text{Normal}[\mu, 1\%^2]$. Then, we use the realized value of $\beta_i^\star$ to generate a cross-section of excess returns as in Equation (7) using the parameters $\mu^\star = 0$ and $\varepsilon_{n,i}^\star \overset{\text{iid}}{\sim} \text{Normal}[0, N \cdot 1\%^2]$. y-axis: estimated anomaly base rate (see Proposition 2.2) under the naïve assumption that $\mu = 0$. Dashed red line is true prior variance used to generate the data, $\sigma^2 = 1\%^2$. x-axis: average strength of true slope coefficients $\mu \in [-0.3\%, 0.3\%]$ per month in Equation (18). Inset: identical figure for $\mu \in [-1\%, 1\%]$. **Bottom Panel:** Simulation showing robustness to Gaussian-prior assumption. Each dot represents the average estimated $\hat{v}_i^2$ across 10,000 simulations of a market with $(N + 1) = 1{,}000$ stocks simulated using a given prior distribution, $\text{Dist}[\cdot] \in \{\text{Normal}, \text{Student's } t, \text{Post-Publication}\}$. Each simulation starts by sampling $\beta_i^\star$ from the appropriate distribution. Then, we use the realized value of $\beta_i^\star$ to generate a cross-section of excess returns as in Equation (7) using the parameters $\mu^\star = 0$ and $\varepsilon_{n,i}^\star \overset{\text{iid}}{\sim} \text{Normal}[0, N \cdot 1\%^2]$. x-axis: prior variance used to simulate the data, σ^2, on a log scale. Left: true slope coefficients drawn from a normal distribution, $\xi_i \overset{\text{iid}}{\sim} \text{Normal}[0, 1]$ and $\beta_i^\star = \sigma \cdot \xi_i$. Middle: true slope coefficients drawn from a Student's-t distribution, $\xi_i \overset{\text{iid}}{\sim} \text{StudentT}[3]$ and $\beta_i^\star = \sigma \cdot \xi_i$. Right: true slope coefficients drawn from post-publication distribution, $\xi_i \overset{\text{iid}}{\sim} \text{PostPub}[5\%]$ and $\beta_i^\star = \sigma \cdot \xi_i$, which requires an in-sample p-value less than 5%. If our Gaussian-prior assumption is robust, the black dots should sit on 45° dashed red line.

with the left chart, the middle chart depicts black dots that neatly line up along the 45° dashed red line, which confirms that our statistical approach is able to recover the correct prior variance even when $\beta_i^\star$ is drawn from a non-Gaussian distribution.

The right panel contains analogous results when sampling $\beta_i^\star$ from the post-publication distribution in Figure 2. To create each dot, we iterate on a 3-step process: First, we sample $\beta_i^\star \sim \text{Normal}[0, \sigma^2]$. Second, we simulate a cross-section of excess returns using this particular $\beta_i^\star$. This is the pre-publication data associated with that $\beta_i^\star$. Third, we check whether the associated in-sample $\hat{\beta}_i$ has a p-value $< 5\%$. If the answer is yes, then it is publishable. If the answer is no, then the predictor never sees the light of day. Whenever we find a publishable $\beta_i^\star$, we simulate a new cross-section of excess returns using this $\beta_i^\star$ value. This is the post-publication data associated with the $\beta_i^\star$. We stop once we have post-publication data for 10,000 realizations of $\beta_i^\star$ at each prior variance. Each of the dots in the right panel represents the average estimated $\hat{v}_i^2$ using the post-publication data for these 10,000 simulations:

$$
\begin{aligned}
\beta_i^\star &= \sigma \cdot \xi_i \\
\xi_i &\overset{\text{iid}}{\sim} \text{PostPub}[5\%].
\end{aligned}
\tag{21}
$$

Just as before, the black dots in the right chart line up neatly along the 45° red dashed line.

If you strongly believe that $\beta_i^\star$ is drawn from a non-Gaussian distribution, it is possible to improve the efficiency of the econometric estimator in Proposition 2.2 by adjusting the penalty function (see Appendix B). e.g., Park and Casella (2008) show that using the LASSO is tantamount to adopting a Laplace prior for $\beta_i^\star$. But, the basic insight underpinning both empirical and objective Bayesian thinking is the same.

Moreover, this connection suggests an alternative economic interpretation of the results in this paper. It is now common for financial economists to employ various kinds of penalized-regression procedures. These tools are also widely used by execution desks. In both cases, the OLS-regression procedure is being modified in order to control for false positives, data mining, p-hacking, etc... The standard view is that penalized-regression procedures are a statistical tool for dealing with the proliferation of variables. And, both empirical researchers and execution desks choose the penalty function that best deals with the anomaly zoo in the setting they are currently facing. But, "any sensible estimator is Bayesian for some prior (Diaconis and Skyrms, 2017)." So, if an empirical researcher or execution desk comes to the conclusion that it is best to use a Ridge regression or the LASSO when pruning away weak predictors, this says something important about the economics of anomaly discovery, either in academia or at the research desk.

Cross-Validation. As we point out above, Stone (1977) shows that using a regularized maximum-likelihood estimator, such as the one in Proposition 2.2, is asymptotically equivalent to
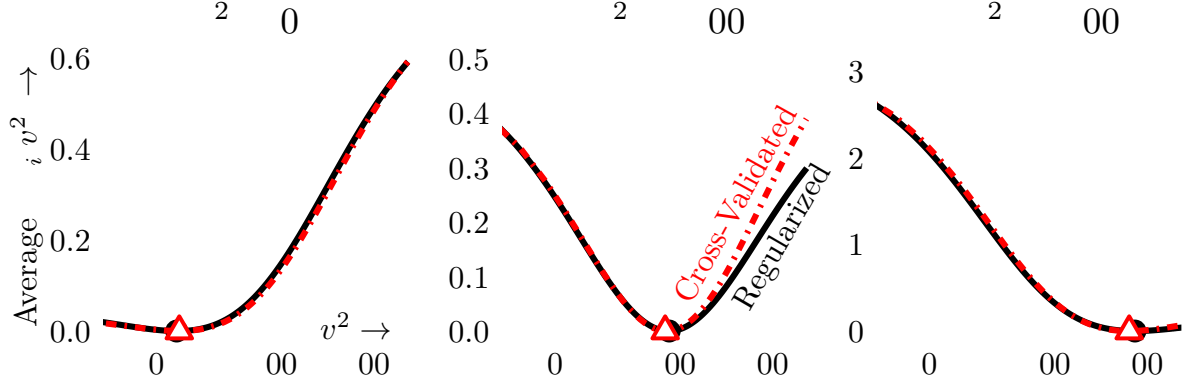
Figure 5. Cross-Validation. Results for 10,000 simulations of a market with $(N+1) = 1,000$ stocks. For each simulation, we generate a new realization of the cross-section of excess returns using the parameters $\mu^\star = 0$, $\beta_i^\star \overset{\text{iid}}{\sim} \text{Normal}[0, \sigma^2]$, and $\varepsilon_{n,i}^\star \overset{\text{iid}}{\sim} \text{Normal}[0, N \cdot 1\%^2]$. Solid black line: average regularized $\text{Fit}_i[v^2]$ at each value of $v^2 \in (0, 16]$. Dashed red line: average cross-validated $\text{Fit}_i[v^2]$ at each value of $v^2 \in (0, 16]$. x-axis: v^2 on a log scale. Each panel reports results for data simulated using $\sigma^2 \in \{0.25\%, 1.00\%, 4.00\%\}$. Large black dot: v^2 value that minimize each regularized curve, which corresponds to estimate defined by Proposition 2.2. Large red/white triangle: v^2 value that minimize each cross-validated curve. The cross-validated analogue to $\text{Fit}_i[v^2]$ is the test-sample mean squared error (MSE) of $\hat{\beta}_i[\lambda]$ where λ is chosen to have the lowest training-sample MSE. The training/test samples are chosen via a 10-fold scheme.

cross-validation in our setting. And, Figure 5 shows that this is indeed the case. The solid black line in each panel corresponds to the results from Figure 3, which reports the average $\text{Fit}_i[v^2]$ at each value of $v^2 \in (0, 16]$ computed across 10,000 simulations. The dashed red line then replicates this analysis using a cross-validated analogue to $\text{Fit}_i[v^2]$. The cross-validated analogue to $\text{Fit}_i[v^2]$ is the test-sample mean squared error (MSE) of $\hat{\beta}_i[\lambda]$ where λ is chosen to have the lowest training-sample MSE. The training/test samples are chosen via a 10-fold scheme. The red triangle corresponds to the value of v^2 that minimizes the cross-validated fit. The fact that this red triangle sits right on top of the value produced by the regularized econometric estimator in Proposition 2.2 shows that these two kinds of estimators produce identical results.

3 Estimation Results

Having described our statistical approach to estimating the anomaly base rate, we next apply it to recover σ^2 each month. We use data on a collection of 85 different variables that were published in the academic literature some time after May 1973.

3.1 Data Description

We begin by describing the data and variables we use in our analysis.

Data Sources. We study the cross-section of monthly returns from May 1973 to June 2015

for each U.S. stock traded on either the NYSE, Amex, or NASDAQ. These data come from the Center for Research in Security Prices (CRSP) monthly stock file. To make sure that our results are not being driven by small illiquid stocks, we exclude any stock with a price below \$1 at the end of the previous month. However, our empirical results are robust to applying other filters to our data set as well (see Appendix D, Table D1). We use balance-sheet data from the Standard and Poor's Compustat database. All items are taken from the fiscal year ending in calendar year $(y - 1)$ for estimation starting in June of year y until May of year $(y + 1)$ predicting returns from July of year y until June of year $(y + 1)$. To alleviate a potential survivorship bias due to back-filling, we require that a firm has at least two years of Compustat data for it to be included in our sample. Let $(N_t + 1)$ denote the number of stocks in our sample in month t.

Return Predictors. We use a collection of 85 different variables that were first documented in the academic literature sometime on or after May 1973. We list each variable along with its publication date in Tables 1a, 1b, and 1c. Let $\mathcal{I}_t$ denote the set of variables discovered prior to month t:

$$\mathcal{I}_t \stackrel{\text{def}}{=} \left\{ i \in \mathcal{I} : \text{publication date for } i\text{th variable} < t \right\}. \tag{22}$$

And, let $I_t \stackrel{\text{def}}{=} |\mathcal{I}_t|$ denote the number of variables discovered prior to month t. So, looking at the first four rows of Table 1a, we have $I_{\text{May73}} = 0$, $I_{\text{Jun73}} = 3$, ..., $I_{\text{Jun77}} = 3$, and $I_{\text{Jul77}} = 4$.

Slope Coefficients. Each month t, we compute the realized returns to a zero-cost strategy based on each variable $i \in \mathcal{I}_t$ by running a separate cross-sectional OLS-regression:

$$R_{n,t} = \hat{\mu}_t + \hat{\beta}_{i,t} \cdot X_{n,i,t-1} + \hat{\varepsilon}_{n,i,t} \qquad \text{for each variable } i = 1, \ldots, I_t. \tag{23}$$

$R_{n,t}$ is the excess return of the nth stock in month t, $\hat{\mu}_t$ is the cross-sectional average excess return for all stocks in our sample during month t, $X_{n,i,t-1}$ is the value of the ith variable for stock n in the previous month normalized to have mean zero and unit variance in the cross-section, $\hat{\beta}_{i,t}$ is the OLS-regression coefficient for the ith variable in month t, and $\hat{\varepsilon}_{n,i,t}$ is the regression residual.

Table 2 provides summary statistics describing these realized returns. There are two things about this table that are worth pointing out. First, contrarian strategies, such as the long-term reversals captured by the variable *'Ret, 36-13'* in row five, will result in estimated values that are negative on average. This is consistent with the modeling assumption that the true $\beta_{i,t}^{\star}$ values are drawn from a mean-zero normal distribution. When we incorporate these sorts of variables in a trading strategy, we will always trade in the appropriate direction. However, the sign of the slope coefficient associated with a cross-sectional long-short strategy is arbitrary.

Second, our estimates for each variable's $\hat{\beta}_{i,t}$ tend to be smaller than the ones reported in the original papers. This is because researchers typically report the excess returns to sorted high-minus-low portfolios. This standard practice is tantamount to running a cross-sectional

	Name	P. Date	Description
1.	*Beta*	1973-05	Rolling CAPM beta
2.	*BetaSq*	1973-05	Rolling CAPM beta, squared
3.	*IdioVolCAPM*	1973-05	Idiosyncratic volatility, CAPM
4.	$Earn/Share$	1977-06	Earnings per share
5.	$Debt/Price$	1979-06	Debt to price
6.	$Divd/Price$	1979-06	Dividend to price
7.	*Mcap*	1981-03	Market capitalization, prev. fiscal year
8.	$Earn/Price$	1982-08	Earnings to price
9.	*Ret*, 36-12	1985-07	Cum. return, months $[-36, -12)$
10.	*AvgSpread*	1986-12	Mean bid-ask spread
11.	$Assets/Mcap$	1988-06	Assets to market cap
12.	*Levrg*	1988-06	Leverage
13.	$Levrg/Price$	1988-06	Leverage to price
14.	$Sales/Cash$	1989-11	Sales to cash
15.	*LtCF*	1989-11	Long-term cash flow
16.	*CurrRatio*	1989-11	Current ratio
17.	$\%\Delta CurrRatio$	1989-11	Perc. change in current ratio
18.	$\%\Delta QuickRatio$	1989-11	Perc. change in quick ratio
19.	$\%\Delta[Sales/Invtry]$	1989-11	Perc. change in sales to inventory
20.	*QuickRatio*	1989-11	Quick ratio
21.	$Sales/Invtry$	1989-11	Sales to inventory
22.	$Sales/Recv$	1989-11	Sales to receivables
23.	*Ret*, 1-0	1990-07	Return, month $[-1, 0)$
24.	*Ret*, 12-1	1990-07	Cum. return, months $[-12, -1)$
25.	*BkVal*	1992-06	Book value
26.	*MonthlyMcap*	1992-06	Market cap, prev. month
27.	$Sales/Price$	1992-06	Sales to price
28.	$\%\Delta[Deprc/PP\&E]$	1992-09	Perc. change in depreciation to PP&E
29.	$D\&A/Assets$	1992-09	D&A to assets

Table 1a. List of Variables. List of variables documented in the academic literature sometime on or after May 1973. Variables are constructed using data from CRSP and Compustat. Name: The name for the variable used throughout this paper. P. Date: The month of publication for the first academic paper about each variable. Description: A description of how variable is constructed for each stock.

