Topics in Financial Forecasting Through Machine Learning

by

Harrison Ham

(Under the Direction of Zhongjin Lu)

Abstract

In this dissertation, I investigate three applications of machine learning in financial forecasting. The first study investigates the best techniques for forecasting corporate earnings, and sheds light on what the most accurate earnings expectation is, and what the market expectation appears to be. We find consistent evidence that the best machine learning forecast outperforms analysts' forecasts. However, the best machine expectation does not beat the analyst forecast by a meaningful amount in most cases, except for two distinct instances: (1) the earnings forecast is for small firms, and (2) the earnings forecast is for a longer horizon. Second, in cases where there are meaningful differences between analyst and machine expectations, earnings response coefficient (ERC) tests imply that investors' expectations appear to be mostly aligned with the best machine forecast. In my second study, I investigate the best interest rate forecasting techniques, and show that existing techniques perform poorly compared to a simple forecast of zero change. In light of this, I propose a new interest rate forecast which focuses on removing the maturity risk premium from forward rates and demonstrate that this new approach outperforms for long horizon

forecasts of interest rates. Given these findings, I decompose excess bond returns to show that the primary driver of excess bond returns for short holding periods is a bonds carry, while for long holding periods its the bonds maturity risk premium. This risk premium is plausibly invariant across both time and across the maturities of forward rates. In my third study, we propose and test the "Sticky Information Cost" (SIC) hypothesis to understand how investors acquire information in uncertain financial markets. SIC asserts that information processing costs for investors are influenced by a firm's slow-changing information environment, closely linked to its fundamental uncertainty. Using direct measures for information processing costs and the return predictability of analysts' biases as a proxy for information acquisition, we find opposite relationships between uncertainty and information acquisition when comparing across firms and over time. These results hold across various uncertainty measures and other earnings-related anomalies, supporting the SIC hypothesis while challenging existing theories.

INDEX WORDS: Machine Learning, Interest Rates, Forecasting, Earnings, Risk Premia

Topics in Financial Forecasting Through Machine Learning

by

Harrison Ham

B.B.A. Finance, University of Georgia, 2020

M.A. Economics, University of Georgia, 2020

A Dissertation Submitted to the Graduate Faculty of the

University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Doctor of Philosophy

Athens, Georgia

2024

Topics in Financial Forecasting Through Machine Learning

by

Harrison Ham

Major Professor: Zhongjin Lu

Committee: Jeffry Netter

Annette Poulsen

John Campbell

Electronic Version Approved:

Ron Walcott

Dean of the Graduate School

The University of Georgia

May 2024

Dedication

To my wife Anna Kathryn, whose support is the only reason I've come this far. To my family who taught me to push the bounds of my education from a young age, and supported me through my entire academic journey. To my professors who never lost faith in me, and pushed me to become the best version of myself. To my friends, who made my time at the University of Georgia the best 8 years of my life. And finally, to my dog Beans, whose loyalty and sweet disposition made every day a little bit easier.

Acknowledgments

I'd like to give a special thank you to my committee members, Zhongjin Lu, Jeffry Netter, Anette Poulsen and John Campbell. They've given me an unbelievable amount of support both during and even before my time in the program. Without their guidance, I would never had made it this far. I would also like to thank all the other professors and PH.D students at the University of Georgia, who have taught me so much, and have encouraged at every step along the way.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The use of machine learning has proliferated across finance over the last few decades. From valuing assets, to modeling firm fundamentals, machines have improved our understanding of many topics by generating more accurate forecasts than traditional methods would allow. Despite machine learning's prominence, there are many topics in finance that largely remained unexplored using the most recent and powerful machine learning tools. On top of this, the overwhelming number of choices that researchers face when implementing machine learning can make it both complicated and time consuming to apply machine learning in an empirical setting. In this chapter, I outline the three papers in my dissertation, and talk about how they shed light on these issues.

1.1 Earnings Forecasting

The primary driver of a firm's value is their ability to generate profits. Even firms that are unprofitable only have value because they have the potential to turn a profit at some point in the future. Because of

this, creating accurate expectations about the future of firm profits is of first order importance to financial researchers. Within the industry, brokers make entire careers generating forecasts of corporate earnings. They then sell these forecasts to investing firms, corporations, and academic researchers alike who use them to form more accurate expectations about the future of different companies.

A large line of accounting and finance research examines the extent to which a firm's earnings can be predicted (i.e., earnings forecasts), as well as investors' expectations of earnings, as revealed by the market reaction to earnings announcements. Productivity gains in the last several decades have led to exponential growth in (1) the amount of data being produced and captured by humans, and (2) the computing power with which to analyze that data. A natural question is the extent to which the advancement in machine learning technology, which capitalizes on these two trends, can improve earnings forecasts, and if so, whether market expectations appear to reflect those superior forecasts.

We examine three related research questions. First, among analyst and many machine learning forecasts, what earnings expectation minimizes ex-post forecast errors (i.e., what *should* the earnings expectation be)? Second, do market earnings expectations appear to align more with machine learning or analysts' forecasts (i.e., what do earnings expectations *appear to be*)? Finally, and perhaps most importantly, how do the answers to these questions evolve over time?

We are the first study to examine whether investors' earnings expectations appear to align more with machine learning or analysts' forecasts. While a few (mostly contemporaneous) studies examine the statistical superiority of machine learning forecasts over analysts' forecasts, their findings are inconclusive. This ambiguity largely stems from the absence of theoretical guidance in implementing ML models, leading to diverse specification choices. Therefore, before we examine our research questions, we first

assess the impact of these specification choices by compiling and examining a comprehensive list of 3,024 machine learning models that represent the range of choices used in the existing literature.

Using a sample of forecasts from 1990 to 2020, we find consistent evidence that the best machine learning forecast outperforms analysts' forecasts. However, the best machine expectation does not beat the analyst forecast by a meaningful amount in most cases, except for two distinct instances: (1) the earnings forecast is for small firms (a "size" effect), and (2) the earnings forecast is for a longer horizon (a "horizon" effect). Second, in cases where there are meaningful differences between analyst and machine expectations, earnings response coefficient (ERC) tests imply that investors' expectations appear to be mostly aligned with the best machine forecast. The alignment with the machine forecast strengthens over time and is especially strong among firms with more sophisticated investors. Third, our time-series analyses suggest that analyst and machine forecasts are converging over time and that analysts' information production remains critical. Taken together, our results suggest that machines rely on analysts' information, analysts appear to rely on machines to reduce their biases, and thus the two are unlikely to diverge significantly even as technology continues to evolve.

1.2 Interest Rate Forecasts

Interest rate forecasting is of first order importance to investors, policy makers, and researchers alike. Through understanding what portion of future interest rates are predictable and why, researchers can get a glimpse into the determinants of the term structure of interest rates. Through understanding the path of future interest rates, investors can more accurately price investment opportunities and manage risk.

And finally, through understanding the fundamental drivers of interest rates, economic policy makers can be more informed when making decisions.

Despite decades of progress, the prior literature largely disagrees about the fundamental determinants of the interest rate term structure and the best ways to form expectations about its future. This disagreement largely stems a difference in forecasting methodologies, and two key discrepancies which cause results to drastically change from one paper to the next. The first, is that data revisions, and look-ahead bias inflate the perceived ability of statistical models to predict future interest rates. The second is that researchers use different benchmarks, which makes it difficult to compare across studies. Because of these discrepancies, it is unclear which forecasting methodologies should be used by both researchers and market participants when forming expectations about future interest rates.

I shed light on this issue by running a comprehensive analysis of the forecasting techniques proposed by the prior literature. In doing so, I demonstrate that the very simple random walk forecast outperforms other statistical models in almost all circumstances. In light of this finding, I then propose a new interest rate forecasting methodology which focuses on removing the risk premium from forward rates under the assumption that the risk premium is invariant in the cross section. Finally, I discuss the theoretical implications of my findings in the context of the spanning hypothesis, and the determinants of bond risk premia.

1.3 Uncertainty and Sticky Information Costs

The relationship between earnings analysts and market participants has analyzed for years. Market participants use analyst forecasts to help form expectations about future earnings. In doing so, they can

incorporate the soft information provided by analysts to more accurately predict what future earnings will be. However, analyst forecasts have been shown to be biased, which can sometimes lead investor expectations astray. In times where this happens, market participants can exert extra effort to try to debias analyst forecasts in order to form even more accurate expectations. However, doing so is costly, which can sometimes cause market participants to not fully unravel analyst biases leading to inefficient expectations. It is unclear however what circumstances give rise to these inefficiencies.

Existing theories highlight the role of uncertainty in shaping individuals' information acquisition decisions but have ambiguous predictions on whether higher uncertainty is associated with more or less information acquisition. Intuitively, with an increasing level of uncertainty, every bit of information becomes more valuable and hence the benefit of acquiring information ("the benefit channel"). Simultaneously, amid heightened uncertainty, the information is potentially more difficult to process, increasing the cost of acquiring information ("the cost channel").

Empirically, whether individuals acquire more or less information when facing heightened uncertainty remains an open question. Existing empirical studies find that investors appear to pay more attention to information when uncertainty is high, supporting the benefit channel. However, despite the potential importance of the cost channel, there is a lack of understanding and evidence on how the cost channel may affect the relation between uncertainty and information acquisition.

In this paper, we propose and test the "Sticky Information Cost" (SIC) hypothesis to understand how investors acquire information in uncertain financial markets. SIC asserts that information processing costs for investors are influenced by a firm's slow-changing information environment, closely linked to its fundamental uncertainty. Using direct measures for information processing costs and the return predictability of analysts' biases as a proxy for information acquisition, we find opposite relationships

between uncertainty and information acquisition when comparing across firms and over time. These results hold across various uncertainty measures and other earnings-related anomalies, supporting the SIC hypothesis while challenging existing theories. Incorporating the SIC into the existing information choice theories provides a new perspective on return anomalies.

Chapter 2

Expectations Matter: When (not) to Use Machine Learning Earnings Forecasts[1]

We comprehensively examine if machine learning technology can meaningfully improve earnings forecasts, and if so, whether market expectations appear to reflect those superior forecasts. First, using a sample of forecasts from 1990 to 2020, we find consistent evidence that the best machine learning forecast outperforms analysts' forecasts. However, the best machine expectation does not beat the analyst forecast by a meaningful amount in most cases, except for two distinct instances: (1) the earnings forecast is for small firms (a "size" effect), and (2) the earnings forecast is for a longer horizon (a "horizon" effect). Second, in cases where there are meaningful differences between analyst and machine expectations, earnings response coefficient (ERC) tests imply that investors' expectations appear to be mostly aligned with the

[1]Co-Authors: John L. Campbell University of Georgia), Zhongjin (Gene) Lu (University of Georgia), Katherine Wood (Bentley University)

best machine forecast. The alignment with the machine forecast strengthens over time and is especially strong among firms with more sophisticated investors. Third, our time-series analyses suggest that analyst and machine forecasts are converging over time and that analysts' information production remains critical. Taken together, our results suggest that machines rely on analysts' information, analysts appear to rely on machines to reduce their biases, and thus the two are unlikely to diverge significantly even as technology continues to evolve.

2.1 Introduction

A large line of accounting and finance research examines the extent to which a firm's earnings can be predicted (i.e., earnings forecasts), as well as investors' expectations of earnings, as revealed by the market reaction to earnings announcements. Productivity gains in the last several decades have led to exponential growth in (1) the amount of data being produced and captured by humans, and (2) the computing power with which to analyze that data. A natural question is the extent to which the advancement in machine learning technology, which capitalizes on these two trends, can improve earnings forecasts, and if so, whether market expectations appear to reflect those superior forecasts.

In this study, we examine three related research questions. First, among analyst and many machine learning forecasts, what earnings expectation minimizes ex-post forecast errors (i.e., what *should* the earnings expectation be)? Second, do market earnings expectations appear to align more with machine learning or analysts' forecasts (i.e., what do earnings expectations *appear to be*)? Finally, and perhaps most importantly, how do the answers to these questions evolve over time?

We are the first study to examine whether investors' earnings expectations appear to align more with machine learning or analysts' forecasts. While a few (mostly contemporaneous) studies examine the statistical superiority of machine learning forecasts over analysts' forecasts, their findings are inconclusive. This ambiguity largely stems from the absence of theoretical guidance in implementing ML models, leading to diverse specification choices. Therefore, before we examine our research questions, we first assess the impact of these specification choices by compiling and examining a comprehensive list of 3,024 machine learning models that represent the range of choices used in the existing literature.

Motivated by Bradshaw et al., 2012, we create the variable ML Superiority, which equals the absolute value of the error in the analysts' forecast minus the absolute value of the error in a given machine learning forecast. Positive values suggest that the machine learning forecast is better at predicting actual earnings relative to the analysts' forecast. We find significant variation in the accuracy of machine forecasts due to specification choice. Nearly 90 percent of the 3,024 machine learning models evaluated underperform analysts' forecasts. Thus, understanding the impact of specification choices is key to making sense of prior papers' results on machine or analysts' superiority. Our results show that utilizing a non-linear algorithm, combined with an outlier-resistant loss function and a temporal train-validation split, significantly improves forecast accuracy. This approach, particularly when focused on refining analysts' forecasts by correcting their predictable biases, consistently yields the most statistically accurate earnings forecast.[2]

Having settled on the most appropriate machine learning specification to use, we then examine our three research questions and offer several key findings. First, using a sample of forecasts from 1990 to 2020, we find consistent evidence that the best machine learning forecast outperforms analysts' forecasts.

[2]Furthermore, when we divide the sample into decade-long subsamples, we find the top performing models consistently use these four key specification choices. Sections 2.2 through 2.4 provide more details behind this methodological exercise. Furthermore, our website contains programming code and statistical estimates to the extent other researchers wish to use these statistically optimal machine learning forecasts.

However, the best machine expectation does not beat the analyst forecast by a meaningful amount in the vast majority of cases, with meaningful differences limited to two distinct instances: (1) the earnings forecast is for small firms (a "size" effect), and (2) the earnings forecast is for a longer horizon (a "horizon" effect). Second, in cases where there are meaningful differences between analyst and machine expectations, earnings response coefficient (ERC) tests imply that investors' expectations appear to be mostly aligned with the best machine forecast. The alignment with the machine forecast strengthens over time and is especially strong among firms with more sophisticated investors. Third, our time-series analyses suggest that analyst and machine forecasts are converging over time.

In additional analysis, we examine whether the improvement of the machine learning models over analysts for longer horizon forecasts (i.e., two years ahead) can significantly improve Implied Cost of Capital (ICC) estimates. ICC is a setting commonly studied in accounting and finance, and a key input for these estimates is long-term earnings forecasts. We follow Lee et al., 2021 and use the measurement error variance (MEV) as the metric for assessing the accuracy of various expected-return proxies (ERP). Our analysis replicates the findings in Lee et al., 2021 that for tracking the cross-sectional variation in expected returns, ICCs based on analysts' forecasts underperform a composite characteristic-based expected-return measure (CER). However, we find that ICCs based on the statistically optimal machine forecasts outperform CER. Furthermore, we observe that the accuracy gain from using the best statistical forecasts declines in firm size, consistent with machine forecasts being materially more accurate than analysts' forecasts among smaller firms.

Motivated by Bertomeu et al., 2021, we also perform feature importance analyses to examine the most important factors that lead to machine forecast dominance and find that the most important features revolve around analysts' forecasts. Specifically, the consensus analysts' earnings forecast is the top

explanatory feature in both short- and long-horizon forecasts, with prior analysts' forecast errors being the next most important for short-horizon forecasts and stock price being the next most important for long-horizon forecasts. These results further emphasize the critical role of analysts' information in enhancing the precision of the machine forecasts. Finally, given the importance of analysts' information, we calculate analyst value add (AVA), defined as the improvement in the machine forecast that includes analyst information relative to the machine forecast that does not include analyst information). We find that AVA is larger for smaller firms and is largest when the firm's reporting is more complex or opaque (Bonsall et al., 2017; Loughran and McDonald, 2023; Loughran and Mcdonald, 2014). Over the time series, we find that AVA is slightly *increasing*, highlighting that the importance of analysts' information does not diminish despite advances in data, computing power, and machine learning technology.

Collectively, our results suggest that the literature's focus on "human versus machine" or even "human plus machine" is misguided. Our results suggest that the "best" machine expectations rely on analysts' information, and analysts appear to rely on machines to reduce their cognitive biases and errors.[3] Both incorporate the other and largely converge on the same expectation, except in cases where analysts are poorer at forecasting (i.e., smaller firms and forecasts over longer horizons). Interestingly, prior research suggests that these instances can be explained by analyst incentives and effort. Specifically, Harford et al., 2019) argue that analysts are poorer at forecasting smaller firms' earnings because small firms tend to be less important to evaluations of analyst performance. Similarly, Ham et al., 2022 argue that analyst accuracy decreases over longer horizons because they are less important to evaluations of analyst performance, but

[3]While no archival study can definitively conclude that analysts use machine learning, anecdotal evidence suggests that Wall Street has used machine learning and large datasets dating back to the 1990s. Specifically, MIS professors in the New York area told us they have been consulting with Wall Street since the late 1990s, and a book by Thomas Bass ("The Predictors") indicates that the global financial firm UBS Group AG was using machine learning in the 1990s. The evidence in our paper is circumstantial evidence that analysts use machine learning to remove cognitive biases and errors given that our time-series analysis finds a convergence between machine and analyst expectations over time, and our feature importance analysis suggests that prior forecast errors are important for machine improvement.

also because analysts tend to be more optimistic over longer horizons. In other words, the observed large differences between analyst and machine expectations can be explained by a lack of analyst effort rather than some superiority on the part of machines. For these reasons, as well as the fact that analysts' value add is slightly *increasing* over time, we see no reason to believe that analysts' forecasts and the best machine forecast will diverge significantly, even as technology continues to innovate.

Our study makes three main contributions to the literature. First, we contribute to the literature on the statistical superiority of machine learning forecast over analysts' forecasts. Table 2.1 presents a summary of papers in this area, showing differences in methodology and ultimately on conclusions as to whether machine or analysts' forecasts appear to dominate across various horizons (e.g., R. T. Ball and Ghysels, 2018; Bradshaw et al., 2012; S. S. Cao et al., 2021; de Silva and Thesmar, 2022; So, 2013; van Binsbergen et al., 2022 among many others). We stand apart as the only study to ask the machine to correct for expected analysts' forecast errors (i.e., the indirect approach following Frankel and Lee, 1998), showing that this approach leads to the most accurate machine forecast in our study. More importantly, our study characterizes the impact of the specification choices and resulting in a wide dispersion of ML model performance. As a result of our exhaustive approach examining over 3,000 model specifications, we provide much more confidence in the results reported in contemporaneous and unpublished studies such as the importance of combining human with machine (S. S. Cao et al., 2021; van Binsbergen et al., 2022), as well as the fact that machines produce better forecasts for smaller firms and longer forecast horizons (e.g., de Silva and Thesmar, 2022). Finally, and in contrast to the impression made by prior studies, we show that the accuracy improvement of machine forecasts over analysts' forecasts is not large much of the time.

Second, we contribute to the literature on investors' expectations of earnings. Prior studies in this area largely assume that analysts' earnings forecasts are the best proxy for market expectations (R. Ball and Brown, 1968; Bernard and Thomas, 1989, 1990; Bradshaw et al., 2012; Kothari, 2001); with some arguing that analysts and investors might have different objective functions (e.g., Basu and Markov, 2004; Z. Gu and Wu, 2003; Weiss et al., 2008) and others arguing that investors might exhibit the same cognitive biases as analysts (e.g., Bertomeu et al., 2021). We are the first to examine whether machine learning can help us learn about investors' earnings expectations and find that investors' earnings expectations appear to align mostly with the best machine forecasts. Still, investors appear to overweight the analysts' forecast relative to the machine and this overweighting decreases over time and is substantially smaller for firms with high institutional ownership.

Finally, we offer a research design contribution to future researchers. We identify the key model specification choices that drive machine forecasting accuracy, offer guidance on when it is most crucial to substitute away from the use of analysts' forecasts in earnings expectations, and provide code and estimates when it is necessary to do so. Specifically, we find that machine forecasts are more necessary for firms in the smallest size quintile and for forecasts of longer horizons—as this is when machine forecasts are more materially accurate than analysts' forecasts.

2.2 Experiment Design

2.2.1 ML Model Specification Choices

When forecasting earnings, researchers must first determine the dataset (y, X), where y is the target variable and X is the predictor set. Given the dataset (y, X), researchers fit a linear or non-linear regression

model ($f(X)$) to minimize the sum of a loss function (L) and a potential regularization term (G) with hyperparameter(s) γ. The model takes the following form:[4]

$$f(X) = \arg\min_{\{\beta\}} \left\{ \frac{1}{N} \sum_{i=1}^{N} L(y_i, \beta | X_i) + G(y_i, \beta | X_i, \gamma) \right\}, \text{for } i \text{ in the estimation window.} \tag{2.1}$$

There are 6 specification choices directly related to model training: the loss function, the ML algorithm (which determines the form of the regularization term), the cross-validation scheme of parameter tuning, the frequency of hyperparameter re-tuning, the frequency of model re-fitting, and the estimation window.

In Table 2.1, we comprehensively review the specification choices made by existing ML earnings forecasting studies.[5] We identify six commonly used ML algorithms: OLS, Lasso, Ridge, Elastic Net (EN), Random Forest (RF), and Gradient Boosted Regression Trees (GBRT), each implemented with a diverse set of specification choices. From these studies, we also identify variations in four potentially important choices not directly related to model training. We detail these specification choices in Table 2.2 and offer an in-depth discussion regarding these choices in Section 2 of the Internet Appendix.

From the full combination of the choices listed in Table 2.2, we derive 3,024 ML models. We evaluate the forecasting performance of this exhaustive list of ML models to assess the impact of ML specification choices.

[4]The Internet Appendix Section 1 provides the OLS and LASSO models as two examples.

[5]Our review focuses on studies that compare the forecast accuracy of ML forecasts versus analysts' forecasts. A related but separate strand of literature studies the statistical earnings forecasts for firms not covered by analysts, such as Hou et al., 2012 and Chattopadhyay et al., 2023.

2.2.2 Model Performance Evaluation Metric

To enable a direct comparison to the earlier literature, we use the superiority measure from Bradshaw et al., 2012 to evaluate the out-of-sample forecasting accuracy. Specifically, we define ML Superiority for the forecast made in month t, for firm i and earnings with fiscal period end T, as follows:[6]

$$\begin{aligned} \text{ML superiority} \equiv & \left| \frac{\text{EPS}_{i,T}}{\text{Price}_{i,t}} - \frac{\text{Analysts Forecast}_t\left(\text{EPS}_{i,T}\right)}{\text{Price}_{i,t}} \right| \\ & - \left| \frac{\text{EPS}_{i,T}}{\text{Price}_{i,t}} - \frac{\text{ML Model Forecast}_t\left(\text{EPS}_{i,T}\right)}{\text{Price}_{i,t}} \right| \end{aligned} \tag{2.2}$$

We use "ML Superiority" to refer to the average ML Superiority of an ML model (over firm-month observations) unless otherwise stated. A more positive ML Superiority means ML forecasts are more accurate than analysts' forecasts. Because the measure is based on the absolute value of EPS forecast errors per dollar stock price, it is economically intuitive and allows comparison across firms. We follow Bradshaw et al., 2012 and winsorize analysts' error and ML error (scaled by price) to the range [-1, 1] throughout our analysis. We drop observations with the stock price less than or equal to $1 to minimize the impact of extreme values, but our results are robust to removing this filter. In all following analyses, the ML Superiority measure is annualized, meaning that all variables in Eq. (2.2) are multiplied by four for FQ earnings.

2.3 Data and Sample Construction

Our dataset consists of the intersection of firms in CRSP, Compustat, and I/B/E/S. Analysts' EPS forecasts (AF) in this paper refer to I/B/E/S median consensus analysts' forecasts, which are publicized

[6] When EPS is scaled by price in the forecasting step, ML Superiority is defined as the difference between $\left| \frac{\text{EPS}_{i,T}}{\text{Price}_{i,t}} - \frac{\text{Analysts Forecast}_t\left(\text{EPS}_{i,T}\right)}{\text{Price}_{i,t}} \right|$ and $\left| \frac{\text{EPS}_{i,T}}{\text{Price}_{i,t}} - \text{ML Model Forecast}_t\left(\frac{\text{EPS}_{i,T}}{\text{Price}_{i,t}}\right) \right|$.

on the third Thursday of each month. The two-year ahead annual earnings (FY2) refers to the forecast horizon of FY2 in I/B/E/S. The next quarterly earnings (FQ) refers to the forecast horizon of quarter 1 (Q1) in I/B/E/S when analysts' forecasts are available, and if not, it corresponds to quarter 2 (Q2).[7] Our FY2 forecasts data begins in January 1983, in line with Bradshaw et al., 2012, and ends in December 2020. Due to limited observations before 1985, the FQ forecast data starts in January 1985 and ends in December 2020.

Our predictor set consists of 77 features that include the WRDS Financial Suite Ratios (de Silva and Thesmar, 2022; van Binsbergen et al., 2022), the current month's stock price, return, six-month momentum, industry momentum, and market capitalization from CRSP, as well as analyst related variables: the current analysts' forecast, the three-month revision of the analysts' forecast, the most recently realized earnings (annual EPS for FY2 and quarterly EPS for FQ), the realized analysts' forecast error, and the distance between the current month and the end of the forecast period.[8] In robustness tests, we also include macroeconomic variables.

We require the current analysts' forecast, the most recently realized earnings, the stock price, and price-to-sales to be non-missing, following Bradshaw et al., 2012. Additionally, we require returns, market capitalization, and the two momentum variables to be non-missing. After cleaning our data, we are left with an average number of 2,421 and 2,338 firms each month in our test dataset for the FY2 and FQ forecasts,

[7] I/B/E/S consensus forecasts are calculated on the third Thursday. So, if Q1 earnings as of the third Thursday are announced before month-end, Q2 EPS becomes the next quarterly earnings at month-end. We find similar results if we use the detail file to compute the consensus forecasts at month-end. We prefer the I/B/E/S consensus file because it is more stable over different historical versions than the detail file (Call et al., 2021). Our results are also similar if we use the consensus mean analyst forecast rather than the consensus median.

[8] The realized analysts' forecast error is calculated using past FY2 (FQ) forecasts that have the same (or the most similar) distance between the forecast date and the end of the forecast period as the current FY (FQ) forecasts. See detailed definitions in Tables 4.9 and 4.10.

respectively. Table A3 describes the filters we apply to arrive at our final dataset. We then winsorize the variables in our predictor set on a monthly basis at the 1% and 99% levels.[9]

Finally, in each estimation window, we standardize the predictors in the training dataset and use the mean and standard deviation of the training dataset to standardize the predictors in the test dataset. Specifically, at the end of each month t, we train the model using the dataset $(y_{i\tau}, X_{i\tau})$ with the target variable $y_{i\tau}$ known between months $t-119$ (or the beginning of our sample if using the expanding training window) and t. We then apply the fitted model to the predictor values as of month t (i.e., predictors in the test set) and generate the predicted value for the target variable, which would be the ML forecast for month t. The forecast accuracy as measured by ML Superiority in Eq. (2.2) is evaluated on an out-of-sample basis. See the Internet Appendix Section 3 for a more detailed discussion of the timeline.

2.4 Impact of Machine Learning Model Specifications

2.4.1 What Specification Choices Matter

We start by presenting the distribution of ML Superiority for our 3,024 estimated machine learning models in Figure 2.1. We find significant variability in the accuracy of ML forecasts, as nearly 90 percent of machine learning models have a negative ML Superiority (i.e., underperforming analysts' forecasts). Thus, understanding the impact of specification choices is key to making sense of prior papers' results on machine or analysts' superiority.[10]

[9] We fill the missing values for variables with the FF38 industry median value and if unavailable, the cross-sectional median value in each month following the winsorization.

[10] Many studies report that ML models outperform analysts' forecasts: for short-horizon earnings from 1 to 3 quarters ahead (R. T. Ball and Ghysels, 2018, K. Cao and You, 2020, van Binsbergen et al., 2022, and Uddin et al., 2022) and for longer horizons from 1 to 3 years ahead (So, 2013, R. T. Ball and Ghysels, 2018, K. Cao and You, 2020, and van Binsbergen et al., 2022). A

To assess the impact of each specification choice listed in Table 2.2, we calculate the average ML Superiority and runtime over all possible combinations of choices across each of the choice sets. For instance, when evaluating estimation window choices, we compute the average ML Superiority and runtime for all ML models with an expanding window versus those with a rolling window. For conciseness, Table 2.3 presents results for choices associated with the highest and lowest ML Superiority within a specification choice set.

Table 2.3 shows that the loss function is the most critical factor affecting ML model performance. Opting for the MAE loss function over the MSE loss function boosts average ML Superiority by 6.32%. This improvement is economically large because given that the median absolute value of EPS over price is 6.07% for FQ (annualized by multiplying by 4) and 6.43% for FY2.[11] This result highlights the importance of handling the outliers in the target variable when implementing ML models for earnings forecasting.

The second most crucial specification choice is the CV scheme for hyperparameter tuning. We observe that using the time-series CV instead of the panel CV scheme (i.e., standard 5-fold CV) increases the average ML Superiority by 1.47%. Our comparative analysis addresses the gap noted in Bertomeu, 2020 that there is no guidance from theoretical or simulation studies on the appropriate CV approach when using accounting data. Our results indicate the assumption of independent observations inherent in standard k-fold CV is violated in the earnings forecasting setting, and support Bertomeu, 2020's advocacy for the use of time-series CV that preserves the data's temporal order.

notable exception is de Silva and Thesmar, 2022, who report that analysts' forecasts are superior for 1 quarter to 1 year ahead earnings, whereas ML forecasts are superior for 2-4 years ahead earnings.

[11]We show in Internet Appendix Section 4 that the best-performing choice within each specification choice set remains the same across the board when utilizing Mean Squared Errors rather than Mean Absolute Errors to compute the superiority measure.

The other three specification choices related to ML model training have a substantially smaller impact on model performance. Different choices of training windows, refitting the model, and tuning the hyperparameters result in a difference in average ML Superiority of 0.27%, 0.10%, and 0.09%, respectively.

For the four specification choice sets not directly related to ML model training, we find that 1) adopting Frankel and Lee, 1998's indirect approach to forecast earnings as opposed to forecasting EPS directly yields an increase in ML Superiority of 1.21%. This result suggests that ML forecasts are more accurate when the machine focuses on correcting predictable analysts' forecast biases. 2) Including analysts' forecasts in the predictor set improves ML Superiority by 0.68%, which corroborates the findings of de Silva and Thesmar, 2022; van Binsbergen et al., 2022 in a more exhaustive and definitive manner. 3) ML Superiority is higher for long-distance forecasts (FY2) than for short-distance forecasts (FQ) by 0.65%, aligning with existing research that shows analysts' forecasts are more accurate for near-term earnings. 4) Price scaling all EPS-related variables in the forecasting step enhances ML Superiority by 0.48%.

Our results in Table 2.3 thus pinpoint three key specification choices: MAE as the loss function, a time-series CV scheme, and the indirect forecasting approach. To highlight the importance of these choices, Figure 2.1 overlays the distribution of the ML Superiority for ML models that employ these three choices over the distribution of the ML Superiority for all models. We observe that under the constrained specification choices, ML models exhibit considerably less variation in performance and consistently outperform analysts' forecasts.

Table 2.3 also reports the computational runtime for each specification choice.[12] Although training and fitting all 3,024 ML models take approximately five years of machine computational runtime, the

[12]The computational runtime for a ML model includes the time spent on tuning hyperparameters, training the model, and generating the ML model forecast at the end of each month. To accelerate the process, we utilize large-scale computing servers at our university, allowing us to run computing jobs concurrently.

difference in runtime caused by each specification choice set is at the largest, 212 computational hours. In situations where faster computing is preferable, we recommend researchers use the three aforementioned specification choices and then select configurations from the remaining choice sets to minimize computing time without substantially impacting model performance. We offer these time-saving specification recommendations in the Internet Appendix Section 4.

Finally, we present the best-performing specification for each ML algorithm.[13] Table 2.4 shows that ML models with the best-performing specifications consistently outperform analysts' forecasts for both FQ and FY2. The ML Superiority varies from 0.091% (OLS) to 0.19% (GBRT) for FQ, and from 0.153% (OLS) to 0.603% (GBRT) for FY2. The GBRT algorithm delivers the highest ML Superiority for both FQ and FY2, indicating the presence of non-linear predictable relationships in the data. To examine whether the top-performing specifications are stable over time, we divide the thirty-year sample into three decade-long subsamples and evaluate the top-performing specifications within each decade. Section 5 of the Internet Appendix presents these results and shows that the top-performing models in each decade consistently use the four key specification choices mentioned above—the GBRT algorithm that exploits non-linear relationships, the MAE loss function that is robust to outliers, the indirect approach that focuses on minimizing analysts' forecast errors, and the time-series CV scheme that preserves time-series order in training and validation samples.

Overall, our findings identify significant variability in ML model performance and demonstrate the importance of four key specification choices that surpass the other specification choices in yielding the most statistically accurate earnings forecast.

[13]Given that most extant studies do not scale EPS in the forecasting step, we restrict our results to models with this scaling choice through the rest of the paper.

2.4.2 The Role of the Analysts' Forecasts

If the ML models with the best-performing specifications consistently outperform analysts' forecasts, what is the enduring relevance of analysts' forecasts in the machine learning era? To answer this question, we examine three scenarios that differ in how the ML models (with the best-performing specifications) encode information from analysts' forecasts. First, we use ML models to forecast EPS directly without analysts' forecasts included in the predictor set (direct approach w/o analysts); second, with analysts' forecasts included in the predictor set (direct approach w/ analysts); and third, with using the ML models to predict analysts' forecast errors (indirect approach).

Table 2.5 shows that when analysts' forecast variables are not included in the predictor set, none of the six ML models outperform analysts' predictions for forecasting FQ EPS, and only one out of six models (GBRT) significantly outperforms analysts for forecasting FY2 EPS. Even so, the outperformance of GBRT is only 0.205% (t-stat=2.39), which represents a modest improvement because it is equivalent to a reduction in absolute EPS forecast errors of 2.1 cents for a $10 stock.

With the inclusion of analyst variables in the predictor set under the direct forecasting approach, we see a substantial increase in ML Superiority across all models. Despite this improvement, five out of the six ML models still underperform analysts' FQ forecasts. The only exception is again the GBRT model, which demonstrates an economically small ML Superiority of 0.0921% (t-stat=3.51). For FY2 forecasts, all models except the RF outperform analysts, with the GBRT standing out as the top model with an economically large ML Superiority of 0.536% (t-stat=7.20).

When using the indirect approach, *all* ML algorithms with the best-performing specifications outperform analysts in a statistically significant way for both FQ and FY2. It is noteworthy that switching from

the direct forecasting approach (w/ analysts) to the indirect forecasting approach significantly narrows the performance gap between the worst and the best algorithm, from 2.5% to 0.1% for FQ and from 1.0% to 0.3% for FY2.

In summary, these findings demonstrate the indispensable role of analysts' forecasts in accurate earnings predictions: even sophisticated ML models struggle to match analysts' forecasts without including analysts' information. Once ML models are allowed to learn from analysts, all six ML models—including the elementary OLS model—–can exceed analysts in terms of accuracy.

2.5 When do ML Earnings Forecasts Excel

Having settled on the most appropriate machine learning specification to use, this section delves deeper into the economic magnitude of the superior performance of ML models in different situations. Prior studies such as Keung et al., 2010 show that investors are skeptical about positive earnings surprises that are smaller than 1 cent. So, we use 1 cent for a $10 stock as our cutoff for economic significance. To give ML models the best chance to outperform analysts, we focus on ML models with the best-performing specifications (see Table 2.4), and discussions are geared toward the top algorithm (i.e., GBRT).

2.5.1 Forecast Horizon Effect in ML Superiority

Our earlier results in Table 2.3 indicate that ML Superiority is smaller for FQ forecasts compared to FY2 forecasts. We conduct a more detailed analysis of the horizon effect in this subsection. Specifically, we follow the methodology in Bradshaw et al., 2012 and analyze how ML Superiority varies as the distance between the forecast date and the earnings announcement changes. For a given distance to an earnings

announcement (1-3 months for FQ and 12-23 months for FY2), Table 2.6 reports the mean ML Superiority for each ML algorithm.

Our results demonstrate a strong horizon effect. For FQ forecasts made one month prior to the earnings announcement, the GBRT model exhibits an ML Superiority of merely 0.086%. Though statistically significant, this ML Superiority is economically small: an accuracy improvement of less than 0.86 cents for forecasting FQ EPS for a $10 stock (annualized by multiplying by 4). All the other ML models exhibit even smaller improvements.

As the distance between the forecast date and earnings announcement increases, we see a notable increase in ML Superiority. For FQ forecasts generated 3 months prior to the earnings announcement, GBRT's ML Superiority quadruples to 0.273%. GBRT's ML Superiority for FY2 forecasts made 12 months before the earnings announcement is 0.518% (5.18 cents for forecasting FY2 EPS for a $10 stock), over six times larger than the ML Superiority for FQ forecasts made one month prior to the earnings announcement. This superiority further increases to 0.674% for forecasts made 19 months ahead of the earnings announcement, then tapers slightly to 0.571% for forecasts made 23 months in advance.[14] These results thus indicate that even the top ML model cannot improve upon analysts' forecasts for near-term earnings in an appreciable way, but it can bring substantial enhancement for FY2 forecasts.

2.5.2 Time-Series Variation in ML Superiority

While advancements in statistical models increase the model forecasting accuracy over time, analysts are also likely to incorporate these technological advancements in their forecasting processes to refine their

[14] Further supporting the horizon effect, we find in untabulated results that the average GBRT's ML Superiority for FY1 forecasts is 0.25%, which is between 0.19% for FQ forecasts and 0.60% for FY2 forecasts.

forecasts. This implies that the gap between analysts' forecasts and a given statistical model might shrink over time.

To test this hypothesis, we first compare analysts to the simplest statistical model: the random walk (RW) model. We plot the 12-month rolling average of the cross-sectional mean of ML Superiority for the RW model in panels A and B of Figure 2.2. In line with the hypothesis of analysts embracing technological improvements, the ML Superiority for the RW model for both FQ and FY2 consistently declines over time. Specifically, it decreases from an average of -1.65% and -0.52% in the first decade to -2.12% and -2.17% in the last decade for the FQ and FY2 forecasts, respectively.

We then compare analysts to the GBRT model under the direct approach without analysts' forecasts in the predictor sets (GBRT Direct w/o analysts). Panels A and B of Figure 2.2 show that the ML Superiority of this more advanced statistical model is higher, but it too shows a consistent downward trend, indicating that analysts' forecasts have improved even compared to more advanced ML models. Lastly, we compare analysts to the GBRT model under the indirect approach (GBRT Indirect, i.e., the top ML model). Once again, we observe a similar downward trend, suggesting a reduction in analysts' predictable errors over time.[15]

These time-series plots also reveal an intriguing business cycle variation in ML Superiority. The ML Superiority of all three statistical models exhibits noticeable decreases around the end of US recessions, as dated by the National Bureau of Economic Research (NBER). These results align with the idea that analysts are more forward-looking than statistical models trained on historical data. These patterns suggest a potential for improving ML model performance by integrating business cycle variables. However, in untabulated analysis, including four macroeconomic variables as used in van Binsbergen et al., 2022 in the

[15]In Internet Appendix Section 6, we present a more rigorous regression analysis of the time trend and find the negative time trend to be statistically significant in all three cases.

predictor set does not lead to any noticeable improvement in ML model performance either on average or during recessions. This result is reasonably expected given that US recessions are retrospectively dated by the NBER, and predicting the precise turning point of a recession using real-time data is notoriously difficult.

These results in Panels A and B of Figure 2.2 also raise an interesting question: given this downward trend in ML Superiority of the top ML model, has the ML model lost its advantage towards the end of our sample? To answer this question, we test the statistical significance of the ML Superiority for the top ML model in rolling 10-year windows in Panels C and D of Figure 2.2. We find that although the machine's edge over human analysts has reduced, the ML Superiority of the top ML model remains positive and statistically significant for both the FQ and FY2 forecasts toward the end of our sample. For example, in the last 10-year window from December 2010 to December 2020, the top ML model achieves an improvement over analysts that are equivalent to a reduction of MAE of 1.05 cent and 3.3 cents per $10 stock price for FQ and FY2 forecasts, respectively.

Taken together, our results suggest that analysts appear to be increasing their accuracy over time, that the best ML model maintains a decreasing, marginal advantage over analysts for now, and that ML models lag behind analysts when the macroeconomy is coming out of a recession.

