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

Jules H. van Binsbergen, Xiao Han, Alejandro Lopez-Lira

解读：王梦涵

2024 年 1 月 2 日

Background

- why are earnings expectations important?
  - Expected future cash flow is a necessary input for pricing a risky asset
- Why are analyst forecast biases important?
  - Influence investors： Market mispricing
  - Influence managers： Stock issuance and repurchase
- Why analyst forecast biases exit?
  - Market uncertainty
  - Individual differences: private information； experience…

Motivation

- 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

Why use random forest regression ？

- It allows nonlinear relationships
  - Allowing for nonlinear effects improves the forecasts （Gu et al.,2020）
  - EPS as a nonlinear function of analysts’ forecasts
- It is designed for high dimensional data
  - Robust to overfitting in the context of cross-sectional returns
  - The variable selection step is innocuous
  - It automatically discards useless forecasting variables and incorporates useful ones

Question

- Whether the analyst forecasts are conditionally biased?
  - Yes
- The effects of analyst forecast biases:
  - Can/How the expectations error affect stock market returns?
    - Find negative stock return predictability
  - What is the relationship between market anomalies and conditional bias?
    - The expectational error component driving anomalies
  - How the new biases affect managers' actions?
    - Managers take advantage of the overly optimistic expectations by issuing stocks

Contribution

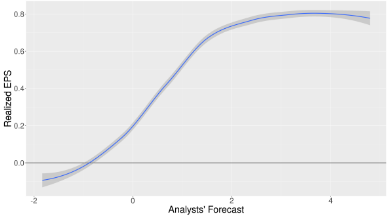
- First to use ML to create an optimal and unbiased proxy for firms earnings’ conditional expectations
  - A real-time benchmark for analyst forecast bias(So,2013)
- Contributions to relevant literature
  - Relationship between anomalies and conditional biases(Engelberg et al.,2018)
  - Providing direct 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)

Build random forest——model

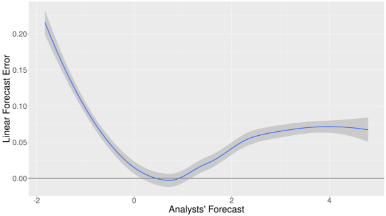
- There are two periods in the economy

$$\widetilde{y}_i = f(x_i) + g(v_i) + z_i + w_i + \widetilde{\epsilon}_i \tag{1}$$

  - v_i,w_i,x_i,and z_i are variables measurable in the first period
  - f and g are nonlinear functions



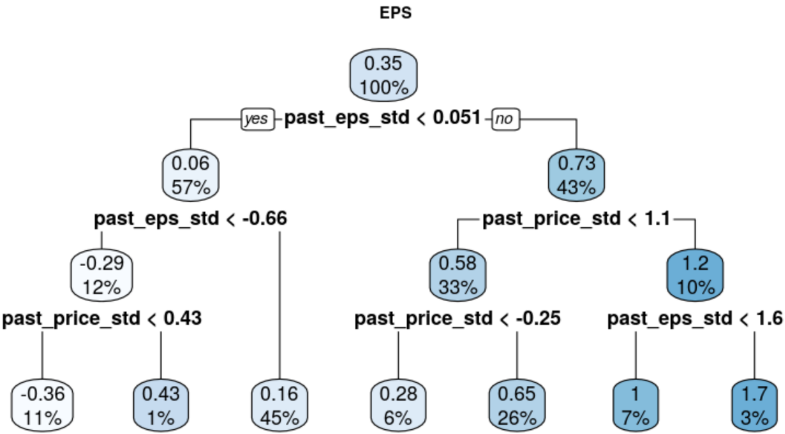
(A) EPS as a nonlinear function of analysts' forecasts



(B) Linear forecast error as a nonlinear function of analysts' forecasts

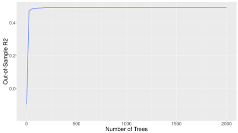
Build random forest——decision tree

- use random samples drawn with replacement
- use a random subset of features at each node

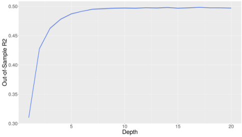


Build random forest——forest

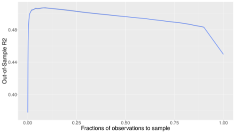
- Choose the hyperparameters using cross-validation
  - training data: from sample inception until January 1986
  - testing data: February 1986



(A) Number of trees in the one-quarter-ahead forecast



(C) Maximum depth of each tree in the one-quarter-ahead forecast



(E) Fraction of the sample in the one-quarter-ahead forecast

- Then the model is trained using rolling windows of 1-year(2-year), keeping the hyperparameters fixed.

