Man vs. Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases*

Jules H. van Binsbergen[1], Xiao Han[2], and Alejandro Lopez-Lira[†3]

[1] *The Wharton School, University of Pennsylvania, NBER and CEPR*

[2] *University of Edinburgh*

[3] *BI Norwegian Business School*

First Version: June 11, 2020

This Version: September 14, 2020

Abstract

We use machine learning to construct a statistically optimal and unbiased benchmark for firms' earnings expectations. We show that analyst expectations are on average biased upwards, and that this bias exhibits substantial time-series and cross-sectional variation. On average, the bias increases in the forecast horizon, and analysts revise their expectations downwards as earnings announcement dates approach. We find that analysts' biases are associated with negative cross-sectional return predictability, and the short legs of many anomalies consist of firms for which the analysts' forecasts are excessively optimistic relative to our benchmark. Managers of companies with the greatest upward biased earnings forecasts are more likely to issue stocks.

Key words: Earnings Forecasts, Machine Learning, Investment Strategies

*We thank Refinitive for their guidance in using the I/B/E/S database. We are grateful for helpful comments and suggestions provided by seminar participants at BI Norwegian Business School.

†Emails: julesv@wharton.upenn.edu; xiao.han@ed.ac.uk; alejandro.lopez-lira@bi.no

1 Introduction

One necessary input for pricing a risky asset is an estimate of expected future cash flows to which the asset owner would be entitled. Common proxies include the most recent realized cash flow, simple linear forecasts, or analysts' forecasts. However, a significant strain of literature documents that forecasts can be biased or predict poorly out-of-sample, thereby limiting their practical usefulness.[1] In this study, we propose a novel approach for constructing a statistically optimal and unbiased benchmark for earnings expectations, which uses machine learning. We demonstrate that our new benchmark is effective out-of-sample.

To provide conditional expectations available in real-time, we use the cross-sectional information of firms' balance sheets, macroeconomic variables, and analysts' predictions. Due to analysts' forecasts belonging to the public information set, the question arises whether these forecasts can be used to improve upon forecasts obtained from other publicly available data sources. Namely, analysts' forecasts may become redundant if other publicly available variables are included in the analysis. Alternatively, analysts may collect valuable private information that is subsequently reflected in their forecasts. We find evidence consistent with the latter: Analysts' forecasts are not redundant relative to the large set of publicly available variables in our algorithm. As such, these forecasts provide a key input to our machine learning approach, as Figure 1 and Figure 2 show.[2]

We use random forest regression for our main analysis. Random forest regression has two significant advantages. First, it naturally allows nonlinear relationships. Second, it is designed for high-dimensional data and is therefore robust to overfitting.[3] We construct one-year- and two-years-ahead forecasts for annual earnings, and one-quarter-, two-quarters-, and three-quarters-ahead forecasts for quarterly earnings. We focus on these particular horizons as analysts' forecasts for other horizons have significantly fewer observations. Given

[1] See Kothari et al. (2016) for an extensive review.

[2] Using mixed data sampling regression, Ball and Ghysels (2018) find that analysts' forecasts provide complementary information to the time-series forecasts of corporate earnings at short horizons of one quarter or less.

[3] See Gu et al. (2020) for an excellent overview of this and other common predictive algorithms in the context of cross-sectional returns. See Bryzgalova et al. (2020) for a novel application of random forest to construct portfolios.

the benchmark expectation provided by our machine learning algorithm, we then calculate the bias in expectations as the difference between the analysts' forecasts and the machine learning forecasts.

We show that analysts' biases increase in the forecast horizon. On average, analysts revise their expectations downwards as the earnings announcement day approaches. These revisions induce negative cross-sectional stock predictability: stocks with more optimistic expectations earn lower subsequent returns. Importantly, the short legs of common anomalies consist of firms for which the analysts' forecasts are excessively optimistic relative to our unbiased benchmark. Finally, we show that managers of those companies with the largest biases seem to take apparent advantage of the over-optimistic expectations by issuing stocks.[4]

Although previous research has used realized earnings to evaluate the bias and efficiency of analyst forecasts, the extant studies do not provide a time series or cross-section of unbiased real-time earnings estimates.[5] Without such forecasts, it is difficult to assess and/or correct for the dynamics of forecast biases before the actual value is realized. Specifically, we cannot know whether the given forecasts are conditionally biased, nor do we know the variance of these biases across stocks and their impact on asset returns.

We fill this void by constructing a statistically optimal time series and cross-section of earnings forecasts. The resulting estimates enable us to compute implied analyst biases, which can be used in cross-sectional stock-pricing sorts. As we show in Figure 3, the realized analysts' bias exhibits substantial time series as well as cross-sectional variation. Therefore, our benchmark expectation diverges from the conventional approach, which uses either the raw analysts' expectations, the realized value, or a simple linear model to form the conditional forecast.[6]

Another strain of the relevant literature sorts stocks cross-sectionally using long-term

[4]We are agnostic on the source of the biases for analysts' forecasts. Scherbina (2004) and Scherbina (2007) show that the proportion of analysts who stop revising their annual earnings forecasts is associated with negative earning surprises and abnormal returns, suggesting that analysts withhold negative information from their forecasts.

[5]See for example Kozak et al. (2018) and Engelberg et al. (2018).

[6]The limitations of a simple linear model to forecast earnings have drawn academics' attention recently. See Babii et al. (2020), for example, who use the sparse-group LASSO panel-data regression to circumvent the issue of using mixed-frequency data (such as macroeconomic, financial, and news time series) and apply their new technique to forecast price-earnings ratios.

earnings forecasts, without comparing these values to a benchmark (e.g., La Porta (1996), Bordalo et al. (2019)). Additionally, Zhou (2018) argues that, despite the useful results that have been documented, it remains difficult to determine whether a belief is biased or exaggerated without a benchmark model.

Finally, studies have posited linear forecasting rules as a solution to the analysts' bias problem. An important contribution to this line of research is So (2013). Using a linear regression framework with variables that have been shown to provide effective forecasting power (Fama and French (2006), Hou et al. (2012)), So (2013) provides a linear forecast, and studies the predictable components of analysts' errors and their impact on asset prices. We differ from So (2013) in three important ways.

First, the linear forecasts in So (2013) are not designed to be statistically optimal. In fact, analysts' forecasts are a better proxy for the conditional expectation than are linear forecasts as measured by the mean squared error. In contrast, our machine learning forecasts are a superior proxy both in-sample and out-of-sample. Second, because linear regressions do not efficiently handle high-dimensional data, a variable selection step is necessary. Often, variables that have been documented as effective predictors are selected in this step. In contrast, our machine learning approach considers the broad set of macroeconomic and firm-specific signals at every point in time. We therefore, do not incur any data leakage. Third and finally, there is no reason to impose the linearity of the conditional expectation function, and indeed we find that allowing for nonlinear effects improves the forecasts, consistent with previous studies using machine learning (Gu et al. (2020)).

Armed with a statistically optimal and unbiased benchmark for firms' earnings expectations and the implied real-time measure for firm-level conditional forecast biases across multiple horizons, we focus on two applications. First, we study the impact of expectations and biases on stock market returns. Second, we evaluate the effect of biases on managers' actions. Concerning the first application, we find significant return predictability associated with our measure of conditional biases. Concerning the second, we find that managers tend to issue more stocks when their firms are subject to more optimistic forecasts relative to our benchmark.

Our work relates to recent work by Hirshleifer and Jiang (2010) and Baker and Wurgler (2013) who argue that managers can take advantage of overpricing on their firms' valuation by issuing stocks. Hirshleifer and Jiang (2010) use firms' stock issuances and repurchases to construct a misvaluation factor, and Stambaugh and Yuan (2017) construct a mispricing-factor based on the net stock issuances. We contribute to this literature by providing direct and novel evidence relating to conditional biases and stock issuances. Since we show that it is feasible to have better forecasts than analysts' forecasts using public information, it seems plausible that managers would have superior forecasts when they exploit their private information.

Finally, there is an extensive literature documenting biases and the importance of expectations for macroeconomic variables using the Survey of Professional Forecasters (SPF) (see Bordalo et al. (2018), Coibion and Gorodnichenko (2015), and Bianchi et al. (2020) for recent expositions).[7] We complement this literature by (1) providing direct evidence of the existence of systematic biases in analysts' earnings forecasts, (2) constructing a more efficient forecast using the publicly available information in each period, and (3) documenting that these biases relate to outcomes in financial markets and corporate policies.

2 Methodology and Data

2.1 Random Forest and earnings forecasts

In this study, we use random forest regressions to forecast future earnings. Random forest regression is a non-linear and non-parametric ensemble method that averages multiple forecasts from (potentially) weak predictors. The ultimate forecast is superior to a forecast following from any individual one predictor (Breiman 2001). We train the algorithm using rolling windows analogous to a rolling regression forecast. The hyper-parameters are chosen using cross-validation: a data-driven method that does not have look-ahead bias by design.

[7] In particular, Bianchi et al. (2020) characterizes the time-varying systematic expectation errors embedded in survey responses using machine-learning techniques. See also Bordalo et al. (2019) and Bordalo et al. (2020) who provide evidence of systematic biases in analysts' forecasts of earnings growth.

We summarize the key parameters of our implementation in Table 1 and discuss the cross-validation method in detail in the Appendix. We explain the algorithm itself in detail in this subsection.

[Insert Table 1 about here]

The building blocks for random forest regression are decision trees with a flowchart structure in which the data are recursively split into non-intersecting regions. At each step, the algorithm splits the data choosing the variable and threshold that best minimizes the mean squared error when the average value of the variable to be forecasted is used as the prediction. Decision trees contain two fundamental substructures: *decision nodes* by which the data are split, and *leaves* that represent the outcomes. At the leaves, the forecast is a constant local model equal to the average for that region.

The decision tree in Figure 5 provides an illustration. The variable we wish to forecast is the earnings-per-share (eps hereafter) for a cross-section of firms. At the first step, the selected explanatory variable is the analysts' forecast (denoted by adj_afeps), and the threshold (or cutoff) value is at -0.206. If we were to end at this step, the forecast eps-value is -0.222 when adj_afeps is less than or equal to -0.206, and 3.232 when adj_afeps is less than or equal to -0.206. In the next step, the algorithm splits each of the previous two sub-spaces in two again. The first subspace (analysts' forecast less than or equal to -0.206) is split into two using the price-to-book ratio (PTB) as an explanatory variable. The threshold value is -0.624. The second subspace (analysts' forecast greater than -0.206) uses short-term debt. We then continue for the predefined number of splits until we arrive at the final nodes. In the final nodes, the prediction is the historical local average of that subspace.

A decision tree model's goal is to partition the data to make optimal constant predictions in each partition (or subspace). Consequently, decision trees are fully non-parametric and allow for arbitrary non-linear interactions. The only parameter for training a decision tree model is the depth, i.e., the maximum length of the path from a root node to leaves. The larger the depth, the more complex the tree, and the more likely it will overfit the data.[8]

More formally, the decision tree model forecast is constant over a disjoint number of regions R_m:

$$\hat{y} = f(x) = \sum_m c_m I_{\{x \in R_m\}}, \tag{1}$$

$$c_m = \frac{1}{N_m} \sum_{\{y_i : x_i \in R_m\}} y_i, \tag{2}$$

and each region is chosen by forming rectangular hyper-regions in the space of the predictors:

$$R_m = \{x_i \in \bigtimes_{i \in I} X_i : x_i \leq k_i^m\}, \tag{3}$$

where $\bigtimes$ denotes a Cartesian product, I is the number of predictors. Thus, each predictor x_i can take values in the set X_i.

The algorithm minimizes numerically the mean squared error in order to best approximate the conditional expectation by choosing the variables and thresholds, and hence the regions R_m in a greedy fashion. Because of their non-parametric nature and flexibility, decision tree models are prone to overfitting when the depth is large. The most common solution is to use an ensemble of many decision trees with shorter depth: specifically, random forest regression models.

Random forest regression models are an ensemble of decision trees that bootstrap the predictions of different decision trees. Each tree is trained on a random sample, usually drawn with replacement. Instead of considering all predictors, decision trees are modified so that they use a strict random subset of features at each node to render the individual decision trees' predictions less correlated.[9] The final prediction of a random forest model is

[8] The standard approach to decrease the risk of overfitting is to stop the algorithm whenever the next split would result in a sample size smaller than a predetermined size, usually five observations for regression. This sample threshold is called the *minimum node size*.

[9] The algorithm allows a fixed set of variables to always be considered at each split. More generally, the algorithm enables us to specify the probability for each predictor to be considered at each split.

obtained by averaging each decision tree's predictions.

Random forest regressions provide a natural measure of the importance of each variable, the so-called *impurity importance* (Ishwaran 2015). The impurity importance for variable X_i is the sum of all mean squared error decreases of all nodes in the forest at which a split on X_i has been used, normalized by the number of trees. The impurity importance measure can be biased, and we use the correction of Nembrini et al. (2018) to address this well-known concern. Finally, we normalize the features' importance of each variable as percentages for ease of interpretation.

There are three main parameters in the random forest algorithm: (1) the number of decision trees; (2) the depth of the decision trees; and (3) the fraction of the sample that is taken in each split.[10]

Since the random forest is a bootstrapping procedure, a high number of decision trees are recommended. Notwithstanding computational time, there is no theoretical downside to using more trees. That said, performance tends to plateau following a large number of trees. Figure 6 and 7 confirm that this indeed holds in our setup: The performance is increasing in the number of trees but reaches a plateau.[11]

[Insert Figure 6 and 7 about here]

The depth of each decision tree determines the overall complexity of the model. More complex models usually over-fit. Nevertheless, because of the inherent randomization, random forests are resilient to over-fitting in a wide variety of circumstances. Figures 8 and 9 show that the performance of the model is increasing in model complexity up until a depth of 7.

[Insert Figure 8 and 9 about here]

[10] There is an additional parameter: the percentage of the predictors considered in each splitting step. The random forest algorithm is not sensitive to its value in our specification.

[11] In the cross-validation step, we measure the performance using the out-of-sample R^2 of the year 1986: $R^2_{oos} = 1 - \frac{\sum(MLF_i - EPS_i)^2}{\sum(EPS_i - \overline{EPS})^2}$. MLF and EPS denote the machine learning forecast and actual realized earnings respectively. The denominator, $\sum(EPS_i - \overline{EPS})^2$, is constant across different specifications.

The last hyper-parameter we have to choose is the fraction of the sample used to train each tree. For example, if that fraction is set to 1%, for each decision tree we would then first take a 1% random sub-sample without replacement as the training sample. We then repeat the process for each remaining tree. Figures 10 and 11 show the relationship between the fraction of the sample used to train each tree and the out-of-sample R^2 in 1986, the year we use for cross-validation. The performance is first increasing in the fraction size and then decreasing. The algorithm benefits from using a small fraction of the sample for each tree.

[Insert Figure 10 and 11 about here]

For the quarterly earnings forecasts and one-year ahead forecast, we train the random forest model using data from the most recent year and forecast earnings in the following periods using only the information available at the current time. For the two-year-ahead forecasts, we train the model using data from the two most recent years because we do not have enough observations when using a 12-month window to train the model.[12] The forecasts are therefore out-of-sample by design. The resulting forecasting regression is:

$$E_t[eps_{i,t+\tau}] = RF[Fundamentals_{i,t}, Macro_t, AF_{i,t}].$$

RF denotes the random forest model using data from the most recent periods. $Fundamentals_{i,t}$, $Macro_t$, and $AF_{i,t}$ denote firm fundamental variables, macroeconomic variables, and analysts' earnings forecasts respectively. The earnings per share of firm i in quarter t+ τ (τ=1 to 3) or year t+τ (τ=1 to 2) is $eps_{i,t+\tau}$. We focus on five forecast horizons, including one-quarter-ahead, two-quarters-ahead, three-quarters-ahead, one-year-ahead, and two-years-ahead because analysts' forecasts for other horizons have significantly fewer observations. As analysts make earnings forecasts every month, we provide our statistically optimal benchmark for every month as well.[13]

[12] Our results remain similar when using longer windows to train the models.

[13] To minimize the impacts of outliers within the model, we winsorize the forecasting variables at the 1% level and standardize them following the recommended guidelines in the literature.

2.2 Variables used for earnings forecasts

We consider an extensive collection of public signals available at each point in time, summarized into three categories: firm-specific variables, macroeconomic variables, and analysts' earnings forecasts.

2.2.1 Firm fundamentals

We consider firm fundamental variables related to future earnings

1. Realized earnings from the last period. Earnings data are obtained from /I/B/E/S
2. Earnings growth, defined as the growth rate in earnings
3. Sales growth, defined as the growth rate in sales and obtained from COMPUSTAT
4. Asset growth, defined as the growth rate in total assets and obtained from COMPUSTAT
5. Investment growth, defined as the growth rate in capital expenditure and obtained from COMPUSTAT
6. Monthly stock prices and returns from CRSP
7. Sixty-seven financial ratios such as book-to-market ratio and dividend yields obtained from the Financial Ratios Suit by Wharton Research Data Services.[14]

2.2.2 Macroeconomic variables

We consider several macroeconomic variables that can affect firms' earnings. We obtain the from the real-time data set provided by the Federal Reserve Bank of Philadelphia.

1. Consumption growth, defined as the log difference of consumption in goods and services
2. GDP growth, defined as the log difference of real GDP
3. Growth of industrial production, defined as the log difference of Industrial Production Index (IPT)
4. Unemployment rate

2.2.3 Analyst forecasts

Analysts' forecasts at time $t-1$ for firm j's earnings at fiscal end period t can be decomposed into public and private signals:[15]

[14] See Appendix for details of these variables' definitions.
[15] See Hughes et al. (2008) and So (2013) among others.

$$AF_{j,t-1}^{t} = \sum_{i=1}^{N} \beta_i X_{j,t-1} + \sum_{i=1}^{M} \gamma_i P_{j,i-1} + \sum_{i=1}^{M} \sigma_i B_{j,i-1}, \tag{4}$$

where $X_{j,t-1}$ are public signals known at $t-1$; $P_{j,i-1}$ are private signals; and $B_{j,i-1}$ are analysts' biases generated by expectation errors or incentive problems. Our machine learning algorithm is designed to use the private signals optimally in analyst forecasts while correcting for their biases.

