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. But 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, the *'anomaly base rate'*. We propose a way to estimate it 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. #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.

JEL Classification: C12, C52, G11

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

*We would like to thank Justin Birru, Svetlana Bryzgalova, Zhi Da, Xavier Gabaix, Niels Gormsen, Sam Hartzmark, Christian Julliard, Ralph Koijen, Bob Korajczyk, Yan Liu, Stefan Nagel, Walt Pohl, Jeff Pontiff, Tarun Ramadorai, Alessio Saretto, Andrea Tamoni, Julian Thimme, Allan Timmermann, Rüdiger Weber, and Dacheng Xiu for extremely helpful comments and suggestions. This paper has also benefited greatly from presentations at the University of Chicago, the University of Illinois, the MFA meetings, AQR Asset-Management Institute's Academic Symposium, the Future of Financial Information Conference, the 5th BI-SHoF Conference, the NBER Summer Institute, the SITE Asset-Pricing Theory and Computation Meetings, the EFA Meetings, the NFA Conference, and the SAFE Asset-Pricing Workshop. Bianca He provided excellent research assistance. Weber also gratefully acknowledges financial support from the University of Chicago, the Fama Research Fund, and the Fama-Miller Center.

Current Version: http://www.alexchinco.com/anomaly-base-rate.pdf

[†]University of Illinois, Gies College of Business. alexchinco@gmail.com

[‡]University of Notre Dame, Mendoza College of Business. aneuhier@nd.edu

[§]University of Chicago, Booth School of Business and the NBER. michael.weber@chicagobooth.edu

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:

$$\begin{aligned} \Pr[z > \omega] &= \Pr[z < -\omega] \\ &= \Phi[-\omega/\sigma]. \end{aligned}$$

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:

$$\begin{aligned} 0 &= -\,2 \cdot \tfrac{1}{N} \cdot \textstyle\sum_n \Big(R_n - \hat{\mu} - \beta \cdot X_{n,i}\Big) \cdot X_{n,i} + 2 \cdot \lambda \cdot \beta \\ &= -\,\tfrac{1}{N} \cdot \textstyle\sum_n (R_n - \hat{\mu}) \cdot X_{n,i} + \beta \cdot \tfrac{1}{N} \cdot \textstyle\sum_n X_{n,i}^2 + \lambda \cdot \beta \\ &= -\,\hat{\beta}_i + (1 + \lambda) \cdot \beta. \end{aligned}$$

Solving for β yields the desired result. □

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

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

The Ridge optimization problem can then be expressed as:

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

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

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

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

$$0 = \mathrm{E}\Big[2 \cdot \tfrac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot \hat{\beta}_i^2\Big] = 2 \cdot \tfrac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot (\sigma^2 + \widehat{se}_i^2).$$

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, $\mathrm{C}_i[v^2]$, to the training error in

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

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

Our goal is to find a functional form for $\mathrm{C}_i[v^2]$ that yields an unbiased estimate of $\mathrm{E}[\hat{v}_i^2] = \sigma^2$.

Note that this corrected optimization problem yields the following first-order condition:

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

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

$$0 = \mathrm{E}\Big[\, 2 \cdot \tfrac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot \hat{\beta}_i^2 \,\Big] - \mathrm{C}_i'[v^2] = 2 \cdot \tfrac{\widehat{se}_i^4}{(v^2 + \widehat{se}_i^2)^3} \cdot (\sigma^2 + \widehat{se}_i^2) - \mathrm{C}_i'[v^2].$$

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

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

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:

$$\mathrm{C}_i[v^2] = 2 \cdot \Big(\tfrac{1}{1 + \widehat{se}_i^2 / v^2}\Big) \cdot \widehat{se}_i^2.$$

$\square$

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

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

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:

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

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* i*th variable,*

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

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

Proof (Proposition B). The $2 \cdot \mathbf{1}\Big[|\hat{\beta}_i| > \sqrt{8} \cdot \widehat{se}_i^2/v\Big] \cdot \widehat{se}_i^2$ 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:

