Can Machines Understand Human Decisions? Dissecting Stock Forecasting Skill

Sean Cao, Xuxi Guo, Houping Xiao, Baozhong Yang

解读：雷印如

2021 年 12 月 26 日

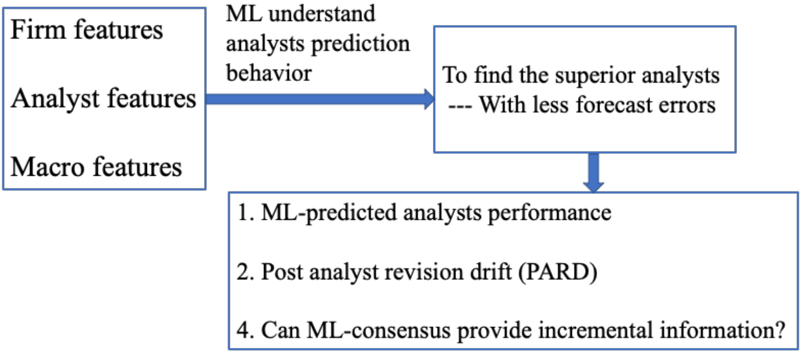
Motivation

- Given the potentially different and complex underlying models, predicting human behavior is empirically challenging.
- The machine learning and AI have achieve success in intellectual tasks that used to be exclusively in the human domain.
- Can machines help understand and predict human decisions? For example, Analyst forecasting.

Why choose analyst earnings forecasts?

- Analysts are important information intermediaries that investors rely on and their decisions have important implications for the financial markets.
- Data available or advantages
  - There is a rich set of observable features associated with analysts and the firms they cover that can be used to understand their skills.
  - Analyst earnings forecasts provide a directly measurable proxy for individual decisions.
  - Past actual earnings provide a natural and accurate benchmark that we can use to determine the accuracy of forecasts and train machine learning models.

Framework


Firm features
Analyst features
Macro features
ML understand analysts prediction behavior
To find the superior analysts --- With less forecast errors
1. ML-predicted analysts performance
2. Post analyst revision drift (PARD)
4. Can ML-consensus provide incremental information?

Literatures

- An emerging literature has applied machine learning in studying issues in financial markets, including
  - Predicting asset prices (Gu, Kelly, and Xiu (2020))
  - Portfolio management (Chen et al., 2020; Cong et al., 2020)
  - Estimating values of artwork (Aubry et al., 2020)
  - Evaluating bank risk (Hanley and Hoberg, 2019)
  - Predicting corporate earnings (van Binsbergen et al., 2020; Cao and You, 2020)

Literatures

- A recent literature shows that investors can benefit from retrieving information from the wisdom of crowds, through
  - Online articles and commentaries (Chen et al., 2014)
  - Crowd forecasts (Da and Huang, 2020)
- This method can accurately aggregate information from multiple individuals that is superior to the commonly used mean (or median) analyst consensus

Contribution

- This study is one of the first to apply machine learning to study human behavior and skill, that provides an interpretable machine learning framework that can be used in other settings.
- This study contributes to the analyst literature by providing a robust, persistent skill measure for analysts that dominates industry-expert-identified star status and historical performance.
- This study is able to explain the post-analyst-revision drift puzzle.

Data

- Quarterly earnings forecasts 1994-2018 from Institutional Brokers' Estimate System (I/B/E/S) dataset.
  - Remove stale forecasts and retain the latest forecast an analyst issues within 90 days before the earnings announcement date
  - Remove analysts with fewer than ten quarters of earning forecasts and firms with fewer than ten quarters of observations in the sample

Data

- The forecast data are then merged with CRSP and Compusat to obtain stock returns and firm characteristics.
  - Drop an observation if the quarter-end stock price is less than $1
- The final sample consists of 2,301,286 analyst -firm -quarter observations of earnings forecasts issued by 19,693 analysts for 10,429 firms.

Construction of Target Variable: Star

- We construct a relative measure of forecast error, Star, to capture the superior performance of analysts
  - Calculate analyst i's absolute forecast error

$$AFE_{i,j,t} = |EPS_{i,j,t}^{forecast} - EPS_{i,j,t}^{*}|$$

  - Sort all the analysts ($N_{j,t}$ in total) to create $Rank_{i,j,t}$
  - Scaled by the number of analysts to be within [0, 1]

$$Score_{i,j,t} = \frac{Rank_{i,j,t} - 1}{N_{j,t}}$$

- We define the target variable Star as 1 if Score is greater than its median (0.5) and 0 otherwise.

