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

Zequn Li, Xiaoxia Lou, Ying Wu, Steve Yang Stevens Institute of Technology

Machine Learning in Asset Pricing

 $\mathbb{E}_t[r_{i,t+1}] = f^*(c_{i,t})$

What is LIME?

Local Interpretable Model-agnostic Explanations

$$\mathbb{E}[f(z)] = g_{i,t}(z) \quad z \in \pi_{c_{i,t}}$$
$$g_{i,t}(z) = a_{i,t} + \sum_{k=1}^{K} b_{i,t}^{(k)} z^{(k)}$$



How is LIME Different from Feature Importance?

- Feature Importance
- One importance score for each characteristic
- Measure overall predictive power
- Static across all stocks and time periods
- LIME Coefficients
- Dynamic interpretation for each stock-time pair
- Measure direction and magnitude of effects
- Enables studying full distribution of characteristic effects

The Predictive Power of Firm Characteristics Varies among Stock Groups, which can be identified with LIME Interpretations from Machine Learning Models.

Full Sample -	0.43	-0.24	0.34	-0.09	0.22	-0.47	0.02	0.51	-0.28	0.36	-0.70	-0.96	0.17	-0.73
Filtered Sample -	-0.17	0.48	0.06	0.47	-0.11	-0.06	-0.81	1.18	-0.20	1.03	-0.10	-0.46	-0.80	0.06
LIME1 -	2.10	0.39	1.10	0.40	0.23	-0.77	-0.33	-0.72	-1.69	-1.03	-2.80	-3.24	-1.01	-3.79
LIME2 -	0.55	0.15	0.16	0.29	0.19	-0.06	0.37	1.30	-0.24	0.26	-0.24	-0.91	0.03	-0.17
LIME3 -	0.31	0.16	0.05	0.12	0.19	-0.14	0.51	1.23	0.29	0.27	0.33	-0.39	0.64	0.07
LIME4 -	0.13	-0.09	0.06	-0.09	0.10	-0.09	0.89	1.30	0.33	0.59	0.11	-0.03	0.87	0.15
LIME5 -	0.39	-1.19	-0.23	-0.66	-0.42	0.21	1.62	1.29	0.39	1.36	0.34	0.36	2.82	0.10
LIME5 - LIME1 -	-1.70	-1.58	-1.33	-1.06	-0.64	0.97	1.95	2.01	2.08	2.39	3.14	3.60	3.83	3.89
	il ep dovol operprof betasa dimon naxiet nomi2m turn momon dovol momilm retual muelt													



Take a picture to •

download the full paper

- ences returns(R_{t+1})
- returns(R_{t+1})

Empirical Methodology

LIME-adjusted Moderation Regression

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

Bivariate Dependent Sort

- $(c_{i,t}^{(k)})$ into quintiles.

Empirical Findings



What is Moderation Effect?

• **Direct Effect**: How one characteristic(X_t) directly influ-

- Example: High Momentum \implies High Return

• Interaction Effect: How two characteristics (X_t, Y_t) jointly affect returns (R_{t+1})

- Example: Size and Value together influence returns differently than each alone

• Moderation Effect: How one variable (M_t) changes the effectiveness of another characteristic's (X_t) on

- Example: The predictive power of momentum varies across different LIME group

• First sort by LIME local coefficients $(b_{i,t}^{(k)})$ into quintiles. • Within each LIME group, sort by firm characteristics

• Results in 5×5 equal-weighted portfolios.

• Create long-short portfolios within each LIME group.

• Evaluate the performance of the long-short portfolios.

• 8 firm characteristics show robust predictability, unaf-

fected by sample filtering

• 14 firm characteristics show varied predictability

• Strongest Positive effects in LIME5 group

• Strongest Negative effects in LIME1 group