Out-of-Sample Alphas Post-Publication

Andrei S. Gonçalves* Johnathan A. Loudis† Richard E. Ogden‡

This Version: February 2025

Link to Most Recent Version

Abstract

Anomaly strategies generate positive and significant CAPM alphas post-publication. Existing explanations include non-market risks, trading costs, and investment frictions. This paper introduces a complementary and novel channel: when a new anomaly strategy is published, investors face uncertainty in identifying the optimal weight to allocate to the anomaly in order to achieve a positive alpha post-publication, making the strategy less appealing. Empirically, we find that the average post-publication alpha of anomaly strategies is close to zero when optimal weights are estimated out-of-sample using pre-publication data. This finding is robust across specifications, including those using empirical Bayesian shrinkage and machine learning to estimate weights. Conceptually, this suggests investors have little incentive to add a new anomaly strategy to their portfolios. While investors can generate positive out-of-sample alphas by combining multiple anomaly strategies via shrinkage methods, we show the demand from such investors is insufficient to eliminate alphas in equilibrium.

JEL Classification: G10, G11, G12, G14

Keywords: Anomalies, CAPM Alphas, Out-of-Sample Alphas

*Fisher College of Business, The Ohio State University, Columbus, OH. E-mail: goncalves.11@osu.edu.
†Mendoza College of Business, University of Notre Dame, Notre Dame, IN. E-mail: jloudis@nd.edu.
‡Rohrer College of Business, Rowan University, Glassboro, NJ. E-mail: ogdenr@rowan.edu.

Introduction

A major pursuit of asset pricing is to understand the performance of anomaly strategies (i.e., systematic strategies that generate alpha relative to the CAPM). The number of papers documenting such strategies has grown significantly over time (Harvey, Liu, and Zhu (2015)). A key stylized fact is that anomaly strategy returns and CAPM alphas decline but remain economically strong and statistically significant post-publication (McLean and Pontiff (2016) and Jensen, Kelly, and Pedersen (2023)). This fact poses a challenge to market efficiency, as investors should eliminate alphas after learning about them through the publication process.

Three common explanations for the persistence of anomaly performance are (i) anomaly strategies are exposed to risk factors beyond market risk (Fama and French (1996)), (ii) little alpha remains after trading costs (Chen and Velikov (2023)), and (iii) frictions prevent investors from eliminating alphas (Shleifer and Vishny (1997)). This paper introduces a complementary and novel channel: even unconstrained investors who avoid trading costs and focus on market risk face an estimation challenge that renders anomaly strategies less attractive. Namely, when a new anomaly strategy is published, investors do not know the optimal weights to combine it with the market index to obtain a positive alpha post-publication.

The logic underlying our main argument is as follows. The CAPM alpha of an anomaly strategy reflects the Sharpe ratio increase from optimally combining it with the market index (Gibbons, Ross, and Shanken (1989)). Standard factor regressions implicitly estimate this alpha based on ex-post optimal weights, leading to in-sample (IS) alphas (Cederburg et al. (2020)). So, investors who learn about a new published anomaly would need to know its post-publication returns to calculate the ex-post optimal weights necessary to harvest its post-publication IS alpha. Since this is infeasible, we estimate post-publication out-of-sample (OS) alphas, which are available to investors in real time. They reflect the Sharpe ratio increase over the post-publication period from combining a given anomaly with the market index using optimal weights estimated from the relevant pre-publication period.

Our main empirical result can be seen in Figure 1, which is based on 177 anomaly long-

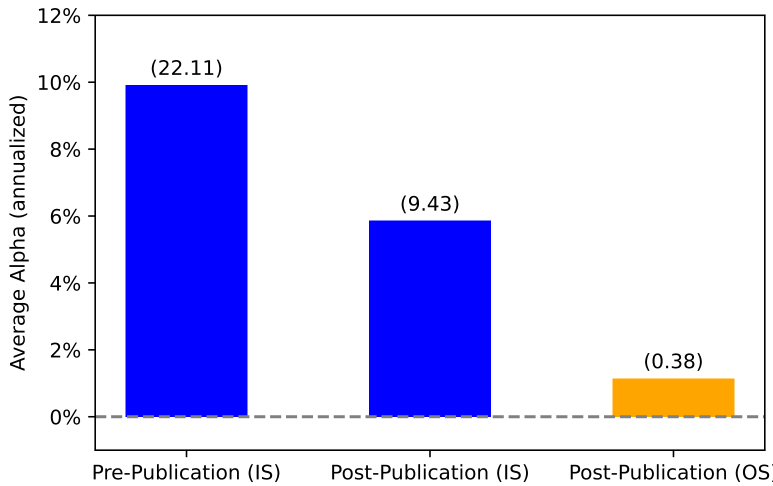


Figure 1
Average CAPM Alphas with their t-stats: In-Sample (IS) vs Out-of-Sample (OS)

The bars plot the average CAPM alpha across in-sample (IS) alphas pre- and post-publication as well as out-of-sample (OS) alphas post-publication. The bootstrap t-statistic for each average alpha is written on the top of its respective bar. The sample goes from 1926 to 2023 and is based on 177 published equity anomalies from Chen and Zimmermann (2022), with the pre-publication period ending in December of the publication year. IS alphas are based on standard CAPM factor regressions whereas OS alphas combine the Gibbons, Ross, and Shanken (1989) expression for alphas with maximum Sharpe ratio weights from the pre-publication period and realized returns from the post-publication period (Equation 6).

short decile portfolios obtained from the Chen and Zimmermann (2022) dataset. The first bar shows that, on average, anomaly strategies deliver substantial IS alphas pre-publication (consistent with the literature). The second bar shows that the average IS alpha remains positive and statistically significant over the post-publication period, albeit substantially smaller than over the pre-publication period (also in line with the literature). The third bar highlights our main result: the average post-publication OS alpha is relatively close to zero and statistically insignificant. As such, allocating capital in real time to an anomaly strategy after its publication has not been as rewarding as the post-publication IS alphas suggest.

The key patterns in Figure 1 are highly robust to empirical implementation decisions. We show that the average OS alpha is relatively close to zero under alternative approaches for selecting anomalies, different ways for defining the pre- and post-publication periods, and alternative methods for constructing the anomaly strategies. OS alphas are also similar if the out-of-sample optimal weights account for the expected anomaly decay post-publication.

While our baseline results rely on a standard frequentist estimation of pre-publication optimal weights, we also explore Bayesian shrinkage for this task. We show that Bayesian shrinkage can produce OS alphas that are similar to the post-publication IS alphas if the investor's prior alpha distribution has some dispersion that is not too narrow nor too wide (e.g., if the prior cross-sectional standard deviation for annual alphas is around 1%). However, an empirical Bayesian approach that estimates the prior on the alpha distribution over time based on the empirical distribution of alphas from previously published anomalies yields alphas that are too dispersed relative to this benchmark. Consequently, the average empirical Bayesian OS alpha is relatively close to zero, as in our baseline analysis.

We also explore machine learning methods to estimate pre-publication optimal weights. Following Davis (2024), we only consider two machine learning methods from the recent finance literature: the KNS method from Kozak, Nagel, and Santosh (2020) and the BPZ method from Bryzgalova, Pelger, and Zhu (2024). Applying these methods, we find that OS alphas remain relatively close to zero on average. While more advanced machine learning methods might perform better, it is unclear they would help answer our particular research question. Our goal is to understand why anomaly alphas persist post-publication, and thus our OS alphas (representing those available to investors) must be out-of-sample not only in terms of the data used, but also in terms of the technology used. While the KNS and BPZ methods are also subject to this limitation, they can be written as a standard LASSO regression (Bryzgalova, Pelger, and Zhu (2024)), which partially alleviates this concern.

The OS alphas of the KNS and BPZ methods do not outperform frequentist OS alphas because shrinkage and anomaly selection have little impact when combining a single anomaly with the market portfolio, as required to estimate anomaly OS alphas. However, prior literature shows that these and related methods can combine multiple anomalies to produce portfolios with Sharpe ratios well above the market Sharpe ratio out-of-sample (e.g., Kozak, Nagel, and Santosh (2020), Bryzgalova, Pelger, and Zhu (2024), and Jensen et al. (2025)). Thus, while the average OS alpha for individual anomalies is relatively close to zero, portfolios using multiple anomalies can achieve strong positive OS alphas. As such, the market may provide an incentive for investors to become "quants" who consider multiple anomalies simultaneously even if it does not provide an incentive for investors to add single anomalies to their passive position on the market portfolio (e.g., to become pure momentum traders).

We show that the incentive for investors to become quants does not alter our conclusion that uncertainty about optimal anomaly weights can generate significant post-publication IS alphas. Specifically, we model mean-variance investors acting as quants who combine multiple anomalies with the market portfolio using a BPZ-like strategy to deal with weight uncertainty. Two channels generate the persistence of post-publication IS alphas. First, as noted by Bretscher, Lewis, and Santosh (2024), an investor's relevant risk measure is the beta relative to their own portfolio. So, quants target alphas relative to the quant wealth portfolio, not the market wealth portfolio. Second, even if all investors are quants holding the market portfolio, they trade less aggressively than a mean-variance investor with Full Information Rational Expectations (FIRE), in line with Davis (2024). This is because quants apply shrinkage when trading on anomalies, leading to more moderate portfolio weights. As such, IS alphas persistent even in an equilibrium in which all investors are quants.

In summary, weight uncertainty leads to low average OS alphas on anomaly strategies, implying investors have little incentive to trade anomalies individually in real time. Weight uncertainty also implies trading anomalies jointly in real time is only profitable with some form of shrinkage, as standard in quant investing. However, investors acting as quants are not aggressive enough to fully eliminate IS alphas, which persist after anomalies are published.

Contribution to the Literature

Our key contribution is to show that the average post-publication OS alpha of anomaly strategies is relatively close to zero and statistically insignificant. To do so, we construct a new OS alpha definition based on ex ante estimable weights that investors can allocate to an anomaly strategy. Our definition is in line with the insight that maximum Sharpe ratio portfolios (and thus alpha) need to rely on weights estimated from data before the evaluation period to be out-of-sample (see Cederburg et al. (2020) and Kan, Wang, and Zheng (2024)).

McLean and Pontiff (2016) show a strong decline in anomaly performance post publication (see also Chen and Zimmermann (2023) and Chen, Lopez-Lira, and Zimmermann (2023)). However, anomalies continue to provide a positive and statistically significant average IS alpha post-publication (Jensen, Kelly, and Pedersen (2023)). In contrast, we find an average post-publication OS alpha that is close to zero economically and statistically. So, our evidence complements these papers as investors cannot achieve IS alphas in real time, only OS alphas. It also explains the persistence of IS alphas since investors are unlikely to "trade away" anomalies if anomaly strategies do not deliver alphas in real time.

The idea that IS alphas are unavailable to investors in real time is not new. For instance, Cederburg et al. (2020) use it to explain the performance of volatility-managed portfolios. Moreover, Kan, Wang, and Zheng (2024) make a related point that the ex-post maximum Sharpe ratio portfolio obtained from a factor model is not achievable to investors and Chabi-Yo, Gonçalves, and Loudis (2024) apply this concept to compare their intertemporal factor model with other models. However, none of these papers link this idea to the factor zoo to explain why anomalies continue to provide IS alphas post-publication, nor do they distinguish between pre- and post-publication data. Our contribution relative to these papers lies in making these connections and introducing an expression to calculate OS alphas.

Distinguishing between pre- and post-publication data avoids publication bias in our OS tests since the post-publication returns of anomalies are not artificially inflated by the publication process. However, our point is broader: the intercept estimate from a factor regression reflects an IS alpha, available only to hypothetical investors with FIRE. Due to weight uncertainty, real investors rely on statistical methods to choose portfolio weights so that they achieve OS alphas that are weaker than the respective IS alphas studied in the literature.[1] This connects our work to recent studies on statistical limits to arbitrage (e.g., Da, Nagel, and Xiu (2022), Davis (2024), and Baba Yara, Boyer, and Davis (2024)). Unlike standard limits

[1]Note that the concept of OS alphas we use is different from ex-ante estimates of future IS alphas (i.e., future factor regression intercepts), which are sometimes studied in the literature (e.g., Jensen, Kelly, and Pedersen (2023) and Marrow and Nagel (2024)). In particular, even if an investor (accurately) forecasts an anomaly to have a positive future intercept on a CAPM regression, this does not imply that the investor knows how much to allocate to the anomaly strategy and market portfolio to harvest the anomaly's alpha.

to arbitrage, which arise from frictions in financial markets (Gromb and Vayanos (2010)), statistical limits to arbitrage reflect statistical difficulties in the investment decision process.

Our paper has a particularly close connection to one paper in this statistical limits to arbitrage literature. Specifically, Davis (2024) shows that inelastic demand in asset pricing systems (Koijen and Yogo (2019)) arises as an equilibrium outcome when statistical arbitrageurs use out-of-sample cross-sectional predictability models to construct maximum Sharpe ratio portfolios. Since these optimal portfolios are not aggressive in trading anomalies given the observed return data, markets are inelastic. Similarly, we argue that average OS alphas are relatively close to zero, reducing investor incentives to trade anomalies individually to harvest alphas. Moreover, the demand from quant investors who combine anomalies to produce portfolios with positive OS alphas is not aggressive enough to eliminate alphas.

Finally, our work can be understood as a cross-sectional asset pricing analogue to the time-series point made by Goyal and Welch (2008) (see also Goyal, Welch, and Zafirov (2024)). In particular, Goyal and Welch (2008) show that most variables that predict time variation in the equity premium IS do not do so OS. This result is a consequence of the uncertainty in predictability parameters faced by investors trying to predict the equity premium in real time. Likewise, we show that anomalies provide a strong average IS alpha, but a close to zero average OS alpha. Moreover, in line with the parameter uncertainty point of Goyal and Welch (2008), our result is a consequence of the uncertainty investors face in the optimal weights needed to combine anomalies and with market index in order to harvest their alphas.

The rest of this paper is organized as follows. Section 1 explains the difference between IS and OS alphas and discusses our main estimation approach as well as the data we use. Section 2 provides our main results, which rely on estimating optimal pre-publication weights using a standard frequentist method. In turn Sections 3 and 4 explore Bayesian shrinkage and machine learning as alternative approaches for estimating optimal pre-publication weights. Section 5 presents a model with quant investors who consider multiple anomalies simultaneously. Section 6 concludes. The Internet Appendix provides the derivation of the expression linking alphas to Sharpe ratios as well as supplementary empirical details and results.

1 Out-of-Sample Alphas: Definition, Estimation, and Data

Subsection 1.1 defines out-of-sample (OS) alphas as the analogue to the usual in-sample (IS) alphas used in the literature. Subsection 1.2 explains how we estimate OS alphas. And Subsection 1.3 describes the data we use for the estimation.

1.1 OS Alphas: Definition

We define OS alphas focusing on the CAPM for simplicity. So, all references to "alpha" refer to CAPM alpha. However, Internet Appendix A generalizes the OS alpha definition to multifactor models.

Let $r_{m,t}$ be returns on the market portfolio in excess of the risk-free rate and $r_{a,t}$ be long-short returns on a given anomaly strategy. Also, assume that, for $t > T_0$, they have the unconditional joint distribution $(r_{m,t}, r_{a,t}) \sim \mathcal{D}ist(\mu, \Sigma)$, with the relation

$$r_{a,t} = \alpha + \beta \cdot r_{m,t} + \epsilon_t \tag{1}$$

In this case, the alpha from Equation 1 can be alternatively written as[2]

$$\alpha = sign[w \cdot \Delta] \cdot \sigma[\epsilon] \cdot \sqrt{|\mathbb{SR}[r_p]^2 - \mathbb{SR}[r_m]^2|} \tag{2}$$

where $\mathbb{SR}[x] = \mathbb{E}[x]/\sigma[x]$ is the Sharpe ratio function, $\Delta = \mathbb{SR}[r_p] - \mathbb{SR}[r_m]$, and $r_{p,t} = (1-w) \cdot r_{m,t} + w \cdot r_{a,t}$ is the maximum Sharpe ratio portfolio combining r_m and r_a so that[3]

$$\omega = \begin{pmatrix} 1-w \\ w \end{pmatrix} = \frac{\Sigma^{-1}\mu}{|1'\Sigma^{-1}\mu|} \tag{3}$$

[2]Equation 2 is a univariate version of the result in Gibbons, Ross, and Shanken (1989) linking pricing errors to mean-variance efficiency (see also the textbook treatment in Chapter 6.6 of Campbell, Lo, and MacKinlay (1997)). Internet Appendix A proves a generalization of Equation 2 to multifactor models.

