Timing Anomalies Through Investor Bias

Jonas Frey[1], Denis Mokanov[2], Kuntara Pukthuanthong[3]

October 7, 2025

Abstract

Motivated by a model linking time-varying expectation bias to anomaly returns, we construct a return predictor from the gap between analysts' earnings growth forecasts and unbiased machine-learning forecasts. Across 179 anomalies, our predictor outperforms the historical-mean benchmark in out-of-sample tests for up to 30% of cases at investment horizons beyond twelve months. Focusing on anomalies that exceed benchmark performance in a validation sample further improves performance. Consistent with our model, the persistence of portfolio-level bias determines whether an anomaly reflects the build-up or resolution of mispricing. A subset of anomalies shifts between these classifications, challenging static taxonomies.

*University of Gothenburg, Department of Economics, Sweden. Email: jonas.frey@economics.gu.se

†NHH Norwegian School of Economics, Bergen, Norway. Email: denis.mokanov@nhh.no

‡University of Missouri, Trulaske College of Business, Columbia, MO USA Email: pukthuanthongk@missouri.edu

1 Introduction

Capital market anomalies, which represent cross-sectional return patterns related to stock characteristics like book-to-market, size, momentum, or accruals, challenge risk-based models by generating returns that are not attributable to the covariance between anomaly and market returns. There is an ongoing debate about the source of these unexplained returns, with much work aimed at disentangling risk-based explanations from mispricing-based explanations (e.g., Pukthuanthong et al., 2019; Holcblat et al., 2022; Frey, 2023; and Böll et al., 2024).

An important feature of anomalies is that their profitability varies substantially over time. Anomaly returns exhibit both cyclical patterns and permanent decays (Groenborg et al., 2025). The temporal variation in returns has motivated a growing literature aimed at exploiting predictable patterns in anomaly returns (e.g., Moreira and Muir, 2017; Haddad et al., 2020; Ehsani and Linnainmaa, 2022; and Neuhierl et al., 2024). In this paper, we advance the study of anomaly timing by proposing a theoretically-motivated return predictor that captures investors' distorted beliefs about the growth of firms' earnings per share (EPS), aggregated at the portfolio level.

To motivate our analysis, we follow Bordalo et al. (2024a) by incorporating distorted cash flow growth expectations into a standard present value framework based on Campbell and Shiller (1988). A novel aspect of our model is that the relation between portfolio-level belief distortions and returns depends on the persistence of these distortions. This feature allows the model to generate dynamics consistent with anomaly profitability being driven by either mispricing being built up ("build-up" anomalies, in the terminology of van Binsbergen et al., 2023) or mispricing being corrected ("resolution" anomalies). Intuitively, if the persistence of belief distortions is low, overoptimistic expectations about cash flow growth lead to overpricing that is corrected in the subsequent period, thus generating negative returns. By contrast, if the distortions are highly persistent, already overoptimistic beliefs are expected to become even more optimistic in the future, producing a build-up of mispricing and positive subsequent returns.

Empirically, we analyze 179 value-weighted decile portfolios from Chen and Zimmermann (2021) over the period between June 1990 and November 2022. We examine the relation between our proxy for portfolio-level belief distortions, which we term the "*ex-ante* long-short bias," and anomaly returns at investment horizons of one, 12, 18, and 24 months, both in-sample and out-of-sample.

We construct our *ex-ante* bias measure as the difference between equity analysts' earnings per share (EPS) forecasts from IBES and unbiased machine learning forecasts obtained using the model proposed by Campbell et al. (2024), scaled by lagged earnings. This belief distortion proxy is available in real time, which allows investors to exploit biased growth expectations to time anomaly returns. An additional advantage of this measure is that it allows us to separate the relation between expected returns and forecast errors from the relation between unforecastable earnings-relevant future news and unexpected realized returns.[1]

To validate our framework, we first test whether the long–short bias measure predicts returns in-sample. Unconditionally, a statistically significant relation between bias and returns holds for 6.7% of anomalies at one month and 11.7% at 24 months. A noteworthy finding that emerges from these unconditional tests is that for the majority of the anomalies we study, long-short bias is positively correlated with future anomaly returns. This indicates that anomaly returns tend to be high following periods during which investors are overoptimistic about the growth prospects of stocks in the long leg of anomalies, consistent with the gradual build-up rather than the resolution of mispricing driving most anomaly returns.

As portfolio rebalancing can cause the portfolio-level relation between bias and returns to vary over time even in the presence of constant stock-level relations, we also conduct con-

[1] When favorable (unfavorable) news about a firm is revealed after an analyst's forecast is submitted, the forecast will appear *ex-post* too pessimistic (optimistic) and coincide with high (low) stock returns. This results in a strong negative correlation between *ex-post* analyst optimism and subsequent unexpected realized returns. However, since such news is unforeseeable at the time of the forecast, *ex-post* optimism does not reflect biased expectations.

ditional return predictability tests in the spirit of Lewellen and Nagel (2006). In particular, we split our sample into quarters and estimate the relation between bias and returns separately for each subsample. At the 24-month investment horizon, we document a statistically significant relation between returns and bias for up to 29% of the anomalies in our sample. An interesting result that emerges from the conditional tests is that the relation between bias and returns changes sign across subsamples for about 90% of the anomalies in our sample, which indicates that anomalies cannot be unconditionally classified as build-up or resolution anomalies, as in van Binsbergen et al. (2023). Instead, our findings indicate that most anomalies display behavior consistent with build-up in some periods and resolution in others.

We next analyze the cross-sectional relation between the persistence of long-short bias and the slope coefficient from the in-sample return predictability regression. Consistent with the predictions of our model, we observe a significant positive cross-sectional correlation, ranging from 0.136 (p-value$= 0.07$) at the one-month investment horizon to 0.214 (p-value< 0.01) at the 24-month investment horizon.

We further validate the relation between bias and returns using a standard out-of-sample return predictability exercise in the spirit of Goyal et al. (2024). For the median anomaly in our sample, our model produces slightly negative values for the out-of-sample coefficient of determination, R^2_{OOS}, across the 1-, 18-, and 24-month investment horizons, with a slightly positive value at the 12-month horizon. Additionally, F-statistics computed following McCracken (2007) indicate that our model forecasts returns more accurately than the historical mean return benchmark at the 5% significance level for roughly 30% of the anomalies. These results are consistent with our in-sample findings and confirm that our model performs well in predicting the returns of a subset of the anomalies in our sample.

Importantly, since our bias measure captures mispricing arising from biased cash flow growth expectations, it is conceptually suited to predicting returns only for anomalies driven by such mispricing. Consequently, the appropriate benchmark for evaluating our model is the proportion of mispricing-driven anomalies that can be timed using our model. While these relevant anomalies cannot be identified *ex-ante*, we use the historical performance of our model to identify them. Specifically, we first evaluate the model on a validation sample and then restrict the investment universe to anomalies for which the model outperforms the historical mean benchmark at the 5% significance level, based on the F-statistics of McCracken (2007). This restriction increases the median anomaly R^2_{OOS} to economically meaningful levels: 0.63% at the 12-month horizon, 1.30% at the 18-month horizon, and 0.95% at the 24-month horizon.

Our paper makes several contributions to the literature. Our theoretical contribution is most closely related to Bordalo et al. (2024a) and Bordalo et al. (2024b), which relate the Fama and French (2015) anomaly returns to distorted cash flow expectations. Our model deviates from the Bordalo et al. (2024a) model by not restricting the relation between expectation errors and anomaly returns to be negative, thus allowing our model to account for build-up anomalies. We also diverge from these papers empirically by using a proxy for *ex-ante* belief distortions, which allows us to distinguish build-up anomalies from resolution anomalies. In addition, we verify the robustness of our in-sample findings by conducting a series of out-of-sample tests.

Our paper is also related to several significant contributions in the fast-growing factor (anomaly) timing literature. Moreira and Muir (2017) propose using the realized volatility of factor returns to time investment in anomalies based on the weak risk-return trade-off. Haddad et al. (2020) propose using the book-to-market ratios of the principal components of factors to predict factor returns, while Ehsani and Linnainmaa (2022) show that past factor returns forecast future returns, which indicates the presence of factor momentum. More recently, Neuhierl et al. (2024) use a large vector of return predictors, including value, volatility, and momentum signals, and find that aggregating these predictions using dimension-reduction techniques, such as partial least squares, maximizes investors' ability to time anomaly returns. Our contribution to this literature is to propose a predictor of returns motivated by economic theory rather than statistical performance. We also propose an anomaly-timing approach that allows investors to use real-time data to alter their investment opportunity set, thereby improving the average predictive performance of their models.

We also extend the anomaly classification results of van Binsbergen et al. (2023) by showing that the majority of anomalies cannot be unconditionally identified as being driven by either the build-up or the resolution of mispricing. Instead, we find that anomalies display behavior consistent with build-up during certain periods and resolution during others.

Finally, our paper builds on the strand of literature that uses statistical models to predict equity analysts' EPS forecast errors. Notable contributions in this regard include So (2013), Van Binsbergen et al. (2022), de Silva and Thesmar (2023), Zhang et al. (2024), and, most importantly for our purposes, Campbell et al. (2024).

2 Theoretical Motivation

According to the return decomposition proposed by Campbell and Shiller (1988), the one-period log return obtained by investing in stock i, $r_{i,t+1}$ can be approximated by:

$$\alpha(p_{i,t+1} - d_{i,t+1}) + g_{i,t+1} - (p_{i,t} - d_{i,t}) + \kappa, \tag{1}$$

where $p_{i,t+1}$ is the log stock price at time $t+1$, $d_{i,t+1}$ is the log dividend, $g_{i,t+1} \equiv d_{i,t+1} - d_{i,t}$ is the dividend growth between time t and $t+1$, and $\kappa > 0$ and $\alpha \in (0,1)$ are constants that depend on the mean price-dividend ratio.[2]

Iterating Equation 1 forward, we obtain the following expression for the log price-dividend ratio:

$$p_{i,t} - d_{i,t} = \frac{\kappa}{1-\alpha} + \sum_{s=0}^{\infty} \alpha^s g_{i,t+1+s} - \sum_{s=0}^{\infty} \alpha^s r_{i,t+1+s}. \tag{2}$$

[2]The constant α is defined as $\frac{\exp(\bar{p}-\bar{d})}{1+\exp(\bar{p}-\bar{d})}$ where $\exp\left(\bar{p}-\bar{d}\right)$ is the mean price-dividend ratio. Given the typical levels of the price-dividend ratio in the data, α takes on the value of approximately 0.96.

The Campbell and Shiller (1988) decomposition is based on an accounting identity, i.e., it holds both in realizations and expectations:

$$p_{i,t} - d_{i,t} = \frac{\kappa}{1-\alpha} + \sum_{s=0}^{\infty} \alpha^s \mathbb{E}_t^{\mathbb{S}} \left[g_{i,t+1+s}\right] - \sum_{s=0}^{\infty} \mathbb{E}_t^{\mathbb{S}} \left[r_{i,t+1+s}\right], \tag{3}$$

Throughout this paper, we denote expectations formed by the market participants by an $\mathbb{S}$ superscript. We allow these beliefs to deviate from rational beliefs, which we denote by a $\mathbb{P}$ superscript.

