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

Jules H. van Binsbergen, Xiao Han, and Alejandro Lopez-Lira
Working paper
2020 09

王念硕

Outline

1. Introduction
   - Background
   - Motivation
   - Research question
   - Related researches
   - Research contents
   - Contribution
   - Framework
2. Research design
   - Variable
   - Data
   - Method
3. Empirical result
   - Analyst forecast error
   - Bias and stock returns
   - Bias and Firm’s Decisions
4. Conclusion

1. Introduction

Background

1. One necessary input for pricing a risky asset is an estimate of expected future earnings to which the asset owner would be entitled. Common proxies include the most recent realized earnings, simple linear forecasts, or analysts' forecasts.
2. A significant strain of literature documents that forecasts can be biased or predict poorly out-of-sample, thereby limiting their practical usefulness.

1. Introduction

Motivation

1. Although previous research has used realized earnings to evaluate the bias and efficiency of analyst forecasts, the extant studies do not provide an unbiased real-time earnings estimates. Without such forecasts, it is difficult to assess and/or correct for the dynamics of forecast biases before the actual value is realized.

1. Introduction

Research question

1. Are machine learning methods more accurate than analysts?

   Yes

2. Can stock returns be predicted by the difference between the analysts' forecasts and the machine learning forecasts?

   Yes

3. Does this difference affect the behavior of executives?

   Yes

1. Introduction

Research Contents

1. Analysts make over-optimistic forecasts and revise their expectations downwards as the earnings announcement day approaches.

2. Machine-learning forecasts are closer to realized earnings than are analysts' forecasts.

3. Stocks with more optimistic earning forecasts earn lower future returns.

4. Firms with more optimistic analysts' forecasts relative to the statistically optimal benchmark issue more stocks.

1. Introduction

Related researches

1. Using a linear regression framework with some variables that have been shown to provide effective forecasting power (Fama and French (2006), Hou et al. (2012)).
2. So (2013) provides a linear forecast, and studies the predictable components of analysts' errors and their impact on asset prices.
3. Hirshleifer and Jiang (2010) and Baker and Wurgler (2013) who argue that managers can take advantage of overpricing on their firms' valuation by issuing stocks.
4. Random forest regression forecast is superior to a forecast following from any individual one predictor (Breiman 2001).