[3]There are infinitely many maximum Sharpe ratio portfolios, all satisfying $\omega = \theta \cdot \Sigma^{-1}\mu$ for $\theta > 0$. The choice of θ does not affect the calculated alphas, as $\mathbb{SR}[r_p]$ and $sign(w \cdot \Delta)$ remain constant for any $\theta > 0$ (see Equation 2). We use the normalization $\theta = 1/|1^{'}\Sigma^{-1}\mu|$, ensuring the absolute sum of ω elements equals one. Normalizing by $\theta = 1/(1^{'}\Sigma^{-1}\mu)$, which forces ω elements to sum to one, would be incorrect if $1^{'}\Sigma^{-1}\mu < 0$, as it would yield the minimum Sharpe ratio portfolio. Although rare in our empirical analysis, this is possible if the global minimum variance portfolio has a negative expected excess return. In this case, the tangency portfolio lies in the inefficient frontier, and the maximum Sharpe ratio portfolio involves a negative position in the tangency portfolio and a positive position in the risk-free asset.

An investor with Full Information Rational Expectations (FIRE) knows μ and Σ, and thus w. As such, FIRE investors can achieve the alpha from Equation 2.[4] Consequently, α is an object of inherent economic interest so that we (researchers) would like to estimate it.

In this context, suppose we observe $r_{m,t}$ and $r_{a,t}$ for $t = T_0 + 1, T_0 + 2, ..., T$. We can consistently estimate α from an Ordinary Least Squares (OLS) regression on Equation 1, which is equivalent to

$$\alpha_{IS} = sign[w_{IS} \cdot \widehat{\Delta}_{IS}] \cdot \widehat{\sigma}[\epsilon] \cdot \sqrt{\left|\widehat{\mathbb{SR}}[r_{IS,p}]^2 - \widehat{\mathbb{SR}}[r_m]^2\right|} \tag{4}$$

where $\widehat{\Delta}_{IS} = \widehat{\mathbb{SR}}[r_{IS,p}] - \widehat{\mathbb{SR}}[r_m]$ and $r_{IS,p,t} = (1 - w_{IS}) \cdot r_{m,t} + w_{IS} \cdot r_{a,t}$, with

$$\omega_{IS} = \begin{pmatrix} 1 - w_{IS} \\ w_{IS} \end{pmatrix} = \frac{\Sigma_{IS}^{-1}\mu_{IS}}{|1'\Sigma_{IS}^{-1}\mu_{IS}|} \tag{5}$$

Our notation attempts to distinguish between two aspects of this estimation process. First, Sharpe ratios ($\mathbb{SR}[\cdot]$) and idiosyncratic volatility ($\sigma[\epsilon]$) need to be estimated and we use hats to denote the respective ex-post moment estimates, which use data over $t = T_0 + 1, T_0 + 2, ..., T$. Second, the weight the investor needs to allocate to the anomaly strategy in order to achieve the given alpha (w) also needs to be estimated, and we use the subscript IS to identify quantities that depend on this weight estimate (w_{IS}). The reason for this distinction is that for investors to achieve α_{IS} they need to know w_{IS} at $t = T_0$, but not the quantities denoted with hats, which they can observe ex-post at $t = T$.

Consequently, the use of data over $t = T_0 + 1, T_0 + 2, ..., T$ to obtain w_{IS} makes α_{IS} an IS quantity, not achievable to real time investors. α_{IS} is still an interesting economic object as it provides a consistent estimate for α, which is the alpha achievable to FIRE investors

[4] We should clarify that, in our terminology, the expression "to achieve the alpha" refers to forming a portfolio that combines the market with the anomaly strategy to achieve the Sharpe ratio increase embedded in the given alpha expression (in this context, Equation 2). An alternative meaning for "to achieve the alpha" would be to form the strategy $r_{a,t} - \beta \cdot r_{m,t}$, which has zero beta and an expected return equal to the given alpha. The reason why the former (and not the later) is the relevant economic definition of "to achieve the alpha" in the context of our analysis is that the goal of the mean-variance investor underlying the CAPM is not simply to invest in the $r_{a,t} - \beta \cdot r_{m,t}$ strategy, but rather to obtain the portfolio with maximum possible Sharpe ratio (and combine it with the risk-free asset given risk aversion). So, an investor who knows β still would need to estimate the optimal weights to combine $r_{a,t} - \beta \cdot r_{m,t}$ with $r_{m,t}$ in order to form the maximum Sharpe ratio portfolio.

as $T \to \infty$. However, measuring the alpha available to real time investors requires an OS estimate for w (i.e., a w estimate using data over $t \leq T_0$). For this purpose, we define

$$\alpha_{OS} \ = sign[w_{OS} \cdot \widehat{\Delta}_{OS}] \cdot \widehat{\sigma}[\epsilon] \cdot \sqrt{\left|\widehat{\mathbb{SR}}[r_{OS,p}]^2 \ - \ \widehat{\mathbb{SR}}[r_m]^2\right|} \tag{6}$$

where $\widehat{\Delta}_{OS} = \widehat{\mathbb{SR}}[r_{OS,p}] - \widehat{\mathbb{SR}}[r_m]$ and $r_{OS,p,t} = (1 - w_{OS}) \cdot r_{m,t} + w_{OS} \cdot r_{a,t}$, with

$$\omega_{OS} \ = \begin{pmatrix} 1 - w_{OS} \\ w_{OS} \end{pmatrix} \ = \ \frac{\Sigma_{OS}^{-1}\mu_{OS}}{|1'\Sigma_{OS}^{-1}\mu_{OS}|} \tag{7}$$

The only difference between α_{IS} and α_{OS} is that α_{IS} uses w_{IS} from Equation 5 (relying on IS estimates for μ and Σ, which use data over $t > T_0$) whereas α_{OS} uses w_{OS} from Equation 7 (relying on OS estimates for μ and Σ, which use data over $t \leq T_0$). Our main finding in this paper is that there is a large decline from α_{IS} to α_{OS}, with the average α_{OS} relatively close to zero and statistically insignificant.

Figure 2 provides a mean-variance diagram in excess return space that captures the essence of this empirical finding from a conceptual point of view. The red dot reflects the market portfolio (r_m) while the green dot reflects the anomaly strategy (r_a). The blue dot captures the maximum Sharpe ratio portfolio that can be formed by combining the market portfolio with the anomaly strategy, $r_p = (1 - w) \cdot r_{m,t} + w \cdot r_{a,t}$. Setting $w = w_{IS}$ maximizes the in-sample Sharpe ratio (i.e., line slope) increase from the red dot to the blue dot. As such, setting $w = w_{OS}$ leads to a lower Sharpe ratio increase whether $w_{OS} > w_{IS}$ or $w_{OS} < w_{IS}$, as can be seen by comparing the slopes of the blue and orange lines. So, the empirical question we tackle is not whether alpha decreases as we move from α_{IS} to α_{OS}, but rather by how much it decreases and how high α_{OS} is. The rest of this section details the data and estimation procedure we use to identify α_{IS} and α_{OS} in our baseline analysis.

1.2 OS Alphas: Estimation

For each anomaly, we observe returns over $t = 1, 2, ..., T_0, T_0+1, ..., T$ and we need to measure α_{IS} and α_{OS} over $t = T_0+1, T_0+2, ..., T$. We refer to $t = 1, 2, ..., T_0$ as the "estimation period" and to $t = T_0+1, T_0+2, ..., T$ as the "evaluation period", with the next subsection explaining how we select T_0 for each anomaly.

For both alphas, we estimate the terms with hats in Equations 4 and 6 (i.e., Sharpe ratios and idiosyncratic volatility) using the respective sample analogues over the evaluation period since investors do not need to know these values at $t = T_0$ to achieve α_{IS} or α_{OS} (but do need their respective weights). The other terms depend on the respective estimates for $w = \Sigma^{-1}\mu/|1'\Sigma^{-1}\mu|$, with

$$\mu = \begin{bmatrix} \mathbb{E}[r_m] \\ \mathbb{E}[r_a] \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \mathbb{V}ar[r_m] & \mathbb{C}ov[r_m, r_a] \\ \mathbb{C}ov[r_m, r_a] & \mathbb{V}ar[r_a] \end{bmatrix} \tag{8}$$

For α_{IS}, we estimate μ and Σ from their respective sample analogues over the evaluation period, μ_{IS} and Σ_{IS}. As such, each anomaly α_{IS} is mathematically equivalent to the OLS estimate of α from the factor regression in Equation 1 over the evaluation period.

For α_{OS}, we also estimate μ and Σ from their respective sample analogues, but over the estimation period. That is,

$$\mu_{OS} = \begin{bmatrix} \overline{r}_m = \frac{\sum_{t=1}^{T_0} r_{m,t}}{T_0} \\ \overline{r}_a = \frac{\sum_{t=1}^{T_0} r_{a,t}}{T_0} \end{bmatrix} \text{ and } \Sigma_{OS} = \begin{bmatrix} \overline{\sigma}_m^2 = \frac{\sum_{t=1}^{T_0}(r_{m,t}-\overline{r}_m)^2}{T_0-1} & \overline{\sigma}_{m,a} = \frac{\sum_{t=1}^{T_0}(r_{m,t}-\overline{r}_m)\cdot(r_{a,t}-\overline{r}_a)}{T_0-1} \\ \overline{\sigma}_{m,a} = \frac{\sum_{t=1}^{T_0}(r_{m,t}-\overline{r}_m)\cdot(r_{a,t}-\overline{r}_a)}{T_0-1} & \overline{\sigma}_a^2 = \frac{\sum_{t=1}^{T_0}(r_{a,t}-\overline{r}_a)^2}{T_0-1} \end{bmatrix} \tag{9}$$

so that the resulting w_{OS} is known to investors at $t = T_0$. Consequently, α_{OS} is achievable to investors who make their portfolio allocation decision at $t = T_0$.

Importantly, there is no mathematical equivalence between α_{OS} and the α from the factor regression in Equation 1. In particular, α_{OS} is not an ex-ante estimate of the factor regression intercept α. So, our definition of OS alpha differs from the concept of ex-ante alpha sometimes studied in the literature (e.g., Jensen, Kelly, and Pedersen (2023) and Marrow and Nagel (2024)). Our definition focuses on whether investors can achieve the alpha, not whether they can forecast it. The reason is that even if an investor (accurately) forecasts an anomaly to have a positive future intercept on a CAPM regression, this does not imply that the investor knows how much to allocate to the anomaly strategy and market portfolio in order to harvest the anomaly's alpha.

We refer to the α_{OS} estimate defined in this section as the "baseline" or "frequentist" OS alpha and report its results in Section 2. However, we also explore more sophisticated ways to estimate μ and Σ out-of-sample. In particular, Section 3 considers Bayesian shrinkage and Section 4 considers machine learning methods. The empirical details are reported in the respective sections.

1.3 OS Alphas: Data and Summary Statistics

Measuring α_{IS} and α_{OS} requires us to decide:

1. How to select anomaly signals for the analysis

2. How to form anomaly strategies (long-short portfolios) for the anomaly signals selected

3. How to specify the weight estimation period (i.e., $t = 1, 2, ..., T_0$) and the alpha evaluation period (i.e., $t = T_0 + 1, T_0 + 2, ..., T$)

For all three decisions, we rely on the Open Source Asset Pricing (OSAP) dataset from Chen and Zimmermann (2022).[5] In terms of decision (1), we select all anomaly signals in OSAP that are labeled as "predictors" by its signal documentation. Additionally, we require anomaly strategies to have at least 5 years of monthly returns available both before and after the publication year. In regards to decision (2), we form each anomaly strategy as the long-short decile portfolio with weights as per the original publication. The decile portfolio return data come directly from the "PredictorAltPorts_Deciles" file of OSAP, which uses stock weighting schemes (typically equal- or value-weighting) from the original papers. Finally, for decision (3), our sample is monthly and $t = T_0$ reflects December of the publication year for the given anomaly. Moreover, for each anomaly, $t = 1$ reflects the first month in the original publication sample or the first month of anomaly strategy return available in the OSAP dataset (whichever is later) and $t = T$ reflects the latest month with anomaly strategy returns

[5] We use the August 2024 version of the OSAP dataset from https://www.openassetpricing.com/.

available in the OSAP dataset (December 2023 for the vast majority of anomaly strategies).[6] Subsection 2.2 shows that our results are robust to the empirical decisions described in this paragraph.

The OSAP dataset includes 212 anomaly signals labeled as "predictors" in their respective original samples, with 179 of these signals having monthly anomaly strategy returns available.[7] Our final baseline sample contains the subset of 177 anomalies that also meet the criteria in the prior paragraph of five years with anomaly strategy return data available both before and after the publication year. The summary statistics for these 177 anomaly strategies are reported in Table 1.[8] The average publication year is 2004, but the sample includes anomalies with publication year as early as 1973 and as late as 2016. The average first year in the publication sample is 1970 while the average first year of available anomaly strategy return is 1955, highlighting that the OSAP dataset typically contains anomaly strategy returns starting even earlier than the original anomaly publication sample. The average anomaly strategy has returns for 640 months in the sample we use in our main analysis, and even the anomaly with the worst coverage has return data for 240 months in this baseline sample. Most of the return data is from the pre-publication period, with an average of 414 pre-publication months versus 226 post-publication months.

Anomaly portfolios earn positive average returns in all subperiods, but average returns are generally lower after publication, in line with the results in McLean and Pontiff (2016) (see also Chen and Zimmermann (2023), Chen, Lopez-Lira, and Zimmermann (2023), and

[6]Note that we start at the first month in the original publication sample even when anomaly strategy returns are available in the OSAP dataset for prior months (which is the case for the vast majority of anomalies). We do so to reflect the data accessible to investors at the time of publication. While data that starts at the first month of the original publication sample was available to investors, data used in the OSAP dataset today prior to the start of the original publication sample may not have been available to investors that easily at the time of publication. Nevertheless, Subsection 2.2 shows that our results are similar if we always use the first month of anomaly strategy return available in the OSAP dataset.

[7]The drop from 207 to 179 anomaly signals is because our anomaly strategy returns are based on decile portfolios, which are not available for all anomalies. For instance, some anomaly signals are discrete variables taking on less than ten possible values.

[8]Throughout the paper, we "annualize" performance metrics when reporting statistics by multiplying average returns and alphas by twelve.

Jensen, Kelly, and Pedersen (2023)). In contrast, the CAPM market risk premium ($\beta \cdot \mathbb{E}[r_m]$) is slightly negative on average and shows little change from the pre-publication to the post-publication period. Consequently, IS alphas strongly decline after publication. Nevertheless, IS alphas are still positive on average and often economically and statistically significant even after publication. Our main argument, however, is that these IS alphas are not accessible to investors, as they would need prior knowledge of the optimal weights required to combine each anomaly with the market portfolio. Since these weights are unknown, the remainder of this paper focuses on OS alphas, calculated using weights estimated from the pre-publication period, making them feasible for investors.

Note that Table 1 reports alpha t-statistics based on Newey and West (1987, 1994), which are common in the literature, as well as bootstrap t-statistics, with the bootstrap procedure described in Internet Appendix B. The two procedures yield similar distributions for the post-publication IS alpha t-statistics in Table 1. They also tend to be similar for individual anomaly strategies, as Figure IA.1 in the Internet Appendix demonstrates. In the following sections, we report only bootstrap t-statistics since Newey and West (1987, 1994) cannot be applied to obtain standard errors for our OS alphas.

2 OS Alphas with Baseline Weight Estimation

This section presents the results from our baseline (or frequentist) estimation of OS alphas. Subsection 2.1 focuses on the main results while Subsection 2.2 presents a robustness analysis with respect to our core empirical decisions.

The key result of this section is that while IS alphas are, on average, large and positive over the post-publication period, OS alphas are, on average, relatively close to zero. Moreover, this result is driven by a decline in the (ex-post) optimal anomaly weight from the pre-publication period to the post-publication period.

2.1 OS Alphas: Main Results

Table 2 provides our main results. The first three rows show the distribution of pre-publication IS alphas, with their respective t-statistics and optimal weights (i.e., maximum Sharpe ratio weights). The most interesting observation from these rows is that the median optimal weight underlying the pre-publication IS alphas is 0.62 so that an investor would typically need to allocate more than 50% of his portfolio to a given anomaly to achieve its pre-publication IS alpha. The next three rows show the distribution of IS alphas post-publication, again with their respective t-statistics and optimal weights. In stark contrast to the pre-publication period, the median optimal weight is 0.36 so that there is a large decline in how much investors should allocate to a given anomaly in comparison to the pre-publication period. While not reported in the table, we also find that the rank correlation (known as Spearman correlation) between the pre-publication optimal weights and post-publication optimal weights is only 0.39.[9] So, an anomaly receiving a relatively high optimal weight in the pre-publication period is not a strong indicator that it should receive a relatively high weight over the post publication period.[10]

Our core point is that, at the time of publication, a typical investor does not have a reliable way to know the optimal weight to hold on a given anomaly strategy over the post-publication period. As a consequence, the OS alphas, which apply the pre-publication optimal weights to generate each strategy over the post-publication period, are relatively close to zero on average (see the next two rows of Table 2). OS alphas for individual anomalies can be large, but they are of relatively similar magnitude on both sides of the OS alpha distribution. For instance, the 10% and 90% quantiles of the OS alpha distribution are -8.8% and 12.8%,

[9] The usual linear correlation (i.e., the Pearson correlation) is even lower at -0.04. However, this low correlation is due to extreme weights (typically in the post-publication weights). If we bound weights to be between 0 and 1, then the linear correlation becomes 0.41, which is similar to the rank correlation.

