Interpreting Cross-Section Returns of Machine Learning Models: Firm Characteristics and Moderation Effect through LIME

Zequn Li, Xiaoxia Lou, Ying Wu, Steve Yang

zli61@stevens.edu

Saturday, 19 October 2024
FMA 2024 Annual Meeting
Session 243 - Machine Learning Application in Asset Pricing

Table of Contents

Section 1

Introduction: The Cross-Section of Expected Stock Returns

The Cross-Section of Expected Returns

Research Challenge

- Firm characteristics predict cross-section expected returns
- Interactions between firm characteristics are complex
- Ten of thousands of potential interactions in the "Factor Zoo"

Motivation

- Explore interactions between firm characteristics
- The predictive power varies across different contexts

The Cross-Section of Expected Returns: Methods

Portfolio Analysis

- Group stocks based on characteristics
- Test if long-short return is zero
- Cannot handle multiple characteristics

Regression Analysis

- Estimates relationship between firm characteristics and expected returns
- Test if coefficients are zero
- Handles multiple characteristics
- Issues with interactions and collinearity

Section 2

Methodology: Machine Learning and LIME(Local Interpretable Model-Agnostic Explanations)

Machine Learning in Empirical Asset Pricing

Predicting Expected Return as a function of Firm Characteristics[1]

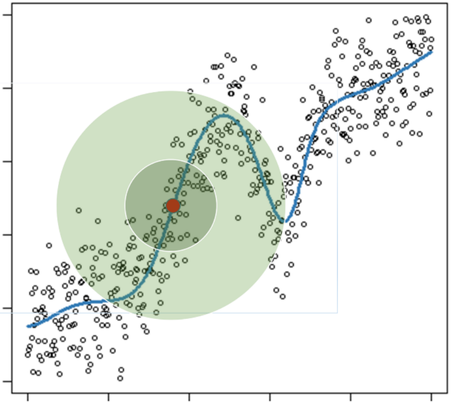
$$\mathbb{E}_t[r_{i,t+1}] = g(c_{i,t}) \tag{1}$$

- The function $g(\cdot)$ is consistent: Same form across firms and time periods
- Depends only on $c_{i,t}$: No information of the history and other stocks

Local Interpretable Model-agnostic Explanations(LIME)

1. **Select a Data Point**
2. **Perturb the Data Point**
3. **Predict perturbed samples**
4. **Fit local linear model**
5. **Interpret coefficients**

Local Interpretable Model-agnostic Explanations(LIME)



LIME: Local Interpretable Model-agnostic Explanations

Formally, we define an explanation of a prediction from a machine learning model g for stock i at time t as an interpretable linear model $g_{i,t}(z)$.

$$\mathbb{E}[g(z)] = g_{i,t}(z) \quad z \in \pi_{c_{i,t}} \tag{2}$$

$$g_{i,t}(z) = a_{i,t} + \sum_{k=1}^{K} b_{i,t}^{(k)} z^{(k)} \tag{3}$$

- z: perturbed Samples
- $\pi_{c_{i,t}}$: neighborhood of $c_{i,t}$
- $a_{i,t}$: intercept
- $b_{i,t}^{(k)}$: local coefficient of the k-th feature
- K: number of features

LIME Coefficient: A Comprehensive Firm Characteristic

- LIME coefficient is new firm characteristic
- Available for each stock i at each time t
- Aggregates information of nonlinearity and interactions

Empirical Methodology: LIME-Adjusted Moderation Regression

Baseline Regression

$$r_{i,t+1} = a + \delta_{k,t} c_{i,t}^{(k)} + \varepsilon_{i,t+1} \tag{4}$$

LIME-adjusted Moderation Regression

$$r_{i,t+1} = a + \delta_{k,t} c_{i,t}^{(k)} + \gamma_{k,t} b_{i,t}^{(k)} c_{i,t}^{(k)} + \xi_{k,t} b_{i,t}^{(k)} + \varepsilon_{i,t+1} \tag{5}$$

Double-Sort Portfolio Analysis: Methodology Overview

Baseline Method: Univariate Sort Portfolio Analysis

- Sorting stocks into 5 equal-weighted portfolios based on firm characteristics $\{\text{k}\}$.
- Going long on top quintile, short on bottom quintile.
- Exaiming the performance and risk of the long-short portfolios.

Bivariate Dependent Sort

- First sort by LIME local coefficients ($b_{i,t}^{(k)}$) into quintiles.
- Within each LIME group, sort by firm characteristics ($c_{i,t}^{(k)}$) into quintiles.
- Results in 5×5 equal-weighted portfolios.
- Create long-short portfolios within each LIME group.
- Evaluate the performance and risk of the long-short portfolios.

Section 3

Empirical Study of U.S. Equities

Predictive Models

Models

- **Naive Model**:Baseline; predicts zero excess return.
- **Linear Model**: Multivariate linear regression.
- **Neural Network(NN3)**: Three hidden layers; captures nonlinearities.
- **Random Forest(RF)**: Ensemble of decision trees

Interpretation Over Prediction

- Focus on interpretability, not prediction accuracy
- Experiments use NN3 and RF models
- Methodology adaptable to other machine learning models

Dataset and Training Methodologies

- **Sample Period and Data Split**:
  - Training: 1964–2021
  - Validation: 12 years rolling window
  - Testing: 1989 – 2021
- **Training Methodology**:
  - Annual retraining
  - Expanding training window
  - Rolling validation
- **Data Source**:
  - CRSP equity returns
  - Risk-free rates: Kenneth French Library
- **Firm Characteristics**:
  - Firm characteristics provided by Gu et al. (2020)
  - Ranked cross-sectionally and scaled to $[-1, 1]$
  - Missing values replaced by cross-sectional medians
  - The largest possible pool of assets

Direct Effects of Baseline Model

		agr	cfp_ia	chcsho	chempia	chinv	egr	grcapx	grltnoa
baseline	coef	0.66	0.32	-0.39	-0.29	-0.34	-0.45	-0.34	-0.42
	t-value	**5.01**	**3.11**	**-3.79**	**-3.59**	**-4.44**	**-3.74**	**-4.34**	**-4.28**
NN3	coef	0.36	0.36	-0.41	-0.19	-0.28	-0.28	-0.36	-0.34
	t-value	**3.43**	**3.96**	**-3.42**	-2.45	**-4.04**	-2.58	**-4.50**	**-3.85**

		hire	indmom	invest	lgr	mom1m	pctacc	sgr	sp
baseline	coef	-0.40	0.41	-0.41	-0.36	-0.53	-0.24	-0.39	0.52
	t-value	**-4.56**	**3.27**	**-4.02**	**-5.57**	**-4.52**	**-3.29**	**-5.40**	**3.19**
NN3	coef	-0.25	0.45	-0.36	-0.29	0.67	-0.22	-0.33	0.39
	t-value	**-3.00**	2.91	**-4.22**	**-4.43**	**3.45**	**-3.21**	**-4.71**	2.39

Additional Direct Effect identified from NN3 model

		acc	dolvol	maxret	mom12m	mvel1
baseline	coef	-0.33	-0.38	-0.01	0.32	-0.39
	t-value	-2.90	-2.80	-0.06	1.70	-2.19
NN3	coef	-0.39	1.00	0.74	0.69	0.77
	t-value	**-3.46**	**4.73**	**3.13**	**4.03**	**4.43**

Moderation Effects of Firm Characteristics Identified by NN3 Model

	NN3 Direct Effect		NN3 Moderation Effect	
	coef	t-value	coef	t-value
betasq	0.08	0.31	-1.54	**-4.07**
chmom	0.24	2.28	0.53	**4.30**
dolvol	1.00	**4.73**	2.46	**4.22**
ep	0.33	1.25	-1.80	**-3.40**
ill	-0.12	-0.78	-1.15	**-3.25**
maxret	0.74	3.13	1.02	**3.59**
mom12m	0.69	**4.03**	1.18	**3.34**
mom1m	0.67	3.45	1.00	**4.20**
mom6m	-0.29	-1.69	2.27	**5.88**
mvel1	0.77	**4.43**	1.78	**4.88**
retvol	0.95	**3.62**	2.02	**4.18**
std_dolvol	0.03	0.30	-1.69	**-3.57**
turn	0.58	2.88	0.83	**4.31**
operprof	0.41	1.98	-1.84	**-3.18**