	Name	P. Date	Description
30.	$Deprc/PP\&E$	1992-09	Depreciation to PP&E
31.	Ret, 6-1	1993-03	Cum. return, months $[-6, -1)$
32.	$\%\Delta Sales$	1994-12	Perc. change in sales
33.	$OpAccr$	1996-07	Operating accruals
34.	$CapitalTOver$	1996-07	Capital turnover
35.	$RetOnEquity$	1996-07	Return on equity
36.	$KaplanZingales$	1997-02	Kaplan-Zingales index
37.	$\%\Delta[\Delta Sales/\Delta Invtry]$	1997-04	Perc. change in ΔSales to ΔInventory
38.	$\%\Delta[\Delta Sales/\Delta Recv]$	1997-04	Perc. change in ΔSales to ΔReceivables
39.	$\%\Delta[\Delta Sales/\Delta XG\&A]$	1997-04	Perc. change in ΔSales to ΔXG&A
40.	$\%\Delta[\Delta GrMgn/\Delta Sales]$	1998-01	Perc. change in ΔGross margin to ΔSales
41.	$LagTOver$	1998-08	Lagged turnover
42.	$\mathrm{Adj}[BkVal/Mcap]$	2000-02	Ind. adjusted book-to-market ratio
43.	$AdjMcap$	2000-02	Ind. adjusted market cap
44.	$SdTOver$	2001-01	Std. deviation of daily turnover
45.	$AdvertRate/Ret$	2001-12	Advertising expense rate to returns
46.	$R\&D/Mcap$	2001-12	R&D to market cap
47.	$R\&D/Sales$	2001-12	R&D to sales
48.	$Advert/Mcap$	2001-12	Advertising expense to market cap
49.	$\Delta Invtry/Assets$	2002-06	Inventory changes to assets
50.	$OpCF/Price$	2004-04	Operating cash flow to price
51.	$Invmt/\mathrm{Lag}[AvgInvmt]$	2004-12	Investment to trailing 3 years average
52.	$NetOpAssets/Sales$	2004-12	Net operating assets to lagged sales
53.	$\%\Delta BkVal$	2005-09	Perc. change in book value
54.	$\%\Delta LtDebt$	2005-09	Perc. change in long-term debt
55.	$Price\text{-}52WkHi$	2005-11	Closeness to previous 52-week high
56.	$IdioVolFF93$	2006-02	Idiosyncratic volatility, FF93
57.	$TotVol$	2006-02	Total volatility

Table 1b. List of Variables, Ctd. List of variables documented in the academic literature sometime on or after May 1973. Variables are constructed using data from CRSP and Compustat. Name: The name for the variable used throughout this paper. P. Date: The month of publication for the first academic paper about each variable. Description: A description of how variable is constructed for each stock.

	Name	P. Date	Description
58.	$\varepsilon^{\%\Delta Mcap}$	2006-08	Residual perc. change in market cap
59.	$NetExtnlFin/_{Assets}$	2006-10	Net external financing to assets
60.	$DailyBeta$	2006-11	Daily rolling CAPM beta
61.	$NetPO/_{Price}$	2007-04	Net payouts to price
62.	$PO/_{Price}$	2007-04	Payouts to price
63.	$NetPO$	2007-04	Net payout ratio
64.	$RetOnInvstCap$	2007-06	Return on invested capital
65.	$\%\Delta Shares$	2008-04	Perc. change in shares outstanding
66.	$ProfMgn$	2008-05	Profit margin
67.	$AdjProfMgn$	2008-05	Ind. adjusted profit margin
68.	$RetOnOpAssets$	2008-05	Return on net operating assets
69.	$AssetTOver$	2008-05	Asset turnover
70.	$AdjAssets$	2008-05	Ind. adjusted total assets
71.	$\%\Delta InvmtX$	2008-07	Perc. change in investments (Xing)
72.	$\%\Delta Invmt$	2008-08	Perc. change in investments
73.	$\Delta AdjShares$	2008-08	Change in split-adjusted shares outstanding
74.	$RetOnCash$	2009-01	Return on cash
75.	$Tangibility$	2009-04	Asset tangibility
76.	$\Delta AdjTOver$	2009-10	Change in market-adjusted turnover
77.	$UnexplVlm$	2009-10	Standardized unexplained volume
78.	$RetOnAssets$	2010-05	Return on assets
79.	$OpLevrg$	2011-01	Operating leverage
80.	$MaxRet$	2011-02	Max monthly return during prev. year
81.	$FreeCF$	2011-05	Free cash flow
82.	$R\&Dcapital$	2011-09	R&D capital
83.	$\%\Delta Invtry$	2012-01	Perc. change in inventory
84.	Ret, 12-6	2012-03	Cum. return, months $[-12, -6)$
85.	$CashHldgs$	2012-04	Cash holdings

Table 1c. List of Variables, Ctd. List of variables documented in the academic literature sometime on or after May 1973. Variables are constructed using data from CRSP and Compustat. Name: The name for the variable used throughout this paper. P. Date: The month of publication for the first academic paper about each variable. Description: A description of how variable is constructed for each stock.
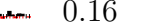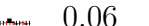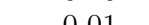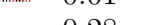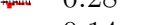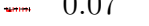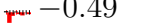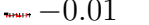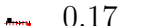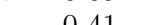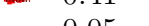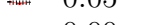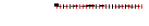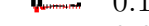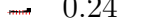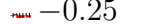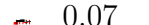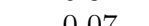
	Avg	Sd		Avg	Sd		Avg	Sd
$Beta$	0.00	2.21	$Deprc/PP\&E$	0.22	1.76	$\varepsilon^{\%\Delta Mcap}$	0.16	1.22
$BetaSq$	0.01	2.21	$Ret,\ 6\text{-}1$	−0.16	2.36	$NetExtnlFin/Assets$	0.00	0.52
$IdioVolCAPM$	0.70	2.88	$\%\Delta Sales$	−0.29	0.84	$DailyBeta$	0.06	1.79
$Earn/Share$	−0.31	2.15	$OpAccr$	0.01	0.97	$NetPO/Price$	−0.15	1.17
$Debt/Price$	0.10	1.48	$CapitalTOver$	−0.17	0.67	$PO/Price$	−0.16	0.99
$Divd/Price$	−0.14	1.74	$RetOnEquity$	−0.44	1.94	$NetPO$	−0.01	0.53
$Mcap$	−0.55	1.69	$KaplanZingales$	−0.03	0.94	$RetOnInvstCap$	−0.28	1.04
$Earn/Price$	−0.21	1.91	$\%\Delta[\Delta Sales/\Delta Invtry]$	0.07	0.49	$\%\Delta Shares$	0.14	0.36
$Ret,\ 36\text{-}12$	−0.56	1.63	$\%\Delta[\Delta Sales/\Delta Recv]$	0.11	0.52	$ProfMgn$	0.07	0.52
$AvgSpread$	0.59	1.68	$\%\Delta[\Delta Sales/\Delta XG\&A]$	−0.10	0.56	$AdjProfMgn$	−0.49	1.34
$Assets/Mcap$	0.25	1.87	$\%\Delta[\Delta GrMgn/\Delta Sales]$	−0.03	0.59	$RetOnOpAssets$	−0.01	1.02
$Levrg$	−0.05	1.42	$LagTOver$	−0.04	2.44	$AssetTOver$	−0.15	0.67
$Levrg/Price$	0.05	1.63	$\mathrm{Adj}[BkVal/Mcap]$	0.39	0.90	$AdjAssets$	0.17	0.97
$Sales/Cash$	0.14	1.97	$AdjMcap$	−0.31	1.43	$\%\Delta InvmtX$	−0.17	0.47
$LtCF$	−0.05	1.37	$SdTOver$	0.04	1.73	$\%\Delta Invmt$	0.09	0.95
$CurrRatio$	−0.20	0.57	$AdvertRate/Ret$	0.31	1.02	$\Delta AdjShares$	−0.41	0.85
$\%\Delta CurrRatio$	−0.20	0.58	$R\&D/Mcap$	0.47	1.69	$RetOnCash$	−0.05	0.87
$\%\Delta QuickRatio$	0.08	0.45	$R\&D/Sales$	0.12	2.00	$Tangibility$	0.00	0.79
$\%\Delta[Sales/Invtry]$	−0.03	1.62	$Advert/Mcap$	0.64	1.20	$\Delta AdjTOver$	0.14	0.39
$QuickRatio$	−0.01	1.45	$\Delta Invtry/Assets$	−0.15	0.52	$UnexplVlm$	0.24	0.55
$Sales/Invtry$	−0.08	0.85	$OpCF/Price$	0.01	1.01	$RetOnAssets$	−0.25	0.96
$Sales/Recv$	0.08	0.72	$Invmt/\mathrm{Lag}[AvgInvmt]$	−0.08	0.43	$OpLevrg$	0.15	0.55
$Ret,\ 1\text{-}0$	−0.45	2.04	$NetOpAssets/Sales$	−0.07	0.89	$MaxRet$	0.07	1.06
$Ret,\ 12\text{-}1$	−0.17	2.40	$\%\Delta BkVal$	−0.26	0.70	$FreeCF$	−0.10	0.89
$BkVal$	0.39	1.45	$\%\Delta LtDebt$	−0.11	0.40	$R\&Dcapital$	0.31	1.30
$MonthlyMcap$	−0.82	1.93	$Price\text{-}52WkHi$	−0.46	2.33	$\%\Delta Invtry$	−0.07	0.42
$Sales/Price$	0.38	1.56	$IdioVolFF93$	0.31	1.66	$Ret,\ 12\text{-}6$	−0.10	0.69
$\%\Delta[Deprc/PP\&E]$	0.00	0.46	$TotVol$	0.28	1.83	$CashHldgs$	0.10	0.99
$D\&A/Assets$	0.32	1.15						

Table 2. Estimated $\hat{\beta}_{i,t}$. Summary statistics describing the realized returns to a zero-cost trading strategy based on each variable $i = 1, \ldots, I_t$. These realized returns correspond to the estimated $\hat{\beta}_{i,t}$ in Equation (23). Sample Period: June 1973 to June 2015. Units: % per month. Sparkline Plots: time series for each variable on a common scale, $-3\% < \hat{\beta}_{i,t} < 3\%$. All plots end in June 2015. Variables discovered later have shorter plots. Red indicated negative returns.

	Avg	Sd		Avg	Sd		Avg	Sd
$Beta$	3.40	4.79	$Deprc/PP\&E$	2.07	3.97	$\varepsilon^{\%\Delta Mcap}$	1.41	2.46
$BetaSq$	3.40	4.79	Ret, 6-1	2.32	4.22	$NetExtnlFin/Assets$	0.23	0.52
$IdioVolCAPM$	4.12	5.30	$\%\Delta Sales$	0.74	1.45	$DailyBeta$	2.15	3.73
$Earn/Share$	2.85	4.32	$OpAccr$	0.87	1.80	$NetPO/Price$	1.32	2.75
$Debt/Price$	1.71	3.34	$CapitalTOver$	0.42	1.10	$PO/Price$	0.93	1.79
$Divd/Price$	2.40	3.80	$RetOnEquity$	2.40	4.19	$NetPO$	0.22	0.50
$Mcap$	2.34	3.87	$KaplanZingales$	0.78	1.82	$RetOnInvstCap$	1.10	2.27
$Earn/Price$	2.35	3.92	$\%\Delta[\Delta Sales/\Delta Invtry]$	0.20	0.60	$\%\Delta Shares$	0.11	0.18
Ret, 36-12	1.98	3.78	$\%\Delta[\Delta Sales/\Delta Recv]$	0.24	0.85	$ProfMgn$	0.22	0.37
$AvgSpread$	2.27	3.83	$\%\Delta[\Delta Sales/\Delta XG\&A]$	0.27	0.45	$AdjProfMgn$	1.92	3.18
$Assets/Mcap$	2.63	4.37	$\%\Delta[\Delta GrMgn/\Delta Sales]$	0.30	0.56	$RetOnOpAssets$	0.88	2.19
$Levrg$	1.54	3.19	$LagTOver$	3.54	4.91	$AssetTOver$	0.41	0.84
$Levrg/Price$	2.00	3.68	$\text{Adj}[BkVal/Mcap]$	0.92	1.74	$AdjAssets$	0.84	2.00
$Sales/Cash$	2.39	4.16	$AdjMcap$	1.27	2.95	$\%\Delta InvmtX$	0.21	0.32
$LtCF$	1.42	2.94	$SdTOver$	2.10	3.45	$\%\Delta Invmt$	0.85	1.51
$CurrRatio$	0.31	0.84	$AdvertRate/Ret$	0.85	2.08	$\Delta AdjShares$	0.82	2.04
$\%\Delta CurrRatio$	0.32	0.89	$R\&D/Mcap$	2.16	3.81	$RetOnCash$	0.69	1.28
$\%\Delta QuickRatio$	0.16	0.37	$R\&D/Sales$	2.53	4.25	$Tangibility$	0.56	1.50
$\%\Delta[Sales/Invtry]$	1.85	3.77	$Advert/Mcap$	1.64	3.30	$\Delta AdjTOver$	0.14	0.35
$QuickRatio$	1.62	3.45	$\Delta Invtry/Assets$	0.25	0.49	$UnexplVlm$	0.32	0.72
$Sales/Invtry$	0.63	1.72	$OpCF/Price$	0.94	2.08	$RetOnAssets$	0.93	1.91
$Sales/Recv$	0.48	1.02	$Invmt/\text{Lag}[AvgInvmt]$	0.15	0.30	$OpLevrg$	0.28	0.45
Ret, 1-0	2.19	4.22	$NetOpAssets/Sales$	0.72	1.75	$MaxRet$	1.07	1.75
Ret, 12-1	2.70	4.42	$\%\Delta BkVal$	0.51	1.67	$FreeCF$	0.74	1.61
$BkVal$	1.84	3.34	$\%\Delta LtDebt$	0.13	0.24	$R\&Dcapital$	1.46	2.72
$MonthlyMcap$	2.70	4.41	$Price\text{-}52WkHi$	2.93	4.55	$\%\Delta Invtry$	0.14	0.23
$Sales/Price$	2.00	3.69	$IdioVolFF93$	1.95	3.53	Ret, 12-6	0.43	0.74
$\%\Delta[Deprc/PP\&E]$	0.17	0.57	$TotVol$	2.35	3.87	$CashHldgs$	0.93	1.53
$D\&A/Assets$	1.29	2.69						

Table 3. Estimated $\hat{v}_{i,t}^2$. Summary statistics describing the variable-specific estimates for the prior variance. We estimate $\hat{v}_{i,t}^2$ separately for each variable $i = 1, \ldots, I_t$ in month t as described in Proposition 2.2. Sample Period: June 1973 to June 2015. Units: $\%^2$ per month. Sparkline Plots: time series for each variable on a common scale, $0\% < \hat{v}_{i,t} < 6\%$. All plots end in June 2015. Variables discovered later have shorter plots.

regression where $\mathrm{Var}[X_{n,i,t-1}] < 1$. For example, going long the top 10% of stocks and short the bottom 10% of stocks corresponds to $\mathrm{Var}[X_{n,i,t-1}] = \frac{1}{10} \cdot (+1)^2 + \frac{1}{10} \cdot (-1)^2 = 1/5$. And, since $\hat{\beta}_{i,t} = \mathrm{Cov}[R_{n,t}, X_{n,i,t-1}]/\mathrm{Var}[X_{n,i,t-1}]$, this approach would result in point estimates that are five times larger than ours. We will adjust for this difference in the variance of the right-hand-side variables when we consider the effects of implementation costs before combining variables into a single strategy in Section 4 below.

3.2 Anomaly Base Rate

Here is how we use this data to estimate the anomaly base rate. Our goal is to mirror the process by which researchers implicitly convert their past experience with other variables in the anomaly zoo into a prior.

