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

Jules van Binsbergen[1] Xiao Han[2] Alejandro Lopez-Lira[3]

[1] University of Pennsylvania, NBER and CEPR

[2]University of Edinburgh

[3]BI Norwegian Business School

December 2020

Motivation

- Are analysts' earnings forecasts statistically optimal?
- Can Machine Learning beat analyst forecasts?
- Are analyst forecasts biased? If so, can we measure the bias in real-time?
- Do analyst forecasts matter for market participants and firm managers?

Earnings Expectations and Biases

- Statistically optimal earnings expectations are important and interesting in their own right.
- In addition, we find expectations and biases impact stock returns and corporate behavior.
- Without a benchmark, we cannot measure cross-sectional and time series variation in forecast biases.

What are the current benchmarks in the literature?

- Ex Post Realizations
  - Does not allow for real-time measurement of conditional biases, because of a lack of a real-time benchmark (Kozak et al. (2018) and Engelberg et al. (2018)).
- No Benchmark
  - Use analysts' long-term earnings growth forecasts as an explanatory variable in empirical studies, without comparing it to a benchmark (La Porta (1996) and Bordalo et al. (2019)).
- Linear Regressions (Fama and French (2006), Hou et al. (2012), So (2013))
  - Linear forecasts are not necessarily optimal.
  - Analysts' forecasts are better than the linear forecasts.
  - Only the variables that have good predictive power are used in the regressions, incurring data leakage problems.

Our Contribution

- We use machine learning to construct a statistically optimal and unbiased benchmark for firms' earnings expectations available in real-time.

- We use random forest regression for our main analysis, which has two significant advantages:
  1. It naturally allows for nonlinear relationships between variables.
  2. It is designed for high-dimensional data which mitigates the risk of in-sample overfitting.

- We study the impact of the real-time expectation biases on stock market returns and corporate financing decisions.

Main Findings Summary

- Analyst earnings expectations exhibit conditional bias: it varies substantially across time and firms. On average they are biased upwards.
- The bias increases in the forecast horizon (on average) and analysts revise their expectations downwards as earnings announcement dates approach.
- Analysts' biases are associated with negative cross-sectional return predictability (profitable trading strategies).
- Managers of firms with the largest upward biases in earnings expectations seem aware of this bias and respond by issuing stocks.

Analysts' Forecasts

- Analysts forecast earnings per share (EPS).
- The forecast horizons are 1Q, 2Q, 3Q, 1Y, and 2Y ahead.
- Analysts forecast up until the announcement date (2-3 months after earnings are realized).

Supervised Machine Learning

- Prediction Machine: Receives public information available at the time and returns a forecast.
- The best forecast available at every period is the conditional expectation.
- Machine Learning is a novel, flexible and robust way to approximate conditional expectations.

Supervised Machine Learning

- Non-linear function of information available at time t.
- Fit the non-linear function using the information up until time t.
- Predicts the value at of earnings at time t+1 (out-of-sample).
- Similar structure as rolling regressions.

Supervised Machine Learning: Random Forest Regression

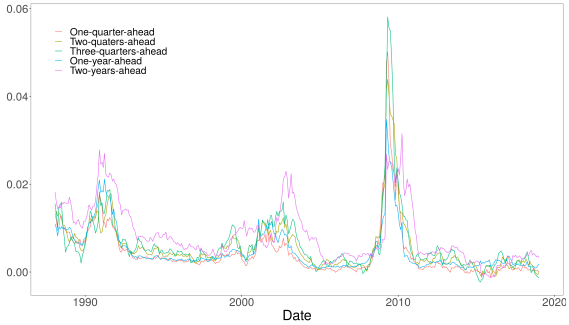
- Flexible, non-parametric, and robust to over-fitting.
- Choose parameters in a data-driven way (cross-validation) before the forecasting period.
- Train using rolling windows.

Data used for forecasting

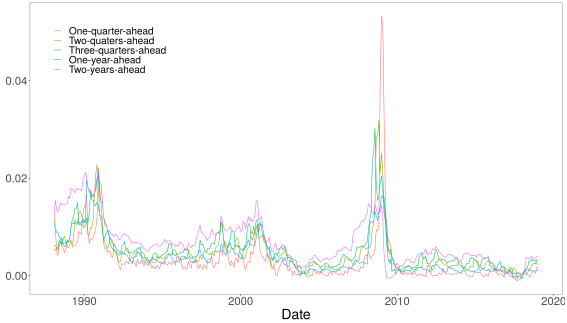
- I/B/E/S database: Analysts' Forecasts and past realized earnings.
- Stock prices and returns, CRSP.
- Firm fundamentals, Compustat.
- Real-time macroeconomic data provided by the Federal Reserve Bank of Philadelphia: Consumption, GDP, and Industrial Production Growth; and Unemployment.