2.5.3 Cross-Sectional Variation in ML Superiority

We next examine cross-sectional variations in ML Superiority. We build on prior literature that documents that the accuracy of analysts' earnings forecasts is associated with firm attributes related to information uncertainty, firm complexity, price informativeness, analysts' incentives, and earnings management. Table 4.1 in Appendix 2.12 provides detailed definitions of these firm attributes. We conduct a multi-

variate regression analysis to investigate the effects of these firm attributes simultaneously. We account for time-series and cross-sectional correlations by estimating Fama and MacBeth, 1973 (Fama-Macbeth) regressions and compute the $t-$statistics based on Newey-West standard errors with 24 lags.[16] For ease of interpretation, we use the normalized rank (i.e., the rank scaled by the number of stocks in a cross-section) of these firm characteristics as independent variables. Therefore, if one characteristic changes from the 25th percentile to the 75th percentile while the other characteristics remain the same, the corresponding change in ML Superiority is 0.5 times the respective multi-variate regression coefficient.

Table 2.7 reports the regression analysis of the ML Superiority of the top ML model (i.e., GBRT Indirect in Table 2.4) on these firm attributes. Consistent with the notion that higher information uncertainty and firm complexity are related to larger biases in analysts' forecasts (Zhang, 2006a), we find that the regression coefficients on 1/Size, the count of business segments, and idiosyncratic volatility are all positive and statistically significant across regressions for both FQ and FY2. We do not find statistically significant coefficients for R&D, although the positive coefficients are consistent with the findings that firms with higher R&D expenses are more difficult to forecast for analysts in Amir et al., 2003 and F. Gu and Wang, 2005.

When using institutional ownership (IO) as a proxy for analysts' incentives, we do not find a statistically significant relationship for either FQ or FY2. The only exception is the FY2 regression with R&D included (Column 5), in which we observe a negative and significant coefficient. This negative relation is consistent with the argument in Ljungqvist et al., 2007 and Frankel et al., 2006 that higher institutional

[16]We conduct separate regressions to investigate the effects of R&D and accrual volatility due to the substantial amounts of missing values for each variable. Because dollar trading volume and the bid-ask spread have a Spearman rank correlation above 0.9, to avoid co-linearity, we choose the bid-ask spread as our proxy for price informativeness in multivariate regressions. Please see Internet Appendix Section 7 to see the full correlation table.

ownership is associated with high analysts' forecast accuracy due to career concerns and brokerage profits (and thus low ML Superiority).

When using net external financing as a proxy for analysts' incentives for optimism, we find a highly statistically significant and positive relation for FY2, but not for FQ. The significant positive relationship for FY2 corroborates the findings in Ljungqvist et al., 2007 and Bradshaw et al., 2016 that analysts' forecasts are less accurate when there are incentives for them to generate investment banking business. The insignificant relationship for FQ supports the view that analysts have more incentives to be accurate for near-term earnings (Ham et al., 2022).

Using the bid-ask spread as a proxy for price informativeness, we find a positive relation across the regressions with varying degrees of statistical significance. This positive relation is consistent with the findings in Kerr et al., 2020 that analysts' forecast accuracy increases with price informativeness. Using accrual volatility (Dechow and Dichev, 2002) as a proxy for earnings management, we find positive coefficients on accrual volatility for both FQ and FY2, although only the FY2 coefficient is statistically significant. Our results are consistent with the findings in J. Abarbanell and Lehavy, 2003; Burgstahler and Eames, 2003 that analysts may not fully uncover management earnings manipulation, due to a lack of capability or willingness.

Comparing the magnitude of these regression coefficients, we observe that Size consistently commands the largest regression coefficients for both FQ and FY2. We therefore further report the ML Superiority of the GBRT Indirect model by size quintiles. Figure 2.3 shows that for stocks in the top size quintile, which account for 87% of total equity market capitalization, the ML Superiority is just 0.062% and 0.064% for FQ and FY2, respectively. These improvements are economically small because they are equivalent to a reduction in MAE of 0.62 and 0.64 cents for a $10 stock. In contrast, among firms in the bottom size

quintile, the ML Superiority is 0.569% for FQ (over 9 times larger than for large-cap firms) and 1.712% for FY2 (almost 30 times larger than for large-cap firms).

2.5.4 Discussion of Statistically Optimal Earnings Expectations

Our paper is the first to systematically examine the long list of machine learning models that represent the range of choices used in the existing literature. As a result of our exhaustive approach examining over 3,000 model specifications, we provide much more confidence in the results reported in some other contemporaneous and unpublished studies such as the importance of combining humans and machines (S. S. Cao et al., 2021; van Binsbergen et al., 2022), as well as the fact that machines produce better forecasts for smaller firms and longer forecast horizons (e.g., de Silva and Thesmar, 2022).

Our paper also uniquely offers three additional findings to this literature. First, we show that the "best" machine expectation uses a forecasting methodology not previously studied in recent machine learning earnings forecasting papers – what we call the indirect approach (designed explicitly to correct analyst forecast errors). Second, our time-series analysis indicates that "machine versus man" or even "machine plus man" is misguided. Our results suggest that the "best" machine forecasts rely on analysts' information, and the analysts likely use machines to help minimize their errors as we show that these two forecasts are converging over time. Finally, contrary to the impression made by recent studies, we demonstrate that for the vast majority of instances, the economic difference between the best machine forecast and analysts' forecasts is small, except for instances among smaller firms and longer horizon forecasts.

While the best machine forecast is statistically more accurate than analysts' forecasts, it remains to be seen whether investors' earnings expectations align more closely with machine or analyst forecasts. We explore this question in the next section.

2.6 Machine Learning Earnings Forecasts and Market Expectations

After determining the statistically optimal ML earnings forecast, we now examine the extent to which investors' expectations appear to be more in line with the statistically optimal machine learning forecast or analysts' forecasts.

To do so, we regress abnormal returns around earnings announcements on the machine learning forecast error and the analysts' forecast error and compare the extent to which each error explains returns as follows:

$$R_{d-1,d+1} = c + \beta \times \text{SUE}_{t-1,d} + \epsilon_{d-1,d+1} \tag{2.3}$$

, where $R_{d-1,d+1}$ is the characteristic-adjusted abnormal 3-day return (Daniel et al., 1997; hereafter DGTW) around the earnings announcement day d, $\text{SUE}_{t-1,d}$ is the earnings surprise measured as the difference between quarterly I/B/E/S reported earnings and the expected earnings (proxied by analysts' or machine FQ forecasts made at month end $t-1$ prior to the announcement) scaled by the price at the end of month $t-1$, and β is the earnings response coefficient (ERC).[17]

To the extent that earnings announcement returns are predominantly driven by revisions in market earnings expectations in reaction to the announcements, employing SUEs based on a more precise proxy

[17]We construct the earnings announcement days following the methodology in Dellavigna and Pollet, 2009; Johnson and So, 2018. The machine and analysts' forecasts are both made at the month end prior to the earnings announcement, but our results are robust to using versions of both forecasts from just before the earnings announcement. Similarly, we use the prior month end forecast and price, however these results are robust to instead using the price and analysts' consensus forecast from the day before the earnings announcement.

for market-expected earnings would lead to larger ERCs and increases in the adjusted R^2 of these regression estimations. If investors are fully rational, market-expected earnings should more closely align with the most statistically accurate forecast (i.e., the machine learning forecast). However, if investors share the same biases as analysts, as suggested by Bordalo et al., 2019, market-expected earnings may correlate more closely with analysts' forecasts.

We first examine the alignment of market-expected earnings with statistically superior earnings forecasts. We find that ML Superiority is strongly associated with ERCs and R^2s, with a rank correlation of 0.974 and 0.963, respectively. To visualize this strong positive association, we sort machine forecasts into quintile groups by their ML Superiority and show the distribution of the ERCs (R^2s) within each quintile in Figure 2.4. Compared to the analysts' forecasts-based SUE, SUEs based on the machine forecasts in the top ML Superiority quintile are predominantly associated with higher ERCs and R^2, while the opposite is true for the machine forecasts in the remaining quintiles. Corroborating our finding that the specification choices of the ML model matter, our findings show that 83% (80%) of machine forecasts are associated with lower ERCs (R^2s) than analysts' forecasts. Overall, Figure 2.4 indicates that market-expected earnings align more closely with the more accurate earnings forecasts.

We more formally test the extent to which investors' expectations align with the statistically optimal machine learning forecast or analysts' forecasts. Columns 1 and 2 of Table 2.8 report the univariate regression results for SUEs derived from analysts' forecasts and the statistically optimal forecast. The ERCs from both regressions are positive and highly significant. Consistent with our previous finding of a high correlation between analysts' FQ forecast and the best statistical FQ forecast, the magnitude of ERC in these two regressions is similar, with the one derived from SUE^{AF} being 18% lower. We then estimate the following bivariate regressions,

$$R_{d-1,d+1} = c + \beta_1 \times \text{SUE}^{ML}_{t-1,d} + \beta_2 \times \left(\frac{\text{ML}_{t-1} - \text{AF}_{t-1}}{P_{t-1}} \right) + \epsilon^{ML}_{d-1,d+1} \tag{2.4}$$

$$R_{d-1,d+1} = c + \gamma_1 \times \text{SUE}^{AF}_{t-1,d} + \gamma_2 \times \left(\frac{\text{ML}_{t-1} - \text{AF}_{t-1}}{P_{t-1}} \right) + \epsilon^{AF}_{d-1,d+1} \tag{2.5}$$

, where ML_{t-1} and AF_{t-1} are the best statistical FQ forecast and the analysts' FQ forecast (made at the month end prior to the announcement), respectively; the second term in the regression is equal to the difference between $\text{SUE}^{AF}_{t-1,d}$ and $\text{SUE}^{ML}_{t-1,d}$.[18]

To see why these regression coefficients are informative, suppose that the true SUE based on market expectation (SUE^{M}) is a weighted average of SUE^{AF} and SUE^{ML},

$$\text{SUE}^{M}_{t-1,d} = w_{AF}\text{SUE}^{AF}_{t-1,d} + w_{ML}\text{SUE}^{ML}_{t-1,d} \tag{2.6}$$

and the ERC based on the true SUE is as follows,

$$R_{d-1,d+1} = c + \beta^{M} \times \text{SUE}^{M}_{t-1,d} + \epsilon^{M}_{d-1,d+1}. \tag{2.7}$$

Given Eqs. (2.6) and (2.7), we can infer the relative weight the market places on analysts' forecasts and the best statistical forecast from the regression coefficients in Eqs. (2.4) and (2.5): $\frac{w_{AF}}{w_{ML}} = -\frac{\beta_2}{\gamma_2}$ and $\frac{\beta_2}{\beta_1} = \frac{w_{AF}}{w_{AF}+w_{ML}}$.[19] If the market expectation is rational in the sense that it places the same weight on analysts' forecasts as the best statistical forecast, then β_2 should be zero. In contrast, if the market

[18] An alternative specification is to regress $R_{d-1,d+1}$ on $\text{SUE}^{AF}_{t-1,d}$ and $\text{SUE}^{ML}_{t-1,d}$, but this specification suffers from multicolinearity as $\text{SUE}^{AF}_{t-1,d}$ and $\text{SUE}^{ML}_{t-1,d}$ are highly correlated.

[19] Substitute $\text{SUE}^{M}_{t-1,d}$ in Eq. (2.7) out using Eq. (2.6), we have

overweights analysts' forecasts relative to the best statistical forecast, then β_2 will be positive and β_2 will be larger when the market overweight analysts' forecasts more.[20]

Columns 3 and 4 of Table 2.8 present these regression results for Eqs. (2.4) and (2.5). We find that $\beta_2 = 0.07$ (t-stat = 5.7), thereby rejecting the hypothesis that the market expectation perfectly aligns with the statistical optimal forecast. A significant positive β_2 indicates that the market expectation places greater weight on analysts' forecasts than the statistical optimal forecast. Nevertheless, Column 4 shows that $\gamma_2 = -0.27$ (t-stat = -17.8), indicating that the market expectation places almost four times as much weight on the statistical optimal forecast as it does on the analysts' forecasts (i.e., $\frac{w_{AF}}{w_{ML}} = -\frac{\beta_2}{\gamma_2} = \frac{.07}{.27} \approx \frac{1}{4}$). Therefore, while the market expectation does not align perfectly with the statistically optimal forecast, it still aligns more with the statistically optimal forecast than with the analysts' forecast.[21]

We further investigate how the relative weight evolves over time and across varying levels of investor sophistication. To do so, we add interaction terms with a time trend and institutional ownership into Eq. (2.4). Column 5 of Table 2.8 shows that at the end of the sample period (i.e., the time trend is equal to 1), the regression coefficient on SUE^{ML} (β_1) is higher and the regression coefficient on the difference between machine and analysts' forecasts (β_2) remains the same. Given that $\frac{\beta_2}{\beta_1} = \frac{w_{AF}}{w_{AF}+w_{ML}}$, our results imply a

$$R_{d-1,d+1} = c + \beta^M \times (w_{AF} + w_{ML})\, \text{SUE}^{ML}_{t-1,d} + \beta^M \times w_{AF} \left(\frac{\text{ML}_{t-1} - \text{AF}}{P_{t-1}} \right) + \epsilon^M_{d-1,d+1}$$

$$R_{d-1,d+1} = c + \beta^M \times (w_{AF} + w_{ML})\, \text{SUE}^{AF}_{t-1,d} - \beta^M \times w_{ML} \left(\frac{\text{ML}_{t-1} - \text{AF}}{P_{t-1}} \right) + \epsilon^M_{d-1,d+1}$$

Therefore, the regression coefficients from Eqs. (2.4) and (2.5) have the following interpretation: $\gamma_1 = \beta_1 = \beta^M \times (w_{AF} + w_{ML})$, $\frac{\beta_2}{\beta_1} = \frac{w_{AF}}{w_{AF}+w_{ML}}$, and $\frac{\beta_2}{\gamma_2} = -\frac{w_{AF}}{w_{ML}}$.

[20] Our earlier results as well as the feature analysis in Section 2.8.1 show that the best-performing ML model relies heavily on analysts' forecasts as an input.

[21] In Internet Appendix Section 8, we address concerns that markets react differently to positive and negative earnings news by re-estimating our ERC analysis in subsamples based on positive and negative earnings surprises, and based on whether or not realized earnings are positive. We show that market expectations put an increased weight on the analyst in times with positive earnings surprises, and when earnings are positive. However, the market expectation still places a larger weight overall on the machine forecast, making it a stronger proxy for market expectations than the analysts' forecast.

gradual shift in market expectations over time, moving away from analysts' forecasts and leaning more towards the statistically optimal forecast. Yet, even at the end of the sample period, market expectations still overweight analysts' forecasts relative to the statistical optimal forecast with a relative weight $\frac{w_{AF}}{w_{ML}+w_{AF}}$ being $\frac{0.08-0.01}{0.24+0.15} \approx \frac{18}{100}$.

In Column 6 of Table 2.8, we use institutional ownership to proxy for investor sophistication. For ease of interpretation, we use an indicator variable that is equal to one when the institutional ownership of a stock is above the cross-sectional median institutional ownership for that month. We find that the regression coefficient on SUE^{ML} (β_1) is higher, and the regression coefficient on the difference between machine and analysts' forecasts (β_2) is lower for firms with above-median institutional ownership. Given that $\frac{\beta_2}{\beta_1} = \frac{w_{AF}}{w_{AF}+w_{ML}}$, these results imply a shift in market expectation away from analysts' forecasts and more towards the statistically optimal forecast for firms with more sophisticated owners. Specifically, the relative weight ($\frac{w_{AF}}{w_{ML}+w_{AF}}$) is equal to $\frac{0.08}{0.30} \approx \frac{26}{100}$ for firms with below median institutional ownership, and $\frac{0.08-0.06}{0.30+0.14} \approx \frac{5}{100}$ for firms with above median institutional ownership.

Overall, our results show that the market does not simply herd on analysts' forecasts. On the one hand, market expectations appear to integrate information beyond analysts' forecasts, akin to the optimal machine forecast. On the other hand, market expectations still tend to emphasize analysts' forecasts more than what the optimal machine forecast would recommend. This tendency to overvalue analysts' forecasts is less pronounced for stocks with higher institutional ownership and diminishes over our sample period, suggesting market expectations increasingly align with statistically optimal forecasts as stock ownership becomes more sophisticated and as time advances. Finally, we cannot conclude from these results that market expectations rely on a machine learning forecast, much less specify which machine learning forecast that might be. All we conclude is that when the market forms its expectations in whatever way that it

does, those expectations are more closely aligned with the most statistically optimal forecast in our sample than with the analysts' forecasts.

2.7 Machine Learning Earnings Forecasts and Implied Cost of Capital

Given that our results in Section 2.5 indicate that machines can offer more improvement over analysts' forecasts for longer horizon (i.e., FY2) earnings, we examine the extent to which the statistically optimal machine learning forecasts improve Implied Cost of Capital (ICC) estimates, as long-term earnings forecasts are a key input to these estimates.

We consider four commonly used ICC measures, proposed by Gebhardt et al., 2001 (GLS), Claus and Thomas, 2001 (CT), Ohlson and Juettner-Nauroth, 2005 (OJ) and Easton, 2004 (PEG), respectively. The first two are based on the residual-income model, whereas the latter two are based on the abnormal-growth-in-earnings model. Following the ICC literature, our analysis focuses on a composite ICC measure that is the equally weighted average of the four measures. Please see the Internet Appendix Section 9 for a detailed description of these measures.

We follow Lee et al., 2021 and use the measurement error variance (MEV) as the metric to evaluate the accuracy of various expected-return proxies (ERP). Lee et al., 2021 have demonstrated that minimizing the MEV is a necessary and sufficient condition for identifying the most accurate ERP. Our analysis focuses on the cross-sectional variation in expected returns because Lee et al., 2021 find that a composite measure of characteristic-based ERP (CER) outperforms ICCs in this setting and recommend that "researchers

utilizing cross-sectional designs in treatment effect studies should rely on characteristic-based ERPs rather than ICCs".

Cross-sectional MEV is defined as follows:

$$Var_t(\omega_{i,t}) = Var_t(\hat{er}_{i,t}) - 2Cov_t(r_{i,t+1,t+12}, \hat{er}_{i,t}) + Var_t(er_{i,t}),$$

where $Var_t(\hat{er}_{i,t})$ is the cross-sectional variance of a given ERP at the end of month t, $Cov_t(r_{i,t+1}, \hat{er}_{i,t})$ is the cross-sectional covariance between a given ERP and next-12-month realized returns, and $Var_t(er_{i,t})$ is the cross-sectional variance of the true expected returns. As the last term, $Var_t(er_{i,t})$, is the same for different ERPs, Lee et al., 2021 calculate $SVar_t = Var_t(\hat{er}_{i,t}) - 2Cov_t(r_{i,t+1}, \hat{er}_{i,t})$ to compare the accuracy of ERPs, with a lower average SVar indicating higher accuracy.

Panel A in Table 2.9 shows the summary statistics of the SVar for the composite ICC measure based on analysts' forecasts (ICC^{AF}), the composite ICC measure based on the statistical optimal forecast (ICC^{ML}), and CER.[22] Consistent with the findings in Lee et al., 2021, we find that CER has a lower average SVar than ICC^{AF}. However, using the statistically optimal earnings forecast improves the accuracy of the composite ICC, with the average SVar for ICC^{ML} being lower than that for CER.

Given our prior finding that the improvement of the statistically optimal forecasts over analysts' forecasts exhibits a pronounced size effect, we conduct our analysis across size quintiles. Specifically, we first sort firms into quintiles based on firm capitalization at the end of each month t and then compute $SVar_t$ in each size quintile. Panel B of Table 2.9 presents the average SVar for ICC^{AF} and ICC^{ML} across the size quintile, as well as their difference. We find that while ICC^{ML} has a lower average SVar than ICC^{AF} in all

[22]We thank Charles Lee, Eric So, and Charlie Wang for providing the CER data on the following website: https://leesowang2021.github.io/data/. Additionally, we report our results without the $1 price filter in the Internet Appendix section 9.

size quintiles, the magnitude of the difference in SVar between ICC^{AF} and ICC^{ML} decreases in firm size. Specifically, the difference is -0.2528 (t-stat = -5.25) in the bottom size quintile and -0.0070 (t-stat=-0.50) in the top size quintile, which represents an 80 times higher accuracy gain in the former group relative to the latter group.

In conclusion, we find that using the statistically optimal earnings forecasts rather than analysts' forecasts to compute ICCs leads to substantial accuracy gains among smaller firms. These results align with our previous results that machine forecasts offer larger improvements over analysts' forecasts for smaller firms and for longer forecast horizons.

2.8 Predictable Errors and Information in Analysts' Forecasts

2.8.1 Feature Importance

The top ML model outperforms analysts by correcting the predictable errors in analysts' forecasts. This prompts an intriguing question: what kind of errors in analysts' forecasts are detected by ML models? Conceptually, the predictable errors in analysts' forecasts may be due to analysts' inability to use all available information or their behavioral biases (Bertomeu, 2020). If the predictable errors are related to the analysts' inability to use all information, then the most important features of the top ML Model should capture information that is not contained in analysts' forecasts. However, if the predictable errors are related to behavioral biases, then analyst-related variables would be the most important features.

We use the drop-column feature importance to identify the most important feature of the top ML model. The basic idea is that if a feature is not important, excluding it from the predictor set should not noticeably decrease the model's out-of-sample performance. The drop-column feature importance

not only accounts for the inter-correlations between features but also accounts for the substitution effect between the dropped feature and the remaining features.[23] Specifically, we calculate the decrease in the ML Superiority measure as defined in Eq. (2.2) when feature k is excluded from the predictor set. We then scale the decrease by the ML Superiority measure when all predictors are used.

$$\%\Delta\text{superiority}_k = \frac{\text{superiority}_{\text{All}} - \text{superiority}_{\text{All}\setminus\{k\}}}{\text{superiority}_{\text{All}}} \tag{2.8}$$

A $\%\Delta\text{superiority}_k$ of 50% indicates that removing variable k from the predictor set reduces the model's superiority (relative to analysts' forecasts) by half.

Figure 2.5 presents the top 10 features for the top ML model for FQ and FY2 forecasts. The current analysts' forecasts (AF) show up as the most influential feature for both horizons. Removing AF from the predictor set results in a 44.5% reduction in ML Superiority for FY2 and a 22.9% reduction in ML Superiority for FQ. The stock price (PRC) and the realized forecast errors of prior analysts' forecasts (ErrAF) rank as the second most influential feature for FY2 and FQ, respectively. The removal of these features results in decreases of 22.9% and 17.7% in ML Superiority. All the other features are less important because removing them reduces ML Superiority by less than 10%. Our finding that AF is the most important feature for predicting analysts' forecast errors leans towards behavioral bias interpretations.

2.8.2 Analysts' Information

Our feature importance analysis highlights the important role of analysts' related variables. We thus introduce a novel analyst value add (AVA) measure to assess the additional information analysts provide to

[23]The Internet Appendix Section 10 provides more analysis of substantial non-linear and interactive relations in predicting analysts' forecast errors.

the best machine forecast. Specifically, our AVA measure is equal to the improvement in forecast accuracy resulting from the inclusion of analysts' forecast-related variables in the predictor set of the best machine forecast,

$$\text{AVA}_{i,t} \equiv \frac{|\text{EPS}_{i,T} - \text{GBRT Direct w/o Analysts}_t\,(\text{EPS}_{i,T})|}{\text{Price}_{i,t}} - \frac{|\text{EPS}_{i,T} - \text{GBRT Direct}_t\,(\text{EPS}_{i,T})|}{\text{Price}_{i,t}} \quad ,$$

where AVA for FQ forecasts are annualized by multiplying by 4.

Building upon our previous findings regarding the notable horizon and size effects in machine superiority over analysts' forecasts, we first evaluate the horizon and size effects in AVA. Figure 2.6 shows that AVA is larger for the shorter horizon and for smaller firms. The more pronounced AVA associated with FQ EPS is consistent with the higher incentives for accurate FQ forecast as noted by Ham et al., 2022, which underscores the challenge for machines to outperform analysts for short-horizon forecasts.[24] The larger AVA associated with smaller firms indicates that despite larger predictable errors, analysts' forecasts also carry considerable information essential for generating the most accurate statistical forecasts, especially among smaller stocks.

We further examine how the complexity and accessibility of financial statements affect analysts' information production. We control for the size and horizon effects in AVA by conducting the analysis within firm size quintiles and forecast horizons. We use three popular complexity measures: LN(Net File Size), LN(Complexity), and the Bog Index (Bonsall et al., 2017; Loughran and McDonald, 2023; Loughran and Mcdonald, 2014, 2016).[25] Table 2.10 presents the univariate regression of AVA on the standardized

[24] Section 11 of the Internet Appendix provides further analysis on the analyst incentives/effort as it relates to firm size and forecast horizon.

[25] We thank these authors for providing the Net File Size and Complexity data on the following website: https://sraf.nd.edu/ and the Bog Index measure on the following website: https://host.kelley.iu.edu/bpm/activities/bogindex.html.

values of these complexity measures, including month fixed effects to control for the time trend. Our findings reveal a generally positive association between AVA and the complexity of financial statements, confirming the notion of analysts as crucial information intermediaries.

We also explore the impact of accessibility improvements using the implementation of EDGAR and XBRL. Beginning in 1993, firms are required to submit their financial statements to the SEC via EDGAR, thereby enhancing access to 10K/10Q filings. Table 2.10 shows that for FQ forecasts, where analysts have higher incentives for accuracy, AVA exhibits a noticeable increase post-EDGAR implementation, ranging from 0.172% in the largest size quintile to 0.638% in the smallest size quintile. These increases are economically significant for the largest and smallest size quintiles, corresponding to improvements in forecasting accuracy (as measured by MAE) of 1.72 cents and 6.38 cents per $10 of stock price, respectively. The EDGAR effect is larger for smaller firms likely because more analysts started following smaller firms once financial information became available for them. Our findings corroborate the results from Gao and Huang, 2020, suggesting enhanced information generation by analysts following EDGAR's implementation. Table 2.10 also assesses the effect of XBRL tag implementation, which is voluntary starting in 2005 and becomes mandatory after 2009. Post-XBRL, analysts' information as measured by AVA decreases for larger firms, which is consistent with more public financial statement information being processed by investors.

Finally, we examine whether there is a broad time series decline in AVA as technological advancement may diminish the role of analysts as crucial information intermediaries. Figure 2.7 plots the 10-year rolling average of the AVA. We find that AVA is quite persistent and slightly *increasing* over time, suggesting that despite technological and data advancements, analysts' information remains vital for accurate forecast-

ing. This finding implies that even the most advanced machine models will continue to rely on analysts' information.

Overall, our research highlights the enduring importance of analysts' forecasts in achieving accurate earnings predictions. This relevance persists even in scenarios where machine expectations markedly outperform analysts and is undiminished in the face of advancing technology and the continuing expansion of data.

2.9 Conclusion

We comprehensively examine the superiority of various machine learning methods and analysts' forecasts in predicting earnings to provide an updated view on which earnings forecast minimizes ex-post forecast errors and which best aligns with investors' earnings expectations. To determine the most appropriate machine learning specification, we evaluate the impact of specification choices used in recent machine learning earnings forecast studies by exhaustively comparing 3,024 models derived from the full combination of nine sets of specification choices and six machine learning algorithms. We find that only a handful of specification choices significantly impact machine learning model forecast accuracy, with most having a minimal effect. Our analysis, complete with codes and estimates, provides a much needed bridge between the earlier literature and the recent (and future) machine learning earnings forecasting studies.

Contrary to the impression created by many recent machine learning studies, we find that even the best machine expectation is only marginally more accurate than analysts' forecasts in most cases, except when the earnings forecast is for small firms or for a longer horizon. Despite this fact however, we show that in cases where analyst and machine expectations differ, market expectations are closer to that of the

machine than to that of the analyst, which is consistent with investors exploiting the marginal improvement brought by machine forecasts. The market expectation is also closest to the strongest machine forecasts, demonstrating the importance of finding the appropriate machine learning specifications. Finally, machine learning models outperform analysts only when the model uses analysts' forecasts as an input. Overall, our results suggest that while the machine does outperform the analysts' forecast in most scenarios, this improvement is marginal among near-term forecasts, and among large-cap firms, making the easily obtainable analysts' forecast a viable option. However, machine learning forecasts can offer significant improvements in longer-term forecasts and for small-cap stocks, while also serving as a stronger proxy for short-term market expectations.

2.10 Figures

Figure 2.1: Distributions of ML Superiority

This figure presents the distribution of the ML Superiority of our 3,024 ML models. The gray bars show the distribution of the ML Superiority measure for all models. The green bars represent the subset of ML models that use the MAE objective function, the indirect forecasting method, and time-series cross validation.

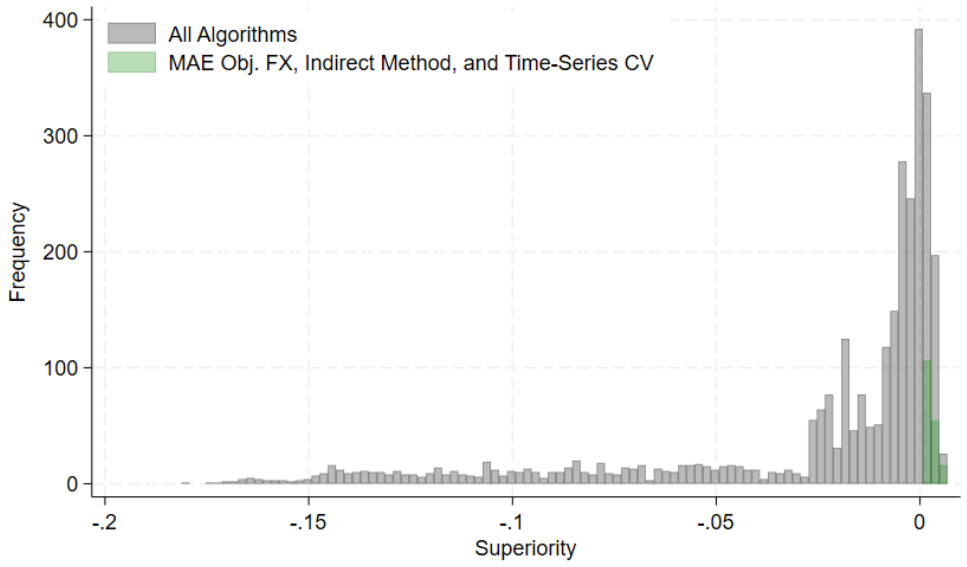
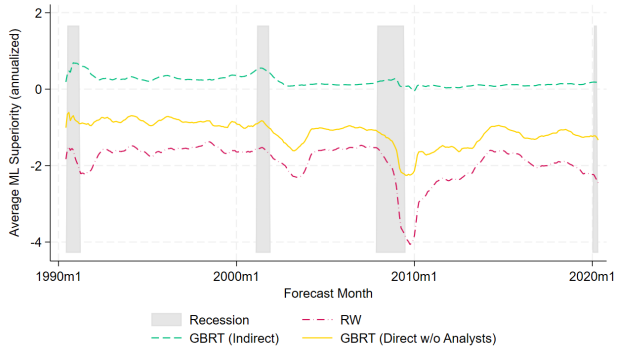
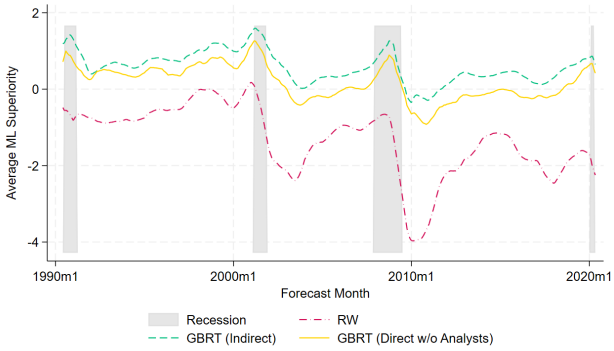


Figure 2.2: ML Superiority over Time

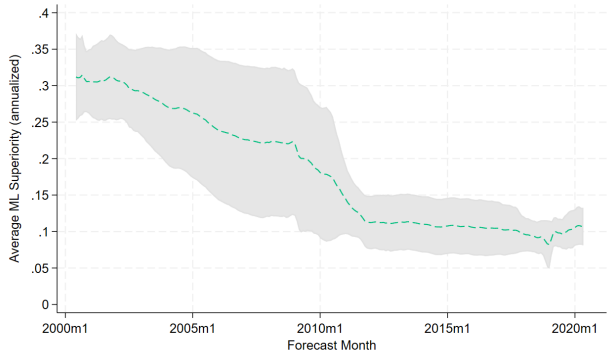
Panels (a) and (b) plot the 12-month rolling average (between months $t-11$ and t) of the cross-sectional mean of the ML Superiority measure for three ML models (with the best-performing specification): GBRT Indirect, GBRT Direct w/o Analysts, and RW (see details in the main text). Panels (c) and (d) plot the 10-year rolling average (between months $t-119$ and t) of the cross-sectional mean of the ML Superiority measure for the GBRT Indirect model. The 95% confidence intervals are generated using Newey and West, 1987 standard errors with 24 lags. Month t is shown on the x-axis. The ML Superiority is in percentage points. Shaded bars in subfigures A and B represent months in which the economy was in a recession as defined by the NBER.



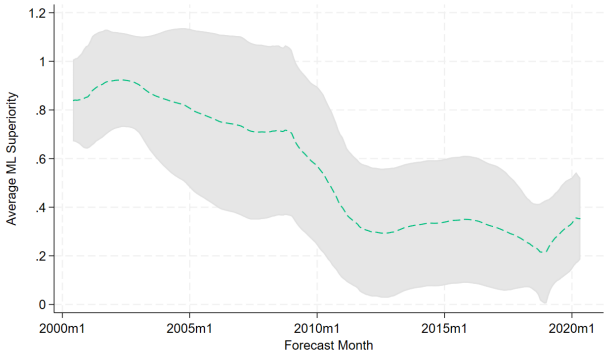
(a) FQ 12-Month Rolling Avg.



(b) FY2 12-Month Rolling Avg.



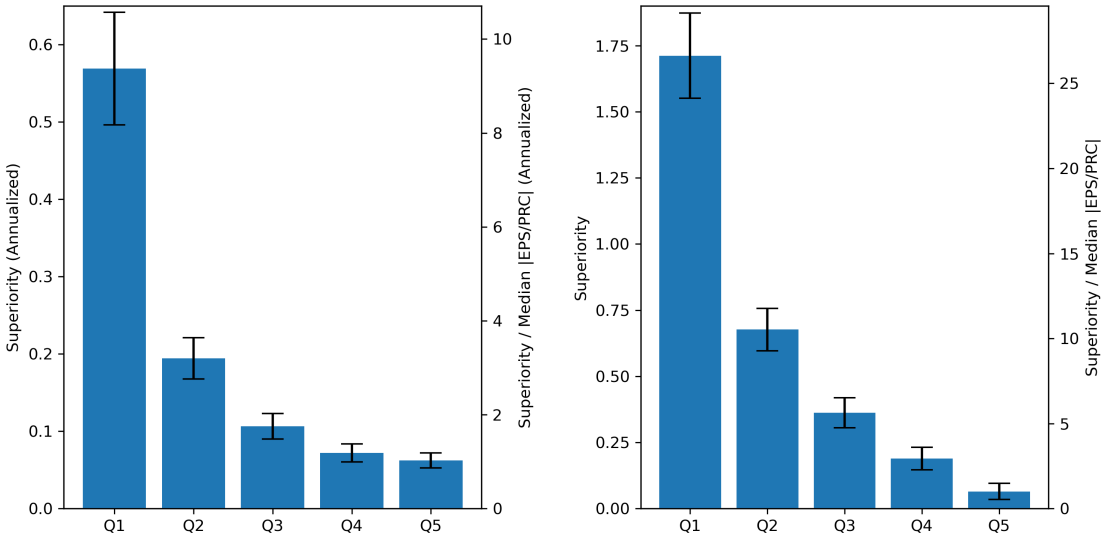
(c) FQ 10 Yr Rolling Average



(d) FY2 10 Yr Rolling Average

Figure 2.3: ML Superiority by Size Quintiles

This figure presents the average ML Superiority by size quintiles. We cross-sectionally sort the firms into size quintiles based on their market capitalization at the end of each month and calculate the panel average ML Superiority within each quintile. The ML Superiority is in percentage points, and for economic magnitude, we also report it as a percentage of the full-sample median |EPS/PRC| on the right y-axis. Whiskers denote 95% confidence bands. Standard errors of the resulting regression coefficients are two-way clustered at the firm and month level.

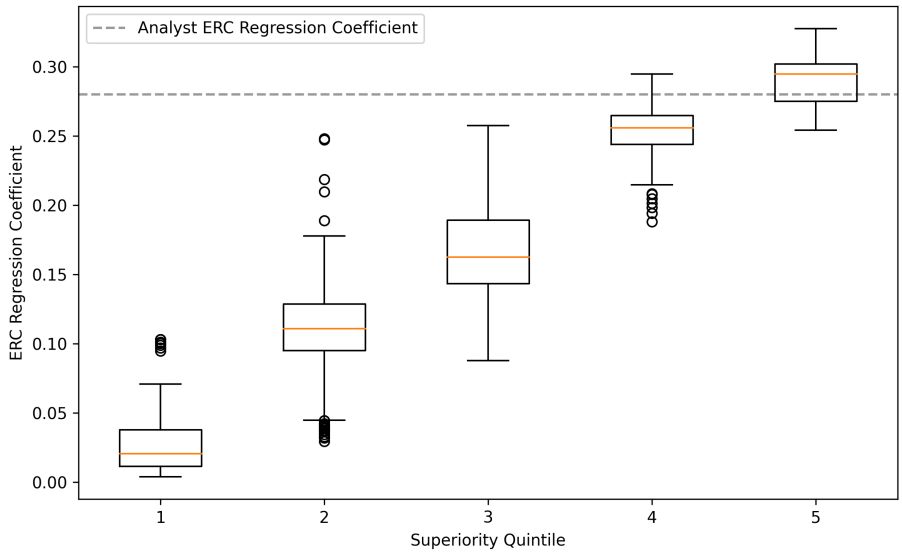


(a) FQ Forecast

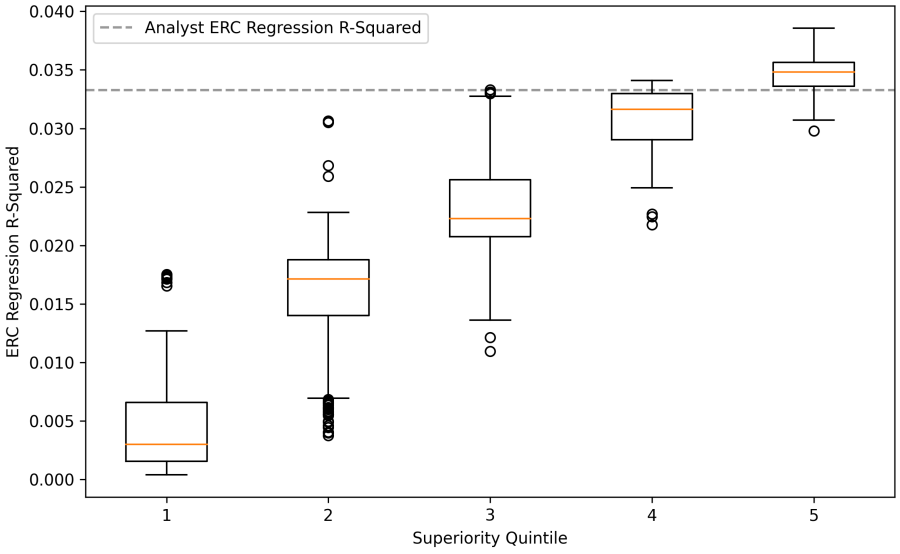
(b) FY2 Forecast

Figure 2.4: Earnings Regression Coefficients and R^2s

This figure reports the regression coefficients (ERCs) and R^2 from regressing 3-day DGTW adjusted announcement returns onto standardized unexpected earnings (SUE) following equation 2.3. We group FQ forecasts derived from 1512 ML models by their average ML Superiority into quintiles, and within each quintile, we present the box plot of the ERCs (Panel A) and R^2s (Panel B).The independent variables are winsorized in the full sample at the 1% and 99% level.



