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

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

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

[2] *Bayes Business School, City, University of London*

[3] *Warrington College of Business, University of Florida*

First Version: June 2020; This Version: January 2022

Review of Financial Studies, Forthcoming

**Abstract**

We introduce a real-time measure of conditional biases to 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, a bias that 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 forecasts. 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. (*JEL* G12, G14)

Key words: Earnings Forecasts, Machine Learning, Investment Strategies

*We are very grateful for the comments of two anonymous referees and the editor, Stefano Giglio. We thank Refinitive for their guidance in using the I/B/E/S database. We are grateful for helpful comments and suggestions provided by Svetlana Bryzgalova (discussant), Jillian Grennan (discussant), Bryan Kelly (discussant), and seminar participants at BI Norwegian Business School, 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 the European Finance Association Annual Meeting 2021. Emails: julesv@wharton.upenn.edu; xiao.han.3@city.ac.uk; alejandro.lopez-lira@warrington.ufl.edu

# 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 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.

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

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

[3] See Gu et al. (2020) for an excellent overview of this and other 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.

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 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

[4]We are agnostic on the source of the biases for analysts' earnings 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).

[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.

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 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.[7] 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. 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, even when using a variable-selection-bias-free linear model, consistent with previous studies using machine learning (Gu et al. (2020)). In particular, investors using linear forecasts after the 2000s would miss the opportunity to earn at least 0.46% of return per month when using the variable-selection-bias-free linear model and even more when using models that have the forward-looking bias.

Armed with a statistically optimal and unbiased benchmark for firms' earnings expectations and the implied real-time measure for firm-level conditional earnings 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

[7]We discuss these results extensively in Internet Appendix Section A10.

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. 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 earnings forecast 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 earnings forecast 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, Gennaioli, Ma and Shleifer (2020), Coibion and Gorodnichenko (2015), and Bianchi et al. (2022) for recent expositions).[8] 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 the appendix.

## 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

[8]In particular, Bianchi et al. (2022) characterizes the time-varying systematic expectation errors embedded in survey responses using machine-learning techniques. See also Bordalo et al. (2019) and Bordalo, Gennaioli, Porta and Shleifer (2020) who provide evidence of systematic biases in analysts' forecasts of earnings growth.

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 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.[9]

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' earnings forecasts dominate the linear earnings 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 predictor (Breiman 2001). We train the algorithm using rolling windows analogous to a rolling regression forecast. The hyper-parameters are chosen using cross-validation:

[9] 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.

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 Internet Appendix A1. 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 5 provides an illustration. The variable we wish to forecast is the earnings-per-share (eps hereafter) for a cross-section of firms. At the first step, the selected explanatory variable is the 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 6 and 7 show the resulting predictive surface.

**[Insert Figures 5, 6 and 7 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.[10]

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:

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

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.

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.[11] 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

[10]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*.

[11]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.

$X_i$ 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.[12]

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 8 and 9 confirm that this indeed holds in our setup: The performance is increasing in the number of trees but reaches a plateau.[13]

**[Insert Figure 8 and 9 about here]**

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 10 and 11 show that the performance of the model is increasing in model complexity up until a depth of 7.

**[Insert Figure 10 and 11 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 subsample without replacement as the training sample for each decision tree. We then repeat the process for each remaining tree. Figures 12 and 13 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 12 and 13 about here]**

[12] 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.

[13] 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.

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 14.

**[Insert Figure 14 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.[14] 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+ $\tau$ ($\tau$=1 to 3) or year t+$\tau$ ($\tau$=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.[15]

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

[15]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)).

## 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
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[16]

### 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:[17]

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

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' earnings forecasts bias generated by expectation errors or incentive problems

[16]See Internet Appendix Section A2 for details of the variables' definitions and Internet Appendix Section A3 for more information on how we merge these databases.

[17]See Hughes et al. (2008) and So (2013) among others.

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 cumulative adjustment factors from the CRSP monthly stock file to adjust the forecast and the actual EPS on the same share basis.[18]

### 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.[19] 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.

[18]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)).

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

### 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, 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.[20] 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]**

[20] 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 15 and 16 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. 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 Figures 15 and 16 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 earnings forecast bias is statistically different from zero for all horizons. Furthermore, we find that analysts are more biased at longer horizons.

Figure 17 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. For comparison, Figure 18 displays the average realized bias. Both the realized and the conditional bias show similar patterns, albeit with different magnitudes, and both figures show spikes during the internet bubble and the financial crisis.

**[Insert Figures 17 and 18 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' earnings expectations, then we should observe negative return predictability. Stocks with a high conditional earnings forecast bias should earn lower returns than stocks with a lower conditional bias.[21]

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.[22] 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 $R^2$s for both regressions have values around 0.01.

**[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

[21]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.

[22]We require at least two non-missing observations of conditional biases across the multiple horizons to measure the average of conditional biases.

and the bias score remain statistically significant after controlling for those variables. We report the individual conditional bias results in Internet Appendix Section A5: the two-quarters, three-quarters, and two-years ahead earnings forecast biases generate significant negative return predictability.[23] 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 Internet Appendix Section A6.

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.

**[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 conditional 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

[23] 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.

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.[24] 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**]

Moreover, we document that consistent with analysts walking down their earnings forecast (on average), and hence their biases, there is an associated decline in magnitude in the realized returns of the long-short portfolio formed on the conditional earnings forecast bias: the majority of the return is concentrated in the first months and the magnitude decreases quickly afterwards. Figure 19 depicts this result.[25]

[**Insert Figure 19 about here**]

Since the magnitude and significance of the results seem large by usual standards, we conduct a placebo test in Internet Appendix Section A8 to shed further light on these results and place them in context. In particular, we replace the machine learning forecast with the future realized value and then compute the conditional bias. The implied returns of these forward-looking (and thus non-tradable) strategies are many times larger in magnitude than the ones implied by our (tradable) machine-learning forecasts. Finally, we show in Internet Appendix Section A10 that the average long-short return earned by sorting on the ML-implied earnings forecast bias is remarkably stable across time horizons (albeit with increased volatility during the recent financial crisis) in contrast with the marked decline in return predictability from the linear models in the existing literature.[26] The stability is

[24] We report the results of the long-short strategy based on individual conditional bias in Internet Appendix Table A5. All strategies but for the one using the one-year-ahead bias exhibit significant alpha.

[25] We present evidence of downward revision in analysts' earnings forecasts in Internet Appendix Section A7.

[26] We report earnings forecast errors of the linear model in Internet Appendix Section A9 and show that linear forecasts have larger forecast errors than analysts' forecasts and random forest forecasts.

also apparent from Figure 20 which displays the cumulative performance of the return of a value-weighted long-short portfolio, that is short on firms with the highest conditional bias and long on firms with the lowest.

**[Insert Figure 20 about here]**

## 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 earnings forecast 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. (2019)).[27] 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.[28] 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 defined as the stocks in the top (bottom) decile portfolio.

[27] Table A16 in Internet Appendix Section A11 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 reports the monthly average returns (in percent) of the long-short anomaly portfolios.

[28] When measuring the anomaly score, we exclude stocks for which we have fewer than 10 rank scores, which occurs when not all the data inputs on the characteristics are available.

**[Insert Table 7 about here]**

Table 7 Panel A presents the average anomaly score for portfolios sorted independently on the conditional earnings forecast bias and the anomaly score.[29] 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 earnings forecast bias and the anomaly score. The long-short portfolio using the anomaly score earns 1.36% per month with a $t$-statistic of 5.74. While the long-short anomaly strategy in each quintile sort on the conditional bias has a similar anomaly score (around 3.6), 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 earnings expectations.[30] Moreover, we document in Internet Appendix Section A12, that the effect of the conditional bias is not subsumed by the anomaly score, as the results remain similar when using the orthogonal component of our conditional

[29]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.

[30]Internet Appendix Table A17 presents alphas with respect to the Fama-French fivefactor model for the portfolios formed by sorting independently on the conditional bias and the anomaly score. We find that the alpha from the long-short anomaly portfolio is larger and more significant across portfolios with a larger conditional bias. More importantly, the anomaly alpha becomes insignificant for portfolios with the smallest bias.

bias measure relative to the anomaly score.

**[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 earnings forecast 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).[31]

## 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 earnings forecast bias in the past 12 months to the fiscal year ending at time $t$.[32] 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.

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 is-

[31]We document in the Internet Appendix Section A12 that the relationship between the conditional bias and the anomaly score is not present out-of-sample when using the earnings forecast bias implied by linear models.

[32]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 Internet Appendix Table A22.

sue 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 earnings forecast 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).

In Internet Appendix Section A14, we document that the predictability of net stock issuances does not decline significantly in the post-2000 period relative to the pre-2000 period. In contrast, we observe a significant decline in the NSI predictability when using the linear earnings forecast bias as proposed in So (2013). Interestingly, the linear forecast, free of forward-looking bias, can predict both in- and out-of-sample net stock issuances. Internet Appendix Section A14 also reports the average net stock issuances for portfolios sorted independently on conditional bias and anomaly scores. Given the independent sort on the anomaly score, we find that stocks in the anomaly short leg have more net stock issuances than stocks in the long leg. In addition, within 9 out of 10 anomaly deciles, we find a significantly positive difference in NSI between stocks with the largest conditional earnings forecast bias and stocks with the smallest conditional bias.

