MAN VS. MACHINE LEARNING:
THE TERM STRUCTURE OF EARNINGS EXPECTATIONS AND CONDITIONAL BIASES

Jules H. van Binsbergen
Xiao Han
Alejandro Lopez-Lira

WORKING PAPER 27843

NBER WORKING PAPER SERIES

MAN VS. MACHINE LEARNING:
THE TERM STRUCTURE OF EARNINGS EXPECTATIONS
AND CONDITIONAL BIASES

Jules H. van Binsbergen
Xiao Han
Alejandro Lopez-Lira

Working Paper 27843
http://www.nber.org/papers/w27843

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2020, Revised August 2021

We thank Refinitive for their guidance in using the I/B/E/S database. We are grateful for helpful comments and suggestions provided by Bryan Kelly (discussant), Svetlana Bryzgalova (discussant) and seminar participants at University of Florida, the NBER Big Data and Securities Markets Fall 2020, the Georgetown Global Virtual Seminar Series on Fintech, the Wolfe Quantitative and Macro Investment Conference, The Future of Financial Information Conference and BI Norwegian Business School. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Man vs. Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases
Jules H. van Binsbergen, Xiao Han, and Alejandro Lopez-Lira
NBER Working Paper No. 27843
September 2020, Revised August 2021
JEL No. D22,D83,D84,G11,G12,G14,G31,G4

ABSTRACT

We introduce a real-time measure of conditional biases in firms' earnings forecasts. The measure is defined as the difference between analysts' expectations and a statistically optimal unbiased machine-learning benchmark. Analysts' conditional expectations are, on average, biased upwards, and the bias increases in the forecast horizon. These biases are associated with negative cross-sectional return predictability, and the short legs of many anomalies contain firms with excessively optimistic earnings. Further, managers of companies with the greatest upward-biased earnings forecasts are more likely to issue stocks. Commonly-used linear earnings models do not work out-of-sample and are inferior to those provided by analysts.

Jules H. van Binsbergen
The Wharton School
University of Pennsylvania
3620 Locust Walk
Philadelphia, PA 19104
and NBER
julesv@wharton.upenn.edu

Alejandro Lopez-Lira
1454 Union Rd
Stuzin Hall 321
Warrington College of Business
University of Florida
Gainesville, FL 32607
Alejandro.Lopez-Lira@warrington.ufl.edu

Xiao Han
University of Edinburgh Business School
29 Buccleuch Place
Edinburgh
United Kingdom
xiao.han@ed.ac.uk

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. Commonly used cash flow proxies include the most recent realized earnings, simple linear forecasts, or analysts' forecasts. However, a significant strain of literature documents these 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, in contrast to linear forecasts, 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 predictions obtained from other publicly available data sources. For example, analysts' forecasts could 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 our algorithm's extensive set of publicly available variables. As such, these forecasts are a crucial input to our machine learning approach.[2] That said, analyst forecasts, which are often biased, can be improved upon by optimally combining them with publicly available information sources.

We use random forest regression for our primary 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-year forecasts for annual earnings. For quarterly forecasts, we use the

[1]See Kothari et al. (2016) for an extensive review. We also document in appendix A12 that the linear forecasts do not have return predictability out-of-sample, consistent with their limited prediction power.

[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 well-known predictive algorithms in the context of cross-sectional returns. See Bryzgalova et al. (2020) for a novel application of tree-based methods to form portfolios.

one-quarter, two-quarter, and three-quarter horizons. We focus on these particular horizons as analysts' forecasts for other horizons have significantly fewer observations. Given 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 induce negative cross-sectional stock return predictability: stocks with overly optimistic expectations earn lower subsequent returns and vice versa. Notably, the short legs of common anomalies consist of firms for which the analysts' forecasts are excessively optimistic relative to our benchmark. Finally, we show that managers of those companies with the largest biases seem to take advantage of the overly optimistic expectations by issuing stocks.[4]

Although previous research uses realized earnings to evaluate the bias and efficiency of analyst forecasts, these extant studies do not use a time series or cross-section of real-time earnings forecasts as a benchmark.[5] Without such forecasts, it is difficult to assess and correct the conditional dynamics of forecast biases before the actual value is realized. Hence, such studies only document an unconditional bias over time and in the cross-section. That is, we cannot know whether the given forecasts are *conditionally* biased, nor do we observe the variation of these biases across stocks and time and their impact on asset returns.

We fill this void by constructing a statistically optimal time-series and cross-section of earnings forecasts. To the best of our knowledge, we are the first to use machine learning to create a real-time proxy for firms earnings' conditional expectations. The resulting estimates enable us to compute real-time implied analyst biases, which can be used in cross-sectional stock-pricing sorts and to study managers' issuance behavior. Therefore, our benchmark expectation diverges from the conventional approach, which uses either the raw analysts' expectations, the past realized earnings value, or a simple linear model to form the conditional

[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 projections.

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

forecast.[6]

Another strain of the relevant literature sorts stocks cross-sectionally using long-term earnings growth forecasts, without comparing these values to a benchmark (e.g., La Porta (1996), Bordalo et al. (2019)). This approach implicitly assumes that the cross-sectional median (or average) is sufficient as a counterfactual. However, given the large cross-sectional variation in earnings, it remains challenging to determine whether beliefs are biased or exaggerated without a fully specified benchmark model (Zhou (2018)).

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 (as in 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. Similarly, Frankel and Lee (1998) suggests a linear model using a few selected variables. We differ from So (2013) and Frankel and Lee (1998) in three important ways.

First, because linear regressions do not efficiently handle high-dimensional data, a variable selection step is necessary. Often, variables that have been documented ex-post as effective predictors are selected in this step, rendering the linear forecast not entirely out-of-sample. We demonstrate in appendix A11 that the variable selection step is not innocuous, and most (if not all) of the return predictability examined in So (2013) using linear forecasts disappears after the 2000s. In contrast, our machine learning approach considers a broad set of macroeconomic and firm-specific signals at every point in time. We, therefore, do not incur any data leakage. As a consequence, the out-of-sample predictability of our machine learning forecasts remains relatively stable throughout the sample.

Second, the linear forecasts in So (2013) are not designed to be statistically optimal. In fact, analysts' forecasts are a better proxy for the conditional expectations than linear forecasts are, as measured by the mean squared error, even after the variable selection step.

[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.

In contrast, our machine learning forecasts are a better proxy out-of-sample.

Third and finally, there is no reason to impose the linearity of the conditional expectation function. 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 exemplify its usefulness by focusing 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 and a high correlation with return anomalies. Regarding the second, we find that managers tend to issue more stocks when their firms are subject to more optimistic forecasts relative to our benchmark.

While these two applications are illustrative of the usefulness of our approach, we also note that part of our contribution is the expectation measure itself, which we make available in the posted data section.

Finally, before explaining the economic and statistical theory and the empirical results, we describe our contribution to the existing literature in the next section.

Related literature

Regarding the relationship between anomalies and conditional biases, Engelberg et al. (2020) document that analysts' price targets and buy/sell recommendations contradict stock return anomaly variables. In contrast, our paper focuses on a different set of analysts that provide earnings forecasts. We find that biases in these cash flow predictions correlate with anomaly returns, suggesting an expectational error component in cash flows driving anomalies.

Previous work also exists on the relationship between analysts' expectations and the stock issuance behavior of firms. Given that this earlier work does not use a real-time conditional benchmark for earnings that the analysts' expectations can be compared to, the conclusions drawn are different from ours. Particularly, Richardson et al. (2004) argue that firms and managers communicate with each other. Analysts start with optimistic forecasts, gradually lower those forecasts as the earnings announcement approaches, undershoot the earnings forecast just before the announcement, allowing firms to outperform the forecast and issue stock shortly after this positive news.

In contrast, our findings are consistent with a different economic mechanism. We use a real-time bias measure and find that firms issue more stocks when the real-time bias is higher, which happens long *before* the end-of-period earnings announcement. Our explanation for this phenomenon is that managers understand when analysts are overly optimistic because managers have private information. Therefore, they take advantage of this optimism in the market and issue stock before earnings are realized, even up to two years before.

We also contribute to the growing literature that documents analysts are skillful and exert effort (for example Grennan and Michaely (2020)) by providing evidence that despite analysts being conditionally biased, they provide unique information above and beyond what can be found in standard accounting and macroeconomic variables. Furthermore, we show how this information can be incorporated efficiently to form better forecasts.

Our work also 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 can construct superior forecasts exploiting 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 publicly available information in each period, and (3) documenting that these biases relate to outcomes in financial markets and corporate policies.

2 Model

This section presents a condensed version of a tractable non-linear model of earnings and earnings expectations that illustrates some reasons linear forecasts are inferior to those provided by machine learning techniques and analysts. In particular, high variance of the relevant non-linear effects causes the linear models to underperform machine learning techniques. The complete model also features asset prices so that it can be used to understand further why our approach produces stable return predictability out-of-sample while linear forecasts do not. This complete model is presented in Appendix A1.

Model

Consider the following setup. There are two periods in the economy. First, there are a measure 1 of assets to be priced, indexed by i. Second, the payoff y of asset i is a random variable forecastable by a combination of linear and non-linear effects. In particular, the actual payoff distribution follows:

$$\tilde{y}_i = f(x_i) + g(v_i) + z_i + w_i + \tilde{\epsilon}_i. \tag{1}$$

Where v_i, w_i, x_i, z_i are variables measurable in the first period and distributed in the cross-section as independent standard normal. f and g are non-linear functions, orthogonal to the space of linear functions in x_i and v_i respectively ($E[xf(x)] = E[vg(v)] = 0$). We assume that analysts use $f(x_i)$ and w_i in their forecasts. However, we assume that they miss out on

[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.

the effects of z_i (which will deliver return predictability) as well as $g(v_i)$. The latter can be motivated either because analysts are not aware of the forecasting power of transformations of v_i or because they only use linear transformations of v_i. $\tilde{y}$ and $\tilde{\epsilon}_i$ are random variables measurable in the second period. $\tilde{\epsilon}_i$ is distributed as an independent standard normal. We assume that the agents have a large enough sample of these variables from past observations so that there is no estimation error of the coefficients. Notice that (due to the orthogonality assumption above) in a linear regression, the true coefficients associated with x_i and v_i are zero. For tractability, the shock to earnings is not priced, and the risk-free rate equals zero.

Our theoretical model includes non-linear effects because, in our empirical specification, we document substantial non-linearities in the earnings process as a function of the explanatory variables. For example, analysts' forecasts are amongst the most important predictors, and Figure 1 shows that EPS is a non-linear function of analysts' forecasts. Hence, using the linear prediction produces substantial errors as shown in Figure 2. Figures 3 and 4 show the same problem arises when using past EPS which is a key ingredient of linear forecasts such as in Frankel and Lee (1998) or So (2013).

[Insert Figure 1 and 2 about here]

[Insert Figure 3 and 4 about here]

We show in the appendix that the earnings forecasting error is weakly decreasing in the number of explanatory variables used, since an ideal conditional expectation function can always disregard useless information. For our application, random forest regression automatically discards useless forecasting variables and incorporates useful ones. Given its flexibility and robustness, it will (asymptotically) always benefit from adding information.

Hence, if we include analysts' expectations (which are in the public information set), any optimal estimator will achieve an error no higher than analysts make. In practice, we find that random forest succeeds when adding analysts' expectations to the information set, while linear models are no better than analysts' forecasts. Because of their flexibility, random forests can approximate any functional form, and (asymptotically) random forests are a consistent estimator of the conditional mean.[8]

We also show in the appendix that under general conditions, as expected, stocks with pessimistic (lower than optimal) predictions should have higher (realized) returns and vice-versa.

Spurious in-sample linear predictability

In the appendix, we also show that even though analysts' forecasts dominate the linear forecasts, return predictability may still arise from the conditional bias measured by the difference between the analysts' forecasts and the *linear* forecasts. It occurs when a variable in which the analyst forecast and the linear forecast differ is associated with return predictability. To make matters worse, if the variable driving the return predictability only works in-sample, the linear model's return predictability will decrease substantially or disappear altogether out-of-sample. In our empirical specification, the linear model return predictability indeed disappears after the 2000s. In contrast, for the machine learning model, the return predictability remains relatively stable.

3 Methodology and Data

In this section, we describe how we apply random forest techniques to earnings. We also describe the data sources that we input to this machine learning algorithm.

3.1 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 and is asymptotically unbiased and can approximate any function. The ultimate forecast is superior to a prediction following from any individual pre-

[8]The property is commonly referred to in the literature as random forests being universal approximators. We confirm in simulations that it applies in our setup.

dictor (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. We summarize the key parameters of our implementation in Table 1 and discuss the cross-validation method in detail in Appendix A4. We explain the algorithm itself thoroughly 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 7 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 past earnings per share (denoted by past_eps_std), and the threshold (or cutoff) value is 0.051. Naturally, the whole sample (100%) is used at this first step. Were we to end at this step, the forecast eps-value is .06 when past_eps_std is less than or equal to 0.051 (which corresponds to 57% of the sample), and 0.73 when past_eps_std is more than or equal to 0.051 (43% of the sample). In the next step, the algorithm splits each of the previous two sub-spaces in two again. The first subspace (past earnings per share less than 0.051) is split in two using past earnings per share as an explanatory variable. The threshold value is -0.66. The second subspace (past earnings per share greater than or equal to 0.051) uses the price per share lower than 1.1. 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. Figures 8 and 9 show the resulting predictive surface.

[Insert Figures 7, 8 and 9 about here]

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 path length from a root node to leaves. The larger the depth, the more complex the tree, and the more likely it will overfit the data.[9]

More formally, the decision tree model forecast ($\hat{y}$) is constant over a disjoint number of regions R_m:

$$\hat{y} = f(x) = \sum_{m} c_m I_{\{x \in R_m\}}, \tag{2}$$

where the constants are given by:

$$c_m = \frac{1}{N_m} \sum_{\{y_i : x_i \in R_m\}} y_i, \tag{3}$$

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

$$R_m = \{x_i \in \bigtimes_{i \in I} X_i : k_{i,l}^m < x_i \leq k_{i,h}^m\}, \tag{4}$$

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

The algorithm minimizes the mean squared error numerically 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 decision trees with shorter depth, specifically random forest regression models.

[9] 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 (Hastie et al. (2001))). This sample threshold is called the *minimum node size*.

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.[10] The final prediction of a random forest model is 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 used in each split.[11]

Since the random forest is a bootstrapping procedure, a high number of decision trees is optimal. Notwithstanding computational time, there is no theoretical downside for using more trees. That said, performance tends to plateau following a large number of trees. Figure 10 and 11 confirm that this indeed holds in our setup: The performance is increasing in the number of trees but reaches a plateau.[12]

[Insert Figure 10 and 11 about here]

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

[11] 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.

[12] 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_i and EPS_i denote the machine learning forecast and actual realized earnings respectively for firm i. $\overline{EPS}$ represents the cross-sectional average of firm earnings. The denominator, $\sum(EPS_i - \overline{EPS})^2$, is constant across different specifications.

The depth of each decision tree determines the overall complexity of the model. Thus, more complex models are more likely to overfit. Nevertheless, because of the inherent randomization, random forests are resilient to over-fitting in a wide variety of circumstances. Figures 12 and 13 show that the performance of the model is increasing in model complexity up until a depth of 7.

[Insert Figure 12 and 13 about here]

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%, we would first take a 1% random sub-sample without replacement as the training sample for each decision tree. We then repeat the process for each remaining tree. Figures 14 and 15 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.

[Insert Figures 14 and 15 about here]

While random forest regressions are non-parametric, we can interpret them using partial dependence plots (PDPs). PDPs explain how features influence the predictions. They display the average marginal effect on the forecast for each value of variable x_i. PDPs show the value the model predicts on average when each data instance has a fixed value for that feature. While a disadvantage is that the averages calculated for the partial dependence plot may include very unlikely data points, we include confidence intervals in the figures to address the uncertainty. Formally PDPs are defined as:

$$\hat{f}_{x_s}(x_s) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x_s, x_c^{(i)}) \approx E_{x_c}\left[\hat{f}(x_s, x_c)\right], \tag{5}$$

where x_s is the variable of interest, and x_c^i is a vector representing realizations of the other variables. We show examples of PDPs in Figures 1 and 2. The technique can also be applied to explain the joint effect of variables, as illustrated in Figure 16.

[Insert Figure 16 about here]

We train the random forest model using data from the most recent year for the quarterly earnings forecasts and one-year ahead forecast. We forecast earnings in the following periods using only the information available at the current time. For the two-year-ahead predictions, 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.[13] 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 i's 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, two quarters, three quarters, one year, and two years, because analysts' forecasts for other horizons have significantly fewer observations. As analysts make earnings forecasts every month, we construct our statistically optimal benchmark monthly.[14]

3.2 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.

3.2.1 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

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

[14]To minimize the impact of outliers within the model, we winsorize the forecasting variables at the 1% level and standardize them following the recommended guidelines in the literature (Hastie et al. (2001)).

2. Monthly stock prices and returns from CRSP

3. Sixty-seven financial ratios such as the book-to-market ratio and dividend yields obtained from the Financial Ratios Suite by Wharton Research Data Services[15]

3.2.2 Macroeconomic variables

We consider several macroeconomic variables that can affect firms' earnings. We obtain these 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

3.2.3 Analyst forecasts

Analysts' forecasts at time t for firm i's earnings at fiscal end period $t+1$ can be decomposed into public and private signals:[16]

$$AF_{i,t}^{t+1} = \sum_{j=1}^{J} \beta_j X_{j,i,t} + \sum_{k=1}^{K} \gamma_k P_{k,i,t} + B_{i,t}, \tag{6}$$

where $X_{j,i,t}$, with $j \in 1, ..., J$, represent the J public signals known at time t about firm i; $P_{k,i,t}$, with $k \in 1, ..., K$ are K private signals about firm i at time t; and $B_{i,t}$ represents the analysts' bias generated by expectation errors or incentive problems for firm i at time t. 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 the I/B/E/S unadjusted actual file (actual realized earnings) with the 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 estimates and the realized EPS value are based on different numbers of shares outstanding. To address this issue, we use the

[15]See Appendix A3 for details of these variables' definitions.
[16]See Hughes et al. (2008) and So (2013) among others.

cumulative adjustment factors from the CRSP monthly stock file to adjust the forecast and the actual EPS on the same share basis.[17]

3.3 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.[18] 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 investor expectations as the differences between the analysts' forecast and the machine learning forecast, scaled by the closing stock price from the most recent month:

$$\text{Biased\_Expectation}_{i,t}^{t+h} = \frac{\text{Analyst\_Forecasts}_{i,t}^{t+h} - \text{ML\_Forecast}_{i,t}^{t+h}}{\text{Price}_{i,t-1}} \tag{7}$$

in which subscript i denotes firm, and t indicates the date when earnings forecasts are made. The superscript $t+h$ represents the forecasting period.