Analysts' Bias

Conditional Bias (ML)



Realized Bias



The Term Structure of Earnings Forecasts via Machine Learning

- Machine Learning (ML) vs Actual Earnings (AE) vs Analyst Forecasts (AF)

Horizon	(AF - AE)
One-quarter-ahead	0.018
t-statistic	3.385
Two-quarters-ahead	0.044
t-statistic	5.107
Three-quarters-ahead	0.061
t-statistic	4.748
One-year-ahead	0.135
t-statistic	4.189
Two-years-ahead	0.348
t-statistic	4.501

The Term Structure of Earnings Forecasts via Machine Learning

- Machine Learning (ML) vs Actual Earnings (AE) vs Analyst Forecasts (AF)

Horizon	(AF - AE)	(ML - AE)
One-quarter-ahead	0.018	-0.008
t-statistic	3.385	-0.997
Two-quarters-ahead	0.044	-0.002
t-statistic	5.107	-0.155
Three-quarters-ahead	0.061	-0.001
t-statistic	4.748	-0.006
One-year-ahead	0.135	0.016
t-statistic	4.189	0.531
Two-years-ahead	0.348	-0.022
t-statistic	4.501	-0.195

The Term Structure of Earnings Forecasts via Machine Learning

- Machine Learning (ML) vs Actual Earnings (AE) vs Analyst Forecasts (AF)

Horizon	$(ML - AE)^2$	$(AF - AE)^2$
One-quarter-ahead	0.061	0.065
Two-quarters-ahead	0.080	0.089
Three-quarters-ahead	0.096	0.111
One-year-ahead	0.687	0.695
Two-years-ahead	1.329	1.699

The Term Structure of Earnings Forecasts via Machine Learning

- Machine Learning (ML) vs Actual Earnings (AE) vs Analyst Forecasts (AF)

Horizon	(ML - AE)/P	(AF - AE)/P	(AF - ML)/P
One-quarter-ahead	0.000	0.006	0.006
t-statistic	0.358	3.594	2.882
Two-quarters-ahead	-0.001	0.006	0.007
t-statistic	-0.226	4.530	3.516
Three-quarters-ahead	-0.001	0.006	0.008
t-statistic	-0.824	6.562	0.774
One-year-ahead	0.003	0.028	0.025
t-statistic	0.632	3.916	3.894
Two-years-ahead	-0.007	0.032	0.040
t-statistic	-0.703	8.106	6.224

Analysts are biased relative to the ML forecast

- Does the market price in this bias?
- Return predictability suggests at least not fully.
- Firm managers do seem to understand forecasts (and market prices) are biased: they are more likely to issue stock.
- Firm managers have all the same public information (including analysts' expectations), but also private information about their firm.

Return Predictability

- Expectations about cash flows are an important input to compute stock fundamental value.
- Analysts revise their expectations downwards as earnings announcement dates approach.
- If expectations are biased, and these biases are not priced in, we should see return predictability.
- Two primary return predictors.
  1. Average Bias: the average of the conditional biases across the multiple horizons.
  2. Bias Score: the arithmetic average of the percentile rankings on each of the five conditional bias measures.
- We find negative return predictability: Indirect evidence that the market pays attention to the analysts' forecasts but does not correct for the bias.

Portfolios Sorted on Conditional Bias

Quintile	1	2	3	4	5	1-5
Panel A: Average Bias						
Mean	1.07	0.70	0.46	-0.04	-0.88	1.95
t-stat	5.03	3.17	1.82	-0.12	-2.05	5.88
CAPM Beta	0.92	0.98	1.11	1.28	1.58	-0.66
Panel B: Bias Score						
Mean	0.96	0.66	0.43	0.07	-0.57	1.53
t-stat	4.76	2.93	1.64	0.22	-1.38	4.90
CAPM Beta	0.89	1.01	1.14	1.28	1.53	-0.63

Time Series Tests with common Asset-Pricing Models

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

	CAPM		FF3		FF5	
	Coeffi	*t*-stat	*Coeffi*	*t*-stat	*Coeffi*	*t*-stat
	Panel A: Average Bias					
Intercept	2.39	8.15	2.52	9.70	2.02	7.21
Mkt_RF	-0.66	-7.81	-0.61	-7.52	-0.42	-5.34
SMB			-0.86	-6.33	-0.62	-4.33
HML			-0.60	-4.10	-1.01	-6.10
RMW					0.84	4.07
CMA					0.53	1.79

Stock-Issuance

- Managers have at least as much information as the investors: public + private signals.
- Managers issue more stock when upward biases are larger.
- Not causal evidence. Other explanations?

Net Stock Issuances and Conditional Biases

Panel A: Net Stock Issuances of Portfolios formed on Biased Expectations						
Quintile	1	2	3	4	5	5-1
Average Bias	0.013	0.011	0.017	0.040	0.073	0.060
t-stat	1.82	1.82	3.33	4.31	5.32	3.44
Bias Score	0.009	0.016	0.020	0.033	0.066	0.058
t-stat	1.33	2.14	3.69	5.17	4.18	3.39

Net Stock Issuances and Conditional Biases

$$NSI_{i,t+1} = \alpha + \beta_1 Bias_{i,t} + \gamma_i \sum_{i=1}^{3} Control_{i,t} + \epsilon_{i,t+1}$$

Panel B: Fama-MacBeth regressions				
	A: Average Bias		B: Bias Score	
	(1)	(2)	(1)	(2)
Bias	1.7048	1.2870	0.1191	0.0510
t-stat	3.86	4.53	6.74	4.82
Controls	No	Yes	No	Yes
R^2	0.0178	0.0921	0.0084	0.0750

The control variables include the log of firm size, the log of book-to-market ratio, and earnings before interest, taxes, and depreciation divided by total assets.

Conclusion

- We provide a novel real-time benchmark for earnings expectations.
- Analyst forecasts are biased, and this bias exhibits large time series and cross-sectional variation.
- There is significant return predictability associated with biased expectations.
- Managers issue more stock when analysts are too optimistic.