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

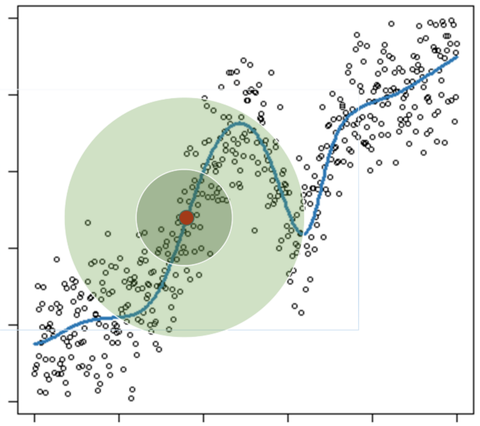
$$\mathbb{E}_t[r_{i,t+1}] = f^*(c_{i,t})$$

What is LIME?

Local Interpretable Model-agnostic Explanations

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

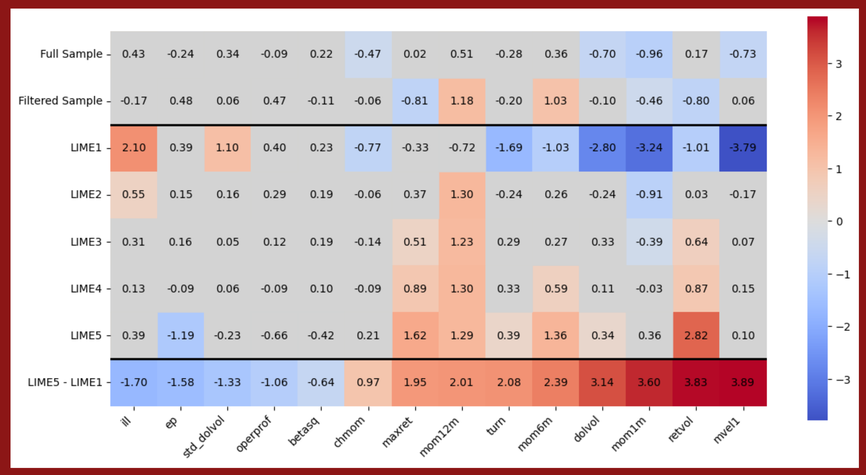
$$g_{i,t}(z) = a_{i,t} + \sum_{k=1}^{K} b_{i,t}^{(k)} z^{(k)}$$



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.



Take a picture to download the full paper

What is Moderation Effect?

- **Direct Effect**: How one characteristic(X_t) directly influences returns(R_{t+1})
  - Example: High Momentum $\Longrightarrow$ 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 returns(R_{t+1})
  - Example: The predictive power of momentum varies across different LIME group

Empirical Methodology

LIME-adjusted Moderation Regression

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

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 of the long-short portfolios.

Empirical Findings

- 8 firm characteristