Expected Growth and Stock Returns: A Machine Learning Approach

Yuan Tian
University of Cincinnati

Abstract

- Expected growth is an important firm fundamental variable but is unobservable and difficult to estimate.
- In this study, I apply machine learning (ML) to forecast growth at the firm level. Compared with conventional linear regression, ML models produce more accurate forecasts out of sample. In particular, the gradient boosting regression performs the best.
- As the first application, I test the relationship between expected growth and expected stock returns. Consistent with theory, expected growth *positively* predicts future returns, controlling for past growth.

Motivation

- Studying expected growth is challenging
  - It is not very persistent and can be lumpy
- The conventional forecasting models used by previous studies have limitations: assume a restrictive linear function form and utilize a small number of predictors
  - The out-of-sample (OOS) forecast accuracy is low
  - Lead to mixed evidence on the relationship between expected growth and expected stock returns

Methodology

- I forecast the firm-level growth using the following models
  - **Ordinary Least Squares (OLS)**: a commonly used benchmark model
  - **Least Absolute Shrinkage and Selection Operator (LASSO)**: a linear regression that performs L1 regularization
  - **Ridge**: a linear regression that performs L2 regularization
  - **Elastic Net**: a hybrid of ridge regression and lasso regression
  - **Random Forest (RF)**: an ensemble method that combines decision trees in parallel
  - **Gradient Boosting Regression (GBR)**: an ensemble method that combines decision trees sequentially
  - **Artificial Neural Network (ANN)**: a deep learning algorithm mimicking the biological neural networks in human brains
- Hyperparameters for ML models are tuned via a five-fold cross-validation procedure
- In each month, I use data from a 60-month rolling window to train the models and then use the latest predictors to make OOS forecasts

Data

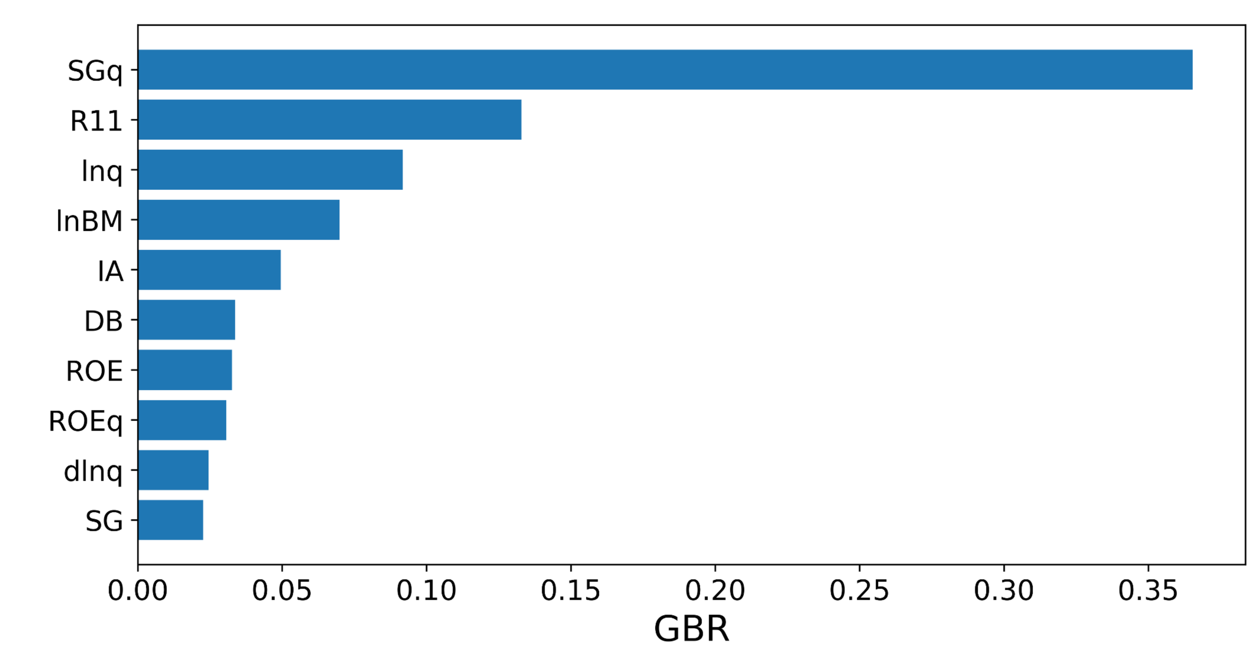
- Sample: US common stocks listed on NYSE, Amex, and NASDAQ
- Target variable: one-year-ahead investment-to-assets (IA)
- Predictors: 43 features covering 6 categories (growth opportunities, market expectation, profitability, risk, financing, and life cycle)
- The OOS predictions are monthly from June 1969 to December 2020