We plug the expression in Equation 3 into Equation 1, which provides us with the following expression for expected returns under rational expectations:

$$\begin{aligned} \mathbb{E}_t^{\mathbb{P}}[r_{i,t+1}] = \mathbb{E}_t^{\mathbb{P}} \left[g_{i,t+1} - \mathbb{E}_t^{\mathbb{S}}[g_{i,t+1}]\right] + \mathbb{E}_t^{\mathbb{P}} \left[\sum_{s=1}^{\infty} \alpha^s \left(\mathbb{E}_{t+1}^{\mathbb{S}}[g_{i,t+1+s}] - \mathbb{E}_t^{\mathbb{S}}[g_{i,t+1+s}]\right)\right] + \\ + \mathbb{E}_t^{\mathbb{S}}[r_{i,t+1}] - \mathbb{E}_t^{\mathbb{P}} \left[\sum_{s=1}^{\infty} \alpha^s \left(\mathbb{E}_{t+1}^{\mathbb{S}} \left[r_{i,t+1+s}\right] - \mathbb{E}_t^{\mathbb{S}} \left[r_{i,t+1+s}\right]\right)\right]. \end{aligned} \tag{4}$$

To put more structure on our framework, we assume that firm i's dividend growth contains a persistent component as in Bansal and Yaron (2004):

$$\begin{aligned} &g_{i,t+1} = \mu_i + x_{i,t} + \sigma_g \varepsilon_{i,t+1}, \\ &x_{i,t} = \rho x_{i,t-1} + \sigma_x e_{i,t}, \\ &\varepsilon_{i,t+1},\ e_{i,t+1} \overset{\text{iid}}{\sim} N(0,1). \end{aligned} \tag{5}$$

Following Bordalo et al. (2024a), we model market participants' belief distortions directly. In particular, we assume that the s-period ahead subjective dividend growth expectations are given by:

$$\begin{aligned} &\mathbb{E}_t^{\mathbb{S}}[g_{i,t+s}] = \mathbb{E}_t^{\mathbb{P}}[g_{i,t+s}] + \frac{\zeta}{(1+s)^p} \nu_{i,t}, \\ &\nu_{i,t} = \varphi \nu_{i,t-1} + \sigma_\nu \eta_{i,t}, \end{aligned} \tag{6}$$

where the p parameter controls the impact of belief distortion shocks across forecasting horizons, $\varphi < 1$ represents the persistence of the distortions, and $\eta_{i,t}$ are standard shocks orthogonal to the fundamental shocks $\varepsilon_{i,t}$ and $e_{i,t}$.

Given the focus of our paper on distorted cash flow expectations, we further assume that agents have rational expectations regarding the evolution of risk premia:

$$\begin{aligned} &\mathbb{E}_t^{\mathbb{S}}[r_{i,t+1}] = \mathbb{E}_t^{\mathbb{P}}[r_{i,t+1}] = \mathbb{E}^{\mathbb{P}}[\varrho_{i,t+1}], \\ &\varrho_{i,t+1} = \phi\varrho_{i,t} + \sigma_\varrho e_{i,t+1}^{\varrho}, \end{aligned} \tag{7}$$

where $e_{i,t+1}^{\varrho}$ are standard normal shocks uncorrelated with the dividend growth and belief distortion shocks.

If market participants' expectations about dividend growth were fully rational, expected returns would be equal to $\phi\mathbb{E}_t^{\mathbb{P}}[\varrho_{i,t}]$.[3] In contrast, in our framework, return predictability arises from predictable dividend growth expectation errors and revisions, in addition to time-varying risk premia. The following proposition details the relation between expected returns and belief distortions in our model:

Proposition 1. *Expected returns at time t are given by:*

$$\mathbb{E}_t^{\mathbb{P}}[r_{i,t+1}] = \phi\mathbb{E}_t^{\mathbb{P}}[\varrho_{i,t}] + \frac{Li_p(\alpha)\zeta\left(\varphi\alpha - 1\right) + \alpha(\zeta - 1)}{\alpha}\nu_{i,t}, \tag{8}$$

where $Li_p(\alpha)$ is the polylogarithm function.

Proof. See Appendix A. □

This proposition shows that the sign of the relation between belief distortions and returns is determined by the relation between the persistence of belief distortions and the $\frac{\zeta}{(1+s)^p}$, which controls the term structure of belief distortions:

- If $\varphi < \frac{\alpha - \alpha\zeta + \zeta Li_p(\alpha)}{\zeta Li_p(\alpha)\alpha}$, expected returns decrease after positive distortion shocks. Following a positive η shock, stock i becomes overvalued. The mispricing is then gradually

[3] Under fully rational expectations, agents form expectations about risk premia using the Kalman filter: $\mathbb{E}_t[\varrho_{i,t}] = \phi\mathbb{E}_{t-1}[\varrho_{i,t-1}] + K\big(r_{i,t} - \phi\mathbb{E}_{t-1}[\varrho_{i,t-1}]\big)$, where K represents the steady-state Kalman gain.

corrected in subsequent periods, which leads to low expected returns.

- Conversely, if $\varphi > \frac{\alpha-\alpha\zeta+\zeta Li_p(\alpha)}{\zeta Li_p(\alpha)\alpha}$, expected returns increase following positive distortion shocks. Intuitively, when the persistence of belief distortions is sufficiently high relative to $\frac{\zeta}{(1+s)^p}$, beliefs are expected to become more distorted in subsequent periods. In other words, overvalued stocks are anticipated to become even more overvalued, which leads to these stocks earning higher returns.

Given the expression in Equation 8, the expected return of a generic long-short anomaly of the type considered in this paper is:

$$\mathbb{E}_t^{\mathbb{P}}[r_{LMS,t+1}] = \phi\mathbb{E}_t^{\mathbb{P}}\left[(\varrho_{L,t} - \varrho_{S,t})\right] + \frac{Li_p(\alpha)\zeta\left(\varphi\alpha - 1\right) + \alpha(\zeta - 1)}{\alpha}\left(\nu_{L,t} - \nu_{S,t}\right), \qquad (9)$$

where the L subscript indicates the long leg of the anomaly and S indicates the short leg. For the remainder of this paper, we use $\Delta_c x_t \equiv x_{L,t} - x_{S,t}$ to denote the cross-sectional difference operator.

3 Data and Methodology

3.1 Data

Throughout this paper, we follow an extensive prior literature (e.g., De la O and Myers, 2021; Bordalo et al., 2024a; and Bordalo et al., 2024b) and assume that the analyst forecasts provided by the Institutional Brokers' Estimate System (IBES) provide reasonable proxies for market participants' beliefs at large. Following Bordalo et al. (2024b), we further assume that payout ratios are constant over time and that dividends are expected to grow at the same rate as earnings. This assumption allows us to focus our analysis on equity analysts' earnings-per-share (EPS) growth forecasts instead of their dividend per share (DPS) forecasts, which do not become widely available until 2002. Restricting our sample to the post-2002 period would significantly reduce the scope of our analysis.

In our analysis, we focus on equity analysts' next-fiscal-year forecasts as previous research has shown that the advantage of machine learning forecasts over analysts is maximized at longer forecasting horizons (e.g., de Silva and Thesmar, 2023 and Zhang et al., 2024). We obtain the median forecasts from the IBES Unadjusted Summary file for the period between 1983 and 2022. IBES identifies the next-fiscal-year forecasts by a fiscal period indicator (FPI) of 2. Given our focus on analysts' consensus forecasts, we restrict our sample to median forecasts based on at least three individual forecasts. Following standard practice in the literature, we adjust the EPS forecasts for stock splits using the CRSP adjustment factor (Diether et al., 2002).

To construct long-short anomaly portfolios, we utilize the 179 characteristics from Chen and Zimmermann (2021), which are used to form value-weighted decile portfolios.[4] These characteristics are combined with monthly stock returns from CRSP to calculate monthly anomaly returns. We then compound the monthly returns to obtain the long-horizon returns ($h = 12, 18, 24$) that serve as the foundation of our analysis. We focus on simple portfolio returns rather than log returns.

3.2 Variable Construction

Our proxy for rational earnings growth expectations is constructed using the machine learning model proposed by Campbell et al. (2024). Campbell et al. (2024) systematically evaluate a comprehensive set of machine learning approaches to determine the model that achieves the highest accuracy in predicting earnings per share.[5][6] We initially train the model on data from 1983 to 1990, with the first set of machine learning forecasts generated in June 1990. Our sample period extends through November 2022. In Appendix B, we provide further

[4] The data can be found at `https://www.openassetpricing.com`.

[5] The code utilized in Campbell et al. (2024) was obtained from `https://bitly.cx/uE5q`.

[6] We verify that using the most accurate model identified by Campbell et al. (2024) does not introduce look-ahead bias by confirming its superior performance within our training sample.

implementation details, including the set of features used to predict earnings, and show that the forecasts generated by the model are (almost) unbiased and significantly more accurate than the analyst forecasts.

To construct anomaly returns, we sort stocks with available earnings forecasts into deciles based on the firm characteristics from Chen and Zimmermann (2021). For each decile, we calculate the value-weighted average return. The h-period return for anomaly j is then defined as the difference between the average returns of the tenth and first deciles:

$$R_{j,t+h} = \frac{\sum_{i\in D10} \omega_{i,t} R_{i,t+h}}{\sum_{i\in D10} \omega_{i,t}} - \frac{\sum_{i\in D1} \omega_{i,t} R_{i,t+h}}{\sum_{i\in D1} \omega_{i,t}},$$

where $\omega_{i,t}$ denotes the market capitalization of stock i at the beginning of month t.

Given the set of median analyst forecasts from IBES and the unbiased forecasts delivered by the Campbell et al. (2024) model, we construct the empirical counterpart of our belief distortion measure as:

$$\nu_{i,t} = \frac{\mathbb{F}_t[e_{i,\tau+1}] - \mathbb{E}_t^{ML}[e_{i,\tau+1}]}{|e_{i,\tau-1}|}, \tag{10}$$

where e represents earnings per share, i indexes firms, τ refers to fiscal years (with the current year being τ), and t indexes calendar months. The operator $\mathbb{F}[\cdot]$ represents the empirical counterpart of $\mathbb{E}^{\mathbb{S}}[\cdot]$, and $\mathbb{E}^{ML}[\cdot]$ corresponds to the empirical counterpart of $\mathbb{E}^{\mathbb{P}}[\cdot]$. For the rest of this paper, we refer to ν as "bias." The particular definition in Equation 10 facilitates the interpretation of our results, as positive values of ν always correspond to consensus forecasts being overly optimistic, independent of the sign of lagged earnings.

We then aggregate stock-level biases into portfolio-level measures by value-weighting individual stock biases. In particular, the month t bias for anomaly j is computed as:

$$\Delta_c \nu_{j,t} = \frac{\sum_{i\in D10} \omega_{i,t} \nu_{i,t}}{\sum_{i\in D10} \omega_{i,t}} - \frac{\sum_{i\in D1} \omega_{i,t} \nu_{i,t}}{\sum_{i\in D1} \omega_{i,t}}.$$

We present the time-series behavior of the cross-sectional median long–short bias across the 179 anomalies in our sample in Figure 1. On average, long-short bias tends to be negative for the majority of anomalies, which indicates that analysts tend to overestimate the earnings growth of stocks in the short leg of anomalies relative to those in the long leg. This finding is consistent with Engelberg et al. (2018).

The figure also shows that the bias typically increases ahead of NBER-defined recessions. For example, our measure identifies the build-up to the dot-com bubble as a period of elevated bias. During recessions, the bias remains high initially but declines as the downturn progresses, suggesting that analyst expectations adjust to macroeconomic conditions with a lag, in line with Gómez-Cram (2022). Following recessions, the bias remains subdued for extended periods and even turns positive for long stretches following the 1990 recession. The median long-short bias is also significantly negatively correlated with the Baker and Wurgler (2006) sentiment index ($\rho = -0.23$, t-statistic $= -4.62$).

The median long-short bias across the 179 anomalies exhibits a gradual decline in magnitude over the sample period from June 1990 to November 2022. It starts at approximately -0.12 in the early 1990s and approaches zero by 2020, indicating that analysts' relative overestimation of short-leg earnings growth has diminished over time. This trend coincides with the observed reduction in anomaly profitability documented in the literature (Chordia et al., 2014), suggesting a weakening of mispricing as belief distortions become less pronounced.

4 Model Validation

4.1 In-sample Return Predictability Results

Before examining the broader implications of our framework for return predictability and portfolio management, we first validate its core predictions. Specifically, Equation 9 provides a direct testable implication for the relation between portfolio bias and returns. We test this relation separately for each anomaly j by estimating the following regression using overlapping returns of 1, 12, 18, and 24 months:

$$R_{j,t+h} = a_{j,h} + b_{j,h}\Delta_c \nu_{j,t} + \varepsilon_{j,t+h}. \tag{11}$$

In Panel A of Figure 2, we report the proportion of anomalies for which the slope coefficient is unconditionally estimated to be positive (blue) and negative (red). The darker segments represent the fraction of anomalies for which the slope coefficients are statistically different from zero at the 5% level, based on t-statistics computed following Newey and West (1987) with lag lengths equal to $h-1$.