Data

- 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)

Analyst forecasts and Real-time biases

- Analyst forecasts(AF)

$$AF_{i,t}^{t+1} = \sum_{j=1}^{J} \beta_j X_{j,i,t} + \sum_{k=1}^{K} \gamma_k P_{k,i,t} + B_{i,t} \tag{2}$$

  - j denotes public signal,k denotes private signal

- The biases in investor expectations(BE)

$$Biased\_Expectation_{i,t}^{t+h} = \frac{Analyst\_Forecasts_{i,t}^{t+h} - ML\_Forecast_{i,t}^{t+h}}{Price_{i,t-1}} \tag{3}$$

  - i denotes firm,t indicates the date,$t+h$ represents the forecasting period

AF VS RF

- RF: ML EPS forecast
- AE: actual realized earnings

The term structure of earnings forecasts via machine learning

	RF	AF	AE	(RF-AE)	(AF-AE)	$(RF-AE)^2$	$(AF-AE)^2$	(AF-RF)/P
One-quarter-ahead	0.290	0.319	0.291	-0.000	0.028	0.076	0.081	0.005
t-stat				-0.17	6.59			6.54
Two-quarters-ahead	0.323	0.376	0.323	-0.001	0.053	0.094	0.102	0.007
t-stat				-0.13	10.31			7.75
Three-quarters-ahead	0.343	0.413	0.341	0.002	0.072	0.121	0.132	0.007
t-stat				0.31	11.55			8.08
1-year-ahead	1.194	1.320	1.167	0.027	0.154	0.670	0.686	0.021
t-stat				1.64	6.24			5.17
2-year-ahead	1.384	1.771	1.387	-0.004	0.384	1.897	2.009	0.035
t-stat				-0.07	8.33			6.57

- Analysts are overoptimistic relative to the machine-learning
- AF get more precise when predicting at shorter horizons
- Conditional biases are statistically different from 0

BE and stock returns——cross-section

- Y: monthly stocks' return

Fama-Macbeth regressions

	A. Average BE (1)	*A. Average BE* (2)	*B. BE score* (1)	*B. BE score* (2)
Bias	-0.054	-0.064	-0.017	-0.028
t-stat	-3.94	-5.08	-4.47	-11.27
ln(size)		-0.079		-0.215
t-stat		-2.22		-6.42
ln(beme)		0.091		0.178
t-stat		1.58		3.14
Ret1		-2.818		-2.987
t-stat		-6.72		-7.12
Ret12_7		0.442		0.220
t-stat		2.88		1.52
IA		-0.003		-0.003
t-stat		-5.67		-5.88
IVOL		-0.224		-0.198
t-stat		-2.04		-1.80
Retvol		0.137		0.168
t-stat		1.19		1.47
Turnover		-0.065		-0.046
t-stat		-1.46		-1.03
Intercept	1.022	2.320	1.865	5.362
t-stat	3.64	4.41	7.89	11.35
R^2 (%)	0.780	5.680	1.242	5.756

- BE are associated with negative stock return predictability.

BE and stock returns——time-series

Portfolios sorted on conditional bias

Quintile	1	2	3	4	5	5-1
A. Average BE						
Mean	1.32	0.98	0.79	0.47	-0.14	-1.46
t-stat	6.53	4.53	3.18	1.62	-0.35	-5.11
CAPM beta	0.90	0.97	1.09	1.22	1.46	0.56
B. BE score						
Mean	1.14	0.93	0.79	0.60	-0.02	-1.16
t-stat	5.66	4.22	3.18	2.06	-0.05	-3.83
CAPM beta	0.90	0.99	1.10	1.21	1.51	0.61

- Value weighted returns decrease with the conditional bias
- CAPM betas tend to increase with higher biased expectations(Antoniou et al.,2015;Hong and Sraer,2016)

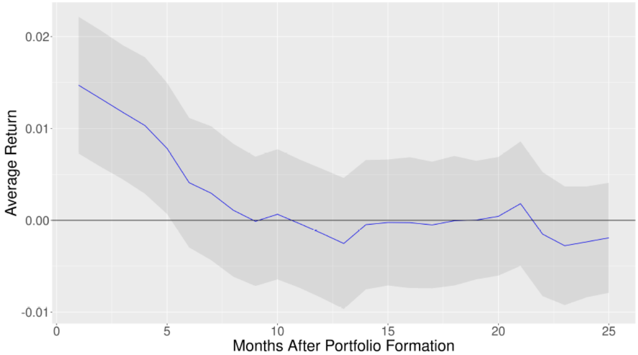
BE and stock returns——time-series

Time-series tests with common asset pricing models

	CAPM		FF3		FF5	
	Coef (β)	t-stat	Coef (β)	t-stat	Coef (β)	t-stat
			A. Average BE			
Intercept	-1.85	-7.18	-1.96	-8.64	-1.54	-5.84
Mkt_RF	0.56	7.53	0.53	7.86	0.38	5.28
SMB			0.80	7.06	0.61	5.17
HML			0.58	5.25	0.95	7.12
RMW					-0.68	-4.10
CMA					-0.53	-1.93

- Returns on this long-short strategy can't be explained by leading asset pricing models

BE and stock returns——time-series



- The negative return decrease quickly after the first month

BE and market anomalies

- 27 significant and robust anomalies
- Average of the rank scores of the 27 anomaly variables