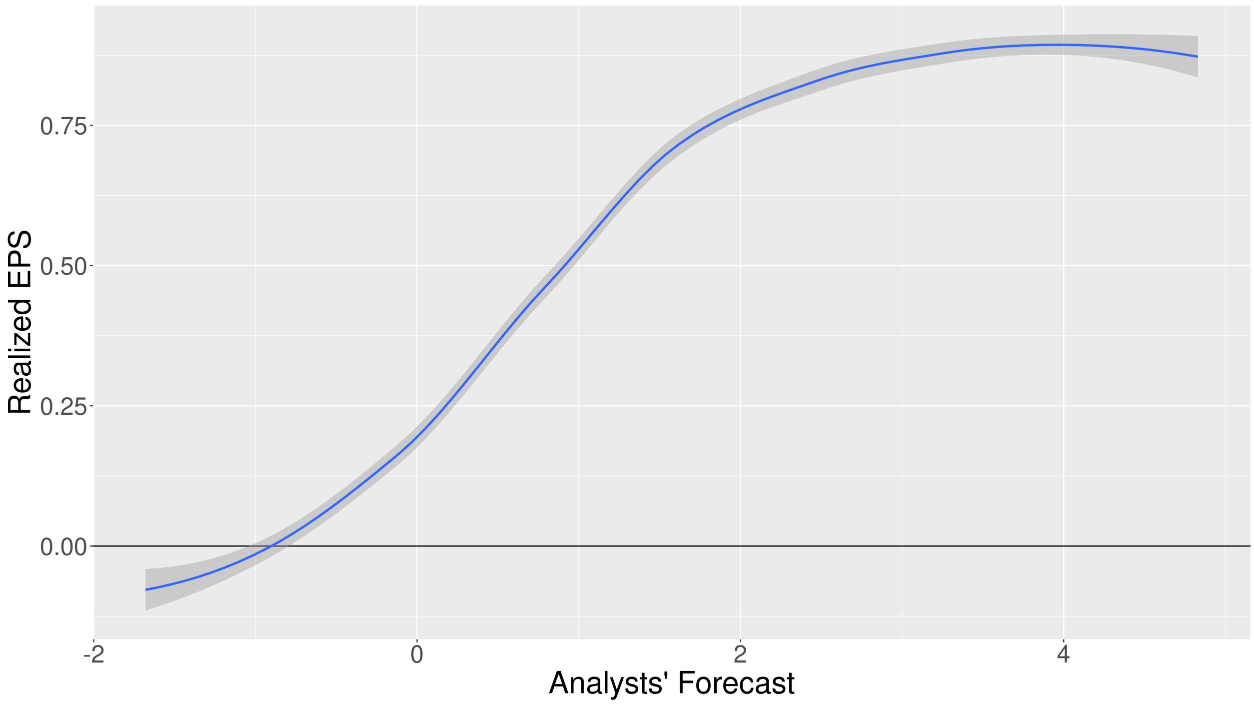
# 6 Conclusion

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 earnings forecast 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.

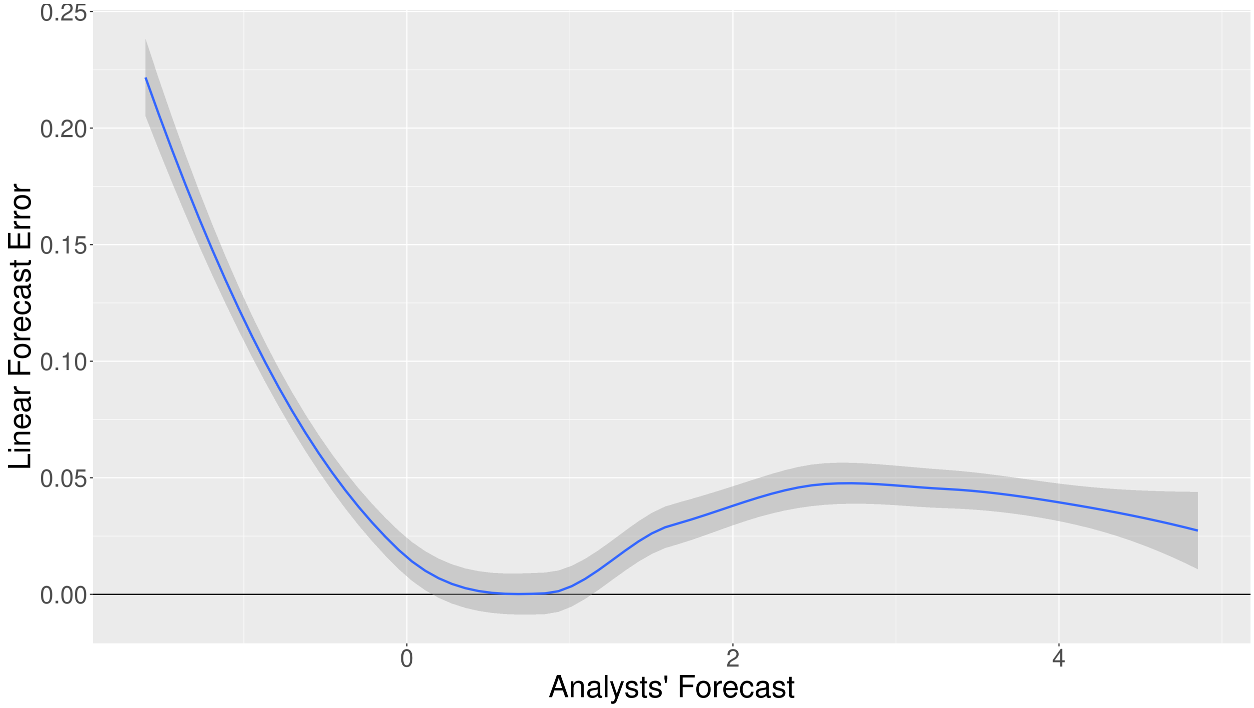
In addition to these asset-pricing results, our findings also have implications for corporate finance. Managers of firms for which the earnings forecast is most upward-biased issue more stocks. This finding indicates that managers are at least partially aware of analyst biases or the associated influence 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



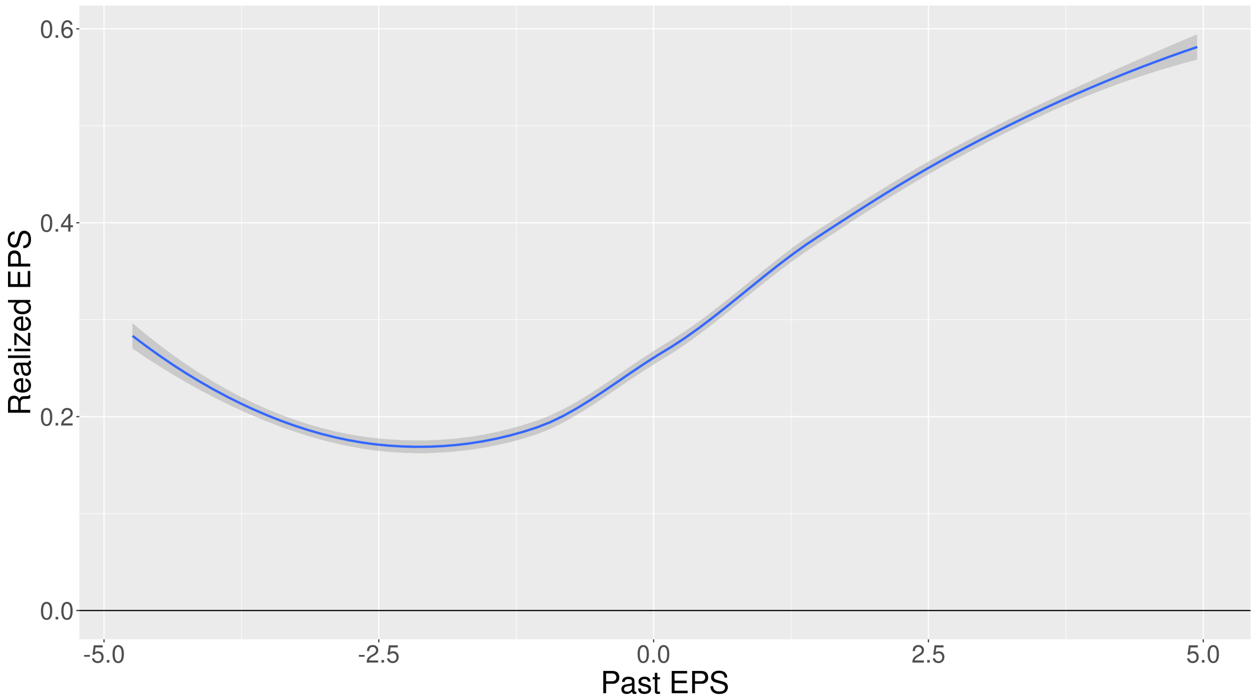

Notes: The figure plots the partial dependence plot of one-quarter-ahead realized EPS on analysts' forecasts. The partial dependence plot is calculated from a random forest regression of EPS on the variables mentioned in Section 3.2. The random forest regression for the figure uses 2000 trees and a minimum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