Empirical Results

GBR generates the lowest forecasting errors

	Forecast Errors			
	Mean	% Improve	Median	% Improve
OLS	0.168		0.105	
Linear ML				
LASSO	0.157	6.55%	0.100	4.76%
Ridge	0.167	0.60%	0.104	0.95%
Elastic Net	0.158	5.95%	0.101	3.81%
Nonlinear ML				
RF	0.162	3.57%	0.099	5.71%
GBR	0.152	9.52%	0.083	20.95%
ANN	0.155	7.74%	0.092	12.38%

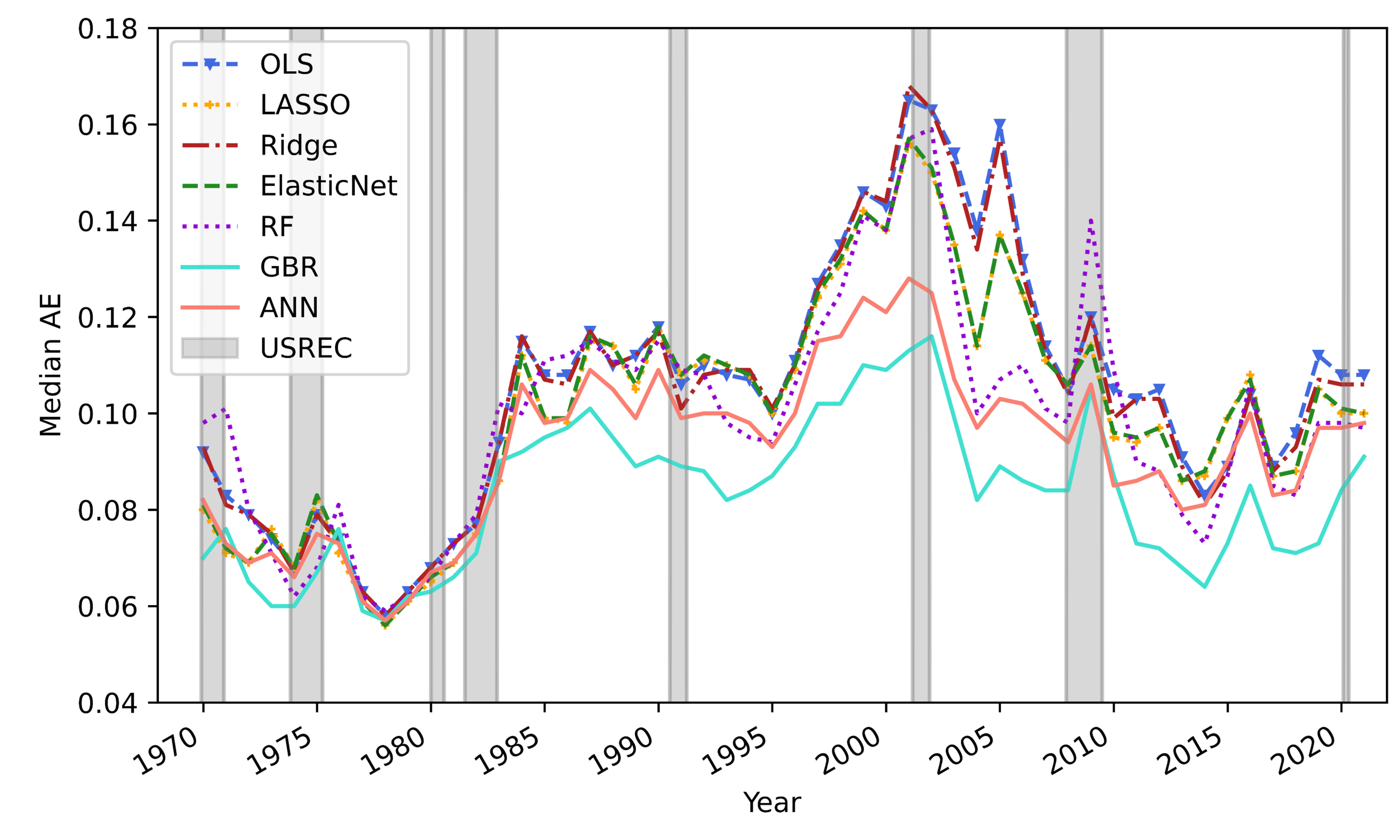
The top-10 most influential predictors for GBR based on the average permutation feature importance