As pointed out by Diether et al. (2002), mistakes occur when matching I/B/E/S unadjusted actual file (actual realized earnings) with I/B/E/S unadjusted summary file (analysts' forecasts), because stock splits may occur between the earnings forecast day and the actual earnings announcement day. In these cases, the forecast and the actual EPS value are based on different numbers of shares outstanding. To address this issue, we use the cumulative adjustment factors from the CRSP monthly stock file to adjust the forecast and the actual EPS on the same share basis.[16]

2.3 Measuring the term structure of real-time biases in analysts' expectations

The I/B/E/S database provides different forecast periods indicated by FPI for analysts' earnings forecasts.[17] The span of the earnings forecast periods is one quarter to five years. The I/B/E/S database also provides forecasts of long-term earnings growth, defined as the expected annual increase in operating earnings over the company's next cycle ranging from three to five years (Bordalo et al.; 2019). At each month t, we measure the biases in investors' expectations as the differences between the analysts' forecast and the machine learning forecast, scaled by the closing stock price from the most recent month:

[16]We do not use the adjusted summary files, because there are rounding errors when I/B/E/S adjusts the share splits for forecasts and actual earnings (Diether et al. (2002)).

[17]For example, the FPI of 1 corresponds to the one-year-ahead earnings forecasts.

$$BiasedExpectation_{i,t}^{t+h} = \frac{Analysts'Forecasts_{i,t}^{t+h} - MLForecast_{i,t}^{t+h}}{Price_{i,t-1}} \tag{5}$$

in which subscript i denotes firm, and t denotes the date when earnings forecasts are made. The superscript $t+h$ denotes forecasting periods, and ML denotes machine learning.

3 Hypotheses

3.1 Biased expectations and cross-section of stock returns

If indeed, our machine learning forecasts provide the rational benchmark for earnings expectations, while investors are affected by (biased) analysts' forecasts, we should observe that the stocks with optimistic earnings forecasts will earn low future returns. After all, overly optimistic earnings forecasts are associated with stock overpricing. Our first hypothesis is, therefore:

Hypothesis 1: Stocks with more optimistic earning forecasts earn lower future returns.

3.2 Biased expectations and market timing

Bordalo et al. (2019) and Bouchaud et al. (2019) show that investors exhibit biases when using current and past earnings information to issue forecasts for the future. Baker and Wurgler (2013) argue that corporate managers have more information about their own firms than investors have, and can use that informational advantage. As such managers could take advantage of investors' expectation biases.

We, therefore, conjecture that managers can identify when investors overestimate or underestimate firms' future cash flows and, further, that managers' expectations will align more closely to our statistically optimal benchmark.[18] For example, managers may issue more stock

[18] Baker and Wurgler (2013) provide a comprehensive review of how rational managers make firm policies in response to mispricing caused by irrational investors.

when investors' expectations are higher than their own, i.e., engage in market timing (Baker and Wurgler; 2002). Therefore, our second hypothesis is:

Hypothesis 2: Firms with more optimistic analysts' forecasts relative to the statistically optimal benchmark issue more stocks in the subsequent periods.

4 Empirical Findings

4.1 Stylized fact: Downward Revisions in Analysts' Earnings Forecasts

Analysts revise their earnings forecasts every month. As the announcement dates approach, analysts should process new information and update their positions to make better forecasts. Table 2 demonstrates that analysts revise their earnings forecasts.

[Insert Table 2 about here]

We find that the average forecast error, defined as the difference between analysts' earnings forecasts per share and the realized earnings per share, is consistently positive for all horizons; the results suggest that analysts make over-optimistic forecasts. Further, the average error decreases as the earnings announcement dates approach; i.e., on average, a downward revision occurs in analysts' forecasts. As expected, the mean squared error also decreases. Analysts make more precise forecasts when the earning announcement dates approach.

For the one-quarter-ahead forecast, the average forecast error decreases from \$0.025, when analysts make the first forecast, to \$0.014, when analysts make the last forecast for the same fiscal quarter end date, which usually follows the quarter end date but precedes the announcement. The mean squared error also declines from 0.075 to 0.061.

A downward revision also occurs in the two-quarters-ahead, the three-quarters-ahead, the one-year-ahead, and the two-years-ahead forecasts. To the extent that investors follow analysts' forecasts and analysts make optimistic expectations (Hribar and McInnis; 2012), these downward updates may result in negative cross-sectional return predictability. Specifically, stocks with more optimistic expectations should earn lower subsequent returns than stocks with less optimistic expectations.

The realized values of earnings are not available when making the forecasts; therefore, the ex-post establishment of biases and their importance is not conducive to forming portfolios in real-time. We cannot know which stocks have biased expectations when using the realized value as a benchmark until that realized value is revealed. In contrast, our statistically optimal benchmark allows us to study the effects of the bias before realization.

4.2 Earnings Forecasts via Machine Learning

Table 3 compares the properties of analysts' earnings forecasts with the statistically optimal forecasts estimated using the Random Forest regressions.

[Insert Table 3 about here]

We find that for forecasts at all horizons, analysts make over-optimistic forecasts on average. The realized analysts' forecasts errors, defined as the difference between the analysts' forecasts and the realized value, increase in the forecast horizon, ranging from 0.018 to 0.348, all of which are statistically significantly different from zero. In sharp contrast, the time-series averages of the differences between the machine-learning forecast and realized earnings are statistically indistinguishable from zero, with an average value of around -0.004 for the quarterly earnings forecasts, 0.016 for the one-year- ahead forecast, and -0.022 for the two-years-ahead forecast.

The mean squared errors of the machine-learning forecast are smaller than the analysts' mean squared errors, demonstrating that our forecasts are, as expected, more precise than are the analysts' forecasts. Overall, our results are consistent regarding machine-learning forecasts' being closer to realized earnings than are analysts' forecasts.

Figure 1 and 2 report the feature importance for the one-year-ahead and one-quarter-ahead earnings forecasts, respectively. The feature importance results are similar for other forecast horizons, and we report them in the Appendix. Analysts' forecasts, past realized earnings, and stock price are the most important variables, and their normalized importance roughly equals 0.20, 0.15, and 0.10, respectively. Other variables such as return on capital employed (ROCE), return on equity (ROE), and pre-tax profit margin (PTPM) also contain useful information for future earnings. Analyst forecasts and stock price get high importance values likely because they reflect the information from many market participants.

[Insert Figure 1 and 2 about here]

We define the conditional expectation bias for every stock as the difference between the analysts' forecast and the machine-learning forecast, scaled by the price in the most recent month, as consistent with the previous literature. The second-to-last column of Table 3 reports the time-series average of the real-time biased expectations. The average conditional bias is statistically different from zero for all horizons. Furthermore, we find that analysts are more biased in longer horizons.

Figure 4 shows the conditional aggregate bias, defined as the average of the individual stocks' expectations. We consider five different forecast horizons and consider the possibility that the aggregate bias is higher during historical bubbles. We find that clear spikes during the Internet bubble around 2001 (Griffin et al. (2011)) and in the financial crisis around 2008; these findings are consistent with the position that analysts are more over-optimistic during bubbles and more pessimistic when stock markets collapse.

[Insert Figure 4 about here]

4.3 Conditional Bias and the Cross-section of Stock Returns

We demonstrated above that analysts are, on average, over-optimistic relative to both the machine-learning benchmark and the realized value, and on average, they update their forecasts downward. If market participants' beliefs align closely with analysts' expectations, then we should observe negative return predictability. Stocks with a higher conditional bias (over-optimistic) should earn lower returns than stocks with a lower conditional bias.[19]

We conduct monthly (Fama-MacBeth) cross-sectional predictive regressions of stock returns on the conditional bias from the previous month, and we report the time-series average of the slope coefficients. Analysts make forecasts on firms' cash flows at multiple horizons; hence we have many conditional biases at every point in time for each firm. For each firm, we use the average of the conditional biases at the multiple horizons, scaled appropriately, as the predictor.[20] For a robustness check, we define the bias score as the arithmetic average of the percentile rankings on each of the five conditional bias measures. We then run a separate predictive regression for the bias score.

Table 4 shows the regression results. The first column in each panel of Table 4 reports the regression without control variables. We find that both the conditional bias and the bias score are associated with negative cross-sectional stock predictability. The coefficient on the conditional bias is -0.0808 with a t-statistic of -4.61, so the zero-investment portfolio associated with this variable has a Sharpe ratio of approximately 0.23. The coefficient on the bias score is also significantly negative with a t-statistic of -6.57. The R^2s for both regressions is approximately 0.01.

[Insert Table 4 about here]

The second column in each panel of Table 4 reports the regressions with control variables, including size, book-to-market ratio, short-term reversal, medium-term momentum, return volatility, share turnover, idiosyncratic volatility; and investment. These variables have been shown to predict stock returns with significant efficacy (Green et al. (2017), Freyberger et al. (2020), and Gu et al. (2020)). We find that the coefficients on both the conditional bias and the bias score remain statistically significant after controlling for those variables. We report the individual conditional bias results in the Appendix: all biases exhibit negative return

[19] We note that, if market participants are using the statistically optimal benchmark and do not follow analyst expectations, we should not find cross-sectional predictability. We document the predictability.

[20] We divide annual forecasts by four to make them comparable to quarterly forecasts. Each month we exclude stocks with fewer than two observations for the five forecasts at different horizons.

predictability. Moreover, conditional biases' return predictability remains consistent when we either scale conditional biases with total assets from the most recent fiscal period or drop stocks whose price are lower than $5. We report these and further robustness checks in the Appendix.

Table 5 reports the correlations between the bias measures and the control variables. We find that the conditional bias and the bias score are, as expected, very positively correlated. Moreover, the conditional bias is negatively correlated with size and momentum; we find the conditional bias positively correlated with book-to-market ratio, idiosyncratic volatility, return volatility, and share turnover. Accordingly, stocks with a smaller size, lower past cumulative returns, and with higher book-to-market ratio, idiosyncratic volatility, return volatility, and share turnover, tend to have more over-optimistic expectations. In the Appendix, we report the summary statistics of these variables.

[**Insert Table 5 about here**]

Additionally, we show that the results from the cross-sectional regressions also hold in time-series regressions. We sort stocks into five quintile portfolios based on the conditional bias. Table 6 reports the portfolio sorts in which we can see two interesting patterns: First, the value-weighted returns decrease in the conditional bias. A long-short portfolio of the extreme quintile results in a return spread of 1.95% per month (t-statistic 5.88) for the average bias and 1.53% per month (t-statistic 4.90) for the bias score. Second, the CAPM betas of these portfolios tend to increase with higher biased expectations, which is consistent with the results of Antoniou et al. (2015) and Hong and Sraer (2016), who show that high-beta stocks are more susceptible to speculative overpricing.

[**Insert Table 6 about here**]

We further examine whether returns on this long-short strategy can be explained by leading asset pricing models. Table 7 Panel A reports the results of using the conditional bias as the portfolio sorting variable. We find that the long-short strategy has a significant CAPM alpha of 2.39% per month, with a significantly negative market beta of −0.66. Columns four to seven show the regression results with the Fama-French three-factor and five-factor models. Neither model can explain the documented return spread. The alpha in the three-factor model is 2.52% with a t-statistic of 9.70; the alpha in the five-factor model is 2.02% with a t-statistic of 7.21. Table 7 Panel B reports the long-short strategy using the average bias as the sorting variable, and we find consistent results.[21] Overall, we conclude that the return predictability of the conditional bias appears in cross-sectional regressions and time-series tests against common multi-factor representations.

[**Insert Table 7 about here**]

Since the magnitude and significance of the results seem large by usual standards, we conduct a placebo test in the Appendix to better understand the results. We replace the machine learning forecast with the future realized value and then compute the conditional bias. The implied returns of these infeasible strategies are overwhelmingly better than the ones presented, with monthly excess returns in the order of 5% and t-statistics above 20, making the previous results look small in comparison.

4.4 Conditional Bias and Market Anomalies

In two recent studies, Engelberg et al. (2018) and Kozak et al. (2018) compare analysts' earnings forecasts to the realized values and both find that analysts tend to have over-optimistic expectations for stocks in the short side of anomalies, which are usually associated with lower returns. However, as previously mentioned, the realized value is not available in time for analysts' forecasts; therefore, we cannot assess whether the bias drives anomalies using the realized value. To shed light on this issue, we use our conditional bias measure to examine whether analysts have more conditional over-optimistic expectations on anomaly

[21] We report the results of the long-short strategy based on individual conditional bias in the Appendix. All strategies but for the one using the one-year-ahead bias exhibit significant alpha.

shorts.

We focus on the 27 significant and robust anomalies considered in Hou et al. (2015). We examine these anomalies for two reasons: *i*) they cover the most prevalent anomalies, including momentum, value-versus-growth, investment, profitability, intangibles, as well as trading frictions; and *ii*) they have been widely used to test leading asset pricing models (Hou et al. (2015), Stambaugh and Yuan (2017), and Daniel et al. (2017)).[22] We follow the literature and sort stocks into ten portfolios based on the decile of each anomaly variable. We define the extreme deciles as the long and the short legs of the anomaly strategies.

Having obtained ranks of stocks based on each anomaly variable, we then combine these ranks to construct an anomaly score defined as the equal-weighted average of the rank scores of the 27 anomaly variables. To calculate the score, for each month, we assign decile ranks to each stock based on the 27 anomaly variables.[23] The anomaly score for an individual stock is calculated as the arithmetic average of its ranking on each of the 27 anomalies. Next, we break stocks into 10 decile portfolios based on the anomaly score. The long legs are defined as the stocks in the top decile portfolio. The short legs are defined as the stocks in the bottom decile portfolio.

[Insert Table 8 about here]

Table 8 Panel A presents the average anomaly score for portfolios sorted independently on the conditional bias and the anomaly score.[24] For each anomaly decile portfolio, the anomaly score ranges from 3.21 to 6.92, with the highest (lowest) score indicating the long (short) leg of the anomaly strategy. Table 8 Panel B reports the average number of stocks in each of 10×5 portfolios. On average, we have around 50 stocks every month in each portfolio. Moreover, the average number of stock per month for the portfolio with the highest conditional biases

[22]Table A13 in Appendix lists the anomalies associated with their academic publications. The sample period spans July 1965 to December 2019, depending on the data availability. We follow the descriptions detailed in Hou et al. (2015) to construct the anomaly variables. The last column in Table A13 reports the monthly average returns (in percent) of the long-short anomaly portfolios.

[23]We exclude stocks with fewer than 10 anomaly variables data.

[24]For the results shown in Tables 8 and 9, we use the average of the conditional biases at different forecast horizons to sort the portfolios. The results remain robust when we use the arithmetic average of the percentile rankings on each of the five conditional bias measures.

and the lowest anomaly score is 93, which is more than double the average number of stocks per month for the portfolio with both the lowest conditional biases and the lowest anomaly score (32 stocks). Stocks with higher conditional biases tend to be anomaly shorts.

Table 9 presents the value-weighted excess returns of the portfolios formed by sorting independently on the conditional bias and the anomaly score.[25] While the long-short anomaly strategy in each quintile sort on the conditional bias has a similar anomaly score (around 3.60), we find that anomalies' payoffs increase when the conditional bias increases. In the quintile group with the greatest conditional bias, the long-short strategy based on anomaly score earns the highest returns (2.22% per month with a t-statistic of 6.11). In sharp contrast, the anomaly payoff becomes 0.30% (t-statistic is 0.95) in the quintile group with the smallest conditional bias. Further, we find that the short-leg portfolio's return decreases from 1.24% per month to −2.27% when we move from the first quintile of the conditional bias to the fifth quintile. These findings suggest that anomaly payoffs tend to arise from overpricing on stocks with the most over-optimistic expectations.

[Insert Table 9 about here]

The last two rows in Table 9 report the conditional biases on each of the 10 decile portfolios formed on the anomaly score. We find that the short-leg portfolio is comprised of stocks with more over-optimistic expectations, suggestive of overpricing. Moreover, the difference in conditional biases between anomaly-short and anomaly-long is 0.006 and significant at the 1% level (with t-statistic of 5.75). This finding is consistent with Engelberg et al. (2018) and Kozak et al. (2018) who find that analysts tend to have over-optimistic expectations for stocks in the short side of anomalies.

4.5 Conditional Bias and Firm's Financing Decisions

Managers have more information about their firm than most investors have, due to the access managers have to private information as well as available public signals. Baker and Wurgler

[25]The long-short portfolio using the anomaly score earns 1.20% per month (the t-statistic is 5.32).

(2013) argue that managers use their additional information to the advantage of existing shareholders and engage in market timing (Baker and Wurgler; 2002). Following Hypothesis 2, we conjecture that managers issue more equity whenever analysts' expectations are more optimistic than the statistically optimal benchmark.

We follow the literature and measure net stock issuances as the logarithm of the ratio of the split-adjusted shares outstanding at the fiscal year ending in calendar year t to the split-adjusted shares outstanding at the fiscal year ending in $t-1$. Because the net stock issuances are measured annually, we match the average of the conditional bias in the past 12 months to the fiscal year ending at time t.[26] Table 10 Panel A reports the value-weighted average net stock issuance for companies sorted in portfolios according to the conditional bias of analysts' forecasts relative to the machine-learning forecast.

The net stock issuances increase monotonically in the conditional bias. Importantly, we find that stocks in the decile portfolio with the most optimistic expectations issue significantly more stocks than do those stocks with the least optimistic expectations. Managers of firms whose earnings forecasts are more optimistic issue on average 6% more of total shares outstanding. The difference is statistically significant at the 1% level.

[Insert Table 10 about here]

Table 10 Panel B reports the Fama-MacBeth regressions of firms' net stock issuances on the conditional bias. As in Baker and Wurgler (2002) and Pontiff and Woodgate (2008), we control for variables such as firm size, the book-to-market ratio, and earnings before interest, taxes, and depreciation divided by total assets. Overall, our findings are consistent with the previous portfolio sorts: managers of firms with a larger conditional bias issue more stocks. We also find that firms with a smaller size, lower book-to-market ratio, and lower profitability tend to issue more stocks, consistent with the results in Baker and Wurgler (2002) and Pontiff and Woodgate (2008).

