

Behavioral Machine Learning?

Computer Predictions of Corporate Earnings also Overreact^{*}

Murray Z. Frank

Jing Gao

Keer Yang

March 22, 2023

Abstract

There is considerable evidence that machine learning algorithms have better predictive abilities than humans in various financial settings. But, the literature has not tested whether these algorithmic predictions are more rational than human predictions. We study the predictions of corporate earnings from several algorithms, notably linear regressions and a popular algorithm called Gradient Boosted Regression Trees (GBRT). On average, GBRT outperformed both linear regressions and human stock analysts, but it still overreacted to news and did not satisfy rational expectation as normally defined. By reducing the learning rate, the magnitude of overreaction can be minimized, but it comes with the cost of poorer out-of-sample prediction accuracy. Human stock analysts who have been trained in machine learning methods overreact less than traditionally trained analysts. Additionally, stock analyst predictions reflect information not otherwise available to machine algorithms. We provide a simple model to illustrate the impact of increased analyst use of machine learning methods on equity market equilibrium.

^{*}Respective affiliations are Murray Frank, murra280@umn.edu, University of Minnesota, Minneapolis, MN. Jing Gao, gao00268@umn.edu, University of Minnesota, Minneapolis, MN. Keer Yang, kkeyang@ucdavis.edu, University of California at Davis, Davis, CA

Contents

1	Introduction	2
2	Data	4
3	Empirical Strategies	5
3.1	Using ML to Make Predictions	5
3.2	Testing for Overreaction	7
3.3	Predictable Forecast Errors	8
4	Empirical Results	9
4.1	Prediction MSE	14
4.2	Machines Do Overreact	16
4.3	Why Do Machine Overreact?	17
4.4	Overreaction to Aggregate or Firm-Specific Shocks?	20
4.5	Linear Regression Models Also Overreact	22
5	Stock Analysts	24
5.1	Technical Analysts	24
5.2	Technical VS Non-Tech Analysts	26
6	Model	30
6.1	Setup	30
6.2	Equilibrium	35
7	Conclusion	38
A	Appendix	41

1 Introduction

According to [Bordalo, Gennaioli and Shleifer \(2022\)](#) there is strong evidence that the expectations of professional forecasters, corporate managers, consumers, and investors are often biased towards overreaction. This evidence seriously challenges the rational expectations and efficient markets hypotheses. This bias is said to be rooted in human psychology, notably selective recall and a tendency to focus on salient information. This perspective can also help to explain certain financial and macroeconomic phenomena that are otherwise difficult to understand, ([Bordalo et al., 2020, 2021a,b](#)).

While there is evidence that machine learning has better predictive power than humans in some finance settings, it is unclear whether machine algorithms produce more rational predictions than humans. To address this question, we study the predictions of corporate earnings from several algorithms, notably linear regressions¹ and a popular algorithm called Gradient Boosted Regression Trees (GBRT), see [Friedman \(2001, 2002\)](#); [Chen and Guestrin \(2016\)](#). We test for overreaction following the approach used in ([Bordalo et al., 2021a](#)). On average, GBRT outperformed both linear regressions and human stock analysts. However, like human forecasts, the machine learning algorithm predictions also overreact to news and do not satisfy rational expectation as normally defined.

It is of interest to see if we can control the degree of overreaction of the algorithm. Our first attempt was to alter the variables in the data set fed to the algorithm. This had little impact on the overreaction. Next we consider the fact that the main results use hyper-parameters selected by cross-validation. We therefore examined alternative, less optimized hyper-parameters to see if we can get rid of overreaction. The overreaction can be reduced by lowering the learning rate, but this also reduces the algorithm's responsiveness to new information. With hyper-parameters that in effect cause the algorithm ignore news, it does not overreact to the news.

We paid considerable attention to comparing the computer algorithmic predictions to those of human stock analysts. When compared to predictions from linear regressions and from professional stock analysts, the machine learning predictions overreaction is not as strong. However, due to the overreaction the machine learning generated predictions reject the hypothesis of rational expectations as normally defined.

Using hand collected data on stock analysts' training we distinguish those with training in

¹In untabulated results we also find that predictions generated using the classical Fama-Macbeth method to estimate parameters ([Fama and MacBeth, 1973](#); [Fama and French, 2006](#)) also overreact.

adopting machine learning methods from those with a more conventional training. Analysts with a machine learning background overreact less to news, even after we standardize on the firms being covered. It is worth stressing that while use of machine learning methods is helpful in some respects it does not strictly dominate human analysts. We show that the human stock analyst predictions contain information beyond that contained in standard databases of macroeconomic variables and firm accounting attributes.

To understand the implications of the growing use of machine learning for equity market equilibrium, we provide a simple model that shows the effect of increasing the number of tech-oriented analysts. Our model is based on [Begenau et al. \(2018\)](#). In the model equilibrium, the presence of a growing number of technically trained stock analysts reduces the magnitude of overreactions. Furthermore, the change to the training of stock analysts affects the amount of equity issued in the equilibrium. The greater rationality by in effect reducing the amount of overoptimism on the part of investors in response to a good shock, means that there is reduced demand for risky assets. Since equity is a risky claim, there is less demand for equity. In the equilibrium this means that less equity is issued due to the increasing importance of technically trained stock analysts. While we do not have a direct test of this implication, empirically we distinguish earlier years (1994-2012) to more recent years (2013-2018) since it has been claimed that in recent years machine learning methods are much more widely used. Overreaction does seem to be weaker in the more recent years.

Related Literature This paper is related to a growing body of literature that investigates the predictive power of machine learning in finance. Our empirical setting is similar to that of [van Binsbergen et al. \(2022\)](#) and [Cao et al. \(2021\)](#), who also compare the predictive power between analysts and machine learning models. Specifically, we find that using a gradient boosting regression tree model leads to a similar magnitude of decrease in mean squared errors as documented in [van Binsbergen et al. \(2022\)](#). However, in our setting to predict annual earnings, the efficacy of the machine learning model is not as strong as in [Cao et al. \(2021\)](#). Though [Cao et al. \(2021\)](#) finds similar evidence that machine learning models perform worse after the 2008 financial crisis. Our paper is the first to document the overreacting behavioral of machine learning corporate earnings forecasts.

[Bianchi et al. \(2022\)](#) uses machine learning as a benchmark to study the belief distortion of individual analysts. Relative to that paper our results are novel in that, rather than using machine

learning as a benchmark to study belief distortion, we study the quality and the rationality of the machine learning predictions themselves. [Bianchi et al. \(2022\)](#) find that machine learning does not overreact in their empirical setting studying the time series of macro variables. Our results differ in that our machine learning forecasts of firm earnings overreact. We also provide novel results that show the difference between the effects of aggregate shocks from firm-specific shocks. Predictions for the aggregate shocks exhibit some overreaction, but notably less than is found for firm-specific shocks.

The rest of the paper is organized as follows. Section 2 describes the sources of information we feed into the prediction algorithms. Empirical strategies are explained in Section 3 and the test for overreaction is described. Section 4 provides key test results. Information about predictions by stock analysts is provided in Section 5. Section 6 provides a formal model that we use to help interpret the evidence. The conclusion is Section 7.

2 Data

Our main empirical analysis uses four data sources: (1) Firm earnings forecasts (EPS) by equity analysts, sourced from IBES; (2) Firm financial data from Compustat, and stock return data from the Center for Research in Security Prices (CRSP); (3) Macro time series downloaded from Federal Reserve Bank of Philadelphia real time data; and (4) Manually collected background information on equity analysts obtained from LinkedIn and cross-references with FINRA brokercheck.

The IBES database provides micro-level data on firm-level forecasts made by equity analysts. These analysts make monthly forecasts before the end of the fiscal year. We use equity analysts' forecasts of firms' annual earnings per share (EPS) with a forecasting horizon ranging from 12 to 23 months. The forecasting horizon is the difference in months, between the fiscal year end month of the realized annual EPS to be predicted, and the month when the analyst made the forecast. For each firm and each month, we take the median value of the forecasts as the median consensus forecast. We use realized annual EPS data from IBES Actual files.

3 Empirical Strategies

3.1 Using ML to Make Predictions

The machine learning predictions mimics the forecasts by equity analysts. Equity analysts made forecasts monthly on the annual EPS of the firm. For EPS of firm i in each fiscal year y , we ask the machine learning model to make predictions each month from 12 to 23 months prior to the end month of the fiscal year, using the most updated information available during the month in fiscal year y . For prediction made in year y at year-month t , we use observations in the past 5 years, from $y - 5$ to $y - 1$, to train the model. Suppose our forecast horizon is h month, $h = 12, 13, \dots, 23$. We made forecast using information available in year y at year-month t , denote as $\Phi_{i,t}$. Because forecast horizon is h , we want to predict EPS of firm i at year-month $t + h$, $t + h$ should be firm i 's fiscal year end month. If the year-month $t + h$ is not firm's fiscal year end month, we do not make predictions.

The information set $\Phi_{i,t}$ of firm i at year-month t includes: (1) last annual EPS, the last three quarters' quarterly EPS; (2) 67 financial ratios from the Financial Ratios Suite by WRDS, which provide firm-month level information that is available at year-month t ; (3) macro variables calculated using Federal Reserve Bank of Philadelphia real-time data; and (4) the median consensus forecast of analysts at each firm-month. We also use the monthly median consensus forecast of analysts to incorporate the idea of the importance of analysts' private information. For each investment horizon h at each month-year t (denote y as the calendar year of the month-year t), we use a training sample consisting of five-year observations prior to the forecast date year y . The training sample includes the $X_{i,\tau}, Y_{i,\tau+h}$ where τ are within year $y - 5$ and $y - 1$, and $Y_{i,\tau+h}$ is publicly available by the end of $y - 1$.² We retrain the model each year to make forecasts for the following year.

If we are given the information set $\Phi_{i,t}$ in year y at year-month t and we want to forecast the annual EPS realization in the fiscal-year end month $t + h$, we can make predictions based on the following model:

$$F_t X_{i,t+h} = f_y(\Phi_{i,t}) \quad (1)$$

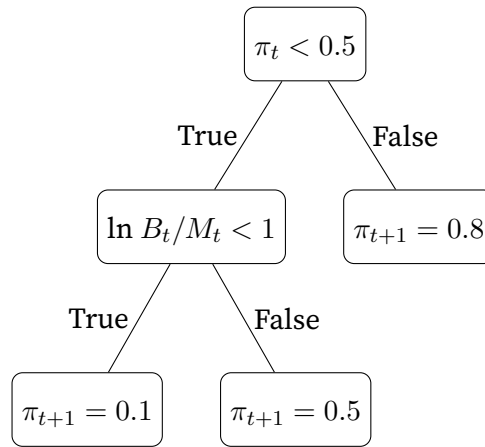
where f_y is trained using the past 5-year observations. $F_t X_{i,t+h}$ is the machine learning model's year-month t forecast for firm i 's annual EPS realization at the fiscal year end month $t + h$. Similar to analysts, the machine learning model could make forecast at any month before fiscal year end

²This restriction is stricter than $\tau + h \leq t$ because earnings are usually reported with a lag.

month.

Gradient boosting regression tree model The machine learning algorithm we used is gradient boosting regression tree model. Gradient boosting regression trees are a particularly prominent and empirically successful prediction method used in many applications. It starts with regression trees and then refines them iteratively by focusing on the errors in the previous iteration. This idea of refinement by focusing on error correction is known as boosting. Gradually an entire forest of trees is constructed. Then an average across the set of trees is used as the model's prediction.

Figure 1: Tree Example



Gradient boosting is the most widely adopted version of boosted forests. It starts by estimating decision trees with fixed shallow depth. Then it computes the residuals for the trees. At the next iteration more weight is devoted to the cases in which the model fit poorly. In the end an ensemble of trees are used to ‘vote’ on the appropriate results. This generally reduces the bias in a simple tree model while maintaining the low variance. The main drawback relative to a simple tree is that forests do not have such simple depictions that show how each variable affects the final result.