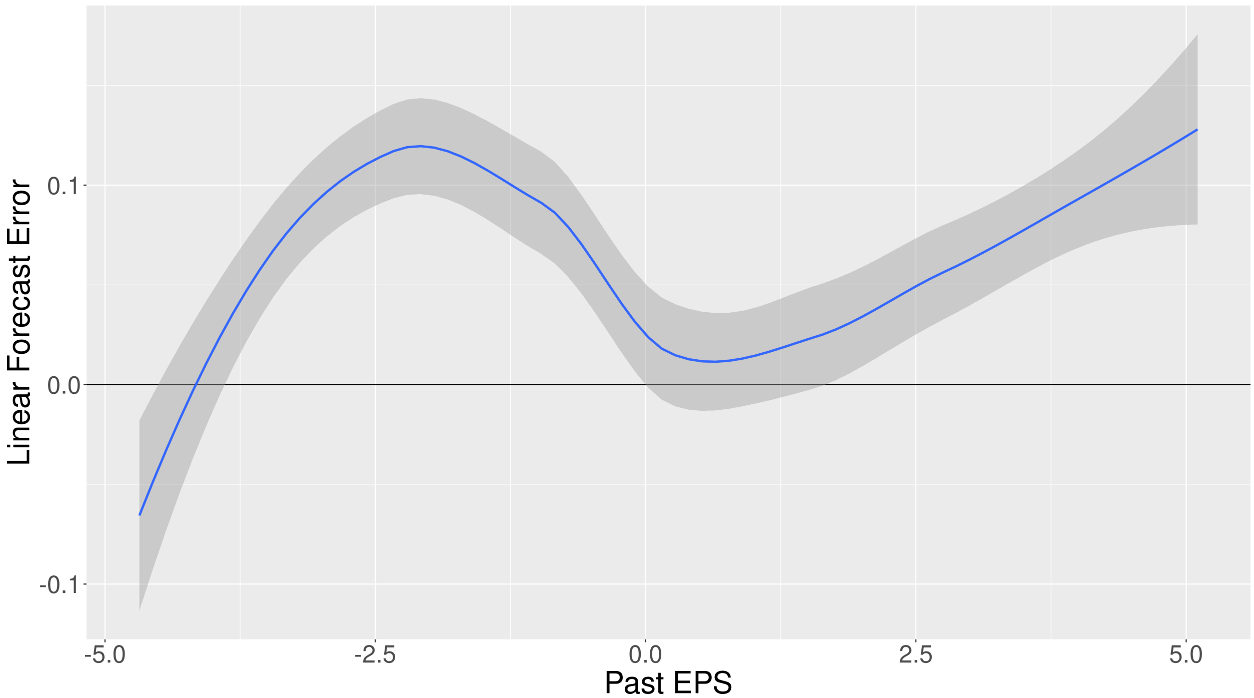

Notes: The figure plots the partial dependence plot of one-quarter-ahead realized linear errors on analysts' forecasts. The linear errors are calculated as the difference between the linear forecast and the realized EPS. The partial dependence plot is calculated from a random forest regression of the 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



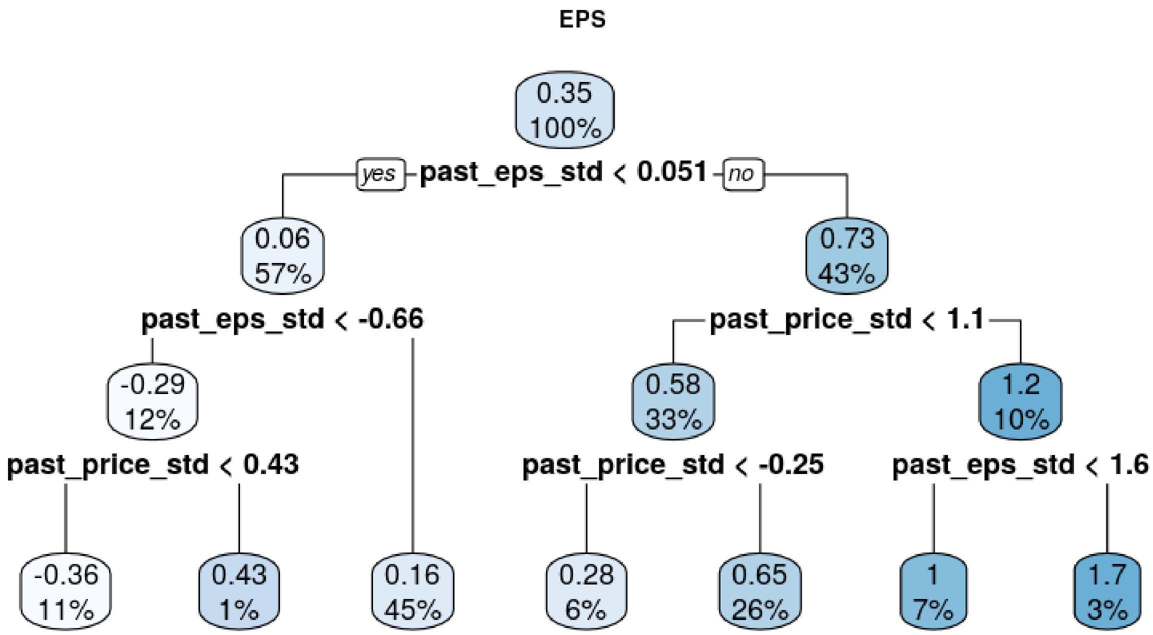

Notes: The figure plots the partial dependence plot of one-quarter-ahead realized EPS on past EPS. The partial dependence plot is calculated from a random forest regression of EPS on the variables mentioned in Section 3.2. The random forest regression for the figure uses 2000 trees and a minimum node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



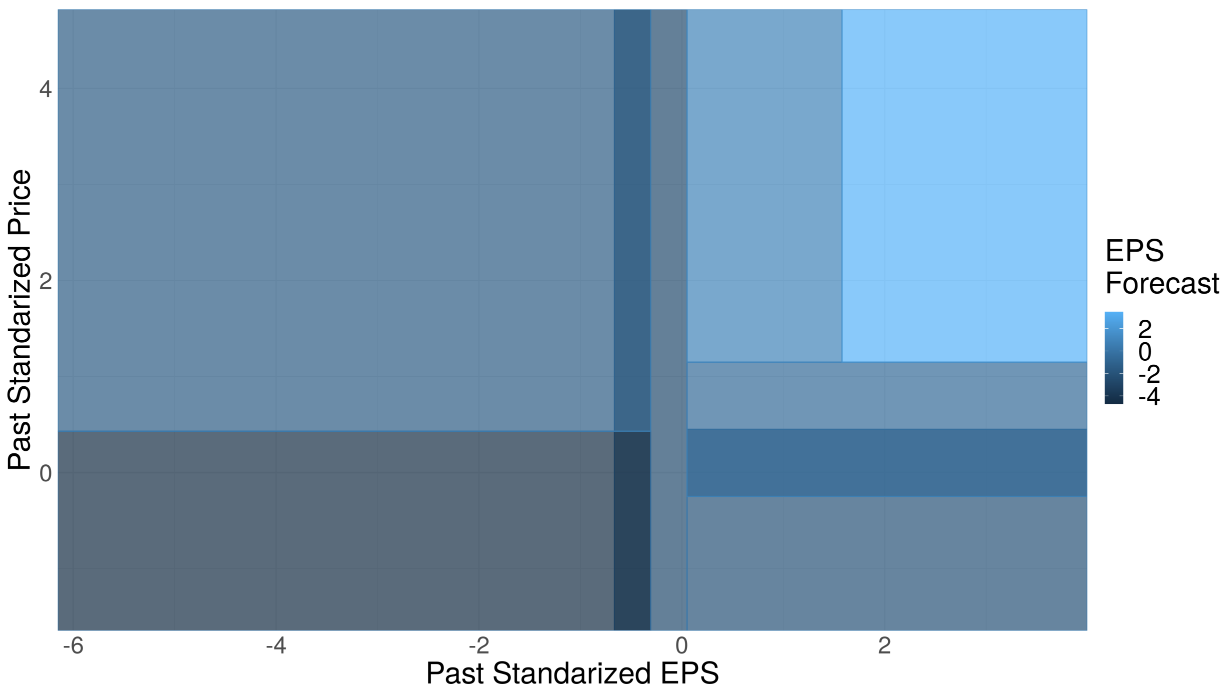

Notes: The figure plots the partial dependence plot of one-quarter-ahead realized linear errors on past EPS. The linear errors are calculated as the difference between the linear forecast and the realized EPS. The partial dependence plot is calculated from a random forest regression of the linear errors 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: Example decision tree



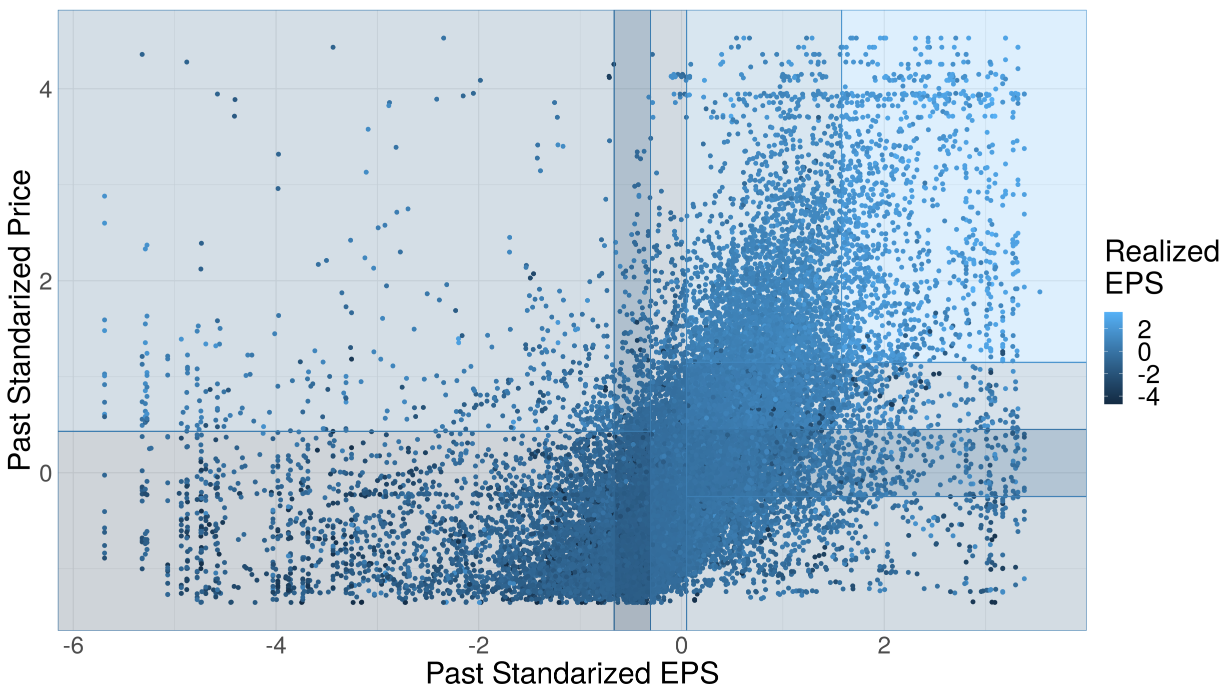

Notes: The figure shows an example decision tree. The variable we wish to forecast is the earnings-per-share (eps hereafter) for a cross-section of firms. At the first step, the selected explanatory variable is the past earnings per share (denoted by past_eps_std), and the threshold (or cutoff) value 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 6: Decision tree predictions