Overall, the results provide modest support for an unconditional relation between bias and returns. The fraction of slope coefficients statistically different from zero ranges from 6.7% at the one-month horizon to 11.7% at the 24-month horizon. The fraction of significant slope coefficients exceeds the roughly 5% expected under the null of no bias-return relation, but these figures should be interpreted cautiously given the lack of independence across anomaly-specific estimates. The increasing significance of the bias-return relation at longer investment horizons suggests that belief distortions gradually influence asset prices over time, with their impact becoming more pronounced as the horizon extends.

However, there is no *ex-ante* reason to expect the relation between long–short bias and anomaly returns to remain stable over time. Even if the relation were constant at the stock level, turnover within long–short portfolios would induce time variation at the portfolio level. These considerations lead us to test a conditional version of the model, in which the bias–return relation is allowed to vary over time.

To conditionally test the relation between bias and returns, we adopt an approach in the spirit of Lewellen and Nagel (2006): we split our sample into four quarters and estimate the model in Equation 11 separately for each subsample.[7] The results of this estimation are reported in Figure 3.

These results provide significantly stronger support for the conditional version of our model. At longer investment horizons, at least 12% (12-month horizon) and up to 29%

[7] The cutoffs for the four subsamples are June 1998, August 2006, and September 2014.

(24-month horizon) of the slope coefficients are estimated to be different from zero at the 5% significance level. The evidence strongly supports a time-varying relation between bias and returns. The time-varying nature of the relation is further underscored by the fact that only 12% to 13% of the slope coefficients are estimated to have the same sign across all four subsamples.

Several additional patterns emerge in Figure 3. First, the share of statistically positive coefficients declines over time, while the share of statistically negative coefficients increases. This shift helps explain the weaker unconditional results reported in Panel A of Figure 2.

Second, the prevalence of positive slope coefficients, particularly in the first half of the sample, indicates that for most anomalies, periods when investors are overoptimistic about the growth prospects of stocks in the long leg relative to those in the short leg tend to be followed by high portfolio returns. This pattern contradicts the view that anomaly returns are primarily driven by the gradual correction of mispricing. Instead, the majority of anomalies exhibit behavior consistent with what van Binsbergen et al. (2023) describe as "build-up" anomalies: already overvalued anomaly longs are expected to become even more overvalued in subsequent periods, which leads to positive returns. Our results extend the insights of van Binsbergen et al. (2023) by showing that most anomalies cannot be classified as static build-up or resolution types. For example, many anomalies resemble build-up anomalies in the pre-1999 period but shift toward resolution-type behavior after 2014.

Overall, the in-sample tests show that long-short bias predicts returns for only a subset of the anomalies in our sample. Several factors may explain this result. First, the majority of the anomalies in our sample may reflect time-varying risk premia rather than mispricing (Frey, 2023). Second, differences between investors' expectations and analysts' forecasts, such as those arising from agency conflicts (Ottaviani and Sørensen, 2006; Gemmi and Valchev, 2023), may obscure the relation between bias and returns. Third, our proxy may fail to capture the entire extent of the bias due to the limitations of the machine learning model used to construct the unbiased earnings forecasts. Given points two and three, we view the predictability estimates reported here as a lower bound on the extent to which bias can predict returns.

While the in-sample regression results in Figure 2 and Figure 3 provide initial support for our framework, we leverage a sharper prediction of our model to further substantiate its theoretical foundation. Specifically, our model predicts that anomalies with greater persistence in bias are associated with more positive slope coefficients in Equation 11.

Figure 4 illustrates the cross-sectional relation between b and φ across the 179 anomalies in our sample, evaluated at investment horizons of 1, 12, 18, and 24 months. Each dot represents a distinct anomaly, with the best-fit line overlaid and the corresponding correlation coefficient, ρ, reported.

The correlation coefficients are estimated to be 0.136 (p-value = 0.07), 0.142 (p-value = 0.06), 0.202 (p-value < 0.01), and 0.214 (p-value < 0.01) for the 1-, 12-, 18-, and 24-month investment horizons, respectively. Importantly, the findings remain robust when excluding anomalies for which φ is estimated to be negative. These results provide strong support for the relation between bias persistence and return predictability underpinning our framework.

In Appendix C.2, we replace the empirical bias series for each portfolio with simulated AR(1) processes calibrated to match the persistence, mean, and volatility of the empirical biases. This exercise confirms that the reported cross-sectional correlations in Equation 11 are unlikely to be artifacts of our estimation procedure. As shown in Figure A3, at the 12-month investment horizon, fewer than 0.5% of the 10,000 simulations produce a cross-sectional correlation coefficient exceeding 0.142 in absolute value.

4.2 Predictability Using Ex-Post and Ex-Ante Bias

A notable feature of our unconditional in-sample predictability results is the roughly equal distribution of positive and negative slope coefficients across the anomalies in our sample. However, at investment horizons beyond 12 months, a significantly larger fraction of the positive coefficients is statistically significant. This contrasts sharply with prior research, such as Bordalo et al. (2024a), which uses bias measures based on *ex-post* earnings realizations and typically finds negative slope coefficients. In Panel B of Figure 2, we reproduce our unconditional tests using *ex-post* bias measures and explore the sources of this divergence in detail in this section.

To do so, we decompose the *ex-post* forecast error into two components: a predictable *ex-ante* component, which reflects information available at the time forecasts are made and is the primary focus of our analysis, and an unforeseeable earnings surprise, which captures the portion of earnings realizations that cannot be anticipated based on available information. It is the correlation between the unforecastable earnings surprise and unexpected returns that drives the divergence between our results and those obtained using *ex-post* bias measures.

Specifically, the sign of the *ex-post* return predictability coefficient depends on the sign of the covariance term:

$$\operatorname{cov}\left(\mathbb{E}_t^{\mathbb{P}}[R_{j,t+h}] + \varepsilon_{j,t+h}^{\perp}, \Delta_c\left(\frac{\mathbb{F}_t[e_{j,\tau+1}] - \mathbb{E}_t^{\mathbb{P}}[e_{j,\tau+1}] - \varepsilon_{j,\tau+1}^{surprise}}{e_{j,\tau-1}}\right)\right), \tag{12}$$

where $\varepsilon_{j,t+h}^{\perp}$ is the unexpected component of realized returns uncorrelated with information at time t, and $\Delta_c\varepsilon_{j,\tau+1}^{surprise}$ captures the cumulative effect of earnings-related news released between t and the earnings announcement period $t+T$:

$$\varepsilon_{j,\tau+1}^{surprise} \equiv e_{j,\tau+1} - \mathbb{E}_t^{\mathbb{P}}[e_{j,\tau+1}] = e_{j,\tau+1} - \sum_{m=0}^{T-1}\left(\mathbb{E}_{t+m+1}^{\mathbb{P}}[e_{j,\tau+1}] - \mathbb{E}_{t+m}^{\mathbb{P}}[e_{j,\tau+1}]\right). \tag{13}$$

The covariance in Equation 12 can be rewritten as:

$$\operatorname{cov}\left(\mathbb{E}_t[R_{j,t+h}], \Delta_c\nu_{j,t}\right) - \operatorname{cov}\left(\varepsilon_{j,t+h}^{\perp}, \frac{\Delta_c\varepsilon_{j,\tau+1}^{surprise}}{e_{j,\tau-1}}\right). \tag{14}$$

Corporate news is released gradually between t and $t+T$, market participants continuously update earnings expectations, and the unexpected component of returns reflects these updates.[8] This produces a strong positive covariance between unexpected returns and earn-

[8] This aligns with Engelberg et al. (2018), who find that anomaly returns are about 50% higher on

ings surprises. Intuitively, if positive (negative) news arrives after time t, it will result in positive (negative) unexpected stock returns and cause analysts' forecasts made at time t to *ex-post* appear to be too pessimistic (optimistic). This covariance between the unforeseeable components of *ex-post* forecast errors and realized returns obscures the relation between expected returns and *ex-ante* bias that we document.

5 Portfolio Bias and Anomaly Return Predictability

The out-of-sample analysis developed in this section complements the in-sample results presented in section 4 and provides further evidence in favor of the framework proposed in this paper.

A key advantage of the out-of-sample approach is that the forecasting regressions specified in Equation 11 are estimated recursively, allowing the model's coefficients to capture the time-varying relation between portfolio bias and returns documented in Section subsection 4.1.

Additionally, the out-of-sample approach provides valuable insights into the ability of market participants to exploit the relation between bias and portfolio returns to enhance the profitability of their investments.

5.1 Out-of-sample Predictability Results

We adopt the standard approach from the out-of-sample return predictability literature (e.g., Haddad et al., 2020 and Neuhierl et al., 2024) by dividing our sample into two distinct periods: an initial training period and a test period. The initial training period spans roughly the first 65% of the full sample. For anomalies with data available throughout the entire sample, the initial training period ends between June 2009 (for the 24-month investment horizon) and May 2011 (for the 1-month investment horizon). We use the remaining data, from June 2011 to November 2022, to test our model. We examine the robustness of our

corporate news days.

findings to alternative sample splits in Appendix C.1.

Starting from this initial training period, we expand the training window by adding one month at a time and re-estimate the following linear model separately for investment horizons of 1, 12, 18, and 24 months:

$$R_{j,t+h} = a_{j,h} + b_{j,t,h}\Delta_c \nu_{j,t} + \varepsilon_{j,t+h} \tag{15}$$

Each time we re-estimate the model, we calculate a new slope coefficient, $\hat{b}_{t,h}$, which is then used to predict h-period cumulative returns as $\hat{R}_{j,t+h} = \hat{a}_{j,h} + \hat{b}_{j,t,h}\Delta_c \nu_{j,t}$. Since the intercept estimate is not the primary focus of our analysis, we reduce the noise in our estimation procedure by estimating the intercept only once using our initial training sample. We then keep it fixed for the duration of our test sample.

To gauge the ability of long-short bias to predict anomaly returns out-of-sample, we compare the ability of the model in Equation 15 to predict anomaly returns to an intercept-only model:

$$R_{j,t+h} = a_{j,t,h} + \varepsilon_{j,t+h} \tag{16}$$

where $a_{j,t} \equiv \frac{\sum_{k=0}^{t-h} R_{j,t+h-k}}{t-h}$ represents average historical h-period returns of anomaly j until period $t-h$.

Given the return predictions of the two models, we construct the out-of-sample coefficient of determination (R^2_{OOS}) that informs us about the relative accuracy of the models in predicting future returns (Campbell and Thompson, 2008):

$$R^2_{OOS,j,h} = 1 - \frac{\sum_{t \in \text{test sample}} (R_{j,t+h} - \hat{R}_{j,t+h})^2}{\sum_{t \in \text{test sample}} (R_{j,t+h} - \bar{R}_{j,t})^2}. \tag{17}$$

To determine the significance of our return predictability results, we also compute the

F-statistic proposed by McCracken (2007):

$$(T_{OOS} - h + 1)\,\frac{\sum_{t\in\text{test sample}}(R_{j,t+h} - \bar{R}_{j,t+h})^2 - \sum_{t\in\text{test sample}}(R_{j,t+h} - \hat{R}_{j,t+h})^2}{\sum_{t\in\text{test sample}}(R_{j,t+h} - \hat{R}_{j,t+h})^2} \tag{18}$$

we determine significance at the 5% level using the cutoff $|F\text{-statistic}| > 1.554$.

We report the middle 90 percent of the cross-sectional distribution of $R_{OOS,j}$ estimates for investment horizons of 1, 12, 18, and 24 months in Panel A of Figure 5. In the figure, we highlight median values using darker dots.

The median values of the out-of-sample coefficient of determination are close to zero across all investment horizons, ranging from -0.48% at the 18-month horizon to 0.08% at the 12-month horizon. These results suggest that, on average, our model forecasts portfolio returns with accuracy comparable to that of the historical mean benchmark. This conclusion is further supported by the McCracken (2007) F-statistic results presented in Panel B of Figure 5, which show that our model outperforms the benchmark at the 5% level for approximately 23% to 30% of the anomalies in our sample at investment horizons exceeding 12 months.