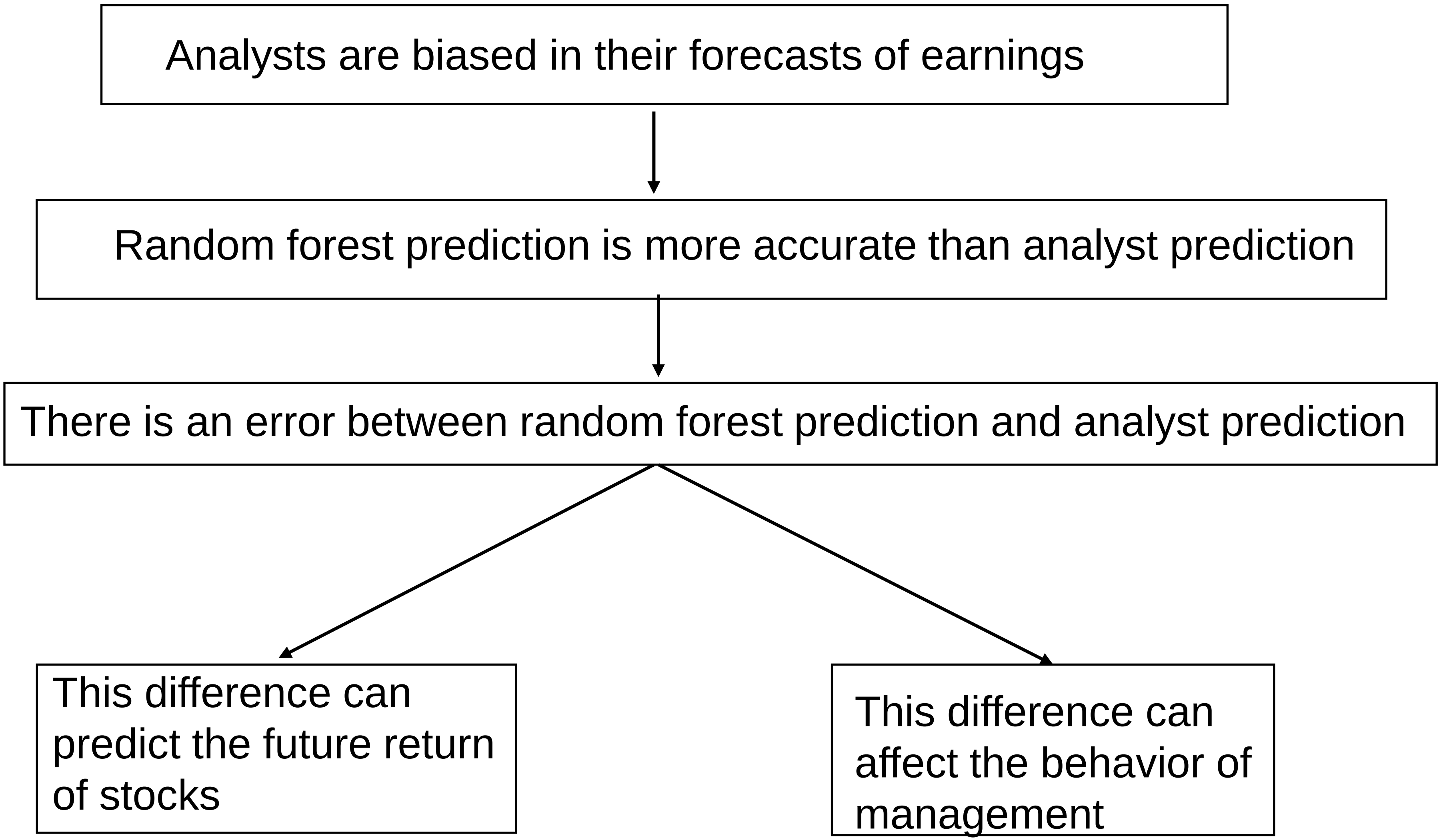
1. Introduction

Contribution

1. We propose a novel approach for constructing a statistically optimal and unbiased benchmark for earnings expectations, which uses machine learning and we demonstrate that our new benchmark is effective out-of-sample.

1. Introduction

Framework



2. Research design: Variable

Explained variable:

Future eps.

Explanatory variables:

Firm-specific variables:

Realized earnings, Earnings growth, Sales growth, Asset growth, Investment growth, stock prices and returns, Sixty-seven financial ratios from the Financial Ratios Suit by Wharton Research Data Services.

Macroeconomic variables:

Consumption growth, GDP growth, Growth of industrial production, Unemployment rate.

Analysts' earnings forecasts:

Analysts' forecasts eps.

2. Research design: Data

Data Source： Wharton Research Data Services、CRSP、 Compustat 、 I/B/E/S、 the Federal Reserve Bank of Philadelphia

Period： 1986 to 2019 monthly data.

Sample: All New York Stock Exchange (NYSE), American Stock Exchange (Amex), and Nasdaq. We keep firms that have both realized earnings and analysts forecasts.

2. Research design: Method

Random Forest

At each step, the algorithm splits the data choosing the variable and threshold that best minimizes the mean squared error when the average value of the variable to be forecasted is used as the prediction.

$$\min \left[\min_{c_1} \sum_{x_1 \in R_1} (y_1 - c_1)^2 + \min_{c_2} \sum_{x_2 \in R_2} (y_2 - c_2)^2\right] \qquad (1)$$