[26]Our results remain robust when matching the average of the conditional bias from the past 24-12 months to the net stock issuances of the fiscal year ending at time t. We report this robustness check in the Appendix.

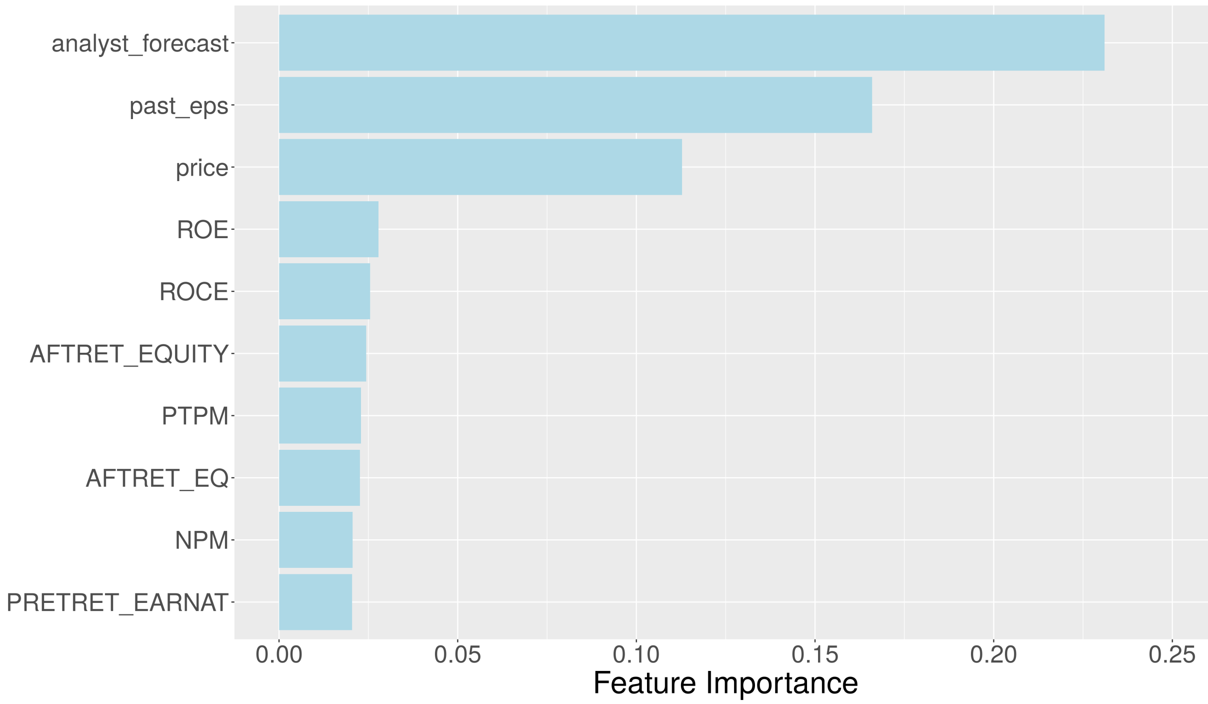
5 Conclusion

The pricing of assets relies significantly on the forecasts of associated cash flows. Analysts' forecasts of earnings are often used as a measure of expectations, despite the common knowledge that these forecasts are on average biased upward. Namely, a structural misalignment obtains between these earnings forecasts and their subsequent lower realizations. In this paper, we develop a novel machine learning forecast algorithm that is statistically optimal, unbiased, and immune to variable selection bias.

This new measure is useful not only as an input to asset-pricing exercises but also as an available real-time benchmark against which other forecasts can be compared. We can therefore construct a real-time measure of analyst biases both in the time series and the cross-section. We find that these biases exhibit considerable variation in both dimensions. Further, cross-sectional asset-pricing sorts based on this real-time measure of analyst biases show that stocks for which the earnings forecast is the most upward- (downward-) biased earn lower (higher) average returns going forward. This finding indicates that the analyst forecast errors may have a nontrivial effect on asset prices.

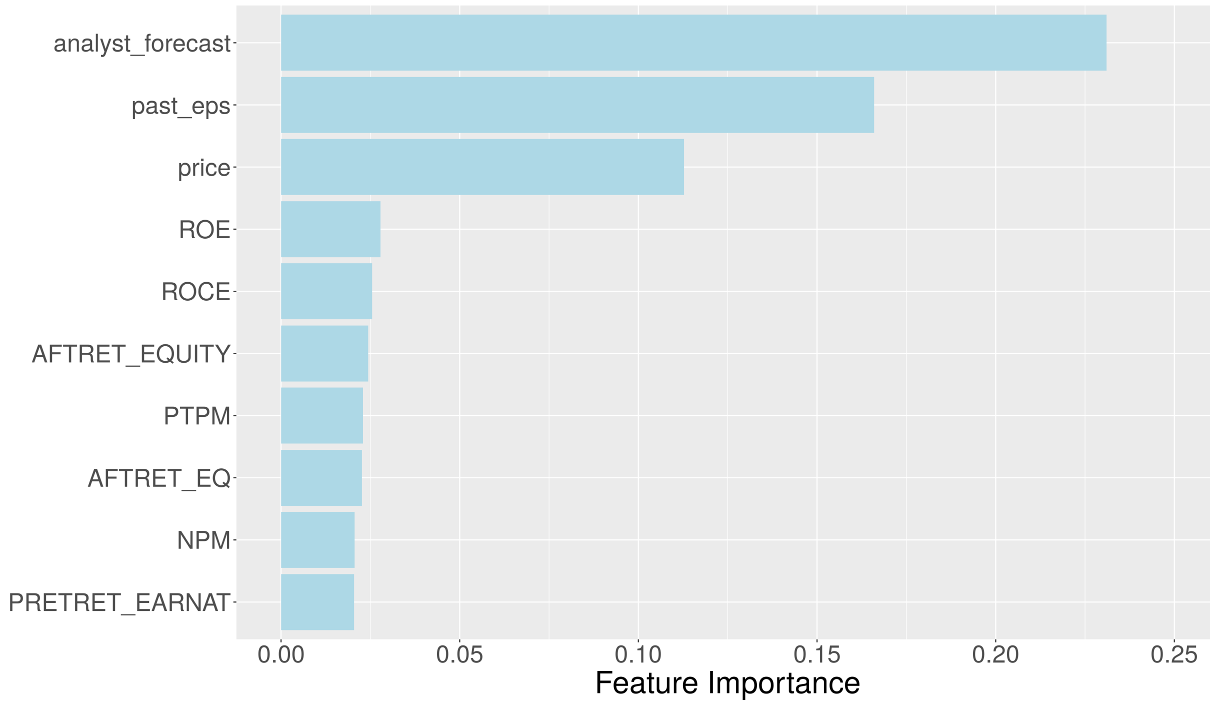
In addition to these asset-pricing results, our findings also have critical implications in corporate finance. Managers of firms for which the earnings forecast is most upward-biased issue more stocks. This finding indicates that managers are at least partially aware of analyst biases or the associated influence on asset prices. This study applies our machine learning approach to earnings, and the approach can easily be extended to other variables.

Figure 1: Feature importance of the one-quarter-ahead forecast



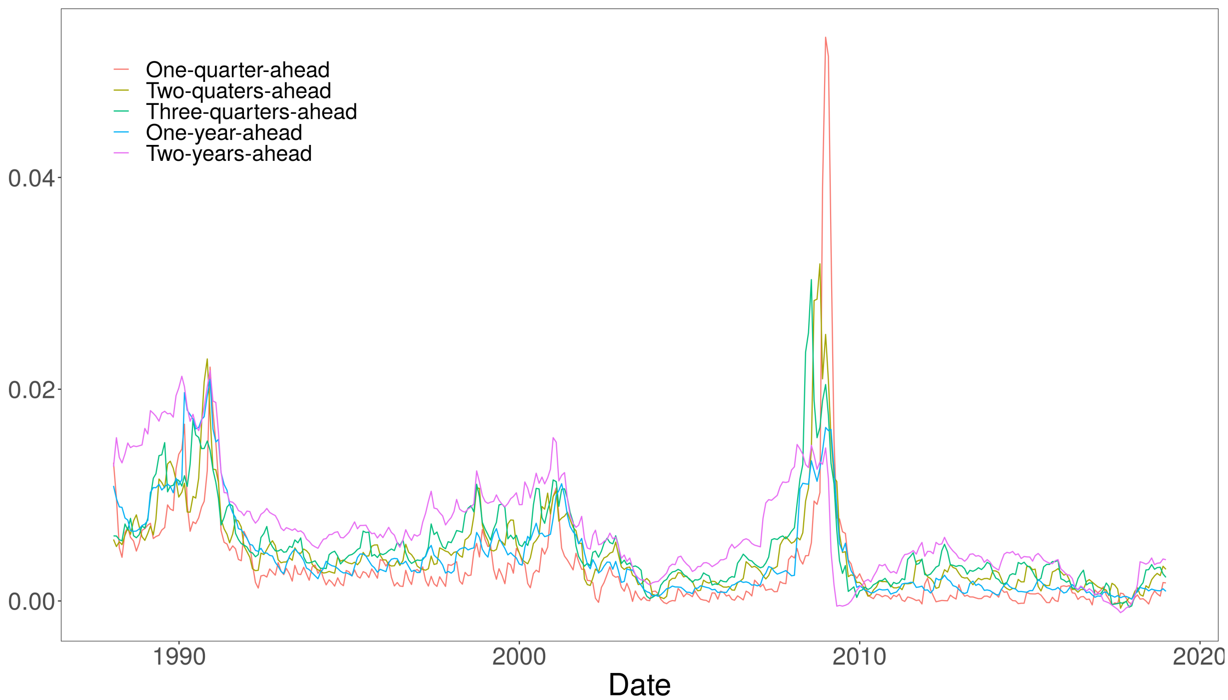
Notes: This figure plots the time-series average of feature importance of the 10 most important variables for the one-quarter-ahead earnings forecasts. The feature importance for each variable is the normalized sum of the reduced mean squared error decrease when splitting on that variable using the method in Nembrini et al. (2018). The feature importance of each variable is normalized so that the features' importance sums up to one.

Figure 2: Feature importance of the one-year-ahead forecast



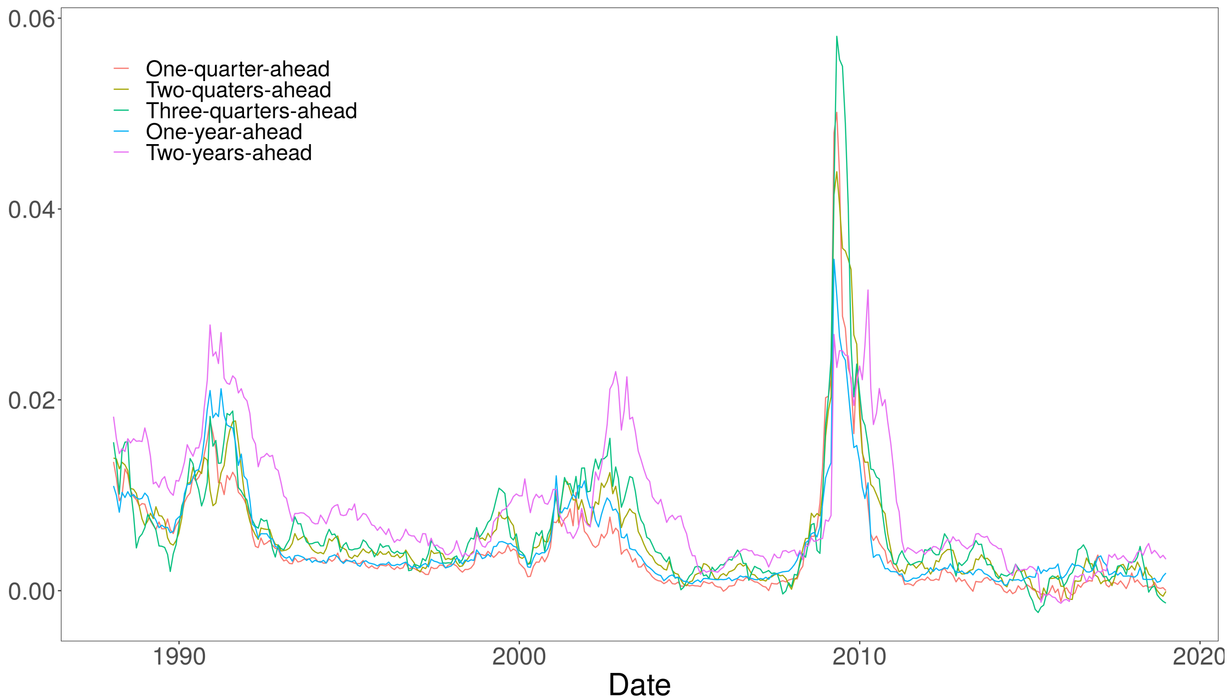
Notes: This figure plots the time-series average of feature importance of the 10 most important variables for the one-year-ahead earnings forecasts. The feature importance for each variable is the normalized sum of the reduced mean squared error decrease when splitting on that variable using the method in Nembrini et al. (2018). The feature importance of each variable is normalized so that the features' importance sums up to one.

Figure 3: Average realized bias of analysts' earnings expectations



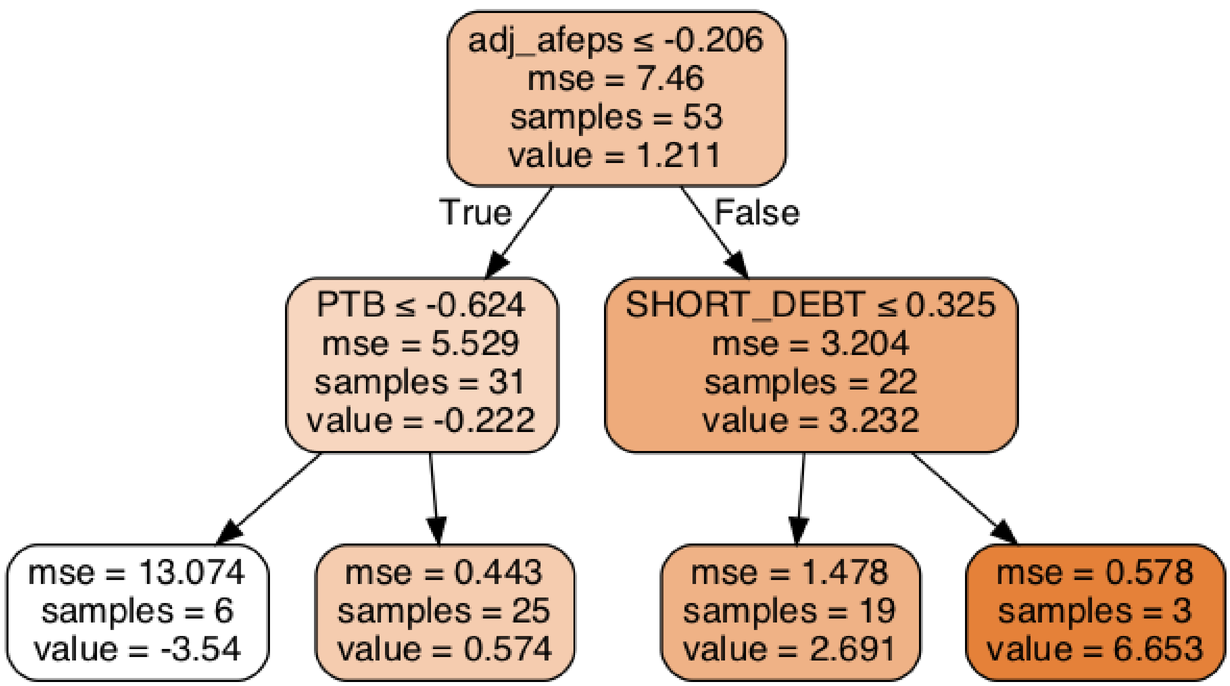
Notes: This figure plots the realized bias of analysts' expectations, which is measured as the average of the bias of expectations of individual firms. We trim the data at the 1% level each period before taking the average. The bias is calculated as the difference between analysts' earnings forecast and the realized value, scaled by the stock price from the most recent period. To ensure the annual earnings forecasts have the same scale as quarterly forecasts, we divide annual forecasts by four.

Figure 4: Average bias of analysts' earnings expectations relative to machine learning forecasts



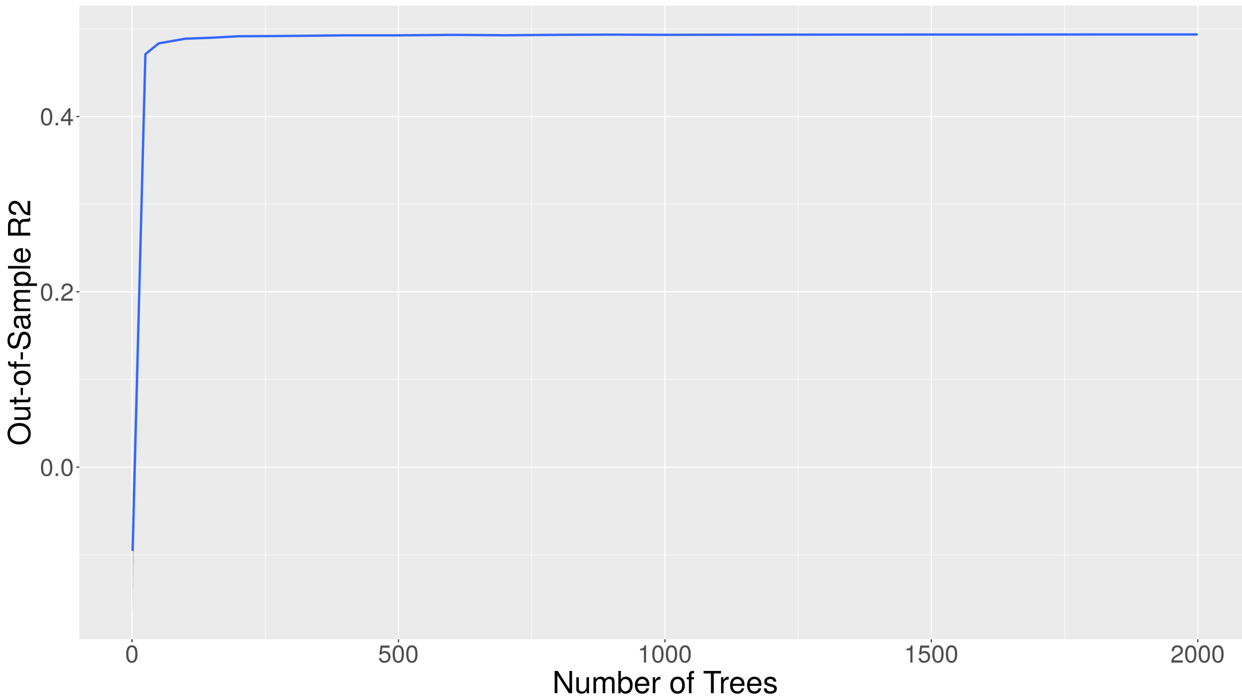
Notes: This figure plots the conditional bias of analysts' expectations, which is measured as the average of the bias of expectations of individual firms. We trim the data at the 1% level each period before taking the average. The bias is calculated as the difference between analysts' earnings forecast and the machine learning forecast, scaled by the stock price from the most recent period. To ensure the annual earnings forecasts have the same scale as quarterly forecasts, we divide annual forecasts by four.

