# Behavioral Machine Learning?

## Computer Predictions of Corporate Earnings also Overreact (WP, 2025)

Murray Z. Frank, Jing Gao, and Keer Yang

Presented by Kang Guo

June 27, 2025

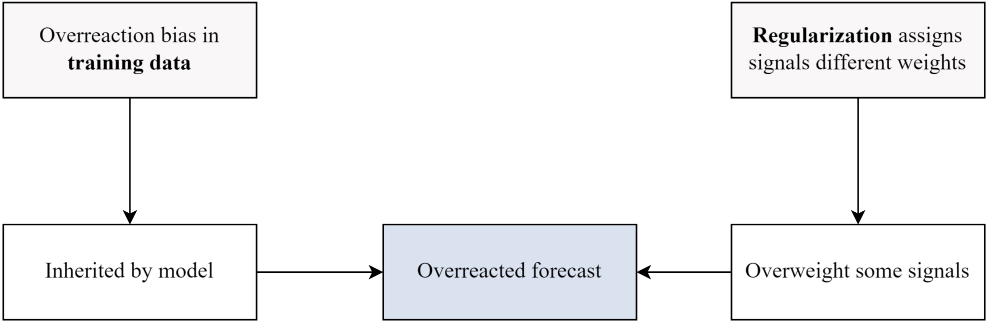
# Overview

# Background

- Algorithms being free from human psychological biases and emotions, should produce more accurate and rational predictions than humans
  - Inflation and GDP growth (Bianchi., 2022)
  - Corporate earnings forecasts (van Binsbergen et al., 2023)
  - Stock price forecasts (Cao et al., 2024)

- Better prediction performance $\neq$ predictions are truly rational
  - Why?
  - **Machine learning predictions may exhibit overreaction**

# Motivation


Overreaction bias in training data
Regularization assigns signals different weights
Inherited by model
Overreacted forecast
Overweight some signals

# Question

- Q1: Do ML forecasts show similar overreaction as human forecasts?
  - Yes
  - XGBoost, random forests, neural network, LLM and OLS
  - ML overreacts much **less** than humans
- Q2: Why do ML forecasts overreact?
  - Bias from the data
  - Bias from the model (regularization)
  - **Bias primarily form the data**
- Q3: Can ML mitigate the overreaction bias observed in human analysts?
  - Analysts using machine learning overreact less than traditional analysts
  - But there is a **tradeoff** between overreaction and forecast accuracy

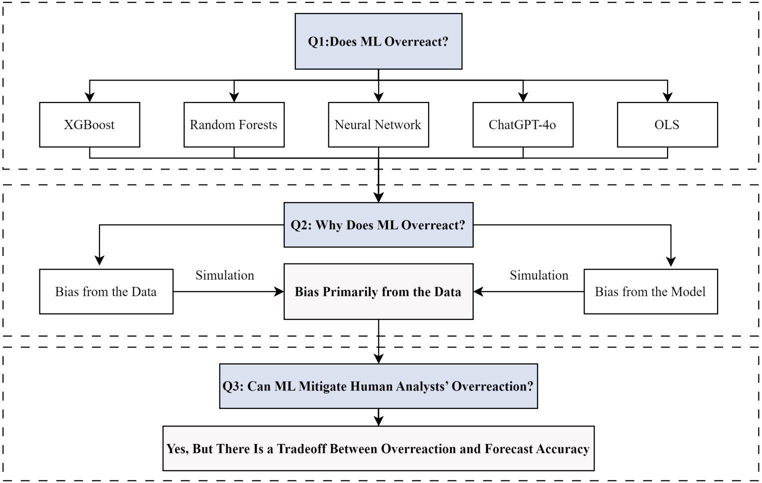
# Contribution

- Belief bias in financial markets
  - Existing: Leverage ML as a benchmark to identify biases in human forecasts
  - **Extension: Investigate biases in ML predictions themselves**

- Measure rational expectations
  - Existing: ML prediction can serve as a proxy for an unbiased expectation
  - **Extension: ML predictions exhibit overreaction like humans**

- Superior accuracy of ML prediction
  - Existing: ML can capture complex nonlinear relationships
  - **Extension: ML learns the irrational behavior of humans**