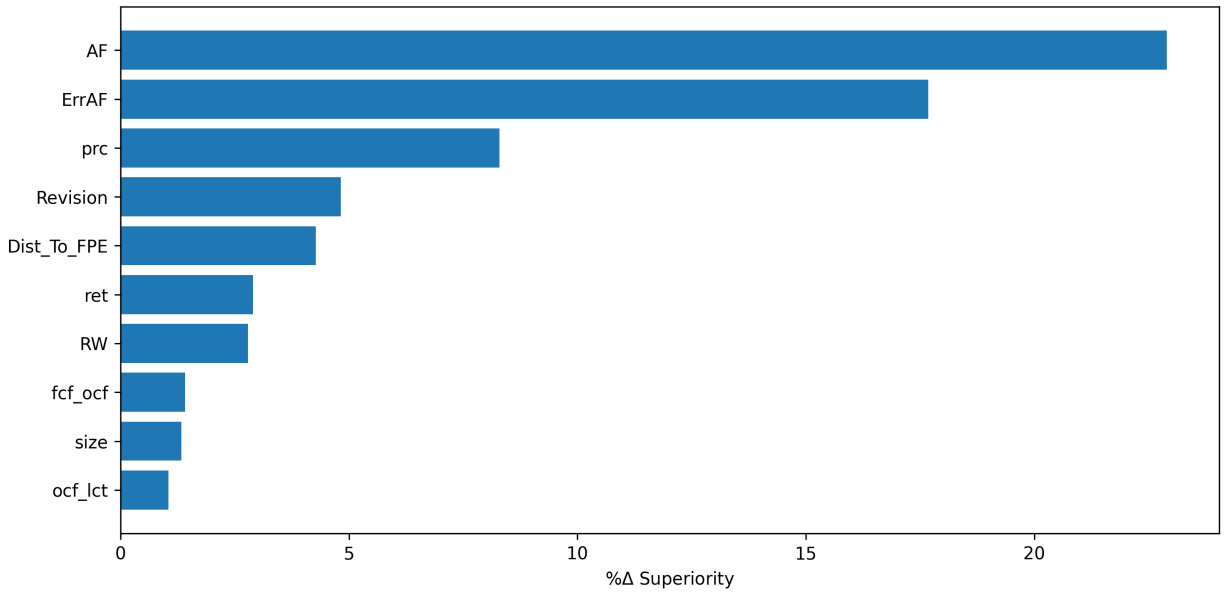
(a) Earnings Response Coefficients



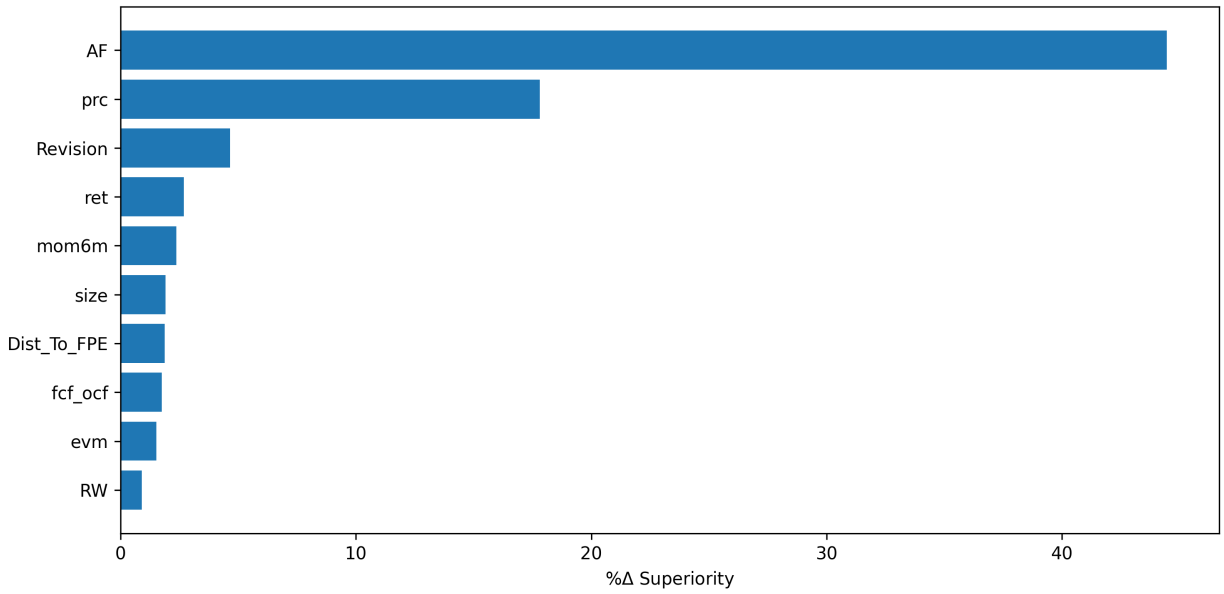
(b) ERC Regression R-Squared

Figure 2.5: Feature Importance Analysis

This figure presents the top 10 features of the top ML model for the FQ and FY2 forecasts, respectively. The feature importance (see Eq. 2.8) is based on the percentage change in ML Superiority when a feature is excluded from the predictor set. The reported numbers are in percentage points.



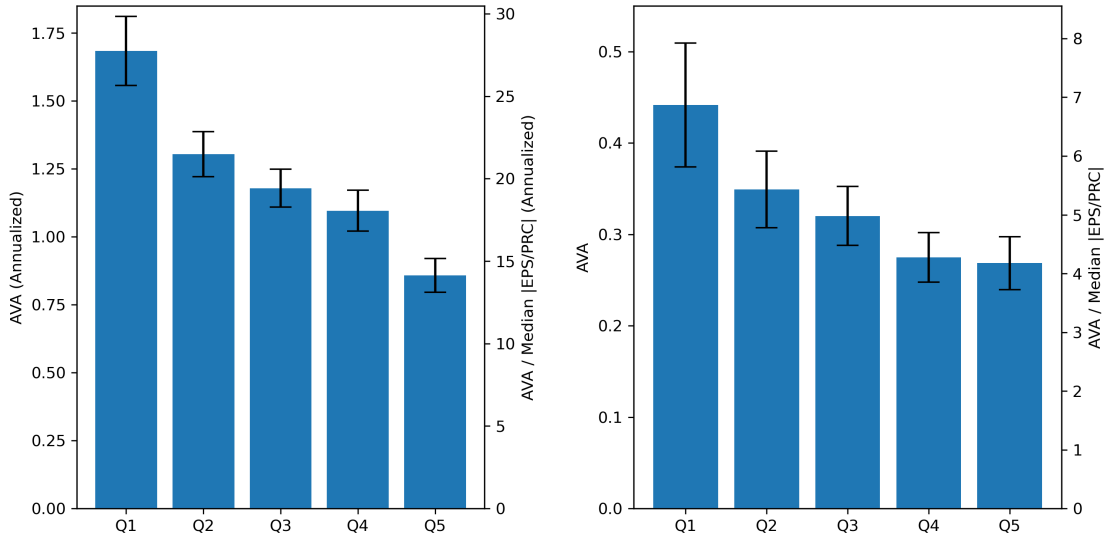
(a) FQ Forecast



(b) FY2 Forecast

Figure 2.6: AVA By Size Quintiles

This figure presents the analysts value added (AVA) by size quintiles. We cross-sectionally sort the firms into size quintiles based on their market capitalization at the end of each month and calculate the panel average ML Superiority within each quintile. The AVA is in percentage points, and for economic magnitude, we also report it as a percentage of the full-sample median |EPS/PRC| on the right y-axis. Whiskers denote 95% confidence bands. Standard errors of the resulting regression coefficients are two-way clustered at the firm and month level.

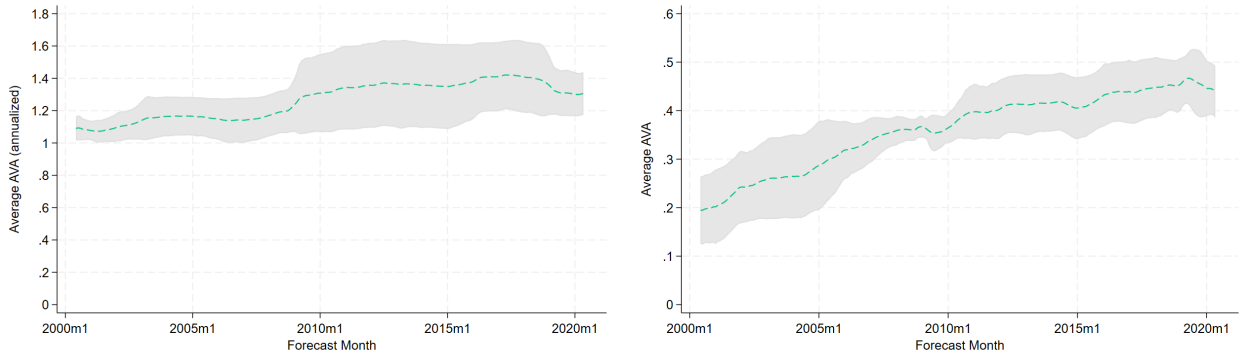


(a) FQ Forecast

(b) FY2 Forecast

Figure 2.7: AVA over Time

Panels (a) and (b) plot the 10-year rolling average (between months $t - 119$ and t) of the cross-sectional mean of the AVA. The 95% confidence intervals are generated using Newey and West, 1987 standard errors with 24 lags. Month t is shown on the x-axis. The AVA is in percentage points.



(a) FQ 10 Yr Rolling Average

(b) FY2 10 Yr Rolling Average

2.11 Tables

Table 2.1: Literature Review

Panel A: Superiority of Analysts' Forecasts Relative to Statistical Model Forecasts

		Analysts' Superiority		Paper's comments on Machine Superiority		
Paper	Journal	Conclusion	Evaluation Metric	Qualitative	Quantitative	Consistency with our paper
Bradshaw et al. (2012)	Review of Accounting Studies	Analyst superiority in short horizon (1 year) and RW superiority in longer horizons (2 and 3 year)	Mean of differences in absolute values of forecast error scaled by price (Eq. 3)	Analysts earnings forecasts consistently beat RW earnings forecasts over short windows, for longer forecast horizons, analysts superiority declines, and at certain horizons, analysts forecasts are dominated by RW forecasts.	For forecasts made in the same month as the earnings announcement, analysts forecasts are more accurate than RW forecasts by 282 basis points. In contrast, 11 months prior, analysts superiority is only 35 basis points...analysts forecasts are significantly more accurate than RW forecasts from 12 through 21 months prior... at month 21, analysts superiority is only 3 basis points, and by months 22 and 23, the RW forecast is significantly more accurate than analysts forecasts on average. However, the difference in accuracy is economically trivial, at 7 and 14 basis points respectively....mean analysts superiority from 24 through 35 months prior. Again, analysts superiority falls monotonically, from 66 basis points at 24 months prior to -41 basis points at 35 months prior	Consistent: RW underperforms AF in short horizons. Differ(Potential Reason): RW no longer outperforms analysts forecasts over longer horizons (Bradshaw et al. focuses on an earlier sample period)
So (2013)	Journal of Financial Economics	Machine superiority (1 year horizon)	Mean forecast errors (EPS level not scaled by price)	The median analyst forecast is generally above the median characteristic forecast, consistent with analysts facing incentives to issue optimistic forecasts.	The average characteristic forecast error per share is 0.112 (t-statistic=1.587), which is consistent with the average difference between realized and forecasted earnings being insignificantly different than zero. In contract, the average analyst error is -0.216 (t-statistic=-4.846) which is consistent with the average analyst forecast being optimistic.	Differ (Potential Reason): AF outperforms OLS when no analyst variables are included in predictor set (Evaluation metric: So examines the mean error rather than the accuracy measures such as MSE or MAE.
van Binsbergen et al. (2022)	Review of Financial Studies	Machine superiority (from 1 quarter to 2 year horizon)	Time series average of squared forecast errors (EPS level not scaled by price)	The mean squared errors of the machine learning forecast are smaller than the analysts mean squared errors (all horizons), demonstrating that our forecasts are more accurate than the forecasts provided by analysts.	The realized analysts forecasts errors...increase in the forecast horizon, ranging from 0.028 to 0.384 on average. All of these are statistically significantly different from zero....the time-series averages of the differences between the machine learning forecast and realized earnings are statistically indistinguishable from zero, with an average absolute value of around 0.001 for the quarterly earnings forecasts, 0.027 for the 1-year-ahead forecast, and -0.004 for the 2-year-ahead forecast.	Consistent: AF underperforms RF for FY2 Differ (Potential Reason): AF outperforms RF for FQ (Evaluation metric: van Binsbergen et al. do not scale MSE such that firms with large EPS likely dominate)
de Silva and Thesmar (2022)	WP	Analyst superiority for the short horizon 1 and 2 quarter and 1 year) and machine superiority in longer horizons (from 2 year to 4 year)	Mean squared forecast errors (scaled by price) normalized by mean squared realized EPS (scaled by price) (p.8)	At quarterly and the one-year horizons, we find the combined forecast cannot meaningfully beat the analyst consensus. At longer horizons, however, the combined forecast dominates by a large amount.	At longer horizons, however, the combined forecast dominates by a large amount: 9 and 21 percentage points of realized MSS at the three- and four-year horizons, respectively. However, comparing these results to those in Table 2 shows the improvement relative to pure econometric forecasts is small: around 1 to 3 percentage points	Consistent: we find consistent results when using MSE (with trimmed y-variable) as the objective function Differ (Potential Reason): AF's superiority over ML for the short horizon depends on the objective function (MSE vs. MAE) and whether we use direct or indirect methods.
Ball and Ghysels (2018)	Management Science	Machine superiority (1 quarter horizon)	Median absolute forecast errors (EPS level not scaled by Price) (Eq. 7)	Forecasts are more accurate...than analysts when forecast dispersion is high and when the firm size is smaller. In addition, we find that combining our MIDAS forecasts with analysts forecasts systematically outperforms analysts alone	Our findings are surprisingly sharp as we find that we are always better off combining MIDAScombination forecasts with those of analysts. At the beginning (end) of the target quarter, the combination of model-based and analyst forecasts reduced the forecast error by 21% (11%) relative to analysts forecasts alone. This means that the MIDAS and analyst forecasts feature complementary information.	Consistent: Our findings are consistent with Ball and Ghysels, despite using different models
Chen et al. (2022)	Journal of Accounting Research	Machine superiority (1 year horizon)	Area Under Curve (predicting the direction of changes in EPS)	Model outperforms conventional benchmarks, RW, and analyst forecasts.	The area under the receiver operating characteristics curve ranges from 67.52% to 68.66%, significantly higher than the 50% of a random guess. The annual size-adjusted returns to hedge portfolios formed based on the prediction of our models range from 5.02% to 9.74%.	Cannot compare (Chen et al. does not forecast EPS)
Cao and You (2021)	WP	Machine superiority (1-3 year horizon)	Statistical difference between time series average of mean absolute analyst error and mean absolute machine error (scaled by MVE)	First COMP_NL has significantly lower mean absolute forecast errors than analyst consensus earnings forecasts for all three forecast horizons	COMP_NL has significantly lower mean absolute forecast errors than analyst consensus earnings forecasts for all three forecast horizons (0.0541, 0.0679, and 0.0776 for COMP_NL vs. 0.0588, 0.0742, and 0.0922 for analyst forecasts).	Consistent: Our findings are consistent with Cao and You, despite forecasting EPS instead of Earnings
Uddin et al. (2022)	Quantitative finance	Machine superiority (1 quarter horizon)	Mean squared forecast errors (not scaled by price), mean absolute percentage error and R2 (for the last two, the minimum of the denominator is set to 0.005)	Combined with our data estimation technique, advanced machine learning algorithms provide a superior prediction of firms earnings.	With the help of CMF estimated data, XGBoost beats analysts consensus forecast by 34%.	Hard to compare: Uddin et al. focuses on a smaller dataset of 117 firms for out-of-sample model performance evaluation

Panel B: Specification Choices Directly Related to Model Training

			Hyper-Parameter Tuning				Model Fitting		
Paper	Model(s)	Loss Function	Method	Frequency	Parameter Range	Dataset(s)	Frequency	Forecast Horizon	Dataset Size
Bradshaw et al. (2012)	RW	N/A	N/A	N/A	N/A	N/A	Annual	1-3 years	1 year rolling window
So (2013)	OLS	MSE	N/A	N/A	N/A	N/A	Annual	1 year	1 year rolling window
van Binsbergen et al. (2022)	Modified RF	MSE	Temporal Training-Validation Split	Beginning of sample	Number Trees (1-2000), Depth (1-20), and Fraction of observations to sample (0.01-1)	Training: 1 year window Validation: 1 month window	Monthly	1-3 quarters and 1-2 years	12 month rolling window
de Silva and Thesmar (2022)	EN, RF, GBR	MSE	5 Fold Cross-Validation	Annual	EN: L1 ratio (0.1,0.99) RF: Number Trees (1000), Depth (4-8), Feature Fraction (0.3-1), Minimum samples in leaf (1-5), Samples to split node (2-10) GBT: Number Trees (500-10k), Depth (1-3), Learning rate (0.001-0.01)	5 year rolling window	Annual	1-4 quarters and 1-4 years	5 year rolling window
Ball and Ghysels (2018)	MIDAS (Mixed Data Sampling)	MSE	N/A	N/A	N/A	N/A	Quarterly	1 quarter	40 quarter rolling window
Chen et al. (2022)	RF, GBR	AUC	Temporal Training-Validation Split	Annual	RF: Number Variables (110-120), Number Trees (500-2000),Minimum samples in leaf (1-4), Fraction of observations to sample (0.5) GBR: Number Trees (500-2000), Learning Rate (0.005, 0.01, 0.05), Max Depth (1-4), Minimum samples in leaf (10), Fraction of observations to sample (0.5)	Training: 2 year rolling window Validation: 1 year rolling window	Annual	1 year	1 year rolling window
Cao and You (2021)	OLS, LASSO, Ridge, RF, GBR, ANN	MSE (All but GBR) Huber (GBR)	5 Fold Cross-Validation	ND	Lasso: L1 penalty (1e-3-1e-1) Ridge: L2 penalty (5e1-1e3) RF: Depth (20, 25, 30, 35), Minimum samples in leaf (15, 20, 25, 50) GBR: Depth (1, 3, 5) Minimum samples in leaf (75, 100, 125, 150) ANN: Activation (RELU, Tanh) Alpha (1e-3, 1e-4, 1e-5), Hidden layers ([64,32,16,8], [32,16,8,4], [16,8,4,2], [64,32,16], [32,16,8], [16,8,4], [8,4,2], [63,32], [32,16], [16,8], [8,4], [4,2], [64], [32], [16], [8], [4])	ND	Annual	1-3 years	10 year rolling window
Uddin et al. (2022)	LASSO, XGB, SVR	MSE	10 Fold Cross Validation	Beginning of sample	Lasso: L1 penalty (1e-5, 1e-4,1e-3, 1e-2, 1e-1) XGBoost: Learning rate (1e-5, 1e-4, 1e-3, 1e-2), Regularization (1e-1, 1e-4, 1e-3, 1e-2) SVR: Epsilon (1e-5, 1e-4, 1e-3, 1e-2), Constant (1, 2, 3, 5, 10)	60 Quarters	Quarterly	1 quarter	Expanding window starting with 59 quarters

Panel C: Other Specifications Choices

Paper	Y Variable	Predictor Set	Missing Values Imputation	Data Transformation	
				X Variables	Y Variables
Bradshaw et al. (2012)	EPS	EPS	ND	ND	ND
So (2013)	EPS	Financial statement variables (Earnings per share, loss indicator, pos./neg. accruals per share, asset growth, dividend indicator, book to market, share price, dividend per share)	Drop observation if missing variables	ND	ND
van Binsbergen et al. (2022)	EPS	WRDS Financial Ratio Suite, Realized Earnings, stock price, stock returns, analysts forecasts, and macro mariables	Replace with industry median	Winsorize and then standardize	ND
de Silva and Thesmar (2022)	Earnings/Price	WRDS Financial Ratio Suite, SIC, stock Returns, stock price, 5-year monthly return volatility, analysts forecasts, number of forecasts, total assets, fiscal year dummies. Include 2 lags of variables	Fill Missing with Zero	Trim at 5x IQR and then standardize	Winsorize forecasts of EPS and EPS at 10 times their interquartile range
Ball and Ghysels (2018)	Change in EPS	Change in inventory, A/R, CAPEX, gross margin, and SG&A, abnormal stock return, return volatility, macro economic variables, analysts forecast	ND	ND	ND
Chen et al. (2022)	Indicator for change in EPS (-1 decrease and 1 increase)	10-K XBRL items, lag of 10-K XBRL items, percentage change with lag of 10-K XBRL items, Nissim and Penman (2001) variables, Ou and Penman (1989), analysts forecast, stock price data	Exclude all custom and uncommon tags. Fill missing values with zero	ND	ND
Cao and You (2021)	Earnings	30 Financial statement variables and their first difference. (Historical earnings and major components (8), Income statement items (5), Balance sheet items (16), OCF)	Drop observation if missing key variables and fill some line items with zero. Then drop any observations if missing variables	ND	ND
Uddin et al. (2022)	EPS	Change in Inventory, A/R, CAPEX, gross margin, and SG&A, abnormal stock return, return volatility, stock price, leverage, total assets, dividends, dividend dummy, net income, negative net income dummy, change in total assets, BTM, MVE, accruals, analysts forecasts	Test different methods: Matrix factorization and coupled matrix factorization	ND	ND

Table 2.2: Nine Sets of Specification Choices Evaluated

This table lists the nine sets of specification choices we evaluate. The complete combination of these specification choices results in 3,024 ML models, with 576 configurations for Lasso, Ridge, Elastic Net, RF, and GBRT and 144 configurations for OLS because it does not require hyperparameter tuning.

Specification	Choices
Loss Function	MAE MSE MSE with trimmed y-variable (1% and 99% in the train set)
Cross-Validation	Time-series Panel
Estimation Window	Rolling Expanding
Hyper-Parameter Tune Frequency	Beginning Annual
Refit Frequency	Yearly Monthly
Forecasting Approach	Direct Indirect*
Predictor Set	With Analyst Variables Without Analyst Variables
Scaling in Forecasting Step	None Price
Forecast Period	FQ FY2

*The indirect forecasting approach predicts EPS in two steps. First, we forecast analysts' forecast error; second, we adjust analysts' forecasts for the predicted errors to arrive at the final EPS forecasts.

Table 2.3: The Impact of Specification Choices

This table analyzes the impact of the 9 specifications choice sets summarized in Table 2.2. For each option within a given specification choice set, we compute the average ML Superiority and machine runtime (in hours) over all the possible combinations of the specifications from the remaining eight sets of specifications. For example, when evaluating the estimation window, we take the average ML Superiority and machine runtime over all models with an expanding window and then do the same for all models with a rolling window. For brevity, we present the results for options with the best and the worst ML Superiority for each choice set and report the resulting difference in the Diff column. The %Diff column reports the values in the Diff column as percentage points of the median $|EPS/PRC|$. ML Superiority is in percentage points, and a higher value means greater outperformance relative to analysts' forecasts. The sample contains forecasts made between June 1990 and December 2020.

			Superiority				Run Time (Hrs)		
	Best Sup.	Worst Sup.	Best	Worst	Diff	% Diff	Best	Worst	Diff
Loss Function	MAE	MSE	-0.37	-6.69	6.32	101.1	226.4	77.8	148.6
Cross-Validation	Time-series	Panel	-1.70	-3.17	1.47	23.5	113.0	158.8	-45.8
Estimation Window	Rolling	Expanding	-2.40	-2.67	0.27	4.4	91.4	168.0	-76.6
Refit Frequency	Monthly	Yearly	-2.49	-2.59	0.10	1.7	151.9	107.4	44.5
Param. Frequency.	Annual	Beginning	-2.39	-2.48	0.09	1.5	241.9	29.9	212.0
Forecasting Approach	Indirect	Direct	-1.51	-2.72	1.21	19.4	126.2	133.4	-7.2
Predictor Set	w/ Analyst	w/o Analyst	-2.72	-3.39	0.68	10.8	133.4	129.4	4.0
Scaling in Forecasting Step	PRC	None	-2.30	-2.78	0.48	7.7	126.2	133.2	-7.0
Forecast Period	FY2	FQ	-2.21	-2.86	0.65	12.8	125.8	133.5	-7.7

Table 2.4: The Best Performing Specifications

This table shows the best-performing choices for specifications related to ML model training for each ML algorithm (using the indirect approach, MAE objective function, and time-series CV scheme). Panel A (B) reports the best-performing specifications for the FQ (FY2) forecasts. The associated ML Superiority is shown in percentage points and the associated runtime is shown in hours. The sample contains forecasts made between June 1990 and December 2020.

Panel A: FQ Specifications

Algorithm	Window	Loss Function	Param Freq	Refit Freq	CV Scheme	Superiority	Run Time
OLS	Rolling	MAE		Monthly		0.113	1.16
Lasso	Rolling	MAE	Beginning	Monthly	Time-series	0.113	1.38
Ridge	Rolling	MAE	Beginning	Monthly	Time-series	0.113	1.39
EN	Rolling	MAE	Annual	Monthly	Time-series	0.113	14.89
RF	Rolling	MAE	Annual	Monthly	Time-series	0.116	23.83
GBRT	Rolling	MAE	Annual	Monthly	Time-series	0.201	73.42

Panel B: FY2 Specifications

Algorithm	Window	Loss Function	Param Freq	Refit Freq	CV Scheme	Superiority	Run Time
OLS	Rolling	MAE		Monthly		0.267	1.05
Lasso	Rolling	MAE	Beginning	Monthly	Time-series	0.300	1.01
Ridge	Rolling	MAE	Beginning	Monthly	Time-series	0.306	1.59
EN	Rolling	MAE	Beginning	Monthly	Time-series	0.309	1.97
RF	Rolling	MAE	Beginning	Monthly	Time-series	0.365	2.47
GBRT	Expanding	MAE	Annual	Monthly	Time-series	0.601	121.35

Table 2.5: The Role of the Analysts' Forecasts

This table provides the average ML Superiority by ML algorithm using the direct method (w/o analysts), direct method (w/ analysts), and indirect method for FQ and FY2 forecasts. The ML Superiority is shown in percentage points. Standard errors are clustered by firm and the year of fiscal period end. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The sample contains forecasts made between June 1990 and December 2020.

	(1) N	(2) RW	(3) OLS	(4) LASSO	(5) Ridge	(6) EN	(7) RF	(8) GBRT
FQ Direct w/o Analysts	1080680	-1.868***	-1.722***	-1.721***	-1.724***	-1.731***	-2.629***	-1.132***
		(-19.34)	(-25.12)	(-25.14)	(-25.10)	(-25.56)	(-25.75)	(-18.42)
FY2 Direct w/o Analysts	973417	-1.211***	-0.460***	-0.434***	-0.461***	-0.420***	-0.558***	0.205**
		(-7.25)	(-4.51)	(-4.50)	(-4.52)	(-4.51)	(-6.14)	(2.39)
FQ Direct w/ Analysts	1080680		-0.113***	-0.117***	-0.120***	-0.113***	-2.369***	0.0921***
			(-4.81)	(-5.04)	(-5.03)	(-5.01)	(-23.06)	(3.51)
FY2 Direct w/ Analysts	973417		0.167**	0.165**	0.168**	0.168**	-0.452***	0.536***
			(2.41)	(2.38)	(2.43)	(2.44)	(-4.84)	(7.20)
FQ Indirect	1080680		0.113***	0.113***	0.113***	0.113***	0.116***	0.201***
			(7.52)	(7.48)	(7.47)	(7.76)	(7.95)	(8.83)
FY2 Indirect	973417		0.267***	0.300***	0.306***	0.309***	0.365***	0.601***
			(4.73)	(5.70)	(5.75)	(5.82)	(6.23)	(7.66)

Table 2.6: Horizon Effect: Distance to Earnings Announcement

This table provides the average ML Superiority by ML algorithm for a given distance to earnings announcements. We compute the ML Superiority for FQ forecasts with distances between 1 and 3 months and FY2 forecasts with distance between 12 months and 23 months. The ML Superiority is shown in percentage points. Standard errors are clustered by firm and the year of fiscal period end. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The sample contains forecasts made between June 1990 and December 2020.

	Distance to Announcement	(1) N	(2) OLS	(3) LASSO	(4) Ridge	(5) EN	(6) RF	(7) GBRT
FQ	1	350470	0.0361***	0.0360***	0.0359***	0.0414***	0.0558***	0.0856***
			(4.41)	(4.42)	(4.34)	(5.07)	(6.43)	(6.19)
	2	357321	0.0855***	0.0854***	0.0856***	0.0871***	0.101***	0.161***
			(6.46)	(6.38)	(6.43)	(6.74)	(7.81)	(7.71)
	3	309094	0.165***	0.165***	0.165***	0.161***	0.153***	0.273***
			(7.70)	(7.68)	(7.67)	(7.89)	(7.68)	(9.36)
FY2	12	49910	0.264***	0.309***	0.304***	0.306***	0.297***	0.518***
			(4.85)	(6.02)	(5.78)	(5.86)	(5.11)	(7.12)
	13	83462	0.201***	0.255***	0.246***	0.249***	0.265***	0.457***
			(3.91)	(5.02)	(4.85)	(4.93)	(4.57)	(6.83)
	14	83113	0.207***	0.259***	0.257***	0.261***	0.287***	0.482***
			(3.96)	(5.04)	(5.01)	(5.08)	(4.62)	(6.44)
	15	82799	0.218***	0.265***	0.266***	0.270***	0.306***	0.514***
			(3.77)	(4.85)	(4.83)	(4.91)	(4.60)	(6.51)
	16	82367	0.299***	0.335***	0.340***	0.344***	0.388***	0.608***
			(4.68)	(5.51)	(5.59)	(5.67)	(5.75)	(7.02)
	17	81897	0.300***	0.325***	0.337***	0.340***	0.401***	0.627***
			(5.12)	(6.00)	(6.11)	(6.18)	(6.34)	(7.56)
	18	81428	0.284***	0.313***	0.323***	0.325***	0.397***	0.642***
			(4.71)	(5.65)	(5.71)	(5.79)	(6.43)	(7.92)
	19	80561	0.304***	0.331***	0.340***	0.342***	0.419***	0.674***
			(5.38)	(6.37)	(6.44)	(6.51)	(7.23)	(8.29)
	20	78987	0.270***	0.296***	0.308***	0.311***	0.396***	0.646***
			(4.65)	(5.60)	(5.73)	(5.82)	(6.93)	(7.97)
	21	76835	0.247***	0.278***	0.287***	0.289***	0.373***	0.639***
			(4.20)	(5.08)	(5.24)	(5.31)	(6.25)	(8.01)
	22	73554	0.257***	0.284***	0.295***	0.298***	0.380***	0.630***
			(3.98)	(4.91)	(5.04)	(5.12)	(6.12)	(7.25)
	23	69194	0.218***	0.241***	0.251***	0.253***	0.340***	0.571***
			(2.97)	(3.81)	(3.81)	(3.89)	(5.67)	(5.84)

Table 2.7: Cross-Sectional Variations in ML Superiority

This table presents Fama-MacBeth regressions of the ML Superiority of the top ML model on size, count of business segments, institutional ownership, idiosyncratic volatility, net external financing, bid-ask spread, R&D, and accrual volatility. All independent variables are the normalized rank (i.e., the rank based on the variable of interest scaled by the number of stocks in the cross-section) between 0 and 1. The ML Superiority is shown in percentage points. Standard errors are computed based on Newey and West, 1987 with 24 monthly lags. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The sample contains forecasts made between June 1990 and December 2020.

	FQ			FY2		
	(1)	(2)	(3)	(4)	(5)	(6)
1/Size	0.422***	0.458***	0.359***	1.271***	1.719***	1.035***
	(4.0)	(4.3)	(4.3)	(5.3)	(7.5)	(6.5)
Count of Business Segments	0.138***	0.180***	0.108***	0.406***	0.575***	0.244***
	(4.2)	(3.7)	(2.9)	(4.0)	(4.7)	(3.0)
Institutional Ownership	-0.050	-0.066	-0.036	-0.175	-0.401***	-0.198*
	(-1.6)	(-1.6)	(-0.9)	(-1.5)	(-3.6)	(-1.9)
Idiosyncratic Volatility	0.218***	0.218***	0.211***	0.837***	0.433***	0.702***
	(3.8)	(3.4)	(6.0)	(5.8)	(3.9)	(5.8)
Net external financing	-0.019	-0.080***	0.035	0.498***	0.535***	0.453***
	(-1.0)	(-3.0)	(1.6)	(7.8)	(6.8)	(6.0)
Bid-Ask Spread	0.123*	0.097	0.122**	0.219*	0.286***	0.175**
	(1.9)	(1.5)	(2.2)	(2.0)	(2.6)	(2.1)
R&D		0.047			0.112	
		(1.2)			(0.7)	
Accrual Volatility			0.016			0.172***
			(0.4)			(3.1)
Observations	868874	482878	513333	793580	436125	477737

Table 2.8: Earnings Response Coefficients

This table reports the regression results of the DGTW adjusted 3-day earnings announcement returns $R_{d-1,d+1}$ onto $\text{SUE}^{AF}_{t-1,d}$ and $\text{SUE}^{ML}_{t-1,d}$, as defined in Eqs. (2.4) and (2.5), as well as their difference defined as Bias $\equiv \text{SUE}^{AF}_{t-1,d} - \text{SUE}^{ML}_{t-1,d} = \frac{\text{ML}_{t-1} - \text{AF}}{P_{t-1}}$. Time is a continuous variable that increases linearly over the course of our sample, starting at 0 in June 1990 and ending at 1 in December 2020. IO is the percentage of 13F institutional ownership. SUE^{AF} and SUE^{ML} are winsorized in the full sample at the 1% and 99% level. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. Standard errors are clustered at the firm and month level.

	(1)	(2)	(3)	(4)	(5)	(6)
SUE^{ML}	0.34***		0.34***		0.24***	0.30***
	(112.66)		(113.19)		(39.00)	(85.64)
SUE^{AF}		0.28***		0.34***		
		(104.10)		(113.19)		
ML-AF			0.07***	-0.27***	0.08***	0.08***
			(13.82)	(-44.55)	(7.61)	(12.45)
SUE^{ML}*Time					0.15***	
					(18.72)	
(ML-AF)*Time					-0.01	
					(-0.48)	
Time					-0.07*	
					(-1.72)	
SUE^{ML}*IO Med.						0.14***
						(20.46)
(ML-AF)*IO Med.						-0.06***
						(-4.36)
IO Med.						0.29***
						(9.39)
Const.	0.11***	0.25***	0.15***	0.15***	0.20***	-0.01
	(7.32)	(16.42)	(9.74)	(9.74)	(6.19)	(-0.25)
R-Squared (%)	3.874	3.327	3.932	3.932	4.041	4.091
# Obs.	314937	314937	314937	314937	314937	313728

Table 2.9: Implied Cost of Capital

This table shows our SVar analysis of the implied cost of capital (ICC) estimates, following the methodology in Lee et al., 2021. In panel A, we provide the summary statistics of SVar for ICC^{AF}, ICC^{ML}, and the characteristic-based expected return (CER). In panel B, we show the average SVar for ICC^{AF} and ICC^{ML} across size quintiles and test whether their differences are statistically different. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. Standard errors of the test statistics are computed based on Newey and West, 1987 with 12 lags. We require the ICC^{AF}, ICC^{ML}, and CER be non-missing for a firm-month to be included in the SVar calculation. All variables are winsorized at the 5 and 95% levels for the full sample following Lee et al., 2021.

Panel A: Summary Statistics

	Count	Mean	P25	P50	P75	Std	T-Stat
Analyst	360	0.0263	-0.1350	0.0573	0.2800	0.4824	0.4083
ML	360	-0.0341	-0.1743	-0.0238	0.1069	0.4244	-0.5288
CER	360	0.0024	-0.1428	0.0329	0.2120	0.4283	0.0399

Panel B: SVAR across Size Quintiles

	Size Q1	Size Q2	Size Q3	Size Q4	Size Q5
AF	0.2669***	0.0860	0.0314	-0.0016	0.0014
	(3.54)	(1.06)	(0.52)	(-0.03)	(0.03)
ML	0.0142	-0.0153	-0.0342	-0.0305	-0.0056
	(0.17)	(-0.16)	(-0.49)	(-0.55)	(-0.12)
ML-AF	-0.2528***	-0.1013**	-0.0655**	-0.0289	-0.0070
	(-5.25)	(-2.12)	(-2.06)	(-1.35)	(-0.50)

Table 2.10: AVA Analysis

This table shows univariate regressions analysis of AVA by size quintile and horizon. We use three measures for complexity of financial statements: the LN(Net File Size), LN(Complexity), and the Bog Index; two measures for accessibility of financial statements: an XBRL indicator that is set to one for the period following a firm's first EDGAR filing with XBRL tags, and an EDGAR indicator that is set to one for the period following a firm's inaugural 10K/Q filing through EDGAR. LN(Net File Size), LN(Complexity), and the Bog Index are all standardized by their respective standard deviations. AVA is shown in percentage points. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The regressions include month fixed effects. Standard Errors are clustered by firm and the fiscal year end. LN(Net File Size), LN(Complexity), the Bog Index are available from February 1996 to December 2020; XBRL and EDGAR indicators are available from June 1990 to December 2020.

	FQ						FY2					
	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q5	(6) Full	(7) Q1	(8) Q2	(9) Q3	(10) Q4	(11) Q5	(12) Full
LN(Net File Size)	0.175***	0.216***	0.185***	0.201***	0.100***	0.071***	0.101	0.133***	0.164***	0.069***	0.052***	0.064***
	(2.9)	(3.5)	(4.2)	(5.6)	(4.5)	(3.4)	(1.5)	(3.7)	(5.8)	(3.3)	(3.5)	(3.4)
LN(Complexity)	0.089***	0.051***	0.054***	0.056***	-0.003	0.011	0.031	0.024**	0.022**	0.008	-0.004	0.002
	(2.6)	(2.7)	(3.1)	(2.9)	(-0.3)	(1.1)	(1.4)	(2.5)	(2.5)	(1.2)	(-0.5)	(0.4)
Bog Index	-0.163	-0.123*	-0.121***	-0.070	-0.051*	-0.108***	0.124**	0.130***	0.130***	0.035**	0.024	0.083***
	(-1.4)	(-1.7)	(-2.8)	(-1.5)	(-1.7)	(-2.6)	(2.3)	(3.3)	(5.0)	(2.4)	(1.4)	(4.3)
XBRL Implemented	-1.625	0.500	0.180	-0.284***	-0.489**	-0.615***	-0.573**	0.131	-0.077	-0.072	0.000	-0.144**
	(-1.0)	(0.8)	(1.1)	(-2.8)	(-2.5)	(-5.9)	(-2.3)	(0.8)	(-1.1)	(-0.7)	(0.0)	(-2.2)
EDGAR Implemented	0.638***	0.563***	0.473***	0.485***	0.172	0.305***	0.013	0.045	0.013	-0.043	-0.044	-0.013
	(3.3)	(2.8)	(3.4)	(4.6)	(1.6)	(3.9)	(0.1)	(0.9)	(0.5)	(-1.0)	(-0.3)	(-0.4)

2.12 Appendix

Table A1: WRDS Financial Ratio Variables

This table provides the definitions of WRDS Financial Ratio Variables. Following van Binsbergen et al., 2022, we exclude Forward P/E to 1-year Growth (PEG) ratio, Forward P/E to Long-term Growth (PEG) ratio, Price/Operating Earnings (Basic, Excl. Extraordinary Income), and Price/Operating Earnings (Diluted, Excl. Extraordinary Income) from the WRDS Financial Suite Ratios due to the large number of missing observations.

Acronym	Definition	Acronym	Definition
accrual	Accruals/Average Assets	int_totdebt	Interest/Average Total Debt
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 Cash Flow/Current Liabilities
capei	Shiller's 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/Total Equity	profit_lct	Profit Before Depreciation/Current Liabilities
debt_assets	Total Debt/Total Assets	ps	Price/Sales
debt_at	Total Debt/Total Assets	ptb	Price/Book
debt_capital	Total Debt/Total Capital	ptpm	Pre-Tax Profit margin
debt_ebitda	Total Debt/EBITDA	quick_ratio	Quick Ratio
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
intcov	After-tax Interest Coverage	totdebt_invcap	Total Debt/Invested Capital
intcov_ratio	Interest Coverage Ratio		

Table A2: Other Variables

This table provides the definitions of the other variables used in generating our ML predictions that are not included in the WRDS Financial Ratio Suite. EPS and ErrAF are target variables, while all other variables are additional predictors.

Acronym	Definition
EPS (FY2/FQ)	Realized Earnings per Share
ErrAF (FY2/FQ)	Realized EPS-Analysts' forecast as of current month
medest2	Analysts' consensus forecast for FY2 horizon
medestqtr	Analysts' consensus forecast for FQ horizon
ibes_earnings_ann	Most recently realized annual earnings as of current month
ibes_earnings_qtr	Most recently realized quarterly earnings as of current month
last_F2ana_fe_med	Most recently realized FY2 horizon analysts' forecast error as of current month
last_Fqtrana_fe_med	Most recently realized FQ horizon analysts' forecast error as of current month
rev_FY2_3m	Revision of analysts' FY2 horizon forecast between current month and 3 months prior
rev_FYqtr_3m	Revision of analysts' FQ horizon forecast between current month and 3 months prior
dist2	Distance between FY2 fiscal period end and current month
distqtr	Distance between FQ fiscal period end and current month
ret	Stock Return
prc	Stock Price
size	LN(Market Capitalization)
mom6m	6 month momentum
indmom	Industry weighted 6 month momentum

Table A3: Sample Construction

This table describes how we arrive at our final sample and shows the effect of each data filter. We identify abnormal forecast period end dates in I/B/E/S (fpedats) following Bordalo et al., 2019, who provide supplementary information and replication codes for the procedure in their online appendix.