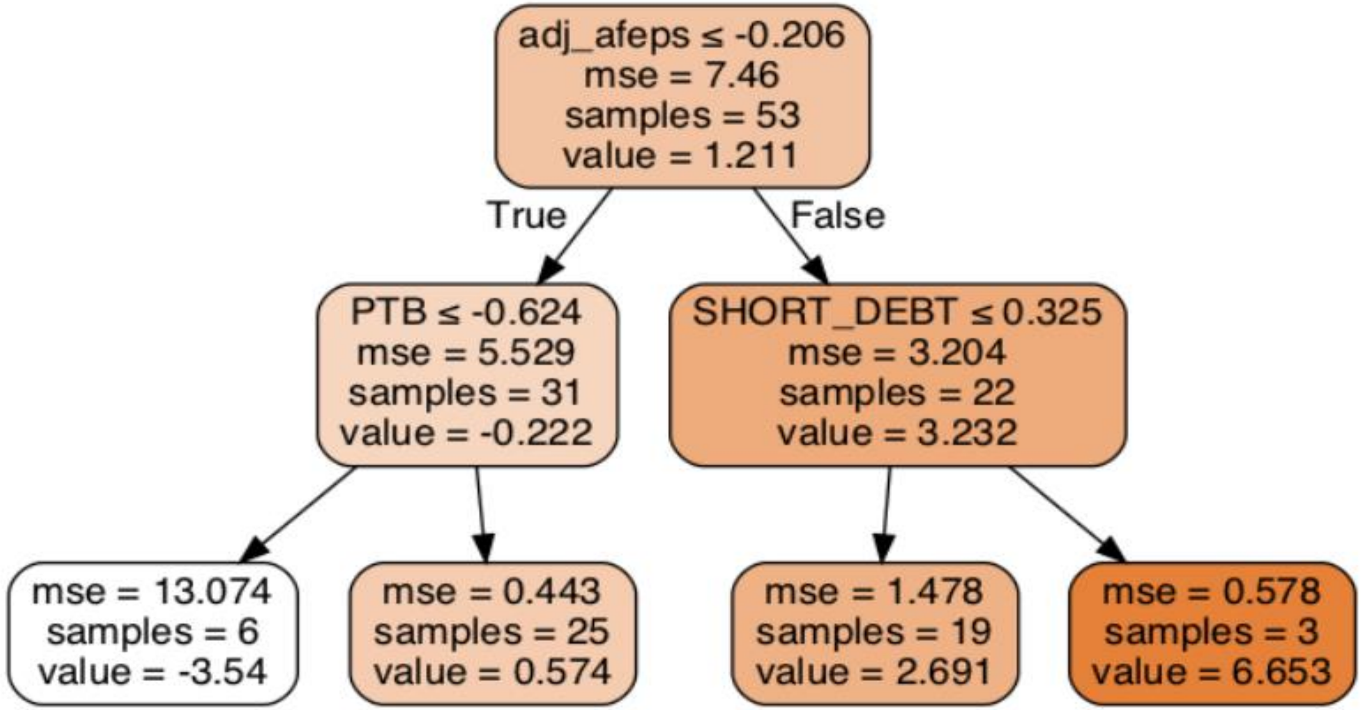
At each leaves, the forecast is equal to the average for that region

$$c_m = \frac{1}{N_m} \sum_{\{y_i : x_i \in R_m\}} y_i \qquad (2)$$

$$\hat{y} = \sum_m c_m I_{\{x_i \in R_m\}} \qquad (3)$$

2. Research design: Method

Figure 5: Example Decision Tree



In the final nodes, the prediction is the historical local average of that subspace.