# Framework


Q1:Does ML Overreact?
XGBoost
Random Forests
Neural Network
ChatGPT-4o
OLS
Q2: Why Does ML Overreact?
Bias from the Data
Simulation
Bias Primarily from the Data
Simulation
Bias from the Model
Q3: Can ML Mitigate Human Analysts' Overreaction?
Yes, But There Is a Tradeoff Between Overreaction and Forecast Accuracy

# Data

- Firm earnings forecasts (EPS) from IBES
- Firm financial data from Compustat
- Stock return data from CRSP
- Macro time series from the Federal Reserve Bank of Philadelphia
- Analyst background information from LinkedIn and FINRA

- Sample: 1994–2018

# Testing for Overreaction

- Following Bordalo et al. (2021), this paper runs the following regression:

$$y_{i,t+1} - F_{i,t+1} = \beta x_{i,t} + \lambda_t + \mu_i + \varepsilon_{i,t} \tag{1}$$

- Dependent variable: next year's forecast errors
  - $y_{i,t+1}$: realized earnings (EPS)
  - $F_{i,t+1}$: predicted earnings (EPS)
- $x_{i,t}$: current-year firm investments (capital expenditure)
- Investments are positively correlated with expected earnings
- **Overreactive expectations** $\Rightarrow \beta < 0$

# Q1: Does ML Overreact? — Regularized Model

- XGBoost forecast errors are **negatively** correlated with investment
- XGBoost overreacts much **less** than human analysts

| | (1)<br>Forecast Error<br>Analysts | (2)<br>Forecast Error<br>Analysts | (3)<br>Forecast Error<br>Machine | (4)<br>Forecast Error<br>Machine |
|---|---|---|---|---|
| Investment | -0.018*<br>(-1.836) | -0.1422***<br>(-10.082) | -0.018**<br>(-2.321) | -0.107***<br>(-8.254) |
| Firm FE | No | Yes | No | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Period | 1994-2018 | 1994-2018 | 1994-2018 | 1994-2018 |
| N | 54541 | 53329 | 54541 | 53329 |
| AdjR2 | 0.02 | 0.32 | 0.03 | 0.22 |

# Q1: Does ML Overreact? — LLM

- ChatGPT forecast errors are **negatively** correlated with investment
- ChatGPT overreacts much **less** than human analysts

| | (1) Forecast Error ChatGPT | (2) Forecast Error Analyst | (3) Forecast Error XGBoost |
|---|---|---|---|
| Investment | $-0.110^{***}$ | $-0.164^{***}$ | $-0.123^{***}$ |
| | (7.165) | (10.173) | (7.846) |
| Firm FE | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| Period | 1994-2018 | 1994-2018 | 1994-2018 |
| N | 28002 | 28002 | 28002 |
| AdjR2 | 0.33 | 0.33 | 0.27 |
| Forecast MSE | 1.350 | 1.478 | 1.436 |

# Q1: Does ML Overreact? — OLS

- OLS forecast errors are **negatively** correlated with investment

| | (1) Forecast Error OLS | (2) Forecast Error OLS | (3) Forecast Error OLS |
|---|---|---|---|
| Investment | -0.112*** | -0.084*** | -0.102*** |
| | (-5.146) | (-3.962) | (-5.190) |
| Firm FE | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| Forecasting Variables: | | | |
| Firm Char | Yes | Yes | Yes |
| Analysts Forecasts | No | Yes | Yes |
| Financial Statement Items | No | No | Yes |
| Period | 1994–2018 | 1994–2018 | 1994–2018 |
| N | 53529 | 53529 | 53529 |
| AdjR2 | 0.44 | 0.25 | 0.26 |

# Q2: Why Does ML Overreact? — Hypotheses

- Biases in ML predictions: **training data and the algorithm itself**
- **H1: The overreaction of OLS stems solely from biased training data**
  - OLS incorporates all observations equally without regularization
  - OLS faithfully learns patterns in the data without assigning differential weights to noise or bias
- **H2: The overreaction of regularized models stems from biases in both the data and the model itself (regularization)**
  - Regularization techniques assign more weights to some signals