Panel A: FQ

Data Filter	Firm-Months
CRSP US common stocks merged with Compustat monthly observations Jan. 1985-Dec. 2020	**2,145,510**
Less: missing analysts' forecasts for all forecast horizons (FY1, FY2, FQ1, FQ2, FQ3)	-669,267
Less: missing FQ analysts' forecasts	-114,483
Less: missing most recently realized quarterly earnings	-16,525
Less: missing stock price, return, market capitalization, the two momentum variables, and price-to-sales	-38,045
Less: abnormal forecast period end	-37,732
Less: announcement month less than or equal to current month	-12,018
Final Dataset	**1,257,440**

Panel B: FY2

Data Filter	Firm-Months
CRSP US common stocks merged with Compustat monthly observations Jan. 1983-Dec. 2020	**2,261,904**
Less: missing analysts' forecasts for all forecast horizons (FY1, FY2, FQ1, FQ2, FQ3)	-727,524
Less: missing FY2 analysts' forecasts	-77,218
Less: missing most recently realized annual earnings	-69,629
Less: missing stock price, return, market capitalization, the two momentum variables, and price-to-sales	-21,069
Less: abnormal forecast period end	-17,757
Less: announcement month less than or equal to current month	-2
Final Dataset	**1,348,705**

Table A4: Variables used in ML Superiority Analysis

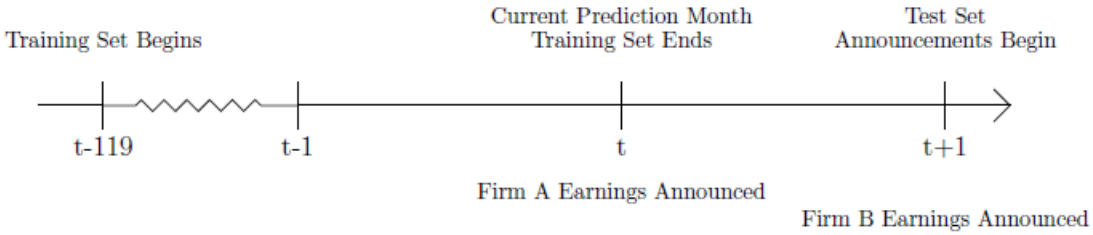
This table provides the definition of the variables used in the analysis of the cross-sectional variations in ML Superiority.

Variable	Definition	Category	Citation
Size	Ln(Market Value of Equity)	Information Uncertainty	Das et al., 1998; Frankel et al., 2006; Kross et al., 1990; Lehavy et al., 2011; Lys and Soo, 1995
Idiosyncratic Volatility	Standard deviation of residuals from CAPM regressions using the past year of daily data.		
Count of Business Segments	Count of the firms' business segments*	Firm Complexity	Amir et al., 2003; Frankel et al., 2006; F. Gu and Wang, 2005; Lehavy et al., 2011
R&D	Research and Development Expense scaled by market value*		
Bid-Ask Spread	Effective bid ask spread based on Corwin-Schulz scaled by stock price.+	Price Informativeness	Kerr et al., 2020
Institutional Ownership	Percent shares held by institutional owners**	Analysts' Incentives	Bradshaw et al., 2016; Frankel et al., 2006; Lehavy et al., 2011; Ljungqvist et al., 2007
Net external financing	Sale of common stock (sstk) minus dividends (dv) minus purchase of common stock (prstkc) plus long-term debt issuance (dltis) minus long-term debt reductions (dltr). Scaled by total assets (at).*+		
Accrual Quality	Estimate a regression for each year and industry of total current accruals on the current value and one year lag and lead of cash flow from operations, change in revenues, and gross value of PPE. Save the regression residuals and replace with missing if there are not at least 20 observations per year and industry. Calculate accrual quality (AQ) as the standard deviation of residuals over 4 years. If more than one observation is missing set AQ to missing.*+	Earnings Management	Dechow and Dichev, 2002; Francis et al., 2004; Lobo et al., 2012

*Data as of the most recently realized fiscal year end.
**Data as of the most recently realized fiscal quarter end.
+Data obtained from website associated with A. Y. Chen and Zimmermann, 2022.

Figure A1: Forecast Timeline

This figure illustrates the timeline in our analysis. As of month t, the training set consists of firm-month observations used to predict earnings that are announced between months *t-119* and *t*. The test set consists of firm-month observations in month t for predicting earnings that are announced after month t. For example, observations corresponding to Firm A's earnings announced at month t would be part of the training set, whereas observations corresponding to Firm B' earnings that are announced in month $t + 1$ would be part of the test set.



Chapter 3

Interest Rate Forecasting: It's Simpler than You Think

I show that existing interest rate forecasting techniques perform poorly compared to a simple forecast of zero change. In light of this, I propose a new interest rate forecast which focuses on removing the maturity risk premium from forward rates and demonstrate that this new approach outperforms for long horizon forecasts of interest rates. Given these findings, I decompose excess bond returns to show that the primary driver of excess bond returns for short holding periods is a bonds carry, while for long holding periods its the bonds maturity risk premium. This risk premium is plausibly invariant across both time and across the maturities of forward rates. Finally, I show that evidence against the spanning hypothesis does not hold when using the strongest yield-based benchmarks.

3.1 Introduction

Despite decades of progress, the prior literature largely disagrees about the fundamental determinants of the interest rate term structure and the best ways to form expectations about its future. This disagreement largely stems a difference in forecasting methodologies, and two key discrepancies which cause results to drastically change from one paper to the next. The first, is that data revisions, and look-ahead bias inflate the perceived ability of statistical models to predict future interest rates. The second is that researchers use different benchmarks, which makes it difficult to compare across studies. Because of these discrepancies, it is unclear which forecasting methodologies should be used by both researchers and market participants when forming expectations about future interest rates.

In this paper, I shed light on this issue by running a comprehensive analysis of the forecasting techniques proposed by the prior literature. In doing so, I demonstrate that the very simple random walk forecast outperforms other statistical models in almost all circumstances. In light of this finding, I then propose a new interest rate forecasting methodology which focuses on removing the risk premium from forward rates under the assumption that the risk premium is invariant in the cross section. Finally, I discuss the theoretical implications of my findings in the context of the spanning hypothesis, and the determinants of bond risk premia.

Interest rate forecasting is of first order importance to investors, policy makers, and researchers alike. Through understanding what portion of future interest rates are predictable and why, researchers can get a glimpse into the determinants of the term structure of interest rates. Through understanding the path of future interest rates, investors can more accurately price investment opportunities and manage risk.

And finally, through understanding the fundamental drivers of interest rates, economic policy makers can be more informed when making decisions.

In this study, I am the first to comprehensively investigate interest rate forecasting techniques in combination with the use of real-time data, and while benchmarking models to the strictest benchmark, the random walk forecast. In doing so, I discover than the random walk, which assumes that interest rates will remain constant, is a far more powerful forecast than even the most advanced statistical forecasts for the large majority of maturities and forecast distances. In fact, when decomposing yield forecasts into forward rate forecasts, I discover that the most advanced statistical model proposed by the prior literature, which uses machine learning in combination with yield, and macroeconomic data (Bianchi, Büchner, and Tamoni, 2020), only statistically outperforms the random walk according to the Deibold Mariano test-statistic for maturities of one month or less, and only at a forecast distance up to twelve months. This finding demonstrates that the the most powerful interest rate forecast in most circumstances is simply assuming that rates will stay the same, with the only exception being that machine learning can offer an advantage for extremely small maturity rate forecasts with a forecast horizon of one year or less.

In light of this evidence, I propose two new interest rate forecasting techniques, the first of which is based on the assumption that the maturity risk premium is invariant across both time and across maturities. This invariant premium (IP) assumption allows me to proxy for the risk premium of all forward rates using the historical mean premium for the shortest forward rate. This forecast is strict however in that it does not allow the maturity risk premium to vary across time. Relaxing this assumption allows me to utilize machine learning to generate a time varying forecast of the risk premium. This cross section invariant premium (CSIP) assumption again allows me to proxy for the risk premium of all forward rates utilizing the forecast of the premium for the shortest forward rate. Both of these models take advantage

of the fact that statistical models outperform the random walk the most for the shortest maturity forward rate. They also differ from the weak expectations hypothesis (WEH) forecast in that they assume the risk premium is constant across forward rates.

I find that for forecast horizons of twenty four months or greater, the invariant premium forecast generates a stronger forecast than both the random walk and the CSIP forecast for nearly all forward rate maturities according to the out of sample R-Squared. When compiling the forward rate forecasts into yield forecasts, the IP forecast outperforms the random walk for all maturities according to the out of sample R-Squared, and statistically outperforms for maturities up to 24 months and 36 months at the two year ahead and three year ahead forecast horizons respectively. This outperformance is much stronger than that of models proposed by the prior literature, and demonstrates that the random walk forecast can be outperformed for long forecast horizons. The strength of the forecast suggests that the maturity risk premium for long forecast horizons is invariant across both time and maturities.

For forecast distances of twelve months or less, I find that the random walk forecast remains a tough benchmark to beat. The CSIP forecast statistically outperforms the random walk only when forecasting forward rates with maturities of at most three months, with the IP forecast performing even worse. This confirms my finding that the random walk is the strongest for shorter forecast distances, except when forecasting extremely low maturity rates.

I next discuss the theoretical implications of my findings in two key contexts. First, I discuss the determinants of bond risk premia in the context of my findings. Following the methodology of Koijen et al., 2013, I decompose excess bond returns into three components: carry, expected price appreciation, and unexpected price shocks. I find that a bond's carry explains the largest forecastable portion of excess bond returns for short holding periods, with the total percentage of excess returns explained being near

83% for one month holding periods. For longer holding periods however, carry does not appear to explain excess bond returns. The expected price appreciation component however shows the opposite trend, explaining the largest portion of excess bond returns for long holding periods explaining as much as 22% of excess bond returns at a five year holding period. This shows that in the short run, excess bond returns are primarily related to the current state of yields, while in the long term, they are primarily related to a maturity premium which is plausibly invariant across both time and the maturities of forward rates.

Finally, I investigate how my findings fit into the long literature surrounding the spanning hypothesis, which theorizes that all available information about future interest rates is reflected by the yield curve. Because of this, non-yield curve related data shouldn't add incremental forecasting power compared to forecasts which use only yield based information. In contrast to the prior literature, I find that the best models in nearly all circumstances are based solely on yield information. Even in the limited set of circumstances where the machine learning based CSIP forecast outperforms, I find that the incremental forecasting power generated by adding macro-economic predictors is insignificant. These findings suggest that prior rejections of the spanning hypothesis do not hold when using the strongest yield based forecasting methods.

I make four key contributions to the literature surrounding the term structure of interest rates. First, I contribute to the growing literature on interest rate forecasts by comprehensively analyzing the available interest rate forecasting techniques, and demonstrate that in most circumstances the random walk forecast outperforms. Second, I further contribute to the literature on interest rate forecasting by proposing two new forecasting techniques based on removing the maturity risk premium from forward rates. I show that these forecasts outperform both the models of the previous literature, and the random walk for forecasts of rates twenty for or more months into the future, but that the random walk remains the strongest forecast

for shorter term horizons in most circumstances. Third, I contribute to the literature on the determinants of bond risk premia by showing that bond premia are primarily defined by the state of current yields for short holding periods, and defined by a possibly time and forward rate maturity invariant maturity risk premium for long holding periods. Finally, I contribute to the literature on the spanning hypothesis by showing that evidence against the spanning hypothesis does not hold when using the strongest yield based forecasts, and when using real-time macro-economic data.

3.2 Data and Prior Literature

Despite decades of progress, the prior literature largely disagrees about the fundamental determinants of the interest rate term structure and the best ways to form expectations about its future. This disagreement largely stems from two key discrepancies which cause results to drastically change from one paper to the next. The first, is that data revisions, and look-ahead bias inflate the perceived ability of statistical models to predict future interest rates. The second is that researchers use different benchmarks, which makes it difficult to compare across studies. In this section, I outline and investigate the performance of commonly used benchmark and statistical models from the prior literature. In doing so, I demonstrate that much of the statistical accuracy found in previous studies does not hold up when using real-time data, and when using the strictest benchmark: the Random Walk.

3.2.1 Data

My first goal is to determine which empirical models perform the best in forecasting interest rates out of sample. I do so by utilizing a combination of yield and real time macroeconomic data, both of which

are commonly used by the prior literature in creating interest rate forecasts. Following the methodology of Bianchi, Büchner, and Tamoni, 2020, I get zero coupon yield curve data from Liu and Wu, 2021, which i use to construct forward rates following Equation 3.3, and principal components using the first 10 years of maturities. I use the difference between the SEH and RW forecast, alongside the first five principal components of yields as predictors in all models.

I investigate whether macroeconomic information adds incremental forecasting power following the methodology of Bianchi, Büchner, and Tamoni, 2020, who use the macroeconomic variables available from McCracken and Ng, 2016. However, the version of their dataset available from Michael McKracken's website contains only final revised versions of macroeconomic variables, which have been shown to over-inflate rate forecasting accuracy through introducing look ahead bias (Wan et al., 2022, D. Huang et al., 2023).[1] Although McKracken's website does contain vintage versions of the dataset, these vintages only begin in 1999, and have several variables that change throughout the sample period, or are removed entirely. Finally, these vintages are aligned such that the macroeconomic variables correspond to their data date, and not their release date, so it is unclear what information would have been available for use in forecasting at different points in time.

To correct for these issues, I generate real-time versions of each of the variables in the McKracken database using the fredapi python package, which allows for the downloading of historical versions of macroeconomic variables which are paired with release and revision dates. In doing so, i ensure that all variable values are known in a given month to prevent look ahead bias, however I also find in untabulated analyses that my results are robust to using only first release data. In total this data consists of 127 Macroeconomic variables which contain information related to output and income, the labor market,

[1] See https://research.stlouisfed.org/econ/mccracken/fred-databases/ for final revised data.

consumption, orders and inventories, money and credit, interest and exchange rates, prices and the stock market.

The sample begins in August 1971 and ends in December 2023. Although both the macroeconomic and yield data are available on a daily basis, I use only the last observation in each month in training my models and generating out of sample forecasts in order to increase processing speed. I do find however in untabulated results that out-of-sample forecast accuracy is not sensitive to the use of daily vs monthly data. I generate forecasts out of sample following the methodology of Bianchi, Büchner, and Tamoni, 2020, and begin out of sample forecasts starting in January 1990 to allow for enough data to train the machine learning models. Following Bianchi, Büchner, Hoogteijling, and Tamoni, 2020, I make sure to leave a gap between my training and testing dataset equal to the forecast distance minus one ($\Delta t - 1$) in order to prevent look ahead bias. Models requiring hyper-parameter training or validation data use a five fold temporal separation of data as discussed in Appendix Section 3.8.1.

3.2.2 Benchmark Models

I begin with the notation of J. Y. Campbell, 2017, but later expand upon it to allow for varying forecasting distances. Consider an n period zero coupon bond which pays $1 at maturity. The price of this bond at time t is then:

$$P_{n,t} = \frac{1}{(1 + Y_{n,t})^n} \tag{3.1}$$

where $Y_{n,t}$ is the yield on the bond at time t. If the bond is held for one period, the corresponding return can then be written as:

$$1 + R_{n,t+1} = \frac{P_{n-1,t+1}}{P_{n,t}} = \frac{(1 + Y_{n,t})^n}{(1 + Y_{n-1,t+1})^{n-1}} \tag{3.2}$$

It is common to work with log return instead of raw returns for ease of notation. The corresponding one period log return can be written as:

$$r_{n,t+1} = p_{n-1,t+1} - p_{n,t} = (n)y_{n,t} - (n-1)y_{n-1,t+1}$$

Log forward rates are the difference in log prices of two bonds with maturities n and $n+1$, and can be calculated using only data at time t:

$$f_{n,t} = p_{n,t} - p_{n+1,t} = (n+1)y_{n+1,t} - (n)y_{n,t} \tag{3.3}$$

The log yield of a bond with n periods left to maturity can be decomposed into the sum of the log forward rates with lower maturities:

$$y_{n,t} = \frac{1}{n}\sum_{i=0}^{n-1} f_{i,t} \tag{3.4}$$

I now expand upon the notation of J. Y. Campbell, 2017 in order to allow me to distinguish between forecasts of varying distances in to the future. I denote the forecast distance Δt as the number of periods between the current period and the period where the forecast is realized. The simplest yield forecast assumes that forward rates follow a random walk, meaning they stay constant across time by maturity. This implies that the best forecast of a forward rate Δt periods into the future is the current forward rate with the same maturity:

$$E_t^{RW} f_{n,t+\Delta t} = f_{n,t} \tag{3.5}$$

This "Random Walk" forecast is only used as a benchmark by a few researchers who find that there are few instances where statistical models can generate stronger forecasting accuracy. Altavilla et al., 2017 find that their statistical model, which anchors yield curve forecasts to those of professional forecasters, only consistently outperforms the random walk at statistically significant levels for forecasts of yields with maturities of 12 months or less, and up to six months into the future. Bauer and Rudebusch, 2020 demonstrate that their dynamic term structure model can incorporate trends in both real interest rates and inflation to generate a forecast of the ten year yield which outperforms the random walk forecast. However, they find that this result is only persistent across sub-samples for forecasts two or more years into the future. Overall, the literature scarcely uses the random walk benchmark, and papers that do use it have demonstrated that it is very difficult to beat. This may be due to the fact that interest rates are very persistent, especially outside of quantitative policy cycles.

A second commonly used benchmark stems from the strong form of the expectations hypothesis, which theorizes instead that as future periods approach, forward rates walk down the yield curve with each incremental period. This implies that the best forecast of a future forward rate $f_{n,t+\Delta t}$ is the current forward rate with a maturity $n + \Delta t$:

$$E_t^{SEH} f_{n,t+\Delta t} = f_{n+\Delta t,t} \tag{3.6}$$

The Strong Expectations Hypothesis (SEH) forecast was most commonly used in the classical literature. Fama, 1976 shows that forecasts of short term rates based on forward rates are just as accurate as ones

created using only current and past spot rate data, and that markets react appropriately to rate conditions when determining forward rate prices. Fama and Bliss, 1987 further extend this analysis by showing that forward rates predict future spot rates for maturities up to 5 Years, and for some maturities up to 4 years in advance. However these studies do not investigate out of sample performance. Gürkaynak and Wright, 2012 investigates whether the SEH holds up in an empirical setting, and find several anomalies that cannot be explained by the SEH model. One issue with the strong version of the expectations hypothesis is that it assumes market participants are risk neutral across different maturities. By relaxing that restriction Cox et al., 1985 show in a theoretical setting that risk averse investors can cause forward rates to deviate from true market expectations. This happens in part due to the liquidity preferences of investors which cause forward rates to be higher than market expectations. Several models relax this assumption by allowing investors to charge a maturity risk premium $\phi_{n,\Delta t,t}$ in excess of their expectations of future rates:

$$f_{n+\Delta t,t} = \phi_{n,\Delta t,t} + E_t f_{n,t+\Delta t} \tag{3.7}$$

So, by subtracting $\phi_{n,\Delta t}$ from the forward rate, you can back out the true market expectation of future rates $E_t y_{1,t+n}$. The simplest estimate of this premium is the weak expectations hypothesis, which assumes that the maturity risk premium does not vary across time. Therefore, under the WEH, the best forecast of future rates can be derived by removing this time invariant premium, which is estimated using the historical mean difference between the SEH forecast, and realized rates:

$$E_t^{WEH} f_{n,t+\Delta t} = f_{n+\Delta t,t} - \sum_{i=t_0}^{t-\Delta t} f_{n+\Delta t,i} - f_{n,i+\Delta t} \tag{3.8}$$

This weak expectations hypothesis (WEH) forecast is very commonly used as a benchmark, and several papers have shown that statistical models can outperform the benchmark's predictions across the yield curve (Bianchi, Büchner, Hoogteijling, and Tamoni, 2020, Wan et al., 2022, Ghysels et al., 2017). Figure 3.1 demonstrates the relationship between forward rates and yields, and gives an example comparing how the random walk and strong expectations hypothesis forecast future forward rates.

One issue with the weak expectations hypothesis is that it assumes the risk premium does not vary across time. Due to this limitation, researchers have developed several alternative term structure forecasting methods. The simplest statistical models assume an affine term structure, which means that all rates are a linear function of certain factors related to future yields. Within this strand of models, some papers use only yields as predictors, while other papers include economic information and survey data as well. The affine model most commonly used as a benchmark uses the principal components of the yield curve to forecast future rates:

$$E_t^{APC} f_{n,t+\Delta t} = \beta_0 + \beta^T X_t + u_t \tag{3.9}$$

where X_t is a vector of principal components derived from the yield curve. This affine principal component (APC) regression typically includes either three or five principal components, and several papers have shown that other statistical models can outperform this simple framework across many forecast distances and maturities (Cieslak and Povala, 2015, D. Huang et al., 2023, Bauer and Hamilton, 2017).

Several papers extend the affine model to include macro-economic information (Ang and Piazzesi, 2003, Moench, 2008, and Bauer and Rudebusch, 2020) or survey data (Altavilla et al., 2017) alongside using yield data as a predictor of future yields. In doing so, they seek to test the spanning hypothesis

empirically, which states that all information related to future yields is reflected by the yield curve. So, if adding macro-economic variables, or other non-yield curve data improves forecasting accuracy beyond that which can be created using only yield data, then the incremental improvement would suggest that some information is not reflected by the yield curve. Bauer and Hamilton, 2017 investigate several of these papers using updated samples and find that much of the prior literature's evidence against the spanning hypothesis does not persist into the updated sample. More recently, Bianchi, Büchner, and Tamoni, 2020 find that their machine learning model's forecasts improve when adding macroeconomic data, which they provides evidence against the spanning hypothesis.

Finally, Duffee, 2002 investigates whether forecasting accuracy can be improved by relaxing the linear assumption behind affine models. He finds that his "essentially affine" model outperforms the standard class of affine models through the relaxation of this assumption, suggesting that there may be strong non-linearities between predictor variables and future yields. More recently, Bianchi, Büchner, and Tamoni, 2020 investigate whether machine learning models can forecast excess bond returns, and find that non-linear models like Neural Networks perform the best in out of sample tests.

3.2.3 Performance Metrics

I investigate the performance of the model forecasts using the out of sample R-Squared benchmarked to the Random Walk. As discussed in section 3.2.2, the random walk is difficult to beat due to the persistence of interest rates. As I will demonstrate, the random walk outperforms other models and benchmarks in almost all circumstances, making it a very strict benchmark. I define the R^2_{OOS} as:

$$R^2_{OOS} = 1 - \frac{\sum_{i=t_0}^{T}(f_{n,i+\Delta t} - \widehat{f_{n,i+\Delta t}})^2}{\sum_{i=t_0}^{T}(f_{n,i+\Delta t} - f_{n,i})^2} \tag{3.10}$$

where $\widehat{f_{n,i+\Delta t}}$ is the model forecast. I test the statistical significance of the difference in accuracy between the model forecasts and the random walk following the methodology of Diebold and Li, 2006 by reporting significance based on Diebold–Mariano test statistics. More specifically, I define the Diebold–Mariano test statistic as $DM_{1,2} = \bar{d}/\hat{\sigma}$:

$$d_t = (f_{n,t+\Delta t} - f_{n,t})^2 - (f_{n,t+\Delta t} - \widehat{f_{n,t+\Delta t}})^2 \tag{3.11}$$

where $\bar{d}$ and $\hat{\sigma}$ denote the mean and the Newey West adjusted standard errors of d_t over the test sample, setting the number of lags equal to the larger of the forecast distance Δt or twelve.

3.3 Empirical Results

I investigate the predictability of interest rates in three steps. First, I analyze the out of sample performance of models from the prior literature, and find that the random walk forecast outperforms other yield forecasts in almost all circumstances. After decomposing yield forecasts into forward rate forecasts, I discover that the random walk is only outperformed for extremely short maturity forward rates. Second, I propose two new interest rate forecasting techniques which utilizes the predictability of short term rates to remove the maturity risk premium from forward rates instead of forecasting yields directly. I show that these new forecasts outperforms the random walk in more circumstances than the models of the prior literature, but that the random walk remains the superior forecast for close horizon forecasts of most maturities. Finally, I test the theoretical implications of the new forecasting model. I show that bond risk premia are primarily related to the bonds' carry, and not related to other factors as suggested by the prior

literature. Finally, I show that prior rejections of the spanning hypothesis do not hold when benchmarked against the strongest yield-based forecasts.

3.3.1 The Random Walk

I first investigate the efficacy of yield forecasts created by the prior literature. I focus on forecasting the difference between the future log yield, and the strong expectations hypothesis forecast of that yield:

$$E_t y_{n,t+\Delta t} = G_{n,\Delta t,t}(C_t, M_t) + E_t^{SEH} y_{n,t+\Delta t} \tag{3.12}$$

where $G_{n,\Delta t,t}()$ is a function of yield based variables C_t, and macro-economic variables M_t. I forecast the target interest rate in excess of the SEH forecasted rate because doing so mirrors the methodology of studies which forecast excess bond returns instead of interest rates. This can be thought of as forecasting the interest rate in excess of the risk neutral expectation of that rate. Or equivalently, forecasting the excess return on a bond with maturity $n + \Delta t$, scaled by $-n$.[2] This methodology can also be thought of as forecasting the risk premium of a yield as shown in Equation 3.7 because the SEH forecast assumes that the risk premium is zero. So the difference between the SEH forecast and the expected future rate can be thought of as the risk premium. This step is important as it improves the forecasting accuracy of machine learning models, and also follows the prior literature while making it clear which yield is being forecasted.

[2] Specifically, you can begin with the excess return on a bond with maturity $n + \Delta t$ and rearrange the terms as follows:

$$xr_{n+\Delta t,t+\Delta t} = -n\widehat{y_{n,t+\Delta t}} + (n+\Delta t)y_{n+\Delta t,t} - y_{\Delta t,t} \tag{3.13}$$

$$xr_{n+\Delta t,t+\Delta t} = -n[\widehat{y_{n,t+\Delta t}} - \frac{1}{n}((n+\Delta t)y_{n+\Delta t,t} - y_{\Delta t,t})] \tag{3.14}$$

$$xr_{n+\Delta t,t+\Delta t} = -n[\widehat{y_{n,t+\Delta t}} - E_t^{SEH} y_{n,t+\Delta t}] \tag{3.15}$$

$$xr_{n+\Delta t,t+\Delta t} = -n(G_{n,\Delta t,t}(C_t, M_t)) \tag{3.16}$$

I first focus on a forecast distance of one year, which matches that used by Bianchi, Büchner, and Tamoni, 2020 allowing for direct comparison, and which is also the most commonly used forecast distance.

In Table 3.1, I report the out of sample R-Squared values for the three most commonly used benchmark models, which include the strong expectations hypothesis forecast, the weak expectations hypothesis forecast, and the five component affine principal component forecast alongside three machine learning algorithms which follow the forecasting techniques of Bianchi, Büchner, and Tamoni, 2020. These include elastic net, random forest, and gradient boosted regression trees. Because the out of sample R-Squared is benchmarked to the random walk, it can be interpreted at the percentage of squared rate changes which can be forecasted, with negative numbers indicating zero additional predictability beyond that of the random walk.

Using the techniques from the prior literature, only the forecasts of the one and three month rate statistically outperform the random walk. On top of that, all models fail to produce a positive R-Squared for forecasts of yields with maturities greater than 15 months, demonstrating how powerful the random walk benchmark truly is. This greatly contrasts with the findings of the prior literature, which has found that excess bond returns are predictable for many maturities due to the use of non-real-time data, or due to using weak benchmarks when evaluating forecast performance. In fact, Bianchi, Büchner, Hoogteijling, and Tamoni, 2020 find that they generate the largest out of sample R-Squared values for longer maturities, when in fact they only find stronger performance because their WEH benchmark performs much worse than the random walk benchmark for longer maturities. This demonstrates the need for a comprehensive analysis of interest rate forecasting techniques to uncover the truth behind the determinants of the term structure of interest rates.

Because the random walk outperforms or is statistically indistinguishable from statistical models for nearly all maturities, it is unclear whether changes in most interest rates are forecastable at all. In fact, log yields are simply the sum of log forward rates as defined in Equation 3.4, so although some forecasts generate positive R-Squared values over the random walk, the accuracy may only be driven by the component of yields comprised of very short term forward rates. To investigate whether this is true, I decompose the yield forecasts into forward rate forecasts following Equation 3.4, and report the out of sample R-Squared of the GBRT forecast benchmarked against the random walk in Table 3.2. I expand the number of forecast distances to include forecasts 1, 6, 12, 24, 36 and 60 months into the future to more comprehensively assess whether the models of the prior literature achieve forecasting accuracy beyond that of the random walk. I focus on the GBRT forecast as it is the forecast which performs the best across multiple forecasting horizons out of the models considered from the prior literate. However my findings are robust to using any of the models presented in Table 3.4.

Across all forecast distances, the GBRT model fails to generate a statistical forecast that is more accurate than the random walk for forward rates with a maturity of three months or greater. In fact, the model fails to generate a positive R-Squared for maturities greater than 12 months for all forecast distances, and only generates a positive R-Squared consistently for maturities of four months or less. This demonstrates that forecasts for most of the forward rates which comprise the yield curve cannot be improved beyond that of the simple random walk using even the most advanced techniques in the prior literature.

3.3.2 The Invariant Premium Forecast

Given the shortcomings of available interest rate forecasting techniques when compared to the random walk I now propose two new simple forecasting methodologies which are based on the assumption that

the maturity risk premium is constant across maturities. This contrasts with the predictions of the weak expectations hypothesis in Equation 3.3 in that it assumes that the maturity risk premium ϕ is constant across maturities, but may plausibly vary across time. In doing so, forward rates can be forecasted by instead forecasting just one maturity risk premium per future period. To create this forecast, I take advantage of the fact that the maturity risk premium is predictable for short term forward rates as seen in Table 3.2. Specifically, the zero month forward rate ($f_{0,t} = y_{1,t}$) has the largest R-Squared of all maturities for all forecast distances regardless of the forecasting technique used. If the maturity risk premium is in fact constant across maturities in the cross section, then this forecast should also predict the maturity risk premium for all maturities. This approach is similar to the anchoring technique of Altavilla et al., 2017 in that it links the entire term structure to the forecast of the forecast of the lowest maturity forward rate. In their model however, they link to survey expectations of the three month yield, but do not imply restrictions on the forecasts of yields of longer maturities, while I require that the risk premium of all maturities be equal to that of the one month rate.

I generate two versions of this invariant premium forecast, the first of which makes a more strict assumption by requiring that the risk premium be constant across both time and across maturities. Doing so allows for using the historical mean of the zero month forward rate premium as a proxy for the premium of all maturities:

$$E_t^{IP} f_{n,t+\Delta t} = f_{n+\Delta t,t} - \sum_{i=t_0}^{t-\Delta t} f_{\Delta t,i} - f_{0,i+\Delta t} \tag{3.17}$$

This invariant premium (IP) forecast is similar to the weak expectations hypothesis forecast in Equation 3.8 in that it removes a simple historical average premium from the current forward rate. However, it

differs in that it uses the historical mean premium of the zero month forward rate for all maturities instead. The second forecast I consider relaxes the assumption that the risk premium is invariant across time, but keeps the assumption that it does not vary across maturities:

$$E_t^{CSIP} f_{n,t+\Delta t} = G_{0,\Delta t,t}(C_t, M_t) + f_{n+\Delta t,t} \tag{3.18}$$

where $G_{0,\Delta t,t}(C_t, M_t)$ is the machine forecast of the maturity risk premium being charged for the shortest maturity forward rate.[3] This Cross Section Invariant Premium (CSIP) Forecast is similar to the premium forecast in Equation 3.12 in that it forecasts the premium contained within forward rates, with the difference being that it forecasts forward rates directly instead of forecasting yields, and that it uses the forecast of the zero month forward rate premium as the forecast for all maturities instead. This can also be thought of as using the machine to find out what premium is being charged for the future one month risk free rate, $y_{1,t+\Delta t}$.

I report the R-Squared for the IP and CSIP forecasts in panels A and B of Table 3.3, respectively. The results show a striking increase in predictive accuracy for longer forecast distances. Using the IP forecast, the R-Squared for forecasts of distance 24 months or greater into the future become positive for maturities up to sixty months, with the thirty-six month ahead forecast being statistically stronger than the random walk for forecasts of maturities up to 23 months. For shorter forecast distances, the CSIP forecast appears to be more powerful than the IP forecast, but shows only mild improvement compared to the methods

[3] In order to give the machine the best chance of beating the random walk, I deviate here from using the GBRT model of Bianchi, Büchner, and Tamoni, 2020 in two ways. First, I use the MAE objective function to train the GBRT instead of using the MSE. Second, I expand the set of predictors fed into the machine forecast to also include the first five principal components of yields, the difference between the WEH and RW forecast, interest rate momentum, and the rolling standard deviation of daily interest rates. I do so because I argue that the random walk model dominates for forecast distances of twelve months or less except for forecasts of extremely short term rates, so utilizing the best interest rate forecast I can find gives the highest chance to prove my argument incorrect.

of the prior literature used in Table 3.2. Specifically, the CSIP forecast now is statistically more accurate than the random walk for maturities up to four months for the one month ahead forecast, three months for the two month ahead forecast, and two months ahead for the twelve month ahead forecast. These improvements demonstrate that in the short term, the random walk is very hard to beat for the large majority of maturities.

For ease of comparison with work in the prior literature, I compile the IP and CSIP forecasts into yield forecasts following Equation 3.4 in Table 3.4. The results largely mirror the trends of Table 3.3, but show that the CSIP forecast outperforms the random walk only for yields with maturities of 12 months or less for forecast distances of one month, with six and twelve month ahead forecasts performing even worse. This demonstrates that the random walk forecast is incredibly powerful for short term interest rate forecasts for all maturities greater than one year. For forecast distances of 24 months or greater however, the IP forecast generates positive R-Squared values compared to the random walk for all maturities. This implies that for forecast distances of 24 months or greater, the maturity risk premium is neither time nor cross-sectionally varying, and that a simple historical average can be used to improve forecasting accuracy beyond that of the random walk.

3.4 Theoretical Implications

Now that I have analyzed which forecasts perform the best at varying forecast distances and maturities, I proceed to discuss the theoretical implications of my findings. First, I analyze the return based implications of my findings, and show that for short holding periods the primary driver of excess bond returns can be defined by the bond's carry following Koijen et al., 2018, while for long holding periods,

the primary driver of returns is the maturity risk premium which is plausibly invariant across both time and constant across the maturities of forward rates. Second, I revisit a classical question in the interest rate forecasting literature: whether or not the yield curve reflects all available information about future yields. Contrary to the prior literature, I find that macro-economic information fails to generate incremental improvements in forecasting power when using the strongest yield based forecasts as a benchmark.

3.4.1 Determinants of Bond Risk Premia

The strength of the random walk forecast suggests that current yields may be the primary factor in determining the size of bond risk premia. This premise similar to the findings of Koijen et al., 2013, who define the "carry" of an asset as the expected return from holding that asset given prices stay the same. Koijen et al., 2013 decompose the return of an asset into three components:

$$\text{Return} = \text{Carry} + E(\text{Price Appreciation}) + \text{Unexpected Price Shock}$$

where the carry is the return from holding an asset given prices stay the same, E(Price Appreciation) is the expected return from changes in the price of the asset, and the rest of the return is an unexpected price shock. For a bond, the carry of a bond with maturity n held for Δt periods is:

$$C_{n,t+\Delta t} = \sum_{i=n-\Delta t}^{n-1} f_{i,t} - \sum_{i=1}^{\Delta t-1} f_{i,t}$$

The carry of a bond can be thought of as the excess return on a bond with maturity n given that yields stay the same. Because the random walk dominates most forecasts, it suggests that the primary driver of excess bond returns should be the bond's carry. If the random walk is the best forecasting model, then the

expected price appreciation component of bond returns should be non-existent. However, if statistical models can outperform the random walk, as suggested by the long horizon IP forecast, then the expected price appreciation component may also exist. To investigate the components of excess bond returns, I calculate the percentage of buy and hold excess bond returns explained by each return component. Specifically, I calculate the percentage of excess returns explained for each component as:

$$\%\ \text{Explained}_{Carry} = 1 - \frac{|C_{n,t+\Delta t} - xr_{n,t+\Delta t}|}{|\bar{xr_n}|} \tag{3.19}$$

$$\%\ \text{Explained}_{E(Price)} = 1 - \frac{|\widehat{xr_{n,t+\Delta t}}^{(Model)} - xr_{n,t+\Delta t}|}{|\bar{xr_n}|} - \%\ \text{Explained}_{Carry} \tag{3.20}$$

$$\%\ \text{Explained}_{Unexpected} = \frac{|\widehat{xr_{n,t+\Delta t}}^{(Model)} - xr_{n,t+\Delta t}|}{|\bar{xr_n}|} \tag{3.21}$$

where $\widehat{xr_{n,t+\Delta t}}^{(Model)}$ is the model forecast of the excess return, and $|\bar{xr_n}|$ is the time series average of the absolute value of the excess return on the bond with maturity n. I compute the % explained by each component for each maturity and forecast distance, and then report the average % explained by each component by forecast distance in Table 3.5. Panel A reports results using the IP forecast to generate the forecasted excess return, and Panel B reports results using the CSIP forecast. Overall, the results show that a bonds carry explains the largest portion of its excess returns for shorter holding periods, with the carry explaining 83.88% of excess returns for a holding period of one month. This value remains statistically significant for holding periods up to twelve months, and becomes negative for holding periods of thirty-six months or greater. This suggests that over long time-frames, the carry of bond does not predict excess returns. When using the IP forecast to calculate the expected price appreciation, the % explained is negative for forecasts within twelve months, and positive and statistically significant for forecasts of twenty-four

months or greater. The percentage explained by expected price appreciation is generally increasing in the size of the holding period, with the expected price appreciation explaining 22.37% of excess returns for a holding period of 60 months. When using the CSIP forecast however, the result appears to be much weaker.

These results suggest that the primary driver of a bond's excess returns is its carry for shorter holding periods, but that the risk premium is the primary driver of excess returns for longer holding periods. This finding is consistent with my findings in section 3.3, which show that the random walk dominates in most circumstances for short forecast distances, while the IP forecast dominates for long horizon forecasts.

3.4.2 The Spanning Hypothesis

Given the outperformance of the entirely yield-based random walk and IP forecasts, it is unclear whether there are any circumstances where non-yield based information is useful when forecasting interest rates. The Spanning Hypothesis states that all available information about future interest rates is reflected by the yield curve. Because of this, the addition of non-yield curve related information to forecasts of future interest rates should not improve forecasting accuracy. As discussed in Section 3.2.2, several papers investigate this claim by running statistical forecasts with and without non-yield information to see if the additional predictors add incremental forecasting power. Several papers find that some information like macroeconomic and lagged time series data can improve forecasting accuracy beyond the yield-only forecasting models. Bauer and Hamilton, 2017 however have shown that much of this evidence does not hold when using updated samples. More recently, Bianchi, Büchner, and Tamoni, 2020 find that their machine learning model's forecasts improve when adding macroeconomic data, which provides evidence against the spanning hypothesis. However, their forecasts are based on revised data, and as shown in

Table 3.1, their forecasting methods do not beat the random walk for most maturities. In fact because the random walk and the IP forecast dominate for most maturities and forecast distances, and both are derived directly from yield based information, most of the evidence against the spanning hypothesis in the prior literature does not hold. In fact the only time where the machine learning model outperforms the yield based random walk or IP forecast is for forecasts of forward rates with maturities of three or less at a one month forecasting distance, with maturities of two or less at a distance of six months, and maturities of one month or less at a distance of twelve months. So if there is evidence against the spanning hypothesis, it would need to be concentrated in this minuscule portion of the term structure to hold merit.

In Table 3.6, I test the spanning hypothesis by benchmarking the CSIP forecast with both macroeconomic and yield predictors against the CSIP forecast using only yields. If the real-time macroeconomic data contains information not already contained in yields then the model which can use the macroeconomic information should outperform in the segment where the CSIP forecast can beat the random walk. However, for the forecast distance-maturity combinations where this happens, the CSIP forecast with macroeconomic data is statistically indistinguishable from the version without. This suggests that the improvement of the CSIP forecast over the random walk likely comes from yield based information, and not from macroeconomic signals. This contrasts with the prior literature in that it suggests that the evidence against the spanning hypothesis is insignificant, and marginal at best. The difference comes from the fact that the yield based models of the prior literature did not extract the information about future yields as completely as the random walk, IP forecast, and the yields-based CSIP forecast.

3.5 Conclusion

In this paper, I investigate the performance of interest rate forecasting techniques when correcting for the use of real-time data, and when using the most stringent benchmark models. I demonstrate that the strongest forecasts of interest rates are actually very simple. For short forecast horizons, the random walk forecast outperforms more advanced statistical models in nearly all circumstances, and for long horizon forecasts, the IP forecast outperforms. Both of these forecasts are simple to derive from current and historical yield information, which demonstrates that advanced statistical techniques may be unnecessary when investigating the term structure of interest rates.