4 Hypotheses

In this section we lay out our main hypotheses.

4.1 Biased expectations and the cross-section of stock returns

If indeed, our machine learning forecasts provide the statistically optimal unbiased benchmark for earnings expectations, but investors are affected by (biased) analysts' forecasts, we should observe that the stocks with optimistic earnings forecasts will earn low future returns. That is,

[17] 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)).

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

overly optimistic earnings forecasts are associated with stock overpricing. Our first hypothesis is, therefore:

Hypothesis 1: Stocks with more optimistic earning forecasts earn lower returns in the subsequent periods.

4.2 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. In addition, Baker and Wurgler (2013) argue that corporate managers have more information about their firms than investors have and can use that informational advantage. Hence, 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 that managers' expectations will align more closely to our statistically optimal benchmark.[19] For example, managers may issue more stock 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.

5 Empirical Findings

5.1 Earnings forecasts via machine learning

Table 2 compares the properties of analysts' earnings forecasts with the statistically optimal forecasts estimated using our machine learning algorithm (Random Forests).

[Insert Table 2 about here]

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

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 0.028 to 0.384 on average. All of these 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 absolute value of around 0.001 for the quarterly earnings forecasts, 0.027 for the one-year- ahead forecast, and -0.004 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 more accurate than the forecasts provided by analysts.

Figure 17 and 18 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 those 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.

[Insert Figure 17 and 18 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 closing stock price in the most recent month, as consistent with the previous literature (Engelberg et al. (2018)). The second-to-last column of Table 2 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 at longer horizons.

Figure 6 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 clear spikes during the Internet bubble of the early 2000s (Griffin et al. (2011)) and in the financial crisis.

[Insert Figure 6 about here]

5.2 Conditional bias and the cross-section of stock returns

We have demonstrated above that analysts are, on average, over-optimistic relative to the machine-learning benchmark and their estimates get more precise when predicting at shorter horizons. If market participants' beliefs align closely with analysts' expectations, then we should observe negative return predictability. Stocks with a high conditional bias should earn lower returns than stocks with a lower conditional bias.[20]

We conduct monthly cross-sectional predictive regressions (following Fama and MacBeth) 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 across the multiple horizons as the predictor. 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 this bias score.

Table 3 shows the regression results. The first column in each panel of Table 3 reports the regression without control variables. We find that both the conditional bias and the bias score are associated with negative cross-sectional stock return predictability. The coefficient on the conditional bias is -0.054 with a t-statistic of -3.94. The coefficient on the bias score is also significantly negative with a t-statistic of -4.47. The R^2s for both regressions have values around 0.01.

[20]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.

[Insert Table 3 about here]

The second column in each panel of Table 3 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: the two-quarters, three-quarters, and two-years ahead forecast biases generate significant negative return predictability.[21] Moreover, conditional biases' return predictability remains consistent when we either scale conditional biases with total assets per share from the most recent fiscal period or drop stocks whose prices are lower than \$5. We report these and further robustness checks in the Appendix.

Table 4 reports the correlations between the bias measures and the control variables. We find that the conditional bias and the bias score are highly positively correlated. Moreover, the conditional bias is negatively correlated with size and momentum. Further, the conditional bias is positively correlated with the book-to-market ratio, idiosyncratic volatility, and return volatility. Accordingly, stocks with a smaller size, lower past cumulative returns, and a higher book-to-market ratio, idiosyncratic volatility, and return volatility, tend to have more over-optimistic expectations. In the appendix, we report the summary statistics of these variables.

[Insert Table 4 about here]

Additionally, we show that the results from the cross-sectional return regressions also hold in time-series regressions. We sort stocks into five quintile portfolios based on the con-

[21] We find that the forecast bias at the one-quarter and one-year-horizon does not predict stock returns significantly. The lack of return predictability is consistent with analysts predicting better for those horizons and arguably with analysts exercising more effort towards generating the one-quarter and one-year-ahead forecasts.

ditional bias. Table 5 reports the portfolio sorts. Two interesting patterns emerge. First, the value-weighted returns decrease in the conditional bias. A long-short portfolio of the extreme quintiles results in a return spread of −1.46% per month (t-statistic −5.11) for the average bias and −1.16% per month (t-statistic −3.83) 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 5 about here]

We further examine whether returns on this long-short strategy can be explained by leading asset pricing models. Table 6 Panel A reports the results of using the average conditional bias as the portfolio sorting variable. We find that the long-short strategy has a significant CAPM alpha of −1.85% per month, with a significantly positive market beta of 0.56. Columns four to seven show the regression results with the Fama-French three-factor (Fama and French (1993)) and five-factor models (Fama and French (2015)). Neither model can explain the documented return spread. The alpha in the three-factor model is −1.96% with a t-statistic of −8.64; the alpha in the five-factor model is −1.54% with a t-statistic of −5.84. Table 6 Panel B reports the long-short strategy using the bias score as the sorting variable, and we find consistent results.[22] 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 6 about here]

Since the magnitude and significance of the results seem large by usual standards, we conduct a placebo test in the appendix to shed further light on these results and place them in context. In particular, we replace the machine learning forecast with the future realized

[22] 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.

value and then compute the conditional bias. The implied returns of these forward-looking (and thus non-tradable) strategies are many times largerin magnitude than the ones implied by our (tradable) machine-learning forecasts.

5.3 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. Both studies find that analysts tend to have over-optimistic expectations for stocks in the short side of anomalies, which earn lower returns. However, as previously mentioned, the realized earnings value cannot be combined in real-time with analyst forecasts to construct a real-time bias measure that in turn is used to sort portfolios on. To shed light on this issue, we use our conditional bias measure to examine whether analysts have more conditional over-optimistic expectations on anomaly 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, 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)).[23] We follow the literature and sort stocks into ten portfolios based on the decile of each anomaly variable. We use the extreme deciles as the long and the short leg 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.[24] 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 this anomaly score. The long (short) leg is

[23]Table A12 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 A12 reports the monthly average returns (in percent) of the long-short anomaly portfolios.

[24]We exclude stocks for which we have fewer than rank scores, which occurs when not all the data inputs on the characteristics are available.

defined as the stocks in the top (bottom) decile portfolio.

[Insert Table 7 about here]

Table 7 Panel A presents the average anomaly score for portfolios sorted independently on the conditional bias and the anomaly score.[25] For each anomaly decile portfolio, the anomaly score ranges from 3.31 to 6.82, with the highest (lowest) score indicating the long (short) leg of the anomaly strategy. Table 7 Panel B reports the average number of stocks in each of the 10×5 portfolios. On average, we have around 50 stocks every month in each portfolio. Moreover, the average number of stocks per month for the portfolio with the highest conditional biases and the lowest anomaly score is 97, 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 (37 stocks). This implies that stocks with higher conditional biases tend to be anomaly shorts, that is, overpriced stocks.

Table 8 presents the value-weighted returns of the portfolios formed by sorting independently on the conditional bias and the anomaly score. The long-short portfolio using the anomaly score earns 1.36% per month (the t-statistic is 5.74). 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 (2.13% per month with a t-statistic of 6.37). In contrast, the anomaly spread equals 0.60% (with a t-statistic of 1.82) in the quintile group with the smallest conditional bias. The difference in average returns between these two quintile portfolios is significantly positive (1.52% per month with a t-statistic of 3.81). Further, we find that the short leg portfolio return decreases from 1.06% per month to −1.29% when we move from the first quintile of the conditional bias to the fifth quintile. These findings are consistent with anomaly payoffs arising from the overpricing of stocks with the most over-optimistic

[25] For the results shown in Tables 7 and 8, 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.

earnings expectations.

[Insert Table 8 about here]

The last two rows in Table 8 report the conditional biases for each of the 10 decile portfolios sorted 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 the anomaly-short and anomaly-long portfolio is 0.005 and significant at the 1% level (with a t-statistic of 4.81).

5.4 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 (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 machine learning benchmark.

We follow Fama and French (2008) to measure firm i's net stock issuances at the fiscal year end t as the natural logarithm of the ratio of the split-adjusted shares outstanding at the fiscal year end t to the split-adjusted shares outstanding at the fiscal year end $t-1$,

$$NSI_{i,t} = log(\frac{\text{Split\_adjusted\_shares}_{i,t}}{\text{Split\_adjusted\_shares}_{i,t-1}}) \tag{8}$$

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 9 Panel A reports the value-weighted average net stock issuance for stocks sorted in portfolios according to the conditional bias of analysts' forecasts as measured relative to our machine-learning forecast.

[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.

The net stock issuances increase monotonically in the conditional bias. Importantly, we find that firms in the quintile portfolio with the most optimistic earnings expectations issue significantly more stocks than firms 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 9 about here**]

Table 9 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 larger conditional bias issue more stocks. We also find that firms with smaller size, lower book-to-market ratios, and lower profitabilities tend to issue more stocks, consistent with the results in Baker and Wurgler (2002) and Pontiff and Woodgate (2008).

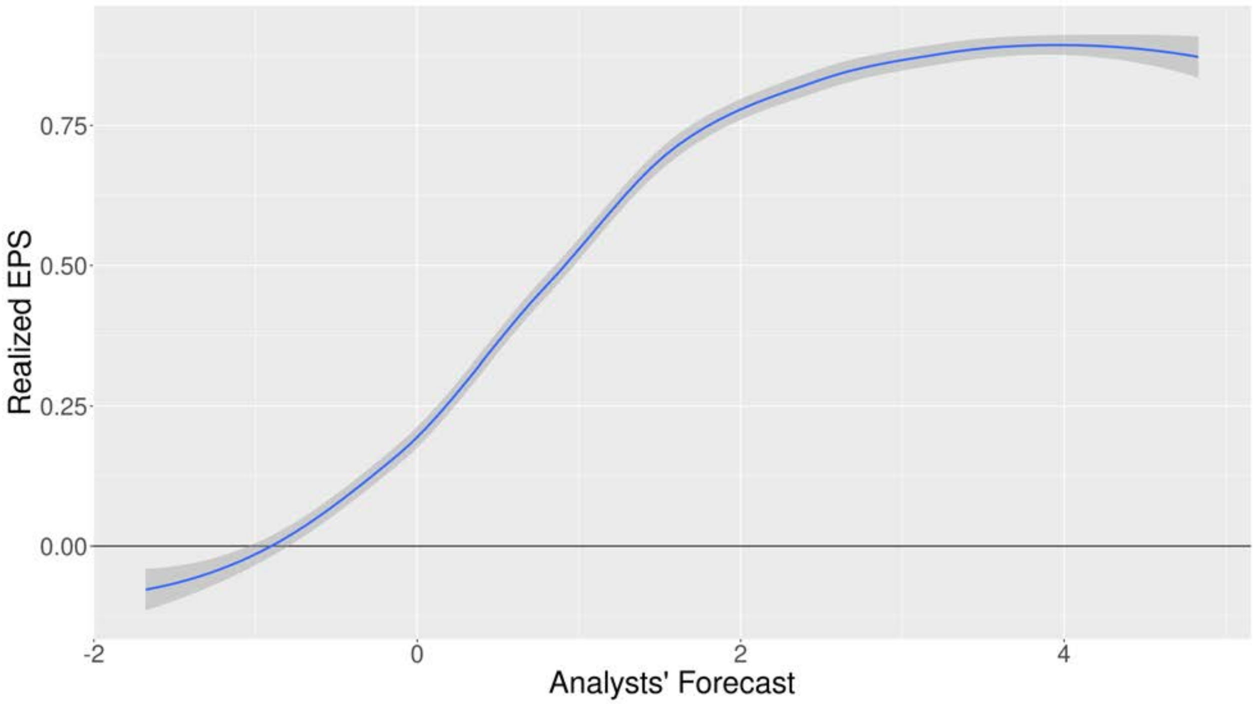
6 Conclusion

The pricing of assets relies significantly on the forecasts of associated cash flows. Analysts' earnings forecasts are often used as a measure of expectations, despite the common knowledge that these forecasts are on average biased upwards: 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 robust to variable selection bias. We demonstrate that, in contrast to linear forecasts, our new benchmark is effective out-of-sample.

This new measure is useful not only as an input to asset-pricing applications 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 forecasts are the most upward- (downward-) biased earn lower (higher) average returns going forward. This finding indicates that analysts' forecast errors may have an effect on asset prices.

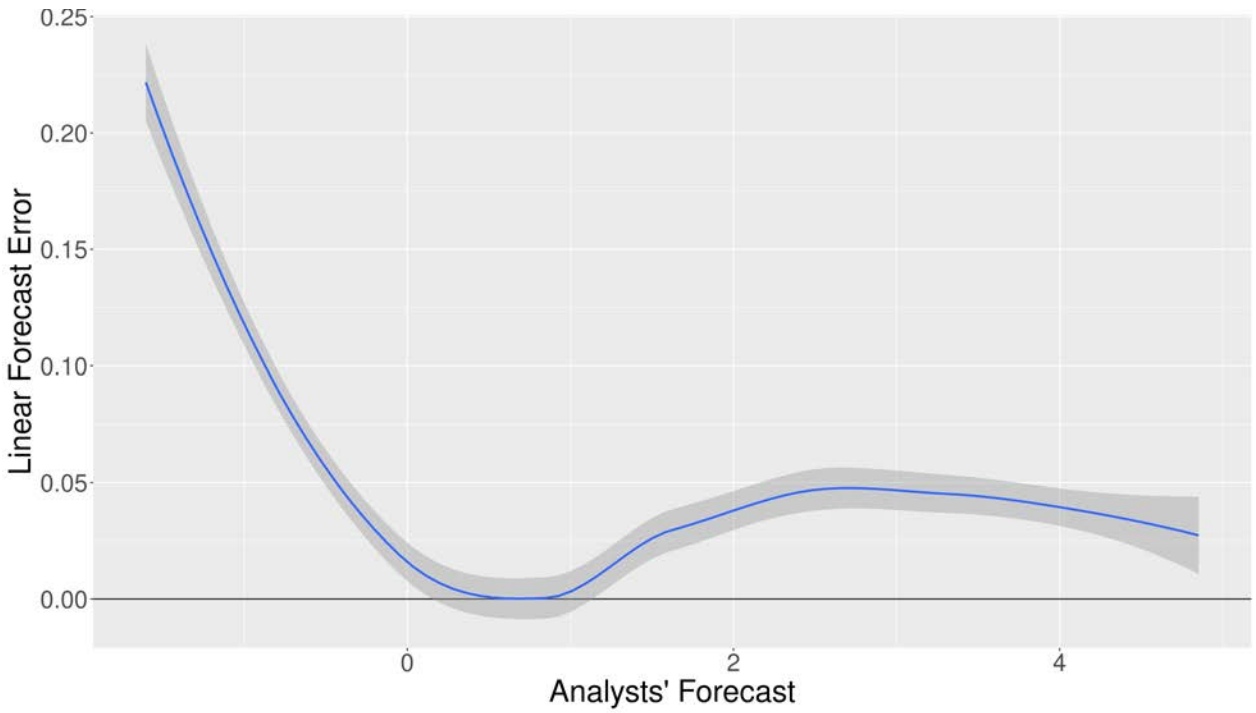
In addition to these asset-pricing results, our findings also have implications for corporate finance. Managers of firms for which the earnings forecast is most upward-biased issue more stocks. This finding indicates that managers are at least partially aware of analyst biases or the associated influence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

Figure 1: EPS as a non-Linear function of analysts' forecasts



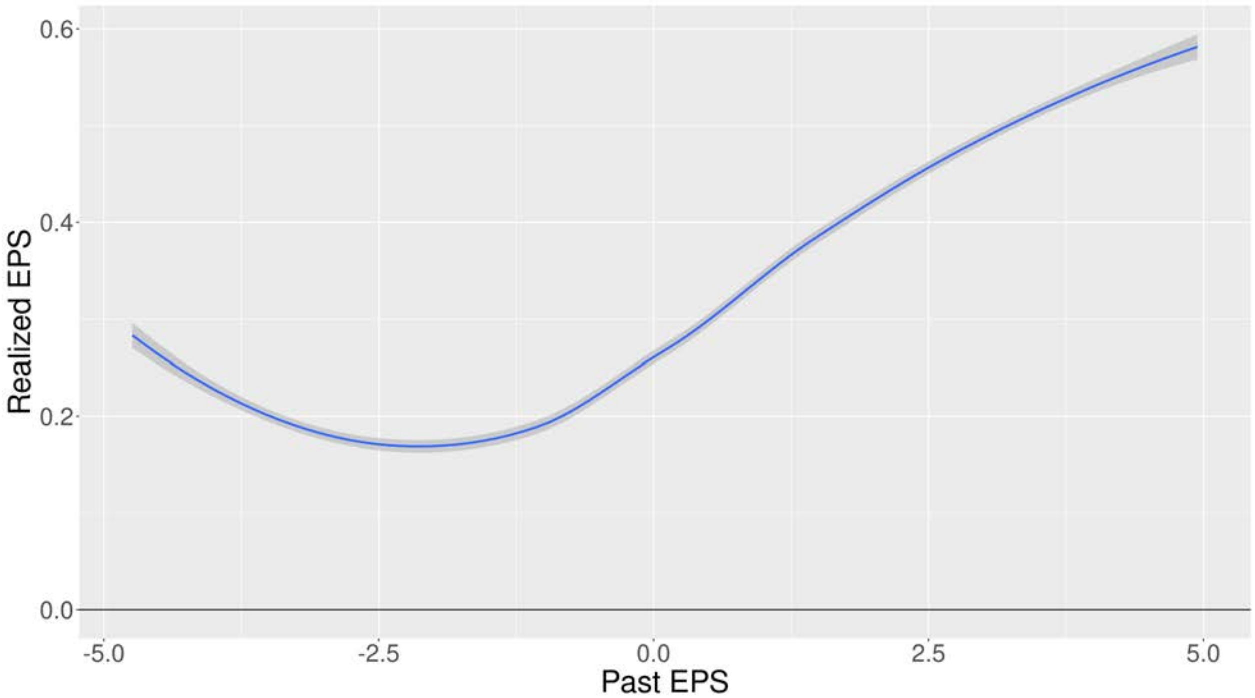
Notes: The figure plots the partial dependence plot of one-quarter-ahead realized EPS on analysts' forecasts. The partial dependence plot is calculated from a random forest regression of EPS on the variables mentioned in Section 3.2. The random forest regression for the figure uses 2000 trees and a minimum node size of 1. The data starts in 1986 and ends in 2019.