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

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

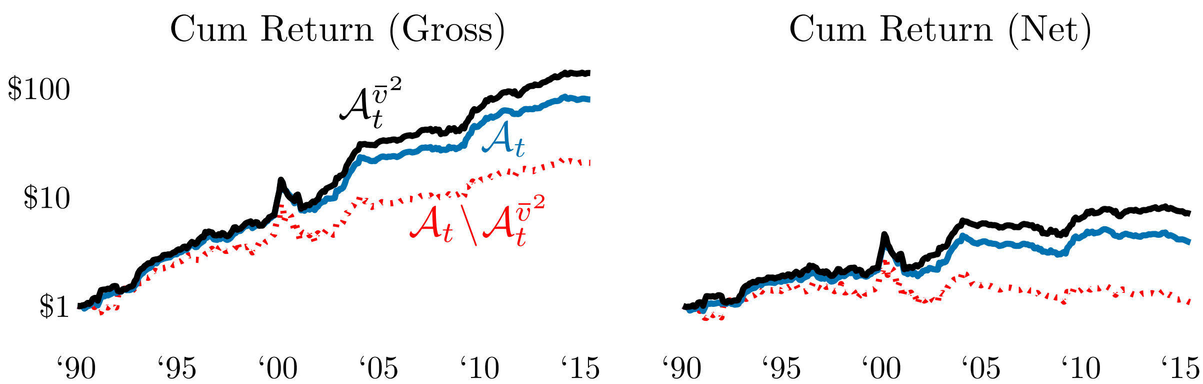

Figure C1. Cumulative Returns. Solid black line, $\mathcal{A}_t^{\bar{v}^2}$: current value of a portfolio that invested \$1 in the base-rate-adjusted strategy on January 1st, 1990. Solid blue line, $\mathcal{A}_t$: current value of a portfolio that invested \$1 in the benchmark strategy on January 1st, 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 $\bar{v}^2_{\neg i,t}$ 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, $\bar{se}^2_{i,t}$, 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:

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

In the equation above, $R_{\mathrm{Mkt},t}$ 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.[8]

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.[9]

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.

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

[9]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
	$\Pr[hiTO_s]$	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
	$\Pr[hiTO_s]$	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
	$\Pr[hiTO_s]$	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. $\Pr[hiTO_s]$: 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
$R_{\text{Mkt},t}$	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)
$R_{\text{Mkt},t}$		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, $\mathcal{A}_t \setminus \mathcal{A}_t^{\bar{v}^2}$. $R_{\text{Mkt},t}$: 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), $t_0(i) = \mathit{Jan2005}$. 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:

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

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)
$\mathbf{1}\left[\, i \in \mathcal{A}_{t_0(i)} \,\right]$	$7.70^{\star}$ (4.10)			
$\mathbf{1}\left[\, i \in \mathcal{A}_{t_0(i)}^{\bar{v}^2} \,\right]$		$10.37^{\star\star}$ (4.77)	$14.81^{\star\star\star}$ (5.33)	$15.56^{\star\star\star}$ (5.50)
$\mathbf{1}\left[\, i \in \mathcal{A}_{t_0(i)} \setminus \mathcal{A}_{t_0(i)}^{\bar{v}^2} \,\right]$				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)
$\lvert\bar{\beta}_{i,t_0(i)}\rvert$			$-1.52^{\star}$ (0.87)	$-1.49^{\star}$ (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. $\mathbf{1}[i \in \mathcal{A}_{t_0(i)}^{\bar{v}^2}]$: indicator variable for ith variable's inclusion in the base-rate-adjusted strategy in month $t_0(i)$; units of % per year. $\mathbf{1}[i \in \mathcal{A}_{t_0(i)} \setminus \mathcal{A}_{t_0(i)}^{\bar{v}^2}]$: 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: * = 10%, ** = 5%, and *** = 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,

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

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:	$\mathrm{E}[R_{s,t}]$	1.99	1.72	2.48
Net of 1.0% minimum performance threshold:	$\mathrm{E}[R_{s,t}]$	1.13	0.74	1.48
	$\mathrm{Sd}[R_{s,t}]$	6.17	4.50	4.02
	$\mathrm{Skew}[R_{s,t}]$	1.26	0.94	1.90
	$\mathrm{Kurt}[R_{s,t}]$	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. $\mathrm{E}[R_{s,t}]$: 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.

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

		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.