Overall, the evidence suggests that the model performs well in predicting returns for certain anomalies, achieving an R^2_{OOS} of over 9% at the 24-month horizon. Additionally, our model yields coefficients of determination of over 5% across investment horizons of over 12 months for well-known anomalies such as the Fama and French (1993) book-to-market and the Hou et al. (2014) return on equity anomalies. However, an investor applying our model indiscriminately across all anomalies in the sample would, on average, underperform slightly relative to the historical mean benchmark. We address this point further in subsection 5.2.

This result aligns with the in-sample return predictability findings discussed in subsection 4.1. Additionally, it is consistent with our prior expectations, as the profitability of most anomalies is unlikely to be driven by mispricing (Frey, 2023). Moreover, our bias measure may not capture all forms of mispricing accurately. To establish an upper bound on our model's ability to predict returns, we repeat the out-of-sample return predictability analysis using an *ex-post* measure of bias based on realized earnings. Using this approach, we find that our model outperforms the historical mean at the 5% level for 65% of anomalies at the 12-month horizon and 58% at the 24-month horizon despite the benefit of hindsight.

Finally, we verify that the deviations of our methodology from the "homologous" specification of Goyal et al. (2024) do not mechanically drive the return predictability results reported in this section. To this end, we conduct a simulation exercise described in Appendix C.2. The simulation results indicate that it is highly unlikely for the findings in this section to arise by chance or to be artifacts of our estimation methodology.

5.2 Restricted Sample Results

A shortcoming of our analysis so far has been the assumption that investors cannot restrict their investment opportunity set to anomalies that can be timed using our model.

To remedy this shortcoming, we propose an alternative investment approach, which involves splitting our sample into three: we use the first half of our sample to train our model, as in subsection 5.1. We then use the next 15% of our sample as a validation sample to determine the subset of anomalies that our model is capable of timing. Then, we use the final 35% of our sample to test the model on these anomalies. During the test stage, we only consider anomalies with McCracken (2007) F-statistics above 1.554 in the validation sample. This leaves us with 29 anomalies at the 12-month investment horizon and 30 anomalies at the 18- and 24-month horizons.

Once we identify the anomalies, we repeat the exercises from the previous section using our restricted sample. We report our findings based on the restricted sample in Figure 6. In Panel A, we show the cross-sectional distribution of the coefficients of determination after the removal of the minimum and maximum values, and in Panel B, we show the proportion of positive McCracken (2007) F-statistics.

We find that restricting the investment opportunity set has an economically meaningful impact on our conclusions. The median R^2_{OOS} estimates increase to between 0.63% at the 12-month horizon and 1.3% at the 18-month horizon. Additionally, about 60% of the anomalies generate McCracken (2007) F-statistics of over 1.554, which is a marked improvement over the results in Figure 5. Overall, the evidence suggests that investors can benefit from using past data to gauge the ability of our model to predict returns.

6 Conclusion

In this paper, we propose a theoretically motivated proxy for the *ex-ante* distortions of investors' cash flow expectations and test its ability to predict returns across a broad cross-section of 179 anomaly portfolios, both in- and out-of-sample.

We find that, in-sample, our proxy forecasts returns for a minority of anomalies (up to 29% in conditional tests) at investment horizons of over 12 months. In-sample results further show that anomaly returns tend to be high following periods in which investors are overoptimistic about cash flow growth in the long legs of anomalies. This pattern indicates that most anomalies behave in a manner consistent with build-up rather than resolution, contrary to conventional intuition.

Out-of-sample tests confirm the findings of our in-sample tests, with our model statistically outperforming the historical mean benchmark for up to 30% of the anomalies in our sample at investment horizons of over 12 months. We further show that the predictive performance of our proxy improves significantly when the investment opportunity set is restricted to anomalies that have historically been predictable using our measure.

The fact that our bias proxy predicts the returns of only a subset of anomalies underscores the limitations of our approach. Our model's inability to predict returns for certain anomalies may stem from a combination of factors, including the anomalies being driven by risk exposure, discrepancies between analysts' forecasts and investors' expectations arising from agency conflicts, or limitations in the unbiased earnings forecasts used in this study.

Further research is required to disentangle the contributions of these factors.

References

Baker, Malcolm and Jeffrey Wurgler (2006). "Investor Sentiment and the Cross-Section of Stock Returns". In: *The Journal of Finance* 61.4, pp. 1645–1680. DOI: https://doi.org/10.1111/j.1540-6261.2006.00885.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2006.00885.x.

Bansal, Ravi and Amir Yaron (2004). "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles". In: *The Journal of Finance* 59.4, pp. 1481–1509. DOI: https://doi.org/10.1111/j.1540-6261.2004.00670.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2004.00670.x.

Böll, Julian et al. (2024). *Following the Footprints: Towards a Taxonomy of the Factor Zoo*. URL: https://papers.ssrn.com/abstract=4702435 (visited on 02/08/2024). Pre-published.

Bordalo, Pedro et al. (2024a). "Belief overreaction and stock market puzzles". In: *Journal of Political Economy* 132.5, pp. 1450–1484.

Bordalo, Pedro et al. (2024b). *Finance without exotic risk*. Tech. rep. National Bureau of Economic Research.

Campbell, John L. et al. (2024). *Expectations Matter: When (not) to Use Machine Learning Earnings Forecasts*. Working Paper.

Campbell, John Y and Samuel B Thompson (2008). "Predicting excess stock returns out of sample: Can anything beat the historical average?" In: *The Review of Financial Studies* 21.4, pp. 1509–1531.

Campbell, John Y. and Robert J. Shiller (1988). "Stock Prices, Earnings, and Expected Dividends". In: *The Journal of Finance* 43.3, pp. 661–676. ISSN: 00221082, 15406261.

Chen, Andrew Y. and Tom Zimmermann (2021). *Open Source Cross-Sectional Asset Pricing*. FEDS Working Paper 2021-040. Available at the Federal Reserve website. Board of Governors of the Federal Reserve System.

Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong (2014). "Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?" In: *Journal of Accounting and Economics* 58.1, pp. 41–58. ISSN: 0165-4101. DOI: https://doi.org/10.1016/j.jacceco.2014.06.001.

De la O, Ricardo and Sean Myers (2021). "Subjective Cash Flow and Discount Rate Expectations". In: *The Journal of Finance* 76.3, pp. 1339–1387. DOI: https://doi.org/10.1111/jofi.13016. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13016.

de Silva, Tim and David Thesmar (2023). "Noise in Expectations: Evidence from Analyst Forecasts". In: *The Review of Financial Studies* 37.5, pp. 1494–1537. ISSN: 0893-9454. DOI: 10.1093/rfs/hhad091. eprint: https://academic.oup.com/rfs/article-pdf/37/5/1494/57262318/hhad091.pdf.

Diether, Karl B., Christopher J. Malloy, and Anna Scherbina (2002). "Differences of Opinion and the Cross Section of Stock Returns". In: *The Journal of Finance* 57.5, pp. 2113–2141. ISSN: 00221082, 15406261.

Ehsani, Sina and Juhani T. Linnainmaa (2022). "Factor Momentum and the Momentum Factor". In: *The Journal of Finance* 77.3, pp. 1877–1919.

Engelberg, Joseph, R. David McLean, and Jeffrey Pontiff (2018). "Anomalies and News". In: *The Journal of Finance* 73.5, pp. 1971–2001. DOI: https://doi.org/10.1111/jofi.12718. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12718.

Fama, Eugene F. and Kenneth R. French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of Financial Economics* 33.1, pp. 3–56. ISSN: 0304-405X. DOI: https://doi.org/10.1016/0304-405X(93)90023-5.

— (2015). "A five-factor asset pricing model". In: *Journal of Financial Economics* 116.1, pp. 1–22. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2014.10.010.

Frey, Jonas (2023). *Which Stock Return Predictors Reflect Mispricing? **. DOI: 10.2139/ssrn.4584121. Social Science Research Network: 4584121. URL: https://papers.ssrn.com/abstract=4584121 (visited on 02/28/2025). Pre-published.

Gemmi, Luca and Rosen Valchev (2023). *Biased Surveys*. Working Paper.

Gómez-Cram, Roberto (2022). "Late to Recessions: Stocks and the Business Cycle". In: *The Journal of Finance* 77.2, pp. 923–966. DOI: https://doi.org/10.1111/jofi.13100. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.13100.

Goyal, Amit, Ivo Welch, and Athanasse Zafirov (2024). "A Comprehensive 2022 Look at the Empirical Performance of Equity Premium Prediction". In: *The Review of Financial Studies* 37.11, pp. 3490–3557. ISSN: 0893-9454. DOI: 10.1093/rfs/hhae044. eprint: https://academic.oup.com/rfs/article-pdf/37/11/3490/59762851/hhae044.pdf.

Groenborg, Niels et al. (2025). "Time-Varying Anomaly Premia: Stable Fact or Disappearing Act?" In: *Working paper*.

Haddad, Valentin, Serhiy Kozak, and Shrihari Santosh (2020). "Factor timing". In: *The Review of Financial Studies* 33.5, pp. 1980–2018.

Holcblat, Benjamin, Abraham Lioui, and Michael Weber (2022). *Anomaly or Possible Risk Factor? Simple-To-Use Tests*. DOI: 10.2139/ssrn.4064159. URL: https://papers.ssrn.com/abstract=4064159 (visited on 07/05/2022). Pre-published.

Hou, Kewei, Chen Xue, and Lu Zhang (2014). "Digesting Anomalies: An Investment Approach". In: *The Review of Financial Studies* 28.3, pp. 650–705. ISSN: 0893-9454. DOI: 10.1093/rfs/hhu068. eprint: https://academic.oup.com/rfs/article-pdf/28/3/650/24450149/hhu068.pdf.

Ke, Guolin et al. (2017). "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc.

Lewellen, Jonathan and Stefan Nagel (2006). "The conditional CAPM does not explain asset-pricing anomalies". In: *Journal of Financial Economics* 82.2, pp. 289–314. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2005.05.012.

Lundberg, Scott M. and Su-In Lee (2017). "A unified approach to interpreting model predictions". In: *CoRR* abs/1705.07874. arXiv: 1705.07874.

McCracken, Michael W. (2007). "Asymptotics for out of sample tests of Granger causality". In: *Journal of Econometrics* 140.2, pp. 719–752. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2006.07.020.

Moreira, Alan and Tyler Muir (2017). "Volatility-Managed Portfolios". In: *The Journal of Finance* 72.4, pp. 1611–1644.

Neuhierl, Andreas et al. (2024). *Timing the Factor Zoo*. Tech. rep. Working Paper.

Newey, Whitney K. and Kenneth D. West (1987). “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”. In: *Econometrica* 55.3, pp. 703–708. ISSN: 00129682, 14680262.

Ottaviani, Marco and Peter Norman Sørensen (2006). “The strategy of professional forecasting”. In: *Journal of Financial Economics* 81.2, pp. 441–466. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2005.08.002.

Pukthuanthong, Kuntara, Richard Roll, and Avanidhar Subrahmanyam (2019). “A Protocol for Factor Identification”. In: *The Review of Financial Studies* 32.4, pp. 1573–1607. ISSN: 08939454, 14657368.

Shapley, Lloyd S (1953). “A Value for n-Person Games”. In: *Contributions to the Theory of Games II*. Ed. by Harold W. Kuhn and Albert W. Tucker. Princeton: Princeton University Press, pp. 307–317.

So, Eric C. (2013). “A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts?” In: *Journal of Financial Economics* 108.3, pp. 615–640. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2013.02.002.

Van Binsbergen, Jules H, Xiao Han, and Alejandro Lopez-Lira (2022). “Man versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases”. In: *The Review of Financial Studies*. ISSN: 0893-9454. DOI: 10.1093/rfs/hhac085. eprint: https://academic.oup.com/rfs/advance-article-pdf/doi/10.1093/rfs/hhac085/46698800/hhac085.pdf.

van Binsbergen, Jules H. et al. (2023). “Dynamic asset (mis)pricing: Build-up versus resolution anomalies”. In: *Journal of Financial Economics* 147.2, pp. 406–431. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2022.11.005.

Zhang, Yingguang, Yandi Zhu, and Juhani T Linnainmaa (2024). “Man versus Machine Learning Revisited”. In: *Available at SSRN 4899584*.