Variable-Specific Estimates. We start by creating separate variable-specific estimates for σ^2 using each variable in the anomaly zoo as of month t. For each variable $i = 1, \ldots, I_t$, we solve the optimization problem outlined in Proposition 2.2 in each month t such that $i \in \mathcal{I}_t$. Table 3 provides summary statistics describing these variable-specific estimates for the prior variance, $\hat{v}_{i,t}^2$. The sparkline plots represent the time series for each variable $i = 1, \ldots, 85$ on a common scale, $0\% < \hat{v}_{i,t} < 6\%$. Every time series ends in June 2015, so shorter sparkline plots correspond to variables that were discovered later in our sample period.

Our goal in showing these sparkline plots is not to give evidence that any one predictor is 'significant'. Rather, the aim is to show that predictor strengths were drawn from a common distribution. Variables in our sample are constructed in very different ways using entirely different datasets—e.g., compare $^{Deprec}/_{PP\&E}$ and $Ret, 6\text{-}1$ in the top two rows of the center panel of Tables 2 and 3. Nevertheless, notice that these two variables generate similar point estimates for $\hat{v}_{i,t}^2$ each month. And, this is true even though, when you look at Table 2, the $\hat{\beta}_{i,t}$ time series for each of these variables is very different. Trading on $^{Deprec}/_{PP\&E}$ and $Ret, 6\text{-}1$ clearly yields a very different pattern of realized returns, and yet the estimated $\hat{v}_{i,t}^2$ implied by each variable is quite similar.

We next use the time series of prior-variance estimates associated with each variable to generate out-of-sample one-month-ahead forecasts. After all, you want to convert your past experience into a prior for future use. So, to make a forecast for the ith variable in month t, we study the 60-month immediately prior to month t, $\{t - 60, \ldots, t - 1\}$. We denote the resulting one-month-ahead forecasted value for month t by $\bar{v}_{i,t}^2$ to distinguish it from the in-sample estimates. We begin making forecasts of the prior volatility implied by the ith variable only after we have observed 60 months of post-publication data to alleviate the publication bias. e.g., for the rolling-CAPM-beta (row 1 in Table 1a), the first forecast is for month $t = \mathrm{Jun}1978$. We know that prior to publication the estimated $\hat{\beta}_i$'s are a combination of both the true slope coefficient $\beta_i^\star$ and the effects of publication bias. By only using post-publication observations to forecast $\bar{v}_{i,t}^2$, we avoid this problem.

29

	#Obs	Avg	Sd	1%	25%	50%	75%	99%
	446	2.19	1.43	0.57	1.16	1.82	2.69	7.30

Table 4. Aggregate $\bar{v}_t^2$. Summary statistics describing the sample average of the rolling one-month-ahead forecast of prior variance, $\bar{v}_t^2$. We separately compute in-sample estimates for the $\hat{v}_{i,t}^2$ associated with each variable $i = 1, \ldots, I_t$ in each month as described in Proposition 2.2. Then, we make an out-of-sample one-month-ahead forecast, $\bar{v}_{i,t}^2$, by fitting an AR(3) model to months $\{t - 60, \ldots, t - 1\}$. Finally, we average these forecasts to compute $\bar{v}_t^2$ as in Equation (26). Units: $\%^2$ per month. Sample Period: June 1978 to June 2015. Sparkline Plot: aggregate time series corresponding to Figure 1 in miniature.

To make each forecast, we fit an AR(3) model to these 60 observations prior to month t:

$$\hat{v}_{i,t'}^2 = \check{a}_i + \sum_{\ell=1}^{3} \check{b}_{i,\ell} \cdot \hat{v}_{i,t'-\ell}^2 + \check{e}_{i,t'} \qquad \text{for months } t' = (t - 60), \ldots, (t - 1). \qquad (24)$$

$\check{a}_i$ and $\{\check{b}_{i,\ell}\}_{\ell=1,2,3}$ denote the coefficients associated with the ith variable, and $\check{e}_{i,t'}$ represents the regression residual in month t'. Note that these coefficients will be different for each forecast date t; we have just suppressed the t subscripts for clarity. We then compute one-month-ahead forecasts for the prior variance in month t by applying these estimated coefficients to the final three months of data prior to month t:

$$\bar{v}_{i,t}^2 \overset{\text{def}}{=} \mathrm{E}_{t-1}[\hat{v}_{i,t}^2] = \check{a}_i + \sum_{\ell=1}^{3} \check{b}_{i,\ell} \cdot \hat{v}_{i,t-\ell}^2. \qquad (25)$$

We use an AR(3) forecasting model in our primary analysis for convenience. Our results are robust to using different numbers of lags as we show in Table D2 of Appendix D. The precise choice of forecasting model is not key to our findings.

Combining Results. The anomaly zoo represents a large data set of parallel situations, one for each variable in the anomaly zoo as of month t, $i = 1, \ldots, I_t$. The idea behind empirical-Bayes methods is that you can combine the signal about σ^2 produced by each of these variables to generate a more precise forecast:

$$\bar{v}_t^2 \overset{\text{def}}{=} \frac{1}{I_t} \cdot \sum_{i \in \mathcal{I}_t} \bar{v}_{i,t}^2. \qquad (26)$$

This combined forecast, $\bar{v}_t^2$, represents your best guess about which prior variance of predictor strengths to use when evaluating variables in month t—i.e., your best guess about the true value σ^2 to use in Equation (6)—based on your past experience with I_t other variables. So, as the number of variables in the anomaly zoo grows, we will be getting more and more signals about the anomaly base rate that we should be using going forward. By analogy, a seasoned researcher who has seen many variables come and go will have more signals with which to inform his beliefs than a junior

researcher with less experience.

To give a sense of the prior beliefs suggested by the data, Table 4 provides summary statistics describing this combined forecast each month, $\bar{v}_t^2$, and Figure 1 plots the $\bar{v}_t^2$ time-series. The yearly average forecasted value for the prior variance peaked in the year 2000 at $\frac{1}{12} \cdot \sum_{\text{Jan'00}}^{\text{Dec'00}} \bar{v}_t^2 = 5.89$, and it reached its low point in 2015 at $\frac{1}{12} \cdot \sum_{\text{Jan'15}}^{\text{Dec'15}} \bar{v}_t^2 = 0.60$. While Figure 1 reveals a large spike in $\bar{v}_t^2$ at the time of the DotCom crash, this variable is not just a crash indicator. For example, note that there is no spike in prior variance around the 1987 crash. We estimate that $\bar{v}_{\text{Oct87}}^2 = 1.53$, which is much lower than the average value over the entire sample period, 2.19.

Relaxing the independence assumption will only affect the precision of our estimates for σ^2. To see why, note that Proposition 2.2 says that $\text{E}[\hat{v}_i^2] = \sigma^2$ for any $i \in \mathcal{I}_t$. The most extreme way to violate the assumption of independent draws for $\beta_i^\star$ in Equation (6) would be to assume that the realized values of $\beta_i^\star$ are the same for all $i \in \mathcal{I}_t$. In this extreme case, we would effectively only have one signal about σ^2. But, we could still use this lone draw, $\hat{v}_1^2 = \hat{v}_2^2 = \cdots = \hat{v}_{I_t}^2$, as a signal about σ^2. It would just be a much noisier signal. When we analyze trading-strategy performance in the following section, we will focus our attention on the sample period starting in January 1990 to minimize this concern.

Macroeconomic Correlations. Of course, there are numerous macroeconomic variables that forecast returns. And, in particular, there is a well-known relationship between expected returns and both realized variance and uncertainty. So, it is important to make sure that, when we estimate $\bar{v}_t^2$, we are not just repackaging and rebranding some existing variable. We consider data on six different alternative variables: the level of the VIX index, realized volatility, the value-weighted market return, log growth rate of seasonally adjusted GDP, the term spread, and the NBER recession indicator. Table 5 provides summary statistics for each of these macroeconomic variables. Note that we do not have data on the VIX or the term spread for our entire sample period. In particular, our data on the VIX begins in January 1990.

We regress our forecasted $\bar{v}_t^2$ in month t on lagged values $\bar{v}_{t-1}^2$, $\bar{v}_{t-2}^2$, and $\bar{v}_{t-3}^2$ as well as values of these other macroeconomic variables using the following specification:

$$\bar{v}_t^2 = \hat{a} + \sum_{\ell=1}^{3} \hat{b}_\ell \cdot \bar{v}_{t-\ell}^2 + \hat{\boldsymbol{c}}^\top \boldsymbol{Z}_t + \hat{e}_t. \tag{27}$$

$\boldsymbol{Z}_t$ denotes a vector of macroeconomic variables in month t.

Table 6 reports the results of these regression specifications. Each column reports the results of a separate regression. The table reveals that, while our forecasts for the prior variance are sometimes related to well-known macroeconomic variables, such as the VIX, they are certainly not subsumed by them. e.g., the three lags of $\bar{v}_t^2$ explain 62.9% of the variation in column (2) whereas the VIX explains only 2.4% of the variation in column (3). Moreover, once other macroeconomic

	VIX_t	$RVol_t$	$R_{\mathrm{Mkt},t}$	$dGDP_t$	$tSpd_t$	$Rcsn_t$
	(1)	(2)	(3)	(4)	(5)	(6)
Avg	19.89	14.47	1.02	1.41	1.81	0.13
Sd	7.59	8.70	4.48	0.89	1.15	0.33
Start	Jan'90	Jun'78	Jun'78	Jun'78	Jan'82	Jun'78
#Obs	306	446	446	446	402	446

Table 5. Macroeconomic Variables. Summary statistics for macroeconomic forecasting variables. VIX_t: level of the VIX index; Jan 1990 to Jun 2015. $RVol_t$: realized volatility on value-weighted market index; Jun 1978 to Jun 2015. $R_{\mathrm{Mkt},t}$: return on the CRSP value-weighted market index; Jun 1978 to Jun 2015. $dGDP_t$: log growth rate of seasonally adjusted GDP; Jun 1978 to Jun 2015. $tSpd_t$: the term spread; Jan 1982 to Jun 2015. $Rcsn_t$: NBER indicator for whether a recession is taking place; Jun 1978 to Jun 2015.

variables are included, the VIX is no longer a significant predictor of prior variance. In other words, $\bar{v}_t^2$ is not just a proxy for the VIX.

In the next section, we are going to incorporate the volatility of prior beliefs into portfolio selection.

4   Trading Strategy

Financial economists implicitly turn their past experience with other variables in the anomaly zoo into prior beliefs based on their knowledge of the research process. To make use of this insight, we employ a simple statistical framework that combines empirical-Bayes methods and the Bayesian interpretation of penalized-regression procedures. In this section, we study trading-strategy performance to demonstrate that our estimates of the anomaly base rate still produces actionable estimates in spite of their transparent simplicity.

4.1   Portfolio Construction

We start with a benchmark strategy that explicitly does not account for the anomaly base rate. This benchmark strategy assumes that any amount of predictability is equally likely ex ante, $\sigma^2 = \infty$. We then adjust this benchmark strategy to account for the anomaly base rate, discarding implausibly strong predictors given the forecasted prior variance.

Variable-Specific Returns. The realized return to a zero-cost long-short portfolio associated with the ith variable in month t, $R_{i,t}$, is just the estimate for $\hat{\beta}_{i,t}$ in that month:

$$R_{i,t} \stackrel{\text{def}}{=} \hat{\beta}_{i,t} = \tfrac{1}{N_t} \cdot \sum_n \left(R_{n,t} - \hat{\mu}_t \right) \cdot X_{n,i,t-1}. \tag{28}$$

Dependent Variable: $\bar{v}_t^2$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Const$	0.37*** (0.08)	0.35*** (0.10)	1.38*** (0.26)	1.82*** (0.13)	2.18*** (0.07)	2.00*** (0.13)	2.05*** (0.14)	2.15*** (0.07)	0.79** (0.40)	0.74 (0.47)	0.37 (0.25)	0.33 (0.29)
$\bar{v}_{t-1}^2$	0.65*** (0.05)	0.69*** (0.06)									0.68*** (0.06)	0.68*** (0.06)
$\bar{v}_{t-2}^2$	0.08 (0.06)	0.04 (0.07)									0.04 (0.07)	0.04 (0.07)
$\bar{v}_{t-3}^2$	0.10** (0.05)	0.10* (0.06)									0.11* (0.06)	0.11* (0.06)
VIX_t			0.04*** (0.01)						0.01 (0.03)			0.00 (0.02)
$RVol_t$				0.02*** (0.01)					0.05*** (0.01)	0.05** (0.02)	0.01* (0.01)	0.01 (0.01)
$R_{\mathrm{Mkt},t}$					0.00 (0.02)				0.02 (0.02)	0.02 (0.02)	0.02 (0.01)	0.02 (0.01)
$dGDP_t$						0.13 (0.08)			0.44** (0.19)	0.44** (0.19)	−0.10 (0.12)	−0.10 (0.12)
$tSpd_t$							0.05 (0.06)		−0.01 (0.08)	−0.01 (0.08)	−0.07 (0.04)	−0.07 (0.05)
$Rcsn_t$								0.26 (0.20)	0.13 (0.37)	0.11 (0.37)	−0.21 (0.23)	−0.22 (0.23)
$Adj\ R^2$	61.0%	62.9%	2.4%	2.0%	−0.2%	0.4%	−0.1%	0.2%	4.4%	4.1%	63.4%	63.2%
Start	Jun'78	Jan'90	Jan'90	Jun'78	Jun'78	Jun'78	Jan'82	Jun'78	Jan'90	Jan'90	Jan'90	Jan'90
#Obs	443	306	306	443	443	443	402	443	306	306	306	306

Table 6. Macroeconomic Correlations. Relationship between the forecasted $\bar{v}_t^2$ in month t and macroeconomic variables. VIX_t: level of the VIX index; Jan 1990 to Jun 2015. $RVol_t$: realized volatility on value-weighted market index; Jun 1978 to Jun 2015. $R_{\mathrm{Mkt},t}$: return on the CRSP value-weighted market index; Jun 1978 to Jun 2015. $dGDP_t$: log growth rate of seasonally adjusted GDP; Jun 1978 to Jun 2015. $tSpd_t$: the term spread; Jan 1982 to Jun 2015. $Rcsn_t$: NBER indicator for whether a recession is taking place; Jun 1978 to Jun 2015. Each column represents results of a separate regression as described in Equation (27). Numbers in parentheses are Newey-West standard errors. Significance: * = 10%, ** = 5%, and *** = 1% under the assumption of a single test.

33

In Equation (28), $\hat{\mu}_t$ denotes the mean excess return of all stocks in month t, $(N_t + 1)$ denotes the number of stocks in month t, and the ith variable has been normalized to have zero mean and unit variance, $0 = \widehat{E}[X_{n,i,t-1}] = \frac{1}{N_t+1} \cdot \sum_n X_{n,i,t-1}$ and $1 = \widehat{\text{Var}}[X_{n,i,t-1}] = \frac{1}{N_t} \cdot \sum_n (X_{n,i,t-1} - \widehat{E}[X_{n,i,t-1}])^2$.