There are three important hyper-parameters in gradient boosting: the depth of the tree d (max_depth), the number of trees in the forest M (n_estimators), and learning rate γ (learning_rate). The default parameters we use have the following values: depth of the tree is 2, number of trees in the forest is 50, and learning rate is 0.1. This set of hyper-paramters are obtained through cross-validation method using sample before 1994. We have systematically carried out the analysis with the hyper-parameters optimized using cross-validation. The hyper-parameters were chosen to optimize the out-of-sample mean squared errors before 1994. Except where noted, the Tables use the default hyper-parameters.

A more formal representation can be found in Section 4.3 when we examine the variation in model hyper-parameters. Gradient boosting estimation was done using the software GradientBoostingRegressor from <https://scikit-learn.org/stable/modules/ensemble.html>, and also using XGBoost from Chen and Guestrin (2016). The results for the two implementations of gradient boosting are very similar. The reported results in the Tables use the algorithm GradientBoostingRegressor from <https://scikit-learn.org/stable/>.

Linear Regressions Similarly, we could use linear regression models to forecast EPS. The forecasting procedures strictly follow the procedure used in gradient boosting forecasts. The only difference here is that the function f_y used to fit the training sample is a linear regression model. We use the same forecasting variable $\Phi_{i,t}$ as in the gradient boosting model.

Mean Squared Errors We use average mean squared errors MSE to measure the predictive performance of each model.

$$MSE = \frac{\sum_i (X_{i,t+1} - F_t X_{i,t+1})^2}{N} \quad (2)$$

3.2 Testing for Overreaction

We follow the framework proposed by Bordalo et al. (2021a) to test for overreaction in analysts' and machine forecasts. We start with the monthly consensus forecasts made by analysts and the monthly forecasts made by machine learning model. For each EPS of firm i at fiscal year $t + 1$, we have machine learning forecasts and consensus analysts forecasts made 12 to 23 months prior to the fiscal year end month. We take the average of the 12 median consensus forecasts. This is the average forecasts made during fiscal year t on the EPS of firm at fiscal year $t + 1$, denoted as $F_t X_{t+1}$.

We merge the firm-year level forecasts to standard firm financial data from COMPUSTAT. We then convert the per share forecasting error to per total assets forecasting error, by multiplying firm's total number of shares and divided by total assets. The results are similar when we instead scaled by total capital stock. All firm level variables are scaled by total assets. We further restrict our sample to forecast made from 1994 to 2018, with at least one analyst forecast, resulting in 54534 firm-year observations.

To test overreaction, we run the following regression.

$$x_{i,t+1} - F_{i,t}x_{i,t+1} = \beta y_{i,t} + \lambda_t + \mu_i + \varepsilon_{i,t} \quad (3)$$

To examine the magnitude of overreaction, we also compare the difference in overreaction coefficients (β) using different forecast models. We use seemingly unrelated regressions to compare the regression coefficients using forecasting errors from different forecast models. We report χ^2 statistics and p-values of Wald tests calculated using covariance estimated from the seemingly unrelated regressions. Because any contamination from measurement errors and fixed effects is the same when we use different forecast errors, we can directly compare the coefficients β . A lower β indicates a lower overreaction coefficient (θ).

3.3 Predictable Forecast Errors

In this section, we provide a simple model of predictable forecast errors to clarify the sources of errors. We analyze how machine learning algorithm may improve the predictive power, and how future forecast errors could be correlated with historical information.

Assume analysts or Machine predicts the the value of EPS x_{it} for firm i in period t . x_{it} follows the following process

$$x_{i,t+1} = \delta x_{i,t} + f(z_{i,t}) + \varepsilon_{i,t+1} \quad (4)$$

where $f(z_{i,t})$ can be interpreted as firm/type specific component, but determined by information at time t . And the forecast is

$$F_t x_{i,t+1} = \hat{\delta} x_{i,t} + \hat{f}(z_{i,t}) + \theta \hat{\delta} \varepsilon_{i,t} \quad (5)$$

The MSE of the forecast is then

$$E(x_{i,t+1} - F_t x_{i,t+1})^2 = E((\delta - \hat{\delta})x_{i,t} - (f(z_{i,t}) - \hat{f}(z_{i,t})) - \theta \hat{\delta} \varepsilon_{i,t} + \varepsilon_{i,t+1})^2 \quad (6)$$

$$= (\delta - \hat{\delta})^2 E(x_{i,t})^2 + E(f(z_{i,t}) - \hat{f}(z_{i,t}))^2 + \theta^2 \hat{\delta}^2 E(\varepsilon_{i,t})^2 + E(\varepsilon_{i,t+1})^2. \quad (7)$$

Notice that to go from Equation 6 to Equation 7, it is assumed that all four components in Equation 6 are uncorrelated with each other.

Without loss of generality assume that $\delta = \hat{\delta}$. Forecast Error is then

$$fe_{i,t+1} \equiv x_{i,t+1} - E_t x_{i,t+1} = \delta x_{i,t} + f(z_{i,t}) + \varepsilon_{i,t+1} - (\hat{\delta} x_{i,t} + \hat{f}(z_{i,t})) - \theta \hat{\delta} \varepsilon_{i,t} \quad (8)$$

$$= (\delta - \hat{\delta}) x_{i,t} - (f(z_{i,t}) - \hat{f}(z_{i,t})) - \theta \hat{\delta} \varepsilon_{i,t} \quad (9)$$

$$= -\theta \hat{\delta} \varepsilon_{i,t} - (f(z_{i,t}) - \hat{f}(z_{i,t})) \quad (10)$$

Calculating the correlation between future forecast errors and current period information, we have

$$\text{cov}(fe_{i,t+1}, x_{i,t}) = \quad (11)$$

$$\text{cov}(-\theta \hat{\delta} \varepsilon_{i,t} - (f(z_{i,t}) - \hat{f}(z_{i,t})), \rho x_{i,t-1} + f(z_{i,t-1}) + \varepsilon_{i,t}) \quad (12)$$

$$= -\theta \hat{\delta} \text{var}(\varepsilon_{i,t}). \quad (13)$$

To go from the Equation 12 to Equation 13, we assume that $f(z_{i,t}) - \hat{f}(z_{i,t})$ is uncorrelated with information prior to time t .

With the existence of firm-level heterogeneity, machine learning could improve predictive power by allowing more factors and more complex functional forms. However, omitting $f(z_{it})$ does not cause more overreaction.

4 Empirical Results

In this section, we first analyze the predictability of machine learning models versus analysts' forecasts by comparing the mean squared errors of each forecasting method. Then, we examine whether the machine learning algorithm exhibits overreacting behavior. We find that machine learning forecasts experience similar overreacting behavior, but with a smaller magnitude. We explore potential explanations by comparing machine learning models with different forecasting variables and regularization parameters. Finally, we examine whether similar overreacting behavior also exists when we use linear regression models.

Figure 2: Earning Predictions and Realizations

This figure shows the relation between predicted earnings and realized earnings. Earnings are scaled by total assets, which is equal to $\frac{\text{Predicted EPS} \times \text{Total Shares Outstanding}}{\text{Total Assets}}$. The x-axis and y-axis are truncated at 0.5 to exclude outliers in the figure being presented.

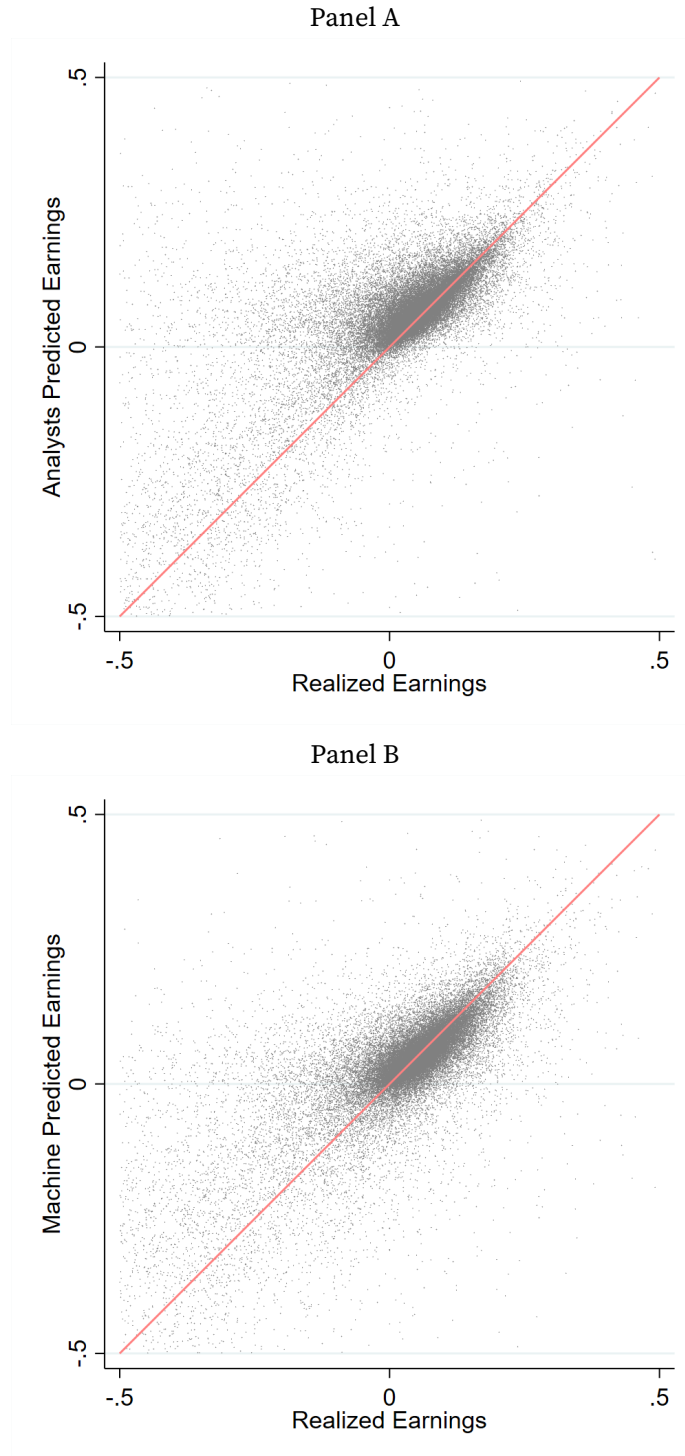


Figure 3: Auto-correlation

This figure shows the density plot of the coefficient estimation of AR(1) models. For each firm, we estimated the following AR(1) model: $x_{t+1} = \rho x_t + \varepsilon_{t+1}$, where earnings were scaled by total assets. We then plotted the distribution of the estimated ρ . Panel A includes all firms for which we could estimate the AR(1) process. In Panel B, we restricted our sample to firms with at least 10-year observations.