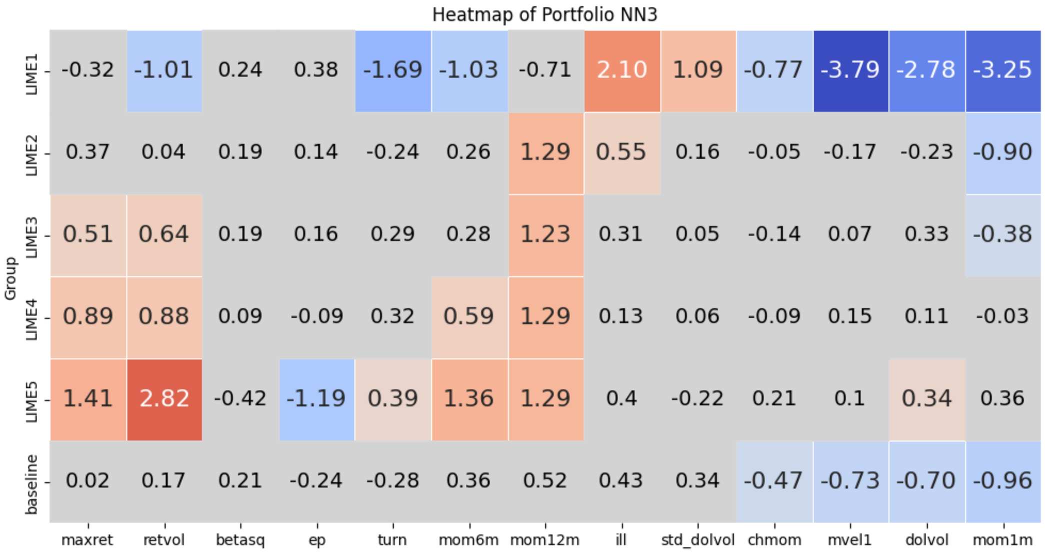
Portfolio Analysis: 6-month Momentum

Portfolio Group	Full Asset Pool	Filtered Sample	LIME 5	LIME 1
Mean Return (HML)	0.36	1.03	1.36	-1.03
t-stats (HML)	1.40	4.21	4.40	-3.13

Key Insights

- Full sample without filtering shows weak momentum
- Filtered Sample performs better than the full asset pool
- LIME5 group shows the strongest momentum
- LIME1 group shows a reversal Pattern

Key Patterns in Portfolio Analysis – NN3 Model



Contribution

Testing Moderation Effect

- Validates interaction statistically
- Quantifies moderation significance

Out-of-Sample Testing with Portfolio Analysis

- Tests predictions with portfolios
- Confirms practical applicability

LIME Coefficient as a Predictive Tool

- Serves as a new firm characteristic.
- Guides portfolio decisions

Conclusions

Innovative Methodological Framework

Integrated Machine Learning with LIME to analyze moderation effects

Enhanced Model Interpretability

Uncovered how firm characteristics interact to predict stock returns.

Empirical Findings

Identified significant firm characteristics predict stock returns through moderation

Practical Implications

Improved investment strategies via bivariate dependent sort portfolio analysis

Thank You!

Comments and Questions are Welcome!

Section 4

Appendix

Out-of-Sample Performance: R^2

$$R^2_{oos} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1})^2} \tag{6}$$

- $\mathcal{T}_3$ is the set of testing samples, where the data never enter into model estimation or tuning.
- Different from the traditional R^2 measure, the denominator of the out-of-sample R^2 is the sum of squared excess returns without demeaning. We compare the model with the naive forecast of zero.

Out-of-Sample Performance

Table: Out-of-Sample Predictive Performance(percentage R^2)

	Naive	FM	FM5	NN3	RF
Full Sample	0.00	-1.02	-0.77	0.49	0.41
Common Stock, lag price > 2	0.00	-1.37	-1.04	0.27	0.26
Size top 1000	0.00	-1.72	-1.36	0.61	1.09
Size bottom 1000	0.00	-0.59	-0.89	-0.49	0.18

Out-of-Sample Prediction Sorted Portfolios I

- We sort all firms into 10 portfolios based on their model-predicted returns and compute the holding period equal-weighted returns for each portfolio.