[10] One may conjecture that an investor applying shrinkage to the pre-publication optimal weight would reach a weight much closer to the ex-post optimal weight post-publication. Sections 3 and 4 show that this is not the case using empirical Bayesian and machine learning methods. The investor's optimal weight would be substantially reduced if the investor considered shrinkage and multiple anomalies jointly, but Section 5 shows that in this case the weights are not aggressive enough to fully eliminate alphas in equilibrium.

respectively. Moreover, for most anomalies, these OS alphas are statistically insignificant.[11]

Figures 3(a) and 3(b) plot the density functions for the post-publication IS alphas and OS alphas, as well as their respective t-statistics. While only a small fraction of the anomaly strategies have a negative IS alpha (12.4%), almost half of the anomaly strategies have a negative OS alpha (47.5%). Similarly, while 49.6% of the IS alphas have a t-stat above 1.64, only 10.7% of the OS alphas do. In terms of average alphas across all anomalies, their t-stats are 9.43 for IS alphas and 0.38 for OS alphas (as shown in Figure 1). So, the average IS alpha is statistically significant while the average OS alpha is not.

Figures 3(c) and 3(d) (as well as the last two rows of Table 2) provide information on the distribution of IS alphas minus OS alphas (which we refer to as IS-OS alphas) and their respective t-statistics. By construction, all IS-OS alphas are positive so that the risk-adjusted return investors face in reality is worse than IS alphas suggest for all anomaly strategies (as highlighted in the diagram of Figure 2 and related commentary in Section 1.1). Given the level of uncertainty in alphas, only 20.3% of the IS-OS alphas have a t-stat above 1.64. So, at a 5% significance level, we can only reject the null of a zero IS-OS alpha in favor of a positive IS-OS alpha for 20.3% of the anomaly strategies we study. However, the p-value for the average IS-OS alpha is 6.18%, so we can reject the hypothesis that the average IS-OS alpha is zero at a 10% level and are close to reject the same hypothesis at a 5% level.[12]

The results outlined above summarize our key findings. The remainder of the paper examines the robustness of these results across various empirical choices and then considers investors who trade on multiple anomalies jointly in Section 5. Some of these empirical choices are related to implementation (e.g., whether to restrict analysis to anomalies available throughout the entire sample period), while others have direct economic implications (e.g., whether to replace pre-publication optimal weights with Bayesian weights, allowing investors to learn from previously published alphas). In both cases, we focus on OS alphas

[11] We find that some anomalies have economically and statistically significant OS alphas. However, investors need to identify these anomalies ex-ante to harvest their alphas. Exploring whether investors can do that in real time is an interesting avenue for future research.

[12] Note that, by construction, IS-OS alpha ≥ 0 for each anomaly strategy, and thus the alternative hypothesis in the statistical test is "average IS-OS alpha > 0" and not "average IS-OS alpha $\neq 0$".

rather than IS-OS alphas, as OS alphas are more relevant from an economic standpoint. Investors do not choose between the ex-post optimal strategy and the ex-ante optimal strategy (with tradeoff captured by IS-OS alphas). Instead, they face the decision to invest solely in the market or in the ex-ante optimal strategy that combines the anomaly with the market (with tradeoff captured by OS alphas). Therefore, from an asset pricing perspective, the distribution of OS alphas is more relevant than that of IS-OS alphas.

2.2 OS Alphas: Robustness Analysis

In this subsection, we show that our main results are robust to the core implementation decisions underlying our OS alpha estimates. When doing so, we hold fixed one key economic decision: the frequentist approach to estimate the optimal weights used in the calculation of post-publication OS alphas. The next two sections explore alternative methods to estimate these optimal weights.

We consider sixteen alternative specifications relative to our baseline empirical procedure. The barplot in Figure 4 reports the resulting average alphas as well as the respective interquartile ranges of the distributions of alphas across anomalies, with the lower and upper bounds indicating the 25th and 75th percentiles. We sort specifications based on the magnitude of their average IS alphas pre-publication and color bars based on whether the average alpha is statistically significant at the 5% level (blue) or not (orange). As is clear from the figure, our key finding (summarized in Figure 1) is robust to our empirical decisions. Specifically, the average post-publication IS alpha is economically and statistically significant in all specifications, albeit much lower than its respective average pre-publication IS alpha. In contrast, the average post-publication OS alpha is relatively close to zero and statistically insignificant in all specifications. Moreover, the interquartile range of alphas only contains positive alpha values for almost all specifications when looking at IS alphas, but always contains both positive and negative alphas when looking at OS alphas. Below, we detail the specifications considered in Figure 4.

The first group of specifications modifies how we select anomalies for the analysis. The objective is to deal with the concern that we may be relying on relatively weak anomalies either because anomalies are weak to start with or because the sample available for those anomalies is short. The specification "Clear Predictors" keeps only the 138 anomalies classified as clear predictors in the OSAP dataset (i.e., anomalies which are clearly treated as predictors in their original publications). The specification "IS Alpha t-stat $\geq$ 3" keeps only the 120 anomalies with an IS alpha t-stat above 3.0 in the pre-publication period (based on Newey and West (1987, 1994)). Specification "10 Years Pre and Post Publication" keeps only the 162 anomalies that have at least 10 years in both the pre-publication and the post-publication periods. And specification "Returns Available since 1963" keeps only the 115 anomalies that have anomaly strategy returns available in the OSAP dataset over the entire period from July 1963 to December 2023 (and uses returns starting in July 1963 for these 115 anomalies even in cases in which the original publication sample starts after July 1963).

The second group of specifications modifies how we form the estimation period (used to estimate weights for OS alphas) and the evaluation period (used to calculate the OS alphas). Specification "Pre-Publication Period $\geq$ 1992" starts the pre-publication period in January 1992 to avoid estimating pre-publication weights with data that is too stale relative to the post-publication period (in this case, we keep only the 156 anomalies published at or after 1996). Specification "All Returns Available in OSAP" starts the pre-publication period in the first month of anomaly strategy return available in the OSAP dataset, which is typically earlier than the first month in the original publication sample used in our baseline analysis. Specification "T_0 from Last Year of Publication Sample" defines the estimation and evaluation periods based on the last year of the publication sample as opposed to the publication year. Specification "Account for Alpha Shock in Transition" ends the pre-publication period one year before the end of the publication sample and starts the post-publication period three years after the publication date (all months in between are dropped from the analysis). This analysis accounts for the evidence in Pénasse (2022) that realized returns are abnormally high in the period around the end of the publication sample as well as right after the publication date (this happens because true alpha declines, inducing an unexpected negative shock in discount rates). And specification "Anomalies Even Before Publication" ignores actual publication dates. Specifically, for each month T_0 (from five years after the anomaly's first month in OSAP to five years before the anomaly's last month in OSAP), we estimate the alphas of all 177 anomaly strategies in our analysis over $t = 1, ..., T_0$ and treat the ones that have a t-statistic above 2.0 as new anomalies "published" at T_0, with their "post-publication" period covering $t = T_0, T_0 + 1, ..., T$.[13] As such, the same anomaly can "be published" at multiple months in this analysis and these months can precede the true publication date of the anomaly. The OS alphas in this specification are subject to publication bias since the "post-publication period" here can cover the original publication sample (which would lead to anomaly returns that are artificially inflated by the publication process tending to select strong anomalies based on in-sample estimates). However, this analysis is useful in illustrating that our core point that OS alphas are substantially lower than IS alphas (from factor regressions) is not restricted to the true post-publication period of anomaly strategies, which we use in our baseline results to avoid the publication bias issue described above.

The third group of specifications modifies how we form the long-short portfolio returns used to define anomaly strategies. The "Equal-Weighted Deciles" and "Value-Weighted-Weighted Deciles" specifications use equal-weighted and value-weighted deciles, respectively, instead of relying on the same weights used in the original publication (the value-weighted deciles are available in the OSAP dataset whereas we construct our own equal-weighted deciles based on OSAP signals because OSAP does not provide these portfolios). The "Original Anomaly Portfolios" specification constructs each long-short portfolio using the exact same procedure as in the original paper instead of using decile portfolios (these anomaly strategy returns are also available directly in the OSAP dataset). And specifications "Dynamic Portfolio Weights, Expanding" and "Dynamic Portfolio Weights, Rolling" follow our baseline frequentist weight estimation procedure at $t = T_0$, but update the weights each

[13]In this "Anomalies Even Before Publication" specification, $t = 1$ reflects the first month of anomaly strategy return available in the OSAP dataset (not the first month in the original publication sample). The reason is that this specification implicitly assumes investors evaluate anomaly signals even before their publication date. So, the original publication sample is irrelevant in the context of this specification.

month from $t = T_0 + 1, ..., T - 1$ to construct $r_{p,t+1}$ in the subsequent month. We consider both an expanding window as well as a 5-year rolling window for the weight estimation.

The final group of specifications allows investors to incorporate anomaly performance decay in their μ estimates. In particular, they replace the second element of μ_{OS} in Equation 9 with $\overline{r}_a^{decay}$. In specification "Account for Mu Decay", we use $\overline{r}_a^{decay} = \theta \cdot \overline{r}_a$, where $0 \leq \theta \leq 1$ so that the θ parameter reflects investors' estimate for the decay in average reutrns after publication (the baseline analysis effectively imposes $\theta = 1$). The value used for θ varies by anomaly based on the information investors have on all other anomalies that have been published as of the time of the given anomaly publication. Specifically, we estimate θ for a given anomaly (let's call it anomaly a) as follows. We start by identifying all anomalies that have been published (with at least five years of returns post-publication) by the end of the pre-publication period of anomaly a. Then, for each of these anomalies, we calculate the ratio between their respective average returns over their own post-publication and pre-publication periods (all within the pre-publication period of anomaly a). Finally, we estimate the θ for anomaly a using the average of these ratios across all of these other anomalies while bounding it to be between zero and one. In specification "Account for Alpha Decay", we use $\overline{r}_a^{decay} = \theta \cdot \alpha_a + \beta_a \cdot \overline{r}_m$, where again $0 \leq \theta \leq 1$ so that the θ parameter reflects investors' estimate for the decay in α after publication (with α_a and β_a estimated over the pre-publication period for anomaly a). To obtain θ, we again start by identifying all anomalies that have been published (with at least five years of returns post-publication) by the end of the pre-publication period of anomaly a. Then, for each of these anomalies, we calculate $\theta = \alpha_{Post}/\alpha_{Pre}$, where α_{Post} and α_{Pre} are estimated, respectively, over the post-publication and pre-publication periods for the given anomaly (again, within the pre-publication period of anomaly a). Finally, as before, we estimate the θ for anomaly a using the average of these ratios across all of these other anomalies while bounding it to be between zero and one.

3 OS Alphas with Bayesian Weight Estimation

The previous section indicates that while IS alphas remain high in the post-publication period, OS alphas, on average, are relatively close to zero. This outcome stems from the fact that the ex-post optimal anomaly weights in the pre-publication period are significantly higher than those in the post-publication period. So, it is reasonable to expect that a Bayesian investor might achieve better OS alphas. This is because a Bayesian investor would naturally adopt more conservative anomaly weights than a frequentist would to deal with the ex ante uncertainty in the optimal weights needed to achieve a given anomaly's alpha. Accordingly, this section examines OS alphas obtained using Bayesian shrinkage. Subsection 3.1 outlines the Bayesian approach and Subsection 3.2 presents the empirical findings.

The key result of this section is that Bayesian shrinkage applied in real time to estimate optimal weights does not change our main conclusions regarding OS alphas. We start by showing that Bayesian shrinkage can produce OS alphas that are similar to the post-publication IS alphas if the investor's prior alpha distribution has some dispersion that is not too narrow nor too wide. However, a Bayesian investor would not know ex ante the optimal prior distribution that would generate the highest OS alphas. So, we implement an empirical Bayesian approach that estimates the prior on the alpha distribution over time based on the empirical distribution of alphas from previously published anomalies using data available in real time. This approach yields relatively wide prior alpha distributions over time, implying little alpha shrinkage. Consequently, the average empirical Bayesian OS alpha is relatively close to zero, as in our baseline analysis.

3.1 Bayesian Weight Estimation

To apply Bayesian shrinkage when estimating optimal anomaly weights, we consider a simple 1-factor index model, $r_{a,t} = \alpha + \beta \cdot r_{m,t} + \epsilon_t$, which implies the μ and Σ expressions in Equation

8 can be written as

$$\mu = \begin{bmatrix} \mathbb{E}[r_m] \\ \alpha + \beta \cdot \mathbb{E}[r_m] \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \mathbb{V}ar[r_m] & \beta \cdot \mathbb{V}ar[r_m] \\ \beta \cdot \mathbb{V}ar[r_m] & \mathbb{V}ar[r_a] \end{bmatrix} \tag{10}$$

Our Bayesian weight estimation (over the pre-publication period) uses OLS to estimate β but shrinks α towards zero (i.e., it uses the prior that anomalies have no alpha), effectively shrinking the anomaly weight towards zero.[14] Specifically, we use

$$\alpha_{BAY} = \left(\frac{1/\sigma^2_{OLS,\alpha}}{1/\sigma^2_\alpha + 1/\sigma^2_{OLS,\alpha}} \right) \cdot \alpha_{OLS} \tag{11}$$

where $\sigma_{OLS,\alpha}$ reflects the standard error of α_{OLS} (estimated by Newey and West (1987, 1994)) whereas σ_α reflects the prior uncertainty about alpha. In turn, we obtain

$$\mu_{BAY} = \begin{bmatrix} \overline{r}_m \\ \alpha_{BAY} + \beta_{OLS} \cdot \overline{r}_m \end{bmatrix} \quad \text{and} \quad \Sigma_{BAY} = \begin{bmatrix} \overline{\sigma}^2_m & \beta_{OLS} \cdot \overline{\sigma}^2_m \\ \beta_{OLS} \cdot \overline{\sigma}^2_m & \overline{\sigma}^2_a \end{bmatrix} \tag{12}$$

and get the (Bayesian) OS anomaly weight by using μ_{BAY} and Σ_{BAY} as a replacement for μ_{OS} and Σ_{OS} in Equation 7.

Our empirical analysis explores three arbitrary priors (i.e., three arbitrary values of σ_α) to illustrate the quantitative effect of Bayesian shrinkage. However, an investor would need to select σ_α for each anomaly using its respective pre-publication sample. We account for that by using an empirical Bayesian estimator as a fourth specification for σ_α. Consider an anomaly with pre-publication sample ending in month T_0. We start by collecting all anomalies that were published up to T_0. We then estimate α_{OLS} for each of these anomalies using the sample ending in T_0. Finally, we obtain σ_α as the standard deviation of α_{OLS} across these anomalies.

[14] We also consider a simple strategy that allocates 50% of the portfolio to the market index and 50% to the anomaly strategy. This approach can be thought of as a generic Bayesian method designed to combat estimation error and often performs better than more complex methods in mean-variance applications with limited time series (see DeMiguel, Garlappi, and Uppal (2009)). However, in the context of our analysis, it leads to OS alpha results that are similar to our baseline OS alpha results (reported in Section 2). So, to conserve space, we do not tabulate the results from this analysis.