Figure 5: Example Decision Tree



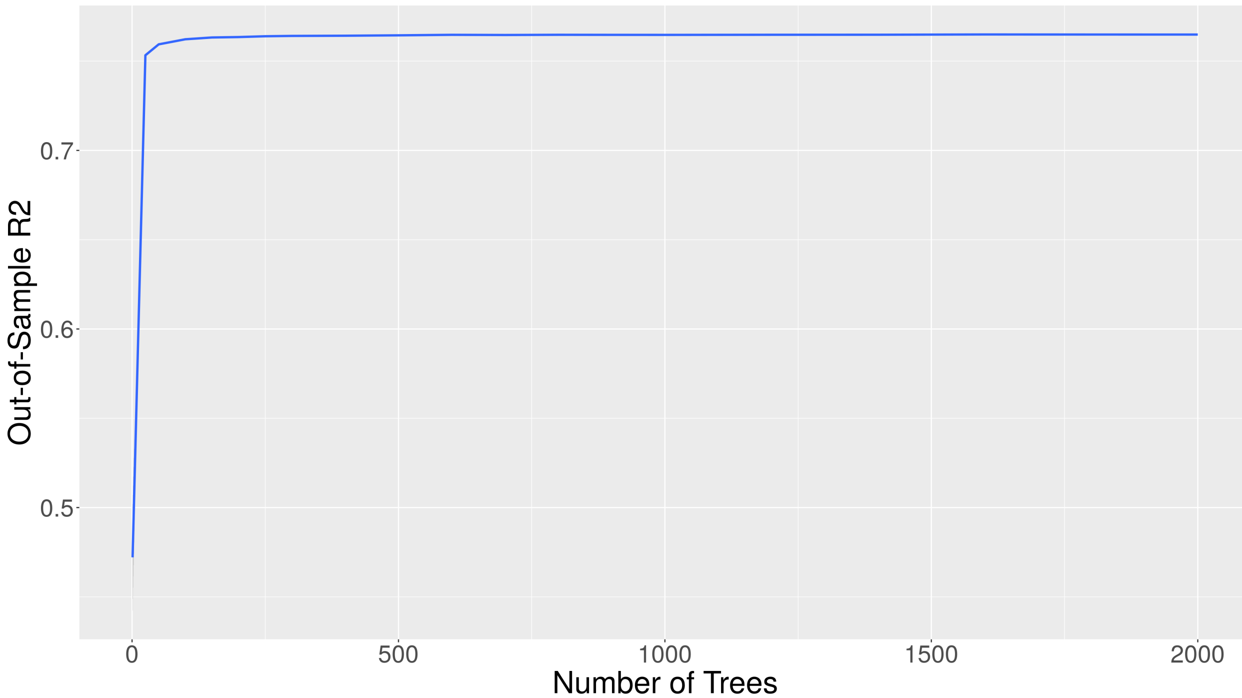
Notes: This Figure shows an example decision tree. The variable we wish to forecast is the earnings-per-share (eps hereafter) for a cross-section of firms. At the first step, the selected explanatory variable is the analysts' forecast (denoted by adj_afeps), and the threshold (or cutoff) value is at -0.206. Were we to end at this step, the forecast eps-value is -0.222 when adj_afeps is less than or equal to -0.206, and 3.232 when adj_afeps is less than or equal to -0.206. In the next step, the algorithm splits each of the previous two sub-spaces in two again. The first subspace (analysts' forecast less than or equal to -0.206) is split in two using the price-to-book ratio (PTB) as an explanatory variable. The threshold value is -0.624. The second subspace (analysts' forecast greater than -0.206) uses short-term debt. We then continue for the predefined number of splits until we arrive at the final nodes. In the final nodes, the prediction is the historical local average of that subspace.

Figure 6: Cross-validation results of the number of trees in the one-quarter-ahead forecast



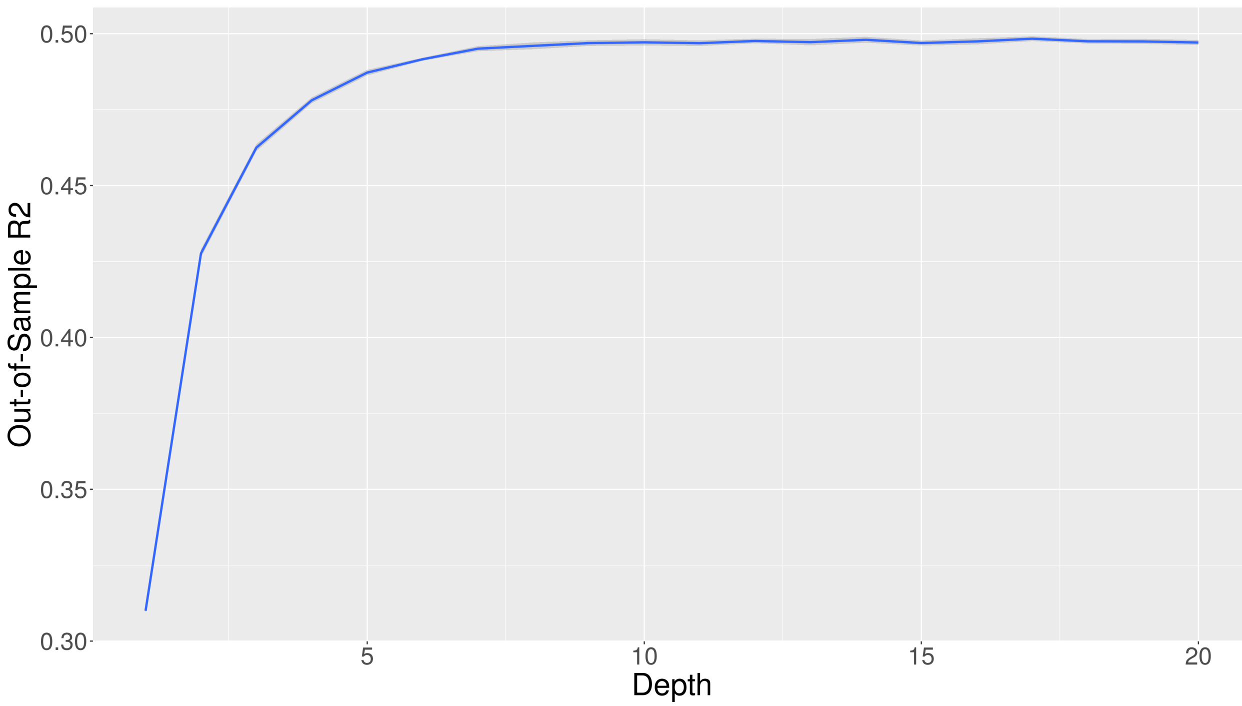
Notes: This figure plots the relation between the number of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-quarter-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using the machine learning forecast divided by the mean squared error of using the realized average value as a forecast. The random forest algorithm is random by design, so we take the average of 100 runs to measure the out-of-sample R^2.

Figure 7: Cross-validation Results of the number of trees in the one-year-ahead forecast



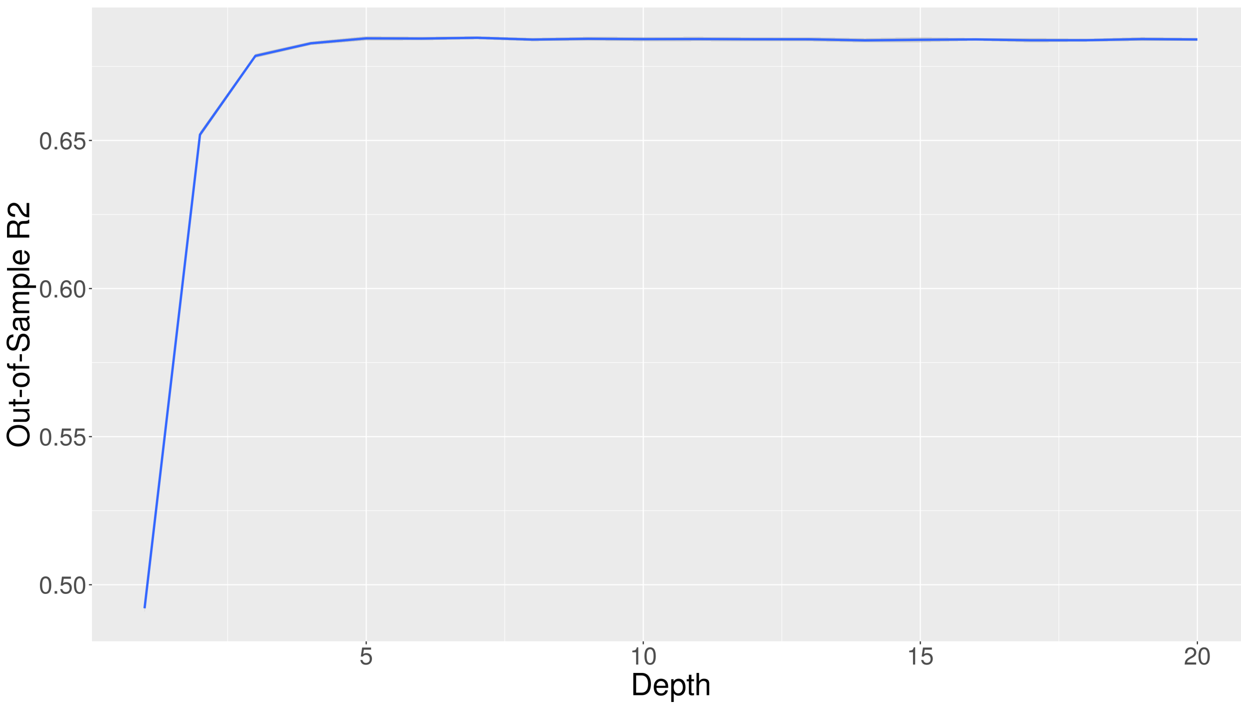
Notes: This figure plots the relation between the number of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-year-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using the machine learning forecast divided by the mean squared error of using the realized average value as a forecast. The random forest algorithm is random by design, so we take the average of 100 runs to measure the out-of-sample R^2.

Figure 8: Cross-validation results of the maximum depth of each tree in the one-quarter-ahead forecast



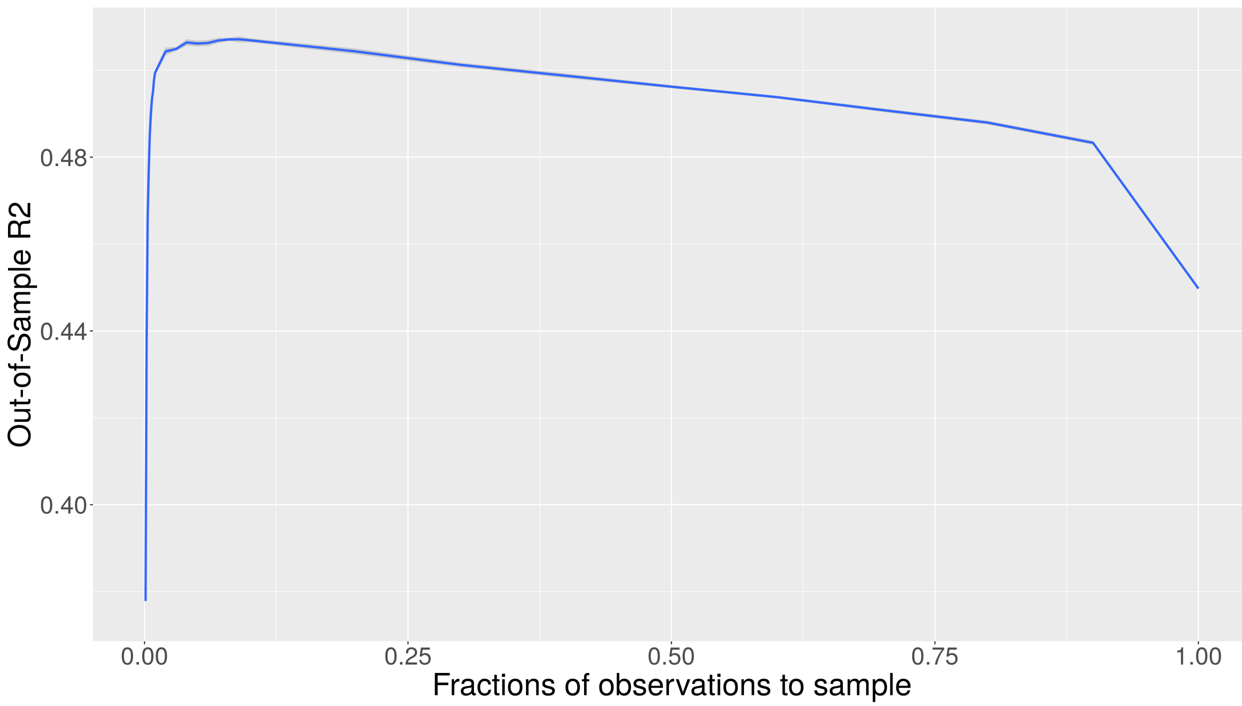
Notes: This figure plots the relation between the depth of of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-quarter-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using the machine learning forecast divided by the mean squared error of using the realized average value as a forecast. The random forest algorithm is random by design, so we take the average of 100 runs to measure the out-of-sample R^2.

Figure 9: Cross-validation results of the maximum depth of each tree in the one-year-ahead forecast



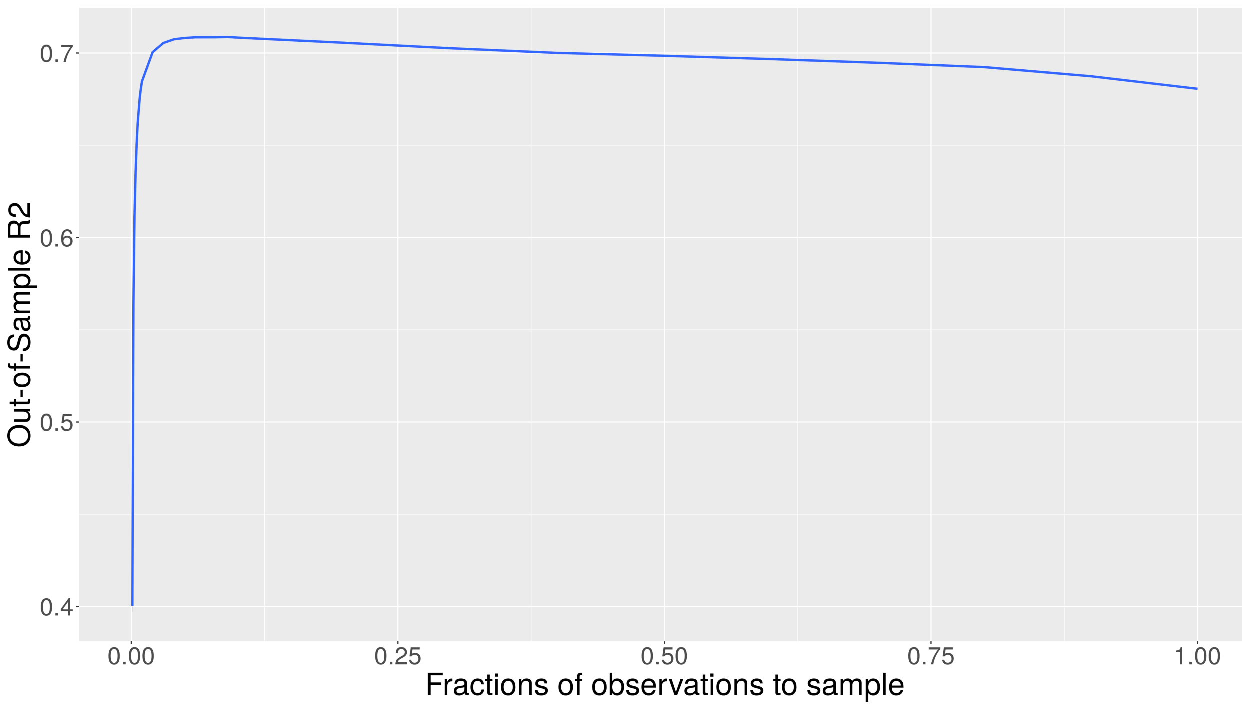
Notes: This figure plots the relation between the depth of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-year-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using the machine learning forecast divided by the mean squared error of using the realized average value as a forecast. The random forest algorithm is random by design, so we take the average of 100 runs to measure the out-of-sample R^2.

Figure 10: Cross-validation results of the fraction of the sample that is taken in each split in the one-quarter-ahead forecast



Notes: This figure plots the relation between the fraction of the sample that is taken in each split used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-quarter-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using the machine learning forecast divided by the mean squared error of using the realized average value as a forecast. The random forest algorithm is random by design, so we take the average of 100 runs to measure the out-of-sample R^2.

Figure 11: Cross-validation results of the fraction of the sample that is taken in each split in the one-year-ahead forecast



Notes: This figure plots the relation between the fraction of the sample that is taken in each split used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-year-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using the machine learning forecast divided by the mean squared error of using the realized average value as a forecast. The random forest algorithm is random by design, so we take the average of 100 runs to measure the out-of-sample R^2.