All of our results are robust to computing $R_{i,t}$ by ranking stocks based on $X_{n,i,t-1}$ and forming top/bottom 30% long-short portfolios à la Fama and French (1993) and Jegadeesh and Titman (1993). We compute variable-specific returns using OLS weights so that our empirical analysis exactly matches our statistical framework. This choice is not essential to our results.

To account for the fact that some variable represent contrarian strategies—e.g., '*Ret, 1-0*', which represents short-run reversals and delivers negative returns on average—we define an indicator variable $direction_i \in \{-1, +1\}$ that flips the sign of the benchmark portfolio's holdings for all contrarian strategies. For example, since there are return reversals at the one-month horizon and momentum at the 12-month horizon, we have that $direction_{Ret, \ 1\text{-}0} = -1$ while $direction_{Ret, \ 12\text{-}1} = +1$. We define this direction indicator only once at the time the variable is discovered and hence, it does not rely on any future information.

To make a one-month-ahead forecasts for these returns, we first fit an AR(3) model to the previous five years of monthly data on $\hat{\beta}_{i,t}$:

$$\hat{\beta}_{i,t'} = \check{a}_i + \sum_{\ell=1}^{3} \check{b}_{i,\ell} \cdot \hat{\beta}_{i,t'-\ell} + \check{e}_{i,t'} \qquad \text{for months } t' = (t-60), \ldots, (t-1). \qquad (29)$$

$\check{a}_i$ and $\{\check{b}_{i,\ell}\}_{\ell=1,2,3}$ denote estimated coefficients, and $\check{e}_{i,t'}$ represents the regression residual in month t'. Note that these coefficients will be different for each forecast date t; we have just suppressed the t subscripts for clarity.

We then apply these estimated coefficients to the final three months of data prior to month t:

$$\bar{\beta}_{i,t} \overset{\text{def}}{=} E_{t-1}[\hat{\beta}_{i,t}] = \check{a}_i + \sum_{\ell=1}^{3} \check{b}_{i,\ell} \cdot \hat{\beta}_{i,t-\ell}. \qquad (30)$$

If the resulting forecast is very different from zero, $|\bar{\beta}_{i,t}| \gg 0$, then we say that the ith variable is a strong signal. By contrast, if $|\bar{\beta}_{i,t}| \approx 0$, then we say it is a weak signal.

The choice of forecasting model is not essential to our results. Just like with our one-month-ahead forecasts for prior variance, we use an AR(3) forecasting model out of convenience. Our results are robust to using alternative number of lags as shown in Table D2 of Appendix D.

Benchmark Strategy. Our benchmark strategy only uses these one-month-ahead forecasts to decide whether or not to invest in the ith variable in month t. It explicitly does not take into consideration any information about the anomaly base rate. It only looks at the data; it does not consider the ex ante probability that the ith variable is a tradable anomaly.

Let $\mathcal{A}_t$ denote the set of 'active' predictors for the benchmark strategy in month t:

$$\mathcal{A}_t \stackrel{\text{def}}{=} \left\{ i \in \mathcal{I}_t : \left| \bar{\beta}_{i,t} \right| > \textit{threshold} \right\}. \tag{31}$$

This is the collection of previously discovered anomalies whose past performance exceeds some minimum threshold. In the analysis below, we are going to set the minimum performance threshold to 1% per month. The solid blue line in the left panel of Figure 6 reports the number of active predictors each month when using this 1%-per-month threshold level.

The benchmark strategy holds an equal-weighted position in all active predictors, $i \in \mathcal{A}_t$, in month t. Its raw returns are given by:

$$R_{\mathcal{A}t,t} \stackrel{\text{def}}{=} \tfrac{1}{|\mathcal{A}_t|} \cdot \sum_{i \in \mathcal{A}_t} R_{i,t} \cdot \textit{direction}_i. \tag{32}$$

You can also compute an analogous 'net' return:

$$\tfrac{1}{|\mathcal{A}_t|} \cdot \sum_{i \in \mathcal{A}_t} \left(R_{i,t} \cdot \textit{direction}_i - \textit{threshold} \right). \tag{33}$$

We will study both kinds of returns in various contexts in the analysis below. If a strategy has positive raw returns but negative net returns, then it is not something that is tradable. The strategy is generating phantom returns that will likely disappear once you start trading.

Base-Rate Adjustment. Imagine you forecasted a large positive realized return in month t for the ith variable, $|\bar{\beta}_{i,t}| \gg 0$. If the forecasted value of $\bar{v}_{i',t}^2$ was small for all other variables $i' \neq i$ in month t and the realized returns for all variables $i = 1, \ldots, I_t$ were drawn from a common distribution, then the ith variable's true magnitude is likely small as well, $|\beta_{i,t}^\star| \approx 0$. So, even though your forecasted value represents a strong signal, you should still be reluctant to trade on it.

But, exactly how reluctant? We use the statistical framework outlined in Section 2 to answer this question. Your forecasted return for the ith variable is a noisy signal about the true realized return next month, $\bar{\beta}_{i,t} \sim \text{Normal}[\beta_{i,t}^\star, N \cdot se_i^2]$, where $se_i^2 > 0$ represents the typical size of your forecast error for the ith variable. So, if realized returns for each variable are drawn from a normal distribution each month with prior variance $\sigma^2 > 0$, $\beta_{i,t}^\star \stackrel{\text{iid}}{\sim} \text{Normal}[0, \sigma^2]$, then your best guess about the ith variable's realized return in month t would be:

$$\text{E}[R_{i,t} | \bar{\beta}_{i,t}, se_i^2, \sigma^2] = \left(1 + se_i^2 / \sigma^2 \right)^{-1} \times \bar{\beta}_{i,t}. \tag{34}$$

This formula is just an application Bayesian-normal learning. It is the same as the formula for the Ridge-regression slope coefficient in Equation (11). And, you could use it to revise your forecast for next month's realized return if you also had forecasts for se_i^2 and σ^2.

Forecasting se_i^2 in month t is easy enough. You can fit an AR(3) model to the squared residuals from the $\hat{\beta}_{i,t}$ forecasting regression in Equation (29):

$$\check{e}_{i,t'}^2 = \hat{\eta}_i + \sum_{\ell=1}^3 \hat{\theta}_{i,\ell} \cdot \check{e}_{i,t'-\ell}^2 + \hat{\omega}_{i,t'} \qquad \text{for months } t' = (t-60), \ldots, (t-1). \qquad (35)$$

$\hat{\eta}_i$ and $\{\hat{\theta}_{i,\ell}\}_{\ell=1,2,3}$ denote coefficients associated with the ith variable when forecasting the value for month t, and $\hat{\omega}_{i,t'}$ represents the regression residual in month t'. Note that these coefficients will be different for each forecast date t; we have just suppressed the t subscripts for clarity.

You can then forecast the standard error in the following month by applying these estimated coefficients to the final three months of data prior to month t:

$$\bar{se}_{i,t}^2 \overset{\text{def}}{=} \mathrm{E}_{t-1}[\check{e}_{i,t}^2] = \hat{\eta}_i + \sum_{\ell=1}^3 \hat{\theta}_{i,\ell} \cdot \check{e}_{i,t-\ell}^2. \qquad (36)$$

Again, forecasts are robust to using different numbers of lags and alternative forecasting models.

Computing a forecast of prior variance for use when evaluating the ith variable in month t requires a bit more subtlety. You cannot use data about the ith variable to inform the prior you use when evaluating the ith variable. That would be circular logic. So, when adjusting the forecast of the ith variable, we compute the average prior-variance forecast using all other variables in the anomaly zoo besides, $i' \neq i$:

$$\bar{v}_{\neg i,t}^2 \overset{\text{def}}{=} \frac{1}{I_t - 1} \cdot \sum_{i' \neq i} \bar{v}_{i',t}^2. \qquad (37)$$

This means that the forecasted return associated with each variable i will be adjusted by a slightly different prior variance:

$$\bar{\beta}_{i,t}^{\bar{v}^2} \overset{\text{def}}{=} \left(1 + \bar{se}_{i,t}^2 / \bar{v}_{\neg i,t}^2\right)^{-1} \times \bar{\beta}_{i,t}. \qquad (38)$$

It also means that you cannot calculate the base-rate adjustment unless you have past experience with at least two other variables. In other words, you need $I_t \geq 3$.

Ideally though, you would have many more than two other variables. For this reason, even though our data sample starts in May 1973, all of the trading-strategy results in this section will compare the benchmark strategy to a base-rate-adjusted strategy using data starting in January 1990 when there are at least $I_t \geq 20$ variables in the anomaly zoo as shown in Table 1a.

Economic Interpretation. The benchmark strategy in Equation (32) does not explicitly use any information about σ^2 or how this prior variance evolves over time. But, that does not mean it is agnostic about this quantity. In fact, the benchmark strategy implicitly makes a very strong assumption about the anomaly base rate—namely, it assumes any level of cross-sectional predictability is just as likely as the next ex ante. The way this assumption shows up mathematically is as follows:

$$\lim_{\sigma^2 \to \infty} \bar{\beta}_{i,t}^{\sigma^2} = \bar{\beta}_{i,t}. \qquad (39)$$
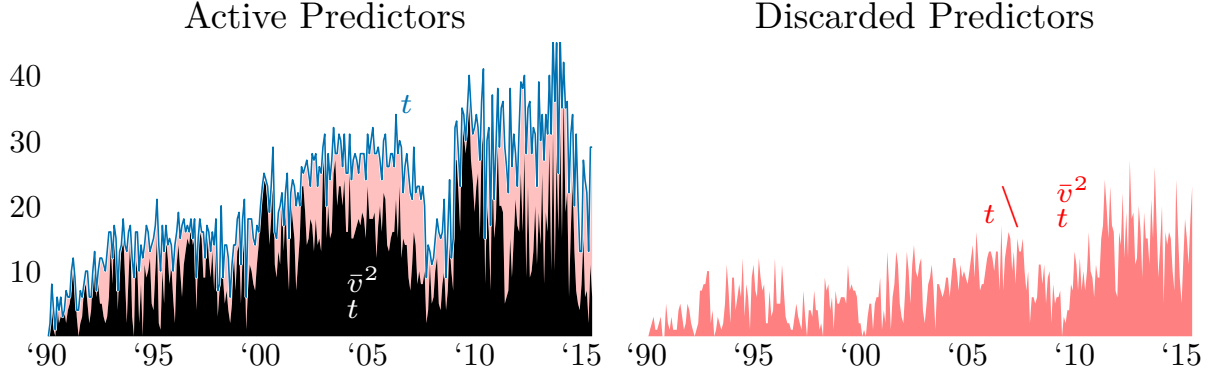
Figure 6. Number of Active Predictors. Black region, $\mathcal{A}_t^{\bar{v}^2}$: number of variables included in the base-rate-adjusted strategy in month t. Blue line, $\mathcal{A}_t$: number of variables included in the benchmark strategy in month t. Red shaded region, $\mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}$: number of variables in the benchmark strategy that were discarded due to base-rate considerations in month t. Height of this region is the same in both panels. Sample Period: Jan 1990 to Jun 2015.

Put differently, in order for it to make sense to completely ignore your prior beliefs and only pay attention to the data presented by today's seminar speaker, your prior beliefs must have been completely uninformative. You have to take a stand on your prior beliefs in order to form posteriors. You cannot punt on this issue. Ignoring the problem is taking a really strong position.

By contrast, if your past experience tells you that there are never any tradable anomalies, then you should discard all evidence to the contrary:

$$\lim_{\sigma^2 \to 0} \bar{\beta}_{i,t}^{\sigma^2} = 0. \tag{40}$$

If $\sigma^2 = 0$ exactly, then every past candidate predictor must have turned out to have no out-of-sample predictive power. But, if you really believed that $\sigma^2 = 0$, then you should not have even bothered showing up to the seminar speaker's talk today. Nothing he could have said would change your mind. The talk would be a wasted hour for you. This is the sense in which believing that $\sigma^2 = 0$ is equivalent to having dogmatic priors.

Base-Rate-Adjusted Strategy. We modify the benchmark trading strategy to account for changes in the prevailing anomaly base rate by choosing which variables to invest in based on $\bar{\beta}_{i,t}^{\bar{v}^2}$ rather than $\bar{\beta}_{i,t}$. Let $\mathcal{A}_t^{\bar{v}^2}$ denote the set of active predictors used by the base-rate-adjusted strategy in month t:

$$\mathcal{A}_t^{\bar{v}^2} \stackrel{\text{def}}{=} \left\{ i \in \mathcal{I}_t : \left| \bar{\beta}_{i,t}^{\bar{v}^2} \right| > threshold \right\}. \tag{41}$$

The solid black region in the left panel of Figure 6 reports the number of active predictors each month for the base-rate-adjusted strategy when the threshold is set to 1% per month. Because the base-rate-adjusted strategy is revising the benchmark return forecasts toward zero, the

base-rate-adjusted strategy will never include more active predictors than the benchmark strategy.

Let $R_{\mathcal{A}_t^{\bar{v}^2},t}$ denote the gross realized return of the base-rate-adjusted strategy at time t:

$$R_{\mathcal{A}_t^{\bar{v}^2},t} \stackrel{\text{def}}{=} \tfrac{1}{|\mathcal{A}_t^{\bar{v}^2}|} \cdot \sum_{i \in \mathcal{A}_t^{\bar{v}^2}} R_{i,t} \cdot direction_i. \tag{42}$$

We also calculate an analogous net return for this strategy:

$$\tfrac{1}{|\mathcal{A}_t^{\bar{v}^2}|} \cdot \sum_{i \in \mathcal{A}_t^{\bar{v}^2}} \left(R_{i,t} \cdot direction_i - threshold \right). \tag{43}$$

These return calculations are the same as for the benchmark strategy, just using $\mathcal{A}_t^{\bar{v}^2}$ rather than $\mathcal{A}_t$.

Let $\mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}$ denote the set of variables discarded due to base-rate considerations:

$$\mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2} \stackrel{\text{def}}{=} \left\{ i \in \mathcal{I}_t : \left| \bar{\beta}_{i,t} \right| > threshold > \left| \bar{\beta}_{i,t}^{\bar{v}^2} \right| \right\}. \tag{44}$$

We plot the number of variables in this discarded set in the right panel of Figure 6. This discarded set of variables represents a clean laboratory for studying the effects of incorporating the anomaly base rate. These variables look like strong signals when considered only on their own merits, but they should turn out to be bad investment ideas after incorporating the right prior beliefs. We will return to this idea when we evaluate trading-strategy performance below.

No Future Information. We want to emphasize that there is no future information being used to construct either the benchmark strategy or the base-rate-adjusted strategy. These trading strategies are both investable. Each takes the set of candidate predictors that have been previously published, and decides which of these variables is worth trading on using the past five years of data. Nothing about the construction of either portfolio requires knowledge about future returns.

Variable Selection. Before we get to these results, it is helpful to get a sense of which variables get included in the benchmark strategy, which get included in the base-rate-adjusted strategy, and which get discarded by the base-rate-adjusted strategy. Table 7 presents a detailed account of variable-specific usage rates. There are three panels. Within each panel, a row corresponds to a single variable just like in Tables 2 and 3. The column labeled $\mathcal{A}_t$ reports the fraction of all post-discovery months that the ith variable was held in the benchmark strategy. The column labeled $\mathcal{A}_t^{\bar{v}^2}$ reports the same statistic but for the base-rate-adjusted strategy.