Tables and Figures

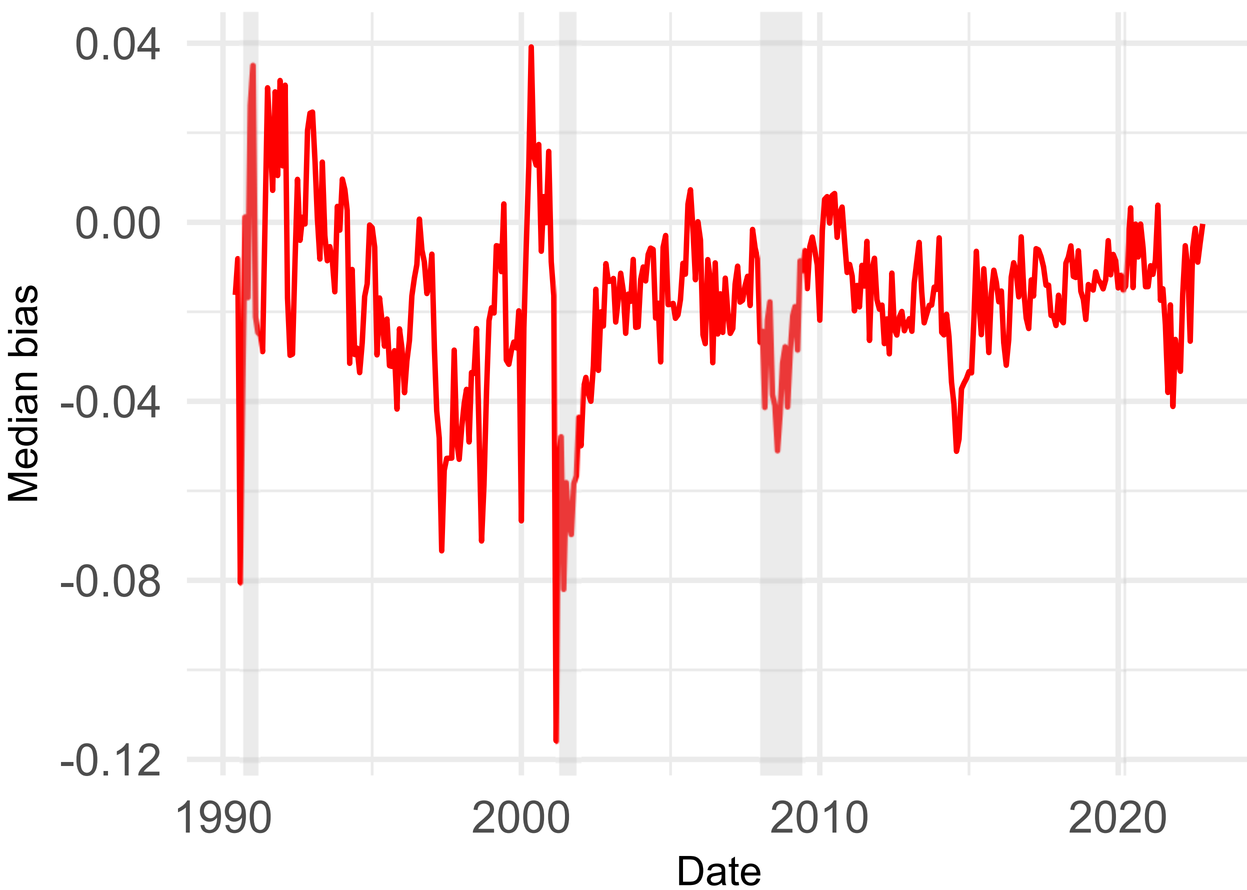
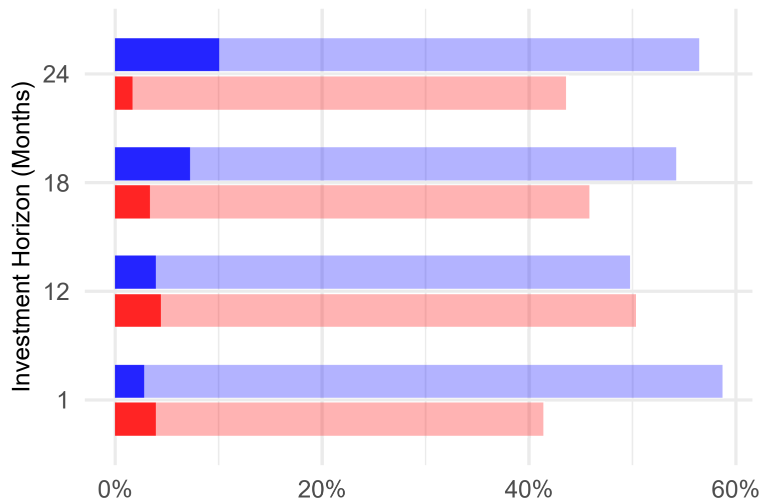
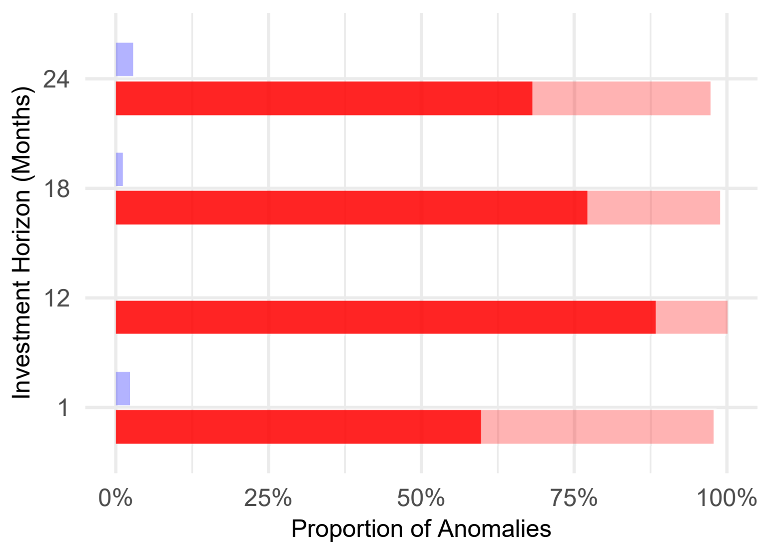


Figure 1: Time-Series Behavior of Bias

This figure illustrates the time-series behavior of the median long-short bias, $\Delta_c \nu_j$, across the 179 anomalies in our sample. Gray bars represent periods classified as recessions by the NBER.



Panel A: Ex-ante Bias Slope Coefficient Estimates, Unconditional



Panel B: Ex-post Bias Slope Coefficient Estimates, Unconditional

Figure 2: In-sample Results, Individual Anomalies

This figure presents results from the predictive regression estimated separately for each anomaly j and investment horizon h: $R_{j,t+h} = a + b\Delta_c\nu_{j,t} + \varepsilon_{j,t+h}$. Panel A reports the proportion of positive (blue bars) and negative (red bars) slope coefficients when the bias term $\Delta_c\nu_{j,t}$ is constructed using a machine learning algorithm. Panel B (the figure just generated) reports analogous results using ex-post earnings realizations to construct the bias measure. Darker segments within each bar indicate the share of coefficients that are statistically significant at the 5% level, based on t-statistics computed following Newey and West (1987) with lag lengths equal to $h - 1$.

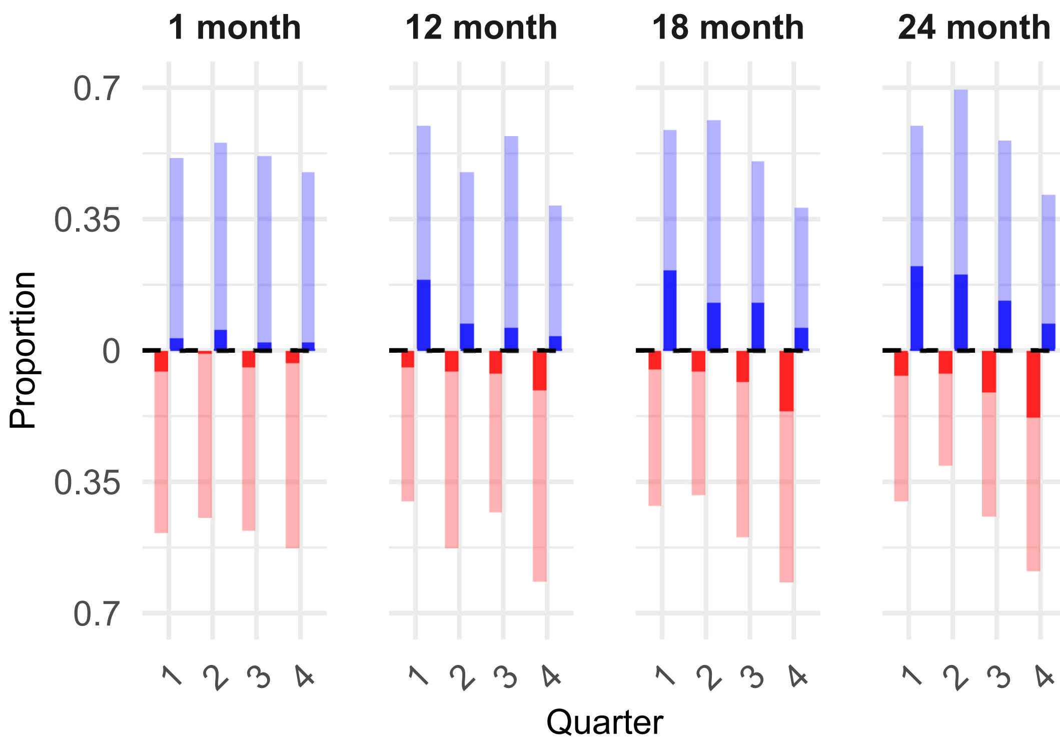
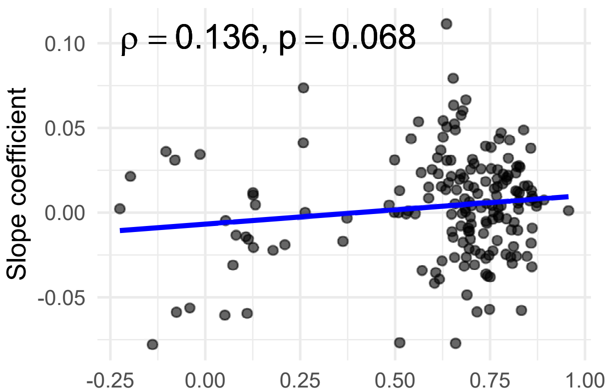
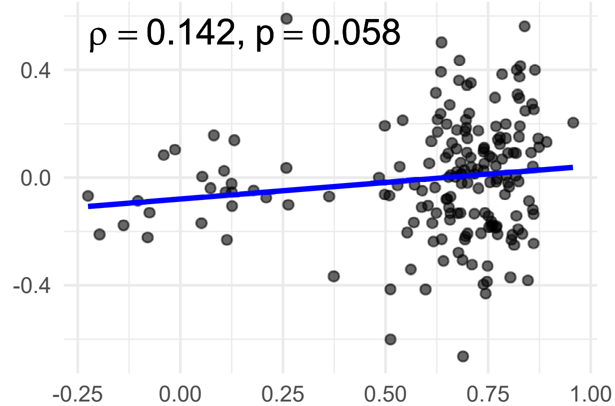


Figure 3: Conditional In-sample Results

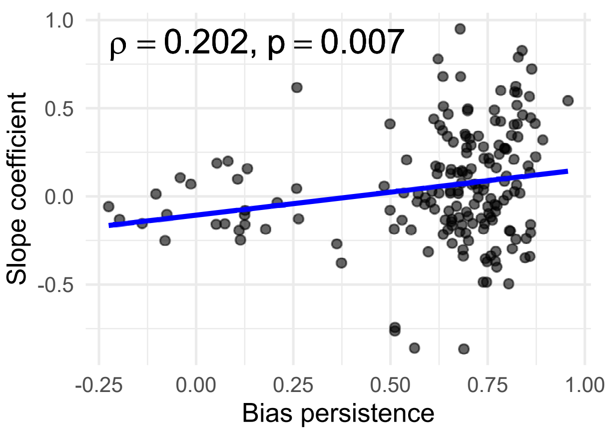
This figure presents results from the predictive regression estimated separately for each anomaly j and investment horizon h: $R_{j,t+h} = a_s + b_s \Delta_c \nu_{j,t} + \varepsilon_{j,t+h}$. The sample is split into quarters, and the figure shows the proportion of positive (blue bars) and negative (red bars) slope coefficients for each quarter of our sample. Darker segments within each bar indicate the share of coefficients that are statistically significant at the 5% level, based on t-statistics computed following Newey and West (1987) with lag lengths equal to $h - 1$.