2. Research design: Method

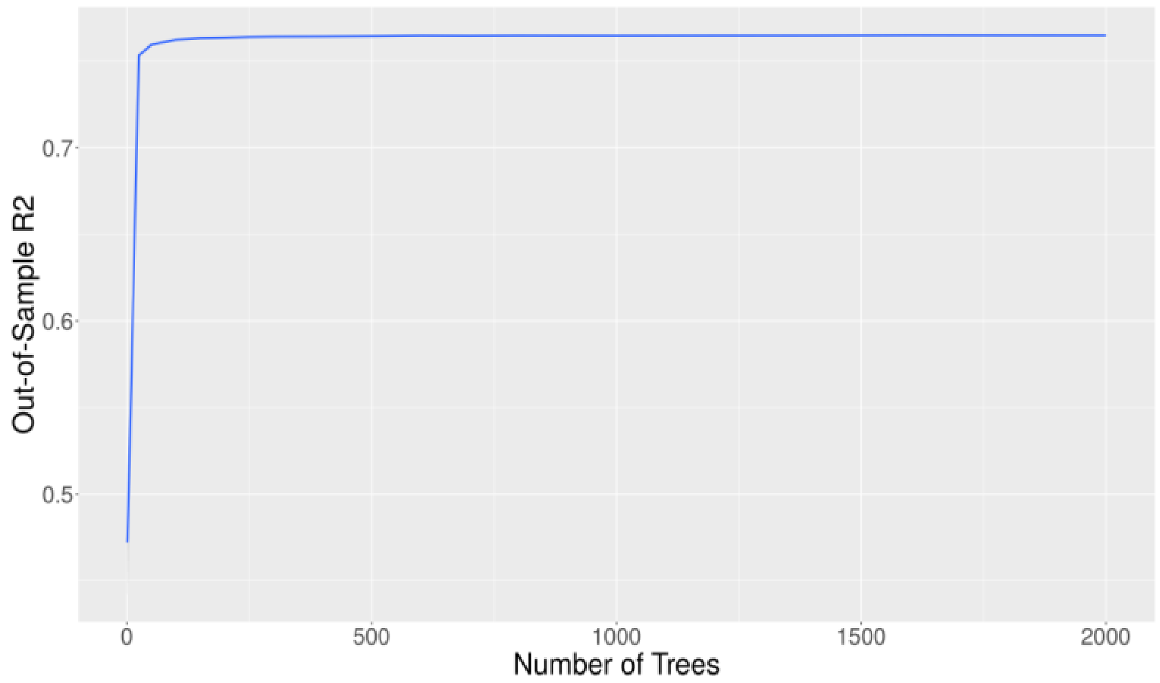
There are three main parameters in the random forest algorithm: (1) the number of decision trees; (2) the depth of the decision trees; and (3) the fraction of the sample that is taken in each split.

$$R^2_{oos} = 1 - \frac{\sum (mlf_i - eps_i)^2}{\sum (eps_i - \overline{eps})^2} \qquad (4)$$

MLF and EPS denote the machine learning forecast and actual realized earnings respectively. The denominator is constant across different specifications.

2. Research design: Method

Figure 7: Cross-validation Results of the number of trees in the one-year-ahead forecast



Number of Trees 2000
Maximum Depth 7
Sample Fraction 1%

2. Research design: Method

We measure the biases in investors' expectations as the differences between the analysts' forecast and the machine learning forecast:

$$BiasedExpectation_{i,t}^{t+h} = \frac{af_{i,t}^{t+h} - mlf_{i,t}^{t+h}}{price_{i,t-1}} \qquad (5)$$

in which subscript i denotes firm, and t denotes the date when earnings forecasts are made.
The superscript t+h denotes forecasting periods, and ML denotes machine learning.

3.1 Empirical result: Analyst forecast error

Panel D: One-year-ahead												
Month-ahead	10	9	8	7	6	5	4	3	2	1	0	-1
N	64026	97403	112926	115749	117769	119325	120842	122121	123511	124963	126707	120647
FE	0.207	0.216	0.209	0.201	0.185	0.166	0.147	0.130	0.107	0.085	0.070	0.067
t-statistic	5.36	6.60	6.92	7.03	6.94	6.63	6.40	6.06	5.56	5.42	5.29	5.15
Sqr_FE	1.220	1.166	1.090	0.970	0.886	0.816	0.688	0.629	0.566	0.465	0.401	0.402

The average forecast error is the difference between analysts' earnings forecasts per share and the realized earnings per share.
The average forecast error is consistently positive for all horizons, which suggests that analysts make over-optimistic forecasts.
The average error decreases as the earnings announcement dates approach.