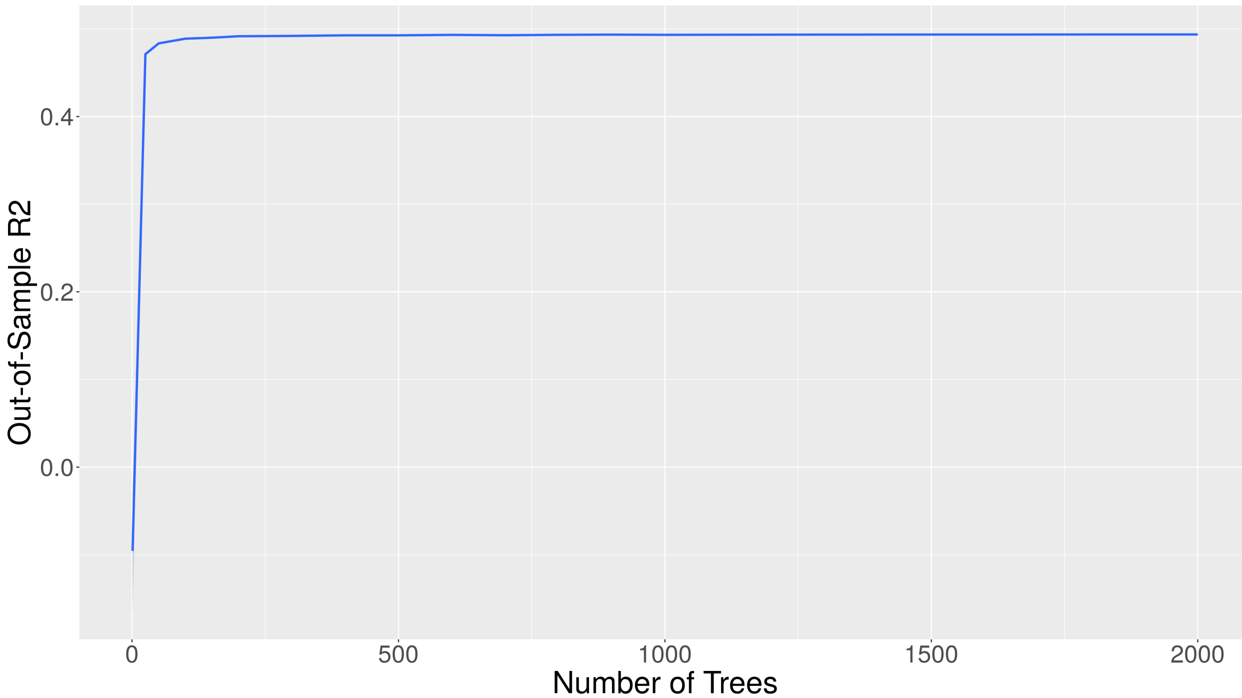

Notes: The figure shows the forecast of the decision tree from Figure 5. The variable we wish to forecast is the earnings-per-share for a cross-section of firms. The prediction is constant within each color box, and corresponds to the historical mean for each sub-space.

Figure 7: Decision tree predictions



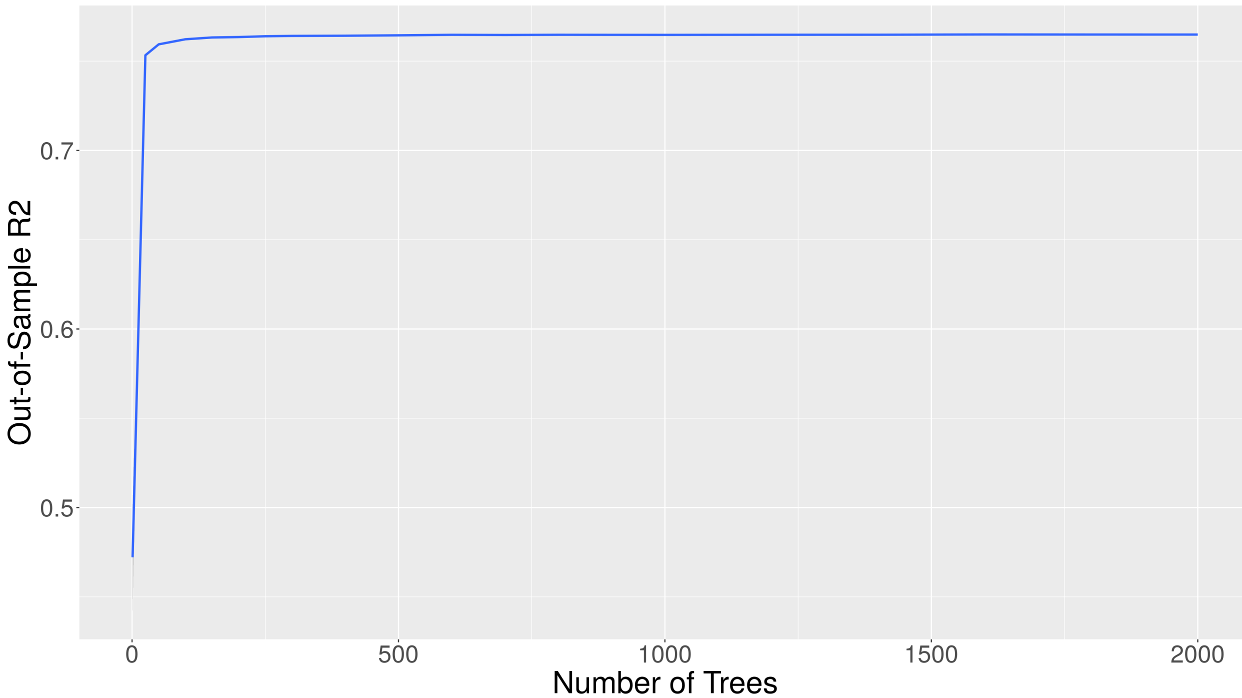

Notes: The figure shows the forecast of the decision tree from Figure 5. The variable we wish to forecast is the earnings-per-share for a cross-section of firms. The prediction is constant within each color box, and corresponds to the historical mean for each sub-space. The realized values are shown with a different color indicating a different value.

Figure 8: Cross-validation results of the number of trees in the one-quarter-ahead forecast



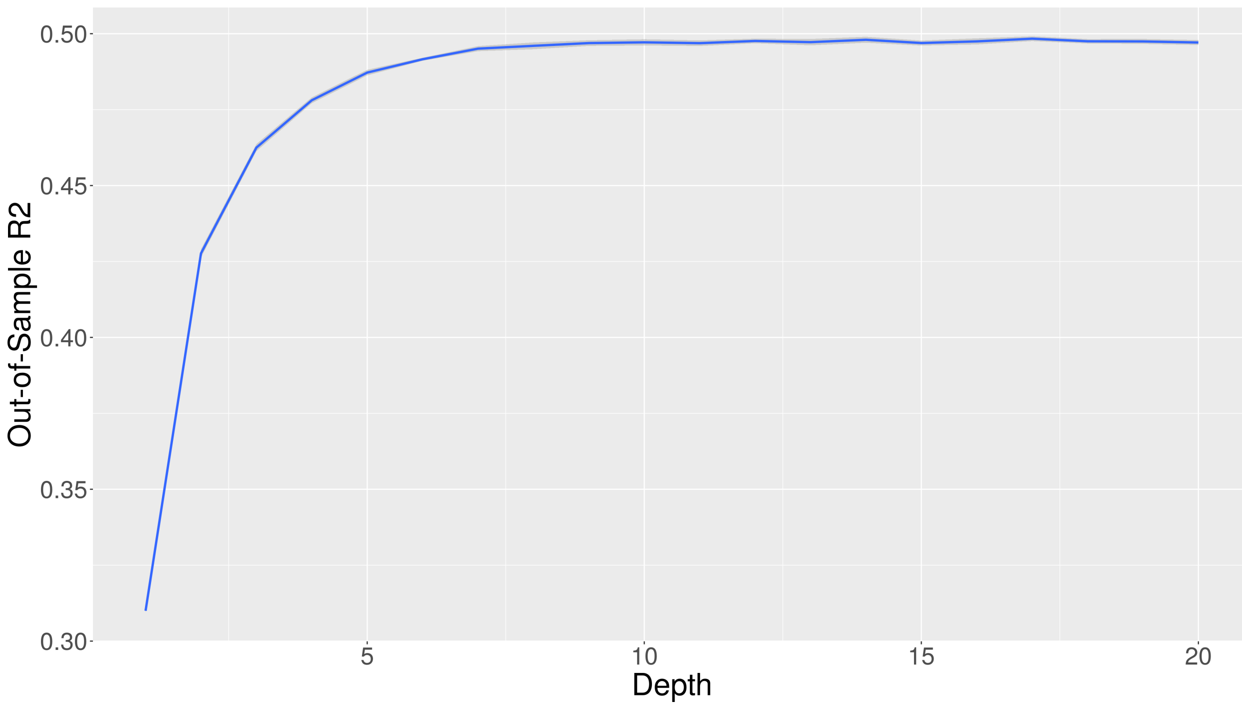

Notes: The figure plots the relation between the number of decision trees used in the random forest for training up to January 1986 and the out-of-sample $R^2$ value for the one-quarter-ahead earnings forecasts made in February 1986 for May 1986. The out-of-sample $R^2$ is defined as 1 minus the mean 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 9: Cross-validation results of the number of trees in the one-year-ahead forecast