Figure 2: Linear forecast error as a non-linear function of analysts' forecasts



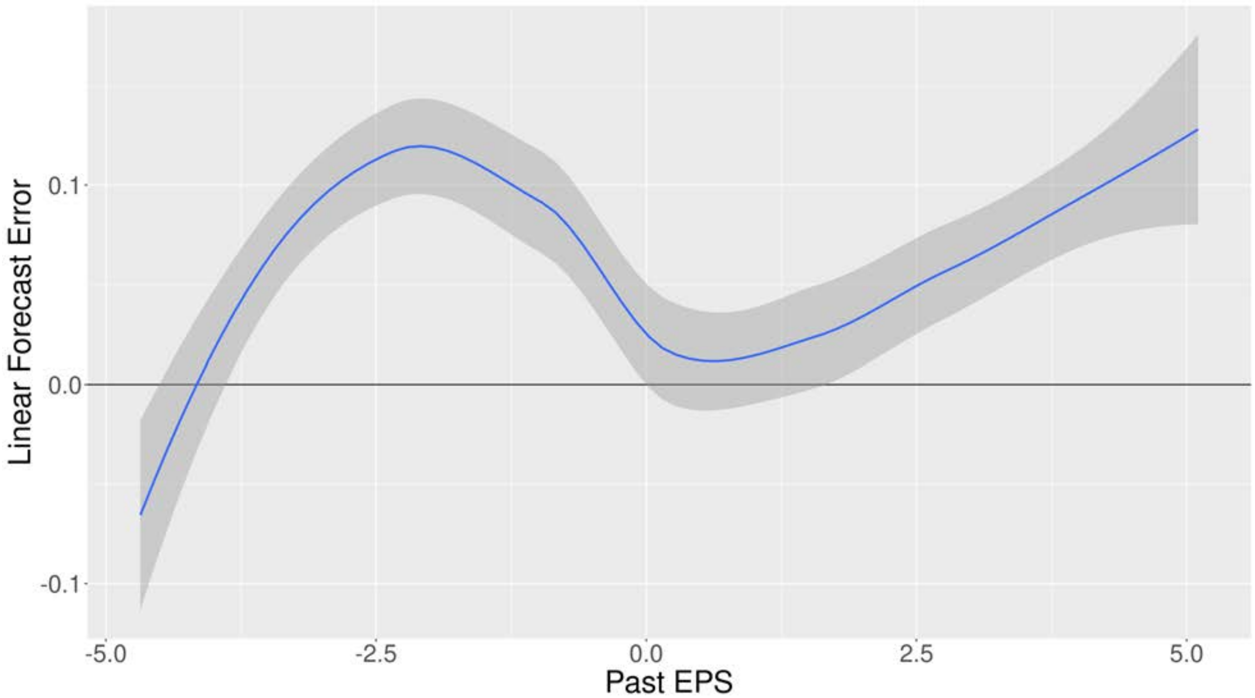
Notes: The figure plots the partial dependence plot of one-quarter-ahead realized linear errors on analysts' forecasts. The linear errors are calculated as the difference between the linear forecast and the realized EPS. The partial dependence plot is calculated from a random forest regression of the linear errors on the variables mentioned in Section 3.2. The random forest regression for the figure uses 2000 trees and a minimum node size of 1. The data starts in 1986 and ends in 2019.

Figure 3: EPS as a non-linear function of past EPS



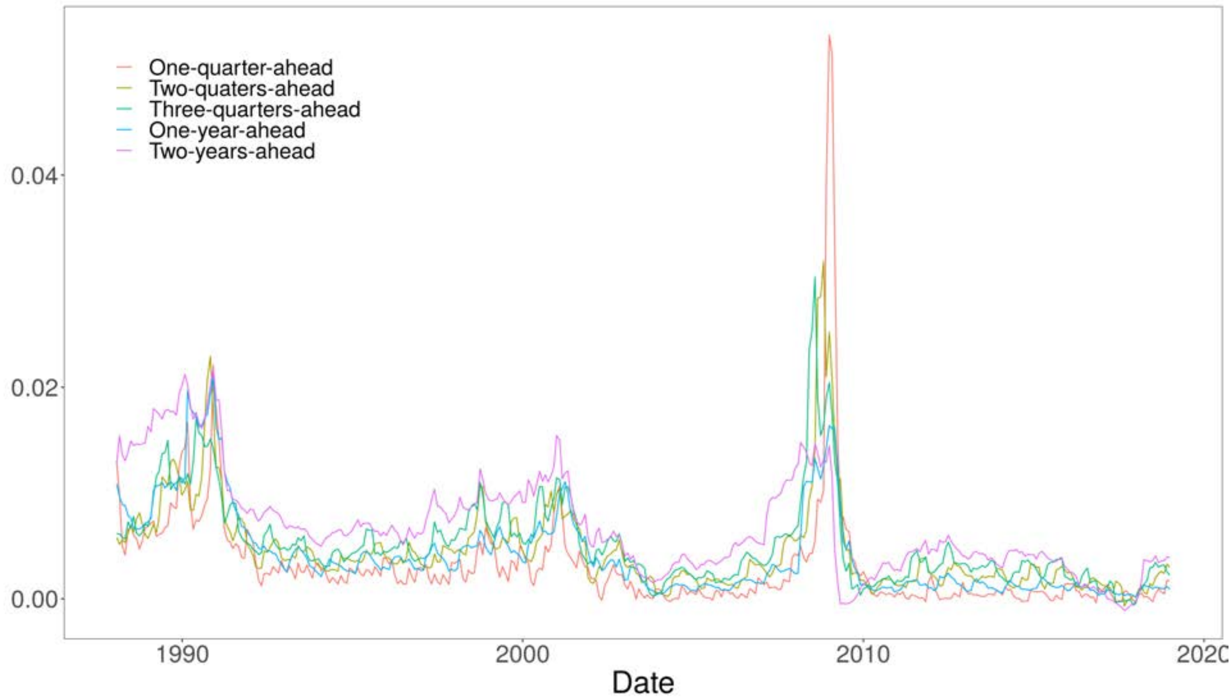
Notes: The figure plots the partial dependence plot of one-quarter-ahead realized EPS on past EPS. The partial dependence plot is calculated from a random forest regression of EPS on the variables mentioned in Section 3.2. The random forest regression for the figure uses 2000 trees and a minimum node size of 1. The data starts in 1986 and ends in 2019.

Figure 4: Linear forecast error as a non-Linear function of past EPS



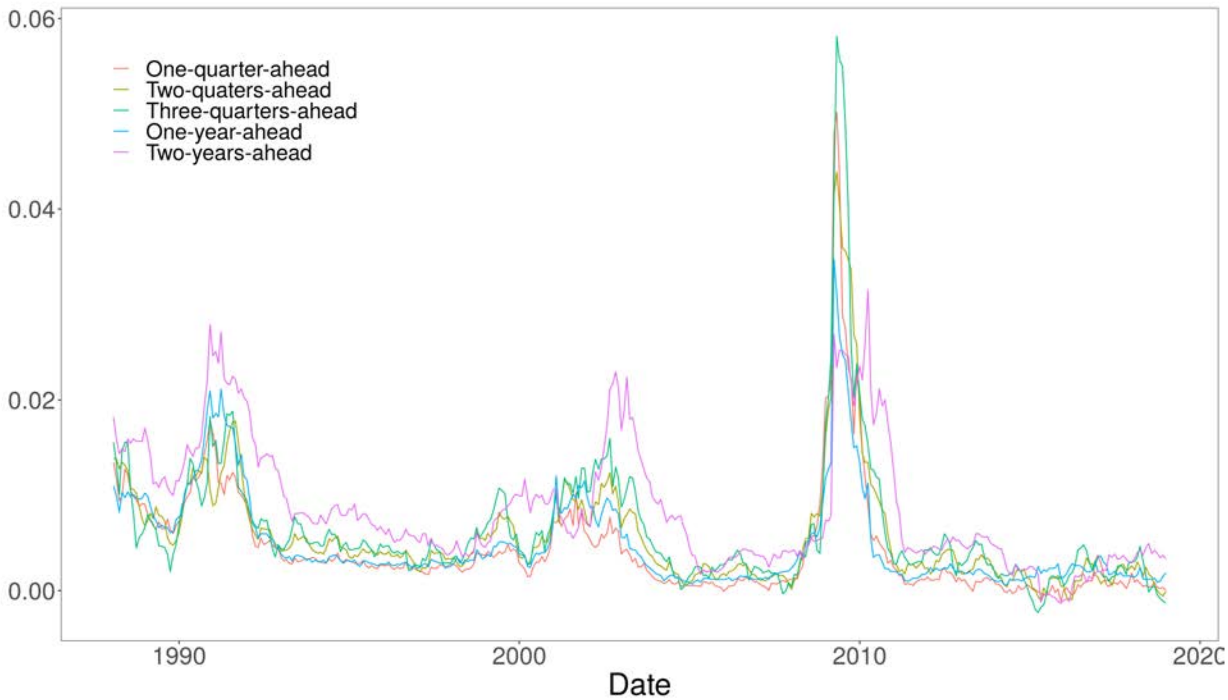
Notes: The figure plots the partial dependence plot of one-quarter-ahead realized linear errors on past EPS. The linear errors are calculated as the difference between the linear forecast and the realized EPS. The partial dependence plot is calculated from a random forest regression of the linear errors on the variables mentioned in Section 3.2. The random forest regression for the figure uses 2000 trees and a minimum node size of 1. The data starts in 1986 and ends in 2019.

Figure 5: Average realized bias of analysts' earnings expectations



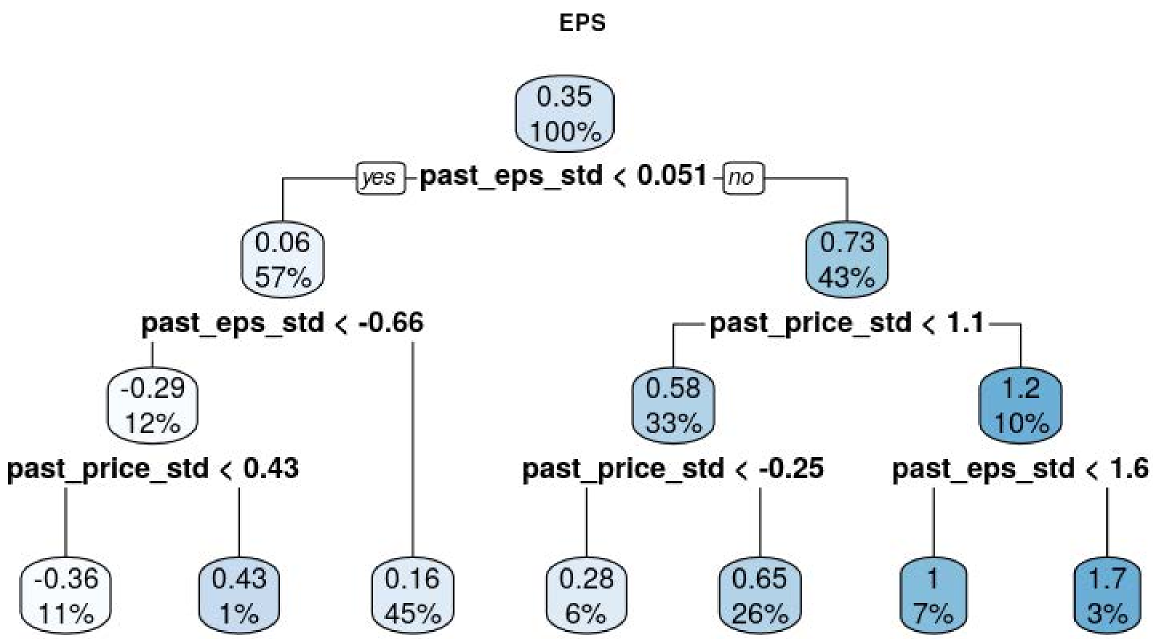
Notes: The figure plots the realized bias of analysts' expectations, which is measured as the average of the bias of expectations of individual firms. We trim the data at the 1% level each period before taking the average. The bias is calculated as the difference between analysts' earnings forecast and 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 6: Average bias of analysts' earnings expectations relative to machine learning forecasts



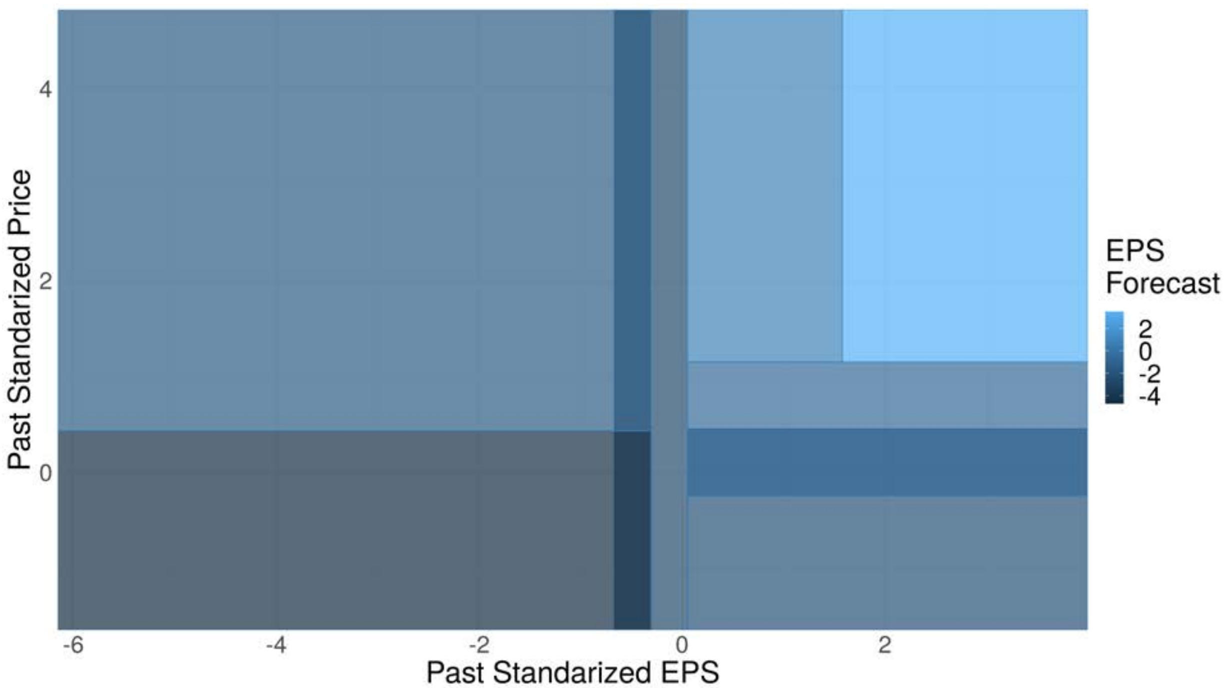
Notes: The figure plots the average conditional bias of analysts' expectations, which is measured as the average of the bias of expectations of individual firms. We trim the data at the 1% level each period before taking the average. The bias is calculated as the difference between analysts' earnings 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 7: Example decision tree



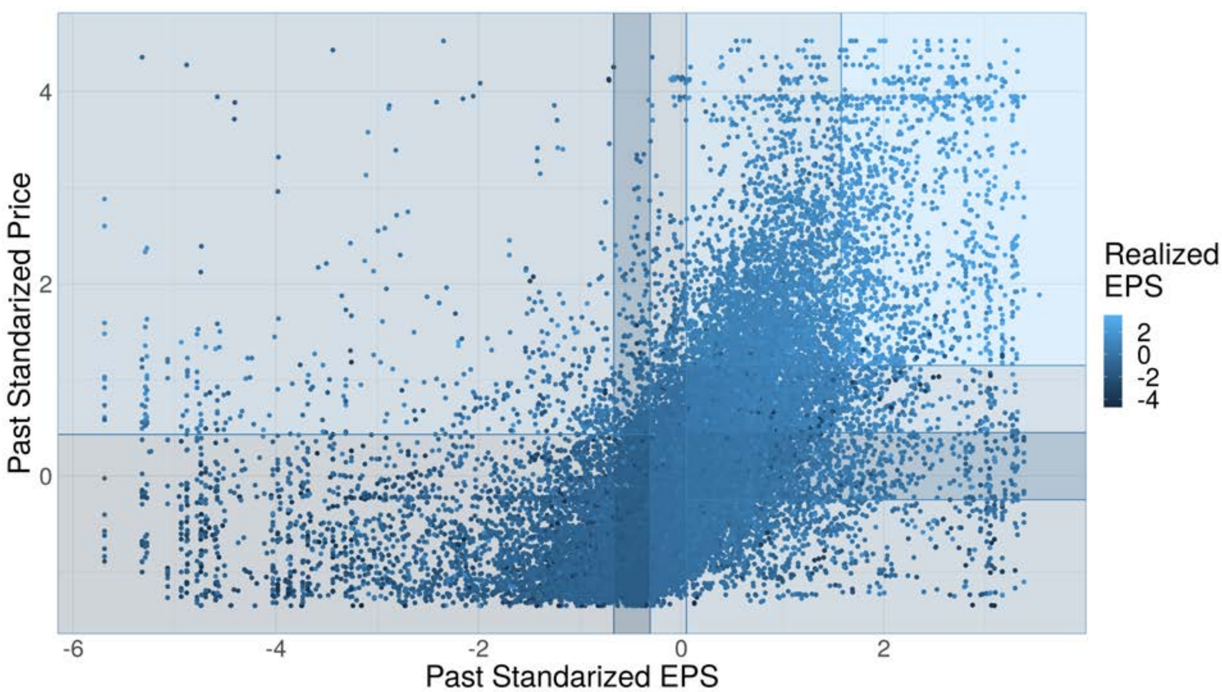
Notes: The figure shows an example decision tree. The variable we wish to forecast is the earnings-per-share (eps hereafter) for a cross-section of firms. At the first step, the selected explanatory variable is the past earnings per share (denoted by past_eps_std), and the threshold (or cutoff) value is at 0.051. Were we to end at this step, the forecasted eps value is .06 when past_eps_std is less than 0.051, and 0.73 when adj_afeps is more than or equal to 0.051. In the next step, the algorithm splits each of the previous two sub-spaces in two again. The first subspace (past earnings per share less than 0.051) is split in two using again the past earnings per share as an explanatory variable. The threshold value is -0.66. The second subspace (past earnings per share greater than 0.051) uses the price per share as the next conditioning variable, and the subspace considered is price per share below the threshold value of 1.1. The percentages show the proportion of the firms that fall in each of the splits. 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 8: Decision tree predictions