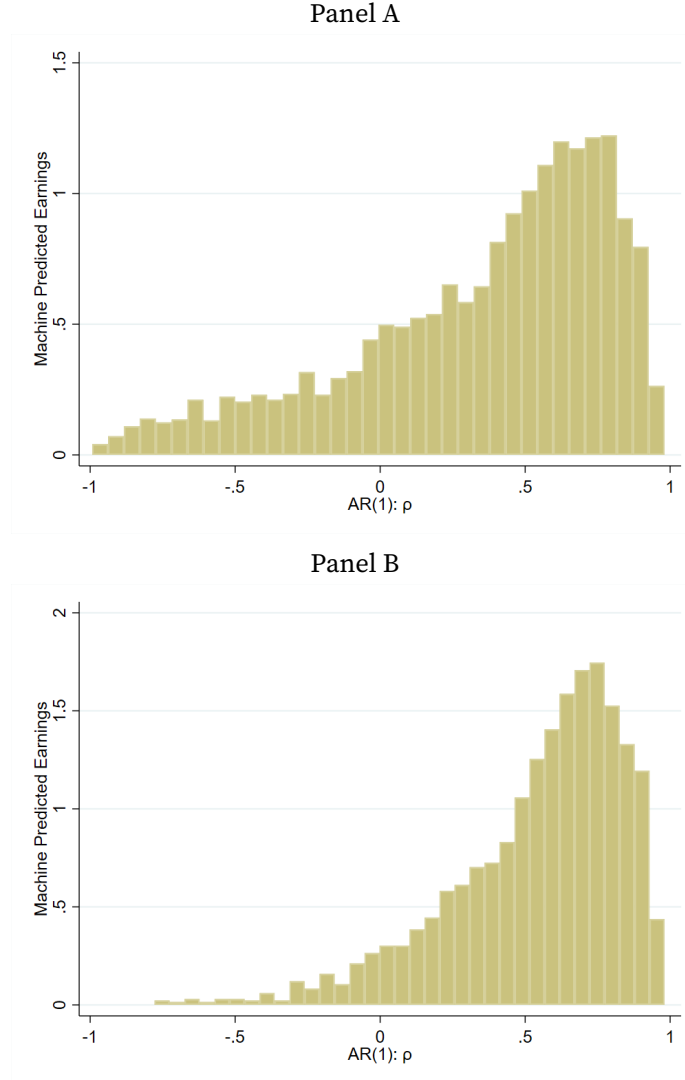


Table 1: Summary Statistics

This table reports the summary statistics of key variables. The mean squared error (MSE) shows the out-of-sample performance of the forecasting models. The MSE for analysts forecasts is 1.312.

(a) % ML Win				
	Forecasting Variables	Learning rate	MSE	Ratio of MSE: Analysts vs ML
(1)	Analysts Forecast	-	1.312	-
Gradient Boosting				
(2)	All	0.1	1.255	1.045
(3)	Lag EPS	0.1	1.892	0.693
(4)	All minus Analysts Forecast	0.1	1.500	0.875
(5)	All plus Financial Statements	0.1	1.253	1.047
(6)	All	0.01	2.579	0.509
(7)	All	0.03	1.576	0.832
(8)	All	0.2	1.259	1.042
Linear Regressions				
(9)	All	-	1.856	0.707
(10)	Lag EPS	-	1.881	0.698
(11)	All minus Analysts Forecast	-	2.611	0.502
(12)	All plus Financial Statements	-	1.850	0.709

Table 1: Summary Statistics (continued)

(b) All Analysts

	Mean	Median	Std Dev	25	75	N
EPS	1.130	0.890	1.916	0.050	1.95	54534
Analyst Forecast	1.424	1.139	1.722	0.413	2.12	54534
ML Forecast	1.155	0.887	1.509	0.203	1.82	54534
Analyst Forecast Error	-0.031	-0.007	0.109	-0.039	0.01	
ML Forecast Error	0.000	0.020	0.158	-0.019	0.05	
Investment	0.058	0.039	0.061	0.019	0.07	54534
Debt Net Issuance	0.017	0.000	0.090	-0.016	0.03	54534

(c) Tech and Non-Tech Analysts

	Mean	Median	Std Dev	25	75	N
Sample period: 1994-2018						
EPS	1.351	1.06	2.169	0.11	2.28	39419
Non-Tech Analyst	1.678	1.341	1.996	0.483	2.5	36155
Tech Analyst	1.67	1.198	2.442	0.253	2.65	14901
Non-Tech Analyst MSE	0.837	0.027	2.98	0.003	0.2	34935
Tech Analyst MSE	1.617	0.039	5.79	0.005	0.34	14585
Investment	0.055	0.036	0.061	0.017	0.07	39773
Sample period: 1994-2012						
EPS	1.248	1.05	1.812	0.23	2.08	25320
Non-Tech Analyst	1.561	1.311	1.613	0.583	2.26	22929
Tech Analyst	1.479	1.1	1.888	0.385	2.25	7922
Non-Tech Analyst MSE	0.748	0.027	2.776	0.003	0.19	22093
Tech Analyst MSE	1.408	0.043	5.28	0.005	0.32	7786
Investment	0.06	0.039	0.064	0.02	0.07	25479
Sample period: 2013-2018						
EPS	1.535	1.1	2.685	-0.12	2.82	14099
Non-Tech Analyst	1.881	1.433	2.513	0.248	3.07	13226
Tech Analyst	1.886	1.395	2.933	0.02	3.2	6979
Non-Tech Analyst MSE	0.989	0.027	3.297	0.003	0.21	12842
Tech Analyst MSE	1.857	0.035	6.315	0.004	0.35	6799
Investment	0.048	0.03	0.056	0.013	0.06	14294

4.1 Prediction MSE

This section provides an overall evidence on the predictability of firm earnings. Table 1 presents evidence on the efficacy of gradient boosting methods, analyst forecasts, and linear regression models. The first column indicates the variables used as forecasting variables. The set of factors labeled ‘All’ is our benchmark set of factors used for prediction, described in Section 2. We also use different sets of factors in the model. The second column represents the learning rate hyperparameter we use to control for over-fitting in gradient boosting tree models; the default number is 0.1. The third column represents the mean squared error of forecasts of firm annual EPS. The last column indicates the ratio of the MSE of the analyst’s forecast and the forecast model used in that row. A ratio larger than one indicates that the model in the row performs better than the median consensus analyst forecast.

Row (1) of Table 1 presents the prediction results for the median consensus analyst forecast, which serves as our benchmark for mean squared error. The mean squared error of the analyst’s forecast is 1.312.

Row (2) of Table 1 presents the prediction results for gradient boosting using all factors. The model performs relatively well, with an MSE of 1.255. The analyst’s forecast has a 4.5% higher squared forecast error. The results suggest that our machine learning model improves predictability, but not as strongly as in other settings (Cao et al. (2021)). The improvement is similar to the estimation results from van Binsbergen et al. (2022).

Row (3) again uses the gradient boosting regression tree model but reduces the set of variables to only lagged EPS. As might be expected, the model estimation performance is significantly reduced. The mean squared error becomes 1.892, and the performance is worse than that of the analyst’s forecasts.

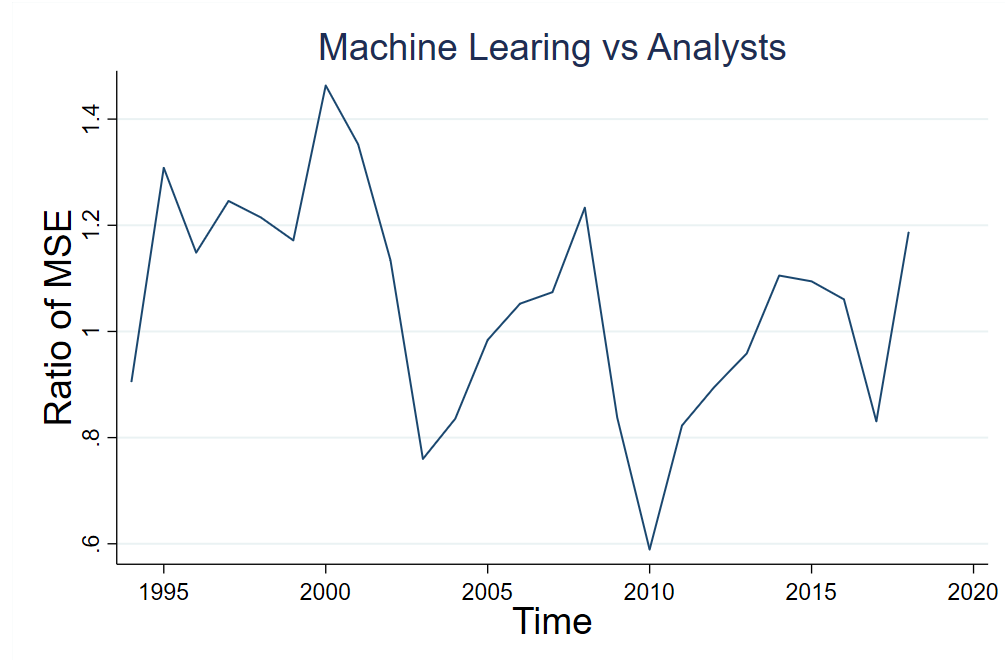
Cao et al. (2021) argue that human forecasts can complement machine learning forecasts. In the main set of forecasts, we already include the median consensus analyst forecast as a predictor. To examine the importance of the analyst’s forecast, we remove it from the factors. Surprisingly, when we remove the analyst’s forecast as a predictor in row (4), the mean squared errors drop to 1.5, which is smaller than the mean squared error of the analyst’s forecast itself. The results suggest that when forecasting firm earnings, the analyst’s private assessment is extremely valuable. The information generated by analysts cannot be replaced by incorporating a large set of public financial ratios.

In row (5), we extend the set of variables to include all firm balance sheet variables. The mean squared errors are very similar to row (1), suggesting that the financial variables we used in the main forecasting set have already captured the firm heterogeneity in explaining earnings.

Rows (6) to (8) capture the predictability when we use different learning rate parameters. When the learning rate is very small, the model does not fit the training sample enough, which leads to relatively poor out-of-sample performance. As we increase the learning rate, the model's performance improves. However, row (8) shows that if the learning rate is too large, the model runs the risk of over-fitting the historical information, leading to a drop in prediction performance.

Figure 4: Ratio of MSE

This figure shows the ratio of mean squared errors of human forecasts and machine forecasts at different years. A ratio smaller than one suggests that humans beat the machines in that year.



We also examine the time-series variations in the relative strength between human analysts and machine learning models. Figure 4 shows the ratio of mean squared errors of human forecasts and machine forecasts at different years. A ratio smaller than one suggests that humans beat the machines in that year. From 1994 to 2018, we observed 10 years where humans beat the machines, including 1994, 2003, 2004, 2005, 2009, 2010, 2011, 2012, 2013, and 2017. One possible explanation for humans beating the machines is that the training samples experienced negative shocks and were, therefore, less useful for future earnings predictions. The annual SP500 return in 2000, 2001,

and 2022 were -10.1%, -13.0%, and 23.4%, respectively. The annual percentage change of the SP500 index in 2008 was -38.5%, while in 2015 was -0.73%. When we include those observations in our 5-year training sample, we find that humans beat the machines.

The results are similar if we use the last three years' observations to train the model, but on average we get a lower decrease in MSE if we use only three years' observations. We also tried to include more years during recessions, as [Cao et al. \(2021\)](#) suggested, and found that the results are similar, with still lower increases in average mean squared errors.

4.2 Machines Do Overreact

In this section, we follow the empirical strategy defined in Section 3.2. The key question is whether the ML generated forecasts exhibit similar overreaction behavior to that found for human reported forecasts.

[Bordalo et al. \(2021a\)](#) run linear regressions in which the dependent variable is the error at date $t + 1$ of stock analyst predictions from IBES. The independent variables are date t values of the forecast, profit, investment, or debt issuance. In each of these regressions firm and year fixed effects are included. According to [Bordalo et al. \(2021a\)](#) under rational expectations the slope coefficients should be zero. Under diagnostic expectations the slope coefficients should be negative. That is because of beliefs that overreact both to good and to bad news. [Bordalo et al. \(2021a\)](#) use linear regressions with fixed effects to carry out this test. We follow [Bordalo et al. \(2021a\)](#) in Table 2. We also add a columns (1) (2) that uses the prediction error from analysts.

Columns (1) and (2) confirm the widely documented empirical findings that human forecasts exhibit overreaction behavior. In Column (2), the regression coefficient is 0.143, indicating that a one standard deviation increase in investment (0.061) leads to an 8% higher disappointment in earnings next year ($0.061 * 0.143/0.109$). Columns (3) and (4) perform the same regressions as Columns (1) and (2), but using machine learning predictions instead of human forecasts. The magnitude of the regressions is reduced compared to Columns (1) and (2), but the results in all columns (1) to (4) are negative and statistically significant.