The sparkline plots then report precisely which months each variable was held by each strategy. The x-axis in each plot represents time in months. Every time series ends in June 2015. So, variables discovered later in our sample period will have shorter sparkline plots. Months in which the ith variable was held by the benchmark strategy are denoted by vertical blue bars on the bottom half of each sparkline plot. Months in which the ith variable was held by the base-rate-adjusted strategy are denoted by vertical black bars on the top half of each sparkline plot. And, months in which the

*i*th variable was discarded by the base-rate-adjusted strategy are denoted by vertical red bars on the top half of each sparkline plot. The color scheme matches the one used in Figure 6.

These sparkline plots show more detailed patterns in variable usage. For example, short-run reversals (*Ret*, 1-0; left panel near bottom) was selected by the benchmark strategy in 77% of months. And, as a result, the bottom half of its sparkline plot is mostly blue. The base-rate-adjusted strategy only selected this variable in 44% of months, though. So, the top half of the sparkline plot is split almost evenly between black bars (included) and red bars (discarded). The sparkline plot further indicates that this variable was more likely to be discarded by the base-rate-adjusted strategy later in our sample period.

4.2   Realized Returns

We now examine the realized returns of these trading strategies and find that adjusting the benchmark strategy for the prevailing anomaly base rate significantly boosts its performance. Our aim is to demonstrate that our simple statistical framework is able to estimate an economically important quantity—namely, the anomaly base rate—with an actionable degree of precision. Showing that it is possible to trade on this information is one way of doing this, which is also of great practical importance. Just think about hitting `Ctrl+H` and replacing 'seminar speaker' with 'quant researcher' on the first two pages.

Active Predictors. We begin by comparing the performance of the active predictors that were selected by each strategy to the inactive predictors that were not. We do this by regressing each variable's realized returns in month t, $R_{i,t} \cdot direction_i$, on an indicator variable for whether the *i*th variable was included in a given strategy. For example, for the benchmark strategy, this means running the following regression:

$$R_{i,t} \cdot direction_i = \hat{a} + \hat{b} \cdot \mathbf{1}\Big[i \in \mathcal{A}_t \Big] + \hat{c} \cdot \bar{s}e_{i,t}^2 + \hat{e}_{i,t}. \tag{45}$$

The regression uses data on each variable in the anomaly zoo $i = 1, \ldots, I_t$ as of month $t = \text{Jan'}90, \ldots, \text{Jun'}15$. We start in January 1990 to ensure that the anomaly zoo is sufficiently large for us to form an accurate estimate for $\bar{v}_{-i,t}^2$. Our results are robust to using alternative start dates. Column (1) in Table 8 shows that active predictors which were selected by the benchmark strategy have 1.56% per month higher excess returns than inactive predictors.

Column (2) displays analogous results for the base-rate-adjusted strategy:

$$R_{i,t} \cdot direction_i = \hat{a} + \hat{b} \cdot \mathbf{1}\Big[i \in \mathcal{A}_t^{\bar{v}^2} \Big] + \hat{c} \cdot \bar{s}e_{i,t}^2 + \hat{e}_{i,t}. \tag{46}$$

Active predictors selected by the base-rate-adjusted strategy have 1.83% per month higher excess returns. But, this is not really a fair comparison. The base-rate-adjusted strategy is trading on

	$\mathcal{A}_t$	$\mathcal{A}_t^{\bar{v}^2}$		$\mathcal{A}_t$	$\mathcal{A}_t^{\bar{v}^2}$		$\mathcal{A}_t$	$\mathcal{A}_t^{\bar{v}^2}$
$Beta$	0.40	0.11	$Deprc/PP\&E$	0.47	0.27	$\varepsilon\%\Delta Mcap$	0.46	0.30
$BetaSq$	0.41	0.12	Ret, 6-1	0.16	0.06	$NetExtnlFin/Assets$	0.03	0.03
$IdioVolCAPM$	0.10	0.04	$\%\Delta Sales$	0.69	0.53	$DailyBeta$	0.36	0.09
$Earn/Share$	0.67	0.26	$OpAccr$	0.29	0.10	$NetPO/Price$	0.47	0.21
$Debt/Price$	0.37	0.23	$CapitalTOver$	0.51	0.40	$PO/Price$	0.48	0.30
$Divd/Price$	0.54	0.28	$RetOnEquity$	0.73	0.45	$NetPO$	0.09	0.04
$Mcap$	0.77	0.59	$KaplanZingales$	0.24	0.18	$RetOnInvstCap$	0.51	0.29
$Earn/Price$	0.61	0.31	$\%\Delta[\Delta Sales/\Delta Invtry]$	0.15	0.11	$\%\Delta Shares$	0.11	0.08
Ret, 36-12	0.75	0.56	$\%\Delta[\Delta Sales/\Delta Recv]$	0.29	0.23	$ProfMgn$	0.10	0.07
$AvgSpread$	0.79	0.55	$\%\Delta[\Delta Sales/\Delta XG\&A]$	0.23	0.21	$AdjProfMgn$	0.14	0.06
$Assets/Mcap$	0.19	0.10	$\%\Delta[\Delta GrMgn/\Delta Sales]$	0.18	0.13	$RetOnOpAssets$	0.29	0.15
$Levrg$	0.22	0.12	$LagTOver$	0.41	0.16	$AssetTOver$	0.42	0.35
$Levrg/Price$	0.37	0.23	$\mathrm{Adj}[BkVal/Mcap]$	0.74	0.66	$AdjAssets$	0.64	0.27
$Sales/Cash$	0.42	0.22	$AdjMcap$	0.68	0.59	$\%\Delta InvmtX$	0.05	0.01
$LtCF$	0.32	0.15	$SdTOver$	0.41	0.06	$\%\Delta Invmt$	0.28	0.18
$CurrRatio$	0.37	0.33	$AdvertRate/Ret$	0.72	0.49	$\Delta AdjShares$	0.78	0.65
$\%\Delta CurrRatio$	0.34	0.29	$R\&D/Mcap$	0.06	0.02	$RetOnCash$	0.26	0.17
$\%\Delta QuickRatio$	0.14	0.12	$R\&D/Sales$	0.23	0.03	$Tangibility$	0.17	0.07
$\%\Delta[Sales/Invtry]$	0.29	0.15	$Advert/Mcap$	0.84	0.77	$\Delta AdjTOver$	0.16	0.02
$QuickRatio$	0.26	0.09	$\Delta Invtry/Assets$	0.46	0.35	$UnexplVlm$	0.88	0.83
$Sales/Invtry$	0.25	0.17	$OpCF/Price$	0.20	0.05	$RetOnAssets$	0.68	0.44
$Sales/Recv$	0.34	0.23	$Invmt/\mathrm{Lag}[AvgInvmt]$	0.20	0.09	$OpLevrg$	0.72	0.46
Ret, 1-0	0.77	0.44	$NetOpAssets/Sales$	0.17	0.06	$MaxRet$	0.53	0.25
Ret, 12-1	0.34	0.14	$\%\Delta BkVal$	0.68	0.54	$FreeCF$	0.68	0.34
$BkVal$	0.11	0.08	$\%\Delta LtDebt$	0.36	0.25	$R\&Dcapital$	0.22	0.02
$MonthlyMcap$	0.85	0.64	$Price\text{-}52WkHi$	0.61	0.31	$\%\Delta Invtry$	0.00	0.00
$Sales/Price$	0.71	0.48	$IdioVolFF93$	0.16	0.05	Ret, 12-6	0.15	0.05
$\%\Delta[Deprc/PP\&E]$	0.05	0.04	$TotVol$	0.15	0.08	$CashHldgs$	0.41	0.15
$D\&A/Assets$	0.68	0.49						

Table 7. Variable Selection. $\mathcal{A}_t$: fraction of all post-discovery months that a variable was held by benchmark strategy. $\mathcal{A}_t^{\bar{v}^2}$: same statistic for the base-rate-adjusted strategy. Sparkline plots: x-axis is time in months. All time series end in June 2015. Variables discovered later have shorter sparkline plots. Blue bar, bottom half: variable is held by benchmark strategy. Black bar, top half: variable is held by base-rate-adjusted strategy. Red bar, top half: variable was discarded by base-rate-adjusted strategy. Sample period: January 1990 to June 2015.

return forecasts that have been shrunk toward zero. So, to be included in the base-rate-adjusted strategy, a variable had to start out with a more extreme return forecast.

To account for this fact, we add the magnitude of each variable's return forecast, $|\bar{\beta}_{i,t}|$, to the previous regression specification:

$$R_{i,t} \cdot direction_i = \hat{a} + \hat{b} \cdot \mathbf{1}\left[i \in \mathcal{A}_t^{\bar{v}^2}\right] + \hat{c} \cdot \bar{se}_{i,t}^2 + \hat{d} \cdot |\bar{\beta}_{i,t}| + \hat{e}_{i,t}. \tag{47}$$

Column (3) reports an even larger increase in performance, $\hat{b} = 2.02\%$ per month, after controlling for the scale of the ith variable's return forecast. Thus, the boost in performance that comes with adjusting for the prevailing anomaly base rate cannot be explained by the base-rate-adjusted strategy simply selecting more extreme predictors.

Our ideal experiment would involve comparing two variables with the same out-of-sample return forecasts in two different base-rate regimes, one with $\sigma^2 \gg 0$ and the other with $\sigma^2 \approx 0$. In this situation, we should find that the variable in the $\sigma^2 \approx 0$ regime consistently under-performs the otherwise identical variable in the $\sigma^2 \gg 0$ regime.

We approximate this ideal experiment by estimating the following regression specification:

$$\begin{aligned}
R_{i,t} \cdot direction_i = \hat{a} &+ \hat{b}_1 \cdot \mathbf{1}\left[i \in \mathcal{A}_t^{\bar{v}^2}\right] \\
&+ \hat{b}_2 \cdot \mathbf{1}\left[i \in \mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}\right] + \hat{c} \cdot \bar{se}_{i,t}^2 + \hat{d} \cdot |\bar{\beta}_{i,t}| + \hat{e}_{i,t}.
\end{aligned} \tag{48}$$

This specification controls for the magnitude and precision of the one-month-ahead return forecasts, $|\bar{\beta}_{i,t}|$ and $\bar{se}_{i,t}^2$. So, the coefficient on $\mathbf{1}\left[i \in \mathcal{A}_t^{\bar{v}^2}\right]$ captures the excess returns to trading on a strong cross-sectional predictor in the high base-rate regime. Whereas, the coefficient on $\mathbf{1}\left[i \in \mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}\right]$ reflects the excess returns to trading on that exact same cross-sectional predictor in the low base-rate regime—i.e., in a regime where $\sigma^2 \approx 0$ leading to lots of shrinkage.

Column (4) displays the results for this regression specification. We find that, for a given level of in-sample signal strength, the ith variable is a much better forecasting tool during the high base-rate regime: $\hat{b}_1 = 2.20\%$ per month vs. $\hat{b}_2 = 0.85\%$ per month. The difference between these coefficient estimates is statistically significant at the 1% level. And, they suggest an economically large difference in trading-strategy performance. If the prevailing anomaly base rate in month t was high enough such that $\bar{\beta}_{i,t}^{\bar{v}^2} > 1\%$, then the point estimates in column (4) of Table 8 suggest that trading on the ith variable would result in a monthly return of 2.20%. By contrast, if the prevailing anomaly base rate in month t was such that $\bar{\beta}_{i,t}^{\bar{v}^2} < 1\%$, then trading on the ith variable would only result in a monthly return of 0.85%. While still positive, note that this predicted value of 0.85% per month is no longer large enough to cover the 1%-per-month minimum return threshold. Thus, we have clear evidence that trading on the same variable in two different base-rate regimes leads to two very different outcomes.

Dependent Variable: $R_{i,t} \cdot direction_i$						
	(1)	(2)	(3)	(4)	(5)	(6)
$Const$	-0.03 (0.18)	0.13 (0.18)	0.23 (0.13)	0.07 (0.14)	0.22 (0.13)	0.20 (0.13)
$\mathbf{1}\big[\,i \in \mathcal{A}_t\,\big]$	$1.56^{\star\star\star}$ (0.23)					
$\mathbf{1}\big[\,i \in \mathcal{A}_t^{\bar{v}^2}\,\big]$		$1.83^{\star\star\star}$ (0.25)	$2.02^{\star\star\star}$ (0.32)	$2.20^{\star\star\star}$ (0.34)		
$\mathbf{1}\big[\,i \in \mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}\,\big]$				$0.85^{\star\star}$ (0.17)		
$\mathbf{1}\big[\,i \in \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]}\,\big]$					$1.67^{\star\star\star}$ (0.30)	
$\mathbf{1}\big[\,i \in \mathcal{A}_t^{\bar{v}^2} \cap \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]}\,\big]$						$2.05^{\star\star\star}$ (0.33)
$\mathbf{1}\big[\,i \in \mathcal{A}_t^{\bar{v}^2} \setminus \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]}\,\big]$						$2.18^{\star\star}$ (0.83)
$\mathbf{1}\big[\,i \in \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]} \setminus \mathcal{A}_t^{\bar{v}^2}\,\big]$						0.48 (0.27)
$\bar{se}^2_{i,t}$	0.01 (0.01)	0.02 (0.01)	$0.03^{\star}$ (0.01)	$0.02^{\star}$ (0.01)	$0.03^{\star}$ (0.01)	$0.03^{\star}$ (0.01)
$|\bar{\beta}_{i,t}|$			-0.08 (0.07)	-0.08 (0.07)	-0.06 (0.07)	-0.08 (0.07)
$Adj.\,R^2$	0.98%	1.05%	1.12%	1.26%	0.86%	1.13%
#Obs	15,913	15,913	15,913	15,913	15,913	15,913

Table 8. Active Predictors. Relationship between the realized returns to a zero-cost long-short portfolio based on the ith variable, $R_{i,t} \cdot direction_i$, and portfolio inclusion. Each column reports results for a separate regression involving 15,913 variable×month observations—one observation for each variable in the anomaly zoo $i = 1, \ldots, I_t$ as of month $t = \mathrm{Jan'90}, \ldots, \mathrm{Jun'15}$. $\mathbf{1}[i \in \mathcal{A}_t]$: indicator variable for the ith variable's inclusion in the benchmark strategy in month t. $\mathbf{1}[i \in \mathcal{A}_t^{\bar{v}^2}]$: indicator variable for the ith variable's inclusion in the base-rate-adjusted strategy in month t. $\mathbf{1}[i \in \mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}]$: indicator variable for the ith variable's inclusion in the benchmark strategy but not the base-rate-adjusted strategy in month t. $\mathbf{1}[i \in \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]}]$: indicator variable for the ith variable's inclusion in the average base-rate-adjusted strategy in month t. $\mathbf{1}[i \in \mathcal{A}_t^{\bar{v}^2} \cap \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]}]$: indicator variable for the ith variable's inclusion in both the true base-rate-adjusted strategy and the average base-rate-adjusted strategy in month t. $\mathbf{1}[i \in \mathcal{A}_t^{\bar{v}^2} \setminus \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]}]$: indicator variable for the ith variable's inclusion in the true base-rate-adjusted strategy but not the average base-rate-adjusted strategy in month t. $\mathbf{1}[i \in \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]} \setminus \mathcal{A}_t^{\bar{v}^2}]$: indicator variable for the ith variable's inclusion in the average base-rate-adjusted strategy but not the true base-rate-adjusted strategy in month t. $|\bar{\beta}_{i,t}|$: absolute value of one-month-ahead return forecast for ith variable in month t. $\bar{se}^2_{i,t}$: variance of one-month-ahead return forecast for ith variable in month t. Numbers in parentheses are standard errors clustered by variable. Significance: $\star = 10\%$, $\star\star = 5\%$, and $\star\star\star = 1\%$ under the assumption of a single test.