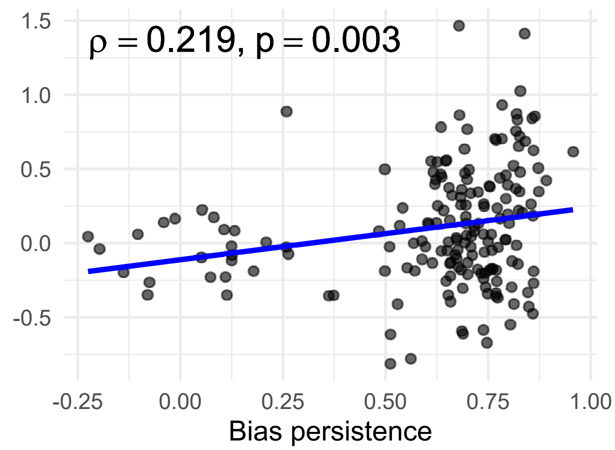
(a) Panel (A): Investment Horizon of 1 Month



(b) Panel (B): Investment Horizon of 12 Months



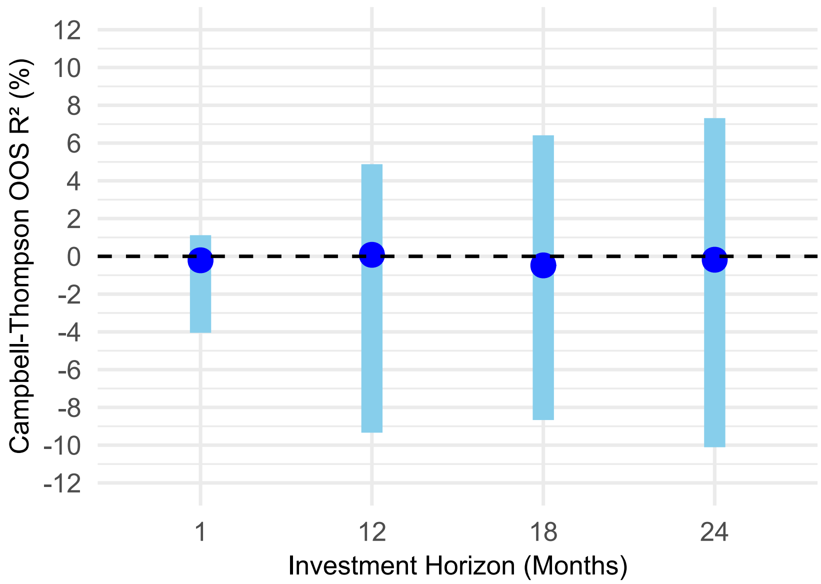
(c) Panel (C): Investment Horizon of 18 Months



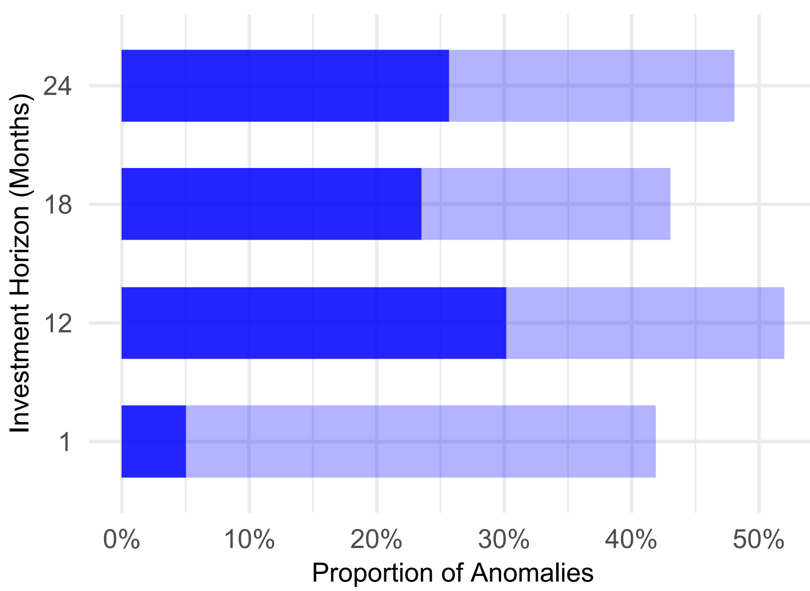
(d) Panel (D): Investment Horizon of 24 Months

Figure 4: Bias Persistence and Return Predictability

This figure reports the cross-sectional correlation between the persistence of the long-short portfolio bias, φ, and the slope coefficient, b, from regressions of value-weighted decile portfolio returns on bias. Results are shown for investment horizons of 1, 12, 18, 24, months. Each dot represents one of the 179 anomalies in our sample.



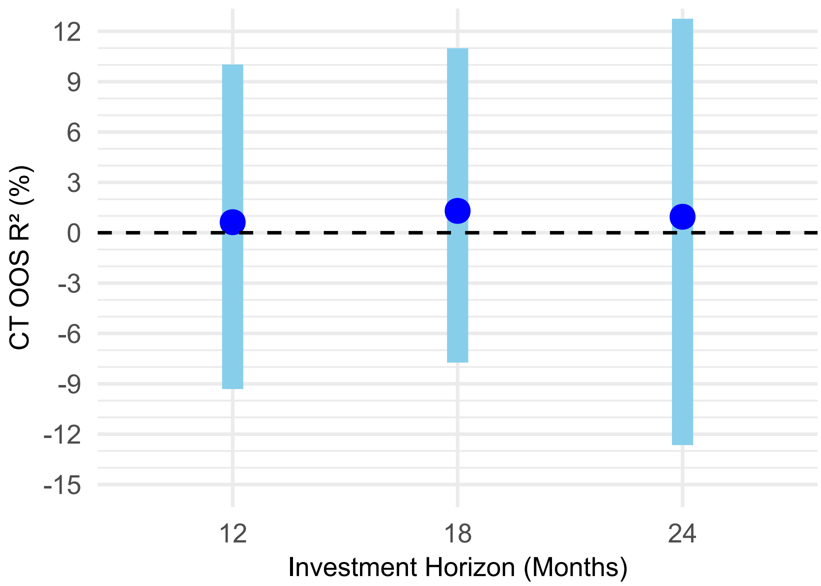
Panel A: R^2_{OOS} Distribution



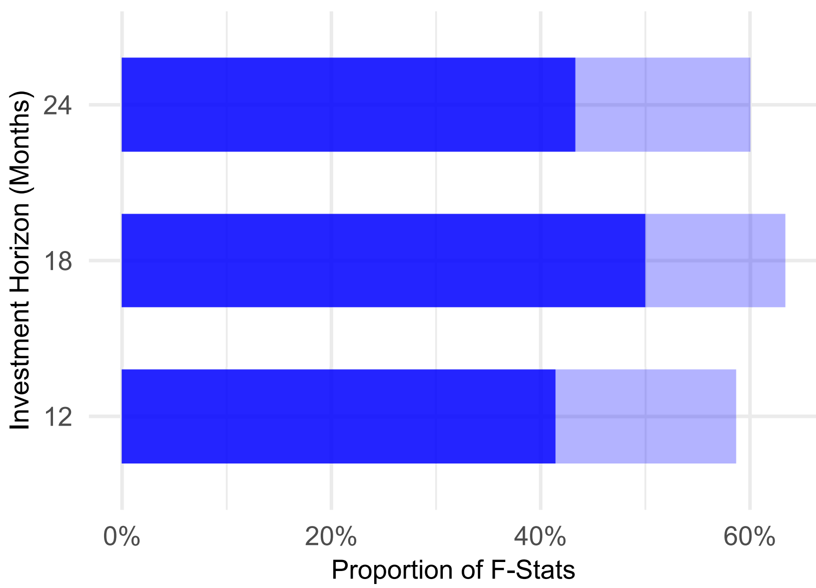
Panel B: Predictability Siginificance

Figure 5: Out of Sample Return Predictability

Panel A shows the distribution of out-of-sample R^2 values, computed following Campbell and Thompson (2008): $R^2_{OOS,j,h} = 1 - \frac{\sum_{t \in \text{test sample}}(R_{j,t+h} - \hat{R}_{j,t+h})^2}{\sum_{t \in \text{test sample}}(R_{j,t+h} - \bar{R}_{j,t})^2}$, where $R_{j,t+h}$ is the h-period return of anomaly portfolio j, $\hat{R}_{j,t+h} \equiv \mathbb{E}[R_{j,t+h} | \Delta_c \nu_{j,t}]$ represents the portfolio return predicted by our model, and $\bar{R}_{j,t}$ is the average h-period portfolio return over period between the beginning of our sample and $t - h$. The maximum and minimum values shown represent the 95th and 5th percentiles, respectively, of R^2_{OOS} distribution across the 179 anomalies in the sample. The dark blue dots indicate the median values of out-of-sample R^2. Panel B shows the fraction of McCracken (2007) F-statistics greater than zero (light blue bars) and greater than 1.554 (darker blue segments). The F-statistics are computed as: $(T_{OOS} - h + 1)\,\frac{\sum_{t \in \text{test sample}}(R_{j,t+h} - \bar{R}_{j,t+h})^2 - \sum_{t \in \text{test sample}}(R_{j,t+h} - \hat{R}_{j,t+h})^2}{\sum_{t \in \text{test sample}}(R_{j,t+h} - \hat{R}_{j,t+h})^2}$.



Panel A: R^2_{OOS} Distribution



Panel B: Predictability Siginificance

Figure 6: Out of Sample Return Predictability, Restricted Sample

Panel A shows the distribution of out-of-sample R^2 values for a restricted sample based on anomalies with McCracken (2007) F-statistic greater than 1.554 in our validation sample. Panel B shows the fraction of anomalies in our restricted sample with McCracken (2007) F-statistics greater than zero (light blue bars) and greater than 1.554 (darker blue segments).

Appendix

A Proofs and Derivations

A.1 Proposition 1

The dependence of expected returns on risk premia is based on the assumption we made regarding agents having rational expectations about future risk premia. Therefore, in this proof we focus on the relation between expected returns and belief distortions.

Predictable forecast errors are given by:

$$\mathbb{E}_t^{\mathbb{P}}[g_{i,t+1} - \mathbb{E}_t^{\mathbb{S}}[g_{i,t+1}]] = -\nu_{i,t}$$

The sum of predictable forecast revisions is given by:

$$\begin{aligned}
&\mathbb{E}_t^{\mathbb{P}}\left[\sum_{s=1}^{\infty}\alpha^s\left(\mathbb{E}_{t+1}^{\mathbb{S}}[g_{i,t+1+s}] - \mathbb{E}_t^{\mathbb{S}}[g_{i,t+1+s}]\right)\right] = \\
&\mathbb{E}_t^{\mathbb{P}}\left[\sum_{s=1}^{\infty}\alpha^s\left(\mathbb{E}_{t+1}^{\mathbb{P}}[g_{i,t+1+s}] + \frac{\zeta}{s^p}\nu_{i,t+1} - \mathbb{E}_t^{\mathbb{P}}[g_{i,t+1+s}] - \frac{\zeta}{(s+1)^p}\nu_{i,t}\right)\right]
\end{aligned}$$

Forecast revisions under the objective measure are not predictable using information available at time t. Therefore, $\mathbb{E}_t^{\mathbb{P}}\left[\mathbb{E}_{t+1}^{\mathbb{P}}[g_{i,t+1+s}] - \mathbb{E}_t^{\mathbb{P}}[g_{i,t+1+s}]\right] = 0$. Then, the sum of predictable forecast revisions is given by:

$$\alpha\zeta\sum_{s=0}^{\infty}\frac{\alpha^s}{(1+s)^p}\varphi\nu_{i,t} - \zeta\sum_{s=1}^{\infty}\frac{\alpha^s}{(1+s)^p}\nu_{i,t} = \left(\zeta\varphi Li_p(\alpha) - \zeta\left(\frac{Li_p(\alpha)}{\alpha} - 1\right)\right)\nu_{i,t}$$

Adding the expression for forecast errors into the expression for the sum of forecast revisions provides us with the expression shown in Equation 8.