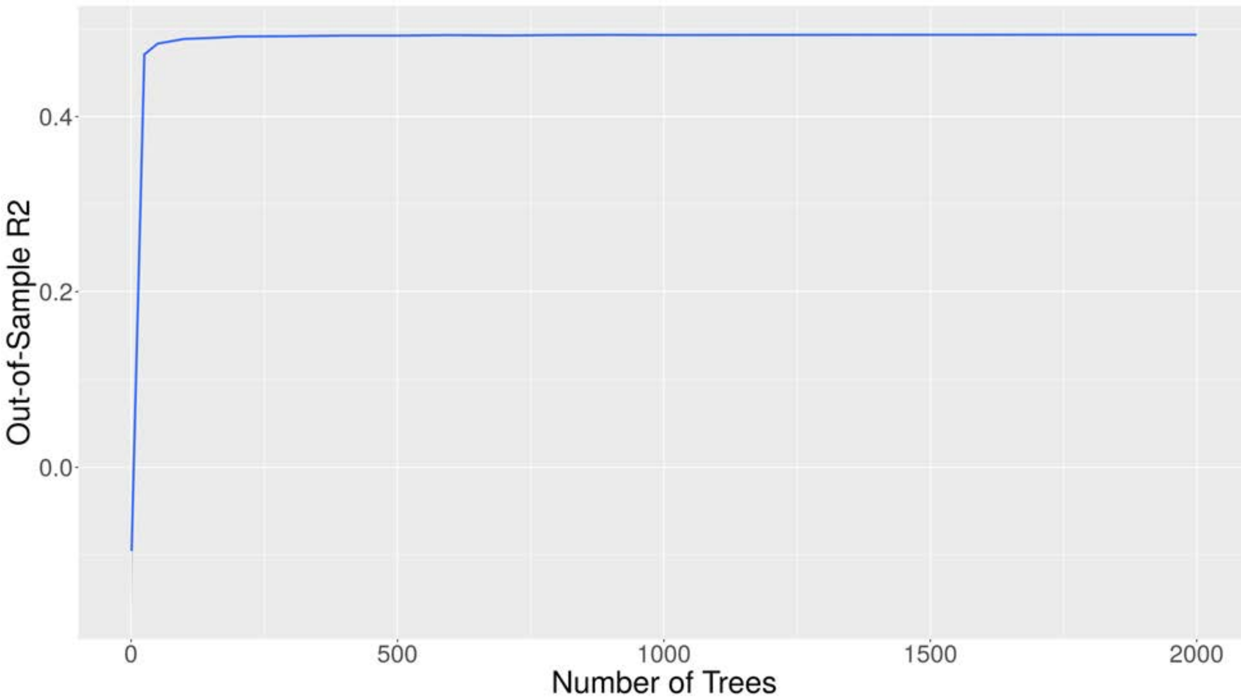
Notes: The figure shows the forecast of the decision tree from Figure 7. The variable we wish to forecast is the earnings-per-share for a cross-section of firms. The prediction is constant within each color box, and corresponds to the historical mean for each sub-space.

Figure 9: Decision tree predictions



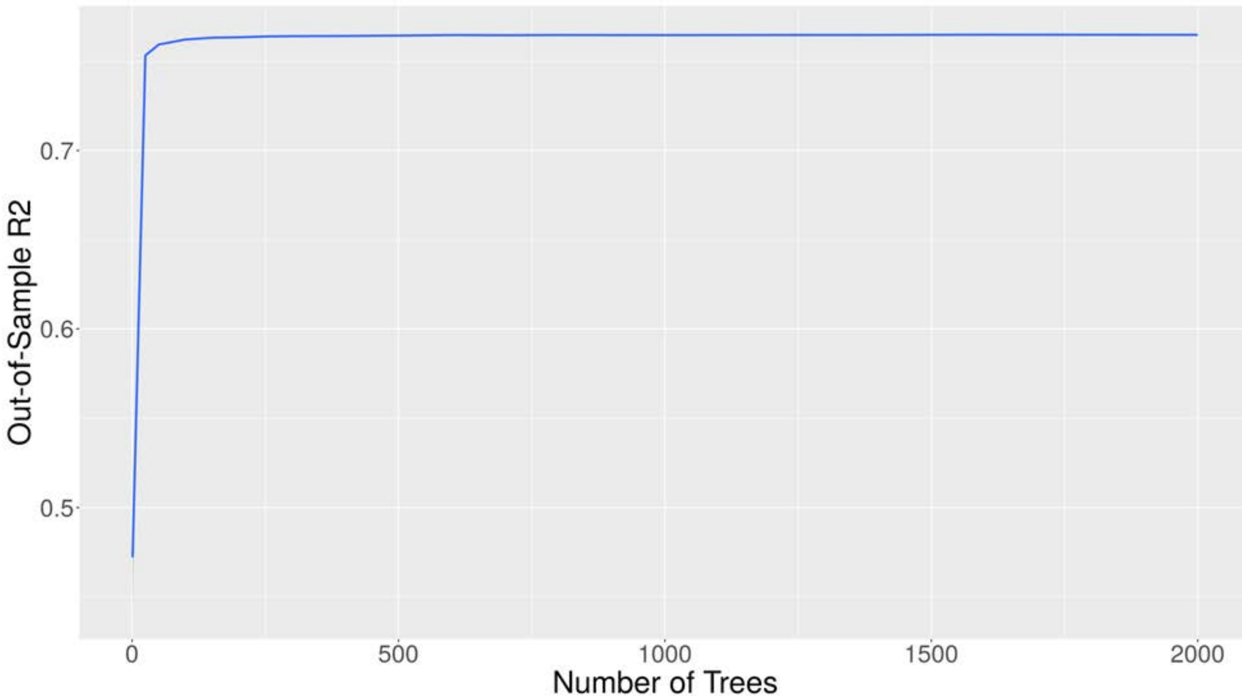
Notes: The figure shows the forecast of the decision tree from Figure 7. The variable we wish to forecast is the earnings-per-share for a cross-section of firms. The prediction is constant within each color box, and corresponds to the historical mean for each sub-space. The realized values are shown with a different color indicating a different value.

Figure 10: Cross-validation results of the number of trees in the one-quarter-ahead forecast



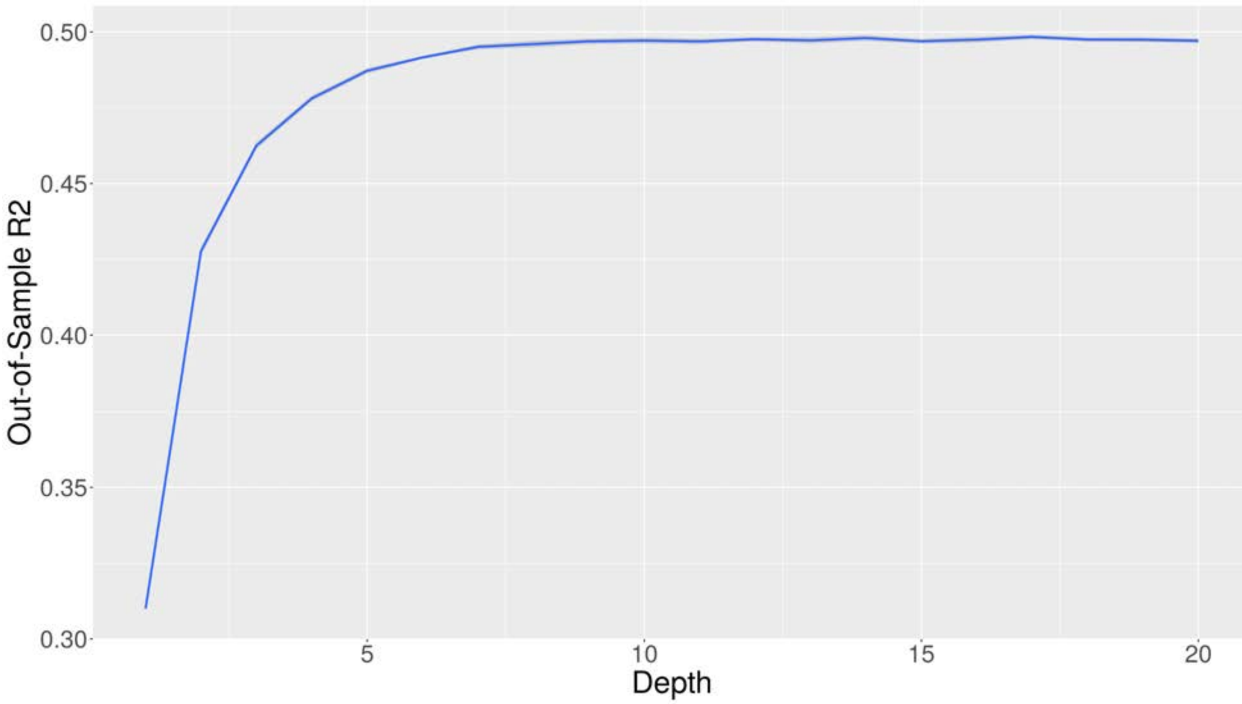
Notes: The figure plots the relation between the number of decision trees used in the random forest for training up to January 1986 and the out-of-sample R^2 value for the one-quarter-ahead earnings forecasts made in February 1986 for May 1986. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by the machine learning forecast divided by the mean squared error of 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 number of trees in the one-year-ahead forecast



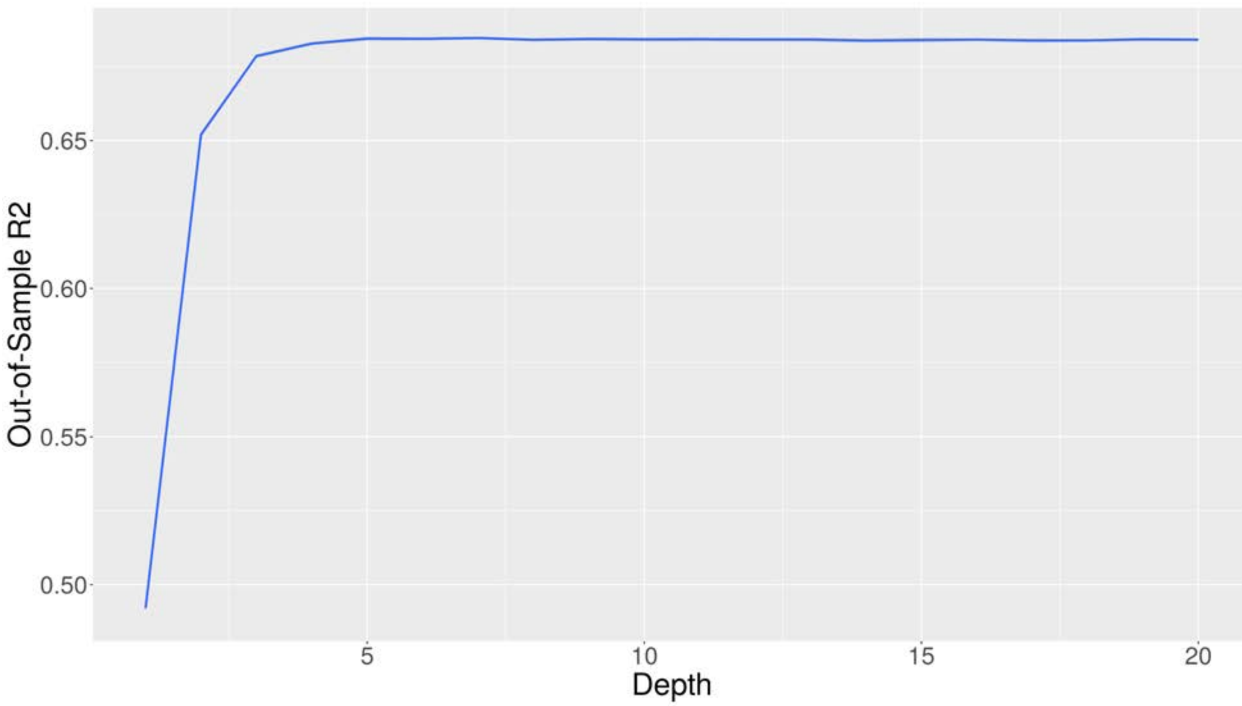
Notes: The figure plots the relation between the number of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-year-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using 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 12: Cross-validation results of the maximum depth of each tree in the one-quarter-ahead forecast



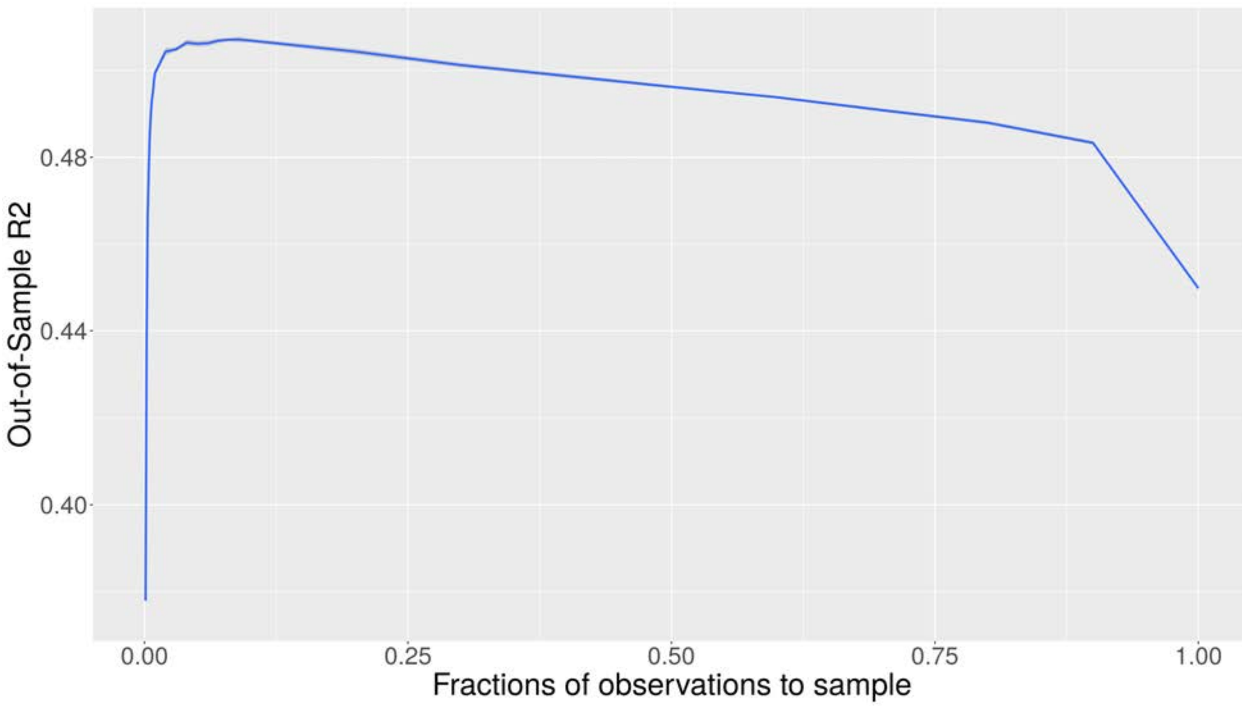
Notes: The figure plots the relation between the depth of of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-quarter-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by 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 13: Cross-validation results of the maximum depth of each tree in the one-year-ahead forecast