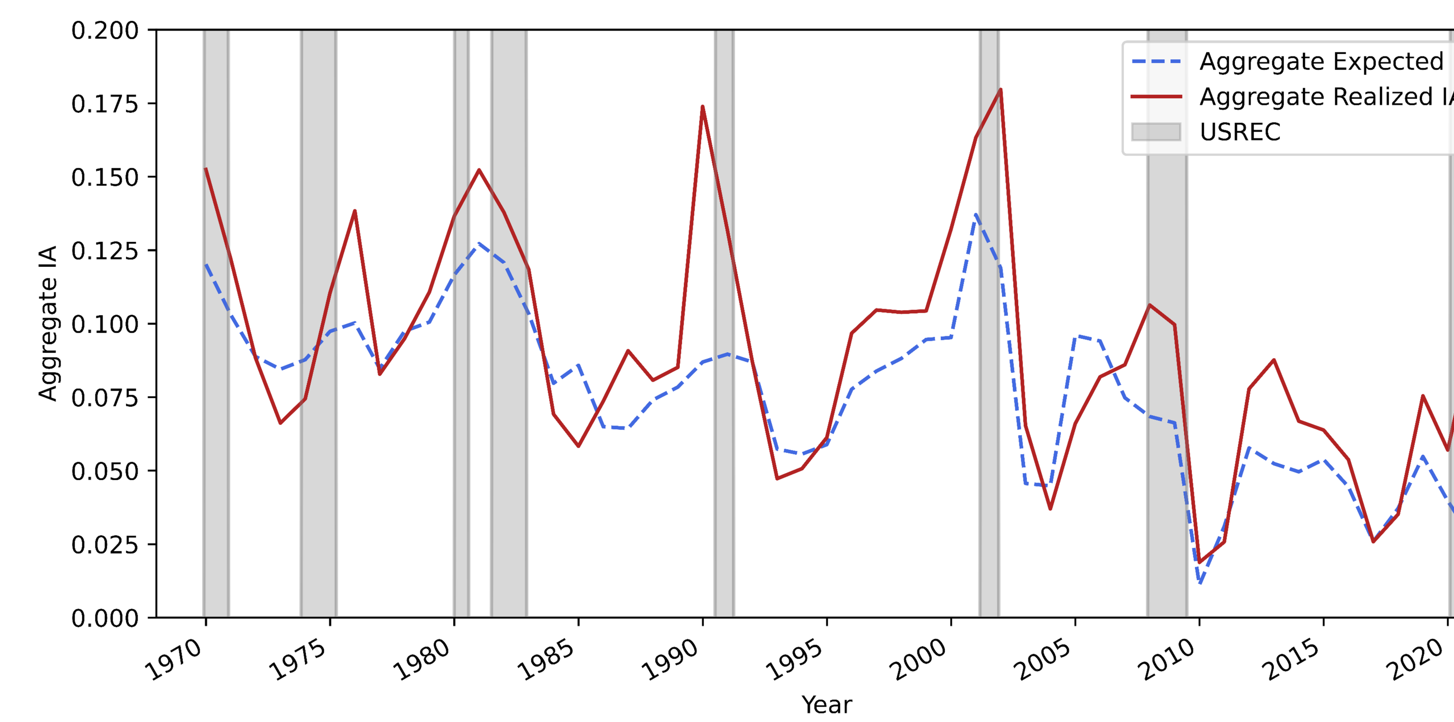
Stock deciles sorted on the expected change in growth

	Low	2	3	4	5	6	7	8	9	High	H-L	t[H-L]
Panel A: Average excess returns, $\bar{R}$												
$\bar{R}$	0.33	0.46	0.54	0.53	0.71	0.53	0.69	0.73	0.71	0.89	0.56	4.49
Panel B: The q-factor model												
α	-0.13	0.01	0.11	-0.07	0.03	-0.07	0.06	0.07	0.02	0.29	0.42	2.60
Panel C: The q^5-factor model												
α	0.00	0.02	0.10	-0.08	-0.01	-0.08	0.03	0.00	-0.12	0.21	0.22	1.53
EG	-0.18	-0.01	0.02	0.03	0.06	0.01	0.03	0.09	0.21	0.11	0.29	2.49
Panel D: The Fama-French five-factor model												
α	-0.18	-0.11	0.03	-0.13	0.02	-0.07	0.02	0.05	0.05	0.33	0.52	4.07
Panel E: The Fama-French six-factor model												
α	-0.15	-0.04	0.08	-0.07	0.01	-0.07	0.02	0.04	0.01	0.26	0.42	3.42

Nonlinear ML models show superior performance over time



The time-series of the aggregate expected IA and aggregate realized IA



Extend portfolio holding periods from 1 month to 6 and 12 months

	Low	2	3	4	5	6	7	8	9	High	H-L	t[H-L]
6-month holding period												
$\bar{R}$	0.38	0.47	0.57	0.62	0.64	0.54	0.61	0.75	0.71	0.81	0.43	3.97
α_q	-0.07	0.04	0.14	0.03	0.02	-0.05	-0.05	0.09	0.03	0.18	0.25	1.87
α_{q5}	0.04	-0.01	0.14	-0.03	0.01	-0.10	-0.12	0.01	-0.02	0.18	0.13	1.09
α_{FF5}	-0.14	-0.08	0.07	-0.03	-0.02	-0.07	-0.07	0.08	0.06	0.22	0.37	3.51
α_{FF6}	-0.11	-0.02	0.10	0.00	-0.01	-0.06	-0.07	0.07	0.05	0.18	0.29	2.84
12-month holding period												
$\bar{R}$	0.40	0.54	0.60	0.64	0.64	0.56	0.62	0.71	0.71	0.75	0.35	3.61
α_q	-0.02	0.10	0.15	0.06	0.02	-0.05	-0.02	0.08	0.04	0.13	0.16	1.25
α_{q5}	0.07	0.04	0.12	-0.02	-0.01	-0.08	-0.07	0.03	0.00	0.15	0.08	0.67
α_{FF5}	-0.11	-0.02	0.08	0.00	-0.01	-0.07	-0.03	0.07	0.06	0.17	0.29	2.80
α_{FF6}	-0.07	0.03	0.10	0.02	-0.01	-0.06	-0.03	0.06	0.05	0.14	0.21	2.10

Conclusion

- ML models generate more accurate predictions for firm-level growth than the conventional linear regression model
- Consistent with the investment CAPM, expected growth predicts stock returns positively controlling for past growth
- The ML-based expected growth measure can be useful in other applications