Table 1: Hyper-parameters for the Random Forest Regression

Notes: This table reports the parameters chosen for the random forest regression. Number of trees is the number of decision trees used. Maximum Depth is the maximum number of splits that each decision tree can use. Sample Fraction is the fraction of observations used to train each decision tree. The minimum node size is the threshold to stop the decision tree whenever the split would result in a sample size smaller than the minimum node size. The hyper-parameters are chosen using cross-validation over 1986 as detailed in the Appendix. The random forest regression is trained using rolling regressions keeping the hyper-parameters fixed.

Number of Trees	2000
Maximum Depth	7
Sample Fraction	1%
Minimum Node Size	5

Table 2: Updates in analysts' beliefs

Notes: This table presents the time-series average of aggregate analysts' forecast errors, defined as the differences between analysts' earnings forecasts and the realized actual earnings. $Month-ahead$ denotes the number of months from the time when analysts make forecasts until the fiscal year/quarter end. N denotes the number of observations. FE and sqr_FE denote the average forecast error and the average square of the error respectively. t-statistics of forecast errors are reported in the corresponding line. The sample period is 1987 January to 2019 December.

Panel A: One-quarter-ahead												
Month-ahead	1	0	-1									
N	377000	431530	389233									
FE	0.025	0.019	0.014									
t-statistic	6.81	6.07	4.95									
Sqr_FE	0.075	0.067	0.061									
Panel B: Two-quarters-ahead												
Month-ahead	4	3	2									
N	341911	386999	352291									
FE	0.053	0.049	0.047									
t-statistic	10.75	10.74	11.39									
Sqr_FE	0.106	0.100	0.094									
Panel C: Three-quarters-ahead												
Month-ahead	7	6	5									
N	308450	352834	322634									
FE	0.069	0.067	0.066									
t-statistic	12.01	12.47	13.21									
Sqr_FE	0.126	0.121	0.116									
Panel D: One-year-ahead												
Month-ahead	10	9	8	7	6	5	4	3	2	1	0	-1
N	64026	97403	112926	115749	117769	119325	120842	122121	123511	124963	126707	120647
FE	0.207	0.216	0.209	0.201	0.185	0.166	0.147	0.130	0.107	0.085	0.070	0.067
t-statistic	5.36	6.60	6.92	7.03	6.94	6.63	6.40	6.06	5.56	5.42	5.29	5.15
Sqr_FE	1.220	1.166	1.090	0.970	0.886	0.816	0.688	0.629	0.566	0.465	0.401	0.402
Panel E: Two-years ahead												
Month-ahead	22	21	20	19	18	17	16	15	14	13	12	11
N	47946	74093	87074	91623	94112	95916	97807	98796	99924	101029	101952	99765
FE	0.424	0.426	0.413	0.410	0.403	0.387	0.373	0.359	0.338	0.307	0.293	0.282
t-statistic	6.27	7.49	7.79	7.88	7.81	7.65	7.56	7.42	7.31	7.21	7.25	7.44
Sqr_FE	2.565	2.361	2.269	2.153	2.086	2.024	1.896	1.829	1.719	1.576	1.494	1.416

Table 3: The term structure of earnings forecasts via machine learning

Notes: This table presents the time series average of machine learning earnings per share forecasts (ML), analysts' earning forecasts (AF), actual realized earnings (AE) —the difference as well as the squared difference between them. N denotes the number of the sample stocks. The scale "P" denotes the stock price from the previous most recent month. We report the Newey-West t-statistics of differences between earnings forecasts and realized earnings. Because the earning forecasts are made monthly, we adjust the quarterly forecasts with three lags and the annual forecasts with 12 lags. The sample period is 1987 January to 2019 December.

Horizon	ML	AF	AE	(ML - AE)	(AF - AE)	(AF - ML)	$(ML-AE)^2$	$(AF-AE)^2$	(ML - AE)/P	(AF - AE)/P	(AF - ML)/P	N
One-quarter-ahead	0.291	0.312	0.294	-0.008	0.018	0.021	0.061	0.065	0.000	0.006	0.006	1300819
t-statistic				-0.997	3.385	2.753			0.358	3.594	2.882	
Two-quarters-ahead	0.305	0.351	0.307	-0.002	0.044	0.045	0.080	0.089	-0.001	0.006	0.007	1188374
t-statistic				-0.155	5.107	4.171			-0.226	4.530	3.516	
Three-quarters-ahead	0.323	0.384	0.324	-0.001	0.061	0.061	0.096	0.111	-0.001	0.006	0.008	1083205
t-statistic				-0.006	4.748	4.624			-0.824	6.562	0.774	
One-year-ahead	1.172	1.291	1.156	0.016	0.135	0.119	0.687	0.695	0.003	0.028	0.025	1429640
t-statistic				0.531	4.189	4.469			0.632	3.916	3.894	
Two-years-ahead	1.329	1.699	1.351	-0.022	0.348	0.370	1.555	1.836	-0.007	0.032	0.040	1156906
t-statistic				-0.195	4.501	3.856			-0.703	8.106	6.224	

Table 4: Fama-Macbeth regressions

Notes: This table reports the Fama-MacBeth cross-sectional regressions of monthly stocks' excess returns on the conditional bias in each forecast horizon, including one-quarter, two-quarters, three-quarters, one-year, and two-years-ahead. "Average BE" denotes the average of the conditional biases, defined as the difference between analysts' forecasts and the machine learning forecasts scaled by the share price from the most recent period, at different forecast horizons. "BE Score" denotes the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. (1) and (2) report the regression results with and without control variables, respectively. The control variables include log of firm size (Lnsize), log of book-to-market ratio (Lnbeme), short-term reversal (Ret_1), medium-term momentum (Ret12_7), investment-to-asset (IA),idiosyncratic volatility (IVOL), return volatility (Retvol), and share turnover (Turnover). We report the time-series average of slope coefficients associated with Fama-MacBeth t-statistics (in parentheses). The sample period is 1987 to 2019.

$$R_{i,t+1} = \alpha + \beta_1 BE_{i,t} + \gamma_i \sum_{i=1}^{8} Control_{i,t} + \epsilon_{i,t+1}$$

	Panel A: Average BE		Panel B: BE Score	
	(1)	(2)	(1)	(2)
BE	−0.0808	−0.0852	−0.0279	−0.0456
	(−4.61)	(−5.30)	(−6.57)	(−15.99)
Lnsize		−0.0009		−0.0029
		(−2.46)		(−8.37)
Lnbeme		0.0012		0.0019
		(2.00)		(3.16)
Ret12_7		0.0038		0.0011
		(2.44)		(0.73)
Ret1		−0.0284		−0.0313
		(−6.62)		(−7.29)
IA		−0.0007		−0.0007
		(−2.60)		(−2.73)
Ivol		−0.1941		−0.1743
		(−1.72)		(−1.53)
Retvol		0.1339		0.1982
		(1.13)		(1.67)
Turnover		−0.0006		−0.0005
		(−1.38)		(−1.18)
Intercept	0.0078	0.0213	0.0215	0.0675
	(2.74)	(3.98)	(8.70)	(13.69)
R^2	0.0105	0.0604	0.0156	0.0629

Table 5: Correlations between the conditional bias and characteristics

Notes: This table presents the time series averages of cross-sectional correlations between the conditional bias and characteristics. BE_Q1, BE_Q2, BE_Q3, BE_A1, and BE_A2 denote conditional biases in analysts' one-quarter two-quarters , three-quarters , one-year, and two-years-ahead earnings forecasts, respectively. "Average BE" denotes the average of the conditional bias at the different forecast horizons. "BE Score" denotes the average of the percentile ranking of the conditional bias of different forecast horizons. The characteristics include log of firm size (Lnsize), log of book-to-market ratio (Lnbeme), short-term reversal (Ret_1), medium-term momentum (Ret12_7), investment-to-asset (IA),idiosyncratic volatility (IVOL), return volatility (Retvol), and share turnover (Turnover). A * denotes that the correlation is not significant at the 1% level or more strict thresholds; all other correlations are significant. The t-statistics are adjusted by Newey-West standard errors. The sample period is 1987 to 2019.

	Average BE	B BE Score	BEQ1	BEQ2	BEQ3	BEA1	BEA2	Lnsize	Lnbeme	Ret12_7	Ret1	IA	Ivol	Retvol	Turnover
BE	1.000														
Average BE	0.459	1.000													
BE_Q1	0.756	0.390	1.000												
BE_Q2	0.827	0.413	0.578	1.000											
BE_Q3	0.835	0.418	0.480	0.565	1.000										
BE_A1	0.733	0.395	0.671	0.531	0.477	1.000									
BE_A2	0.885	0.448	0.529	0.626	0.618	0.546	1.000								
Lnsize	-0.280	-0.518	-0.259	-0.236	-0.216	-0.312	-0.249	1.000							
Lnbeme	0.071	0.125	0.078	0.071	0.059	0.056	0.064	-0.183	1.000						
Ret12_7	-0.134	-0.167	-0.130	-0.115	-0.104	-0.138	-0.109	0.130	-0.050	1.000					
Ret1	−0.011*	-0.043	0.004*	-0.029	-0.032	0.012*	−0.011*	0.074	0.014*	0.018	1.000				
IA	0.007*	0.025	0.001*	0.005*	0.010*	0.011*	0.012*	-0.036	-0.079	−0.012*	-0.011	1.000			
IVOL	0.325	0.394	0.317	0.272	0.243	0.361	0.277	-0.463	-0.057	-0.091	-0.023	0.055	1.000		
Retvol	0.313	0.379	0.305	0.261	0.233	0.347	0.267	-0.425	-0.070	-0.078	-0.024	0.055	0.975	1.000	
Turnover	0.024	0.003*	−0.001*	0.040	0.041	0.001*	0.026*	0.063	-0.174	0.101	0.006*	0.042	0.245	0.277	1.000

Table 6: Portfolios sorted on conditional bias

Notes: This table reports the time series average of excess returns (in percent) on value-weighted portfolios formed on the conditional bias in different forecast horizons. Panel A looks at "Average BE", defined as the average of conditional bias at different forecast horizons. Panel B presents the sorts based on "BE score", defined as the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. The sample period is 1987 to 2019.

Quintile	1	2	3	4	5	1-5
			Panel A: Average BE			
Mean	1.07	0.70	0.46	-0.04	-0.88	1.95
t-stat	5.03	3.17	1.82	-0.12	-2.05	5.88
CAPM Beta	0.92	0.98	1.11	1.28	1.58	-0.66
			Panel B: BE Score			
Mean	0.96	0.66	0.43	0.07	-0.57	1.53
t-stat	4.76	2.93	1.64	0.22	-1.38	4.90
CAPM Beta	0.89	1.01	1.14	1.28	1.53	-0.63

Table 7: Time series tests with common asset-pricing Models

Notes: This table reports the regression of stock returns (in percent) on the long-short portfolio sorted with the conditional bias, on the CAPM, the Fama-French three-factor model (FF3), and the Fama-French five-factor model (FF5). Panel A looks at average conditional bias at different forecast horizons. Panel B presents the sorts based on "BE score", defined as the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. The sample period is 1987 to 2019. The t-statistics are adjusted by the White's heteroscedasticity robust standard errors.

$$LS\_Port_t = \alpha + \sum_{i=1}^{5} \beta_i F_{i,t} + \epsilon_t$$

	CAPM		FF3		FF5	
	Coeffi	t-stat	*Coeffi*	t-stat	*Coeffi*	t-stat
Panel A: Average BE						
Intercept	2.39	8.15	2.52	9.70	2.02	7.21
Mkt_RF	-0.66	-7.81	-0.61	-7.52	-0.42	-5.34
SMB			-0.86	-6.33	-0.62	-4.33
HML			-0.60	-4.10	-1.01	-6.10
RMW					0.84	4.07
CMA					0.53	1.79
Panel B: BE Score						
Intercept	1.94	7.02	2.03	8.01	1.53	5.73
Mkt_RF	-0.63	-7.50	-0.56	-6.58	-0.37	-4.62
SMB			-0.83	-6.89	-0.57	-4.39
HML			-0.44	-3.07	-0.83	-4.93
RMW					0.90	4.63
CMA					0.48	1.63

Table 8: Conditional bias and anomalies

Notes: This table reports the conditional bias for portfolios formed by sorting independently on the average conditional bias (BE) and the anomaly score, defined as the equal-weighted average of the decile ranking on each of the 27 anomaly variables. Panel A looks at the time-series average of anomaly score of each portfolio. Panel B looks at the number of stocks in each portfolio. The sample period is 1987 to 2019.

	Anomaly Decile										
BE Quintile	S	2	3	4	5	6	7	8	9	L	L-S
					Panel A: Conditional Biases						
BE quintile	1	2	3	4	5	6	7	8	9	10	10-1
1	3.31	3.96	4.36	4.68	4.97	5.24	5.52	5.83	6.21	6.90	3.59
2	3.32	3.95	4.36	4.68	4.97	5.24	5.52	5.83	6.21	6.89	3.57
3	3.27	3.95	4.35	4.68	4.96	5.24	5.52	5.83	6.21	6.92	3.65
4	3.21	3.94	4.35	4.68	4.97	5.24	5.52	5.83	6.22	6.95	3.74
5	3.13	3.94	4.35	4.68	4.97	5.24	5.52	5.83	6.22	6.96	3.83
All stocks	3.21	3.95	4.35	4.68	4.97	5.24	5.52	5.83	6.21	6.92	3.72
					Panel B: Number of stocks						
1	32	43	47	52	57	59	61	63	62	58	
2	30	45	53	56	59	60	61	61	58	50	
3	45	54	55	56	56	56	54	53	53	52	
4	66	61	57	54	50	48	48	47	49	54	
5	93	65	55	49	45	43	42	42	46	53	
All stocks	266	267	267	267	267	267	267	267	267	266	

Table 9: Returns on portfolios formed on conditional bias and anomaly score

Notes: This table reports the time-series average of value-weighted excess returns on portfolios formed by sorting independently on the average conditional bias (BE) and the anomaly score, defined as the equal-weighted average of the decile ranking on each of the 27 anomaly variables. The last two columns report the conditional bias (with Newey-West t-statistic) of the ten decile portfolios formed on the anomaly score.

	Anomaly Decile										
BE Quintile	S	2	3	4	5	6	7	8	9	L	L-S
1	1.24	1.00	1.12	1.30	1.09	1.20	1.14	1.27	1.26	1.54	0.30
t-statistic	3.20	3.18	4.06	5.13	4.44	4.92	5.12	5.99	5.92	6.56	0.95
2	0.02	0.58	0.58	0.71	0.82	0.64	0.85	0.82	1.05	0.92	0.90
t-statistic	0.07	1.80	2.21	2.74	3.18	2.72	3.68	3.67	4.72	3.98	2.97
3	-0.28	-0.08	0.45	0.43	0.47	0.64	0.57	0.72	0.78	0.85	1.12
t-statistic	-0.71	-0.26	1.42	1.38	1.59	2.28	2.24	2.71	3.14	3.26	3.86
4	-0.70	-1.04	-0.12	-0.02	-0.20	0.31	0.11	0.38	0.56	0.54	1.24
t-statistic	-1.63	-2.57	-0.33	-0.07	-0.55	0.94	0.33	1.20	1.72	1.59	3.89
5	-2.27	-1.54	-1.14	-1.07	-0.51	-0.16	-0.10	-0.59	-0.09	-0.05	2.22
t-statistic	-4.39	-2.88	-2.46	-2.26	-1.03	-0.32	-0.21	-1.18	-0.19	-0.09	6.11
All Stocks	S	2	3	4	5	6	7	8	9	L	L-S
Excess Return	-0.27	0.11	0.52	0.62	0.61	0.69	0.72	0.76	0.96	0.98	1.25
t-statistic	-0.77	0.36	2.01	2.55	2.52	3.04	3.44	3.76	4.82	4.47	5.23
Average BE	0.014	0.010	0.009	0.008	0.008	0.007	0.007	0.007	0.007	0.008	-0.006
t-statistic	6.48	6.19	6.27	6.09	5.71	6.01	5.82	5.97	6.19	6.06	-5.75

Table 10: Net stock sssuances and conditional biases

Notes: Panel A reports the time series average of net stock issuances of value-weighted portfolios sorted on the conditional bias. "Average BE" denotes the average of the conditional bias at different forecast horizons."BE score" denotes the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. Panel B reports the Fama-MacBeth regressions of firms' net stock issuances on the conditional bias and control variables include the log of firm size (Lnsize), the log of book-to-market ratio (Lnbeme), and earnings before interest, taxes, and depreciation divided by total assets (EBITDA). The sample period is 1987 to 2019. We report the time series average of slope coefficients associated with Newey-West t-statistics.

$$NSI_{i,t+1} = \alpha + \beta_1 BE_{i,t} + \gamma_i \sum_{i=1}^{3} Control_{i,t} + \epsilon_{i,t+1}$$

Panel A: Net Stock Issuances of Portfolios formed on BE						
Quintile	1	2	3	4	5	5-1
Average BE	0.013	0.011	0.017	0.040	0.073	0.060
t-stat	1.82	1.82	3.33	4.31	5.32	3.44
BE score	0.009	0.016	0.020	0.033	0.066	0.058
t-stat	1.33	2.14	3.69	5.17	4.18	3.39

Panel B: Fama-MacBeth regressions				
	A: Average BE		B: BE Score	
	(1)	(2)	(1)	(2)
BE	1.7048	1.2870	0.1191	0.0510
t-stat	3.86	4.53	6.74	4.82
Lnsize		-0.0053		-0.0051
t-stat		-3.25		-2.76
Lnbeme		-0.0239		-0.0230
t-stat		-6.10		-5.70
EBITDA		-0.1086		-0.1129
t-stat		-4.36		-4.34
Intercept	0.0621	0.1186	0.0108	0.0775
t-stat	6.12	3.72	0.95	2.46
R^2	0.0178	0.0921	0.0084	0.0750

References

Antoniou, C., Doukas, J. A. and Subrahmanyam, A. (2015). Investor sentiment, beta, and the cost of equity capital, *Management Science* **62**: 347–367.

Babii, A., Ball, R. T., Ghysels, E., Striaukas, J. et al. (2020). Machine learning panel data regressions with an application to nowcasting price earnings ratios, *Working Paper* .

Baker, M. and Wurgler, J. (2002). Market timing and capital structure, *The journal of finance* **57**(1): 1–32.