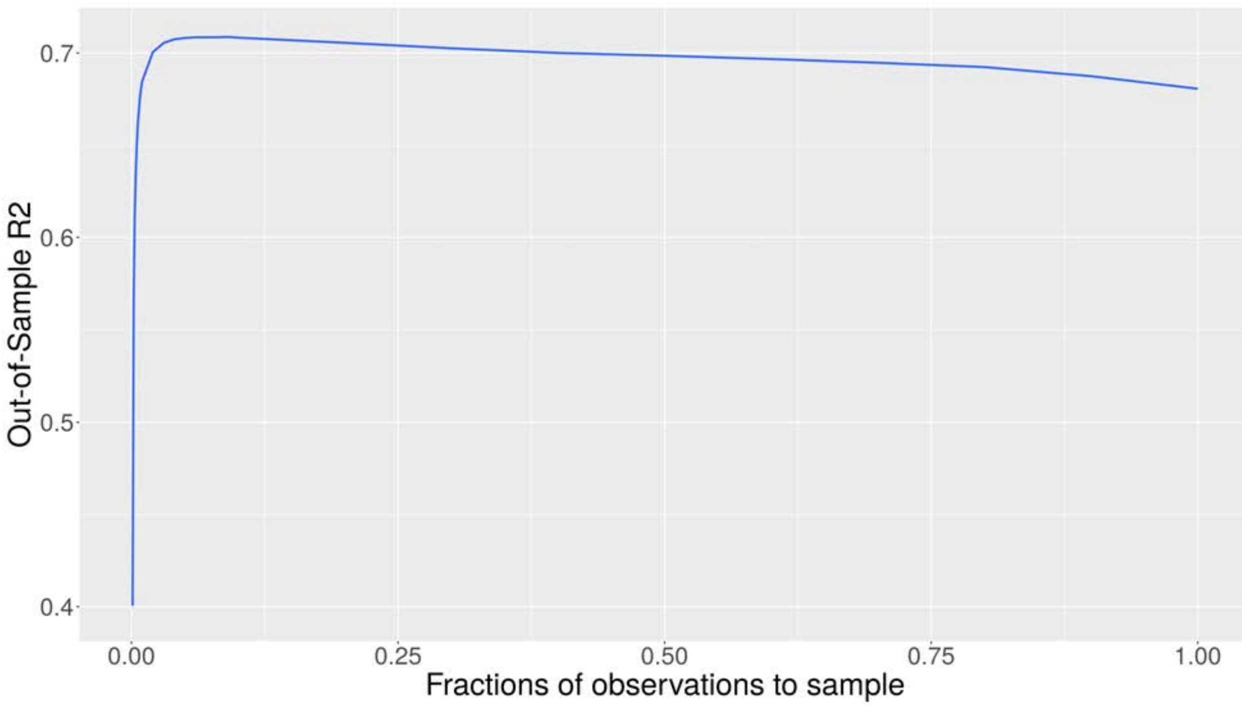
Notes: The figure plots the relation between the depth of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-year-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by using 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 14: Cross-validation results of the fraction of the sample that is taken in each split in the one-quarter-ahead forecast



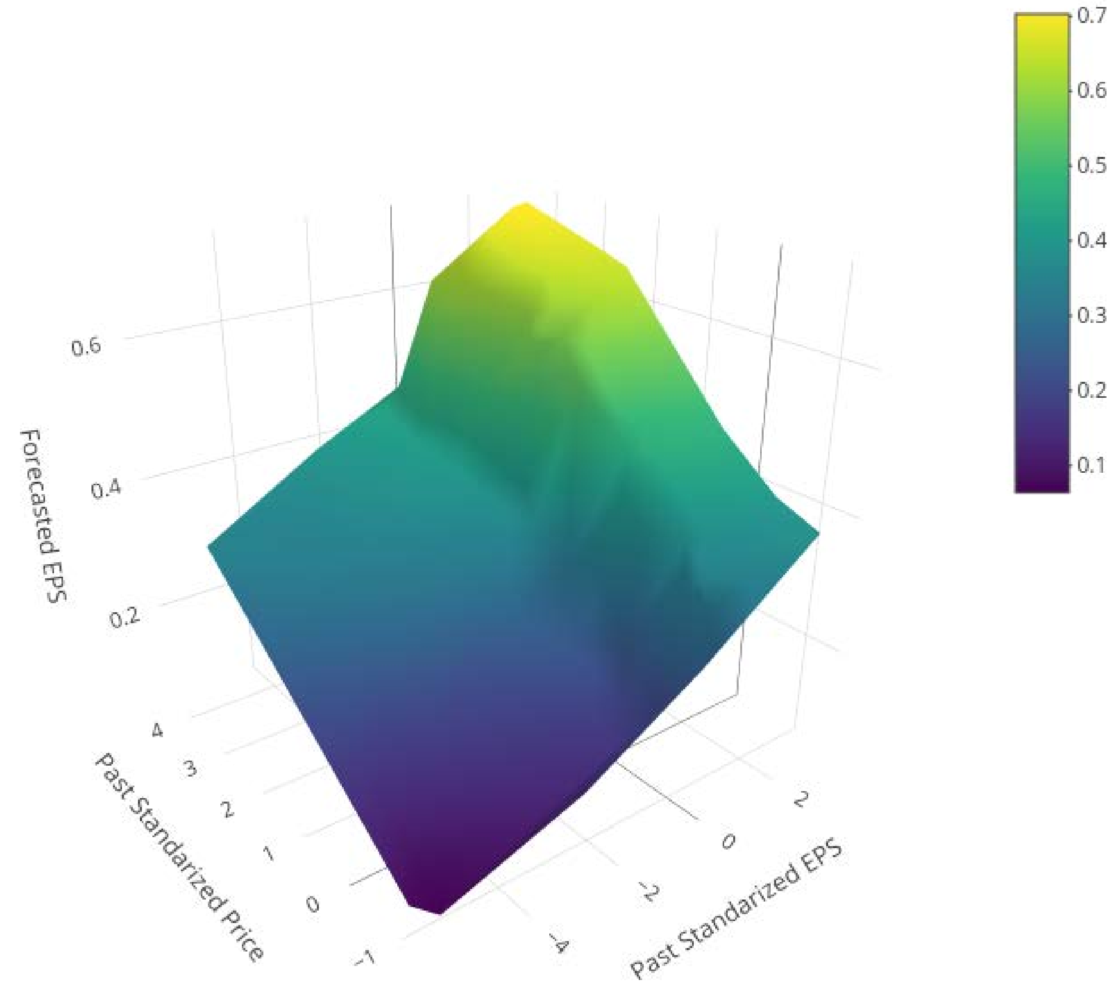
Notes: The figure plots the relation between the fraction of the sample that is taken in each split used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-quarter-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared 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 15: Cross-validation results of the fraction of the sample that is taken in each split in the one-year-ahead forecast



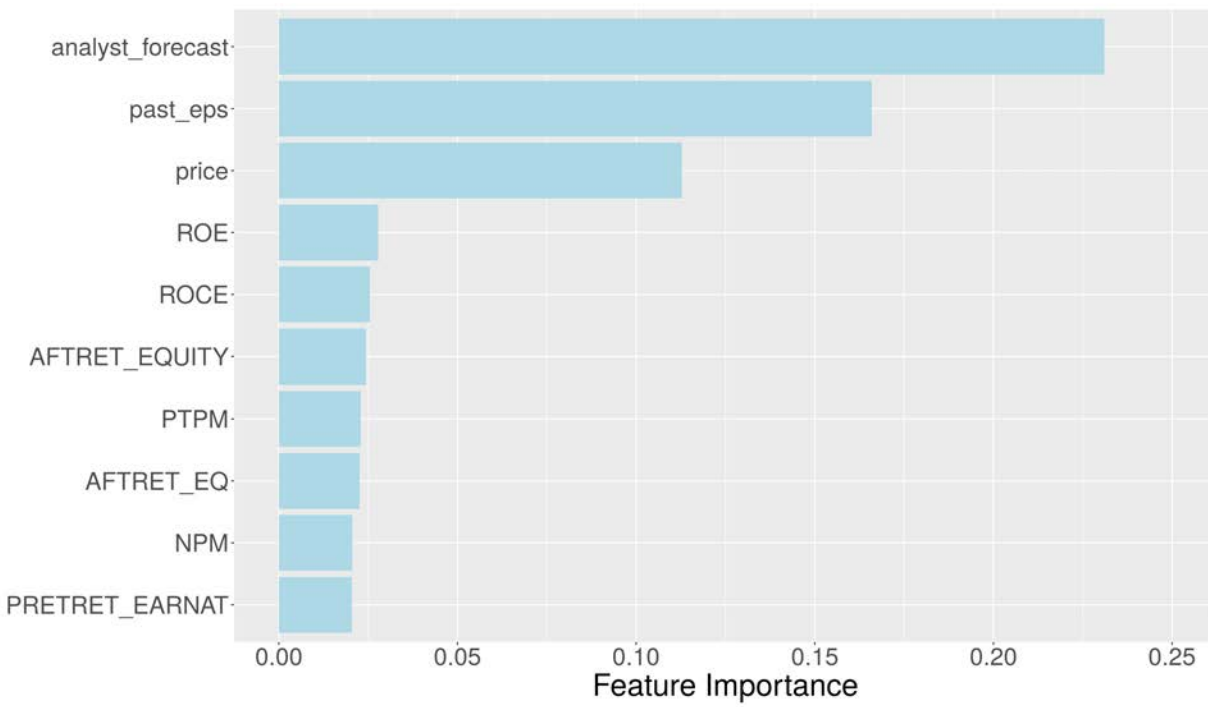
Notes: The figure plots the relation between the fraction of the sample that is taken in each split used in the random forest for training up to 1986 January and the out-of-sample R^2 for the one-year-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared 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.

Figure 16: EPS as a non-linear function of stock price and past EPS



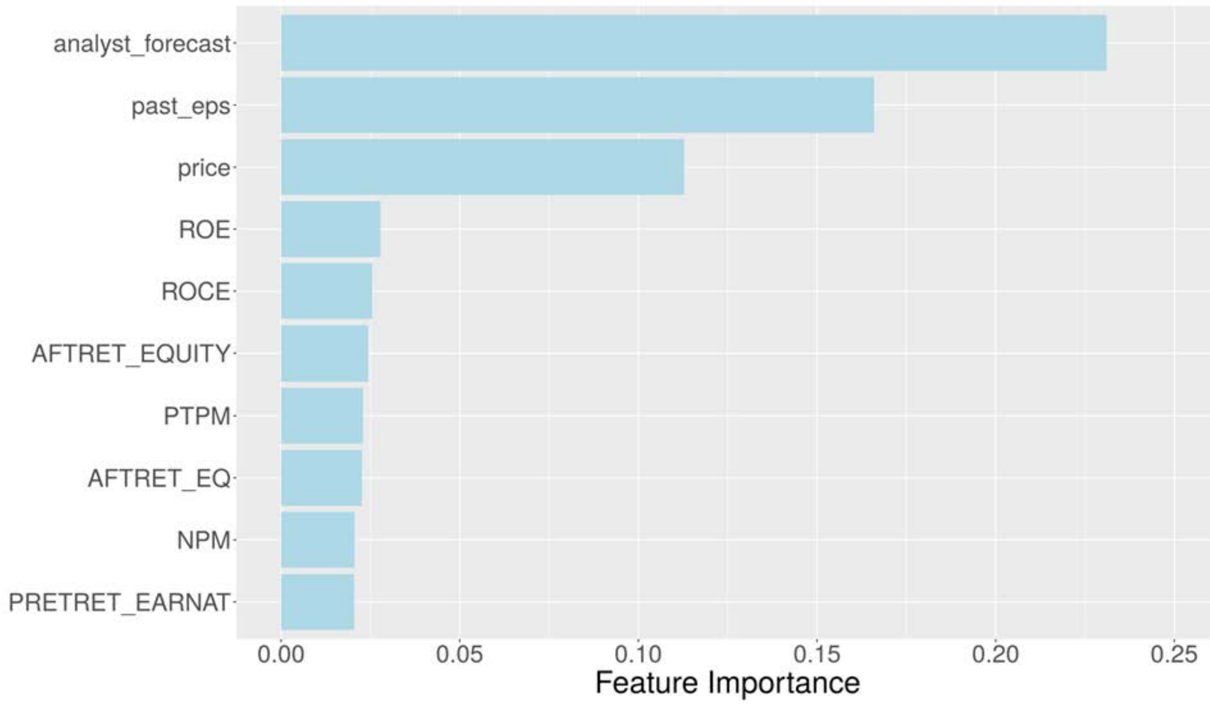
Notes: The figure plots the partial dependence plot of one-quarter-ahead realized EPS on past EPS and stock price. The partial dependence plot is calculated from a random forest regression of EPS on the variables mentioned in Section 3.2. The random forest regression for the figure uses 2000 trees and a minimum node size of 1. The data starts in 1986 and ends in 2019.

Figure 17: Feature importance of the one-quarter-ahead forecast



Notes: The 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 18: Feature importance of the one-year-ahead forecast



Notes: The 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.

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: The term structure of earnings forecasts via machine learning

Notes: This table presents the time series average of machine learning earnings per share forecasts (RF), 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. We report the Newey-West (Newey and West (1987)) 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 when reporting the Newey-West t-statistics. The sample period is January 1986 to December 2019.

	RF	AF	AE	(RF-AE)	(AF-AE)	$(RF-AE)^2$	$(AF-AE)^2$	(AF-RF)/P	N
One-quarter-ahead	0.290	0.319	0.291	-0.000	0.028	0.076	0.081	0.012	1,022,661
t-stat				-0.17	6.59			6.50	
Two-quarters-ahead	0.323	0.376	0.323	-0.001	0.053	0.094	0.102	0.024	1,110,689
t-stat				-0.13	10.31			8.16	
Three-quarters-ahead	0.343	0.413	0.341	0.002	0.072	0.121	0.132	0.031	1,018,958
t-stat				0.31	11.55			10.24	
One-year-ahead	1.194	1.320	1.167	0.027	0.154	0.670	0.686	0.057	1,260,060
t-stat				1.64	6.24			6.04	
Two-years-ahead	1.384	1.771	1.387	-0.004	0.384	1.897	2.009	0.190	1,097,098
t-stat				-0.07	8.33			7.17	

Table 3: Fama-Macbeth regressions

Notes: This table reports the Fama-MacBeth cross-sectional regressions of monthly stocks' returns on the conditional bias at 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 closing stock price from the most recent month, 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 the log of firm size (Lnsize), the log of book-to-market ratio (Lnbeme), the short-term reversal (Ret_1), the medium-term momentum (Ret12_7), the investment-to-asset (IA), the idiosyncratic volatility (IVOL), the return volatility (Retvol), and the share turnover (Turnover). We report the time-series average of slope coefficients associated with Fama-MacBeth t-statistics (in parentheses). The sample period is 1986 to 2019.

$$R_{i,t+1} = \alpha_{t+1} + \beta_1 BE_{i,t} + \gamma_j \sum_{j=1}^{8} Control_{j,i,t} + \epsilon_{i,t+1}$$

	Panel A: Average BE		Panel B: BE Score	
	(1)	(2)	(1)	(2)
Bias	-0.054	-0.064	-0.017	-0.028
t-stat	-3.94	-5.08	-4.47	-11.27
LNsize		-0.079		-0.215
t-stat		-2.22		-6.42
LNbeme		0.091		0.178
t-stat		1.58		3.14
Ret1		-2.818		-2.987
t-stat		-6.72		-7.12
Ret12_7		0.442		0.220
t-stat		2.88		1.52
IA		-0.003		-0.003
t-stat		-5.67		-5.88
IVOL		-0.224		-0.198
t-stat		-2.04		-1.80
Retvol		0.137		0.168
t-stat		1.19		1.47
Turnover		-0.065		-0.046
t-stat		-1.46		-1.03
Intercept	1.022	2.320	1.865	5.362
t-stat	3.64	4.41	7.89	11.35
R^2 (%)	0.780	5.680	1.242	5.756

Table 4: 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 different forecast horizons. "BE Score" denotes the average of the percentile ranking of the conditional bias of different forecast horizons. The characteristics include the log of firm size (Lnsize), the log of book-to-market ratio (Lnbeme), the short-term reversal (Ret_1), the medium-term momentum (Ret12_7), the investment-to-asset (IA), the idiosyncratic volatility (IVOL), the return volatility (Retvol), and the 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 sample period is 1986 to 2019.

Variable	Average BE	BE Score	BE_Q1	BE_Q2	BE_Q3	BE_A1	BE_A2	LNsize	LNbeme	Ret12_7	Ret1	IA	IVOL	RetVol	Turnover
Average BE	1.000														
BE Score	0.407	1.0													
BE_Q1	0.603	0.376	1.0												
BE_Q2	0.680	0.437	0.72	1.0											
BE_Q3	0.664	0.452	0.603	0.692	1.0										
BE_A1	0.689	0.366	0.627	0.624	0.538	1.0									
BE_A2	0.905	0.399	0.361	0.48	0.491	0.388	1.0								
LNsize	-0.223	-0.495	-0.222	-0.259	-0.249	-0.234	-0.191	1.0							
LNbeme	0.083	0.17	0.111	0.115	0.102	0.1	0.059	-0.179	1.0						
Ret12_7	-0.107	-0.18	-0.122	-0.136	-0.128	-0.116	-0.085	0.13	-0.051	1.0					
Ret1	0.002*	-0.032	0.009*	-0.006*	-0.014	0.012*	-0.009*	0.075	0.014	0.018	1.0				
IA	-0.001*	0.017	-0.013	-0.008	0.000*	-0.015	0.013	-0.059	-0.179	-0.017	-0.021	1.0			
IVOL	0.247	0.365	0.272	0.285	0.263	0.28	0.202	-0.466	-0.052	-0.093	-0.023	0.115	1.0		
RetVol	0.238	0.35	0.262	0.272	0.252	0.27	0.194	-0.428	-0.064	-0.08	-0.024	0.118	0.975	1.0	
Turnover	-0.016*	0.007*	-0.012	-0.004*	0.003*	-0.027	0.007*	0.059	-0.168	0.097	0.005*	0.123	0.245	0.277	1.0

Table 5: Portfolios sorted on conditional bias

Notes: This table reports the time series average of returns (in percent) on value-weighted portfolios formed on the conditional bias at 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 1986 to 2019.

Quintile	1	2	3	4	5	5-1
Panel A: Average BE						
Mean	1.32	0.98	0.79	0.47	-0.14	-1.46
t-stat	6.53	4.53	3.18	1.62	-0.35	-5.11
CAPM Beta	0.90	0.97	1.09	1.22	1.46	0.56
Panel B: BE Score						
Mean	1.14	0.93	0.79	0.60	-0.02	-1.16
t-stat	5.66	4.22	3.18	2.06	-0.05	-3.83
CAPM Beta	0.90	0.99	1.10	1.21	1.51	0.61

Table 6: 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 1986 to 2019. The t-statistics are adjusted by the White's heteroscedasticity robust standard errors (White (1980)).