Base Rate Or Just Shrinkage? The base-rate-adjusted strategy improves performance by acting on return forecasts that have been shrunk toward zero. We motivated this shrinkage in Section 2 as a way of incorporating information about the prevailing anomaly base rate. But, so far we have not seen any evidence of this link in the data. Put differently, perhaps any sort of shrinkage will boost performance regardless of how it is motivated?

To address this concern, column (5) in Table 8 shows results for a version of the regression specification in Equation (47) using the average level of prior variance, $\mathrm{E}[\bar{v}_t^2] = 2.19$, rather than the prevailing value of $\bar{v}_t^2$ to construct the base-rate-adjusted portfolio:

$$R_{i,t} \cdot direction_i = \hat{a} + \hat{b} \cdot \mathbf{1}\Big[i \in \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]} \Big] + \hat{c} \cdot \bar{se}_{i,t}^2 + \hat{d} \cdot |\bar{\beta}_{i,t}| + \hat{e}_{i,t}. \tag{49}$$

Using the average level of prior variance mutes the effect of adjusting for the anomaly base rate, $\hat{b} = 1.67\%$ per month rather than 1.83% as in column (3), which suggests that there is more going on than just shrinkage. The time-series variation in the anomaly base rate matters.

But, it is possible to construct an even sharper test by exploiting this time-series variation. When the anomaly base rate is higher than its sample average, there are going to be variables that are strong enough to be included in the true base-rate-adjusted portfolio but not in the average base-rate-adjusted portfolio, $\mathbf{1}\Big[i \in \mathcal{A}_t^{\bar{v}^2} \setminus \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]} \Big]$. The subsequent returns to trading on these variables should look like the subsequent returns to trading on any other active predictor in the base-rate-adjusted strategy. Conversely, when the anomaly base rate is lower than its sample average, there are going to be variables that are strong enough to be included in the average base-rate-adjusted portfolio but not in the true base-rate-adjusted portfolio, $\mathbf{1}\Big[i \in \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]} \setminus \mathcal{A}_t^{\bar{v}^2} \Big]$. These variables should have subsequent returns that look very different.

And, consistent with this logic, when we run a version of the regression in Equation (47) where we split the active set along these lines,

$$\begin{aligned} R_{i,t} \cdot direction_i = \hat{a} &+ \hat{b}_1 \cdot \mathbf{1}\Big[i \in \mathcal{A}_t^{\bar{v}^2} \cap \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]} \Big] \\ &+ \hat{b}_2 \cdot \mathbf{1}\Big[i \in \mathcal{A}_t^{\bar{v}^2} \setminus \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]} \Big] \\ &+ \hat{b}_3 \cdot \mathbf{1}\Big[i \in \mathcal{A}_t^{\mathrm{E}[\bar{v}^2]} \setminus \mathcal{A}_t^{\bar{v}^2} \Big] + \hat{c} \cdot \bar{se}_{i,t}^2 + \hat{d} \cdot |\bar{\beta}_{i,t}| + \hat{e}_{i,t}, \end{aligned} \tag{50}$$

this is exactly what we find. Column (6) in Table 8 shows that $\hat{b}_2 = 2.18\%$ per month while $\hat{b}_3 = 0.48\%$ per month. This difference in coefficient estimates is statistically significant and economically large, confirming that the success of the base-rate-adjusted strategy cannot be replicated by a simple shrinkage-based approach.

Additional Results. In Appendix C, we report additional details about the performance of the benchmark and base-rate-adjusted strategies. We plot the cumulative returns of each strategy

(Figure C1). We present various performance metrics, such as Sharpe ratio and skewness, of each strategy (Table C1). The base-rate-adjusted strategy has a substantially higher Sharpe ratio than the benchmark strategy. But, the base-rate-adjusted strategy is not more negatively skewed. Moreover, the active set associated with the base-rate-adjusted strategy, $\mathcal{A}_t^{\bar{v}^2}$, is no more likely to contain high-turnover predictors than the active set associated with the benchmark strategy, $\mathcal{A}_t$. This fact suggests that the boost in performance coming from the base rate is not due to transaction costs. We also show that the base-rate-adjusted strategy is not just loading on market risk (Table C2).

Finally, we motivated this paper by discussing the problem faced by a researcher who is trying to evaluate statistical evidence concerning a *new* cross-sectional predictor. However, the trading strategies discussed so far have held positions in both brand new and previously discovered variables. e.g., it might invest based on a firm's investment growth at any time on or after December 2004 when Titman et al. (2004) published the paper documenting the variable. So, maybe the improved performance of the base-rate-adjusted strategy relative to the benchmark strategy is only due to different positions in previously discovered variables? In Appendix Table C3, we show this is not the case.

5 Conclusion

Evaluating evidence of cross-sectional predictability is hard. Today's seminar speaker is presenting evidence about how lagged values of some new variable, X_n, are positively related to subsequent excess returns at the 1% level of statistical significance. In order to form the correct posterior beliefs based on this evidence, you need to start out with the correct prior beliefs about X_n being a tradable anomaly and then appropriately revise this prior based on the speaker's statistically significant results for $\hat{\beta} > 0$.

The existing academic literature contains hundreds of other variables that also seem to predict the cross-section of expected returns. These variables are often based on entirely different data sources and have little in common with one another economically. The existence of this so-called 'anomaly zoo' represents something of a crisis in the field of empirical asset pricing. Specifically, it suggests that financial economists are not taking the effects of data mining into account when revising their prior beliefs about each new variable X_n they encounter. One way to address this problem is to demand a higher *t-statistic* cutoff when assessing statistical significance. Researchers can and should be doing this when evaluating evidence of cross-sectional predictability.

But, that being said, it is important to remember that there is more to Bayesian inference than choosing the right *t-statistic*. Anyone who walked into today's seminar room with wildly inaccurate priors is going to walk out of the seminar room with wildly inaccurate posteriors even if the speaker correctly adjusts his empirical results for data mining. There is a fast growing literature on how to adjust econometric techniques in light of widespread data mining. Yet, empirical asset pricing offers

no useful suggestions about which prior beliefs to use when evaluating the next candidate predictor you see. And, this is true in spite of the fact that, in practice, talks about new variables are often dominated by the priors of particular audience members.

In this paper, we follow Winston Churchill's advice about how one should "never let a good crisis go to waste." "Large data sets of parallel situations carry within them their own Bayesian information (Efron and Hastie, 2016, p. 77)" about the appropriate prior. And, what is the anomaly zoo if not a large data set of many parallel situations, one for each variable in the academic literature that seems to predict the cross-section of expected returns? Your past experience with other variables in the anomaly zoo (a few tradable anomalies and many more spurious predictors) contains information about the efficacy of financial economists' current research practices. And, in the absence of an overarching theory of the market conditions responsible for cross-sectional return predictability, you must have been using this information to inform your priors when walking into the seminar room today.

We provide a simple statistical approach to operationalize this insight. Then, we use trading-strategy performances to show that the anomaly-base-rate estimates produced by our simple statistical approach are economically meaningful. We use numerical simulations to demonstrate that our statistical approach can be extended in various ways. So, our hope is that, by clearly outlining this important economic problem and providing a general statistical framework for solving it, we can encourage future research on how to even better estimate the anomaly base rate and draw sharper inference about cross-sectional return predictors.

References

Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control*.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica*.

Bajgrowicz, P. and O. Scaillet (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics*.

Barberis, N. and R. Thaler (2003). A survey of behavioral finance. *Handbook of Financial Economics*.

Barras, L., O. Scaillet, and R. Wermers (2010). False discoveries in mutual-fund performance: Measuring luck in estimated alphas. *Journal of Finance*.

Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*.

Bryzgalova, S. (2017). Spurious factors in linear asset-pricing models. *Working Paper*.

Casella, G. (1985). An introduction to empirical-Bayes data analysis. *American Statistician*.

Chen, A. and T. Zimmermann (2020). Publication bias and the cross-section of stock returns. *Review of Asset Pricing Studies*.

Chinco, A., A. Clark-Joseph, and M. Ye (2018). Sparse signals in the cross-section of returns. *Journal of Finance*.

De Bondt, W. and R. Thaler (1985). Does the stock market overreact? *Journal of Finance*.

DeMiguel, V., L. Garlappi, F. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*.

Diaconis, P. and B. Skyrms (2017). *Ten great ideas about chance*. Princeton University Press.

Efron, B. (2012). *Large-scale inference: empirical-Bayes methods for estimation, testing, and prediction*. Cambridge University Press.

Efron, B. (2013). *Empirical-Bayes modeling, computation and accuracy*. Division of Biostatistics, Stanford University.

Efron, B. and T. Hastie (2016). *Computer age statistical inference*. Cambridge University Press.

Efron, B. and C. Morris (1972). Limiting the risk of Bayes and empirical-Bayes estimators—part ii: The empirical-Bayes case. *Journal of the American Statistical Association*.

Efron, B. and C. Morris (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *Journal of the American Statistical Association*.

Efron, B. and C. Morris (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*.

Fama, E. (1976). *Foundations of finance: Portfolio decisions and securities prices*. Basic Books (AZ).

Fama, E. and K. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*.

Fama, E. and K. French (1996). Multi-factor explanations of asset-pricing anomalies. *Journal of Finance*.

Feng, G., S. Giglio, and D. Xiu (2017). Taming the factor zoo. *Working Paper*.

Ferson, W., S. Sarkissian, and T. Simin (1999). The alpha-factor asset-pricing model: A parable. *Journal of Financial Markets*.

Freyberger, J., A. Neuhierl, and M. Weber (2017). Dissecting characteristics non-parametrically. *Working Paper*.

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Green, J., J. Hand, and F. Zhang (2017). The characteristics that provide independent information about average US monthly stock returns. *Review of Financial Studies*.

Gromb, D. and D. Vayanos (2010). Limits of arbitrage. *Annual Review of Financial Economics*.

Harvey, C. (2017). Presidential address: The scientific outlook in financial economics. *Journal of Finance*.

Harvey, C. and Y. Liu (2018a). Detecting repeatable performance. *Review of Financial Studies*.

Harvey, C. and Y. Liu (2018b). False (and missed) discoveries in financial economics. *Working Paper*.

Harvey, C. and Y. Liu (2018c). Lucky factors. *Working Paper*.

Harvey, C., Y. Liu, and H. Zhu (2016). . . . and the cross-section of expected returns. *Review of Financial Studies*.

Harvey, C. and G. Zhou (1990). Bayesian inference in asset-pricing tests. *Journal of Financial Economics*.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.

Hoerl, A. and R. Kennard (1970). Ridge regression: Applications to non-orthogonal problems. *Technometrics*.

James, W. and C. Stein (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock-market efficiency. *Journal of Finance*.

Jonah, B. (1986). Accident risk and risk-taking behaviour among young drivers. *Accident Analysis & Prevention*.

Kelly, B., S. Pruitt, and Y. Su (2017). Instrumented principal-component analysis. *Working Paper*.

Kelly, B., S. Pruitt, and Y. Su (2018). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*.

Kozak, S., S. Nagel, and S. Santosh (2018a). Interpreting factor models. *Journal of Finance*.

Kozak, S., S. Nagel, and S. Santosh (2018b). Shrinking the cross-section. *Journal of Financial Economics*.

Ledoit, O. and M. Wolf (2017). Non-linear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Review of Financial Studies*.

Lettau, M. and M. Pelger (2018). Estimating latent asset-pricing factors. *Working Paper*.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset-pricing tests. *Journal of Financial Economics*.

Linnainmaa, J. and M. Roberts (2018). The history of the cross-section of stock returns. *Review of Financial Studies*.

Liu, L., R. Moon, and F. Schorfheide (2018). Forecasting with dynamic panel-data models. *Working Paper*.

Lo, A. and C. MacKinlay (1990). Data-snooping biases in tests of financial asset-pricing models. *Review of Financial Studies*.

Maritz, J. (1970). *Empirical Bayes methods*. Methuen.

McLean, D. and J. Pontiff (2016). Does academic research destroy stock-return predictability? *Journal of Finance*.

Moskowitz, T., Y. Ooi, and L. Pedersen (2012). Time-series momentum. *Journal of Financial Economics*.

Novy-Marx, R. and M. Velikov (2015). A taxonomy of anomalies and their trading costs. *Review of Financial Studies*.

Park, T. and G. Casella (2008). The Bayesian LASSO. *Journal of the American Statistical Association*.

Pástor, L. (2000). Portfolio selection and asset-pricing models. *Journal of Finance*.

Robbins, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical and Statistical Probability*.

Shanken, J. (1987). A Bayesian approach to testing portfolio efficiency. *Journal of Financial Economics*.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society*.

Sullivan, R., A. Timmermann, and H. White (1999). Data-snooping, technical trading-rule performance, and the bootstrap. *Journal of Finance*.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society*.

Titman, S., J. Wei, and F. Xie (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis*.

White, H. (2000). A reality check for data snooping. *Econometrica*.

Yan, X. and L. Zheng (2017). Fundamental analysis and the cross-section of stock returns: A data-mining approach. *Review of Financial Studies*.

Zou, H., T. Hastie, and R. Tibshirani (2007). On the "degrees of freedom" of the LASSO. *Annals of Statistics*.