To address concerns about downward dynamic panel bias due to the inclusion of firm fixed effects, we provide GMM estimations in Appendix Table A2. The GMM results give us a larger regression coefficient of 0.472. Both gradient boosting and analyst forecasts exhibit overreacting behavior even when we correct for downward dynamic panel bias. The magnitude is comparable to the coefficient of 0.887 in [Bordalo et al. \(2021a\)](#). The difference in magnitude could result from

Table 2: Predictable Forecast Errors

This table reports OLS regression results of regressing forecast error at $t + 1$ on the information at time t . Forecast error at $t + 1$ is $\text{Forecast Error}_{t+1} := p_{t+1} - F_t p_{t+1}$, where F_t is forecast based on date t data. In columns (1) and (2), forecast $F_t p_{t+1}$ is the average of median monthly earnings forecast of all analysts' forecasts conducted during fiscal year t . In columns (3) and (4), forecast $F_t p_{t+1}$ is the average of median monthly earnings forecast of all machine learning forecasts conducted during fiscal year t . Earnings are converted from EPS forecast to total earnings. Investment_t capital expenditure at time t . Debt issuance_t is net debt issuance at time t . All variables are scaled by total assets.

	(1) Forecast Error Analysts	(2) Forecast Error Analysts	(3) Forecast Error Machine	(4) Forecast Error Machine
Investment	-0.018* (-1.845)	-0.143*** (-10.065)	-0.016** (-1.978)	-0.107*** (-8.175)
Firm FE	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes
Period	1994-2018	1994-2018	1994-2018	1994-2018
N	54536	53324	54536	53324
AdjR2	0.02	0.32	0.03	0.23

different sample firms.

4.3 Why Do Machine Overreact?

To examine why machine learning forecasts exhibit such "irrational" behavior, we will analyze the number of forecasting variables used and the model's regularization parameter.

To examine the magnitude of overreaction, we compare the differences in overreaction coefficients (β) using different forecast models. We use seemingly unrelated regressions to compare the regression coefficients estimated using forecasting errors from different forecast models. We report χ^2 statistics and p-values of Wald tests calculated using the covariance estimated from the seemingly unrelated regressions. Since any contamination from measurement errors and fixed effects is the same when we use different forecast errors, we can directly compare the coefficients β . A lower β indicates a lower overreaction coefficient (θ). The mean squared error (MSE) shows the out-of-sample performance of the forecasting models.

Table 3 shows the results obtained using different information sets to make predictions. Consistent with our simple illustrative model in Section 3.3, ignoring firm heterogeneity could lead to higher squared forecast errors but does not affect overreaction behavior.

In order to further investigate the phenomenon of overreaction in machine learning, we next

Table 3: Predictable Forecast Errors Using Different Forecasting Variables

This table presents the results of an OLS regression analysis of the forecast error at time $t + 1$ and Wald test results from seemingly unrelated regressions. The forecast error at $t + 1$ is defined as $\text{Forecast Error}_{t+1} = p_{t+1} - F_t p_{t+1}$, where $F_t p_{t+1}$ is the average of median monthly earnings forecast based on data from time t , and p_{t+1} represents the actual earnings at time $t + 1$. Forecasts are calculated from various machine learning models that use different forecasting variables. In column (1), lagged annual earnings of the firm are used as the forecasting variables. In column (2), the forecasting variables include firm characteristics and lagged annual and quarterly earnings. In column (3), the forecasting variables include firm characteristics, analysts forecasts, and lagged annual and quarterly earnings. In column (4), the variables include firm characteristics, analysts forecasts, lagged annual and quarterly earnings, and financial statement items. All earnings are converted to total earnings from EPS forecasts and scaled by total assets. Investment_t is capital expenditure at time t . Debt issuance_t is net debt issuance at time t . The χ^2 of the Wald test is used to test the difference of regression coefficients from column (3), with the results of the test presented in the table. The p-value of the test is also included in brackets.

	(1) Forecast Error Machine	(2) Forecast Error Machine	(3) Forecast Error Machine	(4) Forecast Error Machine
Investment	-0.109*** (-5.431)	-0.114*** (-7.692)	-0.107*** (-8.174)	-0.109*** (-8.388)
χ^2	0.0495 [0.824]	1.705 [0.192]	- -	2.221 [0.136]
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Forecasting Variables:				
Lagged Earnings	Yes	Yes	Yes	Yes
Firm Char.	No	Yes	Yes	Yes
Analysts Forecasts	No	No	Yes	Yes
Financial Statement Items	No	No	No	Yes
Period	1994-2018	1994-2018	1994-2018	1994-2018
N	53321	53321	53321	53321
AdjR2	0.43	0.23	0.23	0.22
Forecast MSE	1.891	1.500	1.255	1.253

analyze the impact of machine learning hyper-parameters on the behavior of machine learning forecasts. To fix ideas consider the following representation of the gradient boosting regression tree model,

$$\hat{y} = F_M(x) = \sum_{m=1}^M h_m(x) \quad (14)$$

where \hat{y} is the fitted value, h_m is decision tree regressor with depth of d , and M is the number of trees in the forest. $F_M(x)$ is solved by using a greedy algorithm framework,

$$F_k(x) = F_{k-1}(x) + \gamma h_k \quad (15)$$

where γ is the learning rate. The learning rate shrinks the contribution of each additional tree. h_k is the newly added tree solved by minimize a loss function L given $F_{k-1}(x)$

$$h_k = \arg \min_h \sum_i (L(y_i, F_{k-1}(x_i) + \gamma h(x_i))) \quad (16)$$

There are three main hyper-parameters in gradient boosting: the depth of the tree d (max_depth), the number of trees in the forest M (n_estimators), and learning rate γ (learning_rate).

The default parameters we start with have the following values: depth of the tree is 2, number of trees in the forest is 50, and learning rate is 0.1. We have systematically analyzed the effect of using hyper-parameters optimized using cross-validation. The hyper-parameters were chosen to optimize the out-of-sample mean squared errors before 1994. Except where noted, the Tables use the default hyper-parameters.

Over-reaction is fundamentally about the response to a shock. The responsiveness of the algorithm is primarily controlled by the learning rate. So we are interested in examining the impact of altering the learning rate away from a default or an optimized value. It should be noted that we find that other hyper-parameters do not have much of an affect on the estimated overreaction by the algorithm.

What does the learning rate do? It is a regularization strategy that shrinks the contribution of each additional tree. Larger learning rate leads to a higher probability of over-fitting given the training sample. When we set a high learning rate, the algorithm can capture the deep structure in the training sample which may be just transitory. Our interest is in the relationship between overreacting behavior of the algorithm predictions, and over-fitting parameters in the prediction model.

Table 4: Predictable Forecast Errors Using Hyper-parameters

This table presents the results of an OLS regression analysis of the forecast error at time $t + 1$ and Wald test results from seemingly unrelated regressions. The forecast error at $t + 1$ is defined as $\text{Forecast Error}_{t+1} = p_{t+1} - F_t p_{t+1}$, where $F_t p_{t+1}$ is the average of median monthly earnings forecast based on data from time t , and p_{t+1} represents the actual earnings at time $t + 1$. Forecasts are calculated from various machine learning models that use different hyper-parameters. The default hyper-parameters are set as the following values: depth of the tree is 2 and number of trees in the forest is 50. Table reports the results using different learning rates. Investment_t is capital expenditure at time t . Debt issuance_t is net debt issuance at time t . The χ^2 of the Wald test is used to test the difference of regression coefficients from column (3), with the results of the test presented in the table. The p-value of the test is also included in brackets. The mean squared error (MSE) shows the out-of-sample performance of the forecasting models.

	(1) Forecast Error Machine	(2) Forecast Error Machine	(3) Forecast Error Machine	(4) Forecast Error Machine
Learning rate	0.01	0.03	0.1	0.2
Investment	0.027 (0.781)	-0.075*** (-4.099)	-0.107*** (-8.174)	-0.110*** (-7.718)
χ^2	39.730 [0.000]	10.464 [0.001]	- -	0.576 [0.448]
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Period	1994-2018	1994-2018	1994-2018	1994-2018
N	53321	53321	53321	53321
AdjR2	0.76	0.66	0.23	0.20
Forecast MSE	2.579	1.576	1.255	1.259

Table 4 shows the results obtained using different hyper-parameters to make predictions. We find that a lower learning rate generates predictions that overreact less. This should make intuitive sense. Consider an extreme case. Suppose the algorithm entirely ignored all shocks. Then there would be no response at all. If a prediction does not change, it certainly is not going to change excessively.

4.4 Overreaction to Aggregate or Firm-Specific Shocks?

The results from [Bianchi et al. \(2022\)](#) suggest that macro-economic shocks might affect things differently from firm-specific shocks. So we follow [Chen et al. \(2007\)](#) and decompose machine forecast errors into a market-related variation and a firm-specific variation. We do not examine industry-specific variation because our machine learning models are not trained using industry-specific variables. Consider the following regression:

Table 5: Decomposing Forecast Errors

This table presents the results of an OLS regression analysis of the forecast error at time $t + 1$. We decompose forecast errors into a market-related variation and a firm-specific variation, using the following regression $e_{i,t} = \beta_i + \beta_{i,m}r_{m,t} + \varepsilon_{i,t}$. $e_{i,t}$ is the forecast error of firm i at year t , $r_{m,t}$ is the market forecast error at year t , which is the equal-weighted average forecast error at year t . We define market-related forecast error as $\hat{\beta}_{i,m}r_{m,t}$ and the firm-specific forecast error as $e_{i,t} - \hat{\beta}_{i,m}r_{m,t}$. The dependent variable in column (1) is the market-related forecasting errors of human analysts, while column (2) is the firm-specific forecasting errors of human analysts. The dependent variable in column (3) is the market-related forecasting errors of machine analysts, while column (4) is the firm-specific forecasting errors of machine analysts. Investment_t is capital expenditure at time t .

	(1)	(2)	(3)	(4)
	Forecast Error Analysts Market-Related	Forecast Error Analysts Firm-Specific	Forecast Error Machine Market-Related	Forecast Error Machine Firm-Specific
Investment	-0.040*** (-4.919)	-0.103*** (-9.349)	-0.013** (-2.023)	-0.093*** (-8.178)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Period	1994-2018	1994-2018	1994-2018	1994-2018
N	53321	53321	53321	53321
AdjR2	1.00	0.99	0.96	0.90

$$e_{i,t} = \beta_i + \beta_{i,m}r_{m,t} + \varepsilon_{i,t} \quad (17)$$

where $e_{i,t}$ is the forecast error of firm i at year t , $r_{m,t}$ is the market forecast error at year t , which is the equal-weighted average forecast error at year t . We define market-related forecast error as $\hat{\beta}_{i,m}r_{m,t}$ and the firm-specific forecast error as $e_{i,t} - \hat{\beta}_{i,m}r_{m,t}$. In Table 5, we regress market-related error and firm-specific error on investment separately.

Columns (1) and (2) show the results for analysts' forecasts. The dependent variable in column (1) is the market-related forecasting errors of human analysts, while column (2) is the firm-specific forecasting errors of human analysts. An increase in investment rate leads to disappointment in earnings forecasts by human analysts next year. In columns (3) and (4), we decompose the forecast errors using the gradient boosting model. The dependent variable in column (3) is the market-related forecasting errors of machine analysts, while column (4) is the firm-specific forecasting errors of machine analysts. The overreaction coefficient in column (3) is -0.013 , while the overreaction coefficient in column (1) is -0.040 , suggesting that applying machine learning largely reduces the market-related overreaction. The regression coefficients have similar magnitudes in

columns (2) and (4).

4.5 Linear Regression Models Also Overreact

Our analysis is focused on gradient boosting predictions. That is motivated by the wide use and documented success of the algorithm. “Gradient Tree Boosting or Gradient Boosted Decision Trees (GBDT) is a generalization of boosting to arbitrary differentiable loss functions, see the seminal work of Friedman (2001). GBDT is an accurate and effective off-the-shelf procedure that can be used for both regression and classification problems in a variety of areas...” <https://scikit-learn.org/stable/modules/ensemble.html#gradient-boosting>

It is natural to wonder if our overreaction results are specific to that particular algorithm. They are not. To show this we consider predictions from perhaps the most familiar methods used in finance. In this section, we use linear regression models to forecast EPS. The forecasting procedures strictly follow the procedure used in gradient boosting forecasts. The only difference here is that the function f_y used to fit the training sample is a linear regression model. We use the same forecasting variables $\Phi_{i,t}$ as in the gradient boosting model.