We repeat this process for each anomaly so that σ_α varies across anomalies based on the sequence of published anomalies.[15]

3.2 Bayesian OS Alpha Results

Table 3 and Figure 5 summarize our Bayesian OS alpha results. Panel A and Figure 5(a) consider an extremely informative prior ($\sigma_\alpha = 0.1\%$).[16] In this case, the Bayesian optimal weights are very close to zero. The mean weight is 0.01 and even the 90% quantile of the weight distribution is only 0.04. Consequently, alphas are very close to zero. In particular, the mean alpha is 0.6% and no alpha is statistically different from zero. Panel B and Figure 5(b) consider a prior that is still highly informative, but a little less extreme ($\sigma_\alpha = 1.0\%$). In this case, the Bayesian optimal weights are close to the optimal post-publication IS weights. In particular, the mean weight is 0.33 and the median weight is 0.29. Consequently, the OS alpha distribution gets closer to the IS alpha distribution. Panel C and Figure 5(b) consider a moderately informative prior ($\sigma_\alpha = 5.0\%$). In this case, the overall results are very similar to the baseline OS alpha results. In particular, the Bayesian optimal average and median weights are similar to the pre-publication optimal weights. Consequently, the average and median OS alphas (2.0% and 2.3%, respectively) are relatively close to the average and

[15] In untabulated results (that are similar to the reported results), we also consider a Bayesian estimator that shrinks both α and β toward zero (since each anomaly portfolio is a long-short equity portfolio, it is sensible to use $\beta = 0$ as the benchmark). Specifically, write the regression in Equation 1 as $y = X\theta + \epsilon$, where $\theta' = [\alpha\ \beta]$. Then, the Bayesian estimator for θ is given by

$$\widehat{\theta}_{BAY} = \left(V_o^{-1} + V_{OLS}^{-1}\right)^{-1} \cdot \left[V_o^{-1}\theta_0 + V_{OLS}^{-1}\theta_{OLS}\right]$$

where $\theta_0' = [\alpha_0\ \beta_0] = [0\ 0]$, $\theta_{OLS} = (X'X)^{-1}(X'y)$ is the OLS estimate, and V_{OLS} is the covariance matrix of θ_{OLS} (estimated by Newey and West (1987, 1994)). V_o reflects the prior uncertainty about θ, which is a symmetric matrix with diagonal elements σ_α^2 and σ_β^2, and off-diagonal element $\sigma_{\alpha,\beta}$. Similar to our main Bayesian analysis, we use an empirical Bayesian approach to estimate σ_α^2, σ_β^2, and $\sigma_{\alpha,\beta}$. Consider an anomaly with pre-publication sample ending in month T_0. We start by collecting all anomalies that were published up to T_0. We then estimate θ_{OLS} for each of these anomalies using the sample ending in T_0. Finally, we obtain σ_α and σ_β as the standard deviations of α_{OLS} and β_{OLS} across these anomalies, and $\sigma_{\alpha,\beta}$ as the sample covariance between α_{OLS} and β_{OLS} across these anomalies. We repeat this process for each anomaly so that σ_α^2, σ_β^2, and $\sigma_{\alpha,\beta}$ vary across anomalies based on the sequence of published anomalies.

[16] As with alpha values, we state σ_α in annualized alpha units. For instance, our annual $\sigma_\alpha = 0.1\%$ corresponds to a monthly $\sigma_\alpha = 0.1\%/12$.

median OS alphas obtained in our baseline analysis (1.1% and 1.6%, respectively).

Clearly, whether a Bayesian investor would experience a distribution of OS alphas that is similar to that of IS alphas or not depends on the prior used. In reality, a Bayesian investor would not know the correct prior, but could attempt to infer it from the data on published anomalies. Our empirical Bayesian approach described in the previous subsection captures this idea, with results provided in Panel D of Table 3 and Figure 5(d). The key finding is that the empirical Bayesian approach leads to results that are similar to the moderately informative prior from Panel C. Specifically, the Bayesian optimal average and median weights are similar to the pre-publication optimal weights. As such, the average and median of OS alphas (1.8% and 1.9%, respectively) are relatively close to the average and median OS alphas obtained with the moderately informative prior (2.0% and 2.3%, respectively). The reason is that the average σ_α from the empirical Bayesian approach is 6.3%, which is similar to the $\sigma_\alpha = 5.0\%$ used in the moderately informative prior.[17] Consequently, we conclude that the distribution of OS alphas available to an empirical Bayesian investor is similar to the distribution of OS alphas available to a frequentist investor.

4 OS Alphas with Machine Learning Weight Estimation

While the prior section captures how a Bayesian investor would estimate optimal anomaly weights at the end of their respective pre-publication periods, it has the limitation that the prior distribution of alphas needs to be estimated each year using the available anomalies data. So, a data-driven machine learning approach addressing the uncertainty in identifying optimal anomaly weights could produce higher OS alphas than the Bayesian shrinkage approach from the previous section. Thus, in this section, we explore OS alphas using standard machine learning methods to estimate OS weights. Subsection 4.1 describes the two

[17]One may conjecture that the high average σ_α in the empirical Bayesian method is driven by early anomalies, when the distribution of anomalies was not wide enough for the empirical Bayesian procedure to reasonably estimate σ_α. However, this is not the case. In particular, Figure IA.2 in the Internet Appendix shows that the empirical Bayesian estimates for σ_α are actually lower in the early periods, when less anomaly strategies were present. In the more recent years, they are stable at around $\sigma_\alpha = 7.0\%$, which is even higher than the average σ_α in the empirical Bayesian method.

machine-learning methods we explore and Subsection 4.2 presents the empirical results.

The key result of this section is that applying machine learning methods developed in the recent finance literature does not change our main conclusions regarding OS alphas. Specifically, following Davis (2024), we explore two machine learning methods to compute portfolio weights (the KNS method from Kozak, Nagel, and Santosh (2020) and the BPZ method from Bryzgalova, Pelger, and Zhu (2024)).[18] We find that the average OS alpha across anomalies remains relatively close to zero when we apply these methods. As such, allocating capital in real time to an anomaly strategy after its publication has not been as rewarding as IS alphas suggest even if one uses recent machine learning methods to estimate how much to allocate to anomalies post-publication.

4.1 Machine Learning Weight Estimation

The KNS and BPZ methods we explore are cast in the language of Stochastic Discount Factors (SDFs) since estimating the maximum Sharpe ratio weights of a portfolio is equivalent to estimating the coefficients of a linear SDF. As such, we introduce new notation. Let $f_t = [r_m, r_a]$ be the vector containing the excess returns on the market index and the given anomaly strategy. With this f definition, we have $\mathbb{E}[f] = \mu$ and $\mathbb{V}ar[f] = \Sigma$. Then, consider an SDF given by $M_t = 1 - b^{'}(f_t - \mu)$ so that the asset pricing equation $\mathbb{E}[M_t \cdot f_t] = 0$ yields $b = \Sigma^{-1}\mu$. It is easy to see that the optimal weights in Equation 3 are proportional to b (i.e., $\omega = b/|1^{'}b|$). Moreover, this b can be equivalently written as the solution to the following

[18] Although it may be possible for one to design a machine learning method that performs better than the ones we explore (e.g., by tailoring the machine learning method to recognize the nature of pre-publication vs post-publication data), it is not obvious that such a more advanced method would help us better understand the alphas available to investors in real time. The reason is that typical investors are unlikely to have had knowledge of (or access to) sophisticated machine learning methods historically. Given that our goal is to measure the alphas that anomalies provide to investors after their publication, our OS alphas should be out-of-sample not only in terms of the data used, but also in terms of the technology used. While the KNS and BPZ methods are also subject to this limitation, they can be written as a standard LASSO regression (Bryzgalova, Pelger, and Zhu (2024)), which partially alleviates this concern since LASSO was introduced in Tibshirani (1996), ten years before the median publication year of anomalies in our sample.

optimization problem:

$$b = \underset{\{b\}}{argmin} \quad (\mu - \Sigma b)' \Sigma^{-1} (\mu - \Sigma b) \tag{13}$$

The two machine learning methods we explore rely on generalizations of Equation 13. The first is the KNS method, which uses $\omega_{KNS} = b_{KNS}/|1' b_{KNS}|$ with

$$b_{KNS} = \underset{\{b\}}{argmin} \quad (\mu - \Sigma b)' \Sigma^{-1} (\mu - \Sigma b) \ + \ \lambda_1 \cdot ||b||_1 \ + \ \lambda_2 \cdot ||b||_2 \tag{14}$$

while the second is the BPZ method, which uses $\omega_{BPZ} = b_{BPZ}/|1' b_{BPZ}|$ with

$$b_{BPZ} = \underset{\{b\}}{argmin} \quad ((\mu + \lambda_0 \cdot 1) - \Sigma b)' \Sigma^{-1} ((\mu + \lambda_0 \cdot 1) - \Sigma b) \ + \ \lambda_1 \cdot ||b||_1 \ + \ \lambda_2 \cdot ||b||_2 \tag{15}$$

These methods are regularization techniques and can be equivalently written as a LASSO or LARS regression for computational efficiency (see the simplified discussion in Davis (2024)). For both methods and for each anomaly strategy, we obtain μ and Σ over the respective pre-publication period using the μ_{OS} and Σ_{OS} estimates from Subsection 1.2. We then follow the prior asset pricing literature (e.g., Gu, Kelly, and Xiu (2020), Bryzgalova, Pelger, and Zhu (2024), and Davis (2024)) in using a 4-fold cross validation design to select the hyperparameters (λ_1 and λ_2 in the case of b_{KNS} and λ_0, λ_1, and λ_2 in the case of b_{BPZ}).[19] Finally, we combine the μ_{OS} and Σ_{OS} values and the selected hyperparameters to obtain $\omega_{KNS} = b_{KNS}/|1' b_{KNS}|$ and $\omega_{BPZ} = b_{BPZ}/|1' b_{BPZ}|$ (from Equations 14 and 15) and form the respective portfolios that are evaluated over the post-publication period.

To ensure we have enough data for the 4-fold design of the prior paragraph, we restrict our machine learning analysis to the 151 anomalies that have portfolio return data for at least 20 years in the pre-publication period (so that each fold has a minimum of 5 years of

[19]Consider the b_{KNS} estimation for the momentum anomaly. We start by creating a grid for the hyperparameters (λ_1 and λ_2) and splitting the momentum pre-publication period into four contiguous samples of equal size (known as folds). Then, for each hyperparameter gridpoint, we select one fold, obtain $\omega_{KNS} = b_{KNS}/|1' b_{KNS}|$ (from Equation 14) over the three remaining folds (using μ_{OS} and Σ_{OS} estimated over the respective set of three folds), and calculate the Sharpe ratio of the resulting portfolio over the selected fold. We repeat this process four times (one for each selection of fold) and obtain the average Sharpe ratio (across the four selected folds) for each hyperparameter gridpoint. Finally, we choose the hyperparameter gridpoint with the highest average Sharpe ratio. The process is analogous for b_{BPZ} and other anomalies.

data). The next subsection provides results for these 151 anomalies using the two machine learning weight estimation methods we consider as well as our baseline weight estimation method for comparison.

4.2 Machine Learning OS Alpha Results

Table 4 provides our machine learning OS alpha results. Panel A uses our regular weight estimation but applied to the 151 anomalies used in this machine learning analysis. The results are similar to what we find in our baseline 177 anomalies, except that OS alphas tend to be a little lower in this set of anomalies (e.g., the average alpha here is 0.7% in comparison to 1.1% in our baseline analysis). Panel B provides the results for the KNS method. The KNS optimal weights are similar to the optimal weights from the frequentist method, with a mean of 0.61 and a median of 0.63 (in comparison to a mean of 0.63 and a median of 0.64 from Panel A). Consequently, the KNS OS alpha distribution is similar to the frequentist OS alpha distribution, with an average OS alpha of 0.5% and a median OS alpha of 0.0%. Panel C provides the results for the BPZ method. The BPZ optimal weights are also similar to the frequentist optimal weights, with a mean of 0.62 and a median of 0.62. As such, the BPZ OS alpha distribution also resembles the frequentist OS alpha distribution. In particular, the average OS alpha is 0.4% and the median OS alpha is 1.0%.

The fact that KNS and BPZ do not improve upon our baseline frequentist weight estimation is perhaps not that surprising given our setting. Specifically, these methods are regularization techniques designed to create an SDF (or maximum Sharpe ratio portfolio) out of multiple tradable factors. In our setting, we just have two tradable factors (the market and the given anomaly strategy). Therefore, these methods provide little benefit. We consider investors who explore multiple anomaly strategies jointly in the next section.

5 Portfolios with Multiple Anomalies Simultaneously

The prior sections focus on one anomaly at a time by estimating OS alphas of anomaly strategies individually. This approach tests whether investors who currently hold the market portfolio (i.e., CAPM investors) have an incentive to deviate by adding an anomaly strategy to their portfolio. However, it is possible that the OS alpha of a portfolio that combines multiple anomaly strategies is positive and strong even if this is not the case for the average OS alpha of individual anomalies.[20] The reason is that one may be able to improve on the out-of-sample weight estimation by applying shrinkage and anomaly selection methods, which tend to perform better with multiple anomalies considered jointly. In fact, the prior literature shows that KNS, BPZ, and other related methods can combine multiple anomalies to produce portfolios with Sharpe ratios well above the market Sharpe ratio out-of-sample (e.g., Kozak, Nagel, and Santosh (2020), Bryzgalova, Pelger, and Zhu (2024), and Jensen et al. (2025)). Thus, while the average OS alpha for individual anomalies is relatively close to zero, portfolios using multiple anomalies can achieve strong positive OS alphas. As such, investors may have an incentive to become "quants" who consider multiple anomalies simultaneously to identify optimal weights even though our empirical results indicate that they do not have a clear incentive to add individual anomalies to their passive position on the market portfolio.

In this section, we show that the incentive for investors to become quants does not alter our conclusion that weight uncertainty contributes to the persistence of IS alphas post-publication. In particular, since the results from prior sections show that investors do not have an incentive to trade on individual anomalies, we build a model in which (a subset of) investors act as mean-variance quants who simultaneously trade on multiple anomalies. Importantly, weight uncertainty leads these quants to apply shrinkage to their portfolio weights. Consequently, we show that their demand is not aggressive enough to eliminate IS alphas in equilibrium. The model description and results are below.

[20] Note that there is no difference between the IS alpha of a portfolio and the respective weighted average of IS alphas of the underlying individual anomaly strategies. However, this is not the case for OS alphas given that weights need to be estimated out-of-sample.

There is a representative quant investor (q), who reflects one class of investors. The quant investor can allocate capital to the (zero net supply) risk-free asset with gross return R_f and to N risky assets with gross return vector R. The quant investor has mean-variance preferences so that he forms his wealth portfolio, $R_{q,t+1} = \omega^{'}_{q,t}(R_{f,t+1}, R_{t+1})$, by solving

$$\omega_{q,t} \;=\; argmax \quad \mathbb{E}_{q,t}[R_{q,t+1}] \;-\; 0.5 \cdot \gamma \cdot \mathbb{V}ar_{q,t}[R_{q,t+1}] \qquad s.t. \;\; 1^{'}_{N}\omega_{q,t} = 1 \tag{16}$$

Letting $r_{t+1} = R_{t+1} - 1_N \cdot R_{f,t+1}$ be excess returns relative to the risk-free asset for the N risky assets, we can alternatively write the quant wealth portfolio as $R_{q,t+1} = R_{f,t+1} + \varpi^{'}_{q,t} r_{t+1}$ where $\omega_{q,t} = [1 - 1^{'}_{N}\varpi_{q,t} \,,\, \varpi_{q,t}]$. Using this structure, the quant investor portfolio allocation problem can be alternatively written as

$$\varpi_{q,t} \;=\; argmax \quad \mathbb{E}_{q,t}[\varpi^{'}_{q,t} r_{t+1}] - 0.5 \cdot \gamma \cdot \mathbb{V}ar_{q,t}[\varpi^{'}_{q,t} r_{t+1}] \tag{17}$$

Letting $\mu_{q,t} = \mathbb{E}_{q,t}[r]$ and $\Sigma_{q,t} = \mathbb{V}ar_{q,t}[r]$, the solution to the problem in Equation 17 is the standard mean-variance formula

$$\varpi_{q,t} \;=\; (1/\gamma) \cdot \Sigma^{-1}_{q,t}\mu_{q,t} \tag{18}$$

Let the data generating process be given by $r_{t+1} \sim \mathcal{N}(\mu_t, \Sigma_t)$. Moreover, let $\widehat{\mu}_t$ and $\widehat{\Sigma}_t$ represent maximum likelihood estimates of μ_t and Σ_t given all public information available in the market at time t. We assume the quant investor acts according to the beliefs $\mu_{q,t} = \widehat{\mu}_t - \lambda_\mu \cdot 1_N$ and $\Sigma_{q,t} = \widehat{\Sigma}_t + \lambda_\Sigma \cdot \mathrm{I}_N$. That is, even though the quant investor has access to all publicly available information (and thus to the maximum likelihood estimates $\widehat{\mu}_t$ and $\widehat{\Sigma}_t$), he does not directly use $\mu_{q,t} = \widehat{\mu}_t$ and $\Sigma_{q,t} = \widehat{\Sigma}_t$. Rather, he shifts the maximum likelihood estimates towards some benchmark values (zero for μ_t and an identity matrix for Σ_t).[21] As shown in Bryzgalova, Pelger, and Zhu (2024), these are the Bayesian estimates if the investor

[21] The assumption that the quant investor adjusts his belief relative to the maximum likelihood estimates can be justified conceptually and empirically. Conceptually, even though maximum likelihood yields the most efficient unbiased estimate, an important insights from machine learning (also present in empirical Bayesian methods) is that some biased estimators that rely on shrinkage have large efficiency gains that lead to better out-of-sample performance. Empirically, investors have no incentive to trade on anomalies in real time using standard (i.e., unadjusted) mean-variance optimization as it performs poorly out-of-sample (e.g., DeMiguel, Garlappi, and Uppal (2009)).

has the prior belief $\mu_t \sim \mathcal{N}(\lambda_\mu \cdot \Sigma_t 1_N, \lambda_\Sigma \cdot \Sigma_t^2)$, which is economically reasonable as it implies higher probability of high Sharpe ratios to principle components with larger eigenvalues.[22]

Substituting $\mu_{q,t}$ and $\Sigma_{q,t}$ into Equation 18 yields

$$\varpi_{q,t} \;=\; (1/\gamma) \cdot (\widehat{\Sigma}_t + \lambda_\Sigma \cdot \mathrm{I}_N)^{-1}(\widehat{\mu}_t - \lambda_\mu \cdot 1_N) \tag{19}$$

Letting $r_{q,t+1} = \varpi_{q,t}^{'} r_{t+1}$ represent excess returns on the wealth portfolio of the quant investor, we can isolate $\widehat{\mu}_t$ in Equation 19 to obtain an expression for expected returns:[23]

$$\widehat{\mu}_t \;=\; \gamma \cdot \widehat{\mathbb{C}ov}_t[r_q, r] \;+\; \lambda_\mu \cdot 1_N \;+\; \lambda_\Sigma \cdot \gamma \cdot \varpi_{q,t} \tag{20}$$

where $\widehat{\mathbb{C}ov}_t[r_q, r] = \widehat{\Sigma}_t \varpi_{q,t}$.