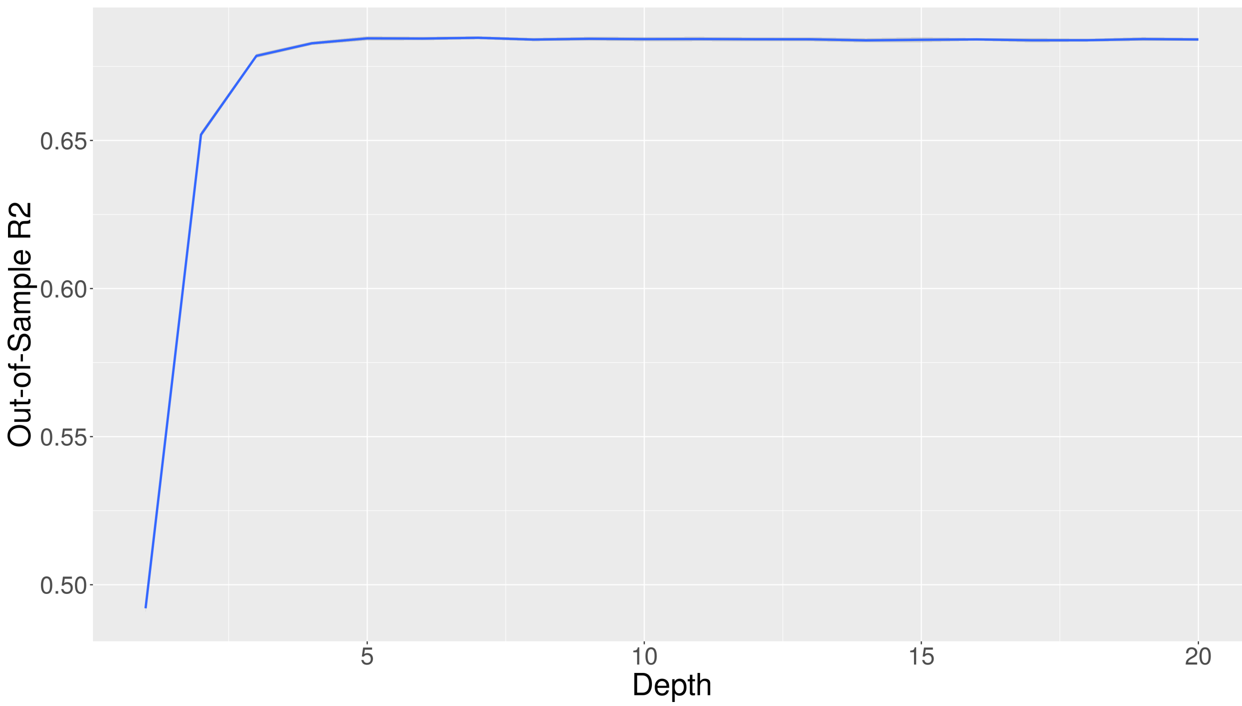

Notes: The figure plots the relation between the number of decision trees used in the random forest for training up to 1986 January and the out-of-sample $R^2$ for the one-year-ahead earnings forecasts in 1986 February. The out-of-sample $R^2$ is defined as 1 minus the mean squared error implied by using 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 10: Cross-validation results of the maximum depth of each tree in the one-quarter-ahead forecast



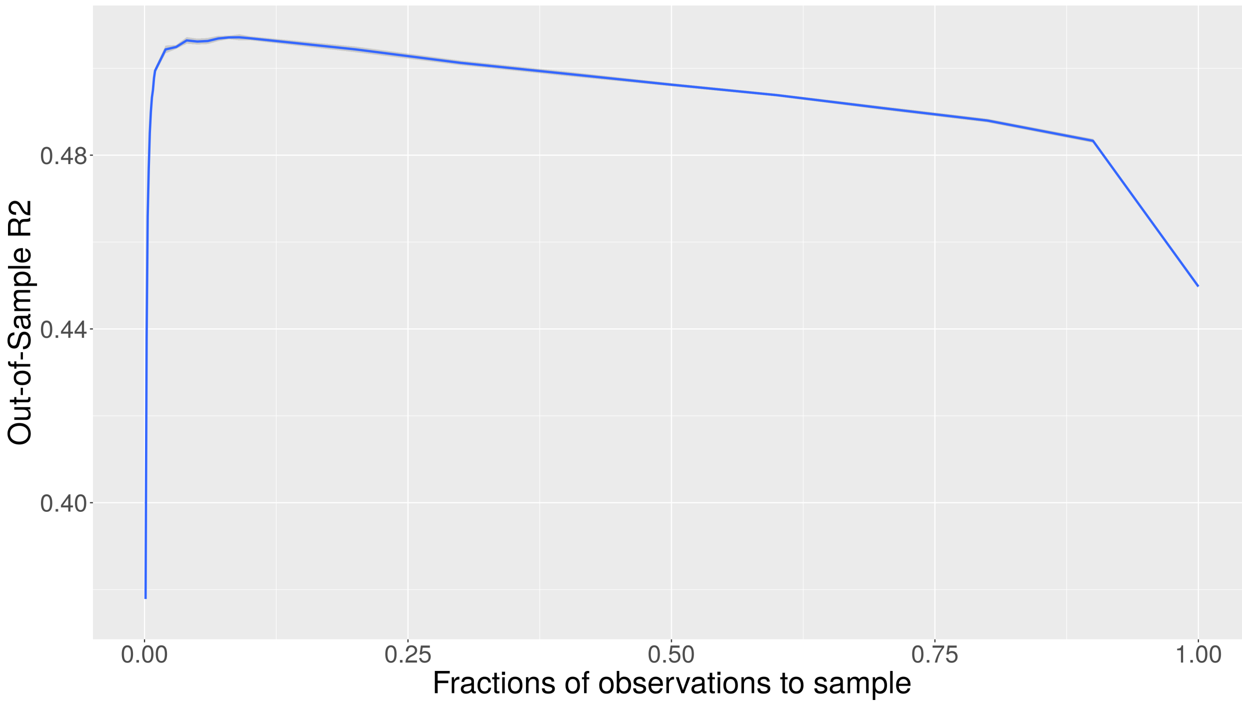

Notes: The figure plots the relation between the depth of of decision trees used in the random forest for training up to 1986 January and the out-of-sample $R^2$ for the one-quarter-ahead earnings forecasts in 1986 February. The out-of-sample $R^2$ is defined as 1 minus the mean squared error implied 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 11: Cross-validation results of the maximum depth of each tree in the one-year-ahead forecast



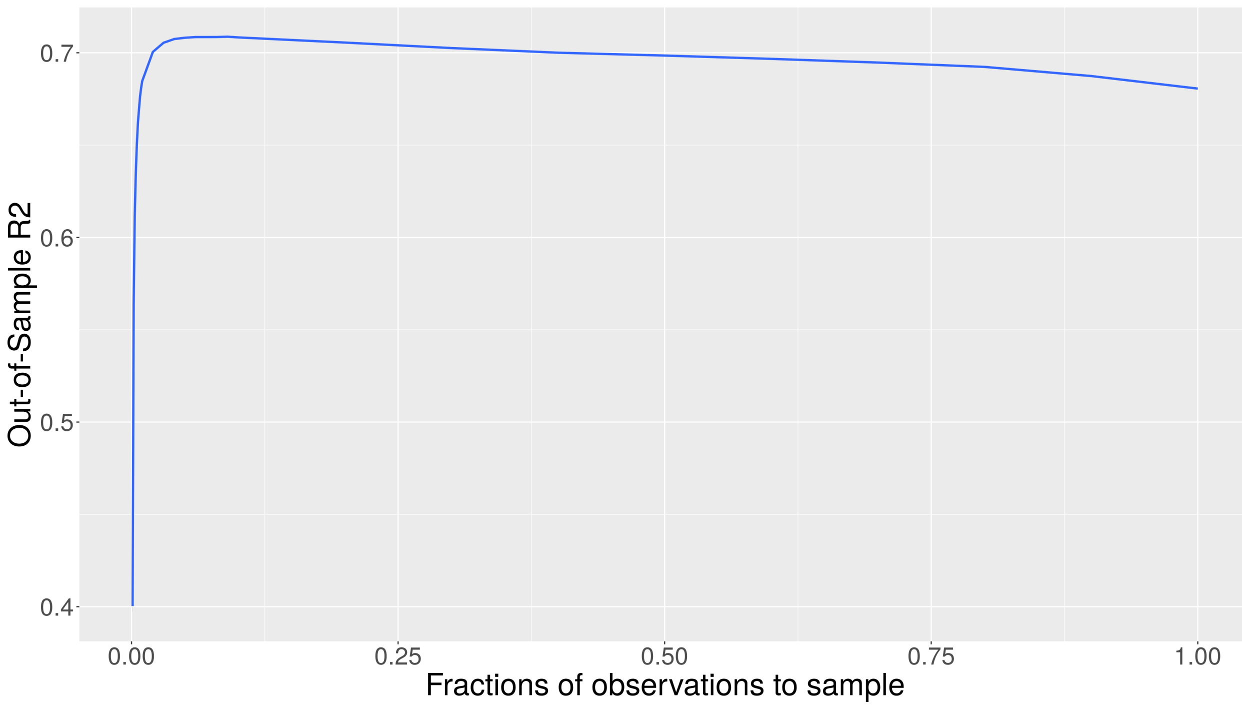

Notes: The figure plots the relation between the depth of decision trees used in the random forest for training up to 1986 January and the out-of-sample $R^2$ for the one-year-ahead earnings forecasts in 1986 February. The out-of-sample $R^2$ is defined as 1 minus the mean squared error implied by using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