Analyst & Forecast Features

- Forecast Values
  - Earning Forecast: the latest earnings forecast for the firm announced by the analyst in the current quarter.
  - Average Consensus: the median of all earnings forecasts issued for the firm and the current quarter in our sample.
  - IBES Consensus: the median of all earnings forecasts issued for the firm and the current quarter in the IBES database.

Analyst & Forecast Features

- Forecast Characteristics
  - Accuracy: the score (scaled rank) of the average absolute forecast errors over the last eight quarters.
  - Consistency: the score (scaled rank) of the standard deviation of absolute forecast errors over the last eight quarter.
  - Forecast Horizon: the number of days between the analyst forecast announcement and actual earnings announcement.

Analyst & Forecast Features

- Effort and Experience
  - Number of Revisions: the number of forecast revisions made by the same analyst for the current quarter's earnings.
  - Firm Forecast Experience: the number of quarters elapsed from the analyst's first estimate for the firm to the current estimate.
  - Industry Forecast Experience: the number of quarters elapsed from the analyst's first estimate for firms within the same Fama-French 12 industry as the firm to the current estimate.

Analyst & Forecast Features

- Portfolio and Resource
  - Company Coverage: the number of firms covered by the analyst in the current quarter.
  - Brokerage Size: the number of analysts in the brokerage firm of the given analyst.
  - Reports Revenue: whether the analyst reports revenue forecasts in the current quarter.
  - Reports Cash Flow: whether the analyst reports cash flow forecasts in the current quarter.

Firm Characteristics and Macroeconomic Variables

- Firm characteristics
  - There are in total 133 firm-level variables included.
- Macro-economic and stock market development
  - Industrial Production Index;
  - Consumer Price Index; Crude Oil Price (WTI)
  - 3-month treasury bill rate
  - 10-Year treasury constant-maturity rate
  - The BAA-AAA yield spread.

Predicting Analyst Skill with Machine Learning

- The prediction model of analyst skill can be written in the following general form

$$Star_{i,j,t} = f(A_{i,j,t-}, F_{j,t-}, T_{j,t-}) + \epsilon_{i,j,t}$$

where $A_{i,j,t-}$ are the firm-analyst level features available prior to the earnings announcement of firm j in quarter t, $F_{j,t-}$ the company-level features, and T_t the macroeconomic features.

Predicting Analyst Skill with Machine Learning

- Rolling-window approach: For each quarter t, we use data in the previous sixteen quarters $t-16, ..., t-1$ for training and validation.
  - Randomly select 70% of the earning forecasts and associated data in the previous sixteen quarters as the training sample
  - Let the remaining 30% to be the validation sample for tuning the hyper-parameters.
- Given that our sample starts in 1994 and the requirement of a four year training-validation period, the models make their first predictions in 1999

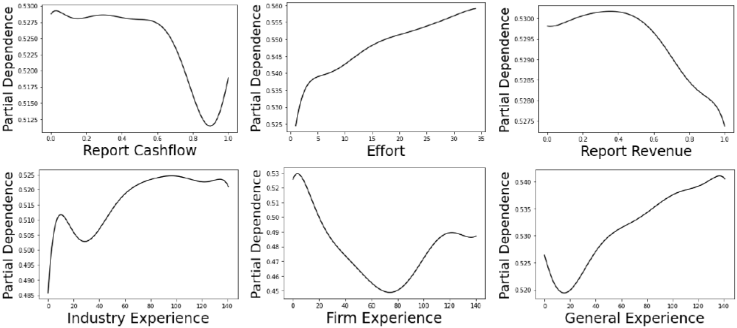
Predicting Analyst Skill with Machine Learning

- The neural networks thus can best capture the complicated interactions among the features in this setting.
- In particular, the CNN model performs the best because it is able to capture “local”interactions among more related variables.

	Models	Accuracy	Precision	Recall	F1 Score
Linear	Logistic Regression	53.81%	54.33%	90.23%	67.84%
	Logistic LASSO	55.49%	55.90%	82.93%	66.78%
Non-Linear	Gradient Boost	58.14%	57.70%	83.95%	68.40%
	Neural Network	68.27%	68.02%	77.72%	75%
	Convolutional Neural Network	69.03%	69.10%	77.04%	72.86%