3.1 Empirical result: Analyst VS Machine Learning

Horizon	ML	AF	AE	(ML - AE)	(AF - AE)	(AF - ML)	$(ML-AE)^2$	$(AF-AE)^2$	(ML - AE)/P	(AF - AE)/P	(AF - ML)/P
One-quarter-ahead	0.291	0.312	0.294	-0.008	0.018	0.021	0.061	0.065	0.000	0.006	0.006
t-statistic				-0.997	3.385	2.753			0.358	3.594	2.882
Two-quarters-ahead	0.305	0.351	0.307	-0.002	0.044	0.045	0.080	0.089	-0.001	0.006	0.007
t-statistic				-0.155	5.107	4.171			-0.226	4.530	3.516
Three-quarters-ahead	0.323	0.384	0.324	-0.001	0.061	0.061	0.096	0.111	-0.001	0.006	0.008
t-statistic				-0.006	4.748	4.624			-0.824	6.562	0.774
One-year-ahead	1.172	1.291	1.156	0.016	0.135	0.119	0.687	0.695	0.003	0.028	0.025
t-statistic				0.531	4.189	4.469			0.632	3.916	3.894
Two-years-ahead	1.329	1.699	1.351	-0.022	0.348	0.370	1.555	1.836	-0.007	0.032	0.040
t-statistic				-0.195	4.501	3.856			-0.703	8.106	6.224

The mean squared errors of the machine-learning forecast are smaller than the analysts' mean squared errors, demonstrating that our forecasts are, as expected, more precise than are the analysts' forecasts.
Overall, our results are consistent regarding machine-learning forecasts' being closer to realized earnings than are analysts' forecasts.

3.2 Empirical result: Bias and stock returns

$$R_{i,t+1} = \alpha + \beta_1 BE_{i,t} + \gamma_i \sum_{i=1}^{8} Control_{i,t} + \epsilon_{i,t+1}$$

	Panel A: Average BE		Panel B: BE Score	
	(1)	(2)	(1)	(2)
BE	−0.0808	−0.0852	−0.0279	−0.0456
	(−4.61)	(−5.30)	(−6.57)	(−15.99)
Lnsize		−0.0009		−0.0029
		(−2.46)		(−8.37)
Lnbeme		0.0012		0.0019
		(2.00)		(3.16)
Ret12_7		0.0038		0.0011
		(2.44)		(0.73)
Ret1		−0.0284		−0.0313
		(−6.62)		(−7.29)
IA		−0.0007		−0.0007
		(−2.60)		(−2.73)
Ivol		−0.1941		−0.1743
		(−1.72)		(−1.53)
Retvol		0.1339		0.1982
		(1.13)		(1.67)

Both the conditional bias and the bias score are associated with negative cross-sectional stock predictability.
We find that the coefficients on both the conditional bias and the bias score remain statistically significant after controlling for other firm variables.

3.2 Empirical result: Bias and stock returns

Quintile	1	2	3	4	5	1-5
			Panel A: Average BE			
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: BE 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

First, the value-weighted returns decrease in the conditional bias. A long-short portfolio of the extreme quintile results in a return spread of 1.95% per month (t-statistic 5.88) for the average bias and 1.53% per month (t-statistic 4.90) for the bias score.
Second, the CAPM betas of these portfolios tend to increase with higher biased expectations.

3.2 Empirical result: Bias and stock returns

	CAPM		FF3		FF5	
	Coeffi	*t*-stat	*Coeffi*	*t*-stat	*Coeffi*	*t*-stat
Panel A: Average BE						
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
Panel B: BE Score						
Intercept	1.94	7.02	2.03	8.01	1.53	5.73
Mkt_RF	-0.63	-7.50	-0.56	-6.58	-0.37	-4.62
SMB			-0.83	-6.89	-0.57	-4.39
HML			-0.44	-3.07	-0.83	-4.93
RMW					0.90	4.63
CMA					0.48	1.63

None of these can explain the documented return spread.

