Summary of "Man versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases"

Jules H. van Binsbergen, Xiao Han, Alejandro Lopez-Lira(RFS,2022)

Guowei Zhao summarised on May. 2024

1 What are the research questions?

- Whether the analyst forecasts are conditionally biased?
- The effects of analyst forecast biases:
  - Can/How the expectations error affect stock market returns?
  - What is the relationship between market anomalies and conditional bias?
  - How the new biases affect managers' actions?

2 Why are the research questions interesting?

- Commonly used forecast proxies behave poorly.
  - Analysts' forecasts are biased(Kothari et al.,2016)
  - Linear forecast predict poorly out-of-sample(So,2013)
- There is no real-time and unbiased forecast as benchmark for analyst forecast biases
  - Realized earnings as benchmark.
  - linear forecasts as benchmark(So,2013)
- This paper construct a unbiased and real-time earnings expectations with random forest regression, which support for nonlinearity and high-dimensional data

3 What is the paper's contribution?

- First to use ML to create an optimal and unbiased proxy for firms earnings'conditional expectations.
  - A real-time benchmark for analyst forecast bias(So,2013)
- Contributions to relevant literature.
  - Relationship between anomalies and conditional biases(Engelberg et al.,2018)
  - Providing direct evidence that firms exploit overpricing by issuing stocks(Hirshleifer et al,2010)
  - Analysts are skillful and exert effort(Grennan and Michaely,2020)
  - Demonstrate the existence of systematic biases in analysts' earnings forecasts(Bianchi et al.2022)

4 What hypotheses are tested in the paper?

- Hypothesis 1: Stocks with more optimistic earnings forecasts tend to earn lower future returns. This suggests that overly optimistic earnings expectations are associated with stock overvaluation, leading to lower subsequent returns.
- Hypothesis 2: Firms with more optimistic analysts' forecasts, relative to a statistically optimal benchmark, are more likely to issue additional equity in the subsequent periods. This implies that managers exploit optimistic market expectations to time stock issuances.

5 Sample

- Monthly for January 1986 to December 2019.
- firm-specific variables:
  - Realized earnings from the last period from /I/B/E/S.
  - Monthly stock prices and returns from CRSP.
  - 67 financial ratios obtained from WRDS.
- macroeconomic variables from the Federal Reserve Bank of Philadelphia.
  - Consumption growth, GDP growth, Growth of industrial production.
  - Unemployment rate.
- AF from I/B/E/S database(five horizons)

6 Dependent and independent variables

- The explained variable (dependent variable) is the future earnings per share (EPS) of firms over different time horizons, such as one quarter, two quarters, and one to two years ahead.
- Past earnings per share (EPS): This represents historical earnings data and is a key predictor.
- Analysts' earnings forecasts: These are publicly available forecasts from financial analysts.
- Firm fundamentals: These include various firm-specific financial indicators, such as realized earnings, stock prices, and returns .

7 Regression/prediction model specification

- The core regression model in the paper is based on random forest regression, a nonparametric machine learning approach. The equation is presented as:

$$E_t[\mathrm{EPS}_{i,t+\tau}] = RF[Fundamentals_{i,t}, \mathrm{Macro}_t, \mathrm{AF}_{i,t}]$$

8 What difficulties arise in drawing inferences from the empirical work?

- Spurious In-sample Linear Predictability: Traditional linear models may show in-sample predictability that does not generalize well out-of-sample. This is particularly problematic when the predictive power of the variables diminishes after a certain period, such as the 2000s. The paper finds that linear models' return predictability significantly decreases out-of-sample compared to machine learning models .
- Bias in Analysts' Forecasts: There is a conditional bias in analysts' forecasts, which complicates inference because these forecasts may deviate from the statistically optimal benchmarks. Furthermore, return predictability could be driven by the bias between analysts' forecasts and linear models, which may not hold across all time periods .
- Nonlinear Relationships: Many traditional models assume linear relationships between variables, which can lead to incorrect inferences. The paper points out that nonlinear models (like random forests) perform better in capturing complex relationships, but they also add complexity to the inference process .

9 Describe at least one publishable and feasible extension of this research.

- What else can we do with this new analyst bias indicator?
  - Examine the relationship between analyst forecast bias and institutional investor attention/shareholding in equities, market sentiment, industry, analysts' analytical style/social connections/educational background, etc.
- What other biases can we build with machine learning?
  - Expectations and biases in macroeconomic variables.