A Technical Appendix

Proof (Proposition 2.1). The result follows from properties of the normal distribution. If $z \sim \text{Normal}[0, \sigma^2]$, then for any $\omega > 0$ we have that:

$$\Pr[z > \omega] = \Pr[z < -\omega]$$
$$= \Phi[-\omega/\sigma].$$

Thus, $\Pr[\,anom_i\,] = \Pr[\,|\beta_i^\star| > threshold\,] = 2 \cdot \Pr[\,\beta_i^\star < -threshold\,] = 2 \cdot \Phi[\,-threshold/\sigma\,]$. □

Derivation (Equation 11). Optimizing Equation (10) results in the first-order condition:

$$0 = -2 \cdot \tfrac{1}{N} \cdot \sum_n \left(R_n - \hat{\mu} - \beta \cdot X_{n,i} \right) \cdot X_{n,i} + 2 \cdot \lambda \cdot \beta$$
$$= -\tfrac{1}{N} \cdot \sum_n (R_n - \hat{\mu}) \cdot X_{n,i} + \beta \cdot \tfrac{1}{N} \cdot \sum_n X_{n,i}^2 + \lambda \cdot \beta$$
$$= -\hat{\beta}_i + (1 + \lambda) \cdot \beta.$$

Solving for β yields the desired result. □

Derivation (Equation 16). Let $\text{S}_i[v^2]$ denote the Ridge shrinkage when $\lambda_i = \widehat{se}_i^2/v^2$:

$$\text{S}_i[v^2] = v^2/(v^2 + \widehat{se}_i^2).$$

The Ridge optimization problem can then be expressed as:

$$\min_{v^2 > 0} \left\{ \text{E}\left[\left(R_n - \hat{\mu} - \text{S}_i[v^2] \cdot \hat{\beta}_i \cdot X_{n,i} \right)^2 \right] \right\}.$$

This optimization problem results in the following first-order condition:

$$0 = 2 \cdot \text{E}\left[\left(R_n - \hat{\mu} - \text{S}_i[v^2] \cdot \hat{\beta}_i \cdot X_{n,i} \right) \cdot \left(\text{S}_i'[v^2] \cdot \hat{\beta}_i \cdot X_{n,i} \right) \right]$$
$$= 2 \cdot \text{S}_i'[v^2] \cdot \left(\hat{\beta}_i \cdot \text{E}[(R_n - \hat{\mu}) \cdot X_{n,i}] - \text{S}_i[v^2] \cdot \text{E}[(\hat{\beta}_i \cdot X_{n,i})^2] \right)$$
$$= 2 \cdot \text{S}_i'[v^2] \cdot \left(\hat{\beta}_i^2 - \text{S}_i[v^2] \cdot \hat{\beta}_i^2 \right)$$
$$= 2 \cdot \text{S}_i'[v^2] \cdot (1 - \text{S}_i[v^2]) \cdot \hat{\beta}_i^2$$
$$= 2 \cdot \frac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot \hat{\beta}_i^2.$$

Taking the expectation with respect to realizations of the true slope coefficient yields:

$$0 = \text{E}\left[2 \cdot \frac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot \hat{\beta}_i^2 \right] = 2 \cdot \frac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot (\sigma^2 + \widehat{se}_i^2).$$

The only way to satisfy this first-order condition is to choose $v^2 = \infty$. □

Proof (Proposition 2.2). Suppose we add a correction term, $\text{C}_i[v^2]$, to the training error in

Equation (15) to undo this in-sample overfitting. The objective function would then become:

$$\hat{v}_i^2 = \arg\min_{v^2 > 0} \left\{ \operatorname{E}\left[\left(R_n - \hat{\mu} - \operatorname{S}_i[v^2] \cdot \hat{\beta}_i \cdot X_{n,i} \right)^2 \right] + \operatorname{C}_i[v^2] \right\}.$$

Our goal is to find a functional form for $\operatorname{C}_i[v^2]$ that yields an unbiased estimate of $\operatorname{E}[\hat{v}_i^2] = \sigma^2$. Note that this corrected optimization problem yields the following first-order condition:

$$
\begin{aligned}
0 &= 2 \cdot \operatorname{E}\left[\left(R_n - \hat{\mu} - \operatorname{S}_i[v^2] \cdot \hat{\beta}_i \cdot X_{n,i} \right) \cdot \operatorname{S}_i'[v^2] \cdot \hat{\beta}_i \cdot X_{n,i} \right] - \operatorname{C}_i'[v^2] \\
&= 2 \cdot \operatorname{S}_i'[v^2] \cdot (1 - \operatorname{S}_i[v^2]) \cdot \hat{\beta}_i^2 - \operatorname{C}_i'[v^2] \\
&= 2 \cdot \frac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot \hat{\beta}_i^2 - \operatorname{C}_i'[v^2].
\end{aligned}
$$

And, taking the expectation of this corrected first-order condition with respect to realizations of the true slope coefficient yields:

$$0 = \operatorname{E}\left[2 \cdot \frac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot \hat{\beta}_i^2 \right] - \operatorname{C}_i'[v^2] = 2 \cdot \frac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot (\sigma^2 + \widehat{se}_i^2) - \operatorname{C}_i'[v^2].$$

By inspection, we see that choosing any $\operatorname{C}_i[v^2]$ with the following first derivative,

$$
\begin{aligned}
\operatorname{C}_i'[v^2] &= 2 \cdot \frac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot (v^2 + \widehat{se}_i^2) \\
&= 2 \cdot \widehat{se}_i^4 \cdot (v^2 + \widehat{se}_i^2)^{-2}
\end{aligned}
$$

will result in a minimum at $v^2 = \sigma^2$. Thus, by appropriately choosing the constant of integration, we can arrive at the desired result:

$$\operatorname{C}_i[v^2] = 2 \cdot \left(\frac{1}{1 + \widehat{se}_i^2 / v^2} \right) \cdot \widehat{se}_i^2.$$

$\square$

B Distributional Assumptions

The statistical approach described in Section 2 models the anomaly-discovery process as independent draws from a normal distribution. The key assumption is that the strength of cross-sectional predictors is drawn from a common distribution. The assumption of normality is not essential.

To see why, consider an alternative setting where the true slope coefficients are drawn from a Laplace distribution:

$$\beta_i^\star \overset{\text{iid}}{\sim} \text{Laplace}[\sqrt{2}/\sigma].$$

The probability density function of this Laplace distribution is given by $\text{pdf}[\beta] = \frac{1}{\sigma \cdot \sqrt{2}} \cdot e^{-\frac{\sqrt{2}}{\sigma} \cdot |\beta|}$, which implies that the mean and variance of the resulting draws are the same as in the original normally distributed case: $\text{E}[\beta_i^\star] = 0$ and $\text{Var}[\beta_i^\star] = \sigma^2$. We now show that, even though the true slope coefficients are being drawn from a different prior distribution, you can apply the exact same logic to estimate the anomaly base rate.

If the true slope coefficients are drawn from a Laplace distribution, then the functional form of our inference problem will change slightly. Now, the negative log likelihood of the true slope coefficient taking on a particular value, $\beta_i^\star = \beta$, given the realized cross-section of returns and lagged values will correspond to

$$
\begin{aligned}
-\log \Pr[\beta | \boldsymbol{R}, \boldsymbol{X}_i] &= \tfrac{1}{2 \cdot (N \cdot \widehat{se}_i^2)} \cdot \sum_n \left(R_n - \hat{\mu} - \beta \cdot X_{n,i} \right)^2 + \tfrac{\sqrt{2}}{\sigma} \times |\beta| + \cdots \\
&= \tfrac{1}{2 \cdot \widehat{se}_i^2} \cdot \left\{ \tfrac{1}{N} \cdot \sum_n \left(R_n - \hat{\mu} - \beta \cdot X_{n,i} \right)^2 + \sqrt{8} \cdot \tfrac{\widehat{se}_i^2}{\sigma} \times |\beta| \right\} + \cdots
\end{aligned}
$$

where the "$\cdots$" represents constants that do not depend on the choice of β. This inference problem suggests using a different penalized-regression procedure than before—i.e., a procedure with an absolute-value penalty rather than a quadratic penalty like a Ridge regression.

The least absolute shrinkage and selection operator (the LASSO; Tibshirani, 1996) is just such a penalized-regression procedure. Estimating the LASSO involves solving the optimization problem below:

$$\hat{\beta}_i[\lambda] \overset{\text{def}}{=} \arg \min_\beta \left\{ \tfrac{1}{N} \cdot \sum_n \left(R_n - \hat{\mu} - \beta \cdot X_{n,i} \right)^2 + \lambda \cdot |\beta| \right\}.$$

Note that this is just the optimization problem given in Equation (5) when replacing β^2 with $|\beta|$. What is more, when there is only one variable that has been standardized to have zero mean and unit variance, it is possible to characterize the solution to this optimization problem analytically:

$$\hat{\beta}_i[\lambda] = \text{Sign}[\hat{\beta}_i] \cdot (|\hat{\beta}_i| - \lambda)_+.$$

Thus, as pointed out in Park and Casella (2008), the LASSO's absolute-value penalty can be interpreted as the effect of imposing Laplace priors on an inference problem when the tuning parameter is chosen as follows:

$$\lambda_i = \sqrt{8} \cdot \widehat{se}_i^2 / \sigma.$$

The proposition below shows that, if the true slope coefficients are drawn from a Laplace distribution instead of a normal distribution, then we can learn about the anomaly base rate by studying the best-fit tuning parameter in the LASSO instead of a Ridge regression. Different prior distribution. Different penalized regression. Same underlying approach.

Proposition B (Econometric Estimator, The LASSO). *Let* $\mathrm{E}[\cdot]$ *denote an expectations operator evaluated with respect to realizations of $\beta_i^\star$ drawn from a Laplace distribution. If $\hat{v}_i^2$ denotes the parameter estimate with the minimum in-sample prediction error subject to an overfitting penalty for the ith variable,*

$$\hat{v}_i^2 \overset{\text{def}}{=} \arg\min_{v^2 > 0} \left\{ \mathrm{Err}_i\left[\widehat{se}_i^2/v^2\right] + 2 \cdot \mathbf{1}\left[|\hat{\beta}_i| > \sqrt{8} \cdot \widehat{se}_i^2/v\right] \cdot \widehat{se}_i^2 \right\},$$

then for all $\sigma^2 > 0$ we have that $\mathrm{E}[\hat{v}_i^2] = \sigma^2$.

Proof (Proposition B). The $2 \cdot \mathbf{1}\left[|\hat{\beta}_i| > \sqrt{8} \cdot \widehat{se}_i^2/v\right] \cdot \widehat{se}_i^2$ term in Proposition B is an information-criterion penalty. This sort of penalty takes the form $2 \cdot (df/N) \times \widehat{\mathrm{Var}}[\varepsilon_{n,i}]$ where df represents the estimator's degrees of freedom. Zou et al. (2007) proves that the number of non-zero slope coefficients is an unbiased estimator for the degrees of freedom when using the LASSO:

$$\mathrm{Pr}\left[|\hat{\beta}_i| > \lambda\right] = \mathrm{df}[\lambda].$$

Thus, since $\widehat{\mathrm{Var}}[\varepsilon_{n,i}] = N \cdot \widehat{se}_i^2$, the generalized information-criterion penalty reduces to the one above when $\lambda_i = \sqrt{8} \cdot \widehat{se}_i^2/\sigma$. $\qquad\square$
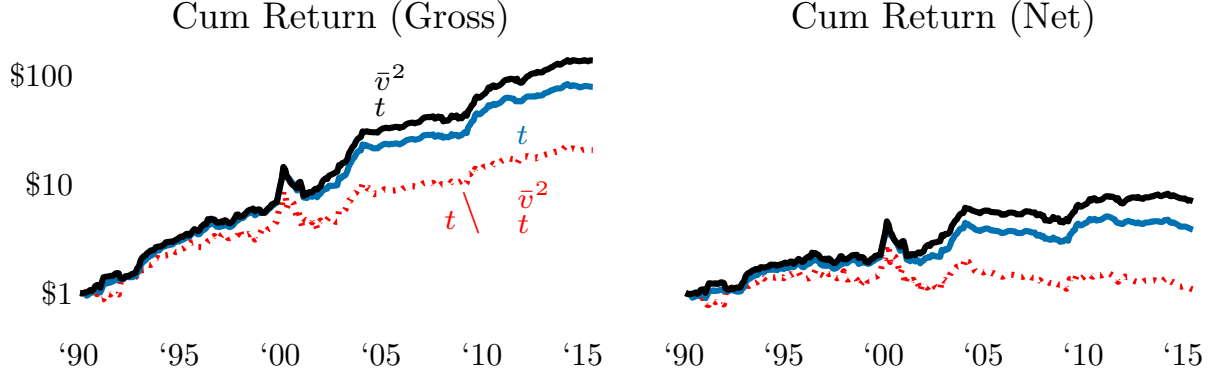
Figure C1. Cumulative Returns. Solid black line, $\mathcal{A}_t^{\bar{v}^2}$: current value of a portfolio that invested \$1 in the base-rate-adjusted strategy on January 1st, 1990. Solid blue line, $\mathcal{A}_t$: current value of a portfolio that invested \$1 in the benchmark strategy on January 1st, 1990. Dotted red line, $\mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}$: current value of a portfolio that invested \$1 in the discarded-variables strategy starting on January 1st, 1990. Left panel: cumulative gross returns. Right panel: cumulative returns net of 1% per month performance threshold. Sample Period: January 1990 to June 2015.

C Additional Results

In this appendix, we present a variety of additional results related to the performance of the benchmark and base-rate-adjusted trading strategies.

Cumulative Returns. Having shown that incorporating information about the prevailing anomaly base rate improves variable selection, we now combine these selections into a single trading strategy. We start by computing the cumulative returns to investing \$1 in either the benchmark strategy or the base-rate-adjusted strategy starting in January 1990. Figure C1 reports the total amount of money that you would have in your account in month t if you followed either strategy, continually reinvesting any capital gains along the way.

The left panel reports cumulative gross returns. By June 2015, a portfolio investing in the base-rate-adjusted strategy (solid black line; $\mathcal{A}_t^{\bar{v}^2}$) would have been worth \$137.42 while a portfolio investing in the benchmark strategy (solid blue line; $\mathcal{A}_t$) would only have been worth \$78.67. What's more, a portfolio that only invested in the variables discarded by the base-rate-adjusted strategy (red dotted line; $\mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}$) would only have been worth \$20.59. The right panel in Figure C1 shows analogous results for cumulative returns net of the 1%-per-month performance threshold.

We recognize that some variables are noisier signals than others. And, with this idea in mind, practitioners often use volatility weights to combine signals (e.g., see Moskowitz et al., 2012). We want to emphasize that these volatility weights are a different phenomenon from what we are studying in the current paper. The prior variance $\bar{v}_{\neg i,t}^2$ measures the ex ante likelihood of encountering a tradable signal. These volatility weights measure the precision of a tradable signal once it has been found.

What's more, our approach to adjusting for the anomaly base rate is accounting for this precision by dividing through by the squared forecasted standard error, $\bar{se}_{i,t}^2$, when adjusting $\bar{\beta}_{i,t}$ for the prevailing anomaly base rate in Equation (38). Thus, the success of the base-rate-adjusted strategy is not just coming from taking large bets on the most volatile predictors.

Performance Metrics. In Table C1, we explore the difference in performance between the

benchmark and base-rate-adjusted strategies in more detail. We report both the mean and standard deviation of each strategy's monthly returns as well as their skewness, kurtosis, and annualized Sharpe ratio. We also report the probability that, if you selected a variable in each strategy's active set of predictors at random, you would select a predictor in the top 33% by turnover.

The first column, middle panel, reveals that, after accounting for implementation costs, the annualized Sharpe ratio of the benchmark strategy is only 0.43 during our sample period. By contrast, when we look at the second column, we can see that the base-rate-adjusted strategy has an annualized Sharpe ratio of 0.57. And, as you would expect, this difference is due to the base-rate-adjusted strategy systematically dropping variables from the benchmark portfolio that only seem to have strong predictive power in-sample.

The third column shows that the net returns of a trading strategy that only invests in the variables that are held by the benchmark strategy but not by the benchmark-adjusted strategy are only 0.18% per month. Moreover, this discarded-variables strategy has an annualized Sharpe ratio of only 0.11. The top and bottom panels show that these conclusions are robust to varying the threshold for implementation costs.