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

	CAPM		FF3		FF5	
	Coef (β)	t-stat	Coef (β)	t-stat	Coef (β)	t-stat
Panel A: Average BE						
Intercept	-1.85	-7.18	-1.96	-8.64	-1.54	-5.84
Mkt_RF	0.56	7.53	0.53	7.86	0.38	5.28
SMB			0.80	7.06	0.61	5.17
HML			0.58	5.25	0.95	7.12
RMW					-0.68	-4.10
CMA					-0.53	-1.93
Panel B: BE Score						
Intercept	-1.58	-5.76	-1.69	-6.91	-1.17	-4.49
Mkt_RF	0.61	7.63	0.56	7.45	0.39	5.27
SMB			0.88	8.17	0.62	5.27
HML			0.56	4.29	0.97	7.05
RMW					-0.91	-5.15
CMA					-0.51	-1.90

Table 7: 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 1986 to 2019.

	Anomaly Decile										
BE Quintile	S	2	3	4	5	6	7	8	9	L	L-S
Panel A: Anomaly Score											
BE quintile	1	2	3	4	5	6	7	8	9	10	10-1
1	3.35	3.97	4.37	4.70	4.99	5.26	5.54	5.85	6.22	6.81	3.46
2	3.34	3.98	4.37	4.70	4.99	5.26	5.54	5.85	6.23	6.82	3.47
3	3.31	3.97	4.37	4.69	4.99	5.26	5.54	5.85	6.22	6.83	3.52
4	3.24	3.96	4.37	4.69	4.99	5.25	5.54	5.85	6.23	6.87	3.62
5	3.21	3.95	4.37	4.69	4.99	5.26	5.53	5.85	6.23	6.91	3.71
All stocks	3.31	3.97	4.37	4.70	4.99	5.26	5.54	5.85	6.23	6.82	3.51
Panel B: Number of stocks											
1	37	47	52	57	63	64	66	67	65	62	
2	34	50	56	62	64	65	66	67	62	54	
3	51	59	61	62	60	59	58	58	56	54	
4	73	65	60	57	55	52	52	52	53	58	
5	97	70	60	53	49	48	46	48	51	60	
All stocks	292	291	289	291	291	288	289	292	286	288	

Table 8: Returns on portfolios formed on conditional bias and anomaly score

Notes: This table reports the time series average of value-weighted 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	1.06	1.00	1.28	1.36	1.38	1.45	1.48	1.34	1.64	1.66	0.60
t-stat	2.73	3.21	4.84	5.40	5.43	6.25	6.90	6.60	7.91	7.09	1.82
2	0.29	0.76	0.99	1.06	0.94	0.90	1.10	1.02	1.33	1.38	1.09
t-stat	0.82	2.66	3.77	4.22	3.78	3.79	4.73	4.50	6.38	6.31	3.74
3	-0.16	0.40	0.64	0.60	0.68	1.11	0.92	1.02	1.21	1.06	1.23
t-stat	-0.43	1.24	2.23	2.14	2.52	4.13	3.65	4.06	4.72	4.06	4.40
4	-0.73	-0.31	0.51	0.58	0.30	0.64	0.74	0.80	1.04	0.81	1.54
t-stat	-1.75	-0.79	1.53	1.59	0.86	1.87	2.33	2.66	3.54	2.58	4.78
5	-1.29	-0.81	-0.41	-0.01	-0.06	0.27	0.25	0.29	0.90	0.84	2.13
t-stat	-2.62	-1.63	-0.97	-0.03	-0.14	0.61	0.59	0.69	2.04	1.99	6.37
5-1	-2.35	-1.81	-1.69	-1.38	-1.44	-1.18	-1.23	-1.05	-0.74	-0.83	1.52
t-stat	-6.04	-4.75	-5.02	-3.66	-3.84	-3.12	-3.36	-2.98	-1.92	-2.37	3.81
All Stocks	S	2	3	4	5	6	7	8	9	L	L-S
Return	-0.06	0.46	0.81	0.95	0.87	1.02	1.04	1.05	1.31	1.30	1.36
t-stat	-0.17	1.56	3.22	3.99	3.66	4.52	4.94	5.11	6.62	5.94	5.74
BE	0.009	0.007	0.005	0.004	0.004	0.004	0.004	0.003	0.004	0.004	-0.005
t-stat	5.83	5.24	6.19	6.05	5.59	5.76	6.02	5.73	5.02	4.71	-4.81

Table 9: 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 1986 to 2019. We report the time series average of slope coefficients associated with Newey-West t-statistics.

$$NSI_{i,t+1} = \alpha_{t+1} + \beta_1 BE_{i,t} + \gamma_j \sum_{j=1}^{3} Control_{j,i,t} + \epsilon_{i,t+1}$$

Panel A: Net Stock Issuances of Portfolios formed on BE						
Quintile	1	2	3	4	5	5-1
Average BE	0.006	0.012	0.017	0.028	0.065	0.059
t-stat	1.16	1.54	2.52	4.13	4.86	4.24
BE score	0.006	0.011	0.018	0.030	0.063	0.057
t-stat	0.99	1.50	3.37	5.58	4.32	3.69

Panel B: Fama-MacBeth regressions				
	A: Average BE		B: BE Score	
	(1)	(2)	(1)	(2)
Bias	0.442	0.355	0.072	0.039
t-stat	2.24	1.94	4.57	2.14
LNsize		-0.503		-0.484
t-stat		-2.91		-2.26
LNbeme		-2.042		-2.013
t-stat		-7.00		-6.41
EBITDA		-10.866		-10.949
t-stat		-4.96		-4.91
Intercept	3.470	9.473	0.526	7.904
t-stat	8.52	3.43	0.57	1.97
R^2 (%)	2.888	8.724	0.913	6.969

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.

Engelberg, J., McLean, R. D. and Pontiff, J. (2020). Analysts and anomalies, *Journal of Accounting and Economics* **69**(1): 101249.
URL: *https://www.sciencedirect.com/science/article/pii/S0165410119300448*

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.

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* **33**: 3–56.

Fama, E. and French, K. (2008). Dissecting anomalies, *Journal of Finance* **63**: 1653–1678.

Fama, E. and French, K. (2015). A five-factor asset pricing model, *Journal of Financial Economics* **116**: 1–22.

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

Frankel, R. and Lee, C. M. (1998). Accounting valuation, market expectation, and cross-sectional stock returns, *Journal of Accounting and economics* **25**(3): 283–319.

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.

Grennan, J. and Michaely, R. (2020). Artificial Intelligence and High-Skilled Work: Evidence from Analysts, *SSRN Electronic Journal* .
URL: *https://papers.ssrn.com/abstract=3681574*

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.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc., New York, NY, USA.

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.

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*

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**: 703–708.

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.

Richardson, S., Teoh, S. H. and Wysocki, P. D. (2004). The walk-down to beatable analyst forecasts: The role of equity issuance and insider trading incentives, *Contemporary accounting research* **21**(4): 885–924.

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.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica: journal of the Econometric Society* pp. 817–838.

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. Model

In this Appendix, we present a tractable non-linear model of earnings and earnings expectations that illustrates some reasons linear forecasts are inferior to those provided by machine learning techniques and analysts. In particular, a high variance of the relevant non-linear effects causes the linear models to behave poorly. The condensed version of this model is presented in the main paper in Section 2. The model also features asset prices, so that it can be used to further understand our return predictability results.

Economy

Consider the following setup. There are two periods in the economy. There is a measure 1 of assets to be priced, indexed by i. The payoff y_i of asset i is a random variable that is forecastable by a combination of linear and non-linear effects. In particular, the true payoff distribution follows:

$$\tilde{y_i} = f(x_i) + g(v_i) + z_i + w_i + \tilde{\epsilon}_i. \tag{A1}$$

Where v_i, w_i, x_i, z_i are variables measurable in the first period and distributed in the cross-section as independent standard normal. f and g are measurable non-linear functions, orthogonal to the space of linear functions in x_i and v_i respectively. That is, f and g satisfy $E[xf(x)] = E[vg(v)] = 0$. This implies that the best linear approximation of the functions are constants given by $E[f(x)]$ and $E[g(v)]$ respectively.[27] We assume $E[(f(x) - E[f(x))^2] = var(f(x)) \equiv \sigma^2_{fx} > 1$ and $var(g(v)) \equiv \sigma^2_{gv}$, and assume that all second moments exist.

We further assume that analysts use $f(x_i)$ and w_i in their forecasts. However, they miss out on the effects of z_i as well as $g(v_i)$ either because they are not aware of the forecasting power of transformations of v_i, or alternatively, because they use linear transformations of v_i

[27]Examples of functions that satisfy the conditions are $f(x) = x^p$ where p is an even positive integer or any symmetric function around zero where $f(x) = f(-x)$.

only. Furthermore, we assume a high variance of $f(x_i)$, which will result in analyst forecasts being more accurate than linear forecasts, despite the linear forecast using *all* variables.

$\tilde{y}$ and $\tilde{\epsilon}_i$ are random variables measurable in the second period. $\tilde{\epsilon}_i$ is distributed as an independent standard normal. We assume that the agents have a large enough sample of these variables from past observations so that there is no estimation error of the coefficients. Notice that (due to the orthogonality assumption above) in a linear regression the true coefficients associated with x_i and v_i are zero. For tractability, we assume that the shock to earnings is not priced and the risk-free rate is equal to zero.

The reason why our theoretical model includes non-linear effects is that in our empirical specification, we document substantial non-linearities in the earnings process as a function of the explanatory variables. For example, analysts' forecasts are amongst the most important predictors, and Figure 1 shows that EPS is a non-linear function of analysts' forecasts. Hence, using the linear prediction produces substantial errors as shown in Figure 2. Figures 3 and 4 show the same problem arises when using past EPS, which is a key ingredient of linear forecasts such as in Frankel and Lee (1998) or So (2013).

[Insert Figure 1 and 2 about here]

[Insert Figure 3 and 4 about here]

As stated above, we assume that the shock to earnings is not priced and the risk-free rate is equal to zero. Let $\tilde{m}$ be the stochastic discount factor (SDF), then $Cov(\tilde{m}, \tilde{\epsilon}_i) = 0 \ \forall i$ and $E[\tilde{m}] = 1$.

Define $\mu_{i,j} = E[\tilde{y}_i | F_{i,j}]$, that is, the conditional expectation of a representative agent when using sigma algebra $F_{i,j}$ to form the expectation. The following result is immediate from the definition of conditional expectation:

Lemma 1 *If* $F_{i,j} \subseteq F_{i,k}$ *then* $E[(\tilde{y}_i - \mu_{i,k})^2] \leq E[(\tilde{y}_i - \mu_{i,j})^2]$.

Lemma 1 has two important implications.

First, including more variables in an ideal estimator will weakly decrease the error, since the estimator can always disregard the useless variables. For our application, random forest regression automatically discards useless variables and incorporates the information of useful ones. Given its flexibility and robustness it will always benefit from adding information, at least asymptotically.[28]

Second, if we include the conditional expectation, $\mu_{i,j}$ as a variable to use for prediction (e.g. analyst forecasts), in an optimal estimator, the error of the estimator must be at least as low as the error when using the conditional expectation $\mu_{i,j}$ as a forecast, since the optimal estimator can always ignore all of the information except for $\mu_{i,j}$.

Naturally, if we include the analysts expectation, which is in the public information set, any optimal estimator will achieve an error no higher than analysts. Formally, any conditional expectation is a function of observable variables, say $E[\tilde{y}_i|F_{i,j}] = G_{i,j}(x,z,w)$ in our setup, and observing $G_{i,j}(x,z,w) = \mu_{i,j}$ provides additional information and Lemma 1 applies. In practice, we find that when adding analysts' expectations, the squared error of the random forest prediction is lower than that of analysts, whereas the squared error of the linear model is higher than that of analysts.

Third, a predictor that is unconditionally biased, if it is not the conditional expectation, will be conditionally biased, since the conditional expectation and the predictor will differ in some information sets.

If all agents in the economy form expectations using the information set $F_{i,j}$, then the price of asset i is $P_i = \mu_{i,j}$ and the expected return from the point of view of the agents is $E[R_i|F_{i,j}] = \frac{E[\tilde{y}_i|F_{i,j}]}{\mu_{i,j}} = 1$.

The actual expectation of y_i is given by $\mu_i^* = E[\tilde{y}_i|F_j^*] = 1 + f(x_i) + g(v_i) + z_i + w_i$. The estimator may be unfeasible if the agents do not know the true functional form or cannot process all the variables. The (actual) expected return is then given by:

[28] Unfortunately, the addition of useless variables is not free due to finite sample sizes. At every step each decision tree chooses a finite number of variables, and if none of the variables provide information, the decision tree will waste a split and predict the mean from the previous node. In practice, random forests are very robust to adding useless features and can be modified to be more selective in the presence of very high-dimensional data.

$$E[R_i] = \frac{\mu_i^*}{\mu_{i,j}} \tag{A2}$$

Naturally, stocks with pessimistic (lower than optimal) predictions will have higher (realized) returns and vice-versa.

We now consider three different ways of forming expectations. First, let us consider linear forecasts: we assume that (1) agents have access to past realizations of the variables, (2) estimate the linear model precisely, but (3) only include first order terms. That is, they run a regression of the form:

$$y = a + b_x x + b_v v + b_z z + b_w w + u, \tag{A3}$$

and estimate a, b_x, b_v, b_z, b_w. For simplicity, we assume that they get accurate coefficients (up to specification) due to a large enough sample size: $a = 1 + E[f(x)] + E[g(v)], b_x = 0, b_v = 0, b_z = 1, b_w = 10$. Hence they form expectations equal to $\mu_l = E[y|\text{linear model}] = a + z + w = 1 + E[f(x)] + E[g(v)] + z + w$, where $E_i[\cdot]$ denotes a cross-sectional expectation. Notice that the resulting conditional expectation is (cross-sectionally) unbiased:

$$E_i[\mu_l] = E_i[a + z + w] = E_i[E[\tilde{y}]] = 1 + E[f(x)] + E[g(v)], \tag{A4}$$

where $E_i[\cdot]$ denotes a cross-sectional expectation. The linear model compensates for the lack of linearity in x and v by adding the unconditional expectation of $f(x)$ and $g(v)$ to the intercept.

Second, let us consider analyst expectations: we assume that analysts form expectations using x, v, and w, exclusively, for example because they can only process a certain amount of information. They also have access to the correct functional form of x, but not v, to illustrate specification uncertainty. Their resulting estimate is $\mu_a = E[y|\text{analyst}] = 1 + E[g(v)] + f(x) + w$.

Third, we form expectations using a non-linear function estimated by applying random forests to the past sample. Because of their flexibility, random forests can approximate any functional form, and (asymptotically) random forest are a consistent estimator of the conditional mean.[29] For simplicity, we consider the estimate to be: $\mu_{ML} = E[y|\text{machine learning}] = 1 + f(x_i) + g(v_i) + z_i + w_i$, but notice that in practice there is a finite (although large) sample size and the estimates are subject to sampling error.

The (asymptotic) mean squared error is $\sigma^2_{fx} + \sigma^2_{gv} + var(\epsilon)$ for the linear model, $var(z) + \sigma^2_{gv} + var(\epsilon)$ for analysts, and $var(\epsilon)$ for the machine learning forecast. We say that a forecast dominates another forecast if the mean squared error of the first is smaller than the mean squared error of the second. To match the empirical results, we assume $\sigma^2_{fx} > var(z) = 1$. Hence, within the model, as in our empirical findings, the machine learning forecast dominates the analyst's forecast, which in turn dominates the linear forecast.

We now assume that the economy-wide expectations of the agents coincide with the analyst expectations. Generally, assets with high bias with respect to the machine learning forecast will get lower returns. Since the machine learning is a better forecast, and approximates better the true conditional expectation, the returns will roughly follow:

$$E[R_i] = \frac{E[y_i|\text{machine learning}]}{E[y_i|\text{analyst}]}, \tag{A5}$$

and firms with overly optimistic forecasts with respect to the machine learning forecast will have lower average returns.

Spurious in-sample linear predictability

Even though analysts' forecasts dominate the linear forecasts, return predictability may still arise from the conditional bias measured by the difference between the analysts' forecasts and the linear forecasts, in two situations.

First consider the case where the linear forecast conditionally dominates the analysts' forecast. For example, for assets with $x = 0$ and $z \neq 0$, the linear model will dominate the analysts' forecast, and stocks with optimistic expectations will have lower returns. This is a consequence of Lemma 1, as non-optimal expectations can be conditionally biased.

[29] This is commonly referred to in the literature as random forest being universal approximators. We confirm in simulations that it applies in our setup.

Second, and more importantly, if the analysts' forecast and the linear forecast have a different loading on the variable z, and z induces a correlation between the payoff and the SDF, return predictability may arise from the conditional bias measured by the difference between the analysts' forecasts and the linear forecasts.

To illustrate the latter point formally, assume now that the SDF, $\tilde{M}$, has $E[\tilde{M}] = 1$, $E[\tilde{M}\tilde{\epsilon}] = 0$ and $Var(\tilde{M}) = 1$.

The payoff of asset i follows:

$$\tilde{y}_i = 1 + f(x_i) + g(v_i) + z_i + w_i + h(z_i)\tilde{f} + \tilde{\epsilon}_i, \tag{A6}$$

where $h : \mathbb{R} \to (0,1)$ is an increasing strictly positive function , $E[\tilde{f}] = 0$, $Var(\tilde{f}) = 1$ and $Corr(\tilde{f}, \tilde{M}) = Cov(\tilde{f}, \tilde{M}) = -a, a > 0$.[30]

We assume that regardless of the way agents form expectations, they are aware of the covariance with the SDF. The (conditional) covariance is then given by

$$Cov(\tilde{y}, \tilde{M}) = h(z_i)Cov(\tilde{f}, \tilde{M}) = -h(z_i)a. \tag{A7}$$

Hence, firms with higher z_i have higher returns, as the price is given by:

$$Price(y_i|F_{i,j}) = E[\tilde{M}\tilde{y}|F_{i,j}] = E[\tilde{y}|F_{i,j}] - h(z_i)a = \mu_{i,j} - h(z_i)a, \tag{A8}$$

and the expected return is given by:

$$E[R_i] = \frac{\mu_i^*}{\mu_{i,j} - h(z_i)a}. \tag{A9}$$

Notice that a simple portfolio sort using z will produce a spread in returns, since firms with lower z have lower returns. Notice as well that the difference between the analysts' forecast and the linear forecast is given by:

$$E[\tilde{y}|\text{analyst}] - E[\tilde{y}|\text{linear model}] =$$

[30]We assume a is small enough that none of the prices are zero.

$$1 + E[g(v)] + f(x) + w - (1 + E[g(v)] + E[f(x)] + z + w) = f(x) - E[f(x)] - z \qquad \text{(A10)}$$

In the model (and in the empirical results) analyst earnings estimates are better than linear forecasts. Nevertheless, the bias in the linear forecast appears to be correlated with differences in expected returns. If expected returns and biases are both correlated with a common variable z, then this return predictability can appear even when economically these biases in and of themselves are not the driver of the return predictability.[31]

To make matters worse, if the variable that is driving the return predictability only works in-sample then the out-of-sample linear model's return predictability will decrease substantially or disappear.[32] In our empirical specification, the linear model return predictability disappears after the 2000s.