Baker, M. and Wurgler, J. (2013). Behavioral corporate finance: An updated survey, *Handbook of the Economics of Finance*, Vol. 2, Elsevier, pp. 357–424.

Balakrishnan, K., Bartov, E. and Faurel, L. (2010). Post loss/profit announcement drift, *Journal of Accounting and Economics* **50**(1): 20–41.

Ball, R. T. and Ghysels, E. (2018). Automated earnings forecasts: Beat analysts or combine and conquer?, *Management Science* **64**(10): 4936–4952.

Barth, M. E., Elliott, J. A. and Finn, M. W. (1999). Market rewards associated with patterns of increasing earnings, *Journal of Accounting Research* **37**(2): 387–413.

Basu, S. (1983). The relationship between earnings' yield, market value and return for nyse common stocks: Further evidence, *Journal of financial economics* **12**(1): 129–156.

Belo, F. and Lin, X. (2012). The inventory growth spread, *The Review of Financial Studies* **25**(1): 278–313.

Bianchi, F., Ludvigson, S. C. and Ma, S. (2020). Belief distortions and macroeconomic fluctuations, *Working Paper 27406*, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w27406*

Bordalo, P., Gennaioli, N., LaPorta, R. and Shleifer, A. (2020). Expectations of fundamentals and stock market puzzles, *Working Paper* .

Bordalo, P., Gennaioli, N., Ma, Y. and Shleifer, A. (2018). Over-reaction in macroeconomic expectations, *Working Paper 24932*, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w24932*

Bordalo, P., Gennaioli, N., Porta, R. L. and Shleifer, A. (2019). Diagnostic expectations and stock returns, *Journal of Finance* **74**(6): 2839–2874.

Bouchaud, J.-p., Krueger, P., Landier, A. and Thesmar, D. (2019). Sticky expectations and the profitability anomaly, *The Journal of Finance* **74**(2): 639–674.

Boudoukh, J., Michaely, R., Richardson, M. and Roberts, M. R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing, *The Journal of Finance* **62**(2): 877–915.

Breiman, L. (2001). Random forests, *Mach. Learn.* **45**(1): 5–32.
URL: *https://doi.org/10.1023/A:1010933404324*

Bryzgalova, S., Pelger, M. and Zhu, J. (2020). Forest through the trees: Building cross-sections of stock returns, *working paper* .

Chan, L. K., Jegadeesh, N. and Lakonishok, J. (1996). Momentum strategies, *The Journal of Finance* **51**(5): 1681–1713.

Chan, L. K., Lakonishok, J. and Sougiannis, T. (2001). The stock market valuation of research and development expenditures, *The Journal of Finance* **56**(6): 2431–2456.

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts, *American Economic Review* **105**(8): 2644–78.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.20110306*

Cooper, M. J., Gulen, H. and Schill, M. J. (2008). Asset growth and the cross-section of stock returns, *The Journal of Finance* **63**(4): 1609–1651.

Daniel, K., Hirshleifer, D. and Sun, L. (2017). Short and long horizon behavioural factors, *National Bureau of Economic Research* .

Daniel, K. and Titman, S. (2006). Market reactions to tangible and intangible information, *The Journal of Finance* **61**(4): 1605–1643.

Dechow, P. M., Sloan, R. G. and Soliman, M. T. (2004). Implied equity duration: A new measure of equity risk, *Review of Accounting Studies* **9**(2-3): 197–228.

Diether, K. B., Malloy, C. J. and Scherbina, A. (2002). Differences of opinion and the cross section of stock returns, *The Journal of Finance* **57**(5): 2113–2141.

Eisfeldt, A. L. and Papanikolaou, D. (2013). Organization capital and the cross-section of expected returns, *The Journal of Finance* **68**(4): 1365–1406.

Engelberg, J., McLean, R. D. and Pontiff, J. (2018). Anomalies and news, *The Journal of Finance* **73**(5): 1971–2001.

Fama, E. F. and French, K. R. (1996). Multifactor explanations of asset pricing anomalies, *The journal of finance* **51**(1): 55–84.

Fama, E. F. and French, K. R. (2006). Profitability, investment and average returns, *Journal of financial economics* **82**(3): 491–518.

Foster, G., Olsen, C. and Shevlin, T. (1984). Earnings releases, anomalies, and the behavior of security returns, *Accounting Review* pp. 574–603.

Freyberger, J., Neuhierl, A. and Weber, M. (2020). Dissecting characteristics nonparametrically, *The Review of Financial Studies* **33**(5): 2326–2377.

Green, J., Hand, J. R. and Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns, *The Review of Financial Studies* **30**(12): 4389–4436.

Griffin, J. M., Harris, J. H., Shu, T. and Topaloglu, S. (2011). Who drove and burst the tech bubble?, *The Journal of Finance* **66**(4): 1251–1290.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2011.01663.x*

Gu, S., Kelly, B. and Xiu, D. (2020). Empirical asset pricing via machine learning, *The Review of Financial Studies* **33**(5): 2223–2273.

Hafzalla, N., Lundholm, R. and Matthew Van Winkle, E. (2011). Percent accruals, *The Accounting Review* **86**(1): 209–236.

Haugen, R. A. and Baker, N. L. (1996). Commonality in the determinants of expected stock returns, *Journal of Financial Economics* **41**(3): 401–439.

Hirshleifer, D., Hou, K., Teoh, S. H. and Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets?, *Journal of Accounting and Economics* **38**: 297–331.

Hirshleifer, D. and Jiang, D. (2010). A financing-based misvaluation factor and the cross-section of expected returns, *The Review of Financial Studies* **23**(9): 3401–3436.

Hong, H. and Sraer, D. (2016). Speculative betas, *Journal of Finance* **71**: 2095–2144.

Hou, K., Van Dijk, M. A. and Zhang, Y. (2012). The implied cost of capital: A new approach, *Journal of Accounting and Economics* **53**(3): 504–526.

Hou, K., Xue, C. and Zhang, L. (2015). Digesting anomalies: An investment approach, *Review of Financial Studies* **28**: 650–705.

Hribar, P. and McInnis, J. (2012). Investor sentiment and analysts' earnings forecast errors, *Management Science* **58**(2): 293–307.

Hughes, J., Liu, J. and Su, W. (2008). On the relation between predictable market returns and predictable analyst forecast errors, *Review of Accounting Studies* **13**(2-3): 266–291.

Ishwaran, H. (2015). The effect of splitting on random forests, *Mach. Learn.* **99**(1): 75–118.
URL: *https://doi.org/10.1007/s10994-014-5451-2*

Kothari, S. P., So, E. and Verdi, R. (2016). Analysts' forecasts and asset pricing: A survey, *Annual Review of Financial Economics* **8**: 197–219.

Kozak, S., Nagel, S. and Santosh, S. (2018). Interpreting factor models, *The Journal of Finance* **73**(3): 1183–1223.

La Porta, R. (1996). Expectations and the cross-section of stock returns, *The Journal of Finance* **51**(5): 1715–1742.

Lakonishok, J., Shleifer, A. and Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk, *Journal of Finance* **49**: 1541–1578.

Lyandres, E., Sun, L. and Zhang, L. (2007). The new issues puzzle: Testing the investment-based explanation, *The Review of Financial Studies* **21**(6): 2825–2855.

Nembrini, S., König, I. R. and Wright, M. N. (2018). The revival of the Gini importance?, *Bioinformatics* **34**(21): 3711–3718.
URL: *https://doi.org/10.1093/bioinformatics/bty373*

Novy-Marx, R. (2010). Operating leverage, *Review of Finance* **15**(1): 103–134.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium, *Journal of Financial Economics* **108**(1): 1–28.

Pontiff, J. and Woodgate, A. (2008). Share issuance and cross-sectional returns, *The Journal of Finance* **63**(2): 921–945.

Rosenberg, B., Reid, K. and Lanstein, R. (1985). Persuasive evidence of market inefficiency, *The Journal of Portfolio Management* **11**(3): 9–16.

Scherbina, A. (2007). Suppressed Negative Information and Future Underperformance*, *Review of Finance* **12**(3): 533–565.
URL: *https://doi.org/10.1093/rof/rfm028*

Scherbina, A. D. (2004). Analyst disagreement, forecast bias and stock returns, *working paper* .

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings?, *Accounting review* pp. 289–315.

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.

Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors, *Review of Financial Studies* **30**: 1270–1315.

Thomas, J. K. and Zhang, H. (2002). Inventory changes and future returns, *Review of Accounting Studies* **7**(2-3): 163–187.

Xing, Y. (2007). Interpreting the value effect through the q-theory: An empirical investigation, *The Review of Financial Studies* **21**(4): 1767–1795.

Zhou, G. (2018). Measuring investor sentiment, *Annual Review of Financial Economics* **10**: 239–259.

Appendix

A1. Sample selection and machine learning tests

In this section, we detail the sample selection and the procedures of machine learning earnings forecasts.

Our first step is to obtain actual realized earnings and analysts' earnings forecasts from the I/B/E/S database.[27] We keep firms that have both realized earnings and analysts forecasts. We focus on one-year- and two-years-ahead forecasts for annual earnings (IBES FPI of 1 and 2), and one-quarter-, two-quarters-, and three-quarters-ahead forecasts for quarterly earnings (IBES FPI of 6, 7, and 8), because analysts' forecasts for other horizons have significantly fewer observations.

We then match the IBES actual file (actual realized earnings) with the summary file (analysts' consensus forecasts) using Ticker and fiscal end date.[28] As pointed out by Diether et al. (2002) and Bouchaud et al. (2019), mistakes occur when matching I/B/E/S actual file with I/B/E/S summary file, because stock splits may occur between the earnings forecast day and the actual earnings announcement day. However, the I/B/E/S adjusted summary files round the forecast and actual earnings to the nearest penny for adjusting the splits. To circumvent these rounding errors, we obtain data from unadjusted actual and summary files. We use the cumulative adjustment factors (CFACSHR) from the CRSP monthly stock file to adjust the forecast and the actual EPS on the same share basis. For example, if forecasts are made at $t-1$ and the actual earnings are announced at t, we measure the adjusted actual earning as,

$$AdjustActual_t = Actual_t * CFACSHR_{t-1}/CFACSHR_t$$

[27]We do not obtain the actual earnings from Compustat, because I/B/E/S use different accounting basis from Compustat to measure actual earnings. Since or primary goal is to construct a statistically optimal and unbiased benchmark for analysts' earnings forecasts, we obtain the realized earnings from /I/B/E/S database.

[28]PENDS denotes the fiscal end date in the actual file and FPEDATS denotes the fiscal end date in the summary file.

For matching /I/B/E/S with CRSP, we follow the guidance provided by the Wharton Research Data Service to create a link table that uses CUSIP. We require firms' historical CUSIP to be same in both /I/B/E/S and CRSP. [29] We keep common stocks (share code 10 and 11) in stock exchanges of NYSE, AMEX, and NASDAQ (exchange code 1, 2, and 3).[30].

Our sample is in monthly frequency, because analysts make earnings forecasts for firms' earnings every month (I/B/E/S estimate date is STATPERS). We therefor provide our statistically optimal forecast for every I/B/E/S estimate date (STATPERS). Specifically, we assume that we are making forecasts at the same date as when analysts make forecasts. We trained the Random Forest model using the information available at the current time, and then forecast earnings in the same fiscal end periods as analysts do. When matching the forecasts variables such as firm characteristics and macroeconomic variables, we require announcement dates of these information are before STATPERS. The forecasts are therefore out-of-sample and are not based on any future information. The resulting forecasting regression is:

$$E_t[eps_{i,t+\tau}] = RF[Fundamentals_{i,t}, Macro_t, AF_{i,t}].$$

RF denotes the random forest model using data from the most recent periods. $Fundamentals_{i,t}$, $Macro_t$, and $AF_{i,t}$ denote firm fundamental variables, macroeconomic variables, and analysts' earnings forecasts respectively. The earnings per share of firm i in quarter t+ τ (τ=1 to 3) or year t+τ (τ=1 to 2) is $eps_{i,t+\tau}$.

For the quarterly earnings forecasts and one-year ahead forecast, we trained the Random Forest model using the data from the most recent year and then forecast earnings in the following periods using information available at the current time. For the two-year ahead forecasts, we trained the model using the data from the two most recent years rather than from the most recent year, because we do not have enough observations when using a 12-month window to train the model. Our forecasts remain consistent when using different

[29] Matching details can be found via " https://wrds-www.wharton.upenn.edu/pages/support/applications/linking-databases/linking-ibes-and-crsp-data/".

[30] We do not delete the smallest firms, because the smallest firms are simply not covered in /I/B/E/S and the intersection of /I/B/E/S and CRSP heavily tilt towards big stocks (Diether et al. (2002))

windows to train the model. Our training data starts in 1986 January, and our first forecast observations are in 1987 January.

A2. WRDS financial ratios

In the Random Forest model, we use financial ratios obtained from the Financial Ratio Suit by Wharton Research Data Service (WRDS) as forecasting variables. According to WRDS, these variables are most commonly used financial ratios by academic researchers and available at both quarterly and annual frequency. The variables can be grouped into the following seven categories: Capitalization, Efficiency, Financial Soundness/Solvency, Liquidity, Profitability, Valuation and Others. Table A1 details the definitions of financial ratios.[31]

We exclude PEG_1yrforward, PEG_ltgforward, pe_op_basic, pe_op_dil from our forecast model, because these variables have too many missing observations. We replace the missing values of other variables as the industry medians. The industries are defined as in Fama-French 49 industry portfolios.

Table A1: WRDS financial ratios

Variable	Definition	Variable	Definition
Accrual	Accruals/Average Assets	invt_act	Inventory/Current Assets
adv_sale	Avertising Expenses/Sales	lt_debt	Long-term Debt/Total Liabilities
aftret_eq	After-tax Return on Average Common Equity	lt_ppent	Total Liabilities/Total Tangible Assets
aftret_equity	After-tax Return on Total Stockholders Equity	npm	Net Profit Margin
aftret_invcapx	After-tax Return on Invested Capital	ocf_lct	Operating CF/Current Liabilities
at_turn	Asset Turnover	opmad	Operating Profit Margin After Depreciation
bm	Book/Market	opmbd	Operating Profit Margin Before Depreciation
capei	Shillers Cyclically Adjusted P/E Ratio	pay_turn	Payables Turnover
capital_ratio	Capitalization Ratio	pcf	Price/Cash flow
cash_conversion	Cash Conversion Cycle (Days)	pe_exi	P/E (Diluted, Excl. EI)
cash_debt	Cash Flow/Total Debt	pe_inc	P/E (Diluted, Incl. EI)
cash_lt	Cash Balance/Total Liabilities	pe_op_basic	Price/Operating Earnings (Basic, Excl. EI)
cash_ratio	Cash Ratio	pe_op_dil	Price/Operating Earnings (Diluted, Excl. EI)
cfm	Cash Flow Margin	PEG_1yrforward	Forward P/E to 1-year Growth (PEG) ratio
curr_debt	Current Liabilities/Total Liabilities	PEG_ltgforward	Forward P/E to Long-term Growth (PEG) ratio
curr_ratio	Current Ratio	PEG_trailing	Trailing P/E to Growth (PEG) ratio
de_ratio	Total Debt/Equity	pretret_earnat	Pre-tax Return on Total Earning Assets
debt_assets	Total Debt/Total Assets	pretret_noa	Pre-tax return on Net Operating Assets
debt_at	Total Debt/Total Assets	profit_lct	Profit Before Depreciation/Current Liabilities
debt_capital	Total Debt/Capital	ps	Price/Sales
debt_ebitda	Total Debt/EBITDA	ptb	Price/Book
debt_invcap	Long-term Debt/Invested Capital	ptpm	Pre-tax Profit Margin
divyield	Dividend Yield	quick_ratio	Quick Ratio (Acid Test)
dltt_be	Long-term Debt/Book Equity	RD_SALE	Research and Development/Sales
dpr	Dividend Payout Ratio	rect_act	Receivables/Current Assets

[31] The formulas to calculate these financial ratios are available at the WRDS website.

continued from previous page

efftax	Effective Tax Rate	rect_turn	Receivables Turnover
equity_invcap	Common Equity/Invested Capital	roa	Return on Assets
evm	Enterprise Value Multiple	roce	Return on Capital Employed
fcf_ocf	Free Cash Flow/Operating Cash Flow	roe	Return on Equity
gpm	Gross Profit Margin	sale_equity	Sales/Stockholders Equity
GProf	Gross Profit/Total Assets	sale_invcap	Sales/Invested Capital
int_debt	Interest/Average Long-term Debt	sale_nwc	Sales/Working Capital
int_totdebt	Interest/Average Total Debt	short_debt	Short-Term Debt/Total Debt
intcov	After-tax Interest Coverage	staff_sale	Labor Expenses/Sales
intcov_ratio	Interest Coverage Ratio	totdebt_invcap	Total Debt/Invested Capital
inv_turn	Inventory Turnover		

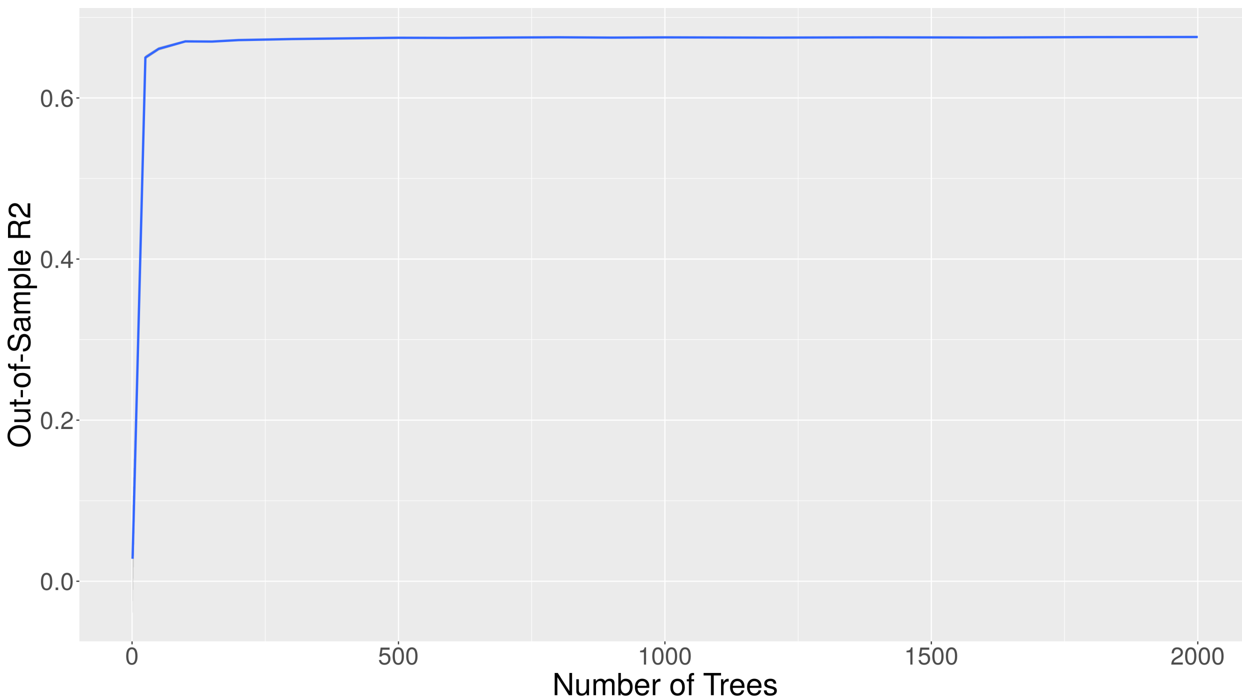
A2. Parameters in Random Forest

We choose the hyper-parameters in a purely data-driven way using cross-validation. We use data up to (and including) 1986 by dividing the data into into two partitions: training and testing (cross-validation). The training data contains the beginning of the sample: from the beginning of the sample until January 1986. The testing data contains a single month: February 1986. The results are similar for other testing periods in 1986. We train the model using the training data for different configurations of the hyper parameters. We evaluate the results in the testing data and pick the parameters that result in the best performance. Notice that the testing data is not using information from future periods. We maintain the hyper parameters chosen in 1986 for the whole sample, and we start our forecasts in 1987. The model is then trained using rolling windows keeping the hyper parameters fixed.