Letting $\varpi_{m,t}$ reflect market capitalization weights, the maximum likelihood estimate for the CAPM alpha is given by

$$\begin{aligned} \widehat{\alpha}_t \;&=\; \widehat{\mu}_t - \widehat{\beta}_t \cdot \widehat{\mu}_{m,t} \\ &=\; \gamma \cdot (\widehat{\mathbb{C}ov}_t[r_q, r] - \widehat{\mathbb{C}ov}_t[r_q, r_m] \cdot \widehat{\beta}_t) \;+\; \lambda_\mu \cdot (1_N - \widehat{\beta}_t) \;+\; \lambda_\Sigma \cdot \gamma \cdot (\varpi_{q,t} - \varpi_{m,t}^{'} \varpi_{q,t} \cdot \widehat{\beta}_t) \end{aligned} \tag{21}$$

where $\widehat{\mu}_{m,t} = \varpi_{m,t}^{'} \widehat{\mu}_t$ is the market expected return and $\widehat{\beta}_t = \widehat{\mathbb{C}ov}_t[r_m, r]/\widehat{\mathbb{V}ar}_t[r_m] = \widehat{\Sigma}_t \varpi_{m,t}/\varpi_{m,t}^{'} \widehat{\Sigma}_t \varpi_{m,t}$ is the vector of market betas for the N risky assets.[24] So, the investor expects assets to have CAPM alphas that differ from zero in the future.

[22] Note that, as shown in Bryzgalova, Pelger, and Zhu (2024), the BPZ objective function in Equation 15 (with $\lambda_1 = 0$) is equivalent to maximizing Sharpe ratio using $\mu_{q,t}$ and $\Sigma_{q,t}$ (and reduces to the approach in Kozak, Nagel, and Santosh (2020) if $\lambda_\mu = 0$). We use $\lambda_1 = 0$ in our model because otherwise the model is not tractable enough to deliver closed-form expressions for CAPM alphas. With $\lambda_1 = 0$, the model has shrinkage, but not anomaly selection. Anomaly selection would further strength our point as it would induce the quant investor to have zero demand for some anomalies, and thus to not reduce their IS alphas.

[23] If our model was based on investors who have FIRE, the analogue of Equation 20 would provide an expression for μ_t instead of its maximum likelihood estimate, $\widehat{\mu}_t$. It is easy to see how demand affects μ_t in equilibrium since high (low) demand induces high (low) prices, which leads to low (high) μ_t. A similar logic applies to $\widehat{\mu}_t$ since it reflects the maximum likelihood estimate based on all publicly available information, including prices. So, high (low) demand induces high (low) prices, which leads the maximum likelihood estimator to indicate a low expected return going forward (i.e., a low $\widehat{\mu}_t$).

[24] Equation 21 follows directly from $\widehat{\alpha}_t = \widehat{\mu}_t - \widehat{\beta}_t \cdot \widehat{\mu}_{m,t}$ with

$$\widehat{\beta}_t \cdot \widehat{\mu}_{m,t} \;=\; \widehat{\beta}_t \cdot \varpi_{m,t}^{'} \widehat{\mu}_t \;=\; \gamma \cdot \widehat{\mathbb{C}ov}_t[r_q, r_m] \cdot \widehat{\beta}_t + \lambda_\mu \cdot \widehat{\beta}_t + \lambda_\Sigma \cdot \gamma \cdot \varpi_{m,t}^{'} \varpi_{q,t} \cdot \widehat{\beta}_t$$

where the second equality uses Equation 20.

Note that

$$\widehat{\alpha}_t = \begin{cases} \gamma \cdot (\widehat{\mathbb{C}ov}_t[r_q, r] - \widehat{\mathbb{C}ov}_t[r_q, r_m] \cdot \widehat{\beta}_t) & \text{if} \quad \lambda_\mu = \lambda_\Sigma = 0 \\ \lambda_\mu \cdot (1_N - \widehat{\beta}_t) \; + \; \lambda_\Sigma \cdot \gamma \cdot (\varpi_{m,t} - \varpi_{m,t}^{'}\varpi_{m,t} \cdot \widehat{\beta}_t) & \text{if} \quad \varpi_{q,t} = \varpi_{m,t} \end{cases} \tag{22}$$

and thus there are two sources of CAPM alpha. First, since the quant investor does not necessarily hold the market portfolio, the "risk factor" that gets priced by the quant investor is r_q and not r_m. This induces $\widehat{\alpha}_t \neq 0_N$ even if the investor uses no shrinkage (i.e., under $\lambda_\mu = \lambda_\Sigma = 0$). Second, even if the quant investor held the market portfolio in equilibrium so that $\varpi_{q,t} = \varpi_{m,t}$ (which would be the case if all active investors were quant investors), we would still have $\widehat{\alpha}_t \neq 0_N$ due to $\lambda_\mu \neq 0$ and $\lambda_\Sigma \neq 0$. That is, because the investor acts based on beliefs that reflect shrinkage, he is less aggressive in trading than he would be otherwise, leaving alphas on the table.

The α_{IS} we study in prior sections is based on realized alphas over a period (the post-publication period) as opposed to $\widehat{\alpha}_t$, which depends on all information available to investors at time t. To obtain an expression for realized alphas in the model, we add the auxiliary assumption that the return distribution remains fixed after time t. In this case, the vector of realized alphas from $t+1$ to $t+H$ is

$$\overline{\alpha} \; = \; \overline{\mu} - \overline{\beta} \cdot \overline{\mu}_m \tag{23}$$

where $\overline{\mu} = (1/H) \cdot \Sigma_{h=1}^{H} r_{t+h}$, $\overline{\mu}_m = (1/H) \cdot \Sigma_{h=1}^{H} r_{m,t+h}$, and $\overline{\beta} = (\Sigma_{h=1}^{H}(r_{t+h} - \overline{\mu}) \cdot (r_{m,t+h} - \overline{\mu}_m))/(\Sigma_{h=1}^{H}(r_{m,t+h} - \overline{\mu}_m)^2)$. Moreover, the quant investors can forecast these realized alphas (through $\widehat{\alpha}_t$) and still does not eliminate them in equilibrium. Specifically, we have

$$\overline{\alpha} \; = \; \widehat{\alpha}_t \; + \; \varepsilon_\alpha \tag{24}$$

where $\mathbb{E}_t[\varepsilon_\alpha] = 0$ so that $\widehat{\alpha}_t$ is an unbiased estimate for $\overline{\alpha}$.[25]

[25]To derive Equation 24, start by noting that if $\widehat{\theta}$ is the maximum likelihood estimate for θ, then $f(\widehat{\theta})$ is the maximum likelihood estimate for $f(\theta)$. In our context, this implies $\widehat{\alpha}_t = \widehat{\mu}_t - \widehat{\beta}_t \cdot \widehat{\mu}_{m,t}$ is the maximum likelihood estimate for $\alpha_t = \mu_t - \beta_t \cdot \mu_{m,t}$, and thus we can write $\widehat{\alpha}_t = \alpha_t + u_\alpha$, where $\mathbb{E}_t[u_\alpha] = 0$. Then, since the return distribution remains fixed after time t, we have $\overline{\alpha} = \alpha_t + \epsilon_\alpha$, where $\mathbb{E}[\epsilon_\alpha | \mathcal{F}_t] = 0$ with $\mathcal{F}_t$ reflecting the information filtration process that includes all public information (which investors know) as well as all conditional return moments (which the investors do not know). Then, the law of iterated expectations implies

Thus, alphas persistent even in an economy where quant investors simultaneously allocate capital to multiple anomalies. The reason is that quant investors need to apply shrinkage to their portfolio weights in order to deal with the fundamental uncertainty present in the estimation of anomaly portfolio weights. As a consequence, their demand is not aggressive enough to eliminate the alphas of anomaly strategies in equilibrium.

6 Conclusion

In this paper, we show that when a new anomaly strategy is published, real time investors would not know the optimal portfolio weights needed to combine the anomaly strategy with the market index in order to achieve a positive alpha in the post-publication period. As such, while the average IS alpha of anomaly strategies is strongly positive post-publication, the average OS alpha, which relies on pre-publication optimal weights, is relatively close to zero. Our evidence is robust to various empirical choices, including the use of empirical Bayesian shrinkage and machine learning methods to estimate pre-publication optimal weights. We also provide a model that implies that even though investors may achieve positive OS alphas by trading on multiple anomalies jointly while using shrinkage methods (i.e., by acting as quants), their demand is not aggressive enough to eliminate alphas.

Combining our empirical results with our model implications, we argue that the following channel contributes to the persistence of alphas of anomaly strategies over time. The low average OS alpha of published anomaly strategies implies investors have little incentive to trade anomalies individually in real time. Moreover, trading anomalies jointly in real time is only profitable with some form of shrinkage and/or anomaly selection (as is standard in quant investing) in order to mitigate the inherent uncertainty present in estimating the optimal weights needed to make investment decisions in real time. However, investors acting as quants are not aggressive enough to fully eliminate alphas. Consequently, alphas persist in equilibrium.

$\mathbb{E}_t[\epsilon_\alpha] = \mathbb{E}_t[\mathbb{E}[\epsilon_\alpha|\mathcal{F}_t]] = 0$, so that $\overline{\alpha} = \widehat{\alpha}_t + \varepsilon_\alpha$, where $\varepsilon_\alpha = \epsilon_\alpha - u_\alpha$ satisfies $\mathbb{E}_t[\varepsilon_\alpha] = 0$.

Our findings have implications that are significant for both theory and practice. From a theoretical perspective, they highlight the importance of writing models in which investors account for the difficulties in estimating optimal weights when making their portfolio allocations. In such models, IS alphas that differ from zero can coexist with smart unconstrained investors. More broadly, for a model to credibly explain the fact that anomaly strategies have a positive average IS alpha, the model also needs to produce the pattern that their average OS alpha is close to zero. From a practical perspective, our findings highlight that investors should evaluate trading strategies based on OS alphas, which cannot be obtained directly from the typical factor regressions used in the empirical literature (as they yield IS alphas).

Our work also opens the door to new questions. For instance, it would be interesting to explore the implications of our findings for capital budgeting decisions. In particular, since expected returns that embed OS alphas capture the true opportunity cost of an equity investment, one can argue that using OS alphas when estimating cost of equity is appropriate, in which case the standard CAPM provides a reasonable way to estimate cost of equity. In contrast, since average realized returns embed IS alphas, one can alternatively argue that using IS alphas when estimating cost of equity is a better approach, in which case multifactor models should be used to estimate cost of equity. Exploring this and other related questions is an interesting avenue for future research.

References

Baba Yara, F., B. H. Boyer, and C. Davis (2024). "Messy Asset Pricing: Can AI Models Lead to a Consensus?" Working Paper.

Bretscher, L., R. Lewis, and S. Santosh (2024). "Investor Betas". Working Paper.

Bryzgalova, S., M. Pelger, and J. Zhu (2024). "Forest Through the Trees: Building Cross-Sections of Stock Returns". In: *Journal of Finance* Forthcoming.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Cederburg, S., M. S. O'Doherty, F. Wang, and X. S. Yan (2020). "On the Performance of Volatility-Managed Portfolios". In: *Journal of Financial Economics* 138.1, pp. 95–117.

Chabi-Yo, F., A. S. Gonçalves, and J. Loudis (2024). "An Intertemporal Risk Factor Model". In: *Management Science*. Forthcoming.

Chen, A. Y. and T. Zimmermann (2023). "Publication Bias and the Cross-Section of Stock Returns". In: *Oxford Research Encyclopedia of Economics and Finance*. Forthcoming.

Chen, A. Y., A. Lopez-Lira, and T. Zimmermann (2023). "Does peer-reviewed theory help predict the cross-section of stock returns?" Working Paper.

Chen, A. Y. and M. Velikov (2023). "Zeroing in on the Expected Returns of Anomalies". In: *Journal of Financial and Quantitative Analysis* 58.3, pp. 968–1004.

Chen, A. Y. and T. Zimmermann (2022). "Open Source Cross-Sectional Asset Pricing". In: *Critical Finance Review* 27.2. Working Paper, pp. 207–264.

Da, R., S. Nagel, and D. Xiu (2022). "The Statistical Limit of Arbitrage". Working Paper.

Davis, C. (2024). "The Elasticity of Quantitative Investment". Working Paper.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" In: *Review of Financial Studies* 22.5, pp. 1915–1953.

Fama, E. F. and K. R. French (1996). "Multifactor Explanations of Asset Pricing Anomalies". In: *Journal of Finance* 55.1, pp. 55–84.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). "A Test of the Efficiency of a Given Portfolio". In: *Econometrica* 57.5, pp. 1121–1152.

Goyal, A. and I. Welch (2008). "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction". In: *Review of Financial Studies* 21.4, pp. 1455–1508.

Goyal, A., I. Welch, and A. Zafirov (2024). "A Comprehensive 2022 Look at the Empirical Performance of Equity Premium Prediction". In: *Review of Financial Studies* Forthcoming.

Gromb, D. and D. Vayanos (2010). "Limits of Arbitrage". In: *Annual Review of Financial Economics* 2, pp. 251–275.

Gu, S., B. Kelly, and D. Xiu (2020). "Empirical Asset Pricing via Machine Learning". In: *Review of Financial Studies* 33.5, pp. 2223–2273.

Harvey, C. R., Y. Liu, and H. Zhu (2015). "... and the Cross-Section of Expected Returns". In: *Review of Financial Studies* 29.1, pp. 5–68.

Jensen, T. I., B. T. Kelly, S. Malamud, and L. H. Pedersen (2025). "Machine Learning and the Implementable Efficient Frontier". In: *Review of Financial Studies* Forthcoming.

Jensen, T. I., B. T. Kelly, and L. H. Pedersen (2023). "Is There a Replication Crisis in Finance?" In: *Journal of Finance* 78.5. Forthcoming.

Kan, R., X. Wang, and X. Zheng (2024). "In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing model". In: *Journal of Financial Economics* 155, p. 103837.

Koijen, R. S. J. and M. Yogo (2019). "A Demand System Approach to Asset Pricing". In: *Journal of Political Economy* 127.4, pp. 1475–1515.

Kozak, S., S. Nagel, and S. Santosh (2020). "Shrinking the Cross-Section". In: *Journal of Financial Economics* 135.2, pp. 271–292.

Marrow, B. and S. Nagel (2024). "Real-Time Discovery and Tracking of Return-Based Anomalies". Working Paper.

McLean, R. D. and J. Pontiff (2016). "Does Academic Research Destroy Stock Return Predictability?" In: *Journal of Finance* 71.1, pp. 5–32.

Newey, W. K. and K. D. West (1987). "A Simple, Positive-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708.

Newey, W. K. and K. D. West (1994). "Automatic Lag Selection in Covariance Matrix Estimation". In: *Review of Economic Studies* 61.4, pp. 631–653.

Pénasse, J. (2022). "Understanding Alpha Decay". In: *Management Science* 68.5, pp. 3966–3973.