In contrast, for the machine learning model the results from the previous section apply and assets with high bias with respect to the machine learning forecast get lower returns:

$$E[R_i] = \frac{E[y_i|\text{machine learning}]}{E[y_i|\text{analyst}] - h(z_i)}. \qquad \text{(A11)}$$

And consistent with the empirical results, the machine-learning return predictability remains stable.

A2. 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.[33] 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

[31]In the model x and z are independent cross-sectionally, x is unrelated to returns but firms with higher z will have higher returns, so a sort in z will produce differences in expected returns mechanically.

[32]In our model it would correspond to a change in the covariance with the SDF to zero. More generally, it can be caused by changes in market efficiency.

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

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.[34] 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,

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

For matching /I/B/E/S with CRSP, we use the link table provided by the Wharton Research Data Service. We require firms' historical CUSIP to be same in both /I/B/E/S and CRSP. We keep common stocks (share code 10 and 11) in stock exchanges of NYSE, AMEX, and NASDAQ (exchange code 1, 2, and 3).[35].

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 therefore 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 for the same fiscal end periods as analysts do. When matching the forecasts

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

[35]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 tilts towards big stocks (Diether et al. (2002))

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 i's 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 train 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 train 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 windows to train the model. Our training data starts in 1985 January, and our first forecast observations are in 1986 January.

A3. 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.[36] Since our predicted variable is earnings per share, we also consider another twenty-six fundamental values per share derived from these financial ratios such as book equity per share and current

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

debt per share for improving the forecasts.

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
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		

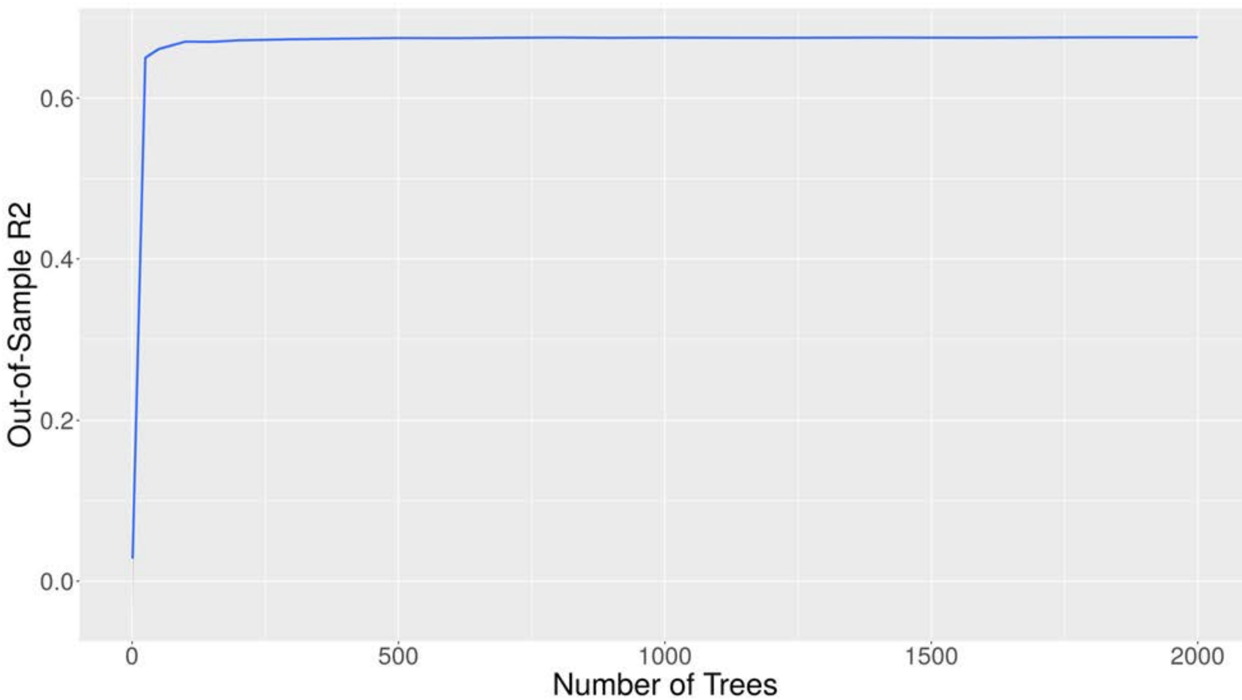
A4. Parameters in random forest

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 two partitions: training and testing (cross-validation). The training data contains the early part of the sample: from the beginning of the sample until December 1985. The testing data contains a single month: January 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 1985 for the whole sample, and we start our forecasts in 1986. 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 R^2 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.

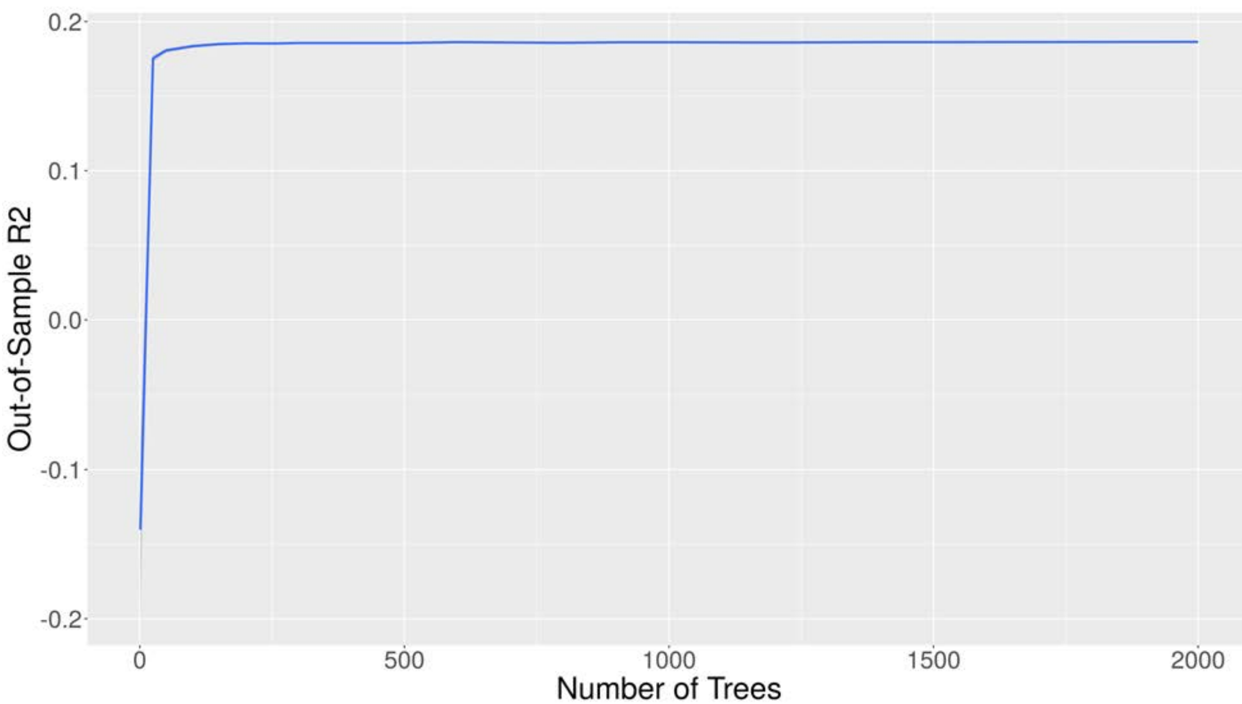
The standard algorithm allows for the specification of the probability of a predictor being chosen at each step. We take advantage of that and implement a two step procedure. First, we run a standard random forest regression, where every variable has the same probability of being chosen, and obtain the 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



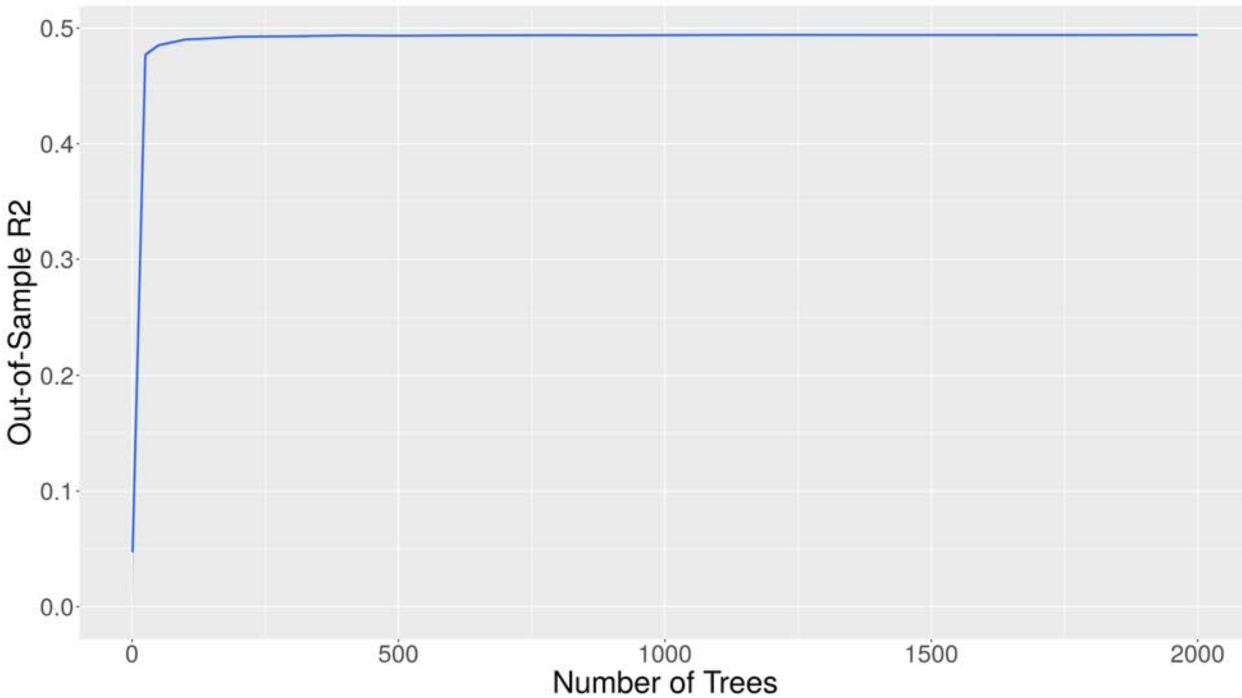
Notes: This figure plots the relation between the number of decision trees used in the random forest for training up to 1986 January and the out-of-sample R^2 for the two-quarters-ahead earnings forecasts in 1986 February. The out-of-sample R^2 is defined as 1 minus the mean squared error implied by 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.

A5. 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-ahead (BE_A2) earnings forecasts. "Average BE" denotes the average of these conditional biases at multiple horizons. "BE Score" denotes the arithmetic average of the percentile rankings on each of the five conditional biases at different forecast horizons. We also report the summary statistics of control variables including 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	1268964	0.0167	0.0983	-0.0268	0.0007	0.0042	0.0137	0.2274
BE Score	1268964	51.1431	23.2862	7.0000	33.2000	48.2000	68.6667	98.6667
BE_Q1	1000603	0.0048	0.0384	-0.0076	-0.0003	0.0005	0.0027	0.0775
BE_Q2	1104296	0.0062	0.0392	-0.0082	0.0002	0.0013	0.0045	0.0851
BE_Q3	1015581	0.0068	0.0466	-0.0174	0.0003	0.0017	0.0056	0.0952
BE_A1	1193713	0.0180	0.1271	-0.0206	0.0003	0.0033	0.0126	0.2625
BE_A2	1081244	0.0339	0.2036	-0.1262	0.0010	0.0118	0.0374	0.4437
LNsize	1268633	13.1098	1.8849	9.3120	11.7508	12.9775	14.3328	17.9642
LNbeme	1153148	-0.7601	0.8561	-3.2706	-1.2255	-0.6680	-0.2000	1.0404
Ret12_7	1207915	0.0817	0.4535	-0.6842	-0.1329	0.0420	0.2228	1.5358
Ret1	1268389	0.0099	0.1574	-0.3819	-0.0625	0.0049	0.0731	0.4828
IA	1174640	0.3021	1.0069	-0.4184	0.0015	0.0893	0.2538	4.3283
IVOL	1268571	0.0247	0.0197	0.0049	0.0125	0.0195	0.0308	0.0954
RetVol	1268016	0.0297	0.0220	0.0067	0.0160	0.0240	0.0366	0.1094
Turnover	1266778	1.5416	12.7502	0.0748	0.4874	0.9836	1.8738	8.3893

A6. Fama-MacBeth regressions with conditional bias in each forecast horizon

Table A3 reports the Fama-MacBeth of monthly stock returns on conditional bias at 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 the two-quarters, three-quarters, and two-years ahead forecast bias negatively predict stock returns; the predictability 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 more optimistic biases earn lower future returns.

Table A5 shows that the return-predictability results from the cross-sectional regressions and portfolio sorts also hold in the time series regressions 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 1986 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)
Bias	-0.029	0.045	-0.184	-0.219	-0.237	-0.278	-0.012	0.005	-0.030	-0.041
t-stat	-0.59	1.11	-3.42	-4.58	-5.12	-6.76	-0.83	0.40	-5.39	-7.80
LNsize		-0.041		-0.080		-0.093		-0.065		-0.118
t-stat		-1.15		-2.25		-2.62		-1.83		-3.29
LNbeme		0.090		0.130		0.130		0.060		0.099
t-stat		1.53		2.24		2.24		1.05		1.77
Ret1		-3.270		-2.559		-2.630		-2.976		-3.012
t-stat		-7.35		-6.12		-6.08		-7.08		-6.98
Ret12_7		0.509		0.357		0.329		0.476		0.338
t-stat		3.31		2.30		2.06		3.08		2.14
IA		-0.003		-0.003		-0.003		-0.003		-0.003
t-stat		-5.24		-5.03		-4.62		-5.46		-4.86
IVOL		-0.298		-0.266		-0.274		-0.197		-0.239
t-stat		-2.61		-2.30		-2.33		-1.81		-2.14
Retvol		0.184		0.186		0.242		0.097		0.214
t-stat		1.54		1.55		1.97		0.84		1.80
Turnover		-0.086		-0.065		-0.087		-0.063		-0.060
t-stat		-1.95		-1.46		-1.98		-1.39		-1.30
Intercept	0.955	1.806	1.048	2.348	1.114	2.483	0.981	2.102	1.166	2.799
t-stat	3.34	3.34	3.70	4.38	3.91	4.64	3.52	4.00	4.23	5.27
R^2 (%)	0.818	5.766	0.846	6.027	0.794	6.193	0.792	5.693	0.841	6.106

Table A4: Portfolios sorted on conditional bias

This table reports the time series average of 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 1986 to 2019.

Quintile	1	2	3	4	5	5-1
Panel A: One-quarter-ahead BE						
Mean	1.04	0.88	0.92	0.97	0.83	-0.21
t-stat	4.51	4.01	3.81	3.52	2.14	-0.79
CAPM Beta	1.00	0.99	1.06	1.14	1.44	0.44
Panel B: Two-quarters-ahead BE						
Mean	1.14	0.91	0.92	0.61	0.21	-0.93
t-stat	5.29	4.26	3.92	2.19	0.52	-3.24
CAPM Beta	0.96	0.96	1.03	1.18	1.49	0.53
Panel C: Three-quarters-ahead BE						
Mean	1.33	1.03	0.80	0.50	-0.03	-1.36
t-stat	6.15	4.86	3.45	1.77	-0.09	-5.21
CAPM Beta	0.95	0.94	1.02	1.20	1.45	0.50
Panel D: One-year-ahead BE						
Mean	0.97	0.91	0.97	0.97	1.01	0.04
t-stat	4.61	4.18	3.96	3.47	2.73	0.14
CAPM Beta	0.94	0.98	1.08	1.18	1.40	0.46
Panel E: Two-years-ahead BE						
Mean	1.39	1.01	0.85	0.64	-0.32	-1.71
t-stat	6.65	4.85	3.59	2.28	-0.88	-6.65
CAPM Beta	0.91	0.93	1.06	1.19	1.42	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 1986 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	-0.52	-2.03	-0.61	-2.74	-0.11	-0.50
Mkt_RF	0.44	5.75	0.40	5.70	0.22	3.55
SMB			0.78	8.56	0.57	5.61
HML			0.51	4.09	0.95	7.10
RMW					-0.78	-5.43
CMA					-0.66	-2.89
Panel B: Two-quarters-ahead BE						
Intercept	-1.30	-4.92	-1.43	-6.08	-1.00	-4.02
Mkt_RF	0.53	6.38	0.51	6.70	0.36	4.94
SMB			0.76	7.48	0.56	4.86
HML			0.64	4.89	0.99	6.98
RMW					-0.74	-4.35
CMA					-0.46	-1.78
Panel C: Three-quarters-ahead BE						
Intercept	-1.71	-7.18	-1.78	-8.34	-1.36	-5.98
Mkt_RF	0.50	7.26	0.46	7.21	0.32	5.06
SMB			0.67	6.78	0.46	4.23
HML			0.42	3.50	0.76	6.53
RMW					-0.74	-4.70
CMA					-0.43	-2.09
Panel D: One-year-ahead BE						
Intercept	-0.28	-1.23	-0.35	-1.64	0.05	0.20
Mkt_RF	0.46	6.80	0.41	6.32	0.27	3.92
SMB			0.70	6.84	0.52	4.79
HML			0.38	3.35	0.72	5.24
RMW					-0.64	-3.97
CMA					-0.49	-1.83
Panel E: Two-years-ahead BE						
Intercept	-2.06	-8.85	-2.17	-10.29	-1.86	-7.93
Mkt_RF	0.51	7.57	0.49	8.07	0.39	6.00
SMB			0.60	5.62	0.47	4.19
HML			0.51	4.71	0.78	5.99
RMW					-0.49	-3.36
CMA					-0.40	-1.65

A7. 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 month, to "predict" stock returns, though realized earnings are not available at time t. Table A6 reports the regressions with average realized biases, and Table A7 and Table A8 report the mean return and alpha on the long-short portfolio strategy based on the realized bias. Overall, we find very consistent results, stocks with more optimistic 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' 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 1986 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)
Bias	-0.108	-0.132	-0.098	-0.110
t-stat	-14.92	-17.32	-38.37	-46.62
LNsize		-0.109		-0.264
t-stat		-2.99		-7.15
LNbeme		0.162		0.107
t-stat		2.80		1.89
Ret1		-3.215		-5.627
t-stat		-7.72		-13.11
Ret12_7		0.289		-0.196
t-stat		1.87		-1.31
IA		-0.003		-0.002
t-stat		-5.33		-3.47
IVOL		-0.177		-0.138
t-stat		-1.61		-1.26
Retvol		0.156		0.133
t-stat		1.35		1.16
Turnover		-0.056		-0.001
t-stat		-1.24		-0.03
Intercept	1.137	2.705	5.777	9.747
t-stat	3.98	5.05	20.73	17.48
R^2 (%)	0.988	6.133	3.368	8.770

Table A7: Portfolios sorted on realized bias

This table reports the time series average of returns (in percent) on value-weighted portfolios formed on 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 1986 to 2019.