Table 6 presents the forecasts calculated from linear regression models that use different forecasting variables. In column (1), lagged annual earnings of the firm were used as the forecasting variables. In column (2), the forecasting variables included firm characteristics and lagged annual and quarterly earnings. In column (3), the forecasting variables included firm characteristics, analysts’ forecasts, and lagged annual and quarterly earnings. In column (4), the variables included firm characteristics, analysts’ forecasts, lagged annual and quarterly earnings, and financial statement items.

Unlike the gradient boosting regression tree model, there is no hyper-parameter governing the level of regularization in OLS. The number of forecasting variables represents the information set used in the model, as well as the potential severity of over-fitting. Table 6 shows that when too many forecasting variables are included in a linear model, the forecast variable has an overreacting coefficient quite similar to that of human analysts. We

In Table A3, we also tested whether the machine learning model could generate overreaction similar to human analysts by increasing the learning rate parameter. Unlike linear models, we found that the machine learning forecast’s overreaction is capped from above. The difference between linear models and the gradient boosting regression tree model is that the gradient boosting model embeds the idea of regularization into the optimization algorithm, making it less likely to

Table 6: Predictable Forecast Errors in Linear Models

This table presents the results of an OLS regression analysis of the forecast error at time $t + 1$ and Wald test results from seemingly unrelated regressions. The forecast error at $t + 1$ is defined as $\text{Forecast Error}_{t+1} = p_{t+1} - F_t p_{t+1}$, where $F_t p_{t+1}$ is the average of monthly earnings forecast based on data from time t , and p_{t+1} represents the actual earnings at time $t + 1$. Forecasts are calculated from linear regression models that use different forecasting variables. In column (1), lagged annual earnings of the firm are used as the forecasting variables. In column (2), the forecasting variables include firm characteristics and lagged annual and quarterly earnings. In column (3), the forecasting variables include firm characteristics, analysts forecasts, and lagged annual and quarterly earnings. In column (4), the variables include firm characteristics, analysts forecasts, lagged annual and quarterly earnings, and financial statement items. All earnings are converted to total earnings from EPS forecasts and scaled by total assets. Investment_t is capital expenditure at time t . Debt issuance_t is net debt issuance at time t . The χ^2 of the Wald test is used to test the difference of regression coefficients from column (3), with the results of the test presented in the table. The p-value of the test is also included in brackets. The mean squared error (MSE) shows the out-of-sample performance of the forecasting models.

	(1) Forecast Error OLS	(2) Forecast Error OLS	(3) Forecast Error OLS	(4) Forecast Error OLS
Investment	-0.138*** (-8.196)	-0.113*** (-5.336)	-0.128*** (-6.971)	-0.133*** (-7.871)
χ^2	10.696 [0.001]	0.368 [0.544]	3.537 [0.060]	6.649 [0.010]
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Forecasting Variables:				
Lagged Earnings	Yes	Yes	Yes	Yes
Firm Char.	No	Yes	Yes	Yes
Analysts Forecasts	No	No	Yes	Yes
Financial Statement Items	No	No	No	Yes
Period	1994-2018	1994-2018	1994-2018	1994-2018
N	53321	53321	53321	53321
AdjR2	0.22	0.44	0.22	0.23
Forecast MSE	1.881	2.611	1.856	1.850

create overfitting forecasts that exhibit overreaction similar to linear models.

5 Stock Analysts

So far, we have found that in a simulation setting, machine learning predictions overreact regardless of the information set used to make the prediction. However, the magnitude of overreaction is smaller for machine learning predictions compared to analysts' forecasts.

In this section, we use real analysts' forecasts to further analyze the impact of adopting machine learning models on the overreacting behavior of forecasts. To do so, we need to measure the probability that analysts adopt machine learning models when making predictions. Given our goal of measuring overreaction, an analyst-level measure is more granular and provides more meaningful inferences compared to a brokerage-level machine learning adoption measure.

We use analysts' technical backgrounds as a proxy for the probability of analysts adopting machine learning models. The underlying assumption is that analysts with advanced statistical training are more likely to adopt machine learning algorithms when making EPS forecasts, once they become available. We compare technical analysts' forecasts with those of non-technical analysts before and after the widespread availability of machine learning. This empirical strategy is similar to a difference-in-differences analysis. However, since technical and non-technical analysts have different stock coverage, we cannot directly implement a difference-in-differences framework. Instead, we compare the magnitude of the overreaction coefficients, represented by β , before and after the widespread availability of machine learning algorithm.

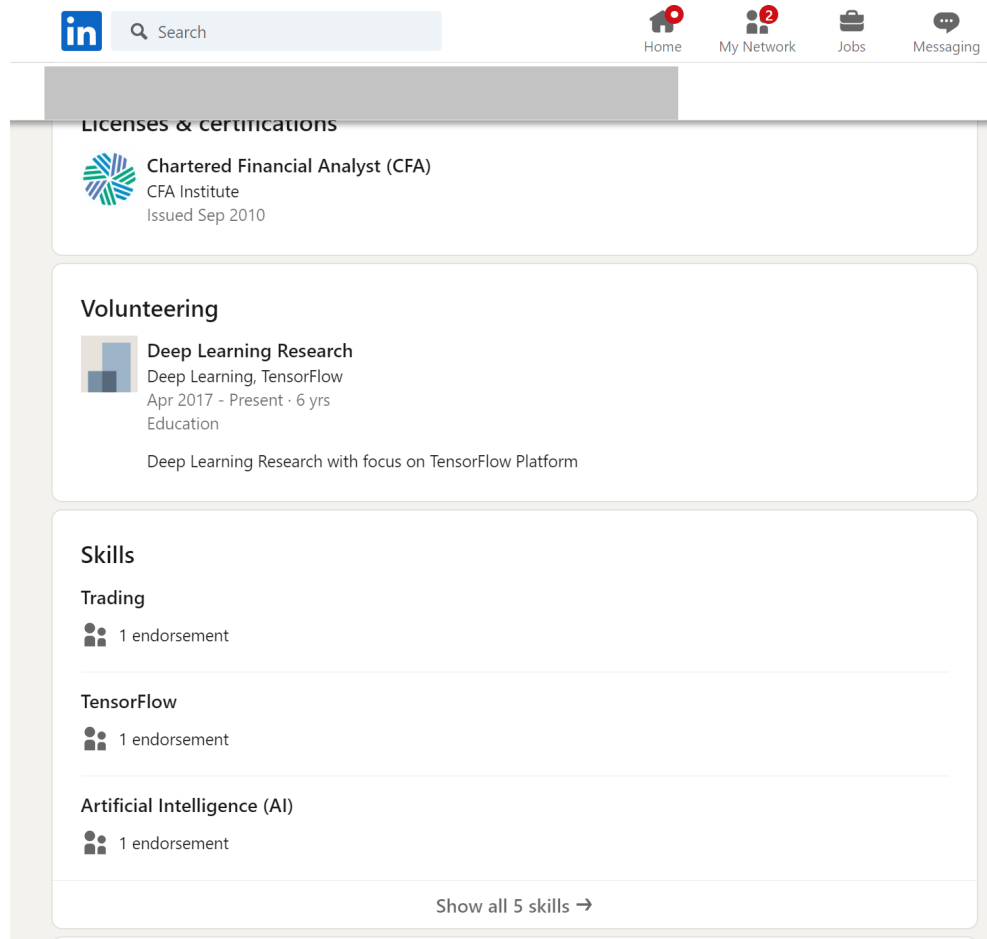
Although the gradient boosting regression tree model becomes popular after the publication of Friedman (2002), it did not become widely accessible until the implementation of several open-source machine learning packages, including scikit-learn and tensorflow. Therefore, we use the year 2012, when scikit-learn became "well-maintained and popular" (<https://en.wikipedia.org/wiki/Scikit-learn>), as the threshold for pre- and post- periods.

5.1 Technical Analysts

We manually collect the technical backgrounds of analysts using their LinkedIn information. Our goal is to measure analysts' statistical tools, and therefore their ability and probability of incorporating machine learning in their EPS forecasts. Here the focus is on the analysts and their backgrounds.

Figure 5: LinkedIn page

This figure gives an example of a technical analysts' LinkedIn page.



To identify analysts' identity and educational background information, we use analysts who gave at least one earnings recommendation in 2018. We cross-check their affiliations on FINRA's brokercheck website and LinkedIn to maximize the likelihood of an accurate manual matching. In total, we successfully linked 858 analysts to their LinkedIn profiles.

To identify analysts' technical backgrounds start with self-reported technical skills, including machine learning, artificial intelligence, and advanced statistics. It should be noted that we do not use 'quantitative modelling' as a keyword since it might refer to a much wider range to prediction methods. Figure 5 gives an example of a technical analysts' LinkedIn page.

We use their higher education background to identify whether analysts have technical skills or not. We manually check the curriculum for each major and identify the majors in Table A1 as technical majors. These majors are mostly STEM majors, but there are some discrepancies.

For example, we do not find advanced statistical courses in the biology major’s curriculum. So analysts with a general biology education are not generally included. However, if an analyst earns a degree in Computational Biology, we find that advanced statistical tools are required. Therefore, we label these analysts as technical analysts.³

In total, we identified 173 analysts with a technical education background. We then distinguish our IBES sample to analysts whom we categorized as technical vs. non-technical. For each firm in each month, we calculated the median consensus forecast by technical analysts $F_t^T x_{it+1}$ and non-technical analysts $F_t^{NT} x_{it+1}$ separately.

To conduct the overreaction test, in each fiscal year $t + 1$, we took the average of the 12 median consensus forecasts made during fiscal year $t + 1$. Starting in 1994 and going to 2018 we have 14,901 firm-year observations for forecasts made by technical analysts and 36155 firm-year observations for forecasts made by non-technical analysts. It should be kept in mind that technical analysts and non-technical analysts may cover different stocks, and this may affect certain comparisons.

5.2 Technical VS Non-Tech Analysts

We first examine whether forecast made by tech analysts and non-tech analysts exhibit different overreacting behavioral after machine learning adoption in 2013. We do so by separately calculating forecasts made by tech analysts and non-tech analysts. We then conduct similar overreacting tests in the previous section.

Table 7a reports OLS regression results of regressing forecast error at $t + 1$ on the information at time t using full sample. In columns (1) and (3), the forecast $E_t p_{t+1}$ is the average earnings forecast of tech analysts’ forecasts conducted during fiscal year t . In columns (2) and (4), the forecast $E_t p_{t+1}$ is the average earnings forecast of non-tech analysts’ forecasts conducted during fiscal year t . Column (1) shows that before 2013, a one percentage point increase in investment led to a 0.181 percentage point stronger disappointment in earnings next year for tech analysts’ forecasts. The negative relationship is statistically significant. Column (2) shows that after 2013, a one percentage point increase in investment led to a 0.09 percentage point stronger disappointment in earnings next year for tech analysts’ forecasts. For tech analysts, the decrease in overreacting coefficients after 2013 is $0.398 = 1 - 0.109/0.181$. Interestingly, the overreaction coefficient becomes insignificant after 2013. The coefficient in column (3) is still statistically significant if we do not cluster at

³Our focus is different from Babina et al. (2021), which focuses on AI-related skills instead of statistical skills. We aim to measure analysts’ ability to understand and use advanced statistics and the probability of using them. Therefore, we do not only focus on AI-related skills.