We choose 2000 trees from the cross-validation procedure but remark that there is little difference after 500. We use the recommended minimum node size of 5. We find that there are no significant differences in the out-of-sample R2 and even a slight reduction after a depth of seven so we choose that parameter. The result is explained in the following way: we train using a rolling window of 12 months for a total of around 10,000 observations. Since each split divides the data into two and we use a minimum node of 5, the maximum number of splits is 10 since $\frac{10^3}{2^{10}} = 9.77$. Figure A1, Figure A2, and Figure A3 show the cross-validation results for the first-period two-quarters-ahead, three-quarters-ahead, and two-years-ahead earnings forecasts.

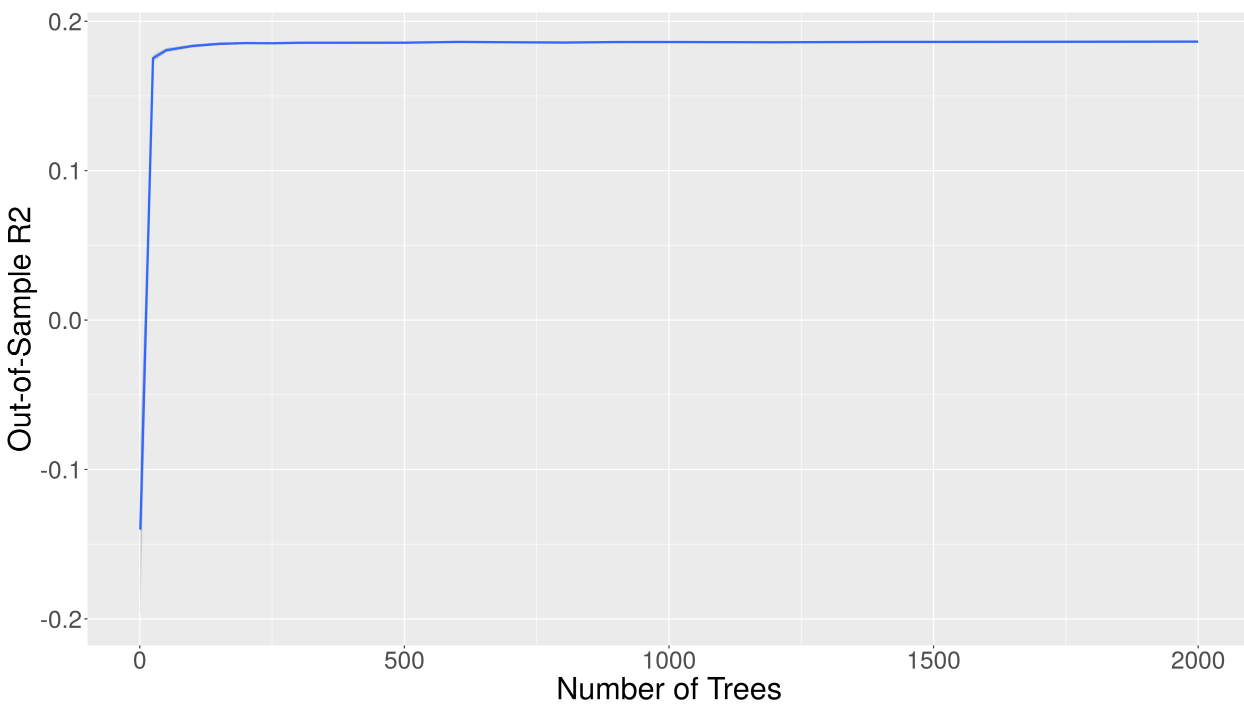
The standard algorithm allows to specify the probability of a predictor being chosen at each step. We take advantage of that and implement a two step procedure. First, we run a standard random forest regression, where every variable has the same probability of being chosen and obtain the variable importance for each of the features. We then run a different random forest where at each split, besides considering the strict random subset, we include the top n features from the first step up until that point in time for consideration at each split. This gives the algorithm the option, but not the obligation, of considering the best predictors from the first stage at each step. We find that adding this step increases the accuracy of the algorithm significantly. We choose $n = 5$ based on cross-validation.

Figure A1: Cross-validation Results of the number of trees in the two-quarters-ahead forecast



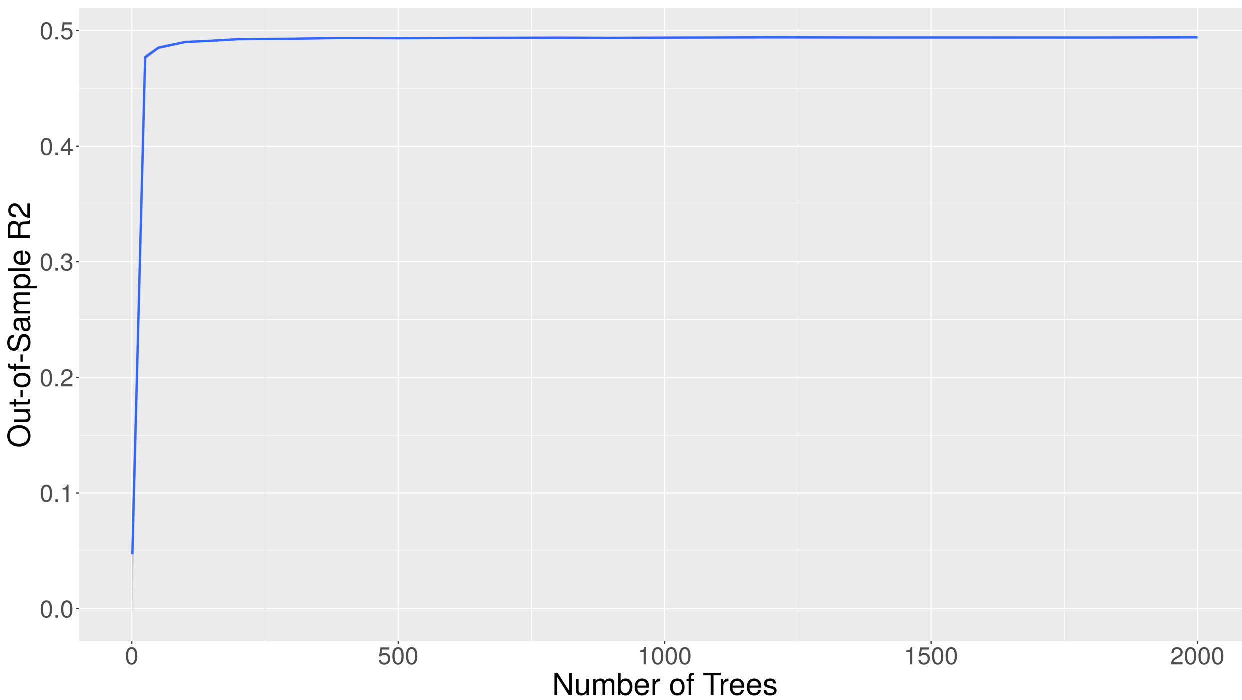
Notes: This figure plots the relation between the number of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the two-quarters-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using the machine learning forecast divided by the mean squared error of using the realized average value as a forecast. The random forest algorithm is random by design, so we take the average of 100 runs to measure the out-of-sample R^2.

Figure A2: Cross-validation Results of the number of trees in the three-quarters-ahead forecast



Notes: This figure plots the relation between the number of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the three-quarters-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using the machine learning forecast divided by the mean squared error of using the realized average value as a forecast. The random forest algorithm is random by design, so we take the average of 100 runs to measure the out-of-sample R^2.

Figure A3: Cross-validation Results of the number of trees in the two-years-ahead forecast



Notes: This figure plots the relation between the number of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the two-years-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using the machine learning forecast divided by the mean squared error of using the realized average value as a forecast. The random forest algorithm is random by design, so we take the average of 100 runs to measure the out-of-sample R^2.

A3. Summary statistics of variables in Fama-MacBeth return regressions

Table A2 reports the summary statistics of conditional biases in analysts' one-quarter (BE_Q1), two-quaters (BE_Q2), three-quarters (BE_Q3), one-year (BE_A1), and two-years (BE_A2) ahead earnings forecasts. Average BE denotes the average of these conditional biases at the multiple horizons. BE score denotes the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. We report the summary statistics we use as control variables, which include the log of firm size (Lnsize), the log of book-to-market ratio (Lnbeme), short-term reversal (Ret_1), medium-term momentum (Ret12_7), investment-to-asset (IA),idiosyncratic volatility (IVOL), return volatility (RetVol), and share turnover (Turnover).

Table A2: Summary statistics

Variable	N	Mean	Std	P1	Q1	Median	Q3	Q99
Average BE	1137094	0.008	0.043	-0.007	0.000	0.002	0.006	0.106
BE Score	1137094	50.265	22.777	8.250	32.600	47.000	67.000	97.750
BE_Q1	1063631	0.005	0.036	-0.006	0.000	0.000	0.003	0.086
BE_Q2	987185	0.007	0.048	-0.016	0.000	0.001	0.005	0.103
BE_Q3	899907	0.008	0.052	-0.024	0.000	0.002	0.006	0.115
BE_A1	1099495	0.022	0.133	-0.008	0.001	0.004	0.014	0.305
BE_A2	921693	0.039	0.206	-0.074	0.003	0.013	0.037	0.450
Lnsize	1137080	13.085	1.878	9.304	11.732	12.948	14.297	17.970
Lnbeme	1029863	-0.772	0.859	-3.288	-1.241	-0.678	-0.209	1.040
Ret12_7	1077283	0.079	0.466	-0.691	-0.138	0.038	0.221	1.563
Ret1	1136967	0.010	0.161	-0.387	-0.063	0.005	0.074	0.497
IA	1049608	0.314	1.047	-0.424	0.001	0.090	0.259	4.518
IVOL	1136942	0.025	0.020	0.005	0.013	0.020	0.031	0.097
RetVol	1136397	0.030	0.022	0.007	0.016	0.024	0.037	0.111
Turnover	1135432	1.571	13.458	0.074	0.492	1.000	1.908	8.534

A4. Fama-MacBeth regressions with conditional bias in each forecast horizon

Table A3 reports the Fama-MacBeth of monthly stock returns on conditional bias in each forecast horizon, including one-quarter, two-quarters, three-quarters, one-year, and two-years-ahead. (1) and (2) report the regression results with and without control variables respectively. We find that for when only consider conditional biases as predictors, they negative predict stock returns. The predictability of two-quarters, three-quarters, and two-years ahead forecast bias also remains robust after controlling for other return predictors.

Table A4 reports the value-weighted portfolio sorts on conditional bias in each forecast horizon. Overall, we find consistent evidence that stocks with larger biases earn lower future returns, though the relation between returns and one-year-ahead conditional bias seems to be flat.

Table A5 shows that the return-predictability results from the cross-sectional regressions and portfolio sorts also hold in time series regression against factor models such as the CAPM and the Fama-French five-factors model.

Table A3: Fama-Macbeth Regressions

Notes: This table reports the Fama-MacBeth cross-sectional regressions of monthly stocks' excess returns on the conditional bias in each forecast horizon: one-quarter, two-quarters, three-quarters, one-year, and two-years-ahead. (1) and (2) report the regression results with and without control variables, respectively. The t-statistics are reported in parentheses. The sample period is 1987 to 2019.

	A: One-quarter		B: Two-quarters		C: Three-quarters		D: One-year		E: Two-years	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
(Intercept)	0.0076	0.0206	0.0082	0.0230	0.0092	0.0246	0.0074	0.0189	0.0101	0.0309
	(2.65)	(3.74)	(2.79)	(4.18)	(3.13)	(4.42)	(2.63)	(3.46)	(3.59)	(5.56)
BE	−0.0602	0.0006	−0.3275	−0.3946	−0.3210	−0.3657	−0.0875	0.0497	−0.2278	−0.2808
	(−1.20)	(0.01)	(−8.47)	(−10.35)	(−9.55)	(−11.36)	(−1.23)	(0.75)	(−6.33)	(−8.70)
Lnsize		−0.0008		−0.0010		−0.0011		−0.0007		−0.0016
		(−2.14)		(−2.87)		(−3.06)		(−1.92)		(−4.16)
Lnbeme		0.0009		0.0017		0.0014		0.0009		0.0014
		(1.50)		(2.69)		(2.27)		(1.40)		(2.32)
Ret12_7		0.0045		0.0021		0.0017		0.0045		0.0022
		(2.86)		(1.31)		(1)		(2.79)		(1.31)
Ret1		−0.0248		−0.0253		−0.0244		−0.0276		−0.0283
		(−5.78)		(−5.91)		(−5.40)		(−6.32)		(−6.17)
IA		−0.0009		−0.0009		−0.0008		−0.0008		−0.0007
		(−3.22)		(−2.85)		(−2.29)		(−2.80)		(−2.39)
Ivol		−0.2309		−0.1667		−0.2259		−0.2113		−0.2290
		(−2.02)		(−1.42)		(−1.86)		(−1.88)		(−1.92)
Retvol		0.1138		0.1733		0.2443		0.0982		0.2314
		(0.95)		(1.40)		(1.88)		(0.82)		(1.82)
Turnover		−0.0006		−0.0009		−0.0009		−0.0007		−0.0007
		(−1.26)		(−1.86)		(−2.00)		(−1.45)		(−1.43)
R^2	0.0096	0.0598	0.0102	0.0646	0.0099	0.0672	0.0102	0.0623	0.0114	0.0665

Table A4: Portfolios sorted on conditional bias

This table reports the time series average of excess returns (in percent) on value-weighted portfolios sorted on the conditional bias at different forecast horizons. Panel A looks at the one-quarter-ahead conditional bias. Panel B looks at the two-quarters-ahead bias. Panel C looks at the three-quarters-ahead bias. Panel D looks at the one-year-ahead bias. Panel E looks at the two-years-ahead bias. The sample period is 1987 to 2019.

Quintile	1	2	3	4	5	1-5
Panel A: One-quarter-ahead BE						
Mean	0.70	0.59	0.69	0.63	0.37	0.33
t-stat	3.19	2.69	2.90	2.16	0.95	1.21
CAPM Beta	0.96	0.99	1.04	1.21	1.48	-0.52
Panel B: Two-quarters-ahead BE						
Mean	1.36	0.70	0.44	0.22	-0.85	2.21
t-stat	5.91	3.21	1.82	0.79	-2.17	7.89
CAPM Beta	0.94	0.98	1.05	1.16	1.50	-0.56
Panel C: Three-quarters-ahead BE						
Mean	1.16	0.67	0.46	0.05	-1.02	2.17
t-stat	4.51	2.95	1.91	0.17	-2.72	8.43
CAPM Beta	0.98	0.93	1.03	1.20	1.40	-0.42
Panel D: One-year-ahead BE						
Mean	0.63	0.61	0.71	0.62	0.70	-0.07
t-stat	3.17	2.86	3.03	2.22	1.98	-0.27
CAPM Beta	0.87	0.93	1.00	1.16	1.28	-0.41
Panel E: Two-years-ahead BE						
Mean	1.01	0.77	0.66	0.37	-0.65	1.66
t-stat	4.29	3.67	2.75	1.27	-1.62	5.51
CAPM Beta	1.00	0.93	1.03	1.21	1.51	-0.51

Table A5: Time series tests with common asset-pricing models

This table reports the regression of stock returns (in percent) on the long-short portfolio sorted with the conditional bias in different horizons, on the CAPM, the Fama-French three-factor model (FF3), and the Fama-French five-factor model (FF5). Panel A looks at the one-quarter-ahead conditional bias. Panel B looks at the two-quarters-ahead bias. Panel C looks at the three-quarters-ahead bias. Panel D looks at the one-year-ahead bias. Panel E looks at the two-years-ahead bias. The sample period is 1987 to 2019. The t-statistics are adjusted by the White's heteroscedasticity robust standard errors.