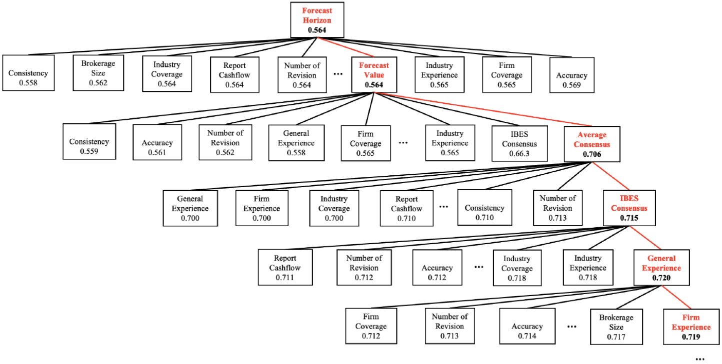
Interpretation of the CNN Model

- CNN is designed to automatically and adaptively learn spatial hierarchies of features that captures local and nonlinear interactions among input features



Interpretation of the CNN Model

- The importance of local interactions among Forecast Value feature group can be seen from the big increase by adding Average Consensus or IBES Consensus



Predicting Analyst Skill with CNN

- We define the predicted analyst skill from the final model as a new variable ML-Star
- In particular, the CNN model performs the best because it is able to capture "local"interactions among more related variables.

Feature	Accuracy	Precision	Recall	F1 Score
$[Analyst, Firm, Macro]$	68.60%	68.37%	77.45%	72.63%
$[Analyst, Macro]$	68.56%	68.53%	77.29%	72.65%
$[Analyst, Firm]$	68.62%	68.44%	77.30%	72.60%
$[Firm, Macro]$	53.93%	53.96%	99.66%	70.01%
$[Analyst]$	69.03%	69.10%	77.04%	72.86%
$[Firm]$	53.93%	53.94%	99.77%	70.03%
$[Macro]$	54.16%	54.16%	99.97%	70.26%

Performance of ML-Predicted Analyst Skill

- Historical Star is the analyst with an average prediction score higher than the median over last year
- All-Star is the analysts awarded by Institutional Investor from the previous year

	(1)	(2)	(3)	(4)
Variables		*Star*		
ML-Star	0.381***	0.382***	0.380***	0.380***
	(123.66)	(81.42)	(123.65)	(81.02)
Historical Star			0.018***	0.018***
			(19.72)	(16.41)
All-Star			0.009***	0.006***
			(6.29)	(3.06)
Quarter FE	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	1,488,430	1,488,430	1,488,430	1,488,430
R-squared	0.145	0.145	0.145	0.145

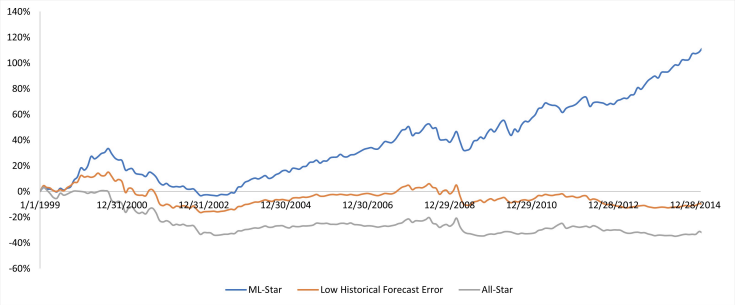
Cross-section of Forecast Accuracy Prediction

- ML model performs better in firms with lower information uncertainty and under non-crisis periods

	ML-Star on Analyst Forecast Accuracy			
	High	Low	Diff	t-Stat
Information Uncertainty				
Bid Ask Spread	0.364***	0.415***	-0.051	(-5.70)
Adj probability of informed trading	0.389***	0.431***	-0.042	(-3.08)
Flesch-Kincaid Grade Level	0.398***	0.383***	0.015	(2.28)
Accruals Quarlity	0.386***	0.409***	-0.023	(-2.64)
Earning Quality	0.416***	0.384***	0.032	(4.49)
Cashflow Volatility	0.365***	0.403***	-0.038	(-6.15)
Return Volatility	0.362***	0.417***	-0.055	(-6.64)
Firm Age	0.397***	0.380***	0.017	(2.56)
Market Condition				
NBER Crisis Dummy	0.366***	0.393***	-0.027	(-1.96)
Oil Price	0.398***	0.382***	0.016	(2.26)
T-Bill Yield	0.393***	0.388***	0.005	(0.64)
Term Spread	0.387***	0.393***	-0.006	(-0.78)
Default Spread	0.388***	0.392***	-0.004	(-0.68)
VIX	0.395***	0.387***	0.008	(0.94)

Post Analyst Revision Drift

- We continue to examine whether the predictability of ML-Star on the cross-section of future stock returns
- The figure shows that the post analyst revision drifts are mainly driven by ML-Star