B Details on the Machine Learning forecasts

Campbell et al. (2024) analyze a multitude of machine learning algorithms and design choices to determine the optimal algorithm for predicting future earnings per share. For our forecast, we use the model that they identify as best-performing.

As a first step, we construct a dataset that contains the same variables as theirs. This dataset includes 67 financial ratios provided by WRDS, earnings and earnings forecast data from IBES, as well as return and price data from CRSP. Table A2 defines all predictor variables we use. We follow Campbell et al. (2024) and train the model to predict the analyst forecast error defined as Realized EPS two years ahead less Analyst forecast, as they find that predicting the forecast error is more accurate than predicting earnings. Since the analyst forecast is known, the predicted forecast error from the model can be transformed into an earnings prediction if required. However, since we are interested in a measure of analyst forecast bias, we use the model's prediction of the analyst forecast error directly.

Since Campbell et al. (2024) do not provide their data generation code, we use code from Zhang et al. (2024) as well as our own code for the data preparation.

Following the optimal specification from Campbell et al. (2024), we use LightGBM (Ke et al., 2017), which is an implementation of the Gradient Boosting Regression Tree approach. The idea behind this approach is as follows: The model starts by building a single decision tree by repeatedly splitting the data based on predictor variables. For instance, the first tree may split data points into those with below and above median analyst forecast in a first step, and then further split these groups based on whether they have below or above median lagged EPS, which results in four subgroups in total. The number of splits a tree makes is called its *depth*. The model's prediction for a given data point is then the mean of all observations in its subgroup.

After the first decision tree is built, the model calculates its prediction errors and continues by building a second tree that attempts to correct these errors. The algorithm continues in this manner with each tree aiming to correct the errors of the preceding tree until a pre-specified number of total trees is reached. To prevent the model from overweighting individual trees, the prediction of each tree is shrunk by multiplying it by a constant called the *learning rate*.

The Gradient Boosting Regression Tree approach is similar to the random forest approach, which has been used to predict earnings in previous work (e.g., Van Binsbergen et al., 2022; Zhang et al., 2024). However, instead of building each tree to correct the errors of the previous trees, random forest builds several independent decision trees and averages their predictions.

LightGBM requires us to make several parameter choices, and we use those that Campbell et al. (2024) found to perform best. Specifically, we use mean absolute error as the objective, timeseries cross-validation, monthly model retraining, and expanding window training. We use 2000 individual trees and re-tune the learning rate and tree depths each year using a Bayesian search. We search over learning rates between 0.0001 and 1, and tree depths between 1 and 3.

Table A3 shows that the machine learning forecasts clearly outperform the analyst forecasts. They have a lower mean absolute error (MAE) on average and also a lower average MAE for more than 60% of firms and in more than 80% of months. Moreover, while the analyst forecasts have a mean error of -0.2681, indicating excess optimism, our machine learning forecasts show a much smaller bias of -0.0457. Given that the standard deviation of realized EPS is 2.57, the bias of the analyst forecasts is slightly more than a 10th of a standard deviation of the underlying, whereas the bias of the machine learning forecast is less than a 50th of the standard deviation of the underlying.

Table A1 shows statistics about the feature importance for the most relevant features, using the SHAP (SHapley Additive exPlanations) value (Lundberg and Lee, 2017). The SHAP value is based on a concept from cooperative game theory originally introduced by Shapley (1953). The Shapley (SHAP) value is a measure of how the payoff of a game (the predictions of a model) would change if a player (feature) were removed from the game

Feature	Mean Abs.	25th Perc. Abs.	Median Abs.	75th Perc. Abs.	Mean	25th Perc.	Median	75th Perc.	Abs. of Mean	Std Signed	Std Abs.
medest2	0.448	0.350	0.468	0.534	-0.428	-0.517	-0.453	-0.326	0.428	0.141	0.137
prc	0.422	0.312	0.409	0.545	0.411	0.284	0.406	0.541	0.411	0.184	0.174
size	0.128	0.083	0.104	0.137	0.123	0.079	0.104	0.137	0.123	0.099	0.095
evm	0.097	0.051	0.090	0.127	-0.091	-0.125	-0.083	-0.043	0.091	0.066	0.064
fcf_ocf	0.086	0.074	0.083	0.096	0.042	0.026	0.045	0.064	0.042	0.032	0.019
pe_exi	0.073	0.039	0.078	0.103	-0.068	-0.100	-0.074	-0.032	0.068	0.044	0.040
cash_conversion	0.065	0.032	0.049	0.084	0.022	-0.001	0.031	0.061	0.022	0.084	0.062
ibes_earnings_ann	0.052	0.032	0.044	0.061	-0.023	-0.041	-0.023	-0.006	0.023	0.031	0.035
ps	0.050	0.035	0.047	0.062	-0.049	-0.061	-0.046	-0.033	0.049	0.022	0.021
rect_turn	0.048	0.019	0.034	0.070	0.037	0.012	0.028	0.059	0.037	0.044	0.039

Table A1: The table shows the feature importance using the SHAP value of the top ten features by mean absolute SHAP value. The SHAP values are calculated by first calculating the mean SHAP value and mean absolute SHAP value of all predictions in a given month and then calculating the summary statistics for these means. Thus, every month has the same weight in the final statistics.

(model).

The table shows that the most relevant feature is the analyst forecast (*medest2*), which is unsurprising given that the model predicts the analyst EPS forecast error, which directly depends on this variable. Almost as important as the analyst forecast is the price (*prc*), which likely matters because the target variable is a per-share measure. Since the magnitude of per-share variables crucially depends on the number of shares, it can only be interpreted in an economically meaningful way together with the share price, which makes the price a crucial interaction term for all predictor variables that are not also on a per-share basis, such as the *size* of the firm, which is the third most important feature. Six out of the seven remaining top ten features are different financial ratios from WRDS, with the most important one being the *Enterprise Value Multiple*. The final top feature is the most recently announced EPS (*ibes_earnings_ann*).

Table A2: Variable Definition

Variable	Definition	Variable	Definition
	A.1 WRDS Financial Ratios (67)		
accrual	Accruals/Average Assets	intcov_ratio	Interest Coverage Ratio

Continued on next page

Table A2 – continued from previous page

Variable	Definition	Variable	Definition
adv_sale	Advertising Expenses/Sales	inv_turn	Inventory Turnover
aftret_eq	After-tax Return on Average Common Equity	invt_act	Inventory/Current Assets
aftret_equity	After-tax Return on Total Stockholders Equity	lt_debt	Long-term Debt/Total Liabilities
aftret_invcapx	After-tax Return on Invested Capital	lt_ppent	Total Liabilities/Total Tangible Assets
at_turn	Asset Turnover	npm	Net Profit Margin
bm	Book/Market	ocf_lct	Operating CF/Current Liabilities
capei	Shillers Cyclically Adjusted P/E Ratio	opmad	Operating Profit Margin After Depreciation
capital_ratio	Capitalization Ratio	opmbd	Operating Profit Margin Before Depreciation
cash_conversion	Cash Conversion Cycle (Days)	pay_turn	Payables Turnover
cash_debt	Cash Flow/ Total Debt	pcf	Price/Cash flow
cash_lt	Cash Balance/Total Liabilities	pe_exi	P/E (Diluted, Excl. EI)
cash_ratio	Cash Ratio	pe_inc	P/E (Diluted, Incl. EI)
cfm	Cash Flow Margin	PEG_trailing	Trailing P/E to Growth (PEG) ratio
curr_debt	Current Liabilities/Total Liabilities	pretret_earnat	Pre-tax Return on Total Earning Assets
curr_ratio	Current Ratio	pretret_noa	Pre-tax return on Net Operating Assets
de_ratio	Total Debt/Equity	profit_lct	Profit Before Depreciation / Current Liabilities

Continued on next page

Table A2 – continued from previous page

Variable	Definition	Variable	Definition
debt_assets	Total Debt/Total Assets (1)	ps	Price/Sales
debt_at	Total Debt/Total Assets (2)	ptb	Price/Book
debt_capital	Total Debt/Capital	ptpm	Pre-tax Profit Margin
debt_ebitda	Total Debt/EBITDA	quick_ratio	Quick Ratio (Acid Test)
debt_invcap	Long-term Debt/Invested Capital	RD_SALE	Research and Development/Sales
divyield	Dividend Yield	rect_act	Receivables/Current Assets
dltt_be	Long-term Debt/Book Equity	rect_turn	Receivables Turnover
dpr	Dividend Payout Ratio	roa	Return on Assets
efftax	Effective Tax Rate	roce	Return on Capital Employed
equity_invcap	Common Equity/ Invested Capital	roe	Return on Equity
evm	Enterprise Value Multiple	sale_equity	Sales/Stockholders Equity
fcf_ocf	Free Cash Flow/Operating Cash Flow	sale_invcap	Sales/Invested Capital
gpm	Gross Profit Margin	sale_nwc	Sales/Working Capital
GProf	Gross Profit/Total Assets	short_debt	Short-Term Debt/Total Debt
int_debt	Interest/ Average Long-term Debt	staff_sale	Labor Expenses/Sales
int_totdebt	Interest/Average Total Debt	totdebt_invcap	Total Debt/Invested Capital
intcov	After-tax Interest Coverage		
	A.2 CRSP (5)		
ret	Monthly Return	prc	Closing Price
size	LN(Market Capitalization)	mom6m	6-month momentum
indmom	Fama French 38 Industry weighted 6-month momentum		
	A.3 IBES (5)		

Continued on next page

Table A2 – continued from previous page

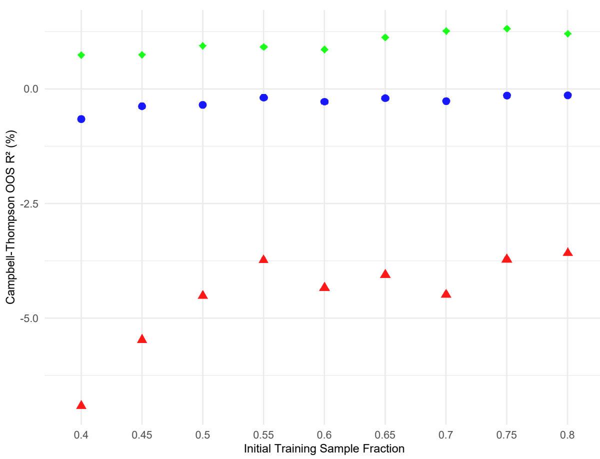
Variable	Definition	Variable	Definition
medest2	Analyst consensus forecast for FY2 horizon	ibes earnings ann	Most recently realized annual earnings
last F2ana fe med	Most recently realized FY2 horizon analyst forecast error	rev FY2 3m	Revision of analyst FY2 horizon forecast between current month and 3 months prior
dist2	Distance between FY2 fiscal period end and current month		

Metric	**Analyst Forecasts**	**ML Forecasts**
Mean Absolute Error	0.7136	0.6729
Mean Error	-0.2681	-0.0457
% of firms with better performance	38.17%	61.83%
% of year/months with better performance	16.37%	83.63%

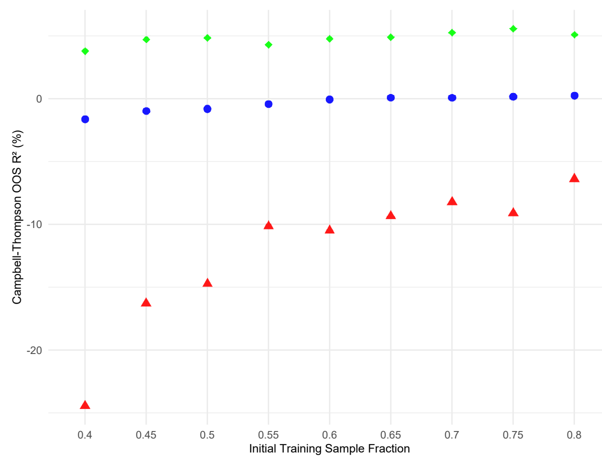
Table A3: The table shows a comparison of the performance between the Analysts and the machine learning model. To calculate % of firms with better performance, we first calculated the average mean absolute forecast error for each firm across all observations for this firm and then checked for which percentage of firms it was lower for the analysts and the machine learning model, respectively. % of year/months with better performance is calculated similarly by comparing the performance in each month, averaged across all firms.