	Panel A: CAPM		Panel B: FF3		Panel C: FF5	
	Coeffi	t-stat	*Coeffi*	t-stat	*Coeffi*	t-stat
Panel A: One-quarter-ahead BE						
Intercept	0.67	2.81	0.73	3.33	0.28	1.23
Mkt_RF	-0.52	-6.79	-0.45	-5.89	-0.29	-4.16
SMB			-0.70	-7.48	-0.45	-4.20
HML			-0.30	-2.74	-0.63	-4.23
RMW					0.85	5.46
CMA					0.35	1.36
Panel B: Two-quarters-ahead BE						
Intercept	2.58	10.32	2.70	11.83	2.31	10.29
Mkt_RF	-0.56	-7.16	-0.55	-7.34	-0.41	-6.09
SMB			-0.53	-5.41	-0.32	-2.78
HML			-0.55	-4.19	-0.84	-5.85
RMW					0.71	4.10
CMA					0.32	1.30
Panel C: Three-quarters-ahead BE						
Intercept	2.44	9.94	2.51	10.39	2.17	9.00
Mkt_RF	-0.42	-5.40	-0.39	-4.83	-0.26	-3.44
SMB			-0.45	-4.62	-0.28	-2.65
HML			-0.34	-2.53	-0.61	-3.41
RMW					0.61	3.63
CMA					0.34	1.34
Panel D: One-year-ahead BE						
Intercept	0.20	0.86	0.26	1.23	0.05	0.21
Mkt_RF	-0.41	-6.21	-0.34	-5.33	-0.25	-3.72
SMB			-0.72	-5.63	-0.65	-5.55
HML			-0.32	-2.68	-0.54	-3.42
RMW					0.27	1.61
CMA					0.38	1.61
Panel E: Two-years-ahead BE						
Intercept	2.01	7.18	2.16	8.50	1.80	6.24
Mkt_RF	-0.51	-6.44	-0.48	-6.54	-0.33	-3.95
SMB			-0.66	-5.53	-0.48	-3.70
HML			-0.56	-4.60	-0.83	-5.22
RMW					0.65	3.70
CMA					0.28	0.87

A5. Cross-sectional return predictability: realized biases

As a placebo tests, we use the realized forecasts biases, defined as the difference between analysts' forecasts and the machine learning forecasts scaled by the share price from the most recent period, to "predict" stock returns, though realized earnings are not available at time t. A6 reports the Fama-MacBeth regression with individual realized bias in different horizon. A7 reports the regressions with average realized biases, and A8 and A9 report the mean return and alpha on the long-short portfolio strategy based on realized average biases. Overall, we find very consistent results, stocks with larger forecast biases earn lower future returns.

Table A6: Fama-Macbeth regressions: realized forecast bias

Notes: This table reports the unfeasible Fama-MacBeth cross-sectional regressions of monthly stocks' excess returns on the realized bias in each forecast horizon: one-quarter, two-quarters, three-quarters, one-year, and two-years ahead. The realized bias is defined as the difference between analysts' forecasts and the realized value scaled by the share price from the most recent period. (1) and (2) report the regression results with and without control variables, respectively. The sample period is 1987 to 2019. It is important to remark that the realized bias are not available at time t and the table is only presented for bench-marking purposes.

	A: One-quarter		B: Two-quarters		C: Three-quarters		D: One-year		E: Two-years	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
(Intercept)	0.0081	0.0216	0.0081	0.0222	0.0095	0.0254	0.0086	0.0231	0.0106	0.0331
	(2.75)	(3.89)	(2.72)	(3.97)	(3.22)	(4.57)	(2.97)	(4.18)	(3.69)	(5.92)
BE	−0.2748	−0.3181	−0.2951	−0.3371	−0.2614	−0.3034	−0.2855	−0.3555	−0.2172	−0.2667
	(−15.79)	(−16.32)	(−16.32)	(−19.17)	(−14.67)	(−16.84)	(−14.80)	(−16.50)	(−12.72)	(−16.26)
Lnsize		−0.0009		−0.0010		−0.0012		−0.0011		−0.0017
		(−2.53)		(−2.64)		(−3.11)		(−2.85)		(−4.49)
Lnbeme		0.0017		0.0019		0.0017		0.0019		0.0016
		(2.79)		(3.11)		(2.65)		(3.09)		(2.66)
Ret12_7		0.0031		0.0019		0.0019		0.0024		0.0026
		(1.93)		(1.21)		(1.13)		(1.48)		(1.54)
Ret1		−0.0321		−0.0303		−0.0281		−0.0324		−0.0318
		(−7.32)		(−6.99)		(−6.27)		(−7.50)		(−7.05)
IA		−0.0008		−0.0006		−0.0005		−0.0007		−0.0004
		(−2.71)		(−1.86)		(−1.41)		(−2.40)		(−1.39)
Ivol		−0.1739		−0.1773		−0.2354		−0.1532		−0.2050
		(−1.52)		(−1.49)		(−1.96)		(−1.36)		(−1.73)
Retvol		0.1491		0.1549		0.2164		0.1501		0.1952
		(1.24)		(1.25)		(1.70)		(1.26)		(1.56)
Turnover		−0.0007		−0.0005		−0.0004		−0.0006		−0.0002
		(−1.51)		(−1.11)		(−0.97)		(−1.23)		(−0.52)
R^2	0.0090	0.0662	0.0107	0.0683	0.0101	0.0692	0.0093	0.0648	0.0100	0.0680

Table A7: Fama-Macbeth regressions: realized forecast bias

Notes: This table reports the unfeasible Fama-MacBeth cross-sectional regressions of monthly stocks' excess returns on the realized bias. We define the realized bias as the difference between analysts' earnings forecasts and actual realized values, scaled by the stock price from the most recent month. "Average BE" denotes the average of the realized biases at different forecast horizons including one-quarter, two-quarters, three-quarters, one-year, and two-years-ahead. "BE score" denotes the arithmetic average of the percentile rankings on each of the five realized biases at different forecast horizons. (1) and (2) report the regression results with and without control variables, respectively. The t-statistics are reported in parentheses. The sample period is 1987 to 2019. It is important to remark that the realized bias are not available at time t and the table is only presented for bench-marking purposes.

	Panel A: Average BE		Panel B: BE_score	
	(1)	(2)	(1)	(2)
BE	−0.1208	−0.1473	−0.0945	−0.1061
	(−14.34)	(−16.47)	(−38.25)	(−45.94)
Lnsize		−0.0012		−0.0026
		(−3.18)		(−6.96)
Lnbeme		0.0019		0.0012
		(3.15)		(2.02)
Ret12_7		0.0026		−0.0019
		(1.66)		(−1.23)
Ret1		−0.0324		−0.0573
		(−7.52)		(−12.89)
IA		−0.0006		−0.0004
		(−2.09)		(−1.57)
Ivol		−0.1731		−0.1286
		(−1.52)		(−1.12)
Retvol		0.1693		0.1437
		(1.41)		(1.20)
Turnover		−0.0006		−0.0002
		(−1.26)		(−0.36)
Intercept	0.0089	0.0251	0.0549	0.0940
	(3.03)	(4.56)	(19.28)	(16.42)
R^2	0.0104	0.0655	0.0340	0.0917

Table A8: Portfolios sorted on realized bias

This table reports the time series average of excess returns (in percent) on value-weighted portfolios formed on the average of the realized analyst' forecast bias. We define the realized bias as the difference between analysts' earnings forecasts and actual realized values, scaled by the stock price from the most recent month. Panel A looks at average conditional bias at different forecast horizons including one-quarter, two-quarters, three-quarters, one-year, and two-years-ahead. Panel B presents the sorts based on "BE score", defined as the arithmetic average of the percentile rankings on each of the five reazlied biases at different forecast horizons. The sample period is 1987 to 2019.

Quintile	1	2	3	4	5	1-5
			Panel A: Average BE			
Mean	2.97	1.30	-0.04	-0.98	-2.00	4.97
t-stat	11.77	6.10	-0.20	-3.61	-5.61	21.77
CAPM Beta	1.03	0.94	0.98	1.16	1.39	-0.36
			Panel B: BE Score			
Mean	2.87	1.40	0.14	-1.01	-2.59	5.46
t-stat	11.61	6.50	0.65	-3.94	-7.83	25.74
CAPM Beta	1.04	0.95	0.95	1.10	1.31	-0.27

Table A9: Time series tests of long-short portfolios sorted on realized bias

This table reports the regression of stock returns (in percent) on the long-short portfolio sorted with the realized bias, on the CAPM, the Fama-French three-factor model (FF3), and the Fama-French five-factor model (FF5). We define the realized bias as the difference between analysts' earnings forecasts and actual realized values, scaled by the stock price from the most recent month. Panel A looks at average conditional bias at different forecast horizons including one-quarter, two-quarters, three-quarters, one-year, and two-years-ahead. Panel B presents the sorts based on "BE score", defined as the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. The sample period is 1987 to 2019. The t-statistics are adjusted by the White's heteroscedasticity robust standard errors.

$$LS\_Port_t = \alpha + \sum_{i=1}^{5} \beta_i FF_{i,t} + \epsilon_t$$

	CAPM		FF3		FF5	
	$Coeffi$	t-stat	$Coeffi$	t-stat	$Coeffi$	t-stat
Panel A: Average BE						
Intercept	5.20	23.24	5.25	23.70	4.90	21.11
Mkt_RF	-0.36	-5.74	-0.33	-4.95	-0.19	-2.93
SMB			-0.39	-4.53	-0.25	-2.87
HML			-0.21	-1.87	-0.54	-4.26
RMW					0.52	4.56
CMA					0.51	2.29
Panel B: BE Score						
Intercept	5.64	26.87	5.67	27.46	5.45	24.53
Mkt_RF	-0.27	-4.50	-0.24	-3.76	-0.15	-2.35
SMB			-0.37	-4.33	-0.29	-3.02
HML			-0.16	-1.51	-0.37	-3.19
RMW					0.32	2.87
CMA					0.34	1.68

A6. Cross-sectional return predictability: other robustness checks

In this section, we check the robustness of Fama-MacBeth regression results in Table 4 by omitting stocks whose prices are lower than \$5 and also by scaling the conditional biases with total asset (per share) from the last fiscal year. Total assets are obtained from Compustat (Item AT) Table A10 and A11 report the two robustness checks results respectively. Overall, we find robust return predictability of conditional biases.

Table A10: Fama-Macbeth regressions: omitting stocks with price lower than \$5

Notes: This table reports the Fama-MacBeth cross-sectional regressions of monthly stocks' excess returns on the conditional bias. "Average BE" denotes the average of the conditional biases at different forecast horizons including one-quarter, two-quarters, three-quarters, one-year, and two-years-ahead. "BE score" denotes the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. (1) and (2) report the regression results with and without control variables, respectively. The t-statistics are reported in parentheses. The sample period is 1987 to 2019. We omit stocks whose previous end of the month price is smaller than 5.

	Panel A: Average BE		Panel B: BE_score	
	(1)	(2)	(1)	(2)
BE	−0.6093	−0.7398	−0.0422	−0.0531
	(−12.36)	(−16.04)	(−12.41)	(−19.47)
Lnsize		−0.0015		−0.0025
		(−4.81)		(−7.94)
Lnbeme		0.0019		0.0022
		(3.38)		(3.83)
Ret12_7		0.0033		0.0021
		(2.21)		(1.47)
Ret1		−0.0187		−0.0210
		(−4.60)		(−5.21)
IA		−0.0004		−0.0004
		(−1.34)		(−1.38)
Ivol		−0.2269		−0.2247
		(−2.04)		(−2.01)
RetVol		0.1643		0.1935
		(1.34)		(1.59)
Turnover		−0.0005		−0.0003
		(−1.07)		(−0.80)
Intercept	0.0102	0.0330	0.0264	0.0653
	(3.95)	(6.95)	(11.20)	(14.02)
R^2	0.0105	0.0676	0.0148	0.0698

Table A11: Fama-Macbeth regressions: scaling conditional biases by total assets per share

Notes: This table reports the Fama-MacBeth cross-sectional regressions of monthly stocks' excess returns on the conditional bias, which is defined as the difference between analysts' earnings forecasts and machine learning forecasts, scaled by the total asset from the most recent fiscal period. "Average BE" denotes the average of the conditional biases at different forecast horizons including one-quarter, two-quarters, three-quarters, one-year, and two-years-ahead. "BE score" denotes the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. (1) and (2) report the regression results with and without control variables, respectively. The t-statistics are reported in parentheses. The sample period is 1987 to 2019.

	Panel A: Average BE		Panel B: BE score	
	(1)	(2)	(1)	(2)
(Intercept)	0.0080	0.0206	0.0232	0.0597
	(2.79)	(3.78)	(9.89)	(11.69)
BE	−0.0249	−0.0413	−0.0310	−0.0440
	(−2.65)	(−6.11)	(−6.65)	(−16.31)
Lnsize		−0.0009		−0.0026
		(−2.38)		(−7.35)
Lnbeme		0.0006		−0.0012
		(0.99)		(−2.05)
Ret12_7		0.0047		0.0034
		(2.92)		(2.20)
Ret1		−0.0287		−0.0311
		(−6.68)		(−7.27)
IA		−0.0007		−0.0007
		(−2.52)		(−2.55)
Ivol		−0.1981		−0.1741
		(−1.75)		(−1.54)
Retvol		0.1221		0.1963
		(1.03)		(1.68)
Turnover		−0.0006		−0.0003
		(−1.29)		(−0.64)
R^2	0.0052	0.0571	0.0180	0.0623

A7. Net stock issuances: robustness check

We check the robustness of results in Table 10 by matching average of conditional bias from the past 24-12 months to net stock issuances of the fiscal year ending in t. Table A12 reports this robustness check. Overall, we find consistent results that Managers of those companies for which analysts' upward biases are greatest take apparent advantage of these biases by issuing stocks.

Table A12: Net stock sssuances and conditional bias

Panel A reports the time series average of net stock issuances of value-weighted portfolios sorted on the conditional bias. "Average BE" denotes the average of the conditional bias at different forecast horizons."BE score" denotes the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. Panel B reports the Fama-MacBeth regressions of firms' net stock issuances on the conditional bias and control variables include the log of firm size (Lnsize), the log of book-to-market ratio (Lnbeme), and earnings before interest, taxes, and depreciation divided by total assets (EBITDA). The sample period is 1987 to 2019. We report the time-series average of slope coefficients associated with Newey-West t-statistics.

Panel A: Net stock issuances of portfolios formed on BE

Quintile	1	2	3	4	5	5-1
Average BE	0.013	0.012	0.018	0.032	0.078	0.065
t-stat	1.98	1.46	3.15	4.1	6.29	4.56
BE score	0.010	0.012	0.022	0.033	0.071	0.062
t-stat	1.42	1.57	4.08	3.86	5.62	5.49

Panel B: Fama-MacBeth regressions

	A: Average BE		B: BE Score	
	(1)	(2)	(1)	(2)
BE	1.7651	1.0731	0.1136	0.0717
t-stat	5.80	4.37	7.32	5.33
Lnsize		-0.0039		-0.0019
t-stat		-2.69		-1.02
Lnbeme		-0.0199		-0.0206
t-stat		-4.90		-5.26
EBITDA		-0.1293		-0.1273
t-stat		-4.83		-4.59
Intercept	0.0345	0.0809	-0.0125	0.0253
t-stat	8.22	3.35	-1.53	0.76
R^2	0.0301	0.0835	0.0197	0.0786

A8. Anomalies

In this study, we follow Hou et al. (2015) as close as possible to define anomaly variables. Table A13 lists the significant anomalies documented in Hou et al. (2015). L-S ret (%) denotes the monthly average return (in percent) of each of the 27 long-short anomaly strategies. The sample period is July 1972 to December 2019, depending on data availability.

Table A13: List of significant anomalies

Anomalies	**Descriptions**	**Sample period**	L-S ret (%)
Sue-1	Earnings surprise (1-month holding period), Foster et al. (1984)	01/1974–12/2019	0.42
Abr-1	Cumulative abnormal stock returns (1-month holding period), Chan et al. (1996)	07/1972–12/2019	0.89
R11-1	Price momentum (11-month prior returns, 1-month holding period), Fama and French (1996)	07/1972–12/2019	1.23
BM	Book-to-market equity, Rosenberg et al. (1985)	07/1972–12/2019	0.46
Dur	Equity duration, Dechow et al. (2004)	07/1972–12/2019	1.27
E/P	Earnings-to-price, Basu (1983)	07/1972–12/2019	0.39
CF/P	Cash flow-to-price, Lakonishok et al. (1994)	07/1972–12/2019	0.33
NO/P	Net payout yield Boudoukh et al. (2007)	07/1972–12/2019	0.30
I/A	Investment-to-assets, Cooper et al. (2008)	07/1972–12/2019	0.45
NOA	Net operating assets, Hirshleifer et al. (2004)	07/1972–12/2019	0.50
ΔPI/A	Changes in property, plant, and equipment plus changes in inventory scaled by assets Lyandres et al. (2007)	07/1972–12/2019	0.41
IG	Investment growth, Xing (2007)	07/1972–12/2019	0.34
CEI	Composite equity issues, Daniel and Titman (2006)	07/1972–12/2019	0.40
NSI	Net stock issues, Pontiff and Woodgate (2008)	07/1972–12/2019	0.59
IvC	Inventory changes, Thomas and Zhang (2002)	07/1972–12/2019	0.51
IvG	Inventory growth, Belo and Lin (2012)	07/1972–12/2019	0.34
OA	Operating accruals, Sloan (1996)	07/1972–12/2019	0.26
POA	Percent operating accruals, Hafzalla et al. (2011)	07/1972–12/2019	0.33
PTA	Percent total accruals, Hafzalla et al. (2011)	07/1972–12/2019	0.30
GP/A	Gross profits-to-assets, Novy-Marx (2013)	07/1972–12/2019	0.21
ROE	Return on equity, Haugen and Baker (1996)	07/1972–12/2019	0.72

continued from previous page

ROA	Return on assets, Balakrishnan et al. (2010)	07/1972−12/2019	0.57
NEI	Number of consecutive quarters with earnings increases, Barth et al. (1999)	07/1972−12/2019	0.30
OC/A	Organizational capital-to-assets, Eisfeldt and Papanikolaou (2013)	07/1972−12/2019	0.26
Ad/M	Advertisement expense-to-market, Chan et al. (2001)	07/1972−12/2019	0.46
RD/M	R&D-to-market, Chan et al. (2001)	07/1972−12/2019	0.78
OL	Operating leverage, Novy-Marx (2010)	07/1972−12/2019	0.23
Average			0.47