BE quintile	Anomaly decile S	2	3	4	5	6	7	8	9	L	L-S
A. Anomaly score											
1	3.35	3.97	4.37	4.70	4.99	5.26	5.54	5.85	6.22	6.81	3.46
2	3.34	3.98	4.37	4.70	4.99	5.26	5.54	5.85	6.23	6.82	3.47
3	3.31	3.97	4.37	4.69	4.99	5.26	5.54	5.85	6.22	6.83	3.52
4	3.24	3.96	4.37	4.69	4.99	5.25	5.54	5.85	6.23	6.87	3.62
5	3.21	3.95	4.37	4.69	4.99	5.26	5.53	5.85	6.23	6.91	3.71
All stocks	3.31	3.97	4.37	4.70	4.99	5.26	5.54	5.85	6.23	6.82	3.51
B. Number of stocks											
1	$\boxed{37}$	47	52	57	63	64	66	67	65	62	
2	34	50	56	62	64	65	66	67	62	54	
3	51	59	61	62	60	59	58	58	56	54	
4	73	65	60	57	55	52	52	52	53	58	
5	$\boxed{97}$	70	60	53	49	48	46	48	51	60	
All stocks	292	291	289	291	291	288	289	292	286	288	

- Stocks with higher BE tend to be anomaly shorts

BE and market anomalies

	Anomaly decile										
BE quintile	S	2	3	4	5	6	7	8	9	L	L-S
1	1.06	1.00	1.28	1.36	1.38	1.45	1.48	1.34	1.64	1.66	0.60
t-stat	2.73	3.21	4.84	5.40	5.43	6.25	6.90	6.60	7.91	7.09	1.82
2	0.29	0.76	0.99	1.06	0.94	0.90	1.10	1.02	1.33	1.38	1.09
t-stat	0.82	2.66	3.77	4.22	3.78	3.79	4.73	4.50	6.38	6.31	3.74
3	-0.16	0.40	0.64	0.60	0.68	1.11	0.92	1.02	1.21	1.06	1.23
t-stat	-0.43	1.24	2.23	2.14	2.52	4.13	3.65	4.06	4.72	4.06	4.40
4	-0.73	-0.31	0.51	0.58	0.30	0.64	0.74	0.80	1.04	0.81	1.54
t-stat	-1.75	-0.79	1.53	1.59	0.86	1.87	2.33	2.66	3.54	2.58	4.78
5	-1.29	-0.81	-0.41	-0.01	-0.06	0.27	0.25	0.29	0.90	0.84	2.13
t-stat	-2.62	-1.63	-0.97	-0.03	-0.14	0.61	0.59	0.69	2.04	1.99	6.37
5-1	-2.35	-1.81	-1.69	-1.38	-1.44	-1.18	-1.23	-1.05	-0.74	-0.83	1.52
t-stat	-6.04	-4.75	-5.02	-3.66	-3.84	-3.12	-3.36	-2.98	-1.92	-2.37	3.81
All stocks	S	2	3	4	5	6	7	8	9	L	L-S
Return	-0.06	0.46	0.81	0.95	0.87	1.02	1.04	1.05	1.31	1.30	1.36
t-stat	-0.17	1.56	3.22	3.99	3.66	4.52	4.94	5.11	6.62	5.94	5.74
BE	0.009	0.007	0.005	0.004	0.004	0.004	0.004	0.003	0.004	0.004	-0.005
t-stat	5.83	5.24	6.19	6.05	5.59	5.76	6.02	5.73	5.02	4.71	-4.81

- Anomaly payoffs arising from the overpricing of stocks with the most overoptimistic earnings expectations
- Short-leg portfolio is comprised of stocks with more overoptimistic

BE and firm's financing decisions

- Annually net stock issuances(NSI)

$$NSI_{i,t} = log(\frac{Split\_adjusted\_shares_{i,t}}{Split\_adjusted\_shares_{i,t-1}}) \quad (4)$$

- Y:NSI

A. Net stock issuances of portfolios formed on BE

Quintile	1	2	3	4	5	5-1
Average BE	0.006	0.012	0.017	0.028	0.065	0.059
t-stat	1.16	1.54	2.52	4.13	4.86	4.24
BE score	0.006	0.011	0.018	0.030	0.063	0.057
t-stat	0.99	1.50	3.37	5.58	4.32	3.69

B. Fama-MacBeth regressions

	A. Average BE		*B. BE score*	
	(1)	(2)	(1)	(2)
Bias	0.442	0.355	0.072	0.039
t-stat	2.24	1.94	4.57	2.14

- Managers of firms with larger BE issue more stocks

Conclusion

- This paper develops a ML forecast algorithm that is statistically optimal and unbiased
  - In contrast to linear forecasts, the new benchmark is effective out-of-sample
- Construct a real-time measure of analyst forecast biases
  - Asset pricing results
    - BE are associated with negative stock return predictability
    - Analysts' forecast errors may have an effect on asset prices
  - The expectational error component driving anomalies
  - Managers of firms for which BE is higher issue more stocks

New ideas

- 补充: 有这个新的分析师偏差指标我们还可以做什么？
  - 研究分析师预测偏差与股票机构投资者关注度/持股比例、市场情绪、行业、分析师分析风格/社会关系/教育背景等的关系
- 我们还可以用机器学习的方式构建什么偏差？
  - 宏观经济变量的预期和偏差
  - ……