Shleifer, A. and R. W. Vishny (1997). "The Limits of Arbitrage". In: *Journal of Finance* 52.1, pp. 35–55.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society: Series B* 58.1, pp. 267–288.

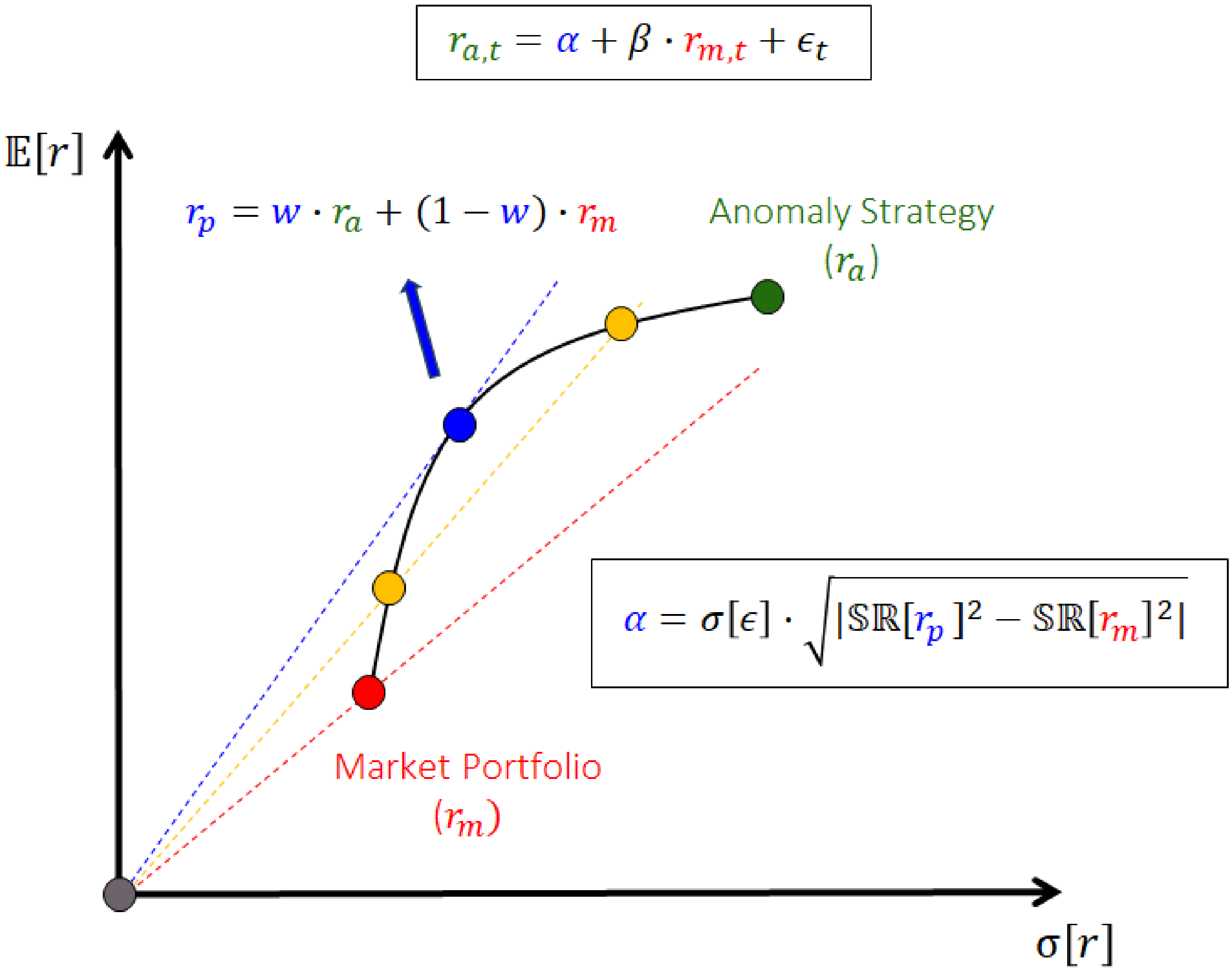


Figure 2
Conceptual Relation Between Alphas and Sharpe Ratios

This figure provides a mean-variance diagram in excess return space that reflects the conceptual relation between alphas and Sharpe ratios. The red dot reflects the market portfolio (r_m) while the green dot reflects the anomaly strategy (r_a). The blue dot captures the maximum Sharpe ratio portfolio that can be formed by combining the market portfolio with the anomaly strategy, $r_p = (1 - w) \cdot r_{m,t} + w \cdot r_{a,t}$. Setting $w = w_{IS}$ maximizes the in-sample Sharpe ratio (i.e., line slope) increase from the red dot to the blue dot. As such, setting $w = w_{OS}$ leads to a lower Sharpe ratio increase whether $w_{OS} > w_{IS}$ or $w_{OS} < w_{IS}$, as can be seen by comparing the slopes of the blue and orange lines. Section 1.1 provides more details on the relation between alphas and Sharpe ratios, including an analysis of this mean-variance diagram.

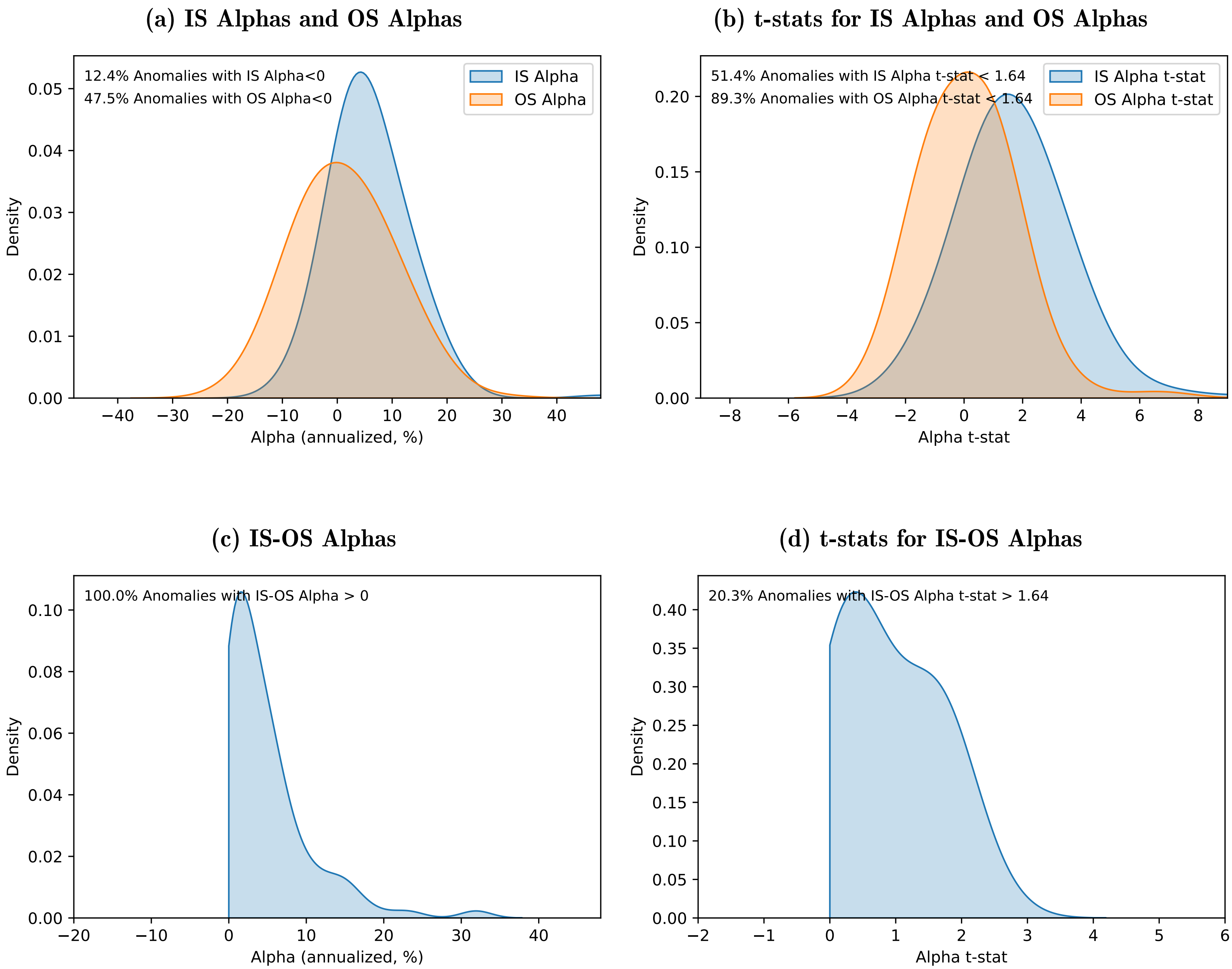


Figure 3
Distribution of Alphas: Baseline Weight Estimation (Main Results)

This figure contains density plots for the distributions of CAPM alphas and their t-statistics across 177 anomaly strategies. We form each anomaly strategy as the long-short decile portfolio with weights as per the original publication. We consider in-sample (IS) alphas and out-of-sample (OS) alphas as well as their differences (IS-OS). IS and OS alphas differ based on the weights used to combine the anomaly strategy with the market portfolio. IS (OS) alphas use weights calculated from the post-publication (pre-publication) period to build a portfolio over the post-publication period, with the pre-publication period ending in December of the publication year. Consequently, IS alphas are equivalent to intercepts from factor regressions over the post-publication period, but OS alphas are not. The t-statistics are based on the bootstrap procedure described in Internet Appendix B. Results are based on monthly returns and annualized (approximately) by multiplying by 12. Section 1 provides more details on alphas while Subsection 2.1 discusses the results from this figure.

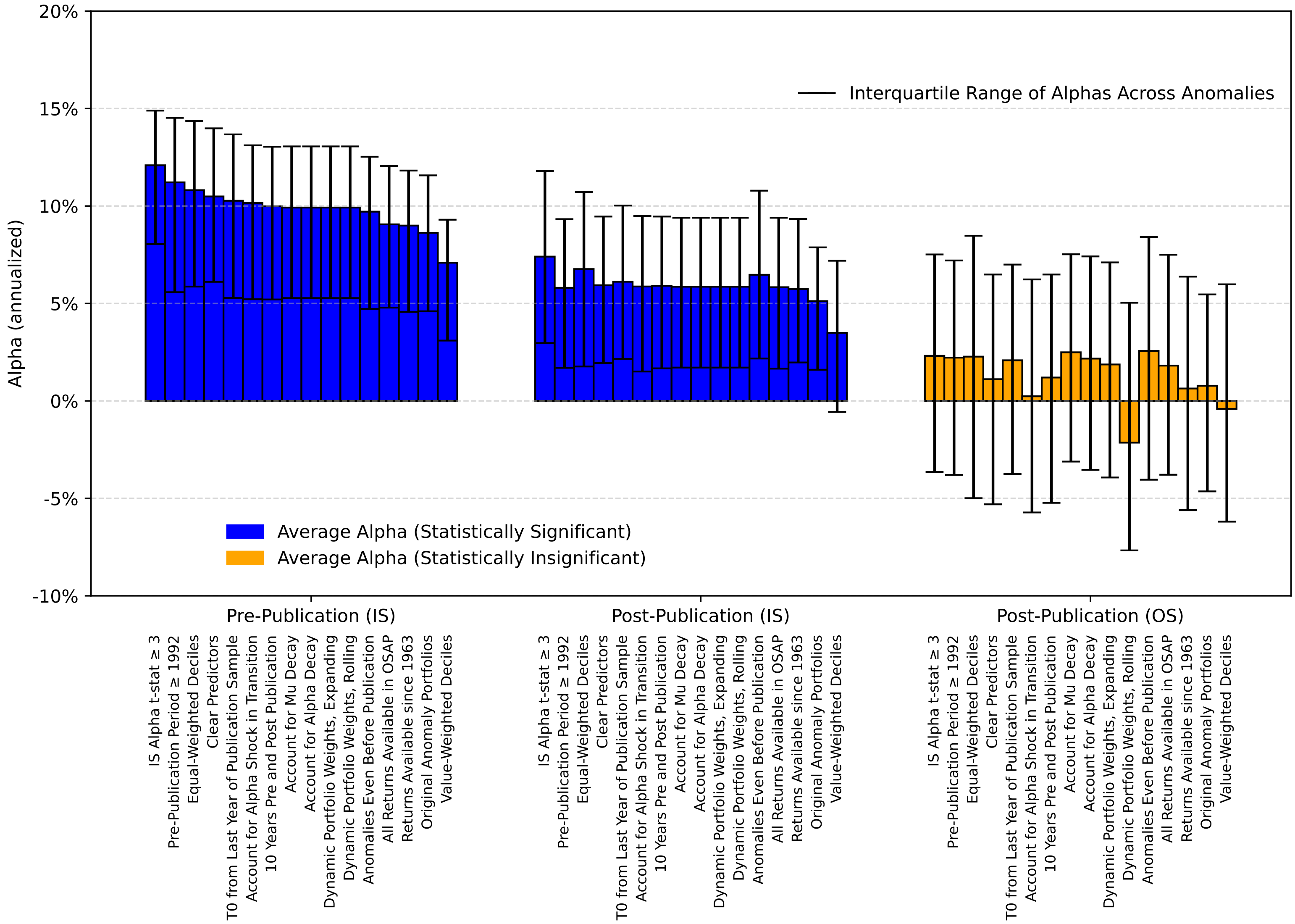


Figure 4
Average and Interquartile Range of Alpha Distribution Across Anomalies (Robustness Analysis)

This figure contains a barplot that summarizes the results from the different specifications we study in our robustness analysis (effectively extending Figure 1 to the different specifications we explore). There are three blocks of bars, with each bar reporting an average alpha. The first block focuses on in-sample (IS) alphas pre-publication, the second block focuses on IS alphas post-publication, and the third block focuses on out-of-sample (OS) alphas post-publication. The intervals around each bar correspond to the interquartile range of the alpha distribution across anomalies, with the lower and upper bounds indicating the 25th and 75th percentiles, respectively. We sort specifications based on the magnitude of their average IS alphas pre-publication and color bars based on whether the average alpha is statistically significant at the 5% level (blue) or not (orange). The t-statistics are based on the bootstrap procedure described in Internet Appendix B. Results are based on monthly returns and annualized (approximately) by multiplying by 12. Section 1 provides more details on alphas while Subsection 2.2 discusses the results from this figure.