Quintile	1	2	3	4	5	1-5
Panel A: Average BE						
Mean	3.21	1.59	0.24	-0.69	-1.73	-4.94
t-stat	13.01	7.53	1.13	-2.58	-5.04	-23.06
CAPM Beta	1.03	0.94	0.96	1.15	1.35	0.32
Panel B: BE Score						
Mean	3.18	1.69	0.41	-0.73	-2.26	-5.44
t-stat	13.06	7.94	1.93	-2.94	-7.01	-26.36
CAPM Beta	1.03	0.94	0.95	1.08	1.29	0.26

Table A8: 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 1986 to 2019. The t-statistics are adjusted by the White's heteroscedasticity robust standard errors.

	CAPM		FF3		FF5	
	$Coeffi$	t-stat	$Coeffi$	t-stat	$Coeffi$	t-stat
Panel A: Average BE						
Intercept	-5.17	-24.66	-5.21	-25.40	-4.88	-21.93
Mkt_RF	0.32	5.82	0.30	5.23	0.18	3.13
SMB			0.40	4.85	0.27	3.16
HML			0.23	2.13	0.53	4.68
RMW					-0.47	-3.86
CMA					-0.49	-2.27
Panel B: BE Score						
Intercept	-5.62	-27.44	-5.64	-28.41	-5.42	-25.18
Mkt_RF	0.26	4.58	0.22	3.78	0.14	2.43
SMB			0.44	5.32	0.36	3.90
HML			0.16	1.49	0.36	3.23
RMW					-0.30	-2.76
CMA					-0.32	-1.69

A8. Cross-sectional return predictability: other robustness checks

In this section, we check the robustness of Fama-MacBeth regression results in Table 3 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 A9 and A10 report the two robustness checks results, respectively. Overall, we find robust return predictability of conditional biases.

Table A9: Fama-Macbeth regressions: omitting stocks with price lower than \$5

Notes: This table reports the Fama-MacBeth cross-sectional regressions of monthly stocks' 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 1986 to 2019. We omit stocks whose closing prices in the previous month are smaller than \$5.

	Panel A: Average BE		Panel B: BE Score	
	(1)	(2)	(1)	(2)
Bias	-0.383	-0.451	-0.027	-0.033
t-stat	-10.94	-14.46	-8.86	-13.57
LNsize		-0.112		-0.180
t-stat		-3.67		-6.00
LNbeme		0.153		0.193
t-stat		2.85		3.60
Ret1		-2.087		-2.163
t-stat		-5.29		-5.51
Ret12_7		0.416		0.328
t-stat		2.89		2.38
IA		-0.002		-0.002
t-stat		-4.35		-4.42
IVOL		-0.254		-0.228
t-stat		-2.32		-2.09
Retvol		0.159		0.153
t-stat		1.33		1.30
Turnover		-0.042		-0.025
t-stat		-0.99		-0.61
Intercept	1.197	3.004	2.208	5.099
t-stat	4.69	6.53	9.87	11.52
R^2 (%)	0.794	6.248	1.151	6.320

Table A10: Fama-Macbeth regressions: scaling conditional biases by total assets per share

Notes: This table reports the Fama-MacBeth cross-sectional regressions of monthly stocks'returns on the conditional bias, which is defined as the difference between analysts' earnings forecasts and machine learning forecasts, scaled by the total asset (per share) 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 1986 to 2019.

	Panel A: Average BE		Panel B: BE Score	
	(1)	(2)	(1)	(2)
Bias	-0.062	-0.079	-0.019	-0.025
t-stat	-4.15	-8.24	-4.36	-10.66
LNsize		-0.093		-0.178
t-stat		-2.63		-5.32
LNbeme		0.009		-0.050
t-stat		0.15		-0.90
Ret1		-2.860		-2.964
t-stat		-6.82		-7.13
Ret12_7		0.509		0.412
t-stat		3.25		2.68
IA		-0.003		-0.003
t-stat		-5.52		-5.79
IVOL		-0.219		-0.209
t-stat		-1.99		-1.91
Retvol		0.136		0.170
t-stat		1.18		1.49
Turnover		-0.056		-0.035
t-stat		-1.25		-0.79
Intercept	1.062	2.455	1.930	4.565
t-stat	3.85	4.67	8.73	9.29
R^2 (%)	0.539	5.448	1.473	5.697

A9. Net stock issuances: robustness check

We check the robustness of results in Table 9 by matching average of conditional bias from the past 24-12 months to net stock issuances of the fiscal year ending in t. Table A11 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 A11: 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 1986 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	0.007	0.010	0.019	0.021	0.071	0.064
t-stat	1.26	1.56	2.33	3.18	4.96	4.67
BE score	0.005	0.013	0.015	0.033	0.065	0.060
t-stat	0.78	1.78	2.81	4.90	5.14	5.35

Panel B: Fama-MacBeth regressions				
	A: Average BE		B: BE Score	
	(1)	(2)	(1)	(2)
Bias	0.514	0.398	0.091	0.063
t-stat	4.11	3.96	8.10	4.60
LNsize		-0.493		-0.307
t-stat		-2.78		-1.42
LNbeme		-1.689		-1.803
t-stat		-5.32		-5.68
EBITDA		-12.163		-12.093
t-stat		-5.18		-5.04
Intercept	3.192	9.585	-0.656	4.485
t-stat	8.15	3.36	-1.01	1.17
R^2 (%)	1.963	7.674	1.201	7.187

A10. Anomalies

In this study, we follow Hou et al. (2015) as close as possible to define anomaly variables. Table A12 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 A12: 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
ROA	Return on assets, Balakrishnan et al. (2010)	07/1972–12/2019	0.57

continued from previous page

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

A11. Out-of-sample linear return predictability in So (2013)

Table A13: Return predictability: linear forecasts pre- and post- 2000

Notes: This table reports the Fama-MacBeth cross-sectional regressions of monthly stocks' returns on the linear conditional bias computed as in So (2013). The sample period is from 1986 to 2019. Pre-2000 covers the period from 1986 to 2000. Post-2000 covers the period 2001-2019.

	All	Pre-2000	Post-2000
Intercept	0.012	0.013	0.010
t-stat	4.36	3.54	2.72
Linear conditional bias	−0.015	−0.030	−0.002
t-stat	−3.79	−3.92	−0.64
R^2	0.004	0.005	0.002

A12. Return predictability: linear forecasts versus random forest versus predicted forecast errors

In this section, we test the return predictability of analysts' forecast bias by using OLS-regression-based earnings forecasts as a benchmark. We also test the return predictability of predicted forecast errors as Frankel and Lee (1998).[37]

Table A14 reports the earnings forecasts via linear regressions. In sharp contrast to the Random Forests forecasts, linear forecasts have larger forecast errors, measured as the mean of squared difference between linear forecasts and realized earnings, than analysts'.

Table A15 reports the return predictability of forecast bias in pre- and post-2000 periods. Panel A and B look at the average forecast bias using random forest forecast and Linear forecast as benchmarks, respectively. We find that the return predictability of random forest forecast bias remains robust when splitting the sample into pre-and post-2000 periods. However, return predictability of linear forecast bias disappears after 2000. Panel C looks at the return predictability of predicted forecast errors in Frankel and Lee (1998). We also find that this return predictability does not work out-of-sample and disappears after 2000.

[37] We follow Frankel and Lee (1998) to predict forecast errors of analysts using the same variables including sale growth, book-to-price ratio, long-term earnings growth forecasts, and an optimism measure. Analysts' forecast errors are defined as the difference between analysts' earnings forecasts and realized earnings scaled by the closing stock price from most recent month. We apply these predicted errors to forecast monthly stock returns.

A13. The meaningful gains of machine learning forecasts

To test which firms that gains of ML forecasts relative to linear and/or analysts'forecasts are more prominent, we consider the following regression of the ratio of the squared errors of linear forecasts/analysts forecasts to the squared errors of random forests,

$$\frac{(LF/AF_{i,t}^{t+1} - AE_{i,t+1})^2}{(RF_{i,t}^{t+1} - AE_{i,t+1})^2} = \alpha_t + \beta_i \sum_{j=1}^{4} Characteristics_{j,i,t} + \epsilon_{i,t}, \tag{A12}$$

where $LF_{i,t}^{t+1}$, $AF_{i,t}^{t+1}$, $RF_{i,t}^{t+1}$ represent linear forecasts, analysts' forecasts, and random forest forecasts at time t for firm i's earnings at fiscal end period $t+1$, respectively. $AE_{i,t+1}$ represents firm i's realized earnings at fiscal end period $t+1$. $Characteristics_{i,t}$ are the log of firm size (LN size), the log of book-to-market ratio (LNbeme), the idiosyncratic volatility (IVOL), and the gross profitability (GP) for firm i at time t. If, for example, the gains of random forest forecasts relative to linear and analysts forecasts are from forecasts for growing firms, then we should expect a negative coefficient on the book-to-market ratio.

Table A16 reports the results. We find that for firms with smaller size, lower book-to-market ratio (growing firms), higher IVOL, and lower profitability, the linear forecasts and analysts' forecasts have larger forecast errors than random forests.

Table A14: Earnings forecasts via linear regression

Notes: This table presents the time series average of OLS regression-based earnings per share forecasts (LF), analysts' earning forecasts (AF), and actual realized earnings (AE) —the difference as well as the squared difference between them. N denotes the number of the sample stocks. We report the Newey-West t-statistics of the 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 when reporting the Newey-West t-statistics. The sample period is 1986 January to 2019 December.

	LF	AF	AE	(LF-AE)	(AF-AE)	$(LF-AE)^2$	$(AF-AE)^2$	N
One-quarter-ahead	0.290	0.319	0.291	-0.001	0.028	0.081	0.081	1,022,661
t-stat				-0.26	6.59			
Two-quarters-ahead	0.316	0.376	0.323	-0.008	0.053	0.124	0.102	1,110,689
t-stat				-0.71	10.31			
One-year-ahead	1.150	1.320	1.167	-0.017	0.154	0.757	0.686	1,260,060
t-stat				-0.90	6.24			
Two-years-ahead	1.293	1.771	1.387	-0.094	0.384	2.573	2.009	1,097,098
t-stat				-1.22	8.33			

Table A15: Return predictability: linear forecasts versus random forest versus predicted forecast error

Notes: This table reports the Fama-MacBeth cross-sectional regressions of monthly stocks' returns on the conditional bias. Panel A and B look at the average of forecast bias at different horizons using random forest forecast and linear forecast as benchmarks, respectively. Panel C looks at the return predictability of predicted forecast errors as in Frankel and Lee (1998). The first and third rows of each panel report the regression results with and without control variables, respectively. The sample period is 1986 to 2019.

Sample Period	Full Sample	Pre-2000	Post-2000
Panel A: Random Forest			
FM without control	-0.054	-0.075	-0.037
t-stat	-3.94	-3.78	-1.99
FM with control	-0.064	-0.086	-0.047
t-stat	-5.08	-5.09	-2.58
Panel B: Linear Forecast			
FM without control	-0.029	-0.044	-0.017
t-stat	-2.39	-2.39	-1.05
FM with control	-0.021	-0.034	-0.011
t-stat	-2.05	-2.33	-0.77
Panel C: Predicted Forecast Error			
FM without control	-0.040	-0.110	0.015
t-stat	-0.75	-2.34	0.17
FM with control	-0.058	-080	-0.40
t-stat	-1.49	-2.35	-0.62

Table A16: Fama-Macbeth regressions

Notes: This table reports the Fama-MacBeth cross-sectional regressions of the ratio of the squared errors of linear forecasts/analysts forecasts to the squared errors of random forests. The regressions are:

$$\frac{(LF/AF_{i,t}^{t+1} - AE_{i,t+1})^2}{(RF_{i,t}^{t+1} - AE_{i,t+1})^2} = \alpha_t + \beta_i \sum_{j=1}^{4} Characteristics_{j,i,t} + \epsilon_{i,t},$$

where $LF_{i,t}^{t+1}$, $AF_{i,t}^{t+1}$, $RF_{i,t}^{t+1}$ represent linear forecasts, analysts' forecasts, and random forest forecasts at time t for firm i's earnings at fiscal end period $t+1$, respectively. $AE_{i,t+1}$ represents firm i's realized earnings at fiscal end period $t+1$. $Characteristics_{i,t}$ are the log of firm size (LNsize), the log of book-to-market ratio (LNbeme), the idiosyncratic volatility (IVOL), and the gross profitability (GP) for firm i at time t. Panel A and B look at the linear forecasts and analysts' forecasts, respectively The sample period is 1986 to 2019.

	Panel A: Linear Regression VS Random Forest				Panel B: Analysts VS Random Forest			
Horizons	One-quarter	Two-quarters	One-year	Two-years	One-quarter	Two-quarters	One-year	Two-years
Intercept	24.55	434.16	42.86	98.69	4.93	12.12	7.79	28.95
t-stat	8.09	1.33	7.38	8.17	11.72	23.36	24.80	33.52
LNsize	-0.43	-16.94	-0.81	-3.68	-0.02	-0.41	-0.22	-1.05
t-stat	-2.96	-1.25	-2.68	-7.13	-0.68	-13.39	-11.28	-18.32
LNbeme	-4.39	-11.49	-3.77	-9.77	-0.21	-0.30	-0.14	-1.85
t-stat	-8.77	-3.88	-10.85	-7.99	-5.61	-4.86	-4.95	-14.53
IVOL	-0.75	-5.81	-1.25	-0.71	0.13	0.21	0.13	0.22
t-stat	-5.89	-1.42	-6.75	-2.56	5.68	6.31	7.28	3.88
GP	-0.01	-0.24	-4.96	-15.10	-0.17	-0.28	-0.20	-1.41
t-stat	-1.74	-1.24	-4.82	-8.06	-1.71	-1.88	-2.52	-4.39
R^2 (%)	0.37	0.41	0.33	0.54	0.38	0.28	0.19	0.30

A14. Downward revisions in analysts' earnings forecasts

To explain the intuition regarding return predictability, we note analysts revise their earnings forecasts every month. As the announcement dates approach, analysts should process new information and update their estimates to make better forecasts. Table A17 demonstrates that analysts revise their earnings forecasts.

[Insert Table A17 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-year-ahead forecast, the average forecast error decreases from 0.215, when analysts make the first forecast, to 0.081, when analysts make their last forecast for that fiscal year. This last forecast is usually made about one month after the fiscal year has ended, though precedes the earnings announcement date for that fiscal year. The mean squared error declines from 1.197 (for the one-year-head forecast) to 0.365 for the last forecast the analysts make.

A downward revision also occurs in the one-quarter-ahead, the two-quarters-ahead, the three-quarters-ahead, and the two-years-ahead forecasts. To the extent that investors follow analysts' forecasts and analysts make optimistic expectations, 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 the earnings realization.

Table A17: 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. The sample period is 1986 January to 2019 December.

Panel A: One-quarter-ahead												
Month-ahead	1	0	-1									
N	356770	406433	185165									
FE	0.027	0.021	0.034									
Sqr_FE	0.078	0.070	0.085									
Panel B: Two-quarters-ahead												
Month-ahead	4	3	2									
N	322922	363517	323760									
FE	0.055	0.050	0.050									
Sqr_FE	0.107	0.100	0.097									
Panel C: Three-quarters-ahead												
Month-ahead	7	6	5									
N	290296	330900	296523									
FE	0.071	0.069	0.069									
Sqr_FE	0.133	0.126	0.121									
Panel D: One-year-ahead												
Month-ahead	10	9	8	7	6	5	4	3	2	1	0	-1
N	62905	91633	104809	106804	108226	109359	110562	111547	112710	113903	115354	80921
FE	0.215	0.223	0.218	0.207	0.192	0.171	0.151	0.133	0.111	0.089	0.073	0.081
Sqr_FE	1.197	1.075	1.017	0.888	0.806	0.730	0.603	0.545	0.490	0.388	0.341	0.365
Panel E: Two-years ahead												
Month-ahead	22	21	20	19	18	17	16	15	14	13	12	11
N	46977	70013	81557	85784	88049	89684	91417	92341	93389	94362	95217	92663
FE	0.434	0.442	0.427	0.425	0.414	0.397	0.382	0.370	0.348	0.316	0.302	0.305
Sqr_FE	2.743	2.475	2.310	2.154	2.066	1.964	1.826	1.754	1.651	1.491	1.423	1.381