Table 7: Tech Analysts vs Non-Tech Analysts Background before and after 2013

This table report OLS regression results of regressing forecast error at $t + 1$ on the information at time t . Forecast error at $t + 1$ is $\text{Forecast Error}_{t+1} := p_{t+1} - F_t p_{t+1}$, where F_t is forecast based on date t data. In columns (1) and (3), forecast $F_t p_{t+1}$ is the average of median monthly earnings forecast of tech analysts conducted during fiscal year t . In columns (2) and (4), forecast $F_t p_{t+1}$ is the average of median monthly earnings forecast of non-tech analysts conducted during fiscal year t . Earnings are converted from EPS forecast to total earnings. Investment_t capital expenditure at time t . All variables are scaled by total assets. Sample periods are reported.

(a) Full Sample				
	(1) Forecast Error Tech	(2) Forecast Error Non-Tech	(3) Forecast Error Tech	(4) Forecast Error Non-Tech
Investment	-0.181*** (-2.731)	-0.127*** (-6.913)	-0.109 (-1.424)	-0.091** (-2.865)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Period	1994-2012	1994-2012	2013-2018	2013-2018
N	7367	21309	6359	12316
AdjR2	0.25	0.28	0.25	0.31
(b) Balanced Sample				
	(1) Forecast Error Tech	(2) Forecast Error Non-Tech	(3) Forecast Error Tech	(4) Forecast Error Non-Tech
Investment	-0.160** (-2.395)	-0.119** (-1.986)	-0.147** (-2.254)	-0.113** (-2.069)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Period	1994-2012	1994-2012	2013-2018	2013-2018
N	4902	4902	5100	5100
AdjR2	0.25	0.26	0.29	0.32

the firm level.

When we examine non-tech analysts's the change in overreacting coefficient, the trend is similar. The overreacting pattern becomes weaker after 2013, perhaps due to the relatively recent trends of relying on machine learning models to make predictions. However, the decrease in overreaction is smaller. For non-tech analysts, the decrease in overreacting coefficients after 2013 is $0.252 = 1 - 0.095/0.127$.

Technical analysts and non-technical analysts may have different sample coverage, and the sample coverage may vary before and after 2013. To address this issue, we first restricted our sample to firms that were covered by both technical and non-technical analysts in the same year. However, as few firms were covered by the same analyst both before and after 2013, we calculated the Mahalanobis distance for firms before and after 2013 based on firm age, total assets, high tech industry, and investment rate. We dropped firms before 2013 that were significantly different from firms in 2013. We then ran similar regressions using our balanced sample, and the results are provided in Table 7b. We found similar results: technical analysts, on average, had a much larger decrease in overreaction after 2013, while overreaction for non-technical analysts remained at similar magnitudes.

Overall, the results suggest that tech analysts become less over-reactive after adopting machine learning models after 2013. The results confirm our prediction that adopting machine learning models could reduce overreacting bias.

To examine whether the difference in overreacting can have a real impact, we test whether less overreaction by tech analysts could lead to lower investment reversal. The empirical framework we use to test investment reversal from overreacting is from [Bordalo et al. \(2021a\)](#). If EPS forecasts used by managers overreact to good news, the investment decision based on the overreacting EPS would be larger. Once earnings are realized, the manager would correct the over-investing by under-investing in the next period. Therefore, the higher the overreacting, the higher the next period investment reversal will be. Since realized forecasting errors capture error terms that are not directly related to the overreacting behavior, we use an instrumental variable approach, as used in [Bordalo et al. \(2021a\)](#). The difference between our setting and [Bordalo et al. \(2021a\)](#) is that we use analysts' forecasts, whereas [Bordalo et al. \(2021a\)](#) use managerial forecasts. We show that analysts' forecasts can indeed affect firms' investment decisions in Table A4, even after controlling for realized profits.

We regress forecast error on investment rate, as shown in Table 7, and use the model to calcu-

Table 8: Investment Reversals

This table reports OLS regression results of regressing changes in investments at $t + 1$ and predicted forecast errors. Predicted Forecast Errors (Tech) are predicted value using regression coefficients in Table 7 for tech analysts. Predicted Forecast Errors (Non-Tech) are predicted values using regression coefficients in Table 7 for non-tech analysts.

	(1) ΔInv_{t+1}	(2) ΔInv_{t+1}	(3) ΔInv_{t+1}	(4) ΔInv_{t+1}
Predicted Forecast Errors (Tech)	6.541*** (5.114)		4.533* (2.840)	
Predicted Forecast Errors (Non-Tech)		6.134*** (4.869)		5.333*** (5.422)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Period	1994-2012	1994-2012	2013-2018	2013-2018
N	54	163	186	349
AdjR2	0.26	0.40	0.07	0.11

late predicted forecast errors. We then regress the change in future investment rates to examine whether the predicted forecast errors lead to investment reversals. Because one stock may have both tech and non-tech analysts' forecasts, to address the problem that managers may receive both forecasts from tech and non-tech analysts, we restrict our sample to stocks covered only by one type of analysts. All regressions are done separately for non-tech analysts' forecasts and tech analysts' forecasts. The results are shown in Table 8.

The results are consistent with our findings of overreaction in Table 7. In column (1), a one-percentage-point increase in predicted forecast errors by tech analysts leads to a 6.543-percentage-point increase in investment rates from 1994 to 2012. After 2013, column (3) shows that a one-percentage-point increase in predicted forecast errors by tech analysts leads to a 5.333-percentage-point increase in investment rates, which is lower than the investment reversal before 2013. These results are consistent with the findings in Table 7 that analysts' forecasts by tech analysts decrease after 2013. Similarly, we find that investment reversals caused by non-tech analysts' overreacting forecasts are smaller after 2013.

For tech analysts, the decrease in investment reversals after 2013 is $0.307 = 1 - 4.533/6.541$. For non-tech analysts, the decrease in investment reversals after 2013 is $0.131 = 1 - 5.333/6.134$. Firms covered by non-tech analysts experience a larger decrease in investment reversals after 2013, the result is consistent with our findings that tech analysts' forecasts have a larger decrease

in overreacting after 2013 compared to non-tech analysts.

Overall, the results suggest that the decrease in overreacting by adopting machine learning algorithm may have real implications, firms covered by those analysts experience lower magnitude of investment reversal behavioral, potentially lead to altered investments. However, such overreaction cannot be totally eliminated given the prevalence of overreacting behavioral in machine learning algorithms.

To rule out the possibility that our results are driven by a change in tech and non-tech analysts' coverage after 2013, we compare the average firm characteristics covered by tech and non-tech analysts before and after 2013, as shown in Table 9. We find that tech analysts increased their number of stock coverage after 2013, while all other firm characteristics, including age, total assets, and tech industry, remained unchanged. The results suggest that our findings are unlikely to be driven by a change in tech analysts' stock coverage.

6 Model

Both our simulation exercise and tech analysts results suggest that adopting machine learning could lead to smaller, but still non-negligible, improvements in forecasting accuracy. We use a model to study the impact of machine learning adoption on making forecasts. This is especially important because we want to understand what happens during the transition to a world of complex machine learning models.

6.1 Setup

Based on [Begenau et al. \(2018\)](#), we build up a repeated static model with one firm. In each period, the firm has a 1-period investment project that gives payoffs depending on the realizations of both the aggregate and idiosyncratic profitability. At the beginning of each period, the aggregate profitability realizes and becomes public to all agents in the economy. Then the firm chooses the number of shares to issue in order to maximize its net revenue. The firm uses the raised capital to finance its 1-period investment opportunity that has a stochastic payoff affected by both idiosyncratic and aggregate shocks. Then investors make portfolio choice decisions to maximize their expected utilities. In particular, the investors do not know the idiosyncratic profitability of the firm's investment project. They will randomly receive private signals about the profitability from two types of analysts, i.e., tech and non-tech analysts. Observing the private signal and equity

Table 9: Tech Analyst Coverage

This table estimates average characteristics of the stocks that tech analysts at analyst-year level. Tech is a indicator variables equalling to 1 if the analyst is tech analyst, 0 otherwise. After 2013 is an indicator variable equaling to 1 if the forecast is made after 2013. Number of Stocks is the total number of stocks covered by the analyst. Age is the average age of the stocks covered by the analyst. Total assets is the average total assets of stocks covered by the analyst. High Tech is the percentage of high-tech stocks covered by the analyst. High-Tech stocks are firms with three digit SIC code of 283, 357, 366, 367, 382, 384, or 737.

(a) Coverage before and after 2013

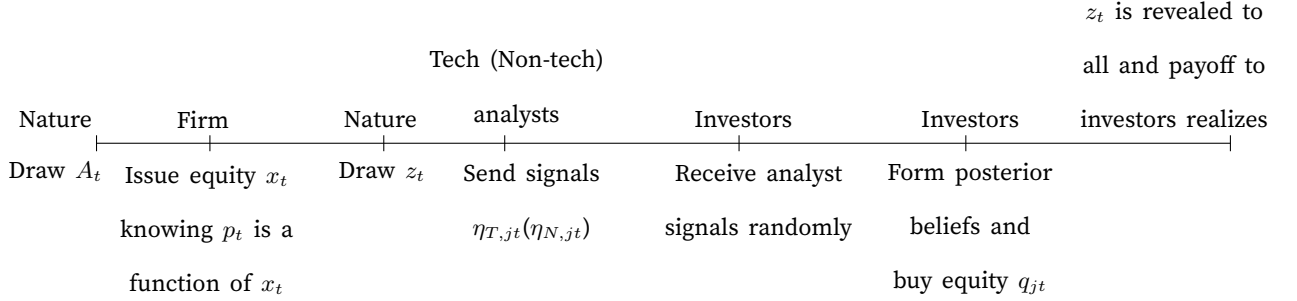
	(1) Num of Stocks	(2) Age	(3) Total Assets	(4) High Tech
Tech*After 2013	0.859*** (3.021)	-0.215 (-0.860)	-0.018 (-0.548)	-0.002 (-0.382)
Analyst FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Period	1994-2018	1994-2018	1994-2018	1994-2018
N	11703	11703	11703	11703
AdjR2	0.58	0.77	0.79	0.92

(b) Tech Analysts Coverage

	(1) Num of Stocks	(2) Age	(3) Total Assets	(4) High Tech
Tech	-0.008 (-0.043)	-1.687*** (-7.362)	-0.194*** (-6.301)	0.220*** (22.969)
Analyst FE	No	No	No	No
Year FE	Yes	Yes	Yes	Yes
Period	1994-2018	1994-2018	1994-2018	1994-2018
N	11716	11716	11716	11716
AdjR2	0.06	0.06	0.05	0.04

price in the market, each investor updates her prior belief about idiosyncratic profitability and forms posterior belief. Investors' expected utilities are formed based on their posterior beliefs. At the end of each period, the payoff and utilities are realized. In the next period, new investors arrive and the same sequence repeats. Figure 6 illustrates the timing within each period t .

Figure 6: Timeline for Each Period



Firm There is one firm that chooses how many shares x_t to issue in order to raise funds for investment. The capital raised from each share is invested in the firm's investment project whose payoff is $A_t z_t$. A_t is the aggregate profitability, while z_t is the idiosyncratic profitability. We introduce A_t just to model the relationship between the aggregate economic state and analysts' forecast error. Otherwise, A_t functions exactly the same as z_t since in the current setting, we only have one firm in the economy. Because we are interested in the analysts' forecasts on the *firm-level* profitability, we assume that A_t is public information and is realized at the very beginning of each period. It follows an AR(1) process $\log A_t = \delta_A \log A_{t-1} + \epsilon_{At}$, with $\epsilon_{At} \sim N(0, \rho_A^{-1})$.⁴

The idiosyncratic productivity, z_t , is not in the firm's information set, i.e., the firm does not know the realized value of z_t when it makes decisions. z_t follows an AR(1) process

$$z_t = \delta z_{t-1} + \epsilon_{zt}, \quad \epsilon_{zt} \sim N(0, \rho_z^{-1}).$$

In fact, as will become clearer later, from the firm's perspective, z_t matters only through equity price p_t . Thus, it is investors' belief about z_t that matters. In this economy, the firm understands the equilibrium behaviors of investors. Conditional on this information set, the firm owner maximizes the current period net revenue from the sale of the firm. Therefore, she solves the following

⁴Because our model is a repeated static model, the dynamic process of A_t does not affect agents' choices in each period. However, the introduction of aggregate profitability and its dynamics enable us to capture the relationship between the performance of machine learning algorithms and the aggregate state that is going to be clearer in Assumption 1.

problem

$$\max_{x_t} E[x_t p_t - \frac{\phi}{2}(\Delta x_t)^2 | \mathcal{F}_t^-], \quad (18)$$

where $\frac{\phi}{2}(\Delta x_t)^2$ is equity issuance cost and $\Delta x_t = x_t - x_{t-1}$. $\mathcal{F}_t^- = \{\mathcal{I}_{t-1}, A_t\}$ is the information set of the firm owner when she makes equity issuance decision and \mathcal{I}_{t-1} represents all the information realized prior to time t .