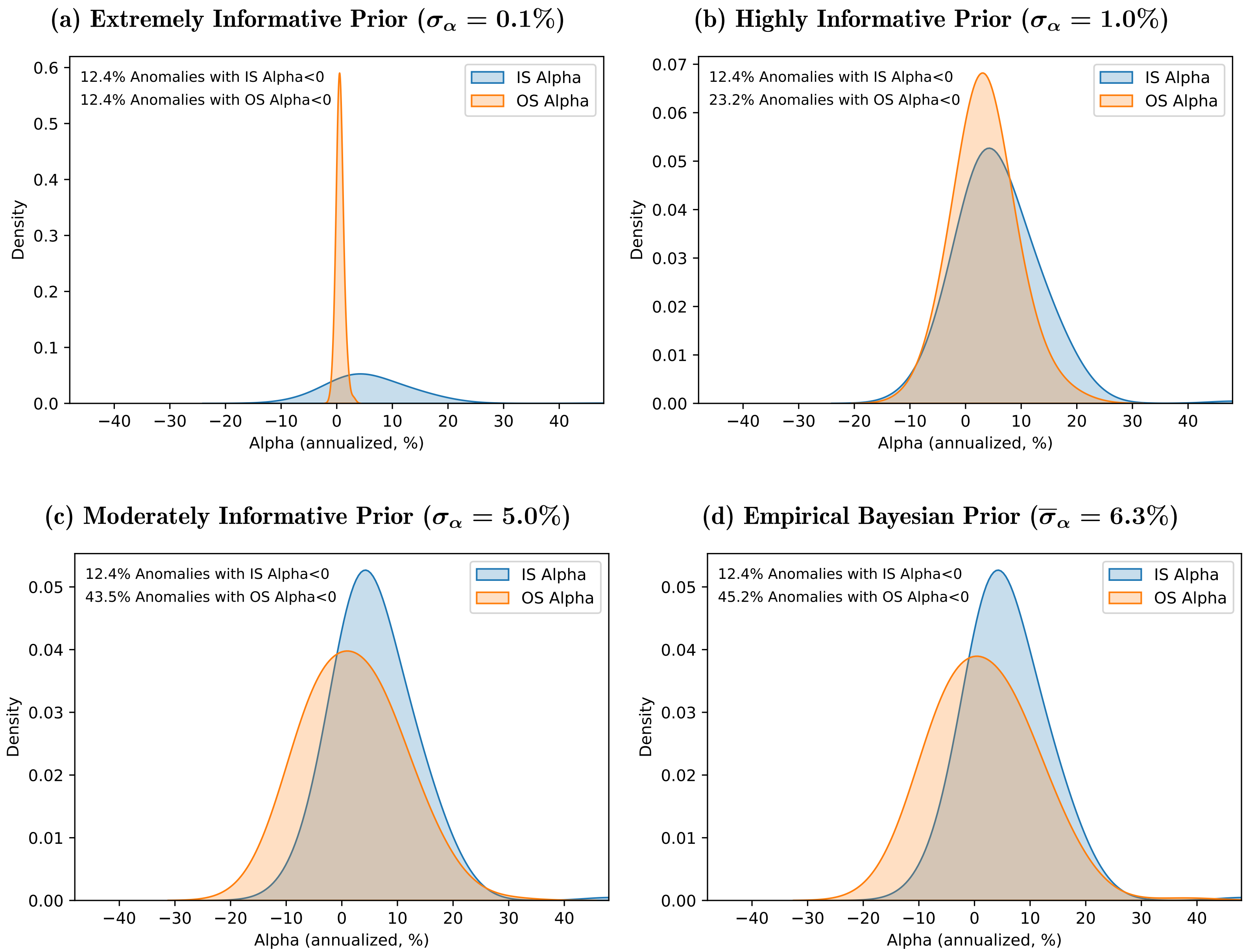
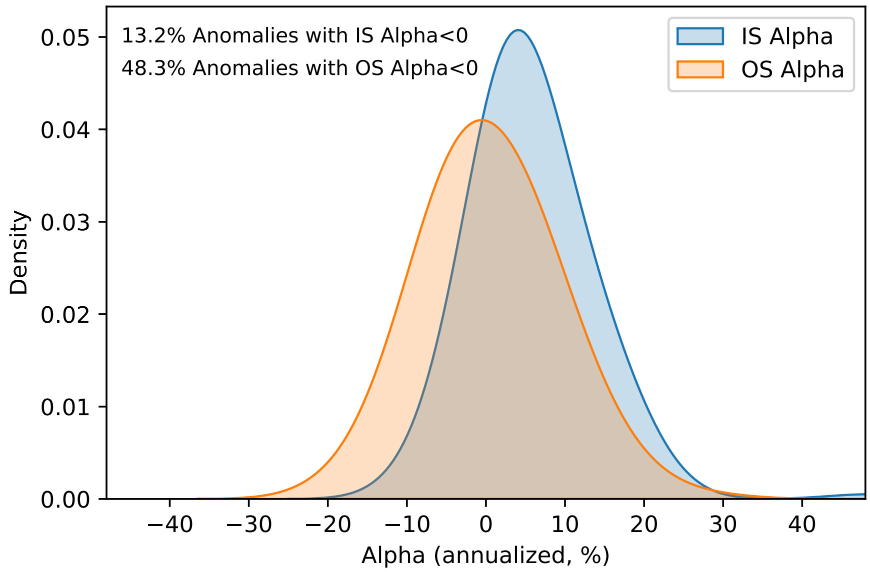


Figure 5
Distribution of Alphas: Bayesian Weight Estimation

This figure contains density plots for the distributions of CAPM alphas across 177 anomaly strategies. We form each anomaly strategy as the long-short decile portfolio with weights as per the original publication. We consider in-sample (IS) alphas and out-of-sample (OS) alphas. IS (OS) alphas use weights calculated from the post-publication (pre-publication) period to build a portfolio over the post-publication period, with the pre-publication period ending in December of the publication year. Consequently, IS alphas are equivalent to intercepts from factor regressions over the post-publication period, but OS alphas are not. The optimal weights for the OS alphas are constructed using Bayesian shrinkage with the figure panels varying by the level of prior informativeness. Panels A to C consider different fixed priors while Panel D considers an empirical Bayesian prior, with the details described in Subsection 3.1. Results are based on monthly returns and annualized (approximately) by multiplying by 12. Section 1 provides more details on alphas while Subsection 3.2 discusses the results from this figure.

(a) Estimating Weights using KNS (Kozak, Nagel, and Santosh (2020))



(b) Estimating Weights using BPZ (Bryzgalova, Pelger, and Zhu (2024))

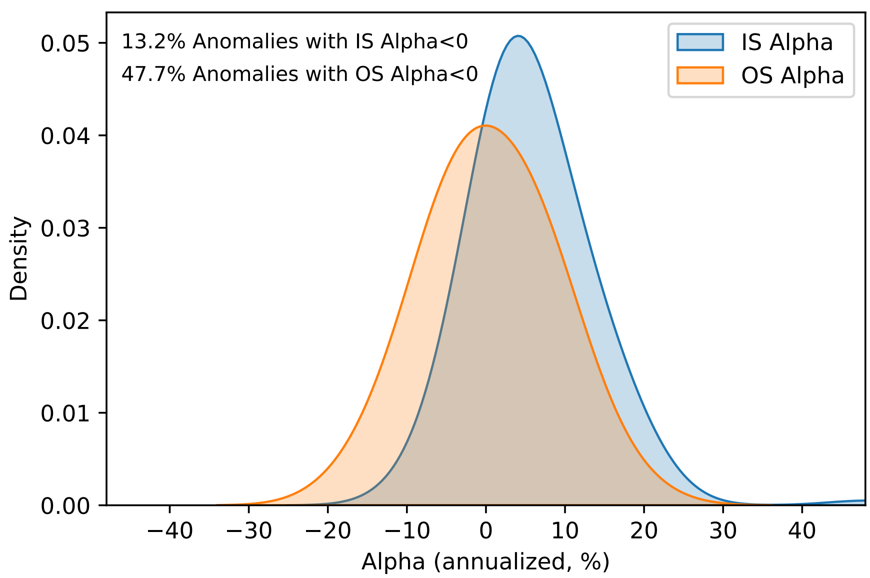


Figure 6
Distribution of Alphas: Machine Learning Weight Estimation

This figure contains density plots for the distributions of CAPM alphas across 151 anomaly strategies. These 151 anomaly strategies reflect the subset of our original 177 anomaly strategies that have return data for at least 20 years in the pre-publication period. We form each anomaly strategy as the long-short decile portfolio with weights as per the original publication. We consider in-sample (IS) alphas and out-of-sample (OS) alphas. IS (OS) alphas use weights calculated from the post-publication (pre-publication) period to build a portfolio over the post-publication period, with the pre-publication period ending in December of the publication year. Consequently, IS alphas are equivalent to intercepts from factor regressions over the post-publication period, but OS alphas are not. The optimal weights for the OS alphas are constructed using Machine Learning methods, with the details described in Subsection 3.1. Following Davis (2024), we only explore two machine learning methods to compute portfolio weights: the KNS method from Kozak, Nagel, and Santosh (2020) (in Panel A) and the BPZ method from Bryzgalova, Pelger, and Zhu (2024) (in Panel B). Results are based on monthly returns and annualized (approximately) by multiplying by 12. Section 1 provides more details on alphas while Subsection 4.2 discusses the results from this figure.

Table 1
Summary Statistics Across our Baseline 177 Anomaly Strategies

This table reports summary statistics for the baseline sample of 177 anomaly strategies we study, with each anomaly strategy reflecting a long-short decile portfolio with weights as per the original anomaly publication. The table includes information on publication years, the first and last year of the publication sample, the first and last year of anomaly strategy returns available in the OSAP dataset, the number of months in the pre- and post-publication periods, average returns, CAPM market risk premiums ($\beta \cdot \mathbb{E}[r_m]$), and in-sample (IS) CAPM alphas. The table also provides t-statistics for IS alphas, based on both Newey and West (1987, 1994) and the bootstrap procedure described in Internet Appendix B. These statistics highlight the performance of anomaly portfolios both before and after their publication, emphasizing changes in returns and alphas over time. Results are based on monthly returns and annualized (approximately) by multiplying by 12. Section 1 provides more details on alphas while Subsection 1.3 discusses the results from this table.

		Mean	**Median**	**Min**	**Q10%**	**Q25%**	**Q75%**	**Q90%**	**Max**
	Publication Year	2004	2006	1973	1993	2001	2009	2012	2016
	First Year (of Publication Sample)	1970	1970	1926	1962	1963	1979	1986	2002
	Last Year (of Publication Sample)	1999	2001	1968	1989	1995	2004	2009	2014
	First Year (of Returns Available)	1955	1953	1926	1927	1931	1971	1983	2001
	Last Year (of Returns Available)	2023	2023	2013	2023	2023	2023	2023	2023
	Full Sample	640	642	240	448	534	726	738	1170
# of Months	**Pre-Publication**	414	426	106	210	294	522	582	1038
	Post-Publication	226	204	84	121	157	264	360	600
	Full Sample	7.4%	6.1%	-4.3%	2.0%	4.0%	9.5%	13.5%	44.5%
Average Return	**Pre-Publication**	9.0%	7.6%	-4.5%	2.5%	4.6%	11.4%	16.2%	46.3%
	Post-Publication	4.5%	3.7%	-5.6%	-0.7%	1.4%	6.7%	10.3%	41.9%
	Full Sample	-1.0%	-0.6%	-9.9%	-3.9%	-1.8%	0.0%	0.8%	6.6%
$\boldsymbol{\beta \cdot \mathbb{E}[r_m]}$	**Pre-Publication**	-0.9%	-0.5%	-9.5%	-3.1%	-1.4%	0.1%	0.7%	11.1%
	Post-Publication	-1.4%	-0.7%	-14.9%	-5.8%	-2.5%	0.4%	2.2%	7.4%
	Full Sample	8.4%	7.3%	-1.2%	2.0%	4.1%	11.5%	15.9%	47.7%
IS Alpha	**Pre-Publication**	9.9%	8.8%	0.7%	3.0%	5.3%	13.1%	17.2%	47.4%
	Post-Publication	5.9%	4.4%	-9.7%	-0.8%	1.7%	9.4%	14.3%	48.5%
	Full Sample	(4.23)	(3.84)	(-0.58)	(1.38)	(2.58)	(5.73)	(7.30)	(15.2)
IS Alpha t-stat	**Pre-Publication**	(4.31)	(3.64)	(0.23)	(1.55)	(2.48)	(5.73)	(7.44)	(14.7)
(Newey-West)	**Post-Publication**	(1.67)	(1.62)	(-1.97)	(-0.29)	(0.56)	(2.65)	(3.64)	(9.42)
	Full Sample	(4.66)	(4.19)	(-0.94)	(1.64)	(2.65)	(6.52)	(8.15)	(16.0)
IS Alpha t-stat	**Pre-Publication**	(4.69)	(4.20)	(0.37)	(1.80)	(2.63)	(6.15)	(8.15)	(15.7)
(Bootstrap)	**Post-Publication**	(1.70)	(1.54)	(-1.82)	(-0.30)	(0.58)	(2.67)	(3.72)	(10.4)

Table 2
Distribution of Alphas: Baseline Weight Estimation (Main Results)

This table reports our main results for the distributions of CAPM alphas and their t-statistics across the 177 anomaly strategies we study, with each anomaly strategy reflecting a long-short decile portfolio with weights as per the original anomaly publication. We consider in-sample (IS) alphas and out-of-sample (OS) alphas as well as their differences (IS-OS). IS (OS) alphas use weights calculated from the post-publication (pre-publication) period to build a portfolio over the post-publication period, with the pre-publication period ending in December of the publication year. Consequently, IS alphas are equivalent to intercepts from factor regressions over the post-publication period, but OS alphas are not. Results are based on monthly returns and annualized (approximately) by multiplying by 12. Section 1 provides more details on alphas while Subsection 2.1 discusses the results from this table.

		Mean	**Median**	**Min**	**Q10%**	**Q25%**	**Q75%**	**Q90%**	**Max**
Pre-Pub	**IS Alpha**	9.9%	8.8%	0.7%	3.0%	5.3%	13.1%	17.2%	47.4%
	IS Alpha t-stat	(4.69)	(4.20)	(0.37)	(1.80)	(2.63)	(6.15)	(8.15)	(15.75)
	Optimal Weight	0.62	0.63	0.15	0.38	0.51	0.76	0.85	1.05
Post-Pub	**IS Alpha**	5.9%	4.4%	-9.7%	-0.8%	1.7%	9.4%	14.3%	48.5%
	IS Alpha t-stat	(1.70)	(1.54)	(-1.82)	(-0.30)	(0.58)	(2.67)	(3.72)	(10.4)
	Optimal Weight	-0.22	0.36	-40.7	-0.17	0.22	0.52	0.63	0.84
	OS Alpha	1.1%	1.6%	-19.9%	-8.8%	-5.3%	6.4%	12.8%	29.4%
	OS Alpha t-stat	(0.15)	(0.28)	(-2.57)	(-1.58)	(-1.00)	(1.06)	(1.65)	(6.90)
	IS-OS Alpha	4.7%	3.2%	0.0%	0.1%	0.8%	6.1%	12.1%	32.0%
	IS-OS Alpha t-stat	(0.89)	(0.66)	(0.00)	(0.05)	(0.20)	(1.56)	(1.87)	(2.64)

Table 3

Distribution of Alphas: Bayesian Weight Estimation

This table reports the results for the distributions of CAPM alphas and their t-statistics across the 177 anomaly strategies we study, with each anomaly strategy reflecting a long-short decile portfolio with weights as per the original anomaly publication. We consider out-of-sample (OS) alphas, which use weights calculated from the pre-publication period to build a portfolio over the post-publication period, with the pre-publication period ending in December of the publication year. Consequently, OS alphas are not equivalent to intercepts from factor regressions over the post-publication period. The optimal weights for the OS alphas are constructed using Bayesian shrinkage with the table panels varying by the level of prior informativeness. Panels A to C consider different fixed priors while Panel D considers an empirical Bayesian prior, with the details described in Subsection 3.1. Results are based on monthly returns and annualized (approximately) by multiplying by 12. Section 1 provides more details on alphas while Subsection 3.2 discusses the results from this table.

PANEL A - Extremely Informative Prior ($\sigma_\alpha = 0.1\%$)

	Mean	Median	Min	Q10%	Q25%	Q75%	Q90%	Max
(Bayesian) Optimal Weight	0.01	0.00	0.00	0.00	0.00	0.02	0.04	0.11
OS Alpha	0.6%	0.5%	-1.0%	-0.1%	0.3%	0.9%	1.3%	3.1%
OS Alpha t-stat	(0.68)	(0.80)	(-1.05)	(-0.39)	(0.48)	(1.04)	(1.26)	(1.99)

PANEL B - Highly Informative Prior ($\sigma_\alpha = 1.0\%$)

	Mean	Median	Min	Q10%	Q25%	Q75%	Q90%	Max
(Bayesian) Optimal Weight	0.33	0.29	0.00	0.05	0.12	0.52	0.68	0.89
OS Alpha	3.5%	3.1%	-9.7%	-2.5%	0.7%	6.2%	9.7%	21.1%
OS Alpha t-stat	(0.81)	(1.03)	(-2.25)	(-0.86)	(0.23)	(1.50)	(2.08)	(7.82)

PANEL C - Moderately Informative Prior ($\sigma_\alpha = 5.0\%$)

	Mean	Median	Min	Q10%	Q25%	Q75%	Q90%	Max
(Bayesian) Optimal Weight	0.58	0.58	0.02	0.28	0.45	0.75	0.84	1.03
OS Alpha	2.0%	2.3%	-14.6%	-7.0%	-4.0%	7.4%	12.7%	29.0%
OS Alpha t-stat	(0.29)	(0.54)	(-2.57)	(-1.47)	(-0.89)	(1.13)	(1.97)	(6.94)

PANEL D - Empirical Bayesian Prior ($\overline{\sigma}_\alpha = 6.3\%$)

	Mean	Median	Min	Q10%	Q25%	Q75%	Q90%	Max
(Bayesian) Optimal Weight	0.59	0.60	0.04	0.30	0.48	0.75	0.84	1.04
OS Alpha	1.8%	1.9%	-15.2%	-7.7%	-4.1%	7.2%	12.6%	37.9%
OS Alpha t-stat	(0.25)	(0.48)	(-2.57)	(-1.48)	(-0.97)	(1.11)	(1.86)	(6.92)

Table 4

Distribution of Alphas: Machine Learning Weight Estimation

This table reports the results for the distributions of CAPM alphas and their t-statistics across 151 of the anomaly strategies we study, with each anomaly strategy reflecting a long-short decile portfolio with weights as per the original anomaly publication. These 151 anomaly strategies reflect the subset of our original 177 anomaly strategies that have return data for at least 20 years in the pre-publication period. We consider out-of-sample (OS) alphas, which use weights calculated from the pre-publication period to build a portfolio over the post-publication period, with the pre-publication period ending in December of the publication year. Consequently, OS alphas are not equivalent to intercepts from factor regressions over the post-publication period. The optimal weights for the OS alphas are constructed using Machine Learning methods, with the details described in Subsection 3.1. Following Davis (2024), we only explore two machine learning methods to compute portfolio weights: the KNS method from Kozak, Nagel, and Santosh (2020) (in Panel B) and the BPZ method from Bryzgalova, Pelger, and Zhu (2024) (in Panel C). For comparison, Panel A provides results using our baseline weight estimation applied to the same 151 anomaly strategies. Results are based on monthly returns and annualized (approximately) by multiplying by 12. Section 1 provides more details on alphas while Subsection 4.2 discusses the results from this table.