All our results are robust to estimating the anomaly base rate using cross-validation rather than the regularized estimator in Proposition 2.2 as shown in Table D3 of Appendix D. And, the boost in performance associated with the base rate is not coming from trading excessively often. The base-rate-adjusted strategy is no more likely than the benchmark strategy to select high-turnover predictors. There is no difference in the fraction of high-turnover strategies, $\Pr[hiTO_s]$, between columns (1) and (2).

Abnormal Returns. We next show that the performance of the base-rate-adjusted strategy is not just the result of exposure to market risk. We run a time-series regression of the net excess returns to the base-rate-adjusted strategy on the excess returns to the value-weighted market:

$$R_{\mathcal{A}_t^{\bar{v}^2}, t} = \hat{a} + \hat{b} \cdot R_{\text{Mkt}, t} + \hat{e}_t.$$

In the equation above, $R_{\text{Mkt},t}$ is the excess return on the market in month t, $\hat{a}$ is the abnormal return to the base-rate-adjusted strategy, $\hat{b}$ is the slope coefficient from this time-series regression, and $\hat{e}_t$ is the regression residual. These data come from Kenneth French's website.[9]

Column (4) in Table C2 shows that market-risk exposure does not account for the base-rate-adjusted trading strategy's good performance. The average net return of the base-rate-adjusted strategy is 0.74% per month; and, after accounting for market-risk exposure, the net abnormal return to this strategy is 0.67% per month. There is hardly any difference.[10]

Columns (2) and (6) in Table C2 replicate the same analysis using the net returns of the benchmark and discarded-variables strategies as the left-hand-side variable. Neither of these alternative strategies has significant net excess returns after adjusting for exposure to the market.

New Variables. Finally, we motivated this paper by discussing the problem faced by a researcher who is trying to evaluate statistical evidence concerning a *new* cross-sectional predictor. So, in the last part of our analysis, we investigate whether our estimate for the prevailing anomaly base rate can help researchers evaluate new predictors.

---

[9]See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[10]These results are not unexpected to the extent that the firm characteristics entered our sample because they had anomalous returns relative to a factor model, often the CAPM, during some period of time.

threshold = 0.50%		$\mathcal{A}_t$ (1)	$\mathcal{A}_t^{\bar{v}^2}$ (2)	$\mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}$ (3)
Gross:	$\mathrm{E}[R_{s,t}]$	1.37	1.46	0.77
Net of minimum performance threshold:	$\mathrm{E}[R_{s,t}]$	0.87	0.96	0.32
	$\mathrm{Sd}[R_{s,t}]$	3.73	3.77	6.37
	$\mathrm{Skew}[R_{s,t}]$	1.88	1.77	1.99
	$\mathrm{Kurt}[R_{s,t}]$	11.16	9.38	17.81
	SR_s	0.80	0.88	0.17
	$\mathrm{Pr}[hiTO_s]$	0.35	0.36	0.28
1.00%				
Gross:	$\mathrm{E}[R_{s,t}]$	1.52	1.72	1.15
Net of minimum performance threshold:	$\mathrm{E}[R_{s,t}]$	0.53	0.74	0.18
	$\mathrm{Sd}[R_{s,t}]$	4.27	4.50	5.66
	$\mathrm{Skew}[R_{s,t}]$	1.31	0.94	0.92
	$\mathrm{Kurt}[R_{s,t}]$	7.75	5.79	8.87
	SR_s	0.43	0.57	0.11
	$\mathrm{Pr}[hiTO_s]$	0.32	0.34	0.29
1.50%				
Gross:	$\mathrm{E}[R_{s,t}]$	1.74	2.06	1.35
Net of minimum performance threshold:	$\mathrm{E}[R_{s,t}]$	0.26	0.68	-0.10
	$\mathrm{Sd}[R_{s,t}]$	4.76	5.09	5.73
	$\mathrm{Skew}[R_{s,t}]$	1.22	0.74	0.42
	$\mathrm{Kurt}[R_{s,t}]$	6.64	5.16	6.61
	SR_s	0.19	0.46	-0.06
	$\mathrm{Pr}[hiTO_s]$	0.31	0.33	0.29

Table C1. Performance Metrics. Performance statistics for the excess returns to three different trading strategies. Column (1): benchmark strategy, $\mathcal{A}_t$. Column (2): base-rate-adjusted strategy, $\mathcal{A}_t^{\bar{v}^2}$. Column (3): strategy that invests in variables held by the benchmark strategy but not the base-rate-adjusted strategy, $\mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}$. All return statistics quoted in % per month. $\mathrm{E}[R_{s,t}]$: mean monthly return. $\mathrm{E}[R_{s,t}]$: mean net monthly return. $\mathrm{Sd}[R_{s,t}]$: standard deviation of net monthly returns. $\mathrm{Skew}[R_{s,t}]$: skewness of net monthly returns. $\mathrm{Kurt}[R_{s,t}]$: kurtosis of net month returns. SR_s: annualized Sharpe ratio using net monthly returns. $\mathrm{Pr}[hiTO_s]$: fraction of predictors included in the strategy that are in top 33% by turnover. Top panel: threshold of 0.50% per month. Middle panel: threshold of 1.00% per month. Bottom panel: threshold of 1.50% per month. Sample period: January 1990 to June 2015.

(a) Summary Statistics

	Avg	Sd	SR
$R_{\mathrm{Mkt},t}$	0.62	4.33	0.50

(b) Regression Results

Dependent Variable: $R_{s,t}$

	$\mathcal{A}_t$		$\mathcal{A}_t^{\bar{v}^2}$		$\mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Const	0.53 (0.34)	0.44 (0.32)	0.74** (0.34)	0.68** (0.32)	0.18 (0.32)	0.04 (0.29)
$R_{\mathrm{Mkt},t}$		0.15*** (0.05)		0.09* (0.08)		0.22** (0.10)
Adj. R^2		1.86%		0.47%		2.62%
#Obs	306	306	306	306	306	306

Table C2. Abnormal Returns. Net abnormal returns relative to the market for three different trading strategies: benchmark strategy, $\mathcal{A}_t$; base-rate-adjusted strategy, $\mathcal{A}_t^{\bar{v}^2}$; set-difference strategy that invests in variables held by the benchmark strategy but not the base-rate-adjusted strategy, $\mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}$. $R_{\mathrm{Mkt},t}$: excess return on the value-weighted market portfolio. **(a) Summary Statistics.** Mean and standard deviation of the excess return on the market in units of % per month as well as the annualized Sharpe ratio. **(b) Regression Results.** Each column reports the results of a separate time-series regression with the net excess returns of a particular strategy as the left-hand side variable. *Const* has units of % per month, and the slope coefficients are dimensionless. Numbers in parentheses are Newey-West standard errors. Statistical significance: $\star = 10\%$, $\star\star = 5\%$, and $\star\star\star = 1\%$ under the assumption of a single test. Sample period: January 1990 to June 2015. All regressions involve 306 monthly observations.

To do this, we first compute the realized returns of each variable-specific trading strategy held by the benchmark strategy in the 10 years immediately after its publication. Let $R_{i,t_0(i)+120}$ denote the annualized return to the variable-specific strategy associated with the ith variable defined in Equation (28) during the 10 years immediately following its publication date as given in Tables 1a, 1b, and 1c. And, let $t_0(i)$ denote the month immediately following the publication of the ith variable, e.g., for investment growth as defined in Titman et al. (2004), $t_0(i) = Jan2005$. There are 62 variables discovered sometime after January 1990 for which we can forecast prior variance.

Then, we regress the post-publication returns of each variable on an indicator for whether the variable would have been held by the benchmark strategy:

$$R_{i,t_0(i)+120} \cdot direction_i = \hat{a} + \hat{b} \cdot \mathbf{1}\Big[i \in \mathcal{A}_{t_0(i)} \Big] + \hat{c} \cdot \bar{se}_{i,t_0(i)}^2 + \hat{e}_i. \tag{52}$$

Column (1) of Table C3 reports the results of this cross-sectional regression, which contains one observation for each of the 62 variables discovered following January 1990. We estimate that

Dependent Variable: $R_{i,t_0(i)+120} \cdot direction_i$				
	(1)	(2)	(3)	(4)
$Const$	2.53 (3.07)	3.20 (2.73)	4.62 (2.80)	3.77 (3.14)
$\mathbf{1}\big[i \in \mathcal{A}_{t_0(i)} \big]$	7.70* (4.10)			
$\mathbf{1}\big[i \in \mathcal{A}^{\bar{v}^2}_{t_0(i)} \big]$		10.37** (4.77)	14.81*** (5.33)	15.56*** (5.50)
$\mathbf{1}\big[i \in \mathcal{A}_{t_0(i)} \setminus \mathcal{A}^{\bar{v}^2}_{t_0(i)} \big]$				3.12 (5.13)
$\bar{se}_{i,t_0(i)}$	−0.66 (0.62)	−0.51 (0.62)	0.01 (0.68)	−0.01 (0.69)
$|\bar{\beta}_{i,t_0(i)}|$			−1.52* (0.87)	−1.49* (0.88)
$Adj.\ R^2$	4.64%	6.42%	9.59%	8.59%
#Obs	62	62	62	62

Table C3. Newly Discovered Variables. Evidence that the anomaly base rate is helpful when evaluating variables at the moment they are added to the academic literature. Each column reports results of a separate cross-sectional regression with 62 observations, one for each variable i discovered after January 1990. $R_{i,t_0(i)+120} \cdot direction_i$: annualized return to zero-cost long-short portfolio associated with ith variable defined in Equation (28) during 10 years immediately following publication. $Const$: intercept term; units of % per year. $\mathbf{1}[i \in \mathcal{A}_{t_0(i)}]$: indicator variable for ith variable's inclusion in the benchmark strategy in the month following its discovery, $t_0(i)$; units of % per year. $\mathbf{1}[i \in \mathcal{A}^{\bar{v}^2}_{t_0(i)}]$: indicator variable for ith variable's inclusion in the base-rate-adjusted strategy in month $t_0(i)$; units of % per year. $\mathbf{1}[i \in \mathcal{A}_{t_0(i)} \setminus \mathcal{A}^{\bar{v}^2}_{t_0(i)}]$: indicator variable for ith variable's inclusion in the benchmark strategy but not the base-rate-adjusted strategy in month $t_0(i)$; units of % per year. Numbers in parentheses are Newey-West standard errors. Statistical significance: $\star = 10\%$, $\star\star = 5\%$, and $\star\star\star = 1\%$.

$\hat{b} = 7.70\%$ per year for the benchmark strategy.

By contrast, when we estimate the same regression using an indicator variable for inclusion in the base-rate-adjusted strategy,

$$R_{i,t_0(i)+120} \cdot direction_i = \hat{a} + \hat{b} \cdot \mathbf{1}\Big[i \in \mathcal{A}^{\bar{v}^2}_{t_0(i)} \Big] + \hat{c} \cdot \bar{se}^2_{i,t_0(i)} + \hat{e}_i, \qquad (53)$$

we get a $\hat{b} = 10.37\%$ per year in column (2). Column (3) indicates that adjusting for the amount of predictive power associated with the variable at the time of publication does not explain this difference.

And, column (4) provides further evidence that a researcher should discount empirical results suggesting the existence of a strong predictor during low anomaly-base-rate regimes. Variables that would be included in the benchmark strategy at the time of discovery but not in the base-rate-adjusted strategy have average returns of only 3.12% per year over the next decade.

D   Robustness Checks

		30/30 LS (1)	$1 filter (2)	$5 filter (3)
Gross:	$E[R_{s,t}]$	1.99	1.72	2.48
Net of 1.0% minimum performance threshold:	$E[R_{s,t}]$	1.13	0.74	1.48
	$Sd[R_{s,t}]$	6.17	4.50	4.02
	$Skew[R_{s,t}]$	1.26	0.94	1.90
	$Kurt[R_{s,t}]$	7.59	5.79	13.13
	SR_s	0.63	0.57	1.28

Table D1. Performance Metrics, Data Sample Robustness of performance statistics for the excess returns of base-rate adjust strategy. Column (1): Statistics for portfolio going long 30% of stocks with highest-predicted returns and going short 30% of stocks with lowest predicted returns. Column (2): $1 price filter for small and illiquid stocks. Column (3): $5 price filter for small and illiquid stocks. All return statistics quoted in % per month. $E[R_{s,t}]$: mean monthly return. $E[R_{s,t}]$: mean net monthly return. $Sd[R_{s,t}]$: standard deviation of net monthly returns. $Skew[R_{s,t}]$: skewness of net monthly returns. $Kurt[R_{s,t}]$: kurtosis of net month returns. SR_s: annualized Sharpe ratio using net monthly returns. Sample period: January 1990 to June 2015.

		AR(1) (1)	AR(3) (2)	AR(5) (3)
Gross:	$E[R_{s,t}]$	2.06	1.72	1.59
Net of 1.0% minimum performance threshold:	$E[R_{s,t}]$	1.08	0.74	0.59
	$Sd[R_{s,t}]$	5.11	4.50	4.46
	$Skew[R_{s,t}]$	0.79	0.94	1.13
	$Kurt[R_{s,t}]$	4.72	5.79	6.40
	SR_s	0.73	0.57	0.46

Table D2. Performance Metrics, Autoregressive Order. Robustness of performance statistics to variations in the autoregressive order to forecast the out-of-sample base rate. Column (1): autoregressive process of order 1. Column (2): autoregressive process of order 3. Column (3): autoregressive process of order 5. All return statistics quoted in % per month. $E[R_{s,t}]$: mean monthly return. $E[R_{s,t}]$: mean net monthly return. $Sd[R_{s,t}]$: standard deviation of net monthly returns. $Skew[R_{s,t}]$: skewness of net monthly returns. $Kurt[R_{s,t}]$: kurtosis of net month returns. SR_s: annualized Sharpe ratio using net monthly returns. Sample period: January 1990 to June 2015.

		Regularized (1)	Cross-Validated (2)
Gross:	$\mathrm{E}[R_{s,t}]$	1.72	1.68
Net of 1.0% minimum performance threshold:	$\mathrm{E}[R_{s,t}]$	0.74	0.69
	$\mathrm{Sd}[R_{s,t}]$	4.50	4.25
	$\mathrm{Skew}[R_{s,t}]$	0.94	0.98
	$\mathrm{Kurt}[R_{s,t}]$	5.79	7.47
	SR_s	0.57	0.70

Table D3. Performance Metrics, Cross-Validated. Robustness of performance statistics to estimating $\hat{v}_{i,t}^2$ using cross-validation rather than the regularized estimator from Proposition 2.2. Column (1): baseline estimates using regularized estimator. Column (2): estimates using 10-fold cross-validation procedure as used in simulation analysis. See Figure 5 and surrounding discussion on page 22 for more details. All return statistics quoted in % per month. $\mathrm{E}[R_{s,t}]$: mean monthly return. $\mathrm{E}[R_{s,t}]$: mean net monthly return. $\mathrm{Sd}[R_{s,t}]$: standard deviation of net monthly returns. $\mathrm{Skew}[R_{s,t}]$: skewness of net monthly returns. $\mathrm{Kurt}[R_{s,t}]$: kurtosis of net month returns. SR_s: annualized Sharpe ratio using net monthly returns. Sample period: January 1990 to June 2015.