Given the firm is the only supplier of its stock, its equity issuance affects the price of the stock p_t . Thus, the firm makes decisions knowing that its issuance affects the price of each share.

Analysts There are two types of analysts with a total measure of 1. $\lambda \in [0, 1]$ of them are tech analysts who apply algorithms to make forecasts, $1 - \lambda$ are non-tech analysts.

Investors There is a continuum of investors with a measure of 1. Each investor is endowed with wealth W_t and decides how to allocate her wealth between riskless and risky assets, taking prices as given. When making decisions, investors do not know the firm's idiosyncratic profitability z_t , but they have a common prior that $z_t \sim N(\mu_{zt}, \rho_z^{-1})$ with $\mu_{zt} = E_{t-1}[z_t] = \delta z_{t-1}$.

Investors update their beliefs on z_t based on private signals they receive from analysts and the stock price. Before making investment decisions, each investor randomly draws an analyst forecast from the pool of analysts. By the law of large numbers, λ investors end up receiving the forecast of tech analysts, and $1 - \lambda$ from non-tech analysts. We call the former investors tech investors, and the latter non-tech investors. Investors do not know the exact type of signal they receive.

Investor j only receives one private signal from one analyst: 1) if the signal is provided by a tech analyst, the signal is $\eta_{T,jt} = z_t + \xi_{T,jt}$ with $\xi_{T,jt} \sim N(0, \rho_\eta^{-1})$; 2) if the signal is provided by non-tech analysts, it is $\eta_{N,jt} = z_t + \xi_{N,jt}$ with $\xi_{N,jt} \sim N(\theta \delta \epsilon_{t-1} + f_t, \rho_\eta^{-1})$. $\xi_{T,jt}$ ($\xi_{N,jt}$) is the difference between the true profitability and the signal value, that is, the total forecast error, and is uncorrelated across investors. The mean of the error term of tech analysts' forecasts is normalized to 0, and that of the non-tech analysts is $\theta \delta \epsilon_{t-1} + f_t$. The first term $\theta \delta \epsilon_{t-1}$ captures the behavioral bias as in the first term of Equation 10, and f_t is a scalar that captures the misspecification error in a reduced form as in the second term of Equation 10. Our empirical finding suggests that the MSE of forecasts varies with the aggregate state, thus we assume f_t to be correlated with the aggregate state A_t . Specifically, the empirical finding suggests that, when the economy is in a recession (in our model, this refers to states where the realized aggregate profitability A_t is lower than its mean),

the MSE of machine learning forecasts is less unbiased compared to the good state. Right now, we do not impose any specific functional relationship between f_t and A_t . The only assumption we have is:

Assumption 1. The absolute value of f_t is non-decreasing in $A_t - \mu_A$, and is smaller when $A_t \leq \mu_A$ compared to when $A_t > \mu_A$.

Investors do not know about the potential biasedness of signals⁵ and therefore believe the private signal they receive is $\eta_{jt} = z_t + \xi_{jt}$, where ξ_{jt} follows a normal distribution with mean 0

$$\xi_{jt} \sim N(0, \rho_\eta^{-1}).$$

Apart from the private signals, investors also learn about z_t from the market price p_t , we assume and later on verify that observing this price is equivalent to observing a normally distributed signal $\eta_{pt} = z_t + e_t$ with $e_t \sim N(0, \rho_{pt}^{-1})$.

Then by Bayes' law, the perceived posterior distribution of z_t is

$$z_t | \eta_{jt}, \eta_{pt} \sim N(\mu_{jt}, \rho_t^{-1}), \quad (19)$$

where $\rho_t^{-1} \equiv (\rho_z + \rho_\eta + \rho_{pt})^{-1}$ and $\mu_{jt} \equiv E[z_t | \eta_{jt}, \eta_{pt}] = \rho_t^{-1}(\rho_z \mu_{zt} + \rho_\eta \eta_{jt} + \rho_{pt} \eta_{pt})$. However, if the investor knows the non-tech signal is biased, the correct μ_{jt} should be $\rho_t^{-1}(\rho_z \mu_{zt} + \rho_\eta(\eta_{jt} - \theta \beta \epsilon_{t-1} - f_t) + \rho_{pt} \eta_{pt})$. Due to this bias, after a positive shock ϵ_{t-1} , $1 - \lambda$ investors who have received non-tech signals overestimate the firm's profitability.

Investor portfolio choice After forming expectations about the firm's profitability, investors decide how to allocate their wealth between riskless and risky assets, taking prices as given.

$$\max_{q_{jt}} U_{jt} = \alpha E_{jt}[W_{jt} | \mathcal{F}_t] - \frac{\alpha^2}{2} V_{jt}[W_{jt} | \mathcal{F}_t], \quad (20)$$

$$s.t. \ W_{jt} = r_t(W_t - q_{jt}p_t) + q_{jt}A_t z_t, \quad (21)$$

⁵Alternatively, we can model an economy where investors know the biasedness of non-tech forecasts, but do not know the specific type of signals they receive. In this case, investors rationally expect they receive a tech signal with probability λ , and a non-tech signal with probability $1 - \lambda$. So from the investors' point of view, the signal follows a normal distribution with mean $(1 - \lambda)(\theta \delta \epsilon_{t-1} + f_t)$. Thus, when an investor receives a signal η_{jt} , she knows that $\eta_{jt} - (1 - \lambda)(\theta \delta \epsilon_{t-1} + f_t) \sim N(0, \rho_\eta^{-1})$. As will become clear in Equation 24, the investor demand for equity is linear in investors' posterior expectation z_t . This implies that the stock price derived from the market clearing condition will correct for the forecast bias. We choose our current setting because if agents in the economy know tech forecasts are biased, they can fix it directly instead of reporting a biased forecast. The empirical evidence indeed shows that the reported forecasts are biased.

where $\mathcal{F}_t = \{\eta_{jt}, p_t, \mathcal{F}_t^-\}$ is investors' information set.

Plug the budget constraint into investor's utility, the maximization problem transforms into

$$\max_{q_{jt}} U_{jt} = \alpha E_j [r_t W_t + q_{jt}(A_t z_t - p_t r_t) | \mathcal{F}_t] - \frac{\alpha^2 A_t^2}{2} q_{jt}^2 V_j(z_t | \mathcal{F}_t). \quad (22)$$

Asset market clearing condition The equity market clears:

$$\int_0^1 q_{jt} dj = x_t + \tilde{x}_t, \quad (23)$$

where \tilde{x}_t is noisy supply and is distributed normally $N(0, \rho_x^{-1})$. The introduction of noisy supply prevents fully revealing z_t with a continuum of traders.

6.2 Equilibrium

Equilibrium definition A Markov equilibrium is stock price $\{p_t\}$, firm equity issuance x_t , investor demand for equity $\{q_{jt}\}_j$ such that:

1. Each period t , given all the information realized before t , \mathcal{I}_{t-1} , and current aggregate profitability A_t , the firm owner chooses equity issuance amount x_t to maximize the firm's net revenue as in Equation 18.
2. Then, each investor $j \in [0, 1]$ receives a private signal about the firm's idiosyncratic profitability z_t , which is randomly drawn from a continuum of analysts (meaning that investor j does not know the type of signal she receives). In this economy, there are two types of analysts: λ tech analysts and $1 - \lambda$ non-tech analysts. The former produces unbiased signals $\eta_{T,jt}$, while the signal from the latter $\eta_{N,jt}$ suffers from overreaction and model misspecification error. Importantly, investor j does not know the biasedness of $\eta_{N,jt}$, and her perceived signal is denoted as η_{jt} .
3. Given all previous information and current aggregate profitability shock $\{\mathcal{I}_{t-1}, A_t\}$, a common prior about the distribution of z_t , a perceived private signal η_{jt} about the firm's profitability z_t and stock price p_t , investor j updates her perceived distribution of z_t according to Bayes' rule. Based on her updated posterior belief, she chooses the amount of equity to purchase, q_{jt} , to maximize her expected utility as in Equation 20.
4. Asset market clears: $\int_0^1 q_{jt} dj = x_t + \tilde{x}_t$.

5. Exogeneous states evolve according to the following processes:

$$\begin{aligned} \log A_t &= \delta_A \log A_{t-1} + \epsilon_{At} \\ z_t &= \delta z_{t-1} + \epsilon_{zt}. \end{aligned}$$

From the investors' optimization problem, we can derive investors' demand for equity from the first order condition of Equation 22:

$$q_{jt} = \frac{A_t \rho_t^{-1} (\rho_z \mu_{zt} + \rho_\eta \eta_{jt} + \rho_{pt} \eta_{pt}) - p_t r_t}{\alpha A_t^2 \rho_t^{-1}}. \quad (24)$$

The diagnostic expectation component $\theta \delta \epsilon_{t-1}$ is embedded in η_{jt} . Recall that all investors believe the private signals they receive are unbiased, so $\eta_{jt} = z_t + \xi_{jt}$ with $\xi_{jt} \sim N(0, \rho_\eta^{-1})$.

From the market clearing condition, we have

$$p_t r_t = A_t \rho_t^{-1} \rho_z \mu_{zt} + A_t \rho_t^{-1} \rho_\eta (1 - \lambda) (\theta \delta \epsilon_{t-1} + f_t) - \alpha A_t^2 \rho_t^{-1} x_t + A_t \rho_t^{-1} \rho_\eta (z_t - A_t \frac{\alpha}{\rho_\eta} x_t) + A_t \rho_t^{-1} \rho_{pt} (z_t + e_t).$$

Both sides should be $z_t + e_t$ measurable, therefore

$$e_t = -A_t \frac{\alpha}{\rho_\eta} \tilde{x}_t.$$

Thus,

$$p_t = \frac{A_t \rho_t^{-1}}{r_t} [\Gamma_t - \alpha A_t x_t + (\rho_\eta + \rho_{pt}) (z_t - A_t \frac{\alpha}{\rho_\eta} \tilde{x}_t)], \quad (25)$$

where $\Gamma_t \equiv \rho_z \mu_{zt} + \rho_\eta (1 - \lambda) (\theta \delta \epsilon_{t-1} + f_t)$, $\rho_{pt} = \frac{\rho_\eta^2}{\alpha^2 A_t^2} \rho_x$.

Equation 25 gives the demand function of the risky asset. Holding everything else constant, a rise in θ leads to a rise in the stock price p_t if the recent firm-level shock ϵ_{t-1} is positive. Intuitively, if investors overestimate the value of z_t following a positive shock, they are willing to pay more in order to get one share of stock.