PANEL A - Estimating Weights using our Baseline Method

	Mean	Median	Min	Q10%	Q25%	Q75%	Q90%	Max
Optimal Weight	0.63	0.64	0.15	0.38	0.53	0.77	0.85	1.05
OS Alpha	0.7%	-0.5%	-19.9%	-9.2%	-5.7%	6.1%	11.3%	29.4%
OS Alpha t-stat	(0.04)	(-0.12)	(-2.57)	(-1.67)	(-1.05)	(1.01)	(1.61)	(6.90)

PANEL B - Estimating Weights using the KNS method (Kozak, Nagel, and Santosh (2020))

	Mean	Median	Min	Q10%	Q25%	Q75%	Q90%	Max
(ML) Optimal Weight	0.61	0.63	0.00	0.29	0.51	0.74	0.84	1.00
OS Alpha	0.5%	0.0%	-19.2%	-8.6%	-5.2%	5.7%	10.7%	27.6%
OS Alpha t-stat	(0.03)	(0.00)	(-2.41)	(-1.50)	(-0.94)	(0.89)	(1.27)	(6.75)

PANEL C - Estimating Weights using the BPZ method (Bryzgalova, Pelger, and Zhu (2024))

	Mean	Median	Min	Q10%	Q25%	Q75%	Q90%	Max
(ML) Optimal Weight	0.62	0.62	0.00	0.38	0.51	0.76	0.84	1.00
OS Alpha	0.4%	0.0%	-17.3%	-9.4%	-4.8%	6.4%	10.0%	19.5%
OS Alpha t-stat	(0.04)	(0.00)	(-2.39)	(-1.49)	(-1.01)	(0.87)	(1.26)	(6.74)

Internet Appendix

"Out-of-Sample Alphas Post-Publication"

By Andrei S. Gonçalves, Johnathan A. Loudis, and Richard E. Ogden

This Internet Appendix is organized as follows. Section A derives the equation that links alphas to Sharpe ratios in the main text. Section B explains our bootstrap procedure used to estimate alpha standard errors used in the main text.

A Derivation of Alpha Equation

This section derives (a generalization to) Equation 2 in the main text.

Suppose we have the factor model such that

$$r_{a,t} = \alpha + \sum_{k=1}^{K} \beta_k \cdot f_{k,t} + \varepsilon_t \tag{IA.1}$$

where $f_{k,t}$ reflects the long-short return on the zero-cost tradable factor k and $r_{a,t}$ reflects the long-short return on a zero-cost tradable strategy (i.e., an anomaly strategy).

Our objective is to derive an expression that connects α to Sharpe ratios. For that, it is useful to note that a Stochastic Discount Factor (SDF) that perfectly prices all $f_{k,t}$ can be written as $SDF_t = a + b \cdot \sum_{k=1}^{K} w_k \cdot f_{k,t} = a + b \cdot f_t^*$, where w_k are the maximum Sharpe ratio weights (see Chapter 6 in Cochrane (2005)). As such, the α from Equation IA.1 is identical to the α from equation[26]

$$r_{a,t} = \alpha + \beta \cdot f_t^* + \epsilon_t \tag{IA.2}$$

Equation 6.6.17 in Campbell, Lo, and MacKinlay (1997) (originally derived in Gibbons, Ross, and Shanken (1989)) shows that the α from Equation IA.2 satisfies

$$\frac{\alpha^2}{\mathbb{V}ar[\epsilon]} = \mathbb{SR}[r_p^*]^2 - \mathbb{SR}[f^*]^2 \tag{IA.3}$$

where $r_{p,t}^* = w \cdot r_{a,t} + (1-w) \cdot f_t^*$ represents the ex-post maximum Sharpe ratio portfolio that can be formed with $r_{a,t}$ and f_t^*. Consequently, we have

$$|\alpha| = \sigma[\epsilon] \cdot \sqrt{\mathbb{SR}[r_p^*]^2 - \mathbb{SR}[f^*]^2} \tag{IA.4}$$

Since $r_{p,t}^*$ is the ex-post maximum Sharpe ratio portfolio that can be formed with f_t^* and r_t, we have $\Delta = \mathbb{SR}[r_p^*] - \mathbb{SR}[f^*] \geq 0$, with this inequality holding with equality if and only if $\alpha = 0$. Moreover, $\alpha > 0$ ($\alpha < 0$) implies $w > 0$ ($w < 0$). Consequently, Equation IA.4 can

[26] Note that the residual terms in Equations IA.1 and IA.2 differ so that the idiosyncratic volatility (and consequently R^2 values) differ between the two equations. For our purpose, it only matters that the alphas are the same.

be alternatively written as

$$\alpha \ = sign[w \cdot \Delta] \cdot \sigma[\epsilon] \cdot \sqrt{\left|\mathbb{SR}[r_p^*]^2 \ - \ \mathbb{SR}[f^*]^2\right|} \tag{IA.5}$$

Note that if $r_{m,t}$ is the only factor (i.e., $f_t^* = r_{m,t}$), then Equation IA.5 becomes

$$\alpha \ = sign[w \cdot \Delta] \cdot \sigma[\epsilon] \cdot \sqrt{\left|\mathbb{SR}[r_p]^2 \ - \ \mathbb{SR}[r_m]^2\right|} \tag{IA.6}$$

where $r_{p,t} = w \cdot r_{a,t} + (1 - w) \cdot r_{m,t}$ so that we have Equation 2 in the main text.

Note also that the coefficient of determination of the factor model in Equation IA.2 is given by $\mathbb{C}or[r_a, f^*]^2 = 1 - \mathbb{V}ar[\epsilon]/\mathbb{V}ar[r_a]$, implying

$$\sigma[\epsilon] \ = \sigma[r_a] \cdot \sqrt{1 - \mathbb{C}or[r_a, f^*]^2} \tag{IA.7}$$

While we do not directly use Equation IA.7 in the main text, it is useful in highlighting that $\sigma[\epsilon]$ is a simple function of the r_a volatility and the correlation between r_a and f^*.

B Bootstrap Procedure for Standard Errors

In the main text, we use bootstrap standard errors for alphas. This section describes our bootstrap procedure.

For the bootstrap standard errors associated with a given anomaly strategy, we proceed as follows. We sample (with replacement) months from the anomaly specific pre-publication and post-publication periods separately (this approach conserves the structure of returns present in the data). We then concatenate the sampled pre-publication and post-publication periods to form a full bootstrap sample for the given anomaly. Using this sample, we calculate full sample IS alphas, pre-publication IS alphas, post-publication IS alphas, and post-publication OS alphas, as well as the difference between post-publication IS and OS alphas (which we refer to as IS-OS alphas). We repeat this process 10,000 times and obtain standard errors from the cross-simulation standard deviation of each of these alphas. In the case of alphas computed using Bayesian optimal weights, we hold fixed the prior distribution of alphas used. In the case of the machine learning methods (KNS and BPZ), we repeat the 4-fold cross validation design to select the hyperparameters in each simulation. Figure IA.1 contrasts (for post-publication IS alphas) our bootstrap t-statistics with t-statistics obtained from Newey and West (1987, 1994), which are common in the literature. The two types of t-statistics tend to be very similar.

For the bootstrap standard errors associated with average alphas across anomalies, we cannot sample pre-publication and post-publication periods separately (as they differ across anomalies), and thus proceed as follows. We start by splitting the sample between early (1963 to 2002) and late (2003 to 2023) periods. Then, we sample (with replacement) months from each of these periods separately and obtain anomaly returns for the sampled months accordingly (we sample $12 \times 40 = 480$ months from the early period and $12 \times 21 = 252$ months from the late period). For each anomaly, we classify each month as a pre-publication month or a post-publication month based on whether it comes from the given anomaly's pre-publication or post-publication period. Finally, we calculate full sample IS alphas, pre-publication IS alphas, post-publication IS alphas, and post-publication OS alphas, as well as the difference between post-publication IS and OS alphas (which we refer to as IS-OS alphas). We then average these values across anomalies. Finally, we repeat this process 10,000 times and obtain standard errors for average alphas from the cross-simulation standard deviation of each of these average alphas. The only exception is the specification "Anomalies Even Before Publication" (in Figure 4) as we repeat the process 1,000 times in that case (due to the long computing time for that specification).

References for Internet Appendix

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Cochrane, J. H. (2005). *Asset Pricing*. Revised Edition. Princeton University Press.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). "A Test of the Efficiency of a Given Portfolio". In: *Econometrica* 57.5, pp. 1121–1152.

Newey, W. K. and K. D. West (1987). "A Simple, Positive-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708.

Newey, W. K. and K. D. West (1994). "Automatic Lag Selection in Covariance Matrix Estimation". In: *Review of Economic Studies* 61.4, pp. 631–653.

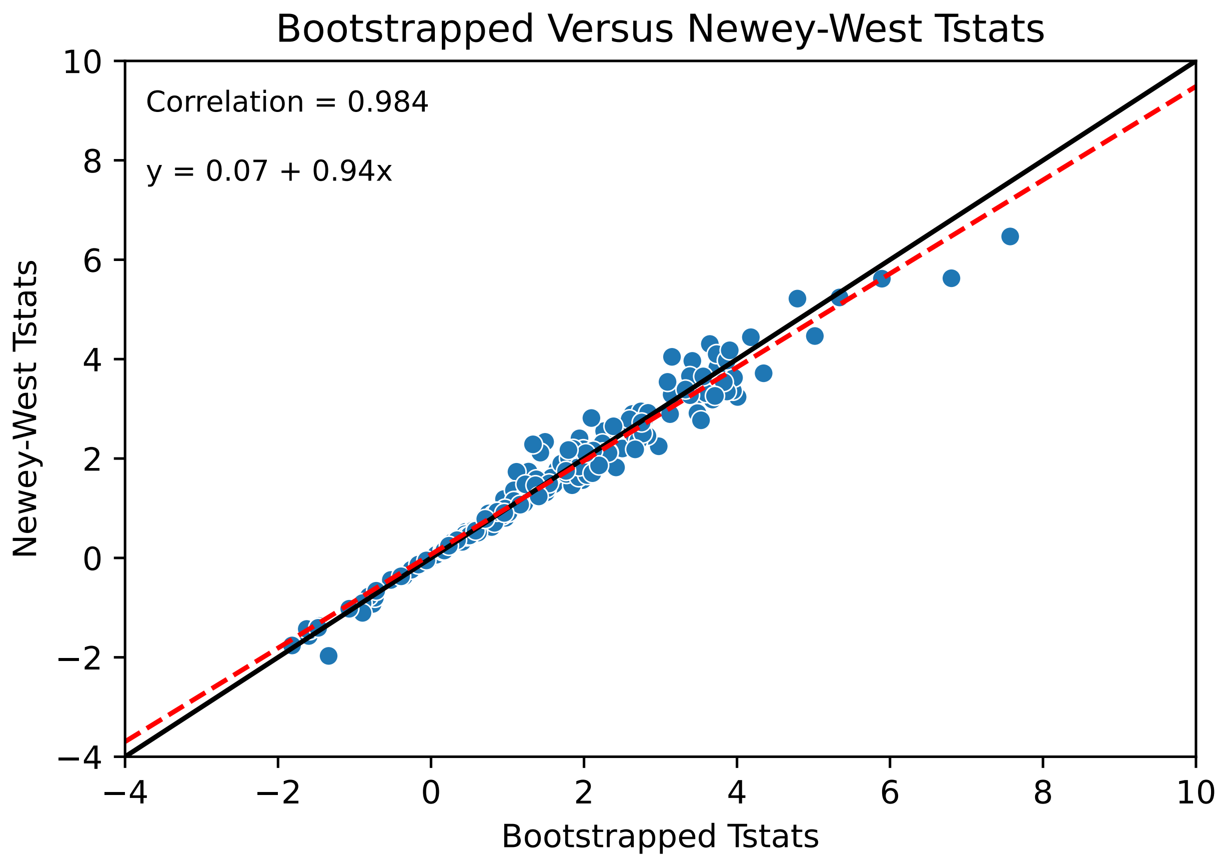


Figure IA.1
Comparing Bootstrap and Newey-West t-stats of IS Alphas Post-Publication

The figure plots t-stats of IS alphas post-publication based on Newey and West (1987, 1994) (in the y-axis) and the bootstrap method described in Section B of this Internet Appendix (in the x-axis). In the main text, we focus on bootstrap t-statistics since we cannot compute t-statistics for OS alphas using Newey and West (1987, 1994). However, the graph shows that the two procedures produce similar t-statistics for IS alphas over the post-publication period. These results are discussed in Subsection 1.3.

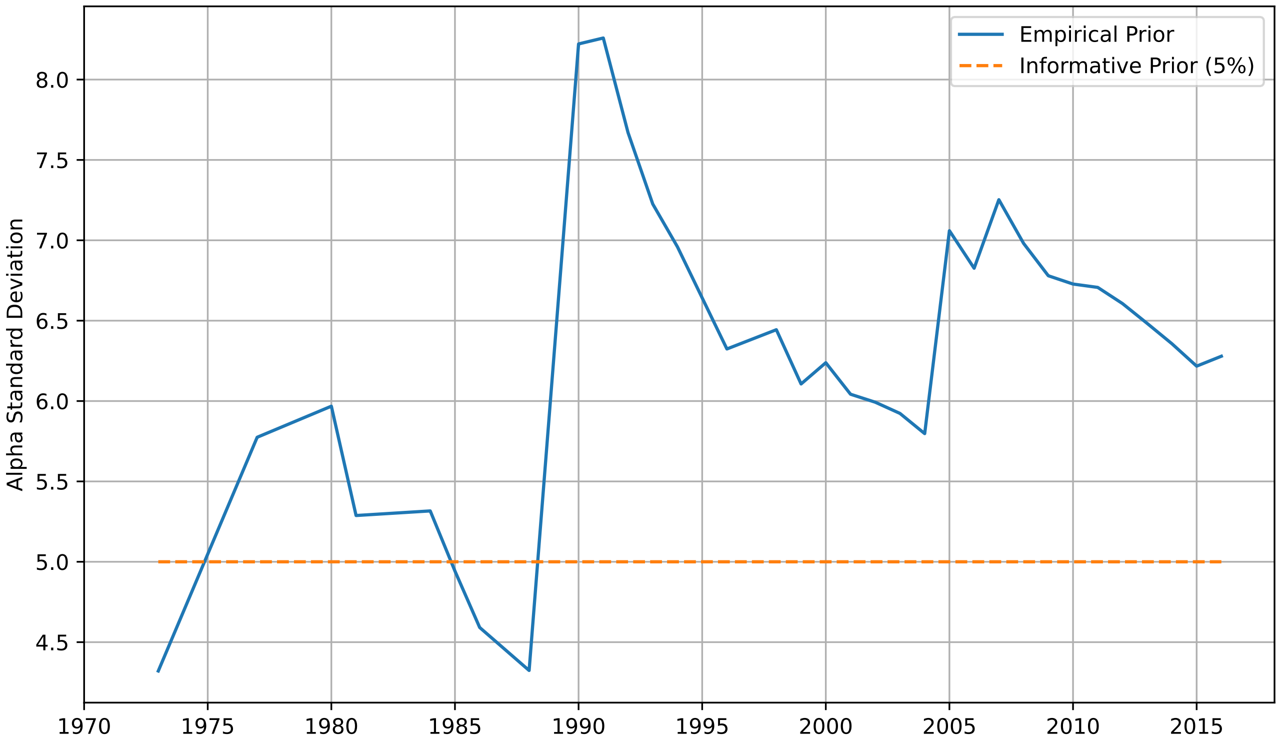


Figure IA.2

Comparing σ_α Values from our Empirical Bayesian Procedure with the Moderately Informative Prior

The figure plots the $\sigma_\alpha = 5.0\%$ of our moderately informative prior (in orange) against the time-series of the σ_α estimated from our empirical Bayesian procedure (in blue). The details of our Bayesian procedure are provided in Subsection 3.1 and these results are discussed in Subsection 3.2.