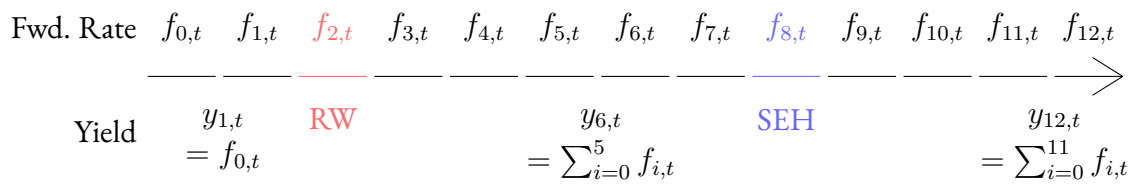
Through these findings, I investigate the determinants of bond risk premia by decomposing excess bond returns into three components. I find that a bond's carry explains the largest forecastable portion of excess bond returns for short holding periods, but that it does not predict returns for long holding periods. The expected price appreciation component however shows the opposite trends explaining the largest portion of excess bond returns for long horizon forecast distances. This shows that in the short run, excess bond returns are primarily related to the current state of yields, while in the long term, they are primarily related to a maturity premium which is plausibly invariant across both time and the maturities of forward rates. It is possible however that there are undiscovered interest rate forecasting techniques which can consistently predict time-series variations in the maturity risk premium. Whether these models exist however is a question for further research.

Finally, I find that the information contained in macroeconomic variables does not significantly improve forecasting accuracy when compared to the strongest yield based forecasting models. This provides

evidence in favor of the spanning hypothesis, and opens up the the need for further research to investigate whether there are circumstances where the spanning hypothesis does not hold.

3.6 Figures

Figure 3.1: Example Forecast Timeline: Forecasting $f_{2,t+6}$



This figure demonstrates the relationship between forward rates and yields, and gives an example comparing how the random walk and strong expectations hypothesis forecast future forward rates. When creating a forecast for the 2 month forward rate in 6 months $f_{2,t+6}$, the random walk forecast assumes that forward rates will remain constant across the yield curve, and thus forecasts the the future three month rate will equal the current one $f_{2,t+6} = f_{2,t}$. The expectations hypothesis assumes that each forward rate will remain constant in time, so as time progresses six months into the future, the new 2 month forward rate will be the previous 8 month forward rate $f_{2,t+6} = f_{8,t}$.
Random Walk:

$$E_t f_{2,t+6} = f_{2,t} \tag{3.22}$$

Expectations hypothesis:

$$E_t f_{2,t+6} = f_{8,t} \tag{3.23}$$

3.7 Tables

Table 3.1: Common Yield Forecasts vs. Random Walk

This table reports the out of sample R-Squared of common yield forecasts benchmarked against the random walk following Equation 3.10. Significance is denoted using asterisks (at the ***1%, **5%, and *10% level) based on the test of Diebold and Mariano, 1995, which tests that the forecast is more accurate than the random walk benchmark. These test-statistics are calculated using standard errors based on Newey and West, 1987 with twelve lags. These out of sample forecasts begin in January 1990, and continue until December 2023.

Maturity (Months)	SEH	WEH	APC	Elastic Net	Random Forest	GBRT
1	-6.15	23.53*	-51.85**	22.39**	25.58**	24.23**
	(-0.32)	(1.85)	(-2.06)	(2.27)	(2.12)	(2.08)
3	-7.48	17.68	-59.84**	16.50*	19.67*	19.76*
	(-0.41)	(1.45)	(-2.36)	(1.81)	(1.70)	(1.69)
6	-8.58	10.55	-66.56***	9.82	12.45	11.77
	(-0.50)	(0.88)	(-2.61)	(1.09)	(1.12)	(1.03)
9	-9.22	5.96	-67.39***	4.69	7.73	7.27
	(-0.56)	(0.51)	(-2.68)	(0.50)	(0.71)	(0.64)
12	-10.22	2.27	-65.47***	-0.01	3.72	3.62
	(-0.65)	(0.20)	(-2.69)	(-0.00)	(0.35)	(0.32)
15	-11.93	-0.92	-64.15***	-4.22	0.52	0.38
	(-0.77)	(-0.08)	(-2.70)	(-0.44)	(0.05)	(0.03)
18	-13.95	-3.55	-64.04***	-8.21	-2.15	-2.34
	(-0.92)	(-0.31)	(-2.73)	(-0.86)	(-0.21)	(-0.21)
24	-18.60	-6.97	-66.84***	-11.30	-5.52	-6.45
	(-1.22)	(-0.64)	(-2.81)	(-1.17)	(-0.56)	(-0.59)
36	-24.34	-11.08	-66.82***	-20.53**	-9.58	-10.65
	(-1.62)	(-1.06)	(-2.89)	(-2.21)	(-1.02)	(-1.08)
48	-24.25*	-13.46	-63.42***	-20.80*	-11.96	-14.72
	(-1.81)	(-1.41)	(-2.98)	(-1.87)	(-1.42)	(-1.56)
60	-25.37**	-13.23	-61.55***	-32.49**	-11.56	-13.81*
	(-1.97)	(-1.51)	(-3.03)	(-2.27)	(-1.54)	(-1.77)
84	-25.96**	-17.19**	-57.64***	-30.46**	-15.24**	-20.94**
	(-2.31)	(-2.09)	(-3.05)	(-2.21)	(-2.13)	(-2.57)
120	-27.00***	-19.76***	-48.09***	-28.44**	-16.97***	-19.75***
	(-2.69)	(-2.69)	(-3.05)	(-2.34)	(-2.73)	(-2.86)

Table 3.2: GBRT vs. Random Walk: Forward Rate Forecasts

This table reports the out of sample R-Squared of the GBRT forecast benchmarked against the random walk following Equation 3.10. Significance is denoted using asterisks (at the ***1%, **5%, and *10% level) based on the test of Diebold and Mariano, 1995, which tests that the forecast is more accurate than the random walk benchmark. These test-statistics are calculated using standard errors based on Newey and West, 1987 with the number of lags set equal to the larger of twelve, or the number of months between the forecast date and the date the rate is realized.

Maturity (Months)	Δ 1 Mo.	Δ 6 Mo.	Δ 12 Mo.	Δ 24 Mo.	Δ 36 Mo.	Δ 60 Mo.
0	19.23***	36.35***	24.23**	16.24	25.16*	7.92
1	12.31*	28.59**	19.14	12.69	22.30	1.89
2	6.39	15.56*	14.59	10.88	19.32	1.20
3	4.29	7.59	7.32	2.99	16.49	-5.79
4	-0.33	3.09	1.75	3.26	14.40	-14.34
5	-1.91	-7.11	-2.75	-0.89	8.65	-9.37
6	-3.67	-7.12	-6.73	0.67	8.80	-16.90
7	-1.61	-11.93	-7.01	1.44	8.25	-17.21
8	-9.95**	-20.73**	-11.87	-5.89	10.25	-30.56*
9	-5.04	-20.39**	-18.77	-5.70	9.35	-24.68
10	-7.25*	-14.92	-11.97	-1.42	5.72	-32.03*
11	-16.66**	-19.42*	-29.78**	-15.35	1.13	-38.87**
14	-8.90**	-27.54***	-40.39**	-25.71	-8.54	-96.67***
17	-16.73**	-39.91***	-39.57***	-30.25**	-26.25	-123.04***
23	-24.31***	-27.05***	-42.52***	-45.00***	-47.13**	-127.04***
35	-29.59***	-195.47***	-316.19***	-127.04***	-116.95***	-267.46***
47	-70.56***	-106.09***	-157.03***	-125.12***	-373.03***	-328.54***
59	-200.02***	-722.42***	-315.84***	-282.95***	-261.15***	-711.04***
83	-81.38***	-278.02***	-395.01***	-539.46***	-370.90***	-473.35***
119	-1056.91***	-1099.52***	-1176.00***	-1088.46***	-592.26***	-449.04***

Table 3.3: Forecasting Forward Rates

This table reports the out of sample R-Squared of forward rate forecasts benchmarked against the random walk following Equation 3.10. Panel A shows results for the invariant premium (IP) forecast, and Panel B shows results for the cross section invariant premium (CSIP) forecast. Significance is denoted using asterisks (at the ***1%, **5%, and *10% level) based on the test of Diebold and Mariano, 1995, which tests that the forecast is more accurate than the random walk benchmark. These test-statistics are calculated using standard errors based on Newey and West, 1987 with the number of lags set equal to the larger of twelve, or the number of months between the forecast date and the date the rate is realized.

Panel A: IP Forecast

Maturity (Months)	Δ 1 Mo.	Δ 6 Mo.	Δ 12 Mo.	Δ 24 Mo.	Δ 36 Mo.	Δ 60 Mo.
0	18.58*	31.37**	23.53*	17.85	33.24**	28.96
1	12.66	23.31*	18.72	18.15	33.86**	30.14
2	4.67	14.87	14.13	18.64	34.69**	31.66
3	-1.78	7.24	10.16	19.20	35.56**	33.21
4	-6.15	0.73	6.93	19.76	36.29**	34.46
5	-8.67	-4.84	4.37	20.27*	36.80**	35.24
6	-9.58	-9.75	2.27	20.72*	37.06**	35.57
7	-9.86	-14.21	0.45	21.13*	37.12**	35.63
8	-10.57	-18.06	-1.15	21.50*	37.07**	35.56
9	-11.95*	-20.99	-2.48	21.84*	36.92**	35.40
10	-13.59**	-22.87	-3.48	22.10*	36.69**	35.16
11	-14.80**	-23.96	-4.11	22.25**	36.33**	34.86
14	-14.58***	-25.80	-4.32	22.08*	33.90*	33.15
17	-15.87***	-24.84*	-2.96	21.05*	30.55*	30.74
23	-12.89***	-15.97	-0.60	18.49	31.31*	30.45
35	-12.27***	-32.91**	-21.78	12.09	18.23	37.91
47	-18.09***	-37.41**	-22.50	10.06	1.94	30.42
59	-13.37***	-27.05*	-18.43	2.80	1.64	7.27
83	-11.00	-42.48***	-31.83*	-20.07	-12.91	34.05
119	-17.39***	-57.11***	-47.64**	-18.11	-23.72	-19.95

Panel B: CSIP Forecast Using GBRT

Maturity (Months)	Δ 1 Mo.	Δ 6 Mo.	Δ 12 Mo.	Δ 24 Mo.	Δ 36 Mo.	Δ 60 Mo.
0	26.90***	37.20**	26.44**	20.39	29.36*	22.78
1	23.09**	30.87**	22.04*	20.44	29.77*	24.32
2	17.07**	24.03**	17.76	20.60	30.33*	26.08
3	11.94*	17.73	14.02	20.83	30.90*	27.75
4	8.31	12.32	10.95	21.09	31.31*	29.01
5	5.98	7.67	8.49	21.38	31.49*	29.72
6	4.76	3.51	6.47	21.69	31.42*	29.92
7	3.98	-0.30	4.72	22.02*	31.17*	29.79
8	2.87	-3.68	3.17	22.39*	30.82*	29.52
9	1.24	-6.31	1.87	22.78*	30.41	29.16
10	-0.61	-8.09	0.90	23.15*	29.93	28.71
11	-2.14	-9.25	0.26	23.43*	29.32	28.17
14	-3.14	-11.56	-0.18	23.56*	25.90	25.90
17	-4.14	-11.03	0.94	22.30*	21.07	23.51
23	-3.41	-3.88	2.93	18.54	19.86	23.88
35	-3.17	-15.51	-16.27	5.75	1.51	33.19
47	-8.02*	-19.17	-15.82	0.06	-16.60	27.44
59	-2.70	-8.63	-12.90	-9.19	-17.80	4.18
83	-0.27	-21.51*	-23.64	-46.20	-35.41	33.40
119	-4.56*	-34.19**	-41.26*	-49.37	-45.34	-16.01

Table 3.4: Composite Yield Forecast Accuracy

This table reports the out of sample R-Squared of yield forecasts benchmarked against the random walk following Equation 3.10. Yield forecasts are generated by compiling rate forecasts into yield forecasts following Equation 3.4. Panel A shows results for the invariant premium (IP) forecast, and Panel B shows results for the cross section invariant premium (CSIP) forecast. Significance is denoted using asterisks (at the ***1%, **5%, and *10% level) based on the test of Diebold and Mariano, 1995, which tests that the forecast is more accurate than the random walk benchmark. These test-statistics are calculated using standard errors based on Newey and West, 1987 with the number of lags set equal to the larger of twelve, or the number of months between the forecast date and the date the rate is realized.

Panel A: IP Forecast

Maturity (Months)	Δ 1 Mo.	Δ 6 Mo.	Δ 12 Mo.	Δ 24 Mo.	Δ 36 Mo.	Δ 60 Mo.
1	18.58*	31.37**	23.53*	17.85	33.24**	28.96
2	15.99	27.44*	21.14*	18.00	33.56**	29.57
3	12.37	23.34*	18.80	18.21	33.96**	30.32
4	8.41	19.32	16.61	18.45	34.39**	31.11
5	4.68	15.52	14.61	18.71	34.81**	31.87
6	1.51	12.00	12.84	18.97	35.18**	32.53
7	-1.02	8.73	11.26	19.22	35.49**	33.08
8	-3.01	5.66	9.85	19.46	35.75**	33.51
9	-4.66	2.77	8.57	19.69	35.95**	33.86
10	-6.14	0.09	7.41	19.92*	36.11**	34.13
11	-7.54	-2.36	6.35	20.13*	36.23**	34.34
12	-8.85	-4.55	5.41	20.33*	36.32**	34.49
15	-11.92	-9.81	3.11	20.77*	36.31**	34.60
18	-13.73*	-13.70	1.42	20.99*	35.92**	34.36
24	-15.61**	-17.79	-0.83	20.90*	35.05**	33.59
36	-15.29***	-21.11	-4.58	19.93	33.39*	34.34
48	-16.31***	-26.70*	-9.11	18.79	29.68	35.20
60	-17.48***	-30.51*	-12.62	16.59	26.05	31.97
84	-20.14***	-42.69**	-22.78	10.76	15.68	28.18
120	-22.59***	-53.19***	-33.81	4.21	5.09	25.37

Panel B: CSIP Forecast Using GBRT

Maturity (Months)	Δ 1 Mo.	Δ 6 Mo.	Δ 12 Mo.	Δ 24 Mo.	Δ 36 Mo.	Δ 60 Mo.
1	26.90***	37.20**	26.44**	20.39	29.36*	22.78
2	25.51***	34.16**	24.26*	20.41	29.57*	23.57
3	23.23**	30.92**	22.10*	20.47	29.84*	24.45
4	20.60**	27.71**	20.06	20.56	30.13*	25.34
5	18.05**	24.67**	18.20	20.66	30.39*	26.16
6	15.79**	21.85*	16.53	20.78	30.61*	26.84
7	13.90**	19.21*	15.05	20.91	30.76*	27.39
8	12.30*	16.73	13.72	21.06	30.86*	27.79
9	10.88*	14.38	12.52	21.22	30.91*	28.10
10	9.55*	12.16	11.42	21.39	30.92*	28.31
11	8.26	10.12	10.42	21.57	30.90*	28.46
12	6.99	8.28	9.52	21.75	30.85*	28.53
15	3.74	3.77	7.33	22.20*	30.43*	28.40
18	1.71	0.38	5.69	22.43*	29.58	27.95
24	-0.76	-3.36	3.49	22.23*	27.74	26.93
36	-1.94	-5.91	0.05	20.31	24.16	27.54
48	-2.95	-9.69	-3.72	17.92	18.40	28.48
60	-3.42	-11.87	-6.83	14.29	12.87	25.35
84	-4.67*	-19.89	-15.27	4.19	-2.08	21.58
120	-5.64**	-26.88*	-24.97	-8.21	-17.72	20.11

Table 3.5: Carry vs. Price Appreciation

This table reports the percentage of excess bond returns explained by return component and holding period following equations, 3.19, 3.20 and 3.21. I first calculate the time series average of each percentage explained by maturity and forecast distance, then average across maturities while computing test-statistics using standard errors based on Newey and West, 1987 with the number of lags set at 24 months by maturity. Significance is denoted using asterisks (at the ***1%, **5%, and *10% level).

Panel A: IP Forecast

	Δ 1 Mo.	Δ 6 Mo.	Δ 12 Mo.	Δ 18 Mo.	Δ 24 Mo.	Δ 36 Mo.	Δ 48 Mo.	Δ 60 Mo.
Carry	83.88***	55.65***	34.82***	17.41	7.81	-1.46	-1.58	-9.22
	(21.88)	(5.65)	(2.61)	(1.10)	(0.46)	(-0.08)	(-0.08)	(-0.42)
E(Price Appreciation)	-1.85***	-6.85***	-3.85***	0.84	5.46***	12.44***	15.77***	22.37***
	(-3.53)	(-3.47)	(-2.65)	(0.78)	(6.50)	(7.09)	(7.15)	(3.98)
Unexpected Price Shock	17.97***	51.20***	69.03***	81.75***	86.73***	89.02***	85.80***	86.85***
	(4.14)	(4.34)	(4.71)	(4.98)	(5.27)	(4.98)	(5.00)	(5.24)

Panel B: CSIP Forecast Using GBRT

	Δ 1 Mo.	Δ 6 Mo.	Δ 12 Mo.	Δ 18 Mo.	Δ 24 Mo.	Δ 36 Mo.	Δ 48 Mo.	Δ 60 Mo.
Carry	83.88***	55.65***	34.82***	17.41	7.81	-1.46	-1.58	-9.22
	(21.88)	(5.65)	(2.61)	(1.10)	(0.46)	(-0.08)	(-0.08)	(-0.42)
E(Price Appreciation)	-0.47***	-3.49***	-3.12**	-0.07	0.46	0.61	7.87***	16.69***
	(-2.96)	(-3.23)	(-2.42)	(-0.06)	(0.21)	(0.16)	(3.89)	(3.72)
Unexpected Price Shock	16.59***	47.83***	68.30***	82.67***	91.73***	100.85***	93.71***	92.53***
	(4.17)	(4.38)	(4.71)	(4.99)	(4.99)	(4.71)	(4.76)	(5.21)

Table 3.6: Test of Spanning Hypothesis

This table reports the out of sample R-Squared of the CSIP forecast with macroeconomic and yield predictors benchmarked against the CSIP forecast using only yield data as predictors following Equation 3.10. Significance is denoted using asterisks (at the ***1%, **5%, and *10% level) based on the test of Diebold and Mariano, 1995, which tests that the forecast is more accurate than the random walk benchmark. These test-statistics are calculated using standard errors based on Newey and West, 1987 with twelve lags.

Maturity (Months)	Δ 1 Mo.	Δ 6 Mo.	Δ 12 Mo.
0	-0.07	3.19	0.76
	(-0.13)	(1.44)	(0.23)
1	0.23	3.22	0.73
	(0.37)	(1.49)	(0.22)
2	0.58	3.26	0.79
	(0.94)	(1.57)	(0.25)
3	0.81	3.26*	0.90
	(1.37)	(1.65)	(0.29)
4	0.90	3.22*	1.04
	(1.62)	(1.71)	(0.35)
5	0.91*	3.16*	1.20
	(1.75)	(1.75)	(0.40)
6	0.86*	3.10*	1.36
	(1.80)	(1.80)	(0.47)
7	0.81*	3.08*	1.53
	(1.82)	(1.86)	(0.53)
8	0.77*	3.05*	1.70
	(1.86)	(1.92)	(0.60)
9	0.74*	3.05**	1.87
	(1.90)	(1.98)	(0.66)
10	0.72*	3.11**	2.04
	(1.94)	(2.08)	(0.72)
11	0.69**	3.20**	2.23
	(1.97)	(2.18)	(0.79)

T-Statistics in Brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: LightGBM Hyper-Parameters

This table reports the hyper-parameter options used in implementing the Gradient Boosted Decision Tree Model.

Learning Rate	1e-4
Max Depth	7
Number of Trees	4000
Early Stopping Rounds	5
Early Stopping Ensemble	5
Minimum observations in leaf	5
Boosting Type	GOSS

3.8 Appendix

3.8.1 Gradient Boosted Regression Trees

In this study I generate predictions using a customized gradient boosted regression tree model which I construct using Microsoft's LightGBM python package. Gradient boosted regression tree models are non-linear non-parametric ensemble models which combine the predictions of many decision trees. Trees are grown in an adaptive way to correct the prediction error from the previous iteration, which is known as boosting (Friedman, 2001). The weighted average of these individual tree models is the final predictor. As Friedman, 2002 shows, subsampling helps reduce the computation time and the overfitting risk in boosting. Instead of randomly selecting a fraction of the data to train the model, I use Gradient-based One-Side Sampling (GOSS) (Ke et al., 2017) to sample observations. At each iteration, GOSS keeps data instances with residual errors in the top a percentile and randomly selects b percent of the remaining instances.[4] GOSS then combines these selected data instances to grow the next tree.

Finally, I implement early stopping to determine the number of trees in the model, which helps prevent overfitting. Early stopping works by checking the validation loss of a holdout set after each tree is added to the model. If the validation loss doesn't improve for a certain number of trees in a row, then the algorithm stops adding trees. By doing this you prevent the model from being too complex. I opt to set the number of trees arbitrarily high to ensure that early stopping comes into effect for each of my models. This is a much faster alternative to tuning the number of trees in the model through cross validation, and tends to have similar (if not superior) results.

[4]The randomly selected data are amplified by the ratio of $\frac{1-a}{b}$ to minimize the influence on the data distribution.

I augment the standard gradient boosted regression tree model through a combination of early stopping and ensembling. Specifically, I average the predictions of five separate GBDT estimators, each of which use a separate 20% of the training data as the validation set for early stopping. By doing this, I make sure that all data is used in training the model parameters, while still allowing for the early stopping function of LightGBM to regulate model complexity, which combats over-fitting. Each of the 5 validation sets are separated temporally, and are non-overlapping. Specifically, the first validation set uses the earliest 20% of observations in the training set, the second validation set uses the next earliest 20% and so on. Other information about the details of the algorithm's implementation can be found in table 3.7.

Chapter 4

Beyond Benefits: Uncertainty and Sticky Information Costs [1]

Motivated by the ambiguous predictions of existing information choice theories, we propose and test the "Sticky Information Cost" (SIC) hypothesis to understand how investors acquire information in uncertain financial markets. SIC asserts that information processing costs for investors are influenced by a firm's slow-changing information environment, closely linked to its fundamental uncertainty. Using direct measures for information processing costs and the return predictability of analysts' biases as a proxy for information acquisition, we find opposite relationships between uncertainty and information acquisition when comparing across firms and over time. These results hold across various uncertainty measures and other earnings-related anomalies, supporting the SIC hypothesis while challenging existing theories. Incorporating the SIC into the existing information choice theories provides a new perspective on return anomalies.

[1]Co-Authors: Zhongjin (Gene) Lu (University of Georgia), Wang Renxuan (China Europe International Business school), Katherine Wood (Bentley University)

4.1 Introduction

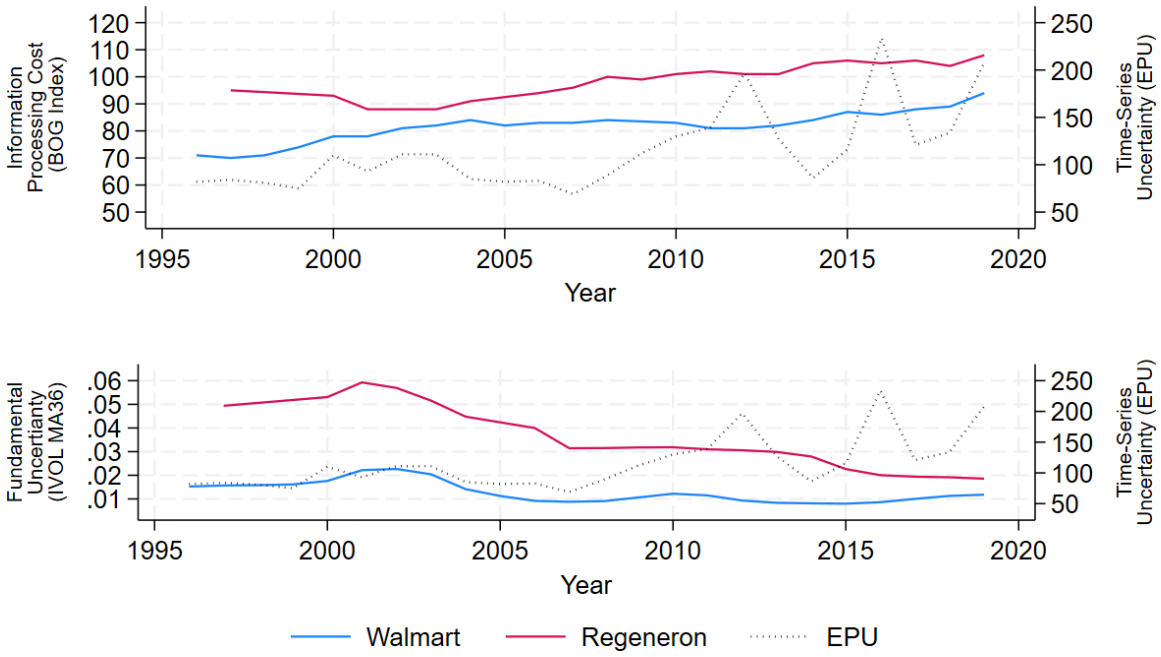
In the era of big data and constant distractions, understanding individuals' information choices holds increasing significance. Existing theories (e.g. Sims, 2003) highlight the role of uncertainty in shaping individuals' information acquisition decisions but have ambiguous predictions on whether higher uncertainty is associated with more or less information acquisition Van Nieuwerburgh and Veldkamp, 2010. Intuitively, with an increasing level of uncertainty, every bit of information becomes more valuable and hence the benefit of acquiring information ("the benefit channel"). Simultaneously, amid heightened uncertainty, the information is potentially more difficult to process, increasing the cost of acquiring information ("the cost channel").

Empirically, whether individuals acquire more or less information when facing heightened uncertainty remains an open question. Existing empirical studies find that investors appear to pay more attention to information when uncertainty is high, supporting the benefit channel Benamar et al., 2021; Bonsall et al., 2020. However, despite the potential importance of the cost channel Blankespoor et al., 2020, there is a lack of understanding and evidence on how the cost channel may affect the relation between uncertainty and information acquisition.

The main contribution of this paper is to propose and test a new hypothesis about how investors' information costs vary with uncertainty when processing firms' fundamental information. Consistent with the predictions of the hypothesis, we document novel empirical facts on how uncertainty and investors' information acquisitions are related, which we show pose challenges to existing theories and support information choice theory as a new perspective for a set of earnings related "anomalies".

The hypothesis, which we term the "Sticky Information Cost" (SIC) hypothesis, posits that investors' marginal costs to process one unit of information (information costs) to forecast a firm's future earnings is related to the firms' *information environment*, which depends on firms' slow-moving characteristics such as its technology and business model Ho and Michaely, 1988.

Figure 4.1: An Illustrative Example of the "Sticky Information Cost" Hypothesis



Note: The graph plots the firm-specific information costs and uncertainty measures of Walmart and Regeneron Pharmaceuticals as of June of each year on the left-hand y-axis and EPU on the right-hand y-axis.

Figure 4.1 illustrates the intuition of the SIC hypothesis. Specifically, the top panel plots the information costs, proxied by the readability of firms' filings (the BOG index as proposed in Bonsall et al., 2017) over time for two firms with different information environment: Regeneron Pharmaceuticals Inc. (NASDAQ:REGN), a pharmaceutical company known for its cutting-edge innovations in biotech and pharma; and Walmart Inc. (NYSE:WMT), an American brick & mortar retail chain with a long-running and straight-forward business model. Intuitively, processing information about the fundamentals of

WMT is simpler than those of REGN, as the latter requires more in-depth knowledge about the promises of the up-to-date technology, and a shorter history of public data than the former.[2] The values of the BOG index for the two firms confirm our intuition — the BOG index of REGN has been persistently higher than that of the WMT over the years.

The SIC hypothesis implies that investors' information processing costs correlate differently with the variations in uncertainty across firms and over time. Across firms, differences in sticky information costs lead investors to acquire less information about firms with higher information costs, resulting in persistently higher price volatility of these firms than those with lower information costs Verrecchia, 1982. Indeed, the bottom panel of Figure 4.1 shows REGN having a persistently higher idiosyncratic volatility than WMT.

At the same time, the SIC hypothesis implies that information processing costs do not respond much to the temporary variations in uncertainty over time because information environment is slow-moving. Figure 4.1 illustrates this point by plotting the EPU (right-hand y-axis), a measure of time-series variation of uncertainty along side the information costs. As the figure demonstrates, the BOG index does not vary with the EPU.

Consequently, the SIC hypothesis predicts divergent cross-sectional and time-series relationships between uncertainty and investors' information acquisition decisions. The strong cross-sectional and the weak time-series correlations between uncertainty and information costs implies that the information cost channel predominantly works in the cross-section, leading to a negative relation between uncertainty; whereas the information benefit channel dominates in the time-series, leading to a positive relation between uncertainty and information acquisition.

[2] Regeneron went public in April, 1991 while Walmart stock started trading in August, 1972.

To empirically test these predictions of the SIC hypothesis, we construct measures of investors' information acquisition and information processing costs. When forecasting firms' future fundamentals, investors encounter a wealth of information, with sell-side analysts' forecasts being a prominent source Kothari et al., 2016. Although sell-side analsyts' forecasts contain value-relevant information, their forecasts are known to contain biases J. S. Abarbanell and Bernard, 1992. Using analysts' forecasts directly would lead to biased forecasts, while de-biasing analysts' forecasts imposes additional costs on investors. We thus measure the extent of investors' information acquisition using the return predictability of analysts' ex-ante human biases (henceforth EHB). If investors choose to spend more costly effort to de-bias analysts' forecasts, we should observe weaker return predictability and vice versa.[3]

Within this setting, we refer to the information costs as investors' marginal costs to process one unit information for de-biasing analysts' forecasts.[4] We measure this cost at the firm level along two dimensions, namely its information scarcity and complexity Zhang, 2006b. If a firm's information is hard to collect (scarce) or to analyze (complex), investors need to deploy more costly capacity to search and integrate one unit of information in their fundamental forecasts. Specifically, we use a firm's age as a measure for information scarcity, as young firms simply have less information available for investors to analyze (Begenau et al., 2018). We use the Bog index and the firms' 10K file sizes as measures of information complexity, which are developed by Bonsall et al., 2017; Loughran and Mcdonald, 2014, 2016 to capture firms' information complexity. Acknowledging that each of the individual measures have noise, we construct an

[3] Empirically, EHB has been shown to persistently predict future returns (e.g. van Binsbergen et al., 2022), confirming that investors continue to rely analysts' forecasts in their investment process Loh and Stulz, 2018, and the processing costs to de-bias analysts' forecasts remain non-trivial.

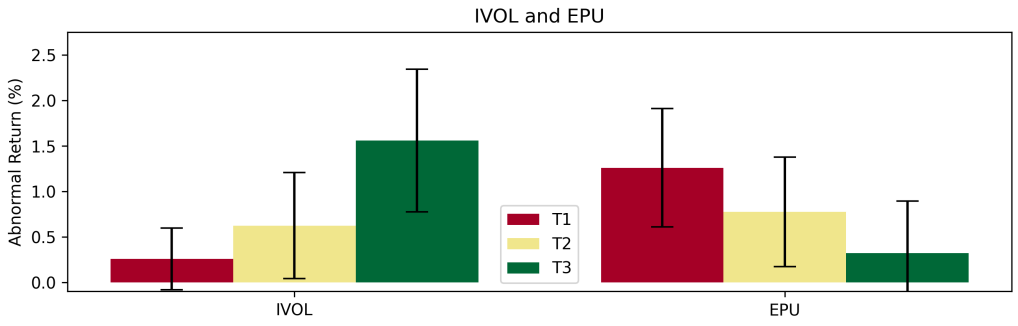
[4] Variations in aggregate marginal information costs are influenced by long-term advancements in information technology, like computer memory costs Farboodi and Veldkamp, 2020. Information costs can be further broken down into awareness, acquisition, and integration costs Blankespoor et al., 2020. In our study, we consider the cost of deciphering analysts' forecasts as an integration cost.

information-cost index that averages these three measures as our main measure of firm-level information costs.

Validating the SIC hypothesis, we find these information cost measures are persistent over time, with large auto-regressive coefficients. Furthermore, they indeed exhibit higher correlations with persistent cross-firm differences in uncertainty than with time-series variation in uncertainty.

Equipped with these measures, we proceed to test the predictions of the SIC hypothesis, which predicts a positive relationship between uncertainty and investors' information acquisition in the time-series, compared to a negative relationship in the cross-section. To test this, we first form long short quintile portfolios based on EHB within each of the uncertainty-sorted terciles and examine how the return predictability of EHB varies. The result is an opposite relationship in the time-series versus the cross-section, as illustrated in Figure 4.2. In this figure, the left half of the panel shows that the (Fama-French Five-Factor) alphas of the long-short portfolios sorted on EHB are the *highest* in the high uncertainty tercile measured by firm-level idiosyncratic volatility ("IVOL"), while the right panel shows that the alphas are the *lowest* in the low uncertainty tercile measured by the Economic Policy Uncertainty index ("EPU").

Figure 4.2: The Opposite Relationships Between Uncertainty and Return Predictability



Note: uncertainty levels are the lowest in tercile 1 (T1) and highest in tercile 3 (T3). The whiskers indicate the 95% confidence interval around point estimates.

We show this contrasting pattern is robust across three different measures of cross-sectional and time-series variations in uncertainty. Furthermore, we decompose the firm-level idiosyncratic volatility (IVOL) measure into a 3-year moving average to capture persistent volatility component stemming from firms' information environment and the deviation from this average for temporary spikes in uncertainty. The data shows stronger return predictability in firms with higher persistent IVOL and weaker predictability in those with abnormal IVOL spikes. These results strongly support the SIC hypothesis and suggests that the information cost channel dominates the cross-sectional relationship between uncertainty and investors' information acquisition decisions.

Next, we test the predictions of the SIC hypothesis using our information cost measures. First, the SIC hypothesis predicts an unambiguous positive relationship between information costs and the return predictability of EHB. Using our empirical proxies, we confirm the EHB return predictability is stronger among firms with higher information costs. Furthermore, *controlling for information costs*, we find that such return predictability vanished almost completely in earnings announcement months and periods with abnormal news coverage, during which uncertainty spikes and information acquisition increases Bonsall et al., 2020. These results further confirm the predictions of the SIC hypothesis, highlighting the importance of considering the properties of information processing costs.

Third, we test the prediction of the SIC hypothesis regarding a broader set of variations in anomaly return predictability. First, we document a strong size effect in the return predictability of EHB, which is consistent with smaller information benefit and higher information cost among small-cap firms. Second, we analyze two prominent earnings-related return anomalies related to analysts' revisions and post-earnings announcement drift. Using these two anomalies to proxy for investors' inefficient processing of earnings-related information, we similarly find a positive relation between uncertainty and information

acquisition in the time series but a negative relation in the cross-section. These results support the SIC hypothesis that the information cost (benefit) channel is the dominant driver of the relation between uncertainty and information acquisition in cross-section (time-series).

Finally, we explore whether alternative explanations based on information demand, behavioral biases, or limits of arbitrage can explain the contrasting cross-sectional versus time-series relationships between uncertainty and the degree to which investors efficiently process analysts' forecasts. Using EDGAR Downloads from Ryans, 2017 as a proxy for information demand, the magnitude of EHB as a proxy for behavioral biases, and the effective bid-ask spreads as a proxy for the trading friction, we find that these alternative stories struggle to explain our empirical findings. Thus, the relation between uncertainty and anomaly returns offers a valuable empirical moment that helps distinguish between these competing theories.

4.1.1 Related Literature

Our paper contributes to the literature concerning the relation between uncertainty and investors' information acquisition. While prior studies (e.g. Andrei et al., 2023; Benamar et al., 2021; Dávila and Parlatore, 2023; Van Nieuwerburgh and Veldkamp, 2010) build theoretical models that analyze this relationship, empirical evidence is limited.[5] Prevailing empirical evidence generally finds a positive relation between uncertainty and investors' information acquisition (Andrei et al., 2023; Benamar et al., 2021; Loh and Stulz, 2018, Andrei et al., 2023), supporting the information benefit channel that higher uncertainty amplifies the marginal benefit of information. In contrast, our findings demonstrate that the costs of information acquisition also play a vital role in driving the relationship between uncertainty and investors'

[5]More broadly, our paper is related to the limited attention literature as attention allocation can be viewed as a form of information choice. See more detailed review of this literature in Baley and Veldkamp, 2021; Mackowiak et al., 2021.

information acquisition.[6] Our study is the first to propose and test the hypothesis that information cost can have distinct relationships with cross-sectional and time-series variations in uncertainty. Our results highlight the necessity to distinguish between these two types of variations in uncertainty to accurately model the information cost channel, which holds significant implications for future research.[7]

Our paper also relates to the literature that aims to understand the role of analysts' forecasts in shaping the markets' earnings expectations, as summarized in Kothari et al., 2016. Specifically, the paper provides an information choice perspective to explain the long-standing puzzle of why investors do not fully unravel analysts' bias Frankel and Lee, 1998; So, 2013; van Binsbergen et al., 2022. Furthermore, we present evidence that shows the SIC hypothesis proposed here can potentially explain a broader range of earnings-related return predictability patterns.

4.2 Theoretical Background and Hypothesis Development

In this section, we describe the SIC hypothesis. We start by applying the standard information choice framework (e.g., Veldkamp, 2011) to model how investors de-bias analysts' forecasts through costly information acquisition. Next, we formally propose the SIC hypothesis and list the predictions of the hypothesis we test.

[6] Our study is thus also related to the burgeoning literature on the role of information acquisition costs and information acquisition (e.g., Blankespoor et al., 2019; D. Chen et al., 2022; Fuster et al., 2022 S. Huang et al., 2022).

[7] Conversely, as both types of uncertainty variations positively influence the information benefit, such a distinction may not be necessary when modeling the information benefit channel.

4.2.1 Theoretical Background: Ambiguous Relationships Between Uncertainty and Information Acquisition

In this model, analysts' forecasts are an important information source for the market's earnings expectations and investors need to do costly de-biasing on analysts' forecasts to obtain more precise signals for firms' future earnings. The model shows that the extent to which investors de-bias analysts' forecasts is endogenously linked to the marginal benefit and cost of information acquisition; both the benefits and costs are related to uncertainty, which result in an ambiguous relationship to investors' information acquisition decision.

Investors' Information Environment

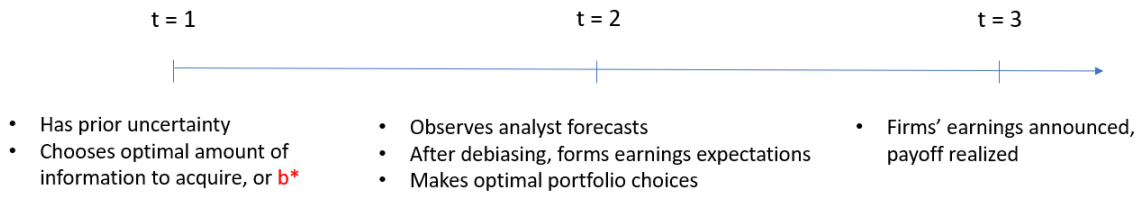
Investors learn about the exogenous, unknown earnings y. For simplicity, we assume earnings are asset payoffs. Investors are endowed with a prior belief that $y \sim N\left(0, \tau_0^{-1}\right)$, with τ_0^{-1} being lower for firms with better information environment. Analysts conduct research to produce forecasts, which are the sum of earnings y and ex-ante predictable errors $B \sim N\left(0, \tau_B^{-1}\right)$.

$$AF = y + B, \tag{4.1}$$

Under the assumption that B is uncorrelated with y, we have $\tau_{AF} = \tau_0 + \tau_B$, with τ_0^{-1} and τ_B^{-1} capturing the variance of prior uncertainty and variance of analysts' bias, respectively.

There are three periods as shown in Figure 4.3. At time 1, investors understand the information environment characterized by τ_0 and τ_B, and they decide how much efforts to put in to unravel analysts'

Figure 4.3: Timeline of investors' information choice and portfolio choice



bias. At time 2, investors observe analysts' forecasts, given the biasing efforts, investors form earnings expectations s,

$$s = AF - b \times B \tag{4.2}$$

where $s \sim N\left(0, \tau_s^{-1}\right)$ with $\tau_s = \tau_0 + \tau_B \times (1-b)^{-2}$.[8] The extent to which investors unravel analysts' bias (i.e., information acquisition) is measured by b, with $b = 1$ meaning fully de-biasing. Based on this signal and exogenously given price and risk-free rates, investors decide how to invest. At time 3, the earnings are realized. Under the simplifying assumptions that stock prices are expected earnings discounted at a constant rate, the volatility of stock prices in our model is characterized by τ_s^{-1}. When investors acquire more information (i.e., unravel analysts' bias more fully), τ_s^{-1} is smaller.[9] In general, due to information acquisition cost, investors do not fully unravel the bias ($b < 1$) and B negatively predicts realized return $y - s$.

[8] We keep the model parsimonious to highlight the key driving forces. In the current set-up, when $b = 1$, investors' signal s is equal to earnings y; but s will be a noisy signal about y if we model analysts' forecasts as $AF = y + B + \epsilon$, with ϵ being noise in analysts' forecasts.

[9] In the data, there are many other forces that drive the stock price volatility such as other sources of information and investors' sentiment.