C Robustness

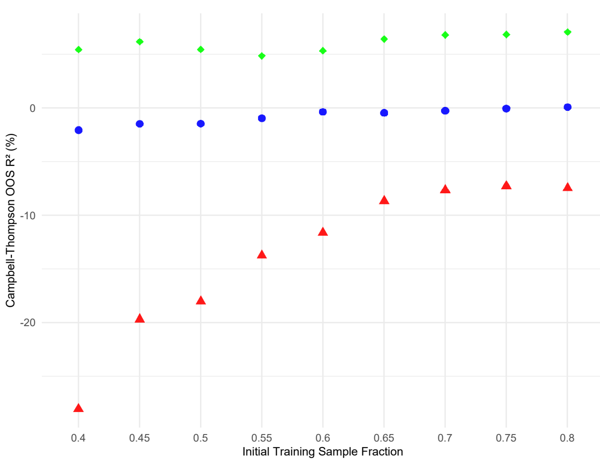
C.1 Additional Results



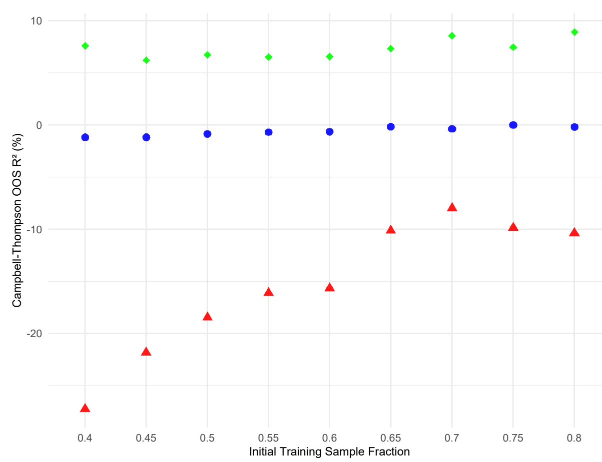
Panel (A): Investment Horizon of 1 Month



Panel (B): Investment Horizon of 12 Months



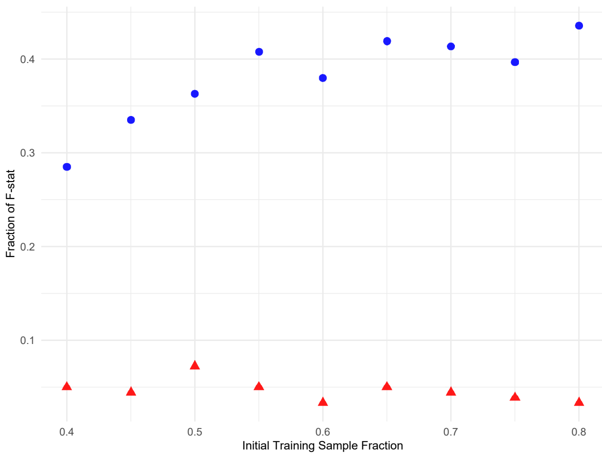
Panel (C): Investment Horizon of 18 Months



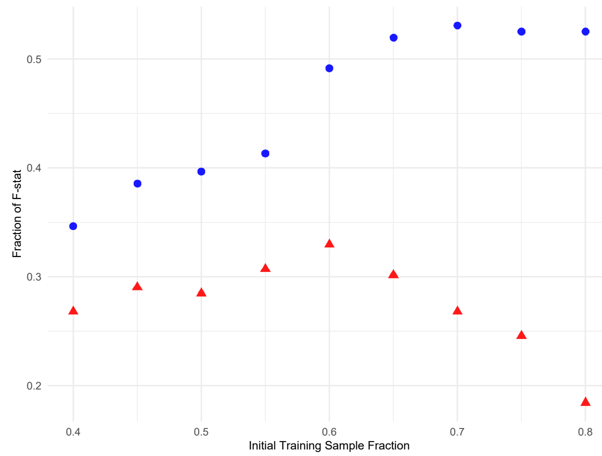
Panel (D): Investment Horizon of 24 Months

Figure A1: Return Predictability Results, Robustness to Sample Splits R^2_{OOS}

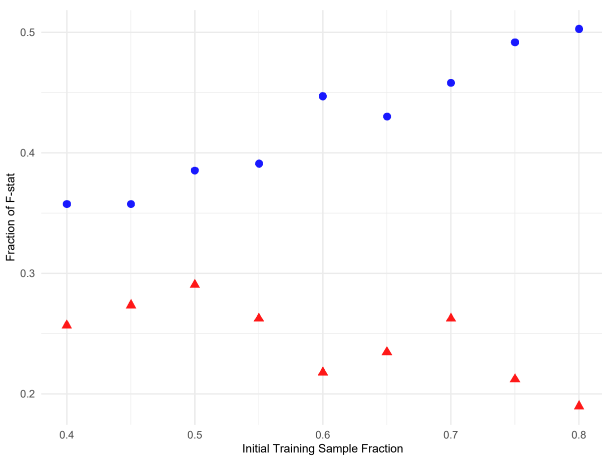
This figure shows the distribution for the out-of-sample coefficients of determination (R^2_{OOS}), computed following Campbell and Thompson (2008), for various fractions of our sample used as an initial training sample in the out-of-sample return predictability exercise outlined in subsection 5.1. The red triangles represent the 5th percentile of the cross-sectional distribution, the blue dots represent the median, and the green diamonds represent the 95th percentile of the distribution.



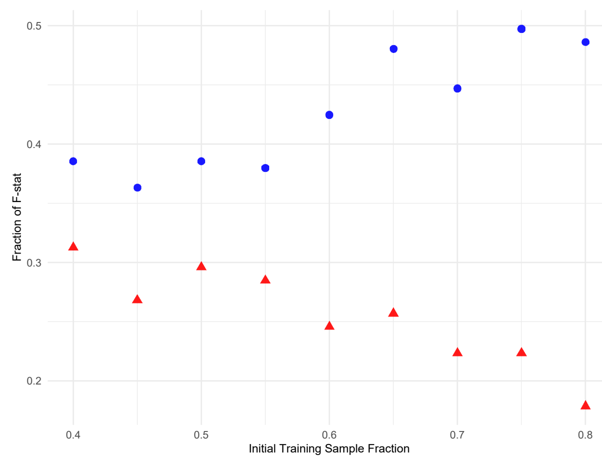
Panel (A): Investment Horizon of 1 Month



Panel (B): Investment Horizon of 12 Months



Panel (C): Investment Horizon of 18 Months



Panel (D): Investment Horizon of 24 Months

Figure A2: Return Predictability Results, Robustness to Sample Splits F-statistic

This figure shows the fraction of F-statistics computed following McCracken (2007) that are greater than zero (blue dots) and greater than 1.554 (red triangles) for various fractions of our sample used as an initial training sample in the out-of-sample return predictability exercise outlined in subsection 5.1.

C.2 Monte Carlo Simulations

In our simulation exercises, we replace the empirical long–short bias series with simulated AR(1) processes calibrated to match the mean, volatility, and persistence of the empirical series. Using these simulated series, we repeat the in-sample and out-of-sample return predictability tests described in the main body of the paper. The figures below plot the resulting distributions based on 10,000 simulations.

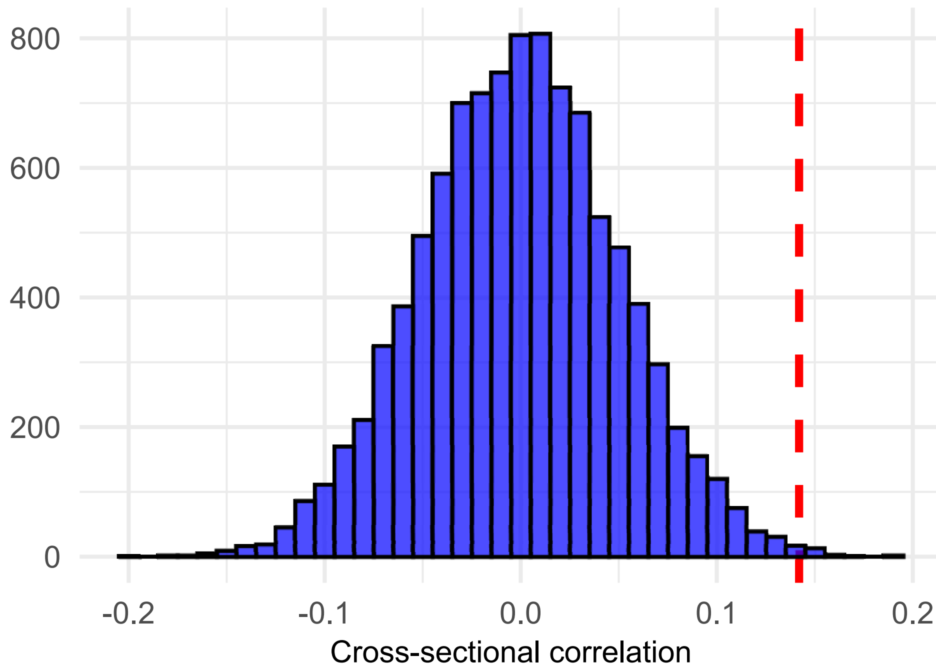


Figure A3: Simulated Cross-Sectional Correlations

This figure reports the distribution of cross-sectional correlation coefficients between simulated bias persistence and predictive regression slope coefficients at the 12-month horizon across the 179 anomalies in our sample. To generate the simulated distribution, we construct 10,000 artificial bias series for each anomaly, calibrated to match the mean, persistence, and volatility of the corresponding empirical long–short bias series. For each simulation, we re-estimate the predictive regression as specified in Equation 11, then compute the cross-sectional correlation between the simulated persistence estimates and the predictive slopes. The empirical correlation coefficient is overlaid as a red dashed line for comparison.

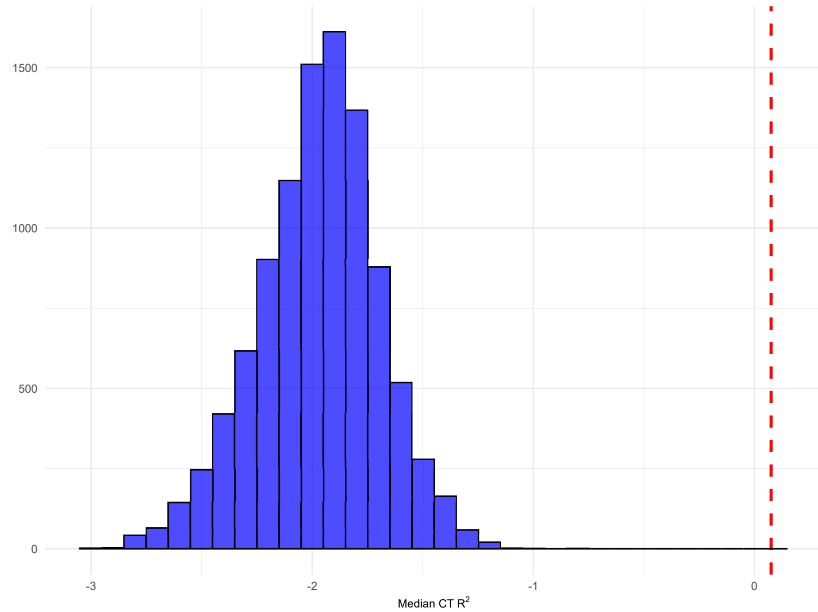


Figure A4: Median R^2_{OOS}, Simulated Data

This figure reports the distribution of the median out-of-sample coefficient of determination (R^2_{oos}) at the 12-month investment horizon computed following Campbell and Thompson (2008). For each of the 179 anomalies, we generate 10,000 artificial bias series calibrated to match the mean, persistence, and volatility of the associated empirical long-short bias series. For each simulation, we repeat our out-of-sample return predictability exercise described in subsection 5.1. The empirical median value is shown using a red dashed line.