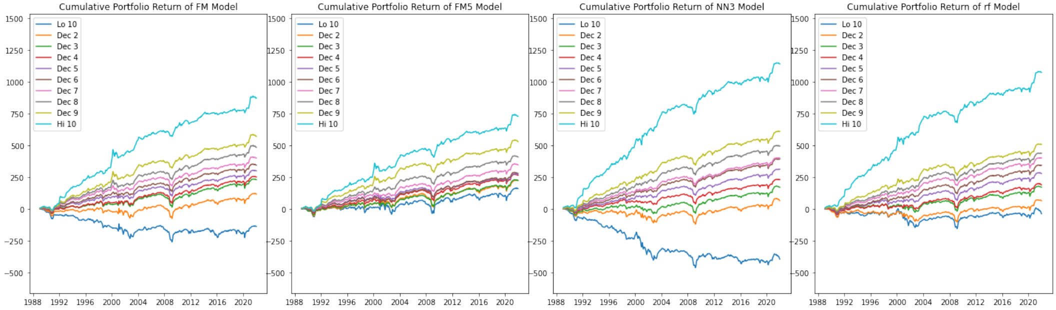


Figure: Machine Learning Portfolio(Full Sample)

Out-of-Sample Prediction Sorted Portfolios II

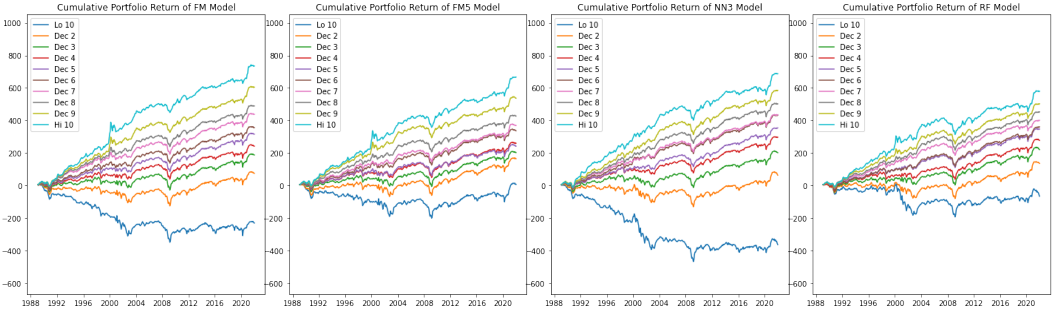
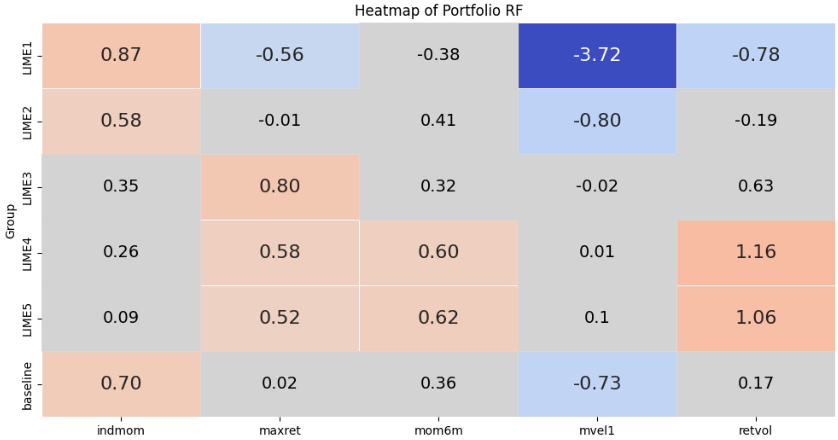


Figure: Machine Learning Portfolio(Common Stock, Lag Price > 2)

Key Patterns in Portfolio Analysis – RF Model



Number of Significant Firm Characteristics Across Models