Investors' optimization problem

Following the standard approach as in Van Nieuwerburgh and Veldkamp, 2010, the Lagrangian problem corresponding to investors' utility optimization generally contains two terms:

$$L\left(b;\tau_0\right)=U\left(\tau_s\right)-c\left(\tau_s\right), \tag{4.3}$$

where $U\left(\tau_s\right)$ is the investors' utility, which is increasing in precision of investors' earnings expectations τ_s; $c\left(\tau_s\right)$ is the cost of obtaining the signal precision τ_s, which is also increasing in τ_s.[10]

The Lagrangian problem in Eq. (4.3) makes clear that investors' optimal information choice τ_s is determined by the marginal benefit and cost of information. Specifically, in our context, investors increase τ_s via unraveling analysts' bias (i.e., increasing b in Eq. (4.2)). The marginal benefit of information, $U^{'}\left(\tau_s\right)$, depends on the prior uncertainty τ_0^{-1}. For both MV and CARA preferences, $U^{'}\left(\tau_s\right)$ is higher when τ_0^{-1} is higher. We refer to this unambiguous relationship between uncertainty and information benefit as the information benefit channel. The common economic intuition is that one extra bit of information cut the prior uncertainty by half and thus the information benefit is higher when prior uncertainty is higher.

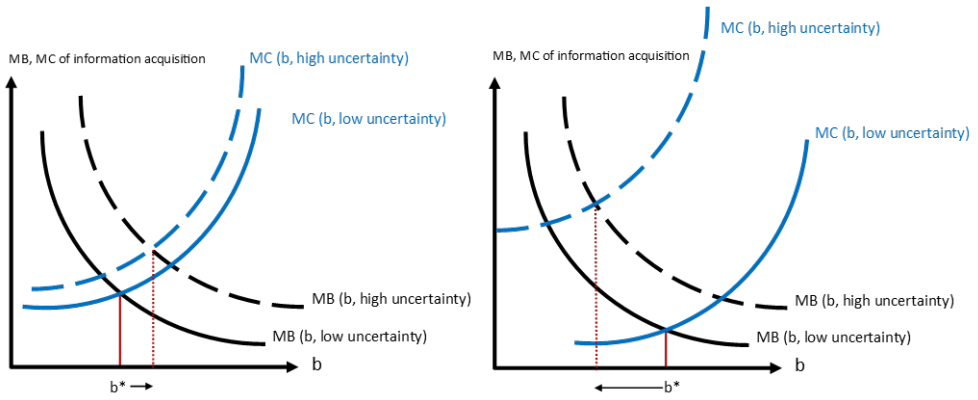
In contrast, how uncertainty affects the marginal cost of information remains ambiguous. We refer to this relationship as the information cost channel, which is different under the two commonly used learning

[10]There are two commonly used preferences: the mean-variance preference (MV) $U\left(W\right)=E_1\left[\rho E_2\left(W\right)-\frac{\rho^2}{2}V_2\left(W\right)\right]$ and the constant absolute risk aversion preference (CARA) $U\left(W\right)=-E_1\left[e^{-\rho W}\right]$, with ρ being the risk aversion. The investors' utility under the optimal portfolio choice is $U\left(\tau_s\right)=c+\frac{1}{2}\frac{\tau_s}{\tau_0}\left(1+\theta^2\right)$ for the former preference and $U\left(\tau_s\right)=-e^{-0.5\theta^2}\times\left(\frac{\tau_s}{\tau_0}\right)^{-\frac{1}{2}}$ for the latter preference, with θ^2 being the squared Sharpe ratio in the economy. There are two commonly used learning technology: the additive precision technology and the entropy learning technology. The resulting cost function is $c\left(\tau_s\right)=a+\lambda\tau_s$ for the former and $c\left(\tau_s\right)=a+\lambda log\frac{\tau_s}{\tau_0}$ for the latter.

technologies. Under the additive learning technology, $c^{'}(\tau_s)$ does not depend on the prior uncertainty τ_0^{-1}, whereas under the entropy learning technology, $c^{'}(\tau_s)$ increases in τ_0^{-1} .

Therefore, in situations in which the marginal cost of information is not related to variations in uncertainty τ_0^{-1}, the information cost channel is muted and the information benefit channel always dominates. This leads to an unambiguously positive relation between uncertainty and the unraveling of analysts' bias. In contrast, in situations in which the marginal cost of information is related to variations in uncertainty τ_0^{-1}, the relationship is ambiguous. Specifically, in the case where the information cost channel dominates, we could observe a negative relationship between uncertainty and the unraveling of analysts' biases—more return predictability of ex-ante analysts' biases.

Figure 4.4: Relation Between Uncertainty and Unraveling of Analysts' Bias



(a) Small Shift (b) Big Shift

Figure 4.4 illustrates the potentially contrasting model predictions regarding the relation between uncertainty and unraveling of analysts' bias. Panel (a) (left) demonstrates the case in which the information benefit channel dominates. As uncertainty increases, the marginal benefits (MB) shifts more than the shift of marginal costs (MC), leading to more de-biasing, i.e. b^* shifts to the right. Panel (b) (right) shows

the opposite prediction when the information cost channel dominates: since MC is more sensitive to uncertainty increases, investors end up de-biasing less, leading to b^* to shift to the left. Although previous literature has presented evidence on the information benefit channel Benamar et al., 2021; Bonsall et al., 2020, the understanding about the information cost channel is limited. The focus of this paper is to propose and test a hypothesis about the information cost channel, which we detail next.

4.2.2 The Sticky Information Cost Hypothesis

Motivated by the intuition that the information processing costs are persistent (Figure 4.1) , we propose the SIC hypothesis:

Hypothesis 1 ***The Sticky Information Cost (SIC) Hypothesis*** *Investors' marginal costs to process one unit of information to forecast a firm's future earnings are slow-moving and vary with the firm's fundamental characteristics, such as technology and business models.*

The hypothesis implies distinctive relationships between investors' information processing costs with the time-series and the cross-sectional variations in uncertainty. Across different firms, firms with more advanced technology and more complicated business models require investors to incur higher marginal costs to de-bias analysts' forecasts. Anecdotally, research analysts working for buy-side firms specializing in investing in bio-tech companies typically require experience in R&D or advanced medical degrees while the entry requirement for analysts' for retail sectors are relatively lower. Simultaneously, these are also firms with higher ex-ante fundamental volatility, e.g. Regeneron vs. Walmart. As a result, the SIC hypothesis implies a positive relationship between information processing costs and uncertainty.

The sticky information costs simultaneously implies that the processing costs do not correlate much with temporary variations in uncertainty over time. Intuitively, if a young technology firm's volatility spikes due to macroeconomic conditions or earnings, the SIC hypothesis posits that the costs for analyzing this firm should not significantly increase as the the firms' fundamental characteristics have not changed significantly.

The SIC hypothesis leads to testable predictions concerning the cross-sectional and time-series relationships between information processing costs, uncertainty and investors information acquisition. First, the SIC hypothesis implies stronger correlation between measures of information costs and uncertainty across firms than over time. We test the predictions using direct measures of information processing costs proposed in the literature.

Second, the SIC hypothesis predicts potentially contrasting relationships between investors' information acquisition and uncertainty in the time-series and across firms. Specifically, the SIC hypothesis implies the information cost channel plays a more dominant role than the information benefit channel in shaping the cross-sectional relationship than the time-series relationship. In the case that the effect of information cost channel is large enough, we could observe a opposite relationships between information acquisition and uncertainty variation in the time-series versus the cross-section. We test these predictions using the return predictability of analysts' ex-ante biases and multiple measures of uncertainty.

Third, the SIC hypothesis implies a positive cross-sectional relationships between information processing costs and information acquisition. Furthermore, time-series variation in uncertainty blurs the relationships between information costs and information acquisition. We test these predictions using our measures of information costs and exogenous time-series variations in uncertainty including firms' earnings announcements and abnormal news coverage.

Finally, the SIC hypothesis provides a new perspective on the variations in return anomalies. It predicts that the return predictability of EHB is weaker among large-cap stocks for which information costs are lower. It also predicts that return anomalies related to investors' inefficient processing of earnings-related information would also exhibit opposite relation with uncertainty in the time series versus in the cross-section.

4.3 Data and Variable Construction

Our sample consists of U.S. common stocks that are covered in the intersection of CRSP, Compustat, and I/B/E/S. We exclude micro-cap stocks, defined as stocks with a market capitalization below the NYSE 20th percentile, and low price stocks, defined as stocks with a price below $5.

We construct the optimal statistical earnings forecasts, following the recommended machine learning (ML) specification in J. L. Campbell et al., 2023.[11] We follow van Binsbergen et al., 2022 and compute the ex-ante measure of the conditional biases in analysts' forecasts as the difference between analysts' forecasts and the ML forecasts in real time, which we refer to as ex-ante human bias (EHB). We generate a composite EHB measure for our analysis that creates a measure with a constant 12 months to maturity.[12] To generate this, we weight one-year-ahead and two-year-ahead EHB. These weights are set up so that the weighted distance from the current month to the fiscal period end is a constant 12 months.[13] Appendix 4.9 provides a detailed description of the input variables and dataset construction used in generating these forecasts as well as a brief description of the methodology used to generate the forecasts. Since ML earnings forecasts

[11] J. L. Campbell et al., 2023 provides a detailed review of the machine learning earnings forecasting literature. The recommended machine learning specification is similar to those used in de Silva and Thesmar, 2022; van Binsbergen et al., 2022.

[12] Our results are robust to alternative specifications of the composite EHB such as the average across the one-quarter, one-year, and two-year ahead EHB measures.

[13] For example, if in month t, the firm is 6 months from the one-year-ahead fiscal period end and therefore 18 months from the two-year-ahead fiscal period end, our composite EHB measure would weight the each individual EHB by 0.5.

require sufficient data in the training sample, our ML earnings forecasts start in June 1990. As a result, our final sample period is from June 1990 through December 2019.

We use the return predictability of EHB to quantify the rationality in market's earnings expectations, as we discussed in the introduction. Detailed variable definitions are provided in Table 4.1. We now turn to our empirical tests.

4.4 Information Costs, Uncertainty and The Market's Information Acquisition

This section tests the predictions from the SIC hypothesis laid out in Section 4.2.2.

4.4.1 Prediction 1: Uncertainty and Information Costs

The SIC hypothesis posits that information processing costs have distinct relationships with cross-sectional and time-series variations in uncertainty. We validate the SIC hypothesis using direct measures of information costs. We describe our empirical measures and then present empirical results on our validation tests.

Direct Measures of Information Costs

Prior literature in finance and accounting (e.g., Begenau et al., 2018; Blankespoor et al., 2020) indicates that information complexity and scarcity are two significant factors influencing information cost. Intuitively, firms with more complex disclosures and less readily available information necessitate higher information processing costs to de-bias analysts' forecasts.

To measure information complexity, we use the Bog index and the log net file size of 10Ks, following the methodologies of Bonsall et al., 2017; Loughran and Mcdonald, 2014, 2016.[14] The Bog index captures the plain English attributes of 10K statements, focusing primarily on the writing clarity of firms' disclosures. In contrast, the log net file size provides a simple and effective gauge of the overall complexity of the firm. As Loughran and Mcdonald, 2016 argue, the readability of 10Ks and the business complexity are ultimately intertwined, so we employ both measures jointly to capture information complexity.

To measure persistent firm-level differences in information scarcity, we use firm age, which is the number of months since the first trading day for each firm. The idea is that as a firm ages, more information becomes available for investors to analyze its fundamentals. As an example of firms' fundamental information, by the end of 2020, IBM (who had their IPO well before EDGAR came into existence) had 105 10K and 10Q filings since the beginning of EDGAR , whereas Tesla, which filed their IPO in 2010, only has 24.

We recognize that accurately measuring information processing costs is challenging, and each of the three individual measures may contain measurement errors. To address this concern, we construct an information cost index (IC index) that integrates the three individual measures. Specifically, at the end of June of each year, we first orthogonalize the cross-sectional normalized rank of each measure against the cross-sectional normalized rank of the Size to control for the impact of firm size. We then average the residuals of these regressions to create the information cost index. This measure is applied from June of year t to May of $t + 1$. Intuitively, firm size is correlated with investors' information processing costs.[15] However, in our main analysis, we want to first show the predictions of SIC directly concerns firms'

[14] Bonsall et al., 2017; Loughran and Mcdonald, 2014, 2016 show that the Bog index and the net file size are superior measures for capturing information complexity than the Fog index. We download these measures directly from their respective websites.

[15] Table 4.9 in Appendix 4.10 shows that there are strong correlations between the components of the IC index and Size.

information costs, and the results we document is not primarily driven by variations in firm sizes. In Section 4.4.4, we discuss separately the relationship between firm size and return predictability and show our results related to the information costs hold also without controlling for size in Table 4.12 in Appendix 4.10.

Information Costs and Variations in Uncertainty

Based on measures of information scarcity and complexity described above, we examine the cross-sectional and time-series relations between information costs and uncertainty with three validation tests.

We start by examining the persistence of the information cost measures (firm age, readability, complexity). First, we regress the measures on their one-year lagged values. The regression coefficients are 0.85 for firm age, 0.92 for the Bog Index, and 0.64 for net file size, which correspond to a half-life of 4.27, 8.31 and 1.55 respectively.[16] These results are consistent with the notion that firm-level information processing costs evolve slowly over time.

We then directly test the predictions of SIC that uncertainty has a stronger association with information costs in the cross section than in the time series. In our first test, we regress firm-level uncertainty (IVOL) on different measures of information costs, controlling for firm size as well as firm- or time- fixed-effects. Regressions with the time fixed-effects capture the cross-sectional relation between uncertainty and information costs whereas those with firm fixed-effects capture the time-series correlations between uncertainty and information costs. As Table 4.2 shows, the coefficient estimates associated with the cross-firm relation (Columns 2, 4, 6) are consistently higher than those with the time-series relation (Columns

[16]We conduct this analysis on an annual basis as the Bog Index and Net File Size are updated annually with the 10K and Firm Age is slow moving.

1, 3, 5). Furthermore, the statistical significance are always stronger for the cross-sectional relation (with time fixed-effects) than that of the time-series relation.

In our second test, we first dissect firm-level idiosyncratic volatility into two distinct components: a persistent component that captures cross-firm differences in fundamental uncertainty and a time-series variation component that captures temporal fluctuations in uncertainty. Empirically, we use the a firm's rolling average of past 36-month IVOL, ("IVOL_{MA36}") to proxy for the former and the difference between the current value of IVOL and the persistent component ("Abnormal IVOL") to proxy for the latter. We then examine whether information measures have a more positive correlation with the persistent component. As Figure 4.5 shows, the IC index, along with its consisting measures, all show positive correlation between IVOL_{MA36}, while having a slightly negative correlation with Abnormal IVOL. These results further confirm SIC.

In summary, our results in this subsection support the hypothesis that information costs are more strongly correlated with persistent cross-firm differences in uncertainty than with time-series variation in uncertainty.

4.4.2 Prediction 2: Uncertainty and Information Acquisition

The SIC hypothesis predicts the information cost channel plays a more dominant role in the cross-sectional relationship between uncertainty and investors' information acquisition as compared to the time-series relationship. We test this prediction using the return predictability of EHB as the measure for information acquisition. Specifically, at the end of each month, we sort stocks into tercile portfolios based on an uncertainty measure. Within each uncertainty group, we further sort stocks into quintile portfolios based on EHB measures (i.e., analysts' ex-ante bias). We then compute the abnormal returns

(i.e., Fama-French Five-Factor alpha's) of the resulting 15 value-weighted portfolios as well as the high-minus-low return on the EHB Q1-Q5 portfolio. Figure 4.2 shows the contrasting relations between the return predictability of EHB and IVOL versus EPU using the high-minus-low portfolios.

The contrasting relationships in Figure 4.2 confirm the prediction of the SIC hypothesis. We further examine if the relationships we show in Figure 4.2 is robust across different measures of uncertainty— be it realized, forward-looking, firm-specific, or aggregate.

In Figure 4.6 and Table 4.3, we use evaluate two measures of firm-specific uncertainty. First we use IVOL, defined as the standard deviation of the residuals from CAPM regressions using the past year of daily data (Ali et al., 2003; Ang et al., 2006), as a measure of firm-specific realized uncertainty. To test whether differences between realized and forward-looking uncertainty drive the pattern in Figure 4.6,we employ a forward-looking firm-level uncertainty measure, the option implied volatility (OIV). OIV is the average implied volatility from a call and put option with 30 days to maturity and a delta of 0.5 (-0.5 for a put option) from the volatility surface file of OptionMetrics on the last day of the month.

Similarly, because the Economic Policy Uncertainty Measure (EPU) provided by Baker et al., 2016 is a backward looking aggregate uncertainty measure, we also use the forward-looking macroeconomic uncertainty (MU) provided by Jurado et al., 2015 and Ludvigson et al., 2021.[17] We prefer MU to the VIX index as the forward-looking aggregate uncertainty measure because VIX is also affected by risk premium. Figure 4.6 provides a visual representation of the long-short portfolio across uncertainty terciles of the information shown in Table 4.3.

Panels A and B of Table 4.3 show how the return predictability of EHB varies across periods with different levels of MU and EPU, respectively. Although MU captures forward-looking economic uncer-

[17]According to Baker et al., 2016, EPU "capture(s) uncertainty about who will make economic policy decisions, what economic policy actions will be undertaken and when, and the economic effects of policy actions (or inaction)."

tainty whereas EPU captures the prevailing economic uncertainty, both panels show consistent patterns that the return predictability of EHB is the weakest when in high uncertainty periods. Specifically, for MU T1 and T2, the long-short portfolio based on EHB ("EHB Q1-Q5") generates an average abnormal returns of 0.717% and 0.609% per month, both statistically significant. In contrast, the long-short portfolio abnormal returns for MU T3 (i.e., periods when MU is highest) are only 0.142%, and statistically insignificant. Similarly, the EHB Q1-Q5 portfolio generates average abnormal returns of 1.199% per month when EPU is lowest ("EPU T1"), and 1.024% per month when EPU is second lowest ("EPU T2"), only 0.463% per month when EPU is highest. Moreover, the difference between the two long-short portfolio returns in EPU T1 and T3 is statistically significantly positive, amounting to 0.736% per month.

To the best of our knowledge, we are the first to document systematic time-series variations in EHB return predictability. These results indicate that investors acquire more information to de-bias analysts' forecasts during periods of higher uncertainty, which is consistent with the information benefit channel being the dominant force of investors' information choice when time-series uncertainty is high. Our results thus corroborate prior findings in Benamar et al., 2021; Bonsall et al., 2020; Hirshleifer and Sheng, 2022, supporting the important role of the information benefit channel in explaining the relation between uncertainty and information acquisition.

Panel C and D of Table 4.3 show how the return predictability of EHB varies with firm-level uncertainty, as measured by OIV and IVOL respectively. Contrary to patterns in Panel A and B, the return predictability of EHB is the strongest in the high-uncertainty group. Specifically, the FF5 alpha of the EHB Q1-Q5 portfolio increases in uncertainty groups, yielding statistically insignificant abnormal returns of 0.118% per month for OIV T1 and 0.181% per month for IVOL T1 compared to statistically significant returns of 1.594% per month for OIV T3 and 1.780% per month for IVOL T3.Having demonstrated that the

difference between realized and forward-looking uncertainty does not drive the contrasting contrasting relationships between uncertainty and the return predictability of EHB in the time-series versus cross-section,we now proceed to examine whether the difference between firm-specific and aggregate uncertainty is driving the contrasting relationship.

Specifically, we we use the time-series and cross-sectional variations in firm-specific uncertainty (IVOL) to test the prediction. If our SIC hypothesis is correct, then even using firm-specific uncertainty, we should observe that the return predictability is positively related to IVOL_{MA36} (the persistent component of IVOL) and but negatively related to Abnormal IVOL (the time-varying component of IVOL).

Table 4.4 presents the abnormal returns of portfolios based on IVOL_{MA36} in Panel A and Abnormal IVOL in Panel B, respectively. Panel A shows, when uncertainty is measured by IVOL_{MA36}, we observe consistent pattern of abnormal returns similar to what is seen when uncertainty is measured by IVOL. EHB exhibits the strongest return predictability within the high uncertainty group (IVOL_{MA36} T3), with the long-short portfolio yielding 1.209% (t-stat = 2.84) abnormal returns per month. Conversely, EHB has the weakest return predictability within IVOL_{MA36} T1, yielding statistically insignificant 0.178% (t-stat = 1.04) per month abnormal return.

Panel B shows that sorting on firms' Abnormal IVOL leads to the opposite results—EHB exhibits the strongest return predictability among firms with the lowest Abnormal IVOL (T1). Indeed, the long-short portfolio based on EHB within the Abnormal IVOL T1 leads to 0.693% (t-stat = 2.35) monthly abnormal returns, compared to a 0.342% (t-stat = 1.18) per month abnormal returns within the Abnormal IVOL T3.

In sum, these results show that the contrasting relationships between uncertainty and the return predictability of analysts' ex-ante biases also hold at the firm level, further corroborating the predictions of SIC.

Our results are related to the finding in Zhang, 2006b that the return predictability of analysts' revisions are positively correlated with IVOL. Our innovation is to jointly consider time-series and cross-sectional variations in uncertainty and we show that the time-series and cross-sectional variations in IVOL have distinct relation with the information acquisition proxied by return predictability of EHB.

4.4.3 Prediction 3: Information Costs and Information Acquisition

The SIC hypothesis predicts an unambiguously negative relationship between information costs and investors' information acquisition. Therefore, we should observe the return predictability of EHB (the negative of the information acquisition) to be stronger among firms with higher information costs. Furthermore, the SIC hypothesis predicts that controlling for information costs, there is a negative relationship between uncertainty and return predictability in the time-series.

To test the cross-sectional relationship, we first sort stocks according to the IC index into terciles, and within each IC tercile, we further sort stocks into quantiles based on EHB. Table 4.5 shows the Fama-French Five-Factor alphas of these 15 portfolios. In alignment with SIC's prediction, these results show that the return predictability of EHB increases with information costs, as measured by the IC index. Specifically, for firms with highest information costs (IC Index T3), the long-short portfolio based on EHB (EHB Q1-Q5) generates a monthly abnormal return of 1.076% (t-stat = 2.91). This monthly abnormal return declines monotonically to 0.850% (t-stat = 2.95) for IC index T2 and finally to 0.405% (t-stat = 1.51).

To test the time-series relationship, we examine whether the temporary variation in uncertainty over time weakens the positive relationships between information costs and return predictability of EHB. We employ two firm-level measures of variation in uncertainty. Our first measure, abnormal news coverage ("High Media"), is calculated from the ratio of the count of news stories in a given month from the Dow

Jones Index to the rolling 36-month average of that count.[18] As shown in Bonsall et al., 2020, firms with abnormal media coverage experience abnormal volatility.

Our second measure, earnings announcement months ("Earn. Annc."), is an indicator variable which equals to one for months in which a firm has an earnings announcement and equal to zero in other months. As shown in Dubinsky et al., 2019, firms experience increase implied volatility during earnings announcement months.[19]

Based on these measures, we test the prediction that, after controlling for cross-sectional differences in firm-level information costs (information scarcity), time-series variations in uncertainty positively correlate with the return predictability of EHB. Specifically, we first create terciles based on firm age, our measure for cross-sectional difference in information scarcity. Within each tercile, we run pooled monthly regressions of one-month-ahead returns on firms EHB, the measures of the temporary change in information scarcity, and their interactions. We run the regressions with and without controls for commonly used determinants of stock expected returns (i.e., firms' market capitalization, operating profits, asset growth, book to market ratio, 6-month price momentum). All analysis includes time fixed-effects. Our variable of interests is the coefficient associated with the interaction variable between EHB and the two measures of information availability.

Table 4.6 presents the results, with Panel A and B showing results for the abnormal news coverage and earnings announcement months, respectively. Both panels consistently support the hypothesis that greater information scarcity strengthens return predictability. Consistent with the cross-sectional results in Table 4.5, EHB return predictability is the strongest among the youngest firms ("T3").

[18] We utilize Ravenpack to obtain a count of news stories. We limit our count of news coverage to the Dow Jones Index and require the story to have a relevance measure of 100, which we base off the methodology and code from X. Chen et al., 2022.

[19] The abnormal news coverage starts after 2001, while the earnings announcement months are available throughout our full sample.

Furthermore, we find positive estimates of the coefficient associated with the interaction between EHB and the time-variations of firm-level uncertainty. For both measures, the coefficients associated with the interaction variables ("EHB x High Media" and "EHB x Earn. Annc") are positive for "T3" and "T3-T1", which means that the significantly negative relationship between EHB and future returns among younger firms is weakened during periods with high uncertainty. In terms of the magnitude, the positive coefficients on the interaction term are close to the coefficients on EHB. These results are to robust whether or not we include controls.[20] Thus, the additional return predictability of EHB among the younger firms is reduced to the large extent during earnings announcement months or when there is abnormal news coverage. These results support the second prediction from the SIC hypothesis laid out in Section 4.2.2.

4.4.4 An Information-Choice Perspective on Return Predictability

Embedding the SIC hypothesis into information choice theories also provide a new perspective on a broader set of variations in return predictability.

Variation of Return Predictability of EHB across Firm Size

Return predictability is known to be stronger among smaller firms. Nevertheless, information choice theory provides an alternative perspective regarding why the return predictability of EHB should be weaker among larger firms. From the information benefit channel, de-biasing analysts' bias brings more benefit as big firms account for a larger share of the investors' total wealth; from the information cost channel, big firms produce more data and therefore have reduced information processing costs of investors relative

[20] In Appendix 4.10, we show that these results still hold when we create information scarcity terciles based on 1/Age residualized to size, showing the effects are not driven by firm size.

to smaller firms Begenau et al., 2018. Therefore, both information benefit and cost channels predict that the return predictability of EHB should decrease with firm sizes.

Table 4.7 presents evidence supporting this prediction. First, we examine the return predictability of EHB among different size segments based on NYSE breakpoints. The abnormal returns of different EHB portfolios are presented in Panel A. Consistent with the hypothesis that firm size correlates with investors' information processing costs, EHB return predictability decreases monotonically in Size. The long-short EHB portfolio (EHB Q1-Q5) has the highest abnormal return among small-cap stocks, yielding 0.930% per month (t-stat = 4.28). The abnormal return declines to 0.690% per month (t-stat = 3.18) for large caps and 0.337% per month (t-stat = 1.72) for the mega caps. The difference in abnormal returns between mid- and mega-cap stocks is 0.593% per month (t-stat = 3.91), which is economically significant.

The Relation Between Uncertainty and Other Analysts' Forecasts Related Anomalies

Besides the size effect, we show the SIC hypothesis prediction holds for the announcement day returns Bernard and Thomas, 1990 and analysts' forecast revisions Givoly and Lakonishok, 1980.

We adopt the same portfolio sorting methodology in Figure 4.6 and show the results for the announcement day returns and analysts' forecast revisions in Figures 4.7 and 4.8, respectively. Notice that the results here are for non-mega cap stocks, whose contrasting patterns are stronger than those with mega-caps.[21] This is driven by our discussion in the previous subsection. Consistent with the pattern we find based on EHB, the return predictability associated with both variables are positively correlated with the persistent, cross-firm variations in uncertainty as measured by OIV, IVOL or IVOL_{MA36}, while simultaneously negatively correlated with the temporal variations in uncertainty as measured by MU, EPU or Abnor-

[21]We present the mega-cap results in Figures 4.12 and 4.13 in Appendix 4.10. .

mal IVOL. These results are consistent with our hypothesis that the information benefit (cost) channel is the dominant driver of the relation between uncertainty and the extent to which investors efficiently process analysts' forecasts in the time-series (cross-sectional) dimension because information costs covary more strongly with the persistent, cross-firm variations in uncertainty. Given that these two variables have been shown to be persistent and robust predictors for future returns and are able to price a broad sets of asset returns Daniel et al., 2020; Kothari et al., 2016, these results provide another piece of evidence supporting the information choice perspective in viewing return predictability. Next, we compare this information-choice based explanation to alternative theories proposed in the literature.

4.5 Alternative Explanations

Existing theories of return predictability emphasize the role of risk exposures, behavioral biases, and limits of arbitrage. In this section, we explore whether these theories explain our key empirical finding—the contrasting relationship between uncertainty and the degree to which the market efficiently process analysts' forecasts in the cross-section versus in the time-series.

First, risk-based theories might account for the contrasting cross-sectional and time-series correlations between uncertainty and EHB return predictability if the risk exposures of the EHB Q1-Q5 long-short portfolio relate oppositely to uncertainty across these two dimensions. However, no theoretical model to date has posited such a mechanism. Empirically, if the FF5 model accurately reflects appropriate risk exposures, our results, based on the FF5 alphas, imply that risk-based theories fall short of explaining our empirical finding.

Second, explanations grounded in behavioral biases suggest two potential explanations. The first possibility is that analysts' biases correlate positively with cross-sectional fluctuations in uncertainty, yet negatively with time-series fluctuations. However, contrary to this conjecture, most behavioral theories (e.g., Hirshleifer, 2001) propose an unambiguously positive link between uncertainty and human biases. Empirically, we directly test this conjecture by regressing the magnitude of our measure of analysts' bias (i.e., EHB) on uncertainty measures. The left-most columns of Table 4.8 show the results. We observe a positive correlation between analysts' bias and both cross-sectional and time-series variations in uncertainty, thereby not supporting this conjecture.

The second possibility is that investors' attention correlates negatively with cross-sectional variations but positively with time-series variations in uncertainty. Behavioral theories suggest that the relationship between uncertainty and attention is complex and depends on whether uncertainty either diverts or draws attention. Empirically, we directly test this possibility by regressing a measure of attention (Human Downloads from EDGAR from Ryans, 2017) on uncertainty measures. Contrary to this possibility, we find that EDGAR downloads are positively related to cross-sectional variations in uncertainty.

Finally, theories based on limits of arbitrage could rationalize the contrasting relationship if trading costs are positively associated with cross-sectional variations but negatively with time-series variations in uncertainty. Contrary to this conjecture, micro-structure theories predict an unambiguous positive relation between uncertainty and trading costs as higher uncertainty leads to increased information asymmetry and thus higher trading costs. Empirically, we directly test this explanation by regressing a trading cost measure (i.e., the effective spread) on uncertainty measures. We find that trading cost is positively

or insignificantly correlated with both time-series variations and cross-sectional variations in uncertainty, which contradicts this hypothesis but aligns with micro-structure theories.[22]

In summary, alternative theories of return predictability struggle to explain the contrasting relationship between uncertainty and the return predictability of EHB. Admittedly, the empirical tests presented in this section do not directly reject fully all variations of models within each theories. However, our analysis here underscores that the primary empirical result detailed in this paper offers a valuable empirical moment that helps distinguish information choice models from these competing theories.

4.6 Conclusion

In this study, we have delved into the intricate relationship between market uncertainty, information costs and the rationality of earnings expectations in financial markets. Our key contributions lie in establishing that while higher uncertainty increases the benefits of information acquisition, it also elevates the associated information processing costs. Specifically, by utilizing the return predictability of analysts' biases as a barometer, we found opposite relationships between uncertainty and investors' information acquisition in the time-series and the cross-section.

Our novel hypothesis, SIC, explains how information processing costs fluctuate with varying levels of uncertainty, shedding light on the complex mechanics of information choice in financial markets. These findings have implications for understanding a broader set of return predictability patterns and pose new empirical moments for future theories to match. Furthermore, another interesting direction is to explore the micro-foundation for our SIC hypothesis.

[22]Variations in EPU does not significantly covary with trading costs. The three other measures do positively covary with trading costs.

4.7 Figures

Figure 4.5: Correlation Matrix of Information Cost Index Components and IVOL components

This figure shows the Spearman correlations for the components of the Composite Information Cost Index and the components of IVOL. As the Information Cost Index consists of measures that update infrequently (the Bog Index, Complexity, and Net File Size update annually and Firm Age is slow moving), the analysis is done as of the end of June in each year. All variables use the normalized rank (i.e., the rank scaled by the number of stocks in the cross-section) orthogonalized to the normalized rank of Size. The Composite Information Cost Index is the average of the residual of the normalized rank of -LN(Firm Age), the Bog Index, and LN(Net File Size) each orthogonalized to the normalized rank of Size. For comparability, IVOL and its components ($IVOL_{MA36}$ and Abnormal IVOL) are also orthogonalized to the normalized rank of Size. The sample period for Firm Age and IVOL (and its components) begins in June 1990, the annual sample period for the Bog Index and Net File Size begins in June 1996 All samples end in December 2019.

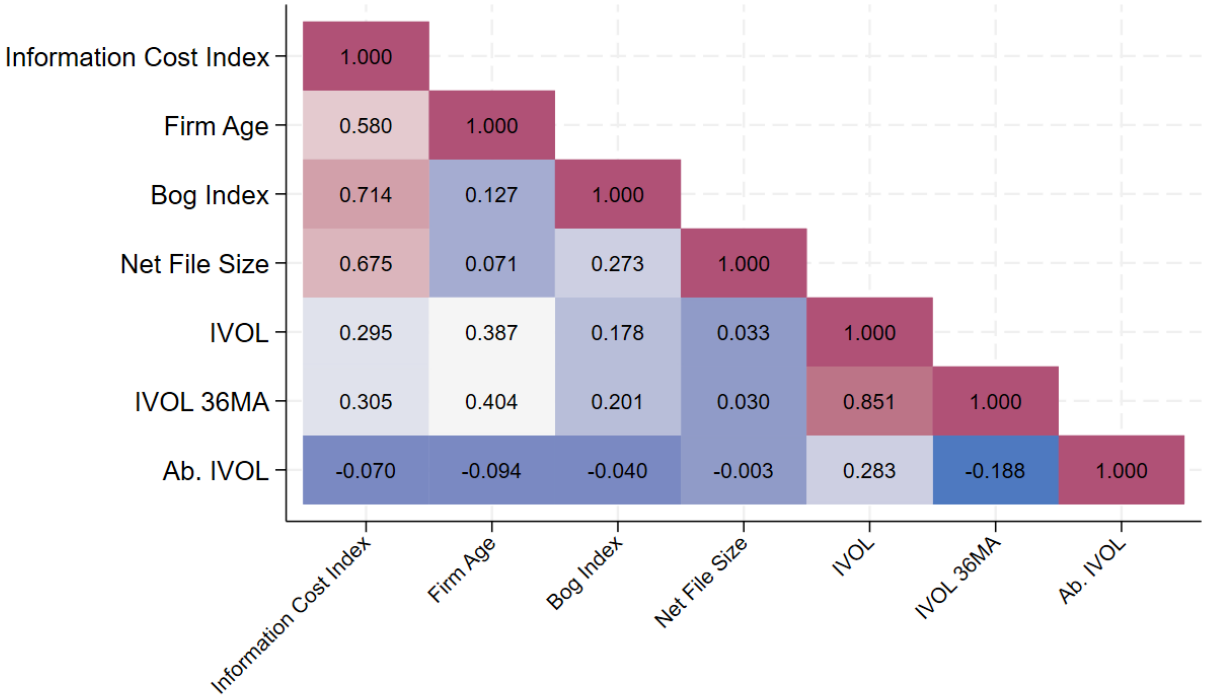


Figure 4.6: Return Predictability of EHB by Uncertainty Terciles

This figure shows the Fama-French Five-Factor alphas of the EHB Q1-Q5 portfolios by uncertainty terciles using six monthly uncertainty measures. The EHB Q1-Q5 portfolios based on OIV , IVOL, IVOL$_{MA36}$, and abnormal IVOL, are made by first cross-sectionally sorting companies into terciles based on the specific uncertainty measure. Then firms are conditionally cross-sectionally sorted into quintiles based on EHB. The EHB Q1-Q5 portfolios based on MU, and EPU are made by sorting observations into terciles in the time series based on each uncertainty measure. Then we conditionally cross-sectionally sort firms into quintiles based on EHB. Portfolios are value weighted and are re-balanced on a monthly basis. Q1 (Q5) contains firms with the lowest (highest) values of EHB. T1 (T3) of each uncertainty tercile contains firms with the lowest (highest) values. The average value and test statistics are presented below their corresponding bars. Whiskers denote 95% confidence bands. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. The sample period is from June 1990 to December 2019, with the exception of OIV which begins in January 1996.

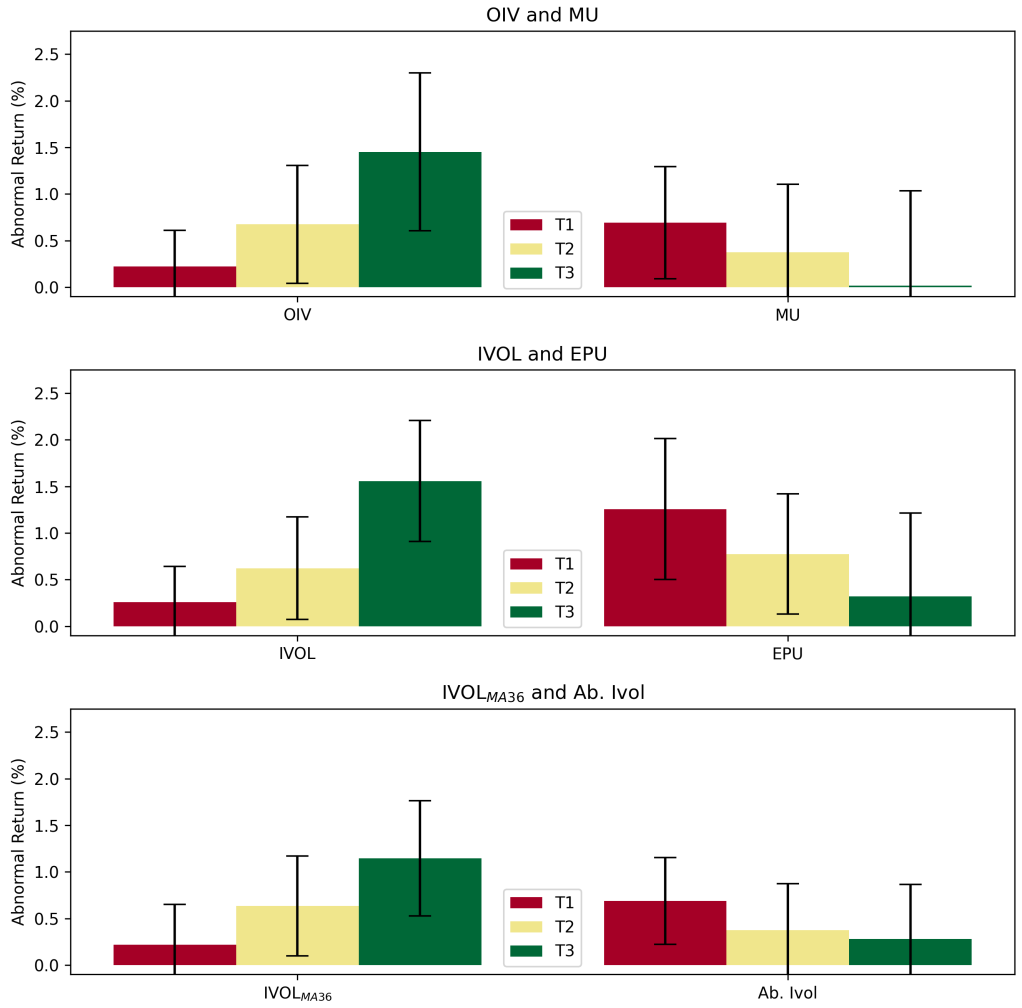


Figure 4.7: Return Predictability of Announcement Return by Uncertainty Terciles: Non-Mega Cap

This figure shows the Fama-French Five-Factor alphas of the Announcement Return Q5-Q1 portfolios by uncertainty terciles using six monthly uncertainty measures. This figure uses only firms in the non-mega cap subsample. The Announcement Return Q5-Q1 portfolios based on OIV , IVOL, IVOL$_{MA36}$, and abnormal IVOL, are made by first cross-sectionally sorting companies into terciles based on each uncertainty measure. Then firms are conditionally cross-sectionally sorted into quintiles based on their announcement return. The Announcement Return Q5-Q1 portfolios based on MU, and EPU are made by sorting observations into terciles in the time series based on each uncertainty measure. Then we conditionally cross-sectionally sort firms into quintiles based on their Announcement Return. Portfolios are value weighted and are re-balanced on a monthly basis. Q1 (Q5) contains firms with the lowest (highest) values of the Announcement Return. T1 (T3) of each uncertainty tercile contains firms with the lowest (highest) values. The average value and test statistics are presented below their corresponding bars. Whiskers denote 95% confidence bands. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. The sample period is from June 1990 to December 2019, with the exception of OIV which begins in January 1996.