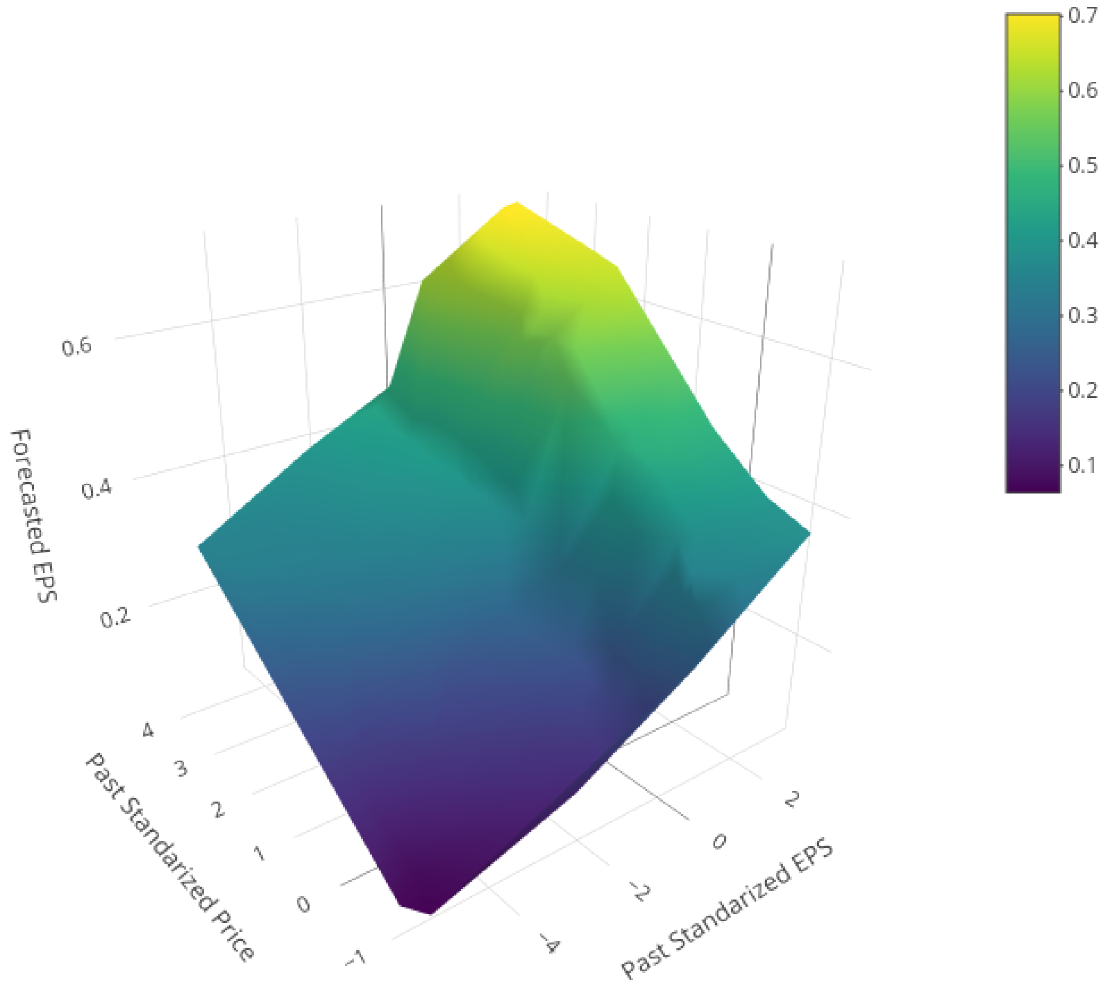

Notes: The figure plots the relation between the fraction of the sample that is taken in each split used in the random forest for training up to 1986 January and the out-of-sample $R^2$ for the one-quarter-ahead earnings forecasts in 1986 February. The out-of-sample $R^2$ is defined as 1 minus the mean 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 13: Cross-validation results of the fraction of the sample that is taken in each split in the one-year-ahead forecast



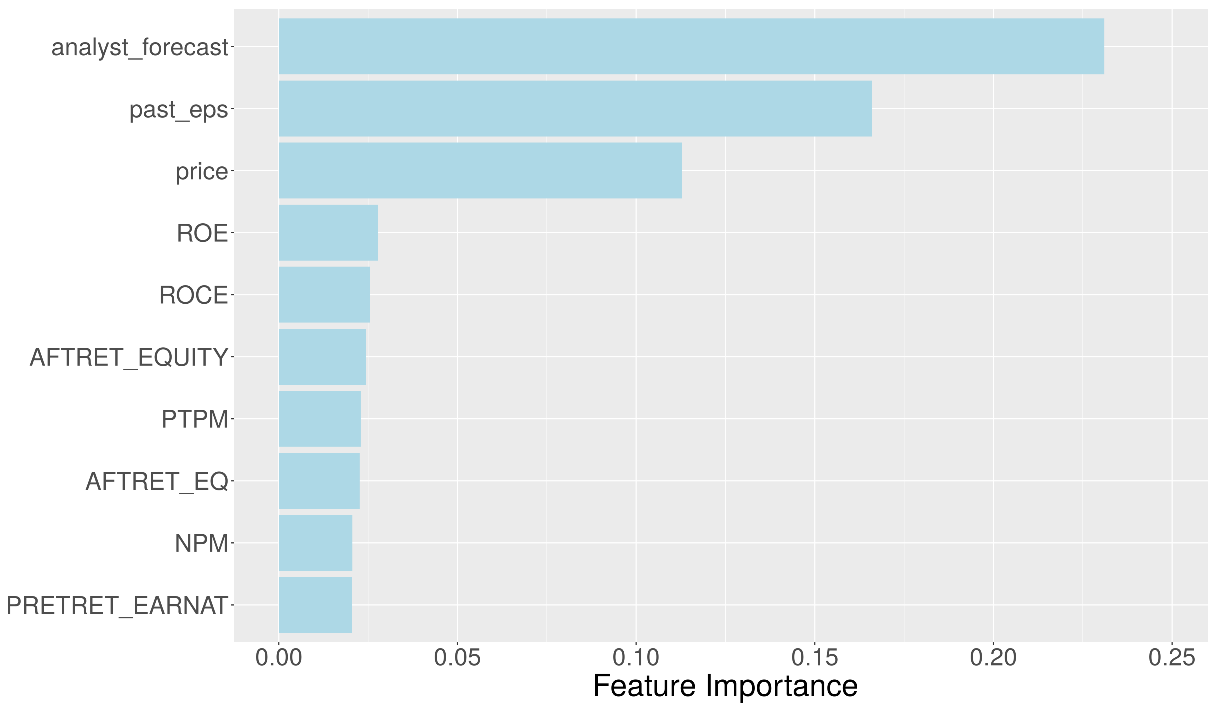

Notes: The figure plots the relation between the fraction of the sample that is taken in each split used in the random forest for training up to 1986 January and the out-of-sample $R^2$ for the one-year-ahead earnings forecasts in 1986 February. The out-of-sample $R^2$ is defined as 1 minus the mean 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 14: EPS as a non-linear function of stock price and past EPS



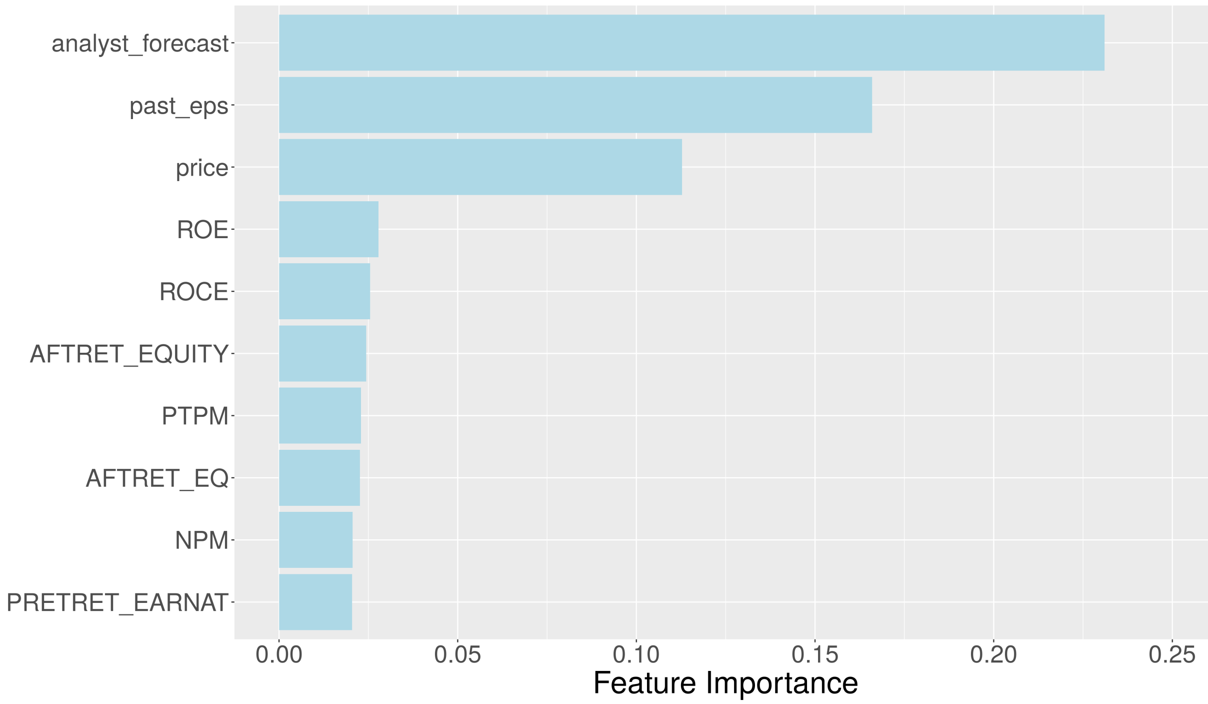

Notes: The figure plots the partial dependence plot of one-quarter-ahead realized EPS on past EPS and stock price. The partial dependence plot is calculated from a random forest regression of EPS on the variables mentioned in Section 3.2. The random forest regression for the figure uses 2000 trees and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