3.2 Empirical result: Placebo test

Quintile	1	2	3	4	5	1-5
			Panel A: Average BE			
Mean	2.97	1.30	-0.04	-0.98	-2.00	4.97
t-stat	11.77	6.10	-0.20	-3.61	-5.61	21.77
CAPM Beta	1.03	0.94	0.98	1.16	1.39	-0.36
			Panel B: BE Score			
Mean	2.87	1.40	0.14	-1.01	-2.59	5.46
t-stat	11.61	6.50	0.65	-3.94	-7.83	25.74
CAPM Beta	1.04	0.95	0.95	1.10	1.31	-0.27

We replace the machine learning forecast with the future realized value and then compute the conditional bias. The implied returns of these infeasible strategies are overwhelmingly better than the ones presented, with monthly excess returns in the order of 5% and t-statistics above 20.

3.2 Empirical result: Placebo test

	Panel A: Average BE		Panel B: BE_score	
	(1)	(2)	(1)	(2)
BE	−0.1208	−0.1473	−0.0945	−0.1061
	(−14.34)	(−16.47)	(−38.25)	(−45.94)
Lnsize		−0.0012		−0.0026
		(−3.18)		(−6.96)
Lnbeme		0.0019		0.0012
		(3.15)		(2.02)
Ret12_7		0.0026		−0.0019
		(1.66)		(−1.23)
Ret1		−0.0324		−0.0573
		(−7.52)		(−12.89)
IA		−0.0006		−0.0004
		(−2.09)		(−1.57)
Ivol		−0.1731		−0.1286
		(−1.52)		(−1.12)
Retvol		0.1693		0.1437
		(1.41)		(1.20)
Turnover		−0.0006		−0.0002
		(−1.26)		(−0.36)
Intercept	0.0089	0.0251	0.0549	0.0940
	(3.03)	(4.56)	(19.28)	(16.42)
R^2	0.0104	0.0655	0.0340	0.0917

The conclusion is consistent with the above.

3.2 Empirical result: Bias and anomalies

Engelberg et al. (2018) and Kozak et al. (2018) both find that analysts tend to have over-optimistic expectations for stocks in the short side of anomalies.
We focus on the 27 significant and robust anomalies considered in Hou et al. (2015).
We then combine these ranks to construct an anomaly score defined as the equal-weighted average of the rank scores of the 27 anomaly variables.

3.2 Empirical result: Bias and anomalies

	Anomaly Decile										
BE Quintile	S	2	3	4	5	6	7	8	9	L	L-S
	Panel A: Conditional Biases										
BE quintile	1	2	3	4	5	6	7	8	9	10	10-1
1	3.31	3.96	4.36	4.68	4.97	5.24	5.52	5.83	6.21	6.90	3.59
2	3.32	3.95	4.36	4.68	4.97	5.24	5.52	5.83	6.21	6.89	3.57
3	3.27	3.95	4.35	4.68	4.96	5.24	5.52	5.83	6.21	6.92	3.65
4	3.21	3.94	4.35	4.68	4.97	5.24	5.52	5.83	6.22	6.95	3.74
5	3.13	3.94	4.35	4.68	4.97	5.24	5.52	5.83	6.22	6.96	3.83
All stocks	3.21	3.95	4.35	4.68	4.97	5.24	5.52	5.83	6.21	6.92	3.72

For each anomaly decile portfolio, the anomaly score ranges from 3.21 to 6.92, with the highest (lowest) score indicating the long (short) leg of the anomaly strategy.
On average, we have around 50 stocks every month in each portfolio.