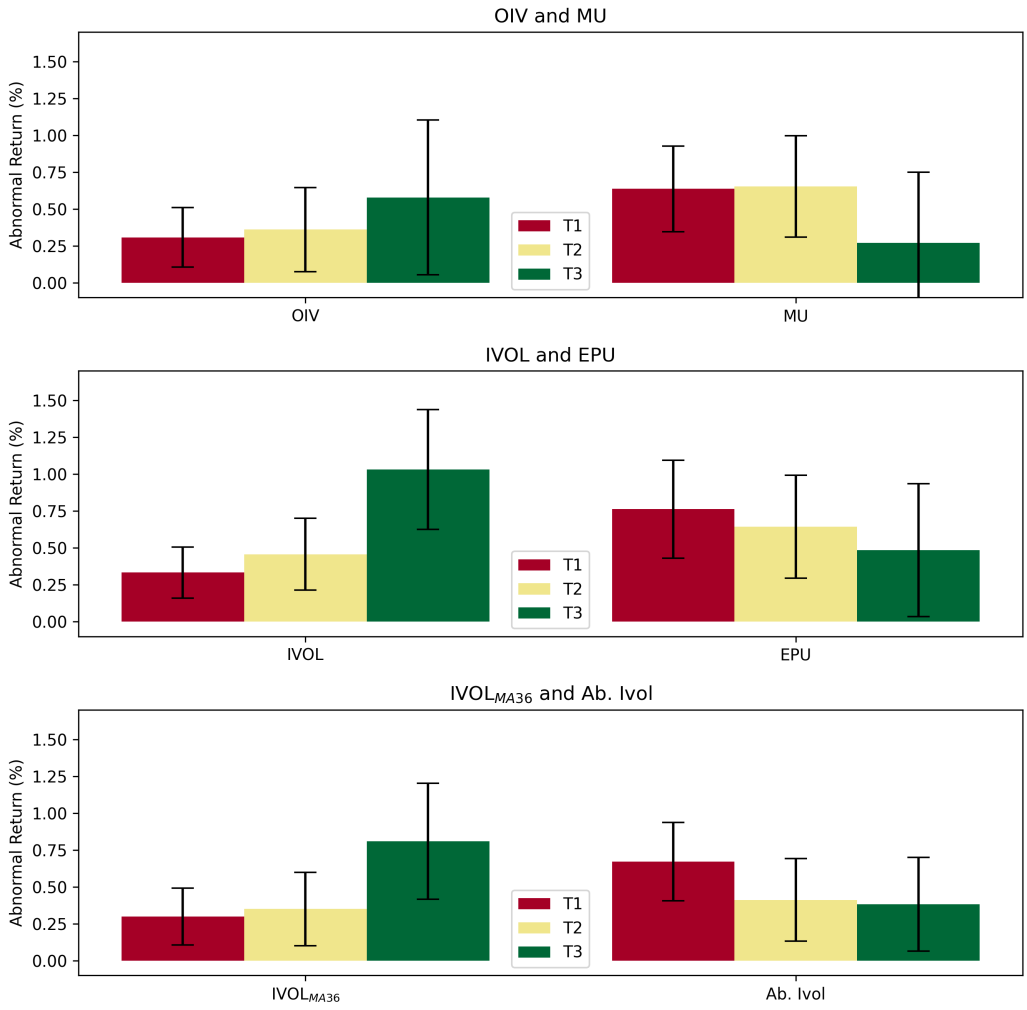
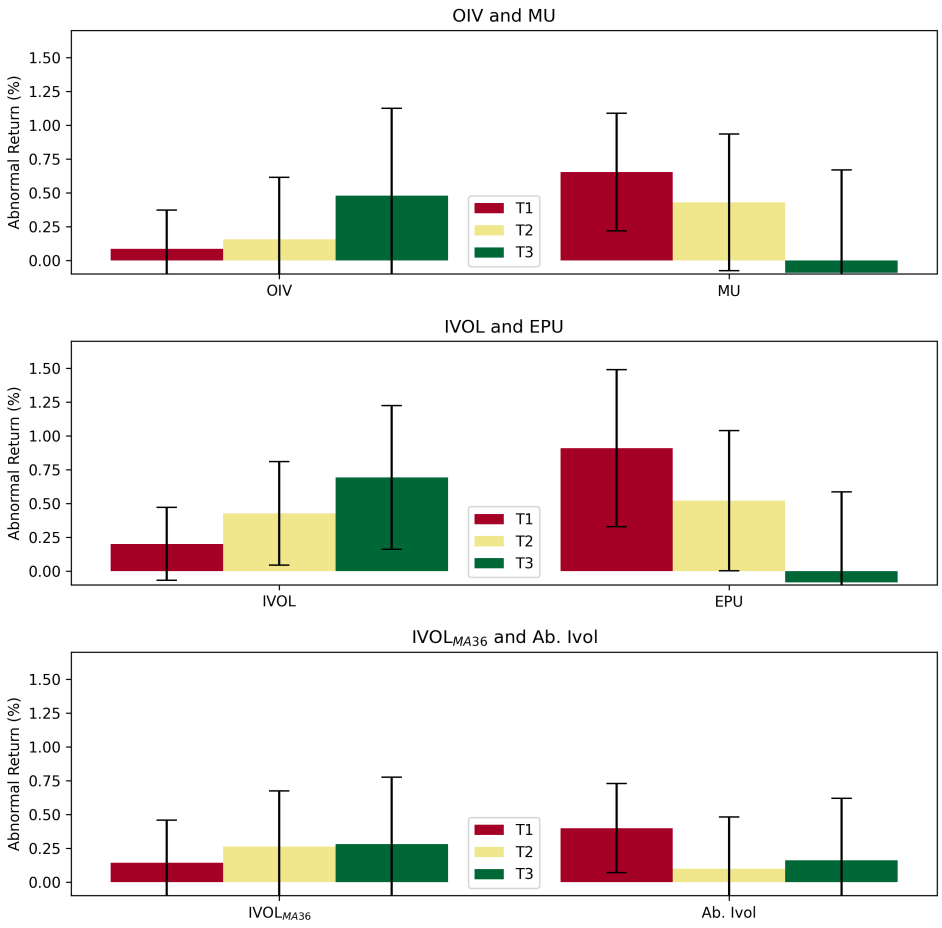


Figure 4.8: Return Predictability of Analysts' Revisions by Uncertainty Terciles: Non-Mega Cap

This figure shows the Fama-French Five Factor alphas of the analysts' revision Q5-Q1 portfolios by uncertainty terciles using six monthly uncertainty measures. This figure uses only firms in the non-mega cap subsample. The analysts' revision Q5-Q1 portfolios based on OIV , IVOL, IVOL$_{MA36}$, and abnormal IVOL, are made by first cross-sectionally sorting companies into terciles based on each uncertainty measure. Then firms are conditionally cross-sectionally sorted into quintiles based on their analysts' revision. The analysts' revision Q5-Q1 portfolios based on MU, and EPU are made by sorting observations into terciles in the time series based on each uncertainty measure. Then we conditionally cross-sectionally sort firms into quintiles based on their analysts' revision. Portfolios are value weighted and are re-balanced on a monthly basis. Q1 (Q5) contains firms with the lowest (highest) values of the analysts' revision. T1 (T3) of each uncertainty tercile contains firms with the lowest (highest) values. The average value and test statistics are in presented below their corresponding bars. Whiskers denote 95% confidence bands. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. The sample period is from June 1990 to December 2019, with the exception of OIV which begins in January 1996.



4.8 Tables

Table 4.1: Key Variable Definitions

This table provides the definition of key variables in the analysis.

Variable	Definition
Ex-Ante Human Bias (EHB)	Analysts' conditional biases (Analysts' forecast-ML Model forecast) (Constant 12 months to Fiscal Period End calculated as the weighted average of FY1 and FY2). EHB is scaled by price
Firm Age	Firm Age (months)
Size	Ln(Market Capitalization) (daily or as of end of month)
Mega Cap	Firms with market capitalization above 80th percentile of NYSE firm size
Small Cap	Firms below median NYSE market capitalization
Mid Cap	Firms above median NYSE market capitalization but not Mega Cap
IVOL	Standard deviation of residuals from CAPM regressions using the past year of daily data. (Require at least 100 non-missing observations.)
OIV	Average of call and put option implied volatility from the volatility surface using 30 day maturity and delta=0.5 (-0.5 for put options) on last day of the month.
IVOL_{MA36}	Moving Average of IVOL from month $t-35$ to t (Trailing IVOL)
Abnormal IVOL	$\frac{IVOL}{IVOL_{MA36}}$
MU	One Month Macro Uncertainty Measure (Ludvigson et al. 2021)
EPU	Economic Policy Uncertainty Index (Baker et al. 2014)
Announcement Return	Sum of risk adjusted returns from two days before and earnings announcement to one day after the announcement
Analysts' Revision	Three month revision in analysts' forecasts for one-quarter ahead earnings forecasts
Bog Index	Plain English Readability Measure Applied to 10Ks (Bonsall et al. 2017)
Net File Size	File Size of 10K excluding ASCII-encoded insertions, HTML, and XBRL (Loughran and McDonald 2014)
Information Cost Index	Average of the cross-sectional normalized ranks of 1/Age, Bog Index, and Net File Size each cross-sectionally orthogonalized to Size. Measure is created in June of year t and is then used until May of year $t+1$
Abnormal Media	Difference between the count of DJ news articles in month t and the 36 month (from $t-35$ to t) moving average
High Media	Indicator variable equal to 1 (-1) for firms in the top (bottom) quintile of abnormal media cross-sectionally orthogonalized to size and 0 otherwise
Earnings Announcement	Indicator equal to 1 for months in which a firm has an earnings announcement
EDGAR Downloads	Count of Human Downloads from EDGAR for a given month (Ryans 2017)
Effective Spread	Monthly Average Effective Spread using TAQ data

Table 4.2: IVOL and Information Cost

This table presents the results of pooled OLS regression of the components of the Information Cost Index (Firm Age, the Bog Index, and Net File Size) on LN(IVOL) and Size. As the Information Cost Index consists of measures that update infrequently (the Bog Index and Net File Size update annually and Firm Age is slow moving), the regression is run as of the end of June in each year. Columns 1 and 2 use -LN(Firm Age), columns 3 and 4 use the Bog Index, and columns 5 and 6 use LN(Net File Size) Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. Standard errors are clustered at the firm and year level. Columns 1, 3, and 5 include firm fixed effects while columns 2, 4, and 6 include time fixed effects. The annual sample period for Firm Age begins in June 1990, the sample period for the Bog Index and Net File Size begins in June 1996. All samples end in December 2019.

	-LN(Age)		Bog Index		LN(Net File Size)	
	(1)	(2)	(3)	(4)	(5)	(6)
LN(IVOL)	0.236***	1.036***	-1.478*	3.653***	-0.063	0.051**
	(3.4)	(20.4)	(-1.8)	(8.3)	(-1.0)	(2.5)
Size	-0.365***	-0.150***	2.001***	0.432***	0.140***	0.102***
	(-16.0)	(-11.0)	(4.9)	(4.4)	(5.8)	(14.3)
Cons.	-1.381***	0.153	63.125***	95.360***	11.559***	12.302***
	(-6.3)	(0.7)	(21.2)	(58.4)	(45.6)	(149.8)
Fixed Effects	Firm	Time	Firm	Time	Firm	Time
Observations	54634	55944	40858	42062	40360	41567

Table 4.3: Uncertainty and the Return Predictability of Analysts' Conditional Biases

This table presents the Fama-French Five-Factor alphas for double sort portfolios created by cross-sectionally sorting companies into terciles based on various uncertainty measures. Panel A shows results using MU, Panel B shows results using EPU, panel C shows results using OIV, and Panel D shows results using IVOL. The EHB Q1-Q5 portfolios based on OIV , and IVOL, are made by first cross-sectionally sorting companies into terciles based on each uncertainty measure then conditionally cross-sectionally sorting firms into quintiles based on EHB. The EHB Q1-Q5 portfolios based on MU, and EPU are made by sorting observations into terciles in the time series based on each uncertainty measure. Then firms are conditionally cross-sectionally sorted into quintiles based on EHB. Portfolios are value weighted and are re-balanced on a monthly basis. Q1 (Q5) contains firms with the lowest (highest) values of the EHB. T1 (T3) of each uncertainty tercile contains firms with the lowest (highest) values. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The sample period is from June 1990 to December 2019, with the exception of OIV which begins in January 1996.

Panel A: MU

	EHB Q1	EHB Q2	EHB Q3	EHB Q4	EHB Q5	EHB Q1-Q5
MU T1	0.234*	0.021	-0.151*	-0.065	-0.458**	0.692**
	(1.92)	(0.27)	(-1.91)	(-0.45)	(-2.13)	(2.25)
MU T2	0.149	0.060	-0.023	-0.100	-0.227	0.377
	(0.94)	(0.70)	(-0.19)	(-0.65)	(-0.89)	(1.01)
MU T3	0.165	-0.204*	0.012	0.042	0.147	0.018
	(0.85)	(-1.89)	(0.11)	(0.19)	(0.38)	(0.03)
MU T1-T3	0.069	0.224*	-0.163	-0.107	-0.605	0.674
	(0.30)	(1.69)	(-1.19)	(-0.41)	(-1.37)	(1.11)

Panel B: EPU

	EHB Q1	EHB Q2	EHB Q3	EHB Q4	EHB Q5	EHB Q1-Q5
EPU T1	0.567***	-0.073	-0.232**	-0.232	-0.690**	1.257***
	(3.73)	(-0.88)	(-2.27)	(-1.37)	(-2.50)	(3.26)
EPU T2	0.231*	-0.038	-0.026	-0.310**	-0.543**	0.774**
	(1.74)	(-0.44)	(-0.27)	(-2.03)	(-2.28)	(2.35)
EPU T3	0.100	0.066	0.020	0.009	-0.220	0.320
	(0.59)	(0.69)	(0.18)	(0.05)	(-0.66)	(0.70)
EPU T1-T3	0.467**	-0.139	-0.251*	-0.241	-0.470	0.937
	(2.04)	(-1.10)	(-1.69)	(-0.96)	(-1.09)	(1.58)

Panel C: OIV

	EHB Q1	EHB Q2	EHB Q3	EHB Q4	EHB Q5	EHB Q1-Q5
OIV T1	0.160	-0.009	-0.066	-0.063	-0.063	0.223
	(1.36)	(-0.10)	(-0.75)	(-0.63)	(-0.45)	(1.13)
OIV T2	0.359**	0.057	-0.176	-0.189	-0.316	0.675**
	(2.20)	(0.42)	(-1.08)	(-1.10)	(-1.32)	(2.09)
OIV T3	0.450*	0.217	-0.014	-0.415*	-1.002***	1.453***
	(1.74)	(1.03)	(-0.07)	(-1.67)	(-3.36)	(3.36)
OIV T1-T3	-0.290	-0.226	-0.052	0.352	0.940***	-1.230***
	(-1.00)	(-0.95)	(-0.22)	(1.32)	(3.45)	(-3.15)

Panel D: IVOL

	EHB Q1	EHB Q2	EHB Q3	EHB Q4	EHB Q5	EHB Q1-Q5
IVOL T1	0.095	0.017	-0.039	-0.046	-0.161	0.256
	(0.92)	(0.21)	(-0.51)	(-0.53)	(-1.21)	(1.30)
IVOL T2	0.335**	-0.028	-0.175	-0.196	-0.286	0.621**
	(2.50)	(-0.26)	(-1.40)	(-1.36)	(-1.38)	(2.21)
IVOL T3	0.641***	0.257	0.050	-0.410**	-0.917***	1.557***
	(3.16)	(1.57)	(0.30)	(-2.21)	(-3.93)	(4.71)
IVOL T1-T3	-0.545**	-0.240	-0.089	0.364*	0.755***	-1.301***
	(-2.51)	(-1.21)	(-0.45)	(1.77)	(3.40)	(-4.49)

Table 4.4: Variations in IVOL and the Return Predictability of Analysts' Conditional Biases

This table presents the Fama-French Five-Factor alphas for double sort portfolios created by cross-sectionally sorting companies into terciles based on the $IVOL_{MA36}$and abnormal IVOL. Then firms are conditionally cross-sectionally sorted into quintiles based on their EHB. Panel A shows results using $IVOL_{MA36}$ and panel B shows results using abnormal IVOL. Portfolios are value weighted and are re-balanced on a monthly basis. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The sample period is from June 1990 to December 2019.

Panel A: $IVOL_{MA36}$

	EHB Q1	EHB Q2	EHB Q3	EHB Q4	EHB Q5	EHB Q1-Q5
$IVOL_{MA36}$ T1	0.023	-0.032	-0.185**	0.001	-0.194	0.217
	(0.23)	(-0.41)	(-2.29)	(0.01)	(-1.20)	(0.98)
$IVOL_{MA36}$ T2	0.359**	0.069	-0.122	-0.162	-0.274	0.633**
	(2.56)	(0.60)	(-0.94)	(-1.09)	(-1.38)	(2.32)
$IVOL_{MA36}$ T3	0.702***	0.246	0.176	-0.246	-0.443**	1.145***
	(3.58)	(1.59)	(1.12)	(-1.35)	(-2.01)	(3.63)
$IVOL_{MA36}$ T1-T3	-0.679***	-0.278	-0.360*	0.247	0.249	-0.928***
	(-3.13)	(-1.48)	(-1.85)	(1.24)	(1.17)	(-3.30)

Panel B: Abnormal IVOL

	EHB Q1	EHB Q2	EHB Q3	EHB Q4	EHB Q5	EHB Q1-Q5
Abnormal IVOL T1	0.497***	0.332***	0.023	-0.023	-0.191	0.689***
	(3.51)	(3.08)	(0.22)	(-0.16)	(-1.17)	(2.91)
Abnormal IVOL T2	0.075	-0.018	-0.113	0.098	-0.298	0.373
	(0.60)	(-0.22)	(-1.20)	(0.78)	(-1.60)	(1.46)
Abnormal IVOL T3	-0.072	-0.126	-0.235*	-0.054	-0.355	0.282
	(-0.58)	(-1.15)	(-1.83)	(-0.29)	(-1.46)	(0.94)
Abnormal IVOL T1-T3	0.570***	0.458***	0.258	0.032	0.163	0.407
	(2.97)	(2.69)	(1.42)	(0.13)	(0.69)	(1.43)

Table 4.5: IC Index and the Return Predictability of EHB

This table presents the Fama-French Five-Factor alphas for double sort portfolios created by cross-sectionally sorting companies into terciles based on the IC index. Then firms are conditionally cross-sectionally sorted into quintiles based on their EHB. Portfolios are value weighted and are re-balanced on a monthly basis. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The sample period is from June 1996 to December 2019.

	EHB Q1	EHB Q2	EHB Q3	EHB Q4	EHB Q5	EHB Q1-Q5
IC Index T1	0.120	0.067	0.075	-0.133	-0.106	0.226
	(0.80)	(0.60)	(0.59)	(-0.82)	(-0.48)	(0.74)
IC Index T2	0.426***	0.027	0.032	0.008	-0.326	0.752**
	(2.72)	(0.25)	(0.30)	(0.05)	(-1.34)	(2.31)
IC Index T3	0.368**	-0.007	-0.083	-0.269	-0.686***	1.054***
	(2.36)	(-0.06)	(-0.63)	(-1.53)	(-2.59)	(2.85)
IC Index T1-T3	-0.248	0.073	0.158	0.137	0.580**	-0.828***
	(-1.26)	(0.44)	(0.89)	(0.66)	(2.51)	(-2.59)

Table 4.6: Information Scarcity and the Return Predictability of EHB

This table presents the results of pooled OLS monthly regressions of one-month-ahead returns (in percentage points) on the normalized rank of EHB (i.e., the rank scaled by the number of stocks in the cross-section) by -LN(Firm Age) Terciles. The terciles are generated by cross-sectionally sorting firms into terciles based on -LN(Firm Age). Columns 2, 4, 6, and 8 include Operating Profitability (Revenue minus cost - administrative expenses - interest expenses, scaled by book value of equity), Asset Growth (proxy for Investments), BTM, 6 month Momentum, and Size as controls. Controls are also the normalized rank (i.e., the rank scaled by the number of stocks in the cross-section) by tercile. In Panel A, High Media is constructed by first obtaining the difference between the current month's DJ news count and the 36 month rolling average. This normalized rank of this measure is then orthogonalized to size before finally being sorted into quintiles. The indicator equals one for observations in the top quintile and negative 1 for those in the bottom quintile. In Panel B, Earnings Announcement is an indicator equal to one for months when a firm has an earnings announcement. Standard errors of the resulting regression coefficient are clustered by firm and month. The regression includes time fixed effects. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The sample period is from December 2001 to December 2019.

Panel A: Abnormal Media

	T1		T2		T3		T3-T1	
EHB	0.002	-0.000	-0.003	-0.004	-0.008**	-0.008**	-0.010***	-0.007***
	(0.6)	(-0.0)	(-0.8)	(-1.2)	(-2.2)	(-2.1)	(-4.8)	(-3.1)
High Media	-0.059	-0.080	-0.124	-0.146	-0.268**	-0.276**	-0.203	-0.190
	(-0.7)	(-0.9)	(-1.4)	(-1.6)	(-2.2)	(-2.2)	(-1.5)	(-1.3)
EHB x High Media	-0.000	0.000	0.002	0.003	0.006***	0.007***	0.006**	0.006**
	(-0.1)	(0.1)	(1.4)	(1.6)	(2.8)	(2.8)	(2.4)	(2.3)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	123722	121292	123356	119475	99315	93298	223037	214590

Panel B: Earnings Announcements

	T1		T2		T3		T3-T1	
EHB	0.002	-0.001	-0.006*	-0.006*	-0.017***	-0.013***	-0.020***	-0.012***
	(0.7)	(-0.2)	(-1.8)	(-1.7)	(-4.8)	(-3.5)	(-8.1)	(-4.4)
Earn. Annc.	-0.005	-0.022	-0.392*	-0.359	-0.274	-0.294	-0.289	-0.252
	(-0.0)	(-0.1)	(-1.8)	(-1.6)	(-1.2)	(-1.2)	(-0.9)	(-1.0)
EHB x Earn. Annc.	-0.000	-0.000	0.006	0.006	0.008*	0.009**	0.008**	0.009***
	(-0.1)	(-0.0)	(1.6)	(1.6)	(1.9)	(2.0)	(2.4)	(2.6)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	227478	223337	218208	211126	209125	151213	436603	374550

Table 4.7: Firm Size and the Return Predictability of EHB

This presents the Fama-French Five-Factor alphas for double sort portfolios created by cross-sectionally sorting companies into groups based on their market value of equity. Then firms are conditionally cross-sectionally sorted into quintiles based on their EHB. The market value of equity groups divide the firms into mega-cap, large-cap, and small-cap groups. Mega-cap firms are defined as firms with market capitalization greater than the 80th percentile of firm sizes on the NYSE. The remaining firms are then defined as small- or large-cap based on whether their size is above the median NYSE market capitalization. Portfolios are value weighted and are re-balanced on a monthly basis. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. Q1 indicates the lowest values and Q5 the highest values for EHB. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. Values are shown in percentage terms The sample period for Panel A is from June 1990 to December 2019. The sample period for Panels B is from June 1996 to December 2019.

	EHB Q1	EHB Q2	EHB Q3	EHB Q4	EHB Q5	EHB Q1-Q5
Small Cap	0.288***	0.225***	-0.015	-0.031	-0.681***	0.968***
	(3.26)	(3.14)	(-0.22)	(-0.34)	(-4.43)	(4.70)
Large Cap	0.319***	0.139*	-0.075	-0.089	-0.381**	0.700***
	(2.78)	(1.80)	(-1.00)	(-0.91)	(-2.27)	(2.95)
Mega Cap	0.214*	0.055	-0.106*	-0.062	-0.094	0.308
	(1.93)	(0.68)	(-1.67)	(-0.80)	(-0.66)	(1.36)
Small-Mega	0.074	0.170*	0.091	0.031	-0.587***	0.660***
	(0.66)	(1.73)	(0.98)	(0.29)	(-4.78)	(3.70)

Table 4.8: Testing Alternative Theories

This table presents the results of monthly regressions of |EHB|, EDGAR Downloads, or Effective Spread on the measures of uncertainty. The |EHB| and Effective Spread analysis use pooled OLS regressions. As EDGAR Downloads are a count measure, a Pseudo Poisson regression is used instead of a pooled OLS regression. Panel A uses IVOL and EPU as the cross-sectional and time-series uncertainty measures while Panel B uses the forward looking OIV and MU. Columns 1, 3, and 5 include only Size as a control and columns 2, 4, and 6 add Firm Age and an indicator equal to one in earnings announcement months. Standard errors are clustered at the firm and month level. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The sample period for |EHB| is from June 1990 to December 2019 for Panel A and January 1996 to December 2019 for Panel B as OIV does not have observations prior to this. The sample for EDGAR Downloads begins in April 2003 and ends in December 2017. The sample for Effective Spread begins in September 2003 and ends in December 2019.

Panel A: IVOL and EPU

	$\|EHB\|$		EDGAR Downloads		Eff. Spread	
Size	-0.167***	-0.191***	0.621***	0.631***	-3.324***	-3.249***
	(-12.0)	(-13.1)	(9.0)	(8.2)	(-22.2)	(-21.0)
EPU	0.002***	0.002***	0.004***	0.004***	-0.003	-0.003
	(3.9)	(3.7)	(4.7)	(4.7)	(-0.7)	(-0.5)
LN(IVOL)	0.652***	0.775***	0.514***	0.467***	4.726***	4.321***
	(17.8)	(18.4)	(6.0)	(8.2)	(7.4)	(6.5)
LN(Age)		0.152***		-0.073		-0.560***
		(11.4)		(-1.2)		(-3.3)
Earn. Annc.		0.053***		0.106**		0.128
		(2.8)		(2.3)		(0.4)
Observations	626870	626870	284143	284143	326669	326669

Panel B: OIV and MU

	$\lvert EHB \rvert$		EDGAR Downloads		Eff. Spread	
Size	-0.097***	-0.115***	0.622***	0.633***	-2.761***	-2.680***
	(-8.6)	(-9.8)	(8.9)	(8.2)	(-18.3)	(-17.2)
MU	1.554***	1.414***	-2.572***	-2.512***	5.831***	6.387***
	(4.6)	(4.3)	(-5.7)	(-6.0)	(2.6)	(2.9)
LN(OIV)	0.873***	0.970***	0.698***	0.669***	5.151***	4.803***
	(19.0)	(19.1)	(8.3)	(11.4)	(10.3)	(9.5)
LN(Age)		0.122***		-0.063		-0.523***
		(9.7)		(-1.1)		(-3.4)
Earn. Annc.		0.092***		0.144***		0.149
		(7.6)		(3.3)		(0.6)
Observations	447112	447112	264813	264813	306895	306895

4.9 Appendix A: EHB Construction

4.9.1 Input Dataset Construction

To generate our composite EHB measure, we first generate forecasts for the next quarter (FQ), one year ahead (FY1), and two years ahead (FY2) earnings using machine learning.[23] The tables below show the variables used in generating the machine learning forecasts. We utilize the methodology in J. L. Campbell et al., 2023 to generate the EHB forecasts using their suggested best specification. Please refer to their paper for a more detailed description of the data generation process.

We apply the same key filters used in J. L. Campbell et al., 2023. Specifically, we require returns, market capitalization, the two momentum variables, the current analysts' forecast, the most recently realized earnings, the stock price, and price-to-sales to be non-missing.[24] Since analysts' forecasts contain private information that adds incremental predictive power for earnings relative to financial statement variables de Silva and Thesmar, 2022; van Binsbergen et al., 2022, we also include the following analysts' forecasts related variables in our predictor set as shown in Table 4.10.

[23] We include FQ forecasts as our results are robust to including them in the EHB calculation.

[24] The requirement of non-missing current analysts' forecast, the most recently realized earnings, stock price, and price-to-sales follows from Bradshaw et al., 2012

Table 4.9: WRDS Financial Ratio Variables

This table provides the definitions of WRDS Financial Ratio Variables. Following van Binsbergen et al., 2022, we exclude Forward P/E to 1-year Growth (PEG) ratio, Forward P/E to Long-term Growth (PEG) ratio, Price/Operating Earnings (Basic, Excl. Extraordinary Income), and Price/Operating Earnings (Diluted, Excl. Extraordinary Income) from the WRDS Financial Suite Ratios due to the large number of missing observations.

Acronym	Definition	Acronym	Definition
accrual	Accruals/Average Assets	int_totdebt	Interest/Average Total Debt
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 Cash Flow/Current Liabilities
capei	Shiller's 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/Total Equity	profit_lct	Profit Before Depreciation/Current Liabilities
debt_assets	Total Debt/Total Assets	ps	Price/Sales
debt_at	Total Debt/Total Assets	ptb	Price/Book
debt_capital	Total Debt/Total Capital	ptpm	Pre-Tax Profit margin
debt_ebitda	Total Debt/EBITDA	quick_ratio	Quick Ratio
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
intcov	After-tax Interest Coverage	totdebt_invcap	Total Debt/Invested Capital
intcov_ratio	Interest Coverage Ratio		

Table 4.10: Other Variables

This table provides the definitions of the variables in our predictor set that are not included in the WRDS Financial Ratio Variables.

Acronym	Definition
EPS (FY2/FQ)	Realized Earnings per Share
ErrAF (FY2/FQ)	Realized EPS-Analysts' forecast as of current month
medest2	Analysts' consensus forecast for FY2 horizon
medestqtr	Analysts' consensus forecast for FQ horizon
ibes_earnings_ann	Most recently realized annual earnings as of current month
ibes_earnings_qtr	Most recently realized quarterly earnings as of current month
last_F2ana_fe_med	Most recently realized FY2 horizon analysts' forecast error as of current month
last_Fqtrana_fe_med	Most recently realized FQ horizon analysts' forecast error as of current month
rev_FY2_3m	Revision of analysts' FY2 horizon forecast between current month and 3 months prior
rev_FYqtr_3m	Revision of analysts' FQ horizon forecast between current month and 3 months prior
dist2	Distance between FY2 fiscal period end and current month
distqtr	Distance between FQ fiscal period end and current month
ret	Stock Return
prc	Stock Price
size	LN(Market Capitalization)
mom6m	6 month momentum
indmom	Industry weighted 6 month momentum

4.9.2 ML Forecast Timing

We construct our train and test datasets carefully to ensure no data leakage. At the end of each month t, for each stock i, and for a specific forecast horizon τ, we construct the earnings prediction. The target variable of interest is the analysts' quarterly, one-year or two-year ahead forecast error (i.e., the realized errors of analysts' forecasts made at month t).[25]

In our training set we make sure that both the target variable and the predictors in the train dataset are known at month t. Specifically, that means the realized earnings used in constructing the target variable are known/announced before or during month t. After we fit the model, including selecting the optimal hyper-parameters, we use the fitted model to generate earnings predictions at month t.

4.9.3 Machine Learning Methodology

We use the gradient boosted decision tree model implemented using LightGBM (LGBM), a popular, off-the-shelf machine learning algorithm, as our main statistical forecasting model, as it provides the best outcome for predicting future earnings J. L. Campbell et al., 2023. LGBM is a non-linear non-parametric ensemble model which combines the predictions of many decision trees. Trees are grown in sequential manner to correct the prediction error from the previous iteration, which is known as boosting (Friedman, 2001). The weighted average of these individual tree models is the final predictor.

Our forecasts begin in June 1990 to allow for enough data to be available at the time of the first forecast to train both the model hyper-parameters, and model parameters. We train our model hyper-parameters using cross-validation, which splits the data in the training window into subsets (creating a training subset and validation subset). Then various combinations of the hyperparameters are tested to identify the best combination of hyperparamters. The ML model is then fit to the data using the selected hyperparameters and forecasts are made using out-of-sample data to ensure no look ahead bias.

[25] For analysts' forecasts from I/B/E/S, we use the consensus median forecasts as of the latest IBES statistical period, STATPERS, before the end of month t.

4.10 Appendix B: Robustness Tables

Figure 4.9: Correlation Matrix of Information Cost Index Components and Size

This figure shows the Spearman correlations for the components of the Composite Information Cost Index and the components of IVOL. As the Information Cost Index consists of measures that update infrequently (the Bog Index, Complexity, and Net File Size update annually and Firm Age is slow moving), the analysis is done as of the end of June in each year. All variables use the normalized rank (i.e., the rank scaled by the number of stocks in the cross-section). The sample period for Firm Age and Size begins in June 1990, the annual sample period for the Bog Index and Net File Size begins in June 1996. All samples end in December 2019.

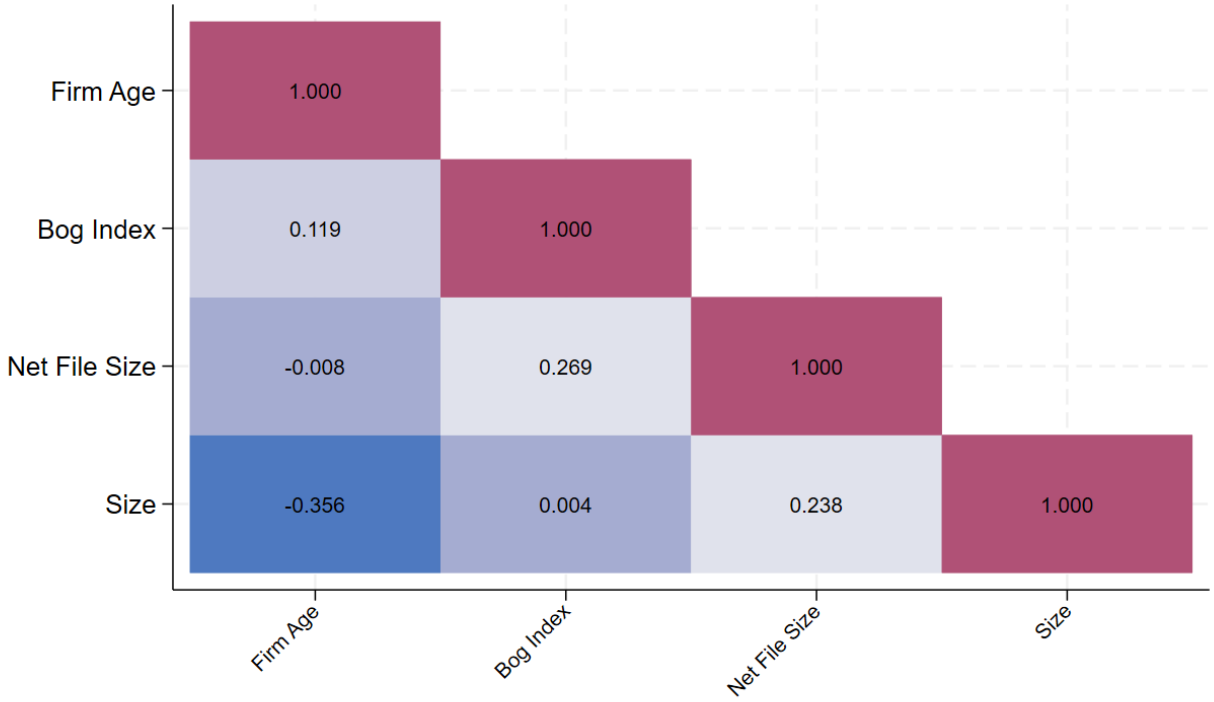


Figure 4.10: Return Predictability of EHB by Uncertainty Terciles: Non-Mega Cap

This figure shows the Fama-French Five-Factor alphas of the EHB Q1-Q5 portfolios by uncertainty terciles using six monthly uncertainty measures. This figure uses only firms in the non-mega cap subsample. The EHB Q1-Q5 portfolios based on OIV , IVOL, IVOL$_{MA36}$, and abnormal IVOL, are made by first cross-sectionally sorting companies into terciles based on each uncertainty measure. Then firms are conditionally cross-sectionally sorted into quintiles based on EHB. The EHB Q1-Q5 portfolios based on MU, and EPU are made by sorting observations into terciles in the time series based on each uncertainty measure. Then we conditionally cross-sectionally sort firms into quintiles based on EHB. Portfolios are value weighted and are re-balanced on a monthly basis. Q1 (Q5) contains firms with the lowest (highest) values of EHB. T1 (T3) of each uncertainty tercile contains firms with the lowest (highest) values. The average value and test statistics are presented below their corresponding bars. Whiskers denote 95% confidence bands. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. The sample period is from June 1990 to December 2019, with the exception of OIV which begins in January 1996.

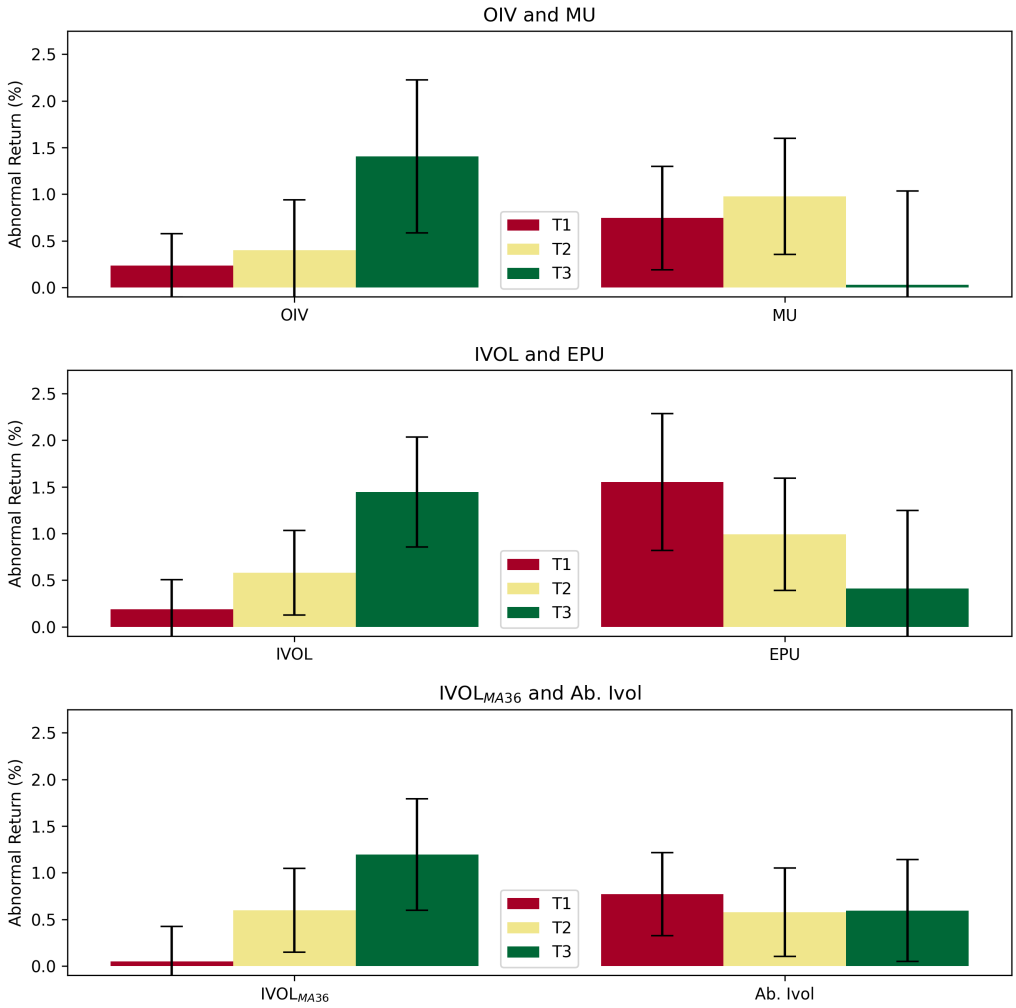


Figure 4.11: Return Predictability of EHB by Uncertainty Terciles: Mega Cap

This figure shows the Fama-French Five-Factor alphas of the EHB Q1-Q5 portfolios by uncertainty terciles using six monthly uncertainty measures. This figure uses only firms in the mega cap subsample. The EHB Q1-Q5 portfolios based on OIV , IVOL, IVOL$_{MA36}$, and abnormal IVOL, are made by first cross-sectionally sorting companies into terciles based on each uncertainty measure. Then firms are conditionally cross-sectionally sorted into quintiles based on EHB. The EHB Q1-Q5 portfolios based on MU, and EPU are made by sorting observations into terciles in the time series based on each uncertainty measure. Then we conditionally cross-sectionally sort firms into quintiles based on EHB. Portfolios are value weighted and are re-balanced on a monthly basis. Q1 (Q5) contains firms with the lowest (highest) values of EHB. T1 (T3) of each uncertainty tercile contains firms with the lowest (highest) values. The average value and test statistics are presented below their corresponding bars. Whiskers denote 95% confidence bands. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. The sample period is from June 1990 to December 2019, with the exception of OIV which begins in January 1996.