Figure 15: Feature importance of the one-quarter-ahead forecast



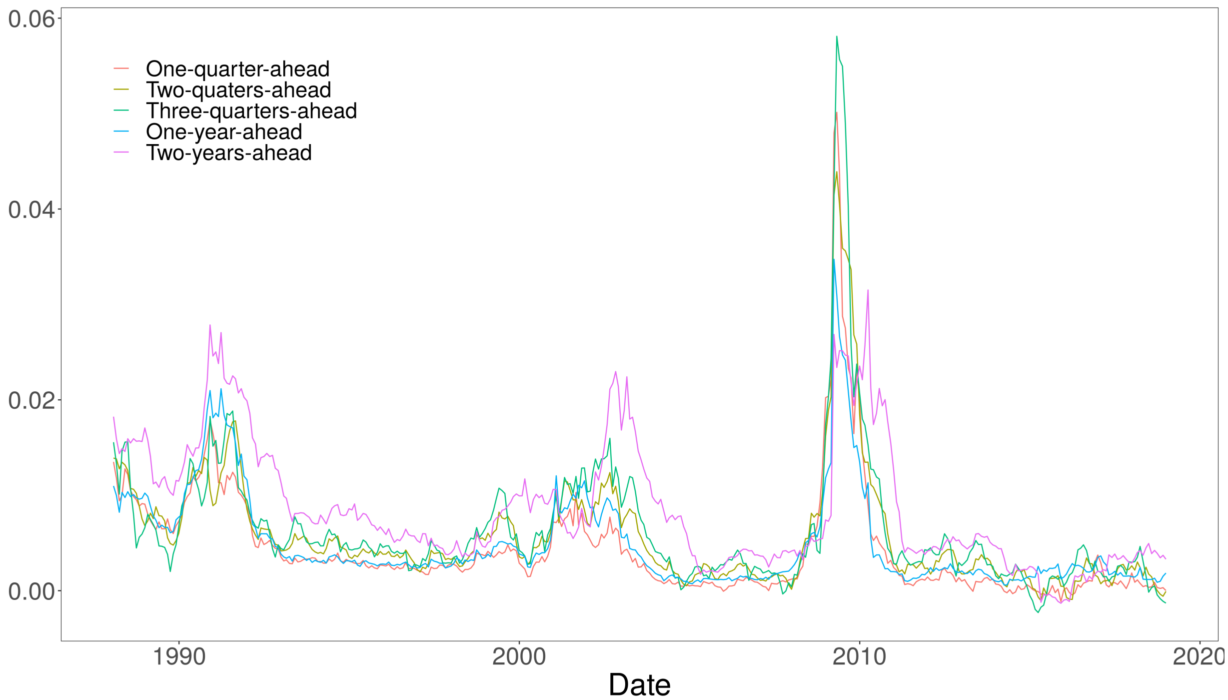

Notes: The figure plots the time-series average of feature importance of the 10 most important variables for the one-quarter-ahead earnings forecasts. The feature importance for each variable is the normalized sum of the reduced mean squared error decrease when splitting on that variable using the method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

Figure 16: Feature importance of the one-year-ahead forecast



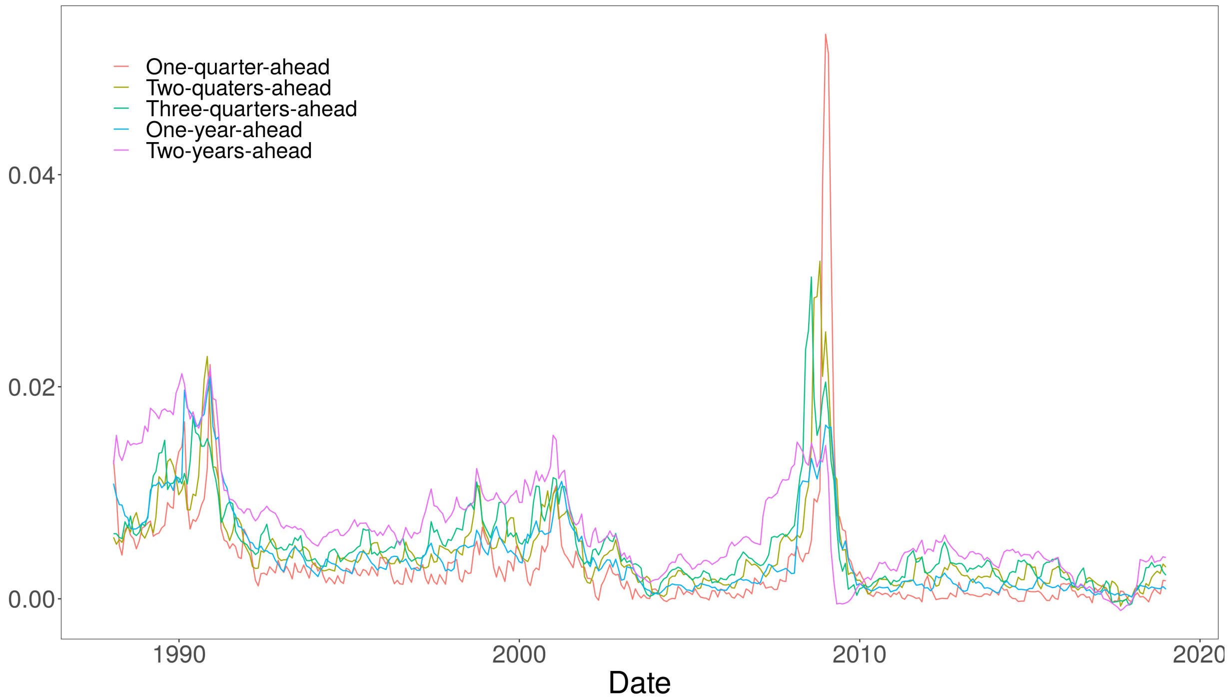

Notes: The figure plots the time-series average of feature importance of the 10 most important variables for the one-year-ahead earnings forecasts. The feature importance for each variable is the normalized sum of the reduced mean squared error decrease when splitting on that variable using the method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

Figure 17: Average bias of analysts' earnings expectations relative to machine learning forecasts



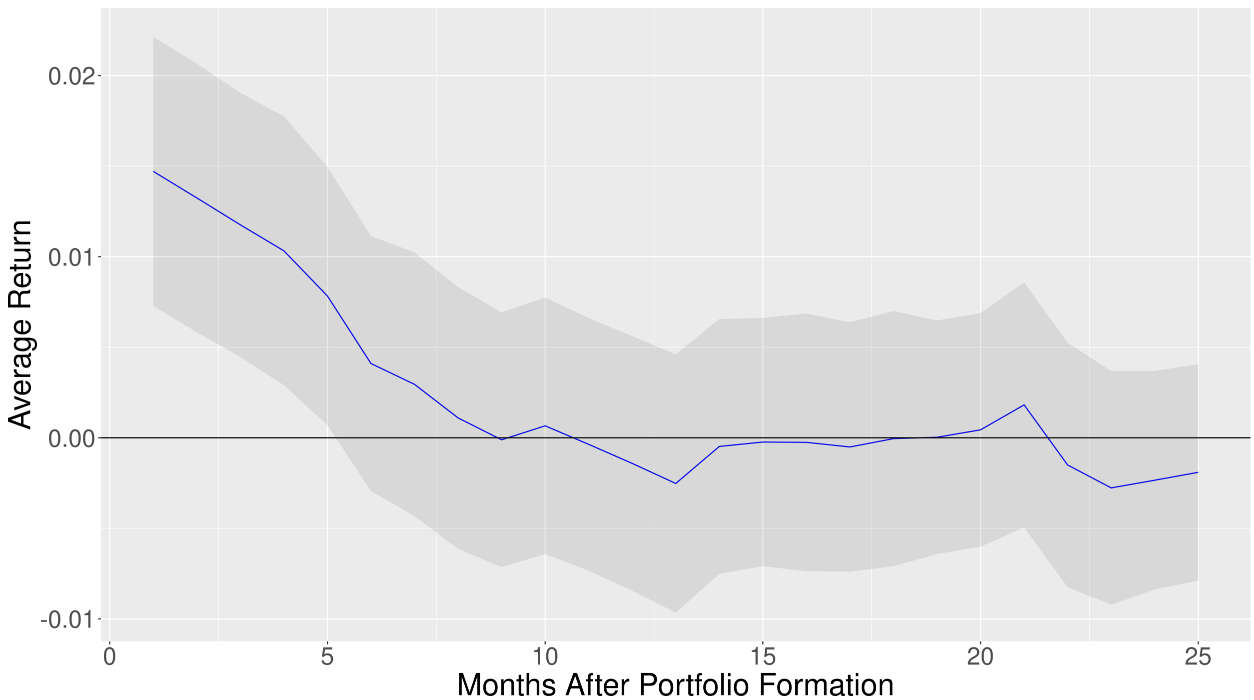

Notes: The figure plots the average conditional bias of analysts' earnings expectations, which is measured as the average of the bias of expectations of individual firms. We trim the data at the 1% level each period before taking the average. The bias is calculated as the difference between analysts' 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 18: Average realized bias of analysts' earnings expectations