3.2 Empirical result: Bias and anomalies

	Anomaly Decile										
BE Quintile	S	2	3	4	5	6	7	8	9	L	L-S
1	1.24	1.00	1.12	1.30	1.09	1.20	1.14	1.27	1.26	1.54	0.30
t-statistic	3.20	3.18	4.06	5.13	4.44	4.92	5.12	5.99	5.92	6.56	0.95
2	0.02	0.58	0.58	0.71	0.82	0.64	0.85	0.82	1.05	0.92	0.90
t-statistic	0.07	1.80	2.21	2.74	3.18	2.72	3.68	3.67	4.72	3.98	2.97
3	-0.28	-0.08	0.45	0.43	0.47	0.64	0.57	0.72	0.78	0.85	1.12
t-statistic	-0.71	-0.26	1.42	1.38	1.59	2.28	2.24	2.71	3.14	3.26	3.86
4	-0.70	-1.04	-0.12	-0.02	-0.20	0.31	0.11	0.38	0.56	0.54	1.24
t-statistic	-1.63	-2.57	-0.33	-0.07	-0.55	0.94	0.33	1.20	1.72	1.59	3.89
5	-2.27	-1.54	-1.14	-1.07	-0.51	-0.16	-0.10	-0.59	-0.09	-0.05	2.22
t-statistic	-4.39	-2.88	-2.46	-2.26	-1.03	-0.32	-0.21	-1.18	-0.19	-0.09	6.11
All Stocks	S	2	3	4	5	6	7	8	9	L	L-S
Excess Return	-0.27	0.11	0.52	0.62	0.61	0.69	0.72	0.76	0.96	0.98	1.25
t-statistic	-0.77	0.36	2.01	2.55	2.52	3.04	3.44	3.76	4.82	4.47	5.23
Average BE	0.014	0.010	0.009	0.008	0.008	0.007	0.007	0.007	0.007	0.008	-0.006
t-statistic	6.48	6.19	6.27	6.09	5.71	6.01	5.82	5.97	6.19	6.06	-5.75

Anomaly payoffs tend to arise from overpricing on stocks with the most over-optimistic expectations.
The short-leg portfolio is comprised of stocks with more over-optimistic expectations, suggestive of overpricing.

3.3 Empirical result: Bias and Firm’s Decisions

Corporate managers have more information about their own firms than investors have, and can use that informational advantage.
Managers’ expectations will align more closely to our statistically optimal benchmark.
Managers can identify the biases when investors overestimate or underestimate firms’ future cash flows.
Managers may issue more stock when investors’ expectations are higher than their own.

3.3 Empirical result: Bias and Firm's Decisions

Panel A: Net Stock Issuances of Portfolios formed on BE

Quintile	1	2	3	4	5	5-1
Average BE	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
BE 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

Panel B: Fama-MacBeth regressions

	A: Average BE		B: BE Score	
	(1)	(2)	(1)	(2)
BE	1.7048	1.2870	0.1191	0.0510
t-stat	3.86	4.53	6.74	4.82
Lnsize		-0.0053		-0.0051
t-stat		-3.25		-2.76
Lnbeme		-0.0239		-0.0230
t-stat		-6.10		-5.70
EBITDA		-0.1086		-0.1129
t-stat		-4.36		-4.34
Intercept	0.0621	0.1186	0.0108	0.0775
t-stat	6.12	3.72	0.95	2.46
R^2	0.0178	0.0921	0.0084	0.0750

The net stock issuances increase monotonically in the conditional bias. Managers of firms whose earnings forecasts are more optimistic issue on average 6% more of total shares outstanding.

4. Conclusion

1. The analysts' forecast income is upward biased, and the random forest prediction is more accurate.
2. Using machine learning prediction as a benchmark to measure analysts' error can predict future stock returns.
3. Managers of companies with high earnings forecast errors will issue more stocks.

4. Comment & Inspiration

1. The parameters of random forest do not change with time. If the optimal parameters are searched again every time the window scrolls, the result may be better.
2. When the stock price is overvalued, we can study other behaviors of executives, such as financing decision, investment scale and investment efficiency.