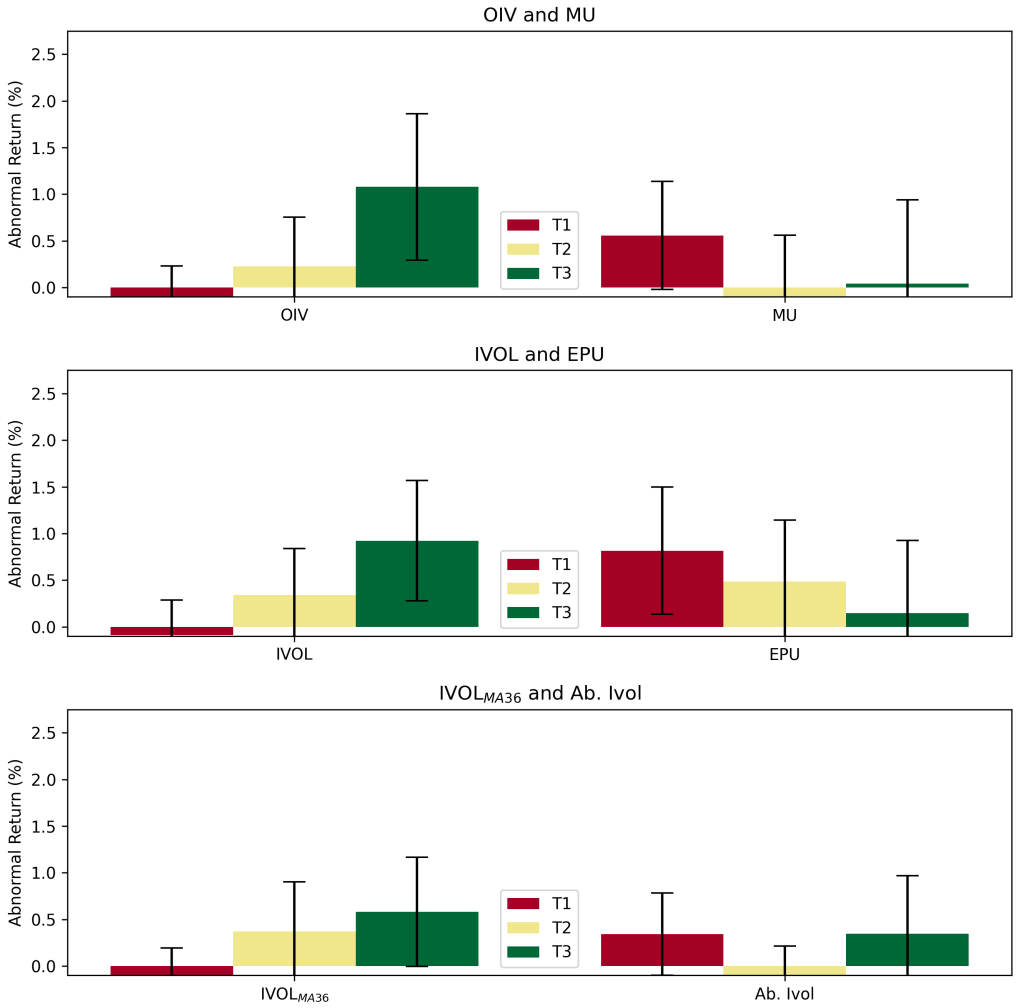


Figure 4.12: Return Predictability of Announcement Return by Uncertainty Terciles - Mega Cap

This figure shows the Fama-French Five-Factor alphas of the Announcement Return Q1-Q5 portfolios by uncertainty terciles using six monthly uncertainty measures. This figure includes only firms in the Mega-Cap subsample. The Announcement Return Q1-Q5 portfolios based on OIV , IVOL, IVOL$_{MA36}$, and abnormal IVOL, are made by first cross-sectionally sorting companies into terciles based on each uncertainty measure. Then firms are conditionally cross-sectionally sorted into quintiles based on their announcement return. The Announcement Return Q1-Q5 portfolios based on MU, and EPU are made by sorting observations into terciles in the time series based on each uncertainty measure. Then we conditionally cross-sectionally sort firms into quintiles based on their Announcement Return. Portfolios are value weighted and are re-balanced on a monthly basis. Q1 (Q5) contains firms with the lowest (highest) values of the Announcement Return. T1 (T3) of each uncertainty tercile contains firms with the lowest (highest) values. The average value and test statistics are presented below their corresponding bars. Whiskers denote 95% confidence bands. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. The sample period is from June 1990 to December 2019, with the exception of OIV which begins in January 1996.

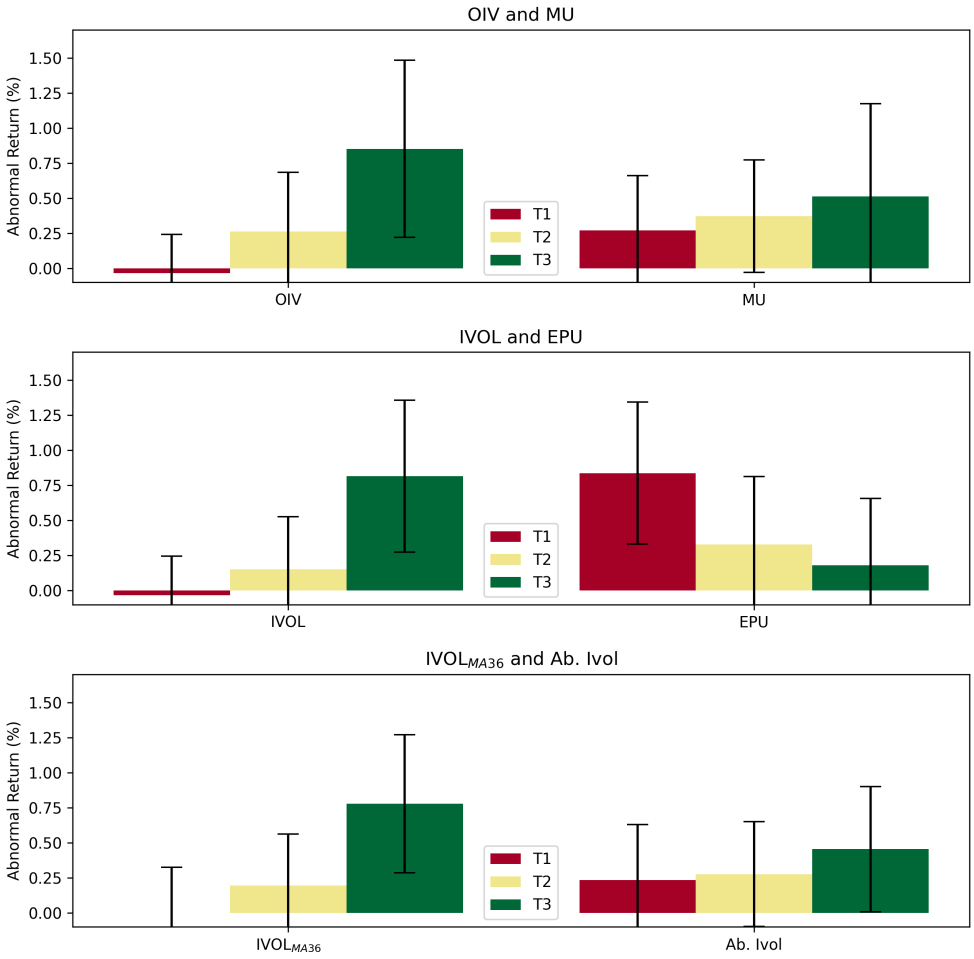


Figure 4.13: Return Predictability of Analysts' Revisions by Uncertainty Terciles - Mega Cap

This figure shows the Fama-French Five-Factor alphas of the analysts' revision Q1-Q5 portfolios by uncertainty terciles using six monthly uncertainty measures. This figure includes only firms in the Mega-Cap subsample. The analysts' revision Q1-Q5 portfolios based on OIV , IVOL, IVOL$_{MA36}$, and abnormal IVOL, are made by first cross-sectionally sorting companies into terciles based on each uncertainty measure. Then firms are conditionally cross-sectionally sorted into quintiles based on the analysts' revision. The analysts' revision Q1-Q5 portfolios based on MU, and EPU are made by sorting observations into terciles in the time series based on each uncertainty measure. Then we conditionally cross-sectionally sort firms into quintiles based on their analysts' revision. Portfolios are value weighted and are re-balanced on a monthly basis. Q1 (Q5) contains firms with the lowest (highest) values of the analysts' revision. T1 (T3) of each uncertainty tercile contains firms with the lowest (highest) values. The average value and test statistics are in presented below their corresponding bars. Whiskers denote 95% confidence bands. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. The sample period is from June 1990 to December 2019, with the exception of OIV which begins in January 1996.

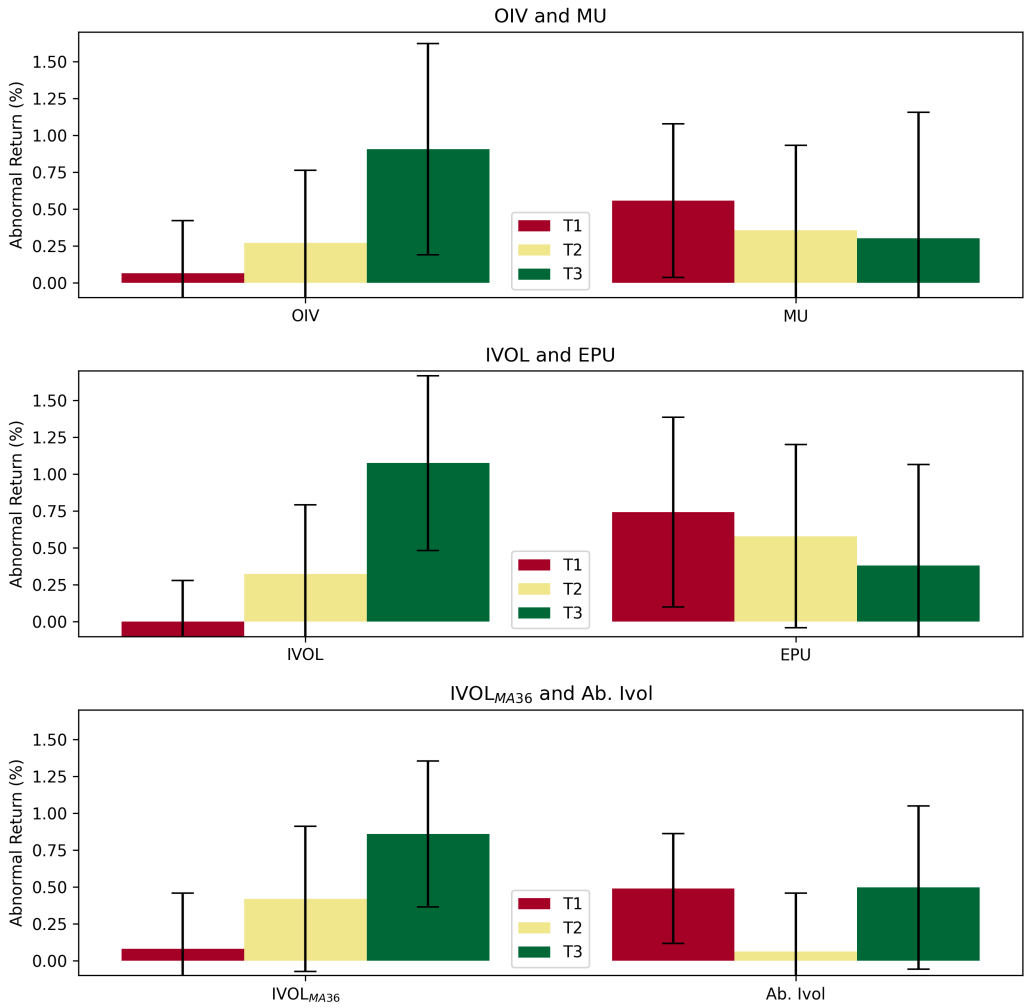


Table 4.11: Information Cost Index Component's Persistence

This table presents the results of panel regressions of the Bog Index, log Net File Size, and log Firm Age on their one year lagged values. We use the values at the end of June in each year. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. Standard Errors are clustered by the firm and year. The sample period for Firm Age begins in June 1990, the sample period for the Bog Index and Net File Size begins in June 1996. All samples end in December 2019.

	(1) -LN(Firm Age)	(2) Bog Index	(3) LN(Net File Size)
-LN(Firm Age)(t-1)	0.847***		
	(223.9)		
Bog Index(t-1)		0.918***	
		(43.5)	
LN(Net File Size) (t-1)			0.644***
			(37.8)
Cons.	-0.910***	7.525***	4.617***
	(-46.2)	(4.5)	(21.4)
Observations	47327	34171	33599

Table 4.12: IC Index Without Size Residualization and the Return Predictability of EHB

This table presents the Fama-French Five-Factor alphas for double sort portfolios created by cross-sectionally sorting companies into terciles based on the IC index (without residualizing the components to size). Then firms are conditionally cross-sectionally sorted into quintiles based on their EHB. Portfolios are value weighted and are re-balanced on a monthly basis. Standard errors of the resulting regression coefficients are computed based on Newey and West, 1987 with 12 lags. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The sample period is from June 1996 to December 2019.

	EHB Q1	EHB Q2	EHB Q3	EHB Q4	EHB Q5	EHB Q1-Q5
IC Index (Not Orth.) T1	0.154	0.054	0.110	-0.007	-0.039	0.193
	(1.04)	(0.48)	(0.91)	(-0.04)	(-0.18)	(0.63)
IC Index (Not Orth.) T2	0.475***	0.025	-0.138	-0.026	-0.371	0.846**
	(2.98)	(0.23)	(-1.20)	(-0.15)	(-1.53)	(2.53)
IC Index (Not Orth.) T3	0.353**	-0.052	0.004	-0.282	-0.665**	1.018***
	(2.22)	(-0.50)	(0.03)	(-1.61)	(-2.48)	(2.73)
IC Index (Not Orth.) T1-T3	-0.199	0.105	0.106	0.275	0.625***	-0.825***
	(-1.04)	(0.65)	(0.60)	(1.34)	(2.73)	(-2.62)

Table 4.13: OIV and Information Cost

This table presents the results of pooled OLS regression of LN(OIV) on the components of the Information Cost Index (Firm Age, the Bog Index, and Net File Size) and Size. As the Information Cost Index consists of measures that update infrequently (the Bog Index and Net File Size update annually and Firm Age is slow moving), the regression is run as of the end of June in each year. Columns 1 and 2 use -LN(Firm Age), columns 3 and 4 use the Bog Index, and columns 5 and 6 use LN(Net File Size) Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. Standard errors are clustered at the firm and year level. Columns 1, 3, and 5 include firm fixed effects while columns 2, 4, and 6 include time fixed effects. The annual sample period for Firm Age begins in June 1990, the sample period for the Bog Index and Net File Size begins in June 1996. All samples end in December 2019.

	(1)	(2)	(3)	(4)	(5)	(6)
-LN(Firm Age)	0.103***	0.100***				
	(2.9)	(12.4)				
Bog Index			-0.004	0.007***		
			(-1.4)	(6.8)		
LN(Net File Size)					-0.020	0.021*
					(-0.6)	(1.8)
Size	-0.175***	-0.113***	-0.206***	-0.137***	-0.209***	-0.143***
	(-7.3)	(-18.1)	(-9.7)	(-25.1)	(-10.0)	(-24.6)
Cons.	0.900***	0.394***	0.987***	-0.546***	0.899**	-0.155
	(4.4)	(7.9)	(3.4)	(-5.7)	(2.2)	(-1.0)
Fixed Effects	Firm	Time	Firm	Time	Firm	Time
Observations	37613	38553	35399	36339	34950	35870

Table 4.14: Information Scarcity, Information Flows, and the Return Predictability of Analysts' Conditional Biases

This table presents the results of pooled OLS monthly regressions of one-month-ahead returns (in percentage points) on the normalized rank of EHB (i.e., the rank scaled by the number of stocks in the cross-section) by -LN(Firm Age) orthogonalized to Size Terciles. The terciles are generated by cross-sectionally sorting firms into terciles based on first regressing the normalized rank of -LN(Firm Age) on the normalized rank of Size and then sorting into terciles. Columns 2, 4, 6, and 8 include Operating Profitability (Revenue minus cost - administrative expenses - interest expenses, scaled by book value of equity), Asset Growth (proxy for Investments), BTM, 6 month Momentum, and Size as controls. Controls are also the normalized rank (i.e., the rank scaled by the number of stocks in the cross-section) by tercile. In Panel A, High Media is constructed by first obtaining the difference between the current month's DJ news count and the 36 month rolling average. This normalized rank of this measure is then orthogonalized to size before finally being sorted into quintiles. The indicator equals one for observations in the top quintile and negative 1 for those in the bottom quintile. In Panel B, Earnings Announcement is an indicator equal to one for months when a firm has an earnings announcement. Standard errors of the resulting regression coefficient are clustered by firm and month. The regression includes time fixed effects. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively.

Panel A: Abnormal Media

	T1		T2		T3		T3-T1	
EHB	0.002	0.001	-0.004	-0.006**	-0.007*	-0.005	-0.009***	-0.006***
	(0.6)	(0.2)	(-1.3)	(-2.0)	(-1.8)	(-1.6)	(-4.5)	(-2.7)
High Media	-0.105	-0.141	-0.271**	-0.269**	-0.149	-0.195	-0.045	-0.053
	(-1.1)	(-1.5)	(-2.5)	(-2.4)	(-1.2)	(-1.5)	(-0.3)	(-0.3)
EHB x High Media	-0.000	0.000	0.005**	0.005**	0.005**	0.006**	0.005*	0.006*
	(-0.1)	(0.3)	(2.4)	(2.3)	(2.2)	(2.4)	(1.9)	(1.9)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	123490	120991	123073	119265	99830	93809	223320	214800

Panel B: Earnings Announcements

	T1		T2		T3		T3-T1	
EHB	0.002	-0.001	-0.007*	-0.007*	-0.017***	-0.012***	-0.019***	-0.011***
	(0.6)	(-0.3)	(-1.9)	(-1.9)	(-4.6)	(-3.4)	(-7.7)	(-4.3)
Earn. Annc.	-0.045	-0.060	-0.264	-0.198	-0.376	-0.504**	-0.422	-0.490*
	(-0.2)	(-0.3)	(-1.1)	(-0.8)	(-1.6)	(-2.2)	(-1.2)	(-1.8)
EHB x Earn. Annc.	0.001	0.001	0.005	0.004	0.009**	0.011***	0.007**	0.009***
	(0.3)	(0.4)	(1.1)	(0.9)	(2.1)	(2.7)	(2.1)	(2.8)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	222582	218261	221398	214441	210831	152974	433413	371235

Table 4.15: Testing Alternative Theories using AIA

This table presents the results of probit daily regressions of AIA on the measures of uncertainty. AIA is an indicator variable equal to 1 when Bloomberg News Heat-Daily Max Readership Measure is 3-4 and 0 otherwise. Columns 1 and 3 include Size as a control and columns 2 and 4 add Firm Age and an indicator equal to one in earnings announcement months. All dependent variables are measured at the monthly level. Standard errors are clustered at the firm and day level. Statistical significance is denoted as ***, **, and * for p<0.10, p<0.05, and p<0.01, respectively. The sample period is from March 2010 to December 2019.

	(1)	(2)	(3)	(4)
Size	0.283***	0.284***	0.278***	0.288***
	(54.5)	(53.4)	(50.7)	(51.5)
EPU	-0.001***	-0.000***		
	(-3.8)	(-3.0)		
LN(IVOL)	0.375***	0.375***		
	(26.2)	(24.7)		
LN(Age)		-0.001		-0.013**
		(-0.2)		(-2.3)
Earn. Annc.		0.212***		0.246***
		(21.7)		(25.3)
MU			-1.872***	-1.863***
			(-10.6)	(-10.5)
LN(OIV)			0.366***	0.399***
			(22.2)	(23.1)
Observations	4089977	4089977	4014906	4014906

Bibliography

Abarbanell, J., & Lehavy, R. (2003). Can Stock Recommendations Predict Earnings Management and Analysts Earnings Forecast Errors? *Journal of Accounting Research*, *41*(1), 1–31. https://doi.org/10.1111/1475-679X.00093

Abarbanell, J. S., & Bernard, V. L. (1992). Tests of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior. *The Journal of Finance*, *47*(3), 1181–1207. https://doi.org/10.1111/j.1540-6261.1992.tb04010.x

Ali, A., Hwang, L.-S., & Trombley, M. A. (2003). Residual Income Based Valuation Predicts Future Stock Returns: Evidence on Mispricing vs. Risk Explanations. *The Accounting Review*, *78*(2), 377–396. https://doi.org/10.2308/accr.2003.78.2.377

Altavilla, C., Giacomini, R., & Ragusa, G. (2017). Anchoring the yield curve using survey expectations. *Journal of Applied Econometrics*, *32*(6), 1055–1068. https://doi.org/10.1002/jae.2588

Amir, E., Lev, B., & Sougiannis, T. (2003). Do financial analysts get intangibles? *European Accounting Review*, *12*(4), 635–659. https://doi.org/10.1080/0963818032000141879

Andrei, D., Friedman, H., & Ozel, N. B. (2023). Economic uncertainty and investor attention. *Journal of Financial Economics*, *149*(2), 179–217. https://doi.org/10.1016/j.jfineco.2023.05.003

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, *61*(1), 259–299.

Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, *50*(4), 745–787. https://doi.org/10.1016/s0304-3932(03)00032-1

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Baley, I., & Veldkamp, L. (2021, July). Bayesian Learning.

Ball, R., & Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers [Publisher: [Accounting Research Center, Booth School of Business, University of Chicago, Wiley]]. *Journal of Accounting Research*, *6*(2), 159–178. https://doi.org/10.2307/2490232

Ball, R. T., & Ghysels, E. (2018). Automated Earnings Forecasts: Beat Analysts or Combine and Conquer? *Management Science*, *64*(10), 4936–4952. https://doi.org/10.1287/mnsc.2017.2864

Basu, S., & Markov, S. (2004). Loss function assumptions in rational expectations tests on financial analysts earnings forecasts. *Journal of Accounting and Economics*, *38*, 171–203. https://doi.org/10.1016/j.jacceco.2004.09.001

Bauer, M. D., & Hamilton, J. D. (2017). Robust Bond Risk Premia. *The Review of Financial Studies*, *31*(2), 399–448. https://doi.org/10.1093/rfs/hhx096

Bauer, M. D., & Rudebusch, G. D. (2020). Interest rates under falling stars. *American Economic Review*, *110*(5), 1316–54. https://doi.org/10.1257/aer.20171822

Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics*, *97*, 71–87. https://doi.org/10.1016/j.jmoneco.2018.05.013

Benamar, H., Foucault, T., & Vega, C. (2021). Demand for Information, Uncertainty, and the Response of U.S. Treasury Securities to News. *The Review of Financial Studies*, *34*(7), 3403–3455. https://doi.org/10.1093/rfs/hhaa072

Bernard, V. L., & Thomas, J. K. (1989). Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium? *Journal of Accounting Research*, *27*, 1–36. https://doi.org/10.2307/2491062

Bernard, V. L., & Thomas, J. K. (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics*, *13*(4), 305–340. https://doi.org/10.1016/0165-4101(90)90008-R

Bertomeu, J. (2020). Machine learning improves accounting: Discussion, implementation and research opportunities. *Review of Accounting Studies*, *25*(3), 1135–1155. https://doi.org/10.1007/s11142-020-09554-9

Bertomeu, J., Cheynel, E., Floyd, E., & Pan, W. (2021). Using machine learning to detect misstatements. *Review of Accounting Studies*, *26*(2), 468–519. https://doi.org/10.1007/s11142-020-09563-8

Bianchi, D., Büchner, M., Hoogteijling, T., & Tamoni, A. (2020). Corrigendum: Bond Risk Premiums with Machine Learning. *The Review of Financial Studies*, *34*(2), 1090–1103. https://doi.org/10.1093/rfs/hhaa098

Bianchi, D., Büchner, M., & Tamoni, A. (2020). Bond Risk Premiums with Machine Learning. *The Review of Financial Studies*, *34*(2), 1046–1089. https://doi.org/10.1093/rfs/hhaa062

Blankespoor, E., deHaan, E., & Marinovic, I. (2020). Disclosure processing costs, investors information choice, and equity market outcomes: A review. *Journal of Accounting and Economics*, *70*(2-3), 101344. https://doi.org/10.1016/j.jacceco.2020.101344

Blankespoor, E., Dehaan, E., Wertz, J., & Zhu, C. (2019). Why Do Individual Investors Disregard Accounting Information? The Roles of Information Awareness and Acquisition Costs. *Journal of Accounting Research*, *57*(1), 53–84. https://doi.org/10.1111/1475-679X.12248

Bonsall, S. B., Green, J., & Muller, K. A. (2020). Market uncertainty and the importance of media coverage at earnings announcements. *Journal of Accounting and Economics*, *69*(1), 101264. https://doi.org/10.1016/j.jacceco.2019.101264

Bonsall, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics*, *63*(2), 329–357. https://doi.org/10.1016/j.jacceco.2017.03.002

Bordalo, P., Gennaioli, N., Porta, R. L., & Shleifer, A. (2019). Diagnostic Expectations and Stock Returns. *The Journal of Finance*. https://doi.org/10.1111/jofi.12833

Bradshaw, M. T., Drake, M. S., Myers, J. N., & Myers, L. A. (2012). A re-examination of analysts' superiority over time-series forecasts of annual earnings. *Review of Accounting Studies*, *17*(4), 944–968. https://doi.org/10.1007/s11142-012-9185-8

Bradshaw, M. T., Lee, L. F., & Peterson, K. (2016). The Interactive Role of Difficulty and Incentives in Explaining the Annual Earnings Forecast Walkdown [Publisher: American Accounting Association]. *The Accounting Review*, *91*(4), 995–1021.

Burgstahler, D. C., & Eames, M. J. (2003). Earnings Management to Avoid Losses and Earnings Decreases: Are Analysts Fooled? *Contemporary Accounting Research*, *20*(2), 253–294. https://doi.org/10.1506/BXXP-RGTD-H0PM-9XAL

Call, A. C., Hewitt, M., Watkins, J., & Yohn, T. L. (2021). Analysts' annual earnings forecasts and changes to the I/B/E/S database. *Review of Accounting Studies*, *26*(1), 1–36. https://doi.org/10.1007/s11142-020-09560-x

Campbell, J. L., Ham, H., Lu, Z., & Wood, K. (2023, June). Expectations Matter: When (not) to Use Machine Learning Earnings Forecasts. https://doi.org/10.2139/ssrn.4495297

Campbell, J. Y. (2017). *Financial decisions and markets: A course in asset pricing*. Princeton University Press.

Cao, K., & You, H. (2020, November). Fundamental Analysis Via Machine Learning. https://doi.org/10.2139/ssrn.3706532

Cao, S. S., Jiang, W., Wang, J. L., & Yang, B. (2021). From Man vs. Machine to Man + Machine: The Art and AI of Stock Analyses. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3840538

Chattopadhyay, A., Fang, B., & Mohanram, P. (2023). Machine Learning, Earnings Forecasting, and Implied Cost of Capital - US and International Evidence. *Working Paper*.

Chen, A. Y., & Zimmermann, T. (2022). Open Source Cross-Sectional Asset Pricing. *Critical Finance Review*, 69.

Chen, D., Ma, Y., Martin, X., & Michaely, R. (2022). On the fast track: Information acquisition costs and information production. *Journal of Financial Economics*, *143*(2), 794–823. https://doi.org/10.1016/j.jfineco.2021.06.025

Chen, X., He, W., Tao, L., & Yu, J. (2022). Attention and Underreaction-Related Anomalies. *Management Science*, mnsc.2022.4332. https://doi.org/10.1287/mnsc.2022.4332

Cieslak, A., & Povala, P. (2015). Expected Returns in Treasury Bonds. *Review of Financial Studies*, *28*(10), 2859–2901. https://doi.org/10.1093/rfs/hhv032

Claus, J., & Thomas, J. (2001). Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets. *The Journal of Finance*, *56*(5), 1629–1666. https://doi.org/10.1111/0022-1082.00384

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, *53*(2), 385–407.

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. *The Journal of Finance*, *52*(3), 1035–1058. https://doi.org/10.2307/2329515

Daniel, K., Hirshleifer, D., & Sun, L. (2020). Short- and Long-Horizon Behavioral Factors (L. Cohen, Ed.). *The Review of Financial Studies*, *33*(4), 1673–1736. https://doi.org/10.1093/rfs/hhz069

Das, S., Levine, C. B., & Sivaramakrishnan, K. (1998). Earnings Predictability and Bias in Analysts' Earnings Forecasts [Publisher: American Accounting Association]. *The Accounting Review*, *73*(2), 277–294.

Dávila, E., & Parlatore, C. (2023). Volatility and informativeness. *Journal of Financial Economics*, *147*(3), 550–572. https://doi.org/10.1016/j.jfineco.2022.12.005

Dechow, P. M., & Dichev, I. D. (2002). The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors [Publisher: American Accounting Association]. *The Accounting Review*, *77*, 35–59.

Dellavigna, S., & Pollet, J. M. (2009). Investor Inattention and Friday Earnings Announcements. *The Journal of Finance*, *64*(2), 709–749. https://doi.org/10.1111/j.1540-6261.2009.01447.x

de Silva, T., & Thesmar, D. (2022, October). Noise in Expectations: Evidence from Analyst Forecasts. https://doi.org/10.2139/ssrn.3762348

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, *130*(2), 337–364. https://doi.org/10.1016/j.jeconom.2005.03.005

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy [Publisher: Taylor & Francis]. *Journal of Business & Economic Statistics*, *13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Dubinsky, A., Johannes, M., Kaeck, A., & Seeger, N. J. (2019). Option Pricing of Earnings Announcement Risks. *The Review of Financial Studies*, *32*(2), 646–687. https://doi.org/10.1093/rfs/hhy060

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, *57*(1), 405–443. https://doi.org/10.1111/1540-6261.00426

Easton, P. D. (2004). PE Ratios, PEG Ratios, and Estimating the Implied Expected Rate of Return on Equity Capital. *The Accounting Review*, *79*(1), 73–95. https://doi.org/10.2308/accr.2004.79.1.73

Fama, E. F. (1976). Forward rates as predictors of future spot rates. *Journal of Financial Economics*, *3*(4), 361–377. https://doi.org/10.1016/0304-405x(76)90027-1

Fama, E. F., & Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, *77*(4), 680–692.

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, *81*(3), 607–636.

Farboodi, M., & Veldkamp, L. (2020). Long-Run Growth of Financial Data Technology. *American Economic Review*, *110*(8), 2485–2523. https://doi.org/10.1257/aer.20171349

Francis, J., LaFond, R., Olsson, P. M., & Schipper, K. (2004). Costs of Equity and Earnings Attributes [Publisher: American Accounting Association]. *The Accounting Review*, *79*(4), 967–1010.

Frankel, R., Kothari, S. P., & Weber, J. (2006). Determinants of the informativeness of analyst research. *Journal of Accounting and Economics*, *41*(1), 29–54. https://doi.org/10.1016/j.jacceco.2005.10.004

Frankel, R., & Lee, C. M. (1998). Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics*, *25*(3), 283–319. https://doi.org/10.1016/S0165-4101(98)00026-3

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5). https://doi.org/10.1214/aos/1013203451

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Fuster, A., Perez-Truglia, R., Wiederholt, M., & Zafar, B. (2022). Expectations with Endogenous Information Acquisition: An Experimental Investigation. *The Review of Economics and Statistics*, *104*(5), 1059–1078. https://doi.org/10.1162/rest_a_00994

Gao, M., & Huang, J. (2020). Informing the Market: The Effect of Modern Information Technologies on Information Production. *The Review of Financial Studies*, *33*(4), 1367–1411. https://doi.org/10.1093/rfs/hhz100

Gebhardt, W. R., Lee, C. M. C., & Swaminathan, B. (2001). Toward an Implied Cost of Capital. *Journal of Accounting Research*, *39*(1), 135–176. https://doi.org/10.1111/1475-679X.00007

Ghysels, E., Horan, C., & Moench, E. (2017). Forecasting through the rearview mirror: Data revisions and bond return predictability. *The Review of Financial Studies*, *31*(2), 678–714. https://doi.org/10.1093/rfs/hhx098

Givoly, D., & Lakonishok, J. (1980). Financial analysts' forecasts of earnings: Their value to investors. *Journal of Banking & Finance*, *4*(3), 221–233. https://doi.org/10.1016/0378-4266(80)90020-5

Gu, F., & Wang, W. (2005). Intangible Assets, Information Complexity, and Analysts Earnings Forecasts. *Journal of Business Finance & Accounting*, *32*(9-10), 1673–1702. https://doi.org/10.1111/j.0306-686X.2005.00644.x

Gu, Z., & Wu, J. S. (2003). Earnings skewness and analyst forecast bias. *Journal of Accounting and Economics*, *35*(1), 5–29. https://doi.org/10.1016/S0165-4101(02)00095-2

Gürkaynak, R. S., & Wright, J. H. (2012). Macroeconomics and the term structure. *Journal of Economic Literature*, *50*(2), 331–367.

Ham, C. G., Kaplan, Z. R., & Lemayian, Z. R. (2022). Rationalizing forecast inefficiency. *Review of Accounting Studies*, *27*(1), 313–343. https://doi.org/10.1007/s11142-021-09622-8

Harford, J., Jiang, F., Wang, R., & Xie, F. (2019). Analyst Career Concerns, Effort Allocation, and Firms Information Environment. *The Review of Financial Studies*, *32*(6), 2179–2224. https://doi.org/10.1093/rfs/hhy101

Hirshleifer, D. (2001). Investor Psychology and Asset Pricing. *The Journal of Finance*, *56*(4), 1533–1597. https://doi.org/10.1111/0022-1082.00379

Hirshleifer, D., & Sheng, J. (2022). Macro news and micro news: Complements or substitutes? *Journal of Financial Economics*, *145*(3), 1006–1024. https://doi.org/10.1016/j.jfineco.2021.09.012

Ho, T. S. Y., & Michaely, R. (1988). Information Quality and Market Efficiency [Publisher: [Cambridge University Press, University of Washington School of Business Administration]]. *The Journal of Financial and Quantitative Analysis*, *23*(1), 53–70. https://doi.org/10.2307/2331024

Hou, K., van Dijk, M. A., & Zhang, Y. (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics*, *53*(3), 504–526. https://doi.org/10.1016/j.jacceco.2011.12.001

Huang, D., Jiang, F., Li, K., Tong, G., & Zhou, G. (2023). Are bond returns predictable with real-time macro data? *Journal of Econometrics*, *237*(2, Part C), 105438. https://doi.org/https://doi.org/10.1016/j.jeconom.2022.09.008

Huang, S., Xiong, Y., & Yang, L. (2022). Skill Acquisition and Data Sales. *Management Science*, *68*(8), 6116–6144. https://doi.org/10.1287/mnsc.2021.4117

Johnson, T. L., & So, E. C. (2018). Asymmetric Trading Costs Prior to Earnings Announcements: Implications for Price Discovery and Returns. *Journal of Accounting Research*, *56*(1), 217–263. https://doi.org/10.1111/1475-679X.12189

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring Uncertainty. *American Economic Review*, *105*(3), 1177–1216. https://doi.org/10.1257/aer.20131193

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, *30*.

Kerr, J., Sadka, G., & Sadka, R. (2020). Illiquidity and Price Informativeness [Publisher: INFORMS]. *Management Science*, *66*(1), 334–351. https://doi.org/10.1287/mnsc.2018.3154

Keung, E., Lin, Z.-X., & Shih, M. (2010). Does the Stock Market See a Zero or Small Positive Earnings Surprise as a Red Flag? *Journal of Accounting Research*, *48*(1), 105–136. https://doi.org/10.1111/j.1475-679X.2009.00354.x

Koijen, R. S. J., Moskowitz, T. J., Pedersen, L. H., & Vrugt, E. B. (2013, August). *Carry* (Working Paper No. 19325). National Bureau of Economic Research.

Koijen, R. S., Moskowitz, T. J., Pedersen, L. H., & Vrugt, E. B. (2018). Carry. *Journal of Financial Economics*, *127*(2), 197–225.

Kothari, S. P. (2001). Capital markets research in accounting. *Journal of Accounting and Economics*, 127.

Kothari, S. P., So, E., & Verdi, R. (2016). Analysts Forecasts and Asset Pricing: A Survey. *Annual Review of Financial Economics*, *8*(1), 197–219. https://doi.org/10.1146/annurev-financial-121415-032930

Kross, W., Ro, B., & Schroeder, D. (1990). Earnings Expectations: The Analysts' Information Advantage [Publisher: American Accounting Association]. *The Accounting Review*, *65*(2), 461–476.

Lee, C. M. C., So, E. C., & Wang, C. C. Y. (2021). Evaluating Firm-Level Expected-Return Proxies: Implications for Estimating Treatment Effects. *The Review of Financial Studies*, *34*(4), 1907–1951. https://doi.org/10.1093/rfs/hhaa066

Lehavy, R., Li, F., & Merkley, K. (2011). The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts. *The Accounting Review*, *86*(3), 1087–1115. https://doi.org/10.2308/accr.00000043

Liu, Y., & Wu, J. C. (2021). Reconstructing the yield curve. *Journal of Financial Economics*, *142*(3), 1395–1425. https://doi.org/https://doi.org/10.1016/j.jfineco.2021.05.059

Ljungqvist, A., Marston, F., Starks, L. T., Wei, K. D., & Yan, H. (2007). Conflicts of interest in sell-side research and the moderating role of institutional investors. *Journal of Financial Economics*, *85*(2), 420–456. https://doi.org/10.1016/j.jfineco.2005.12.004

Lobo, G. J., Song, M., & Stanford, M. (2012). Accruals quality and analyst coverage. *Journal of Banking & Finance*, *36*(2), 497–508. https://doi.org/10.1016/j.jbankfin.2011.08.006

Loh, R. K., & Stulz, R. M. (2018). Is Sell-Side Research More Valuable in Bad Times? *The Journal of Finance*, *73*(3), 959–1013. https://doi.org/10.1111/jofi.12611

Loughran, T., & McDonald, B. (2023). Measuring Firm Complexity. *Journal of Financial and Quantitative Analysis*, 1–28. https://doi.org/10.1017/S0022109023000716

Loughran, T., & Mcdonald, B. (2014). Measuring Readability in Financial Disclosures. *The Journal of Finance*, *69*(4), 1643–1671. https://doi.org/10.1111/jofi.12162

Loughran, T., & Mcdonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, *54*(4), 1187–1230. https://doi.org/10.1111/1475-679X.12123

Ludvigson, S. C., Ma, S., & Ng, S. (2021). Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response? *American Economic Journal: Macroeconomics*, *13*(4), 369–410. https://doi.org/10.1257/mac.20190171

Lys, T., & Soo, L. G. (1995). Analysts' Forecast Precision as a Response to Competition [Publisher: SAGE Publications Inc]. *Journal of Accounting, Auditing & Finance*, *10*(4), 751–765. https://doi.org/10.1177/0148558X9501000404

Mackowiak, B., Matejka, F., & Wiederholt, M. (2021). Rational Inattention: A Review. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3871534

McCracken, M. W., & Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

Moench, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics*, *146*(1), 26–43. https://doi.org/10.1016/j.jeconom.2008.06.002

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703.

Ohlson, J. A., & Juettner-Nauroth, B. E. (2005). Expected EPS and EPS Growth as Determinantsof Value. *Review of Accounting Studies*, *10*(2-3), 349–365. https://doi.org/10.1007/s11142-005-1535-3

Ryans, J. (2017). Using the EDGAR Log File Data Set. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2913612

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690. https://doi.org/10.1016/S0304-3932(03)00029-1

So, E. C. (2013). A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts? *Journal of Financial Economics*, *108*(3), 615–640. https://doi.org/10.1016/j.jfineco.2013.02.002

Uddin, A., Tao, X., Chou, C.-C., & Yu, D. (2022). Are missing values important for earnings forecasts? A machine learning perspective. *Quantitative Finance*, *22*(6), 1113–1132. https://doi.org/10.1080/14697688.2021.1963825

Van Nieuwerburgh, S., & Veldkamp, L. (2010). Information Acquisition and Under-Diversification. *The Review of Economic Studies*, *77*(2), 779–805. https://doi.org/10.1111/j.1467-937X.2009.00583.x

van Binsbergen, J. H., Han, X., & Lopez-Lira, A. (2022). Man versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases. *The Review of Financial Studies*, hhac085. https://doi.org/10.1093/rfs/hhac085

Veldkamp, L. (2011, September). *Information Choice in Macroeconomics and Finance*. Princeton University Press.

Verrecchia, R. E. (1982). Information Acquisition in a Noisy Rational Expectations Economy [Publisher: [Wiley, Econometric Society]]. *Econometrica*, *50*(6), 1415–1430. https://doi.org/10.2307/1913389

Wan, R., Fulop, A., & Li, J. (2022). Real-time bayesian learning and bond return predictability [ANNALS ISSUE IN "BAYESIAN METHODS IN ECONOMICS AND FINANCE"]. *Journal of Econometrics*, *230*(1), 114–130. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.04.052

Weiss, D., Naik, P. A., & Tsai, C.-L. (2008). Extracting Forward-Looking Information from Security Prices: A New Approach [Publisher: American Accounting Association]. *The Accounting Review*, *83*(4), 1101–1124.

Zhang, X. F. (2006a). Information Uncertainty and Analyst Forecast Behavior. *Contemporary Accounting Research*, *23*(2), 565–590. https://doi.org/10.1506/92CB-P8G9-2A31-PV0R

Zhang, X. F. (2006b). Information Uncertainty and Stock Returns. *The Journal of Finance*, *61*(1), 105–137. https://doi.org/10.1111/j.1540-6261.2006.00831.x