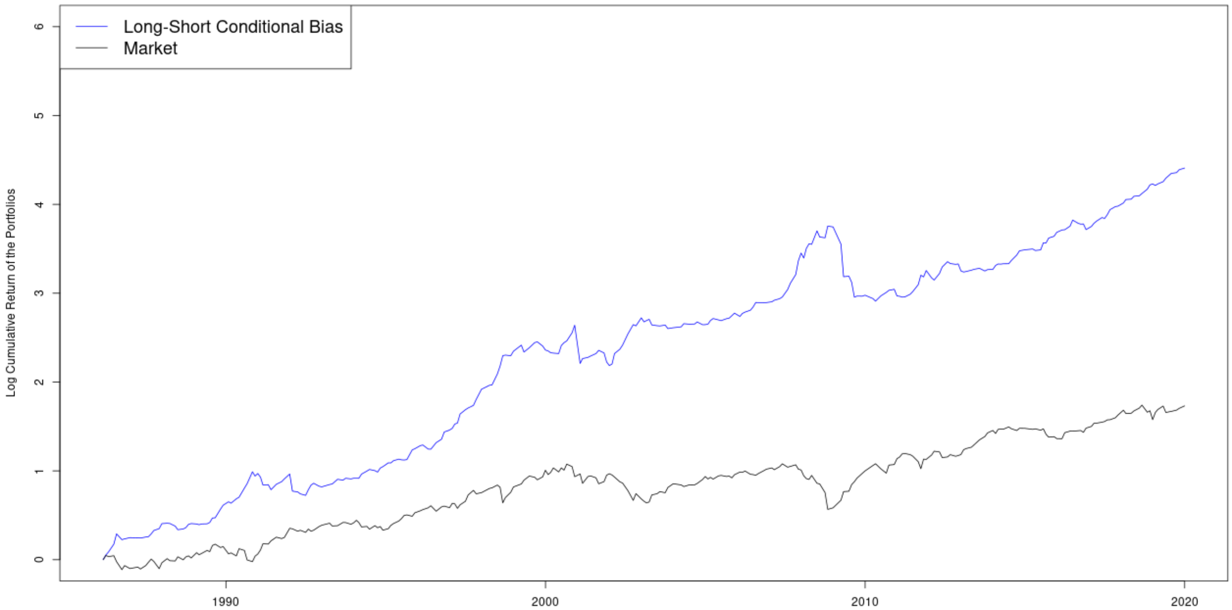

Notes: The figure plots the realized bias of analysts' earnings expectations, which is measured as the average of the bias of expectations of individual firms. We trim the data at the 1% level each period before taking the average. The bias is calculated as the difference between analysts' earnings forecast 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 19: Average Conditional Bias Portfolio Returns for Different Horizons




Notes: This figure plots the (log) cumulative performance of the return of a value-weighted long-short portfolio that is short on firms with the highest conditional earnings forecast bias (using ML forecasts as a benchmark) and long on the firms with the lowest bias. The figure also plots the market return for comparison. The market return data comes from Kenneth R. French's website. The sample period is 1986 to 2019.

Figure 20: Cumulative Performance of the Portfolios Sorted on Conditional Bias




Notes: This figure plots the (log) cumulative performance of the return of a value-weighted long-short portfolio that is short on firms with the highest conditional earnings forecast bias and long on the firms with the lowest. The figure also plots the market return for comparison. The market return data comes from Kenneth R. French's website. The sample period is 1986 to 2019.

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 | 0.005 | 1,022,661 |
| $t$-stat | | | | -0.17 | 6.59 | | | 6.54 | |
| Two-quarters-ahead | 0.323 | 0.376 | 0.323 | -0.001 | 0.053 | 0.094 | 0.102 | 0.007 | 1,110,689 |
| $t$-stat | | | | -0.13 | 10.31 | | | 7.75 | |
| Three-quarters-ahead | 0.343 | 0.413 | 0.341 | 0.002 | 0.072 | 0.121 | 0.132 | 0.007 | 1,018,958 |
| $t$-stat | | | | 0.31 | 11.55 | | | 8.08 | |
| One-year-ahead | 1.194 | 1.320 | 1.167 | 0.027 | 0.154 | 0.670 | 0.686 | 0.021 | 1,260,060 |
| $t$-stat | | | | 1.64 | 6.24 | | | 5.17 | |
| Two-years-ahead | 1.384 | 1.771 | 1.387 | -0.004 | 0.384 | 1.897 | 2.009 | 0.035 | 1,097,098 |
| $t$-stat | | | | -0.07 | 8.33 | | | 6.57 | |

Table 3: Fama–Macbeth regressions

Notes: This table reports the Fama–Macbeth cross-sectional regressions of monthly stocks' returns on the conditional earnings forecast 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. We multiply the coefficient on the bias score by 100 to make it easier to compare. (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 earnings forecast 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 earnings forecast 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 ($\beta$) | $t$-stat | Coef ($\beta$) | $t$-stat | Coef ($\beta$) | $t$-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 earnings forecast 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 earnings forecast 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 rows 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 |

# Appendix A. Model

In this appendix, we present a tractable nonlinear 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 nonlinear 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.[33] We assume $E[(f(x) - E[f(x))^2] = var(f(x)) \equiv \sigma_{fx}^2 > 1$ and $var(g(v)) \equiv \sigma_{gv}^2$, 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 $v_i$ 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

[33] 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 $f(x) = f(-x)$.

Table 9: Net stock issuances and conditional biases

Notes: Panel A reports the time series average of net stock issuances of value-weighted portfolios sorted on the conditional earnings forecast 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. We multiply the coefficient on the bias score by 100 to make it easier to compare. 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 | | -0.109 | | -0.109 |
| $t$-stat | | -4.96 | | -4.91 |
| Intercept | 0.035 | 0.095 | 0.005 | 0.079 |
| $t$-stat | 8.52 | 3.43 | 0.57 | 1.97 |
| $R^2$ (%) | 2.888 | 8.724 | 0.913 | 6.969 |

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.[34]

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

[34] 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.

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:

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

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.[35] 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.

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

**Spurious in-sample linear predictability**

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

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' earnings forecasts and the linear earnings 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 $Corr(\tilde{f}, \tilde{M}) = Cov(\tilde{f}, \tilde{M}) = -a, a > 0$.[36]

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'

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

forecast and the linear forecast is given by:

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

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

In the model (and in the empirical results) analyst earnings estimates are better than linear forecasts. Nevertheless, the bias in the linear earnings 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.[37]

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.[38] 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:

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

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

# References


Antoniou, C., Doukas, J. A. and Subrahmanyam, A. (2015). Investor sentiment, beta, and the cost of equity capital, *Management Science* **62**: 347–67.

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, University of North Carolina at Chapel Hill.* .

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

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


[37] 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.

[38] 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.

Ball, R. T. and Ghysels, E. (2018). Automated earnings forecasts: Beat analysts or combine and conquer?, *Management Science* **64**: 4936–52.

Bianchi, F., Ludvigson, S. C. and Ma, S. (2022). Belief distortions and macroeconomic fluctuations, *American Economic Review* **112**: 2269–315.

Bordalo, P., Gennaioli, N., Ma, Y. and Shleifer, A. (2020). Overreaction in macroeconomic expectations, *American Economic Review* **110**: 2748–82.

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

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

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

Breiman, L. (2001). Random forests, *Machine Learning* **45**: 5–32.

Bryzgalova, S., Pelger, M. and Zhu, J. (2020). Forest through the trees: Building cross-sections of stock returns, *Working Paper, London Business School.* .

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts, *American Economic Review* **105**: 2644–78.

Daniel, K., Hirshleifer, D. and Sun, L. (2019). Short- and long-horizon behavioural factors, *Review of Financial Studies* **33**: 1673–736.

Diether, K. B., Malloy, C. J. and Scherbina, A. (2002). Differences of opinion and the cross section of stock returns, *Journal of Finance* **57**: 2113–41.

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

Engelberg, J., McLean, R. D. and Pontiff, J. (2020). Analysts and anomalies, *Journal of Accounting and Economics* **69**: 101249.

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

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

Fama, E. F. and French, K. R. (2008). Dissecting anomalies, *Journal of Finance* **63**: 1653–78.

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

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

Freyberger, J., Neuhierl, A. and Weber, M. (2020). Dissecting characteristics nonparametrically, *Review of Financial Studies* **33**: 2326–77.

Green, J., Hand, J. R. and Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns, *Review of Financial Studies* **30**: 4389–436.

Grennan, J. and Michaely, R. (2020). Artificial intelligence and high-skilled work: Evidence from analysts, *Working Paper, Santa Clara University.* .

Griffin, J. M., Harris, J. H., Shu, T. and Topaloglu, S. (2011). Who drove and burst the tech bubble?, *Journal of Finance* **66**: 1251–90.
**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, *Review of Financial Studies* **33**: 2223–73.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The elements of statistical learning*, In *Springer series in statistics*, Springer, New York.

Hirshleifer, D. and Jiang, D. (2010). A financing-based misvaluation factor and the cross-section of expected returns, *Review of Financial Studies* **23**: 3401–36.

Hong, H. and Sraer, D. (2016). Speculative betas, *Journal of Finance* **71**: 2095–144.

Hou, K., Van Dijk, M. A. and Zhang, Y. (2012). The implied cost of capital: A new approach, *Journal of Accounting and Economics* **53**: 504–26.

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**: 266–91.

Ishwaran, H. (2015). The effect of splitting on random forests, *Machine Learning* **99**: 75–118.

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, *Journal of Finance* **73**: 1183–223.

La Porta, R. (1996). Expectations and the cross-section of stock returns, *Journal of Finance* **51**: 1715–42.

Nembrini, S., König, I. R. and Wright, M. N. (2018). The revival of the Gini importance?, *Bioinformatics* **34**: 3711–18.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**: 703–8.

Pontiff, J. and Woodgate, A. (2008). Share issuance and cross-sectional returns, *Journal of Finance* **63**: 921–45.

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**: 885–924.

Scherbina, A. (2004). Analyst disagreement, forecast bias and stock returns, *Working Paper, Brandeis University.* .

Scherbina, A. (2007). Suppressed negative information and future underperformance, *Review of Finance* **12**: 533–65.

So, E. C. (2013). A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts?, *Journal of Financial Economics* **108**: 615–40.

Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors, *Review of Financial Studies* **30**: 1270–315.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* **48**: 817–38.

Zhou, G. (2018). Measuring investor sentiment, *Annual Review of Financial Economics* **10**: 239–59.