Post Analyst Revision Drift

- We create a 60-day long(short) signal if an ML-Star analyst whose revision is greater than previous forecast

	(1)	(2)	(3)	(4)	(5)
	Algorithm Return Based on PARD				
Variables	FF3	FF4	FF5	Q4	SY4
Mkt-RF	0.174***	0.269***	0.237***	0.261***	0.301***
	(6.71)	(12.12)	(7.86)	(9.12)	(10.36)
SMB	-0.060*	-0.108***	-0.011	0.028	-0.041
	(-1.70)	(-3.84)	(-0.29)	(0.85)	(-1.30)
HML	-0.092***	-0.008	-0.212***		
	(-2.72)	(-0.29)	(-4.47)		
Mom		0.204***			
		(11.61)			
RMW			0.173***		
			(3.33)		
CMA			0.143**		
			(2.16)		
R_IA				-0.106**	
				(-2.08)	
R_ROE				0.278***	
				(6.49)	
MGMT					-0.031
					(-0.86)
PERF					0.225***
					(9.61)
Constant	0.002**	0.001	0.001	0.001	-0.000
	(1.97)	(1.22)	(0.83)	(0.71)	(-0.38)
Observations	236	236	236	236	216
R-squared	0.193	0.491	0.242	0.295	0.413
adj R-squared	0.183	0.482	0.225	0.282	0.402

Machine-Learning-Based Analyst Consensus

- The aggregated opinion of these analysts will be more accurate consensuses compare to the consensuses by convention
  - We define ConsensusML as the median of forecasts of ML-Star covered the firm
  - the consensus is defined as the median of all forecasts of the analysts covered by the firm
- Then, we test whether the new consensus provides additional information in predicting forecast earning

Machine Learning Consensus and Earnings

- The ConsensusML contains additional information in predicting earnings

	(1)	(2)	(3)	(4)
Dependent Variable		*Earnings*		
Consensus_ML - Consensus	2.133***	2.142***	2.058***	2.034***
	(4.25)	(4.32)	(4.73)	(4.93)
Consensus	1.064***	1.059***	1.112***	1.091***
	(41.69)	(48.97)	(22.46)	(25.18)
Liquidity	-0.003		0.004**	
	(-1.32)		(2.55)	
Momentum	0.030***		0.012**	
	(6.32)		(2.13)	
Log_Size	-0.004		-0.007	
	(-0.80)		(-0.73)	
Book to Market	-0.010*		0.028*	
	(-1.97)		(1.90)	
Coverage	0.001**		-0.001**	
	(2.24)		(-2.23)	
Quarter FE	Yes	Yes	Yes	Yes
Fund FE	No	No	Yes	Yes
Observations	156,635	203,759	156,158	203,118
Adj R-squared	0.790	0.771	0.808	0.791

ML Consensus and Announcement Returns

- Incremental information of ConsensusML from I/B/E/S is not priced in the market before earning announcement but is priced shortly after then.

Variables	(1) *CAR [-1, +1]*	(2) *CAR [+2, +7]*	(3) *CAR [+8, +14]*
Consensus_ML - Consensus	0.019**	-0.001	-0.003
	(2.60)	(-0.16)	(-0.51)
Consensus	0.002***	0.003***	0.002***
	(2.70)	(3.02)	(3.46)
Liquidity	-0.001	-0.001*	0.000
	(-1.46)	(-1.96)	(0.54)
Momentum	0.001	-0.004**	-0.002
	(0.86)	(-1.99)	(-1.24)
Log_Size	-0.012***	-0.007***	-0.007***
	(-11.45)	(-5.98)	(-6.86)
Book to Market	-0.001	-0.001	-0.002*
	(-1.26)	(-0.84)	(-1.82)
Coverage	-0.000	0.000	0.000
	(-0.81)	(0.82)	(0.06)
Quarter FE	Yes	Yes	Yes
Fund FE	Yes	Yes	Yes
Observations	154,783	154,767	154,662
adj R-squared	0.0293	0.0410	0.0446

Conclusion

- This paper uses ML models to analyze analysts' forecasting performance and predict their skill.
  - ML-predicted skilled analysts persistently outperform expert-picked star analysts
  - Machines rely on nonlinear interactions of analyst characteristics to make predictions
  - The puzzle of post-analyst-revision drifts can be explained by our ML model in that such drifts are concentrated in ML-predicted skilled analysts
  - Investment strategies formed based on revisions of ML-predicted skilled analysts and the smart consensus both generate significant abnormal returns

Consideration

- Machine learning in predicting mutual fund managers