# Q2: Why Does ML Overreact? — Simulation

- 200 simulated panels
- **It's hard to support the first hypothesis and the claim made in the abstract**

| | $\beta$ | Mean($\|fe\|$) | %significant | %negative (conditional) |
|---|---|---|---|---|
| Panel A: 33.3% overreaction | | | | |
| XGBoost | -0.001 | 16.13 | 17% | 58% |
| OLS | -0.113 | 62.05 | 72% | 47% |
| Panel B: No overreaction | | | | |
| XGBoost | -0.001 | 16.14 | 18% | 51% |
| OLS | -0.116 | 62.99 | 71% | 48% |

# Q3: Can ML mitigate human analysts' overreaction? — Yes

- **H3: Analysts with statistical training tend to use machine learning**
  - $\Rightarrow$ Identify their technical background by their educational qualifications
- Scikit-learn became popular in 2012

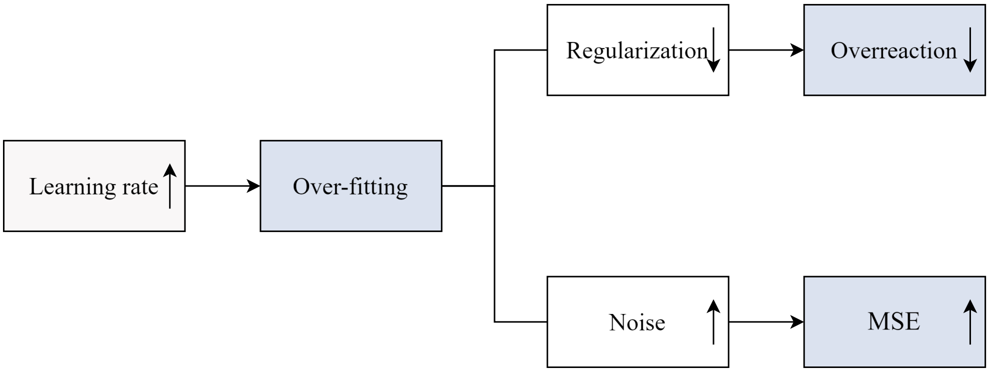
| | (1)<br>Forecast Error<br>Tech | (2)<br>Forecast Error<br>Non-Tech | (3)<br>Forecast Error<br>Tech | (4)<br>Forecast Error<br>Non-Tech |
|---|---|---|---|---|
| Investment | -0.181*** | -0.127*** | -0.109 | -0.090** |
| | (2.736) | (6.922) | (1.426) | (2.864) |
| Firm FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Period | 1994-2012 | 1994-2012 | 2013-2018 | 2013-2018 |
| N | 7367 | 21309 | 6358 | 12316 |
| AdjR2 | 0.25 | 0.27 | 0.25 | 0.31 |

# Q3: Can ML mitigate analysts' overreaction? — Cost

- **Higher learning rate reduces overreaction, but compromises predictive accuracy**

| | (1)<br>Forecast Error<br>Neural Network | (2)<br>Forecast Error<br>Neural Network | (3)<br>Forecast Error<br>Neural Network |
|---|---|---|---|
| Investment | -0.136*** | -0.010 | 0.051 |
| | (-6.981) | (-0.308) | (1.056) |
| Firm FE | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| Period | 1994-2018 | 1994-2018 | 1994-2018 |
| N | 53329 | 53329 | 53329 |
| AdjR2 | 0.25 | 0.36 | 0.72 |
| learning rate | 0.1 | 0.2 | 0.5 |
| Forecast MSE | 1.648 | 2.504 | 3.932 |

# Q3: Why is there a tradeoff?



# Conclusion

- Machine learning predictions exhibit overreaction
- Overreaction bias primarily comes from the training data
- ML can mitigate the overreaction bias observed in human analysts, but there is a **tradeoff** between **overreaction** and **forecast accuracy**

# Extension

- Other behavioral biases
  - Underreaction
  - Overconfidence
  - Herding

- Are forecast errors reflected in prices?
  - Higher forecast error may be correlated with idiosyncratic risk

- Does behavioral similarity to humans enhance model prediction accuracy?
  - Similarity to humans: forecast errors between machines and human analysts