From the first order condition of the firm's maximization problem, we can also obtain the supply function. The firm owner does not use any signals to update her belief on z_t , thus the first

order condition gives

$$\begin{aligned} E[p_t + x_t \frac{\partial p_t}{\partial x_t} | \mathcal{F}_t^-] - \phi(x_t - x_{t-1}) &= 0, \\ E[p_t | \mathcal{F}_t^-] - x_t \frac{\alpha A_t^2 \rho_t^{-1}}{r_t} - \phi(x_t - x_{t-1}) &= 0. \end{aligned}$$

Plug Equation 25 into the equation above. Given that the noisy supply has a mean of 0, we have

$$\frac{A_t \rho_t^{-1}}{r_t} [\Gamma_t - \alpha A_t x_t + (\rho_\eta + \rho_{pt}) \mu_{zt}] - x_t \frac{\alpha A_t^2 \rho_t^{-1}}{r_t} - \phi(x_t - x_{t-1}) = 0. \quad (26)$$

Equation 26 gives us the equilibrium value of x_t :

Proposition 6.1 *In equilibrium, the number of shares issued is*

$$x_t = \frac{\frac{A_t}{\rho_t r_t} [(\rho_z + \rho_\eta + \rho_p) \mu_{zt} + \rho_\eta (1 - \lambda) (\theta \delta \epsilon_{t-1} + f_t)] + \phi x_{t-1}}{2 \frac{\alpha A_t^2}{\rho_t r_t} + \phi}, \quad (27)$$

where $\rho = (\rho_z + \rho_\eta + \rho_p)$, $\rho_{pt} = \frac{\rho_\eta^2}{\alpha^2 A_t^2} \rho_x$.

Equation 27 gives the total shares of equity the firm issues in equilibrium. The first term in the numerator captures the effect of expected z_t which combines the information from the prior belief, private signals and stock price. The remaining terms in the numerator represent the marginal equity issuance costs (interior solution). The denominator captures the sensitivity of price to demand plus the convex equity issuance cost.

Corollary 6.2 *The equilibrium level of equity issuance exhibits less overreaction to recent shocks when the economy has more tech analysts, i.e. when λ increases.*

Proof of Corollary 6.2

The overreaction of equity issuance to the recent shock is

$$\frac{\partial x_t}{\partial \epsilon_{t-1}} = \frac{\frac{A_t}{\rho_t r_t} \rho_\eta (1 - \lambda) \theta \delta}{2 \frac{\alpha A_t^2}{\rho_t r_t} + \phi}.$$

The response of this overreaction to changes in λ is

$$\frac{\partial x_t}{\partial \epsilon_{t-1} \partial \lambda} = - \frac{\frac{A_t}{\rho_t r_t} \rho_\eta \theta \delta}{2 \frac{\alpha A_t^2}{\rho_t r_t} + \phi} < 0.$$

Therefore, the overreaction decreases as λ rises. \square

This result is expected given that tech analysts exhibit lower overreaction to recent idiosyncratic shocks. So if we transit from an economy with few tech analysts to an economy with more tech analysts, the equity issuance we observe would be more rational in the sense that it overreacts less to firm-level shocks.

Corollary 6.3 *In the special case where the equity issuance cost equals 0: i) Holding f_t fixed, x_t is decreasing in A_t ; ii) Holding A_t fixed, x_t is non-decreasing in f_t . The sign of $\frac{\partial x_t}{\partial A_t}$ is ambiguous once we account for the correlation between A_t and f_t .*

Proof of Corollary 6.3

If $\phi = 0$, then

$$x_t = \frac{(\rho_z + \rho_\eta + \frac{\rho_\eta^2}{\alpha^2 A_t^2} \rho_x) \mu_{zt} + \rho_\eta (1 - \lambda) (\theta \delta \epsilon_{t-1} + f_t)}{2\alpha A_t}.$$

Holding f_t constant, a rise in A_t leads to an increase in the denominator and a decline in the numerator, both of which contributes to a lower x_t given that $\frac{\rho_\eta^2}{\alpha^2 A_t^2} \rho_x$, x_t and αA_t are all non-negative⁶.

Holding A_t constant, $\frac{\partial x_t}{\partial f_t} = \frac{\rho_\eta (1 - \lambda)}{2\alpha A_t} \geq 0$.

In the simplified problem above where the equity issuance cost is 0, A_t affects the firm's equity issuance from two channels: the direct effects on the conditional mean and variance of the investment payoff and the indirect effect on the model misspecification error. The former tends to have a net negative impact on firm equity issuance, while the latter tends to have a positive effect because higher f_t translates into a higher expected payoff. The overall impact is ambiguous. In the extreme case where the response of f_t to A_t is negligible, x_t is decreasing in A_t , while in the extreme case where a small increase in A_t leads to a huge rise in f_t which approaches ∞ , x_t is increasing in A_t .

7 Conclusion

A key behavioral challenge to the rational model is the substantial evidence that a great many people overreact to news (Bordalo et al., 2022). We show that it is not just people that overreact. High

⁶ x_t is the total equity issued, not the change of equity issuance Δx_t . Our parameters will be set such that $x_t \geq 0$.

quality machine algorithms also overreact - albeit less strongly than do people. This raises the issue of the underlying source of overreaction. We find that it appears to be connects to the learning rate. By reducing the learning rate below the rate chosen by cross-validation, overreaction is reduced, as the algorithm produces predictions that less fully reflect new information altogether.

Analysts who depend on machine learning algorithms overreact much less than do more traditional analysts. Despite this, replacing traditional analysts with machine learning algorithms is not costless. Traditional stock analysts, despite their behavioral limitations do have information that is not otherwise readily available for use in the algorithms.

References

- Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson**, “Artificial intelligence, firm growth, and product innovation,” *Firm Growth, and Product Innovation* (November 9, 2021), 2021.
- Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp**, “Big data in finance and the growth of large firms,” *Journal of Monetary Economics*, 2018, 97, 71–87.
- Bianchi, Francesco, Sydney C Ludvigson, and Sai Ma**, “Belief distortions and macroeconomic fluctuations,” *American Economic Review*, 2022, 112 (7), 2269–2315.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer**, “Overreaction and Diagnostic Expectations in Macroeconomics,” *Journal of Economic Perspectives*, 2022, 36 (3), 223–44.
- , —, —, and **Stephen J Terry**, “Real credit cycles,” Technical Report, National Bureau of Economic Research, Paper 28416 2021.
- , —, **Spencer Yongwook Kwon, and Andrei Shleifer**, “Diagnostic bubbles,” *Journal of Financial Economics*, 2021, 141 (3), 1060–1077.
- , —, **Yueran Ma, and Andrei Shleifer**, “Overreaction in macroeconomic expectations,” *American Economic Review*, 2020, 110 (9), 2748–82.
- Cao, Sean, Wei Jiang, Junbo L Wang, and Baozhong Yang**, “From man vs. machine to man+ machine: The art and AI of stock analyses,” *National Bureau of Economic Research*, 2021.
- Chen, Qi, Itay Goldstein, and Wei Jiang**, “Price informativeness and investment sensitivity to stock price,” *The Review of Financial Studies*, 2007, 20 (3), 619–650.
- Chen, Tianqi and Carlos Guestrin**, “XGBoost: A scalable tree boosting system,” in “Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining” ACM 2016, pp. 785–794.
- Fama, Eugene F and James D MacBeth**, “Risk, return, and equilibrium: Empirical tests,” *Journal of Political Economy*, 1973, 81 (3), 607–636.
- and **Kenneth R French**, “Profitability, investment and average returns,” *Journal of Financial Economics*, 2006, 82 (3), 491–518.

Friedman, Jerome H, “Greedy function approximation: a gradient boosting machine,” *Annals of statistics*, 2001, pp. 1189–1232.

—, “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, 2002, 38 (4), 367–378.

van Binsbergen, Jules H, Xiao Han, and Alejandro Lopez-Lira, “Man vs. Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases,” *Review of Financial Studies*, 2022.

A Appendix

Table A1: Technical Majors

Transportation	Finance, Physics
Systems engineering	Finance, IT
Quantitative Econ, Financial Modeling	Finance and Information Systems
quantitative analysis	Environmental Engineering
Quantitative finance	Engineering
Quant econ course	Electrical, Ocean, Civil Engineering
Physics and Economics	Electrical Engineering, CS
Physics and applied math	Electronics and communication engineering
Physics	Electrical engineering
PhD Pharmaceutical sciences	Electrical Electronic Engineering
PhD in Pharmacology	Electrical Computer Engineering
PhD in neuroscience	Electrical and electronics engineering
PhD Biomedical/Medical Engineering	Data analyst course
Computer	Computer Science
Mechanical Engineering	Computer engineering
Medical Chemistry	Computer engineering
Mechanical Engineering	Computational biology, bioinformatics
Mathematics, Finance	Computational finance, information system
Mathematics, economics	Civil Engineering
Mathematics, actuarial studies	Civil engineering
Mathematics minor	Chemistry, environmental engineering
Mathematics and statistics	Chemistry engineering
Mathematics and physics	Chemistry and Applied mathematics
Mathematics	Chemistry
Mathematical science	Chemical Engineering
Mathematical economics analysis	Biophysics and biochemistry
Mathematical economics	biophysics
Mathematic, Physics	Bio-Organic Chemistry
Mathematic and economics	biomedical engineering
Math and Economics	BioMathematics
Materials Science	Biology and chemistry
Material metallurgical engineering	Biology and Biochemistry
Masters in Data Science	Biochemistry
Master of engineering	behavioral science
M.A. in Neuroscience	Information systems
Industrial Engineering	
Industrial and system engineering	
Industrial and management engineering	
general engineering	
Financial Modeling, Engineering	

Table A2: Predictable Forecast Errors

This table report GMM regression results of regressing forecast error at $t + 1$ on the information at time t . Forecast error at $t + 1$ is $\text{Forecast Error}_{t+1} := p_{t+1} - E_t p_{t+1}$, where E_t is forecast based on date t data. In columns (1), forecast $E_t p_{t+1}$ is the median earnings forecast of all analysts forecasts conducted during fiscal year t . In columns (3), forecast $E_t p_{t+1}$ is the median earnings forecast of all machine learning forecasts conducted during fiscal year t . Earnings are converted from EPS forecast to total earnings. Investment_t capital expenditure at time t . Debt issuance_t is net debt issuance at time t . All variables are scaled by total assets.

	(1) Forecast Error Analysts	(2) Forecast Error Machine
Investment	-0.472*** (0.003)	-0.345** (0.025)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Period	1994-2018	1994-2018
N	54536	53324

Table A3: Predictable Forecast Errors Using Large Learning Rates

This table presents the results of an OLS regression analysis of the forecast error at time $t + 1$ and Wald test results from seemingly unrelated regressions. The forecast error at $t + 1$ is defined as $\text{Forecast Error}_{t+1} = p_{t+1} - F_t p_{t+1}$, where $F_t p_{t+1}$ is the average of median monthly earnings forecast based on data from time t , and p_{t+1} represents the actual earnings at time $t + 1$. Forecasts are calculated from various machine learning models that use different hyper-parameters. The default hyper-parameters are set as the following values: depth of the tree is 2 and number of trees in the forest is 50. Table reports the results using different learning rates. Investment_t is capital expenditure at time t . The χ^2 of the Wald test is used to test the difference of regression coefficients from default hyper parameters, with the results of the test presented in the table. The p-value of the test is also included in brackets.

	(1) Forecast Error Analysts	(2) Forecast Error Analysts	(3) Forecast Error Machine	(4) Forecast Error Machine
Learning rate	0.15	0.25	0.3	0.5
Investment	-0.106*** (-7.766)	-0.107*** (-7.285)	-0.105*** (-7.262)	-0.099*** (-6.132)
χ^2	0.019 [0.890]	0.012 [0.911]	0.072 [0.788]	1.261 [0.261]
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Period	1994-2018	1994-2018	1994-2018	1994-2018
N	53321	53321	53321	53321
AdjR2	0.21	0.20	0.20	0.18

Table A4: Analysts Forecast Errors

This table report OLS regression regressing firm investment and debt on analysts forecast errors.

	(1) Inv _{t+1}	(2) Inv _{t+1}	(3) Debt Net Iss. _{t+1}	(4) Debt Iss. _{t+1}
Analysts Forecast Errors	-0.007*** (-2.668)	-0.007*** (-3.446)	-0.028*** (-6.466)	-0.032*** (-6.017)
Profit	0.035*** (10.683)	0.020*** (6.617)	-0.001 (-0.323)	0.017*** (4.174)
Firm FE	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes
Period	1994-2018	1994-2018	1994-2018	1994-2018
N	64218	62788	64218	62788
AdjR2	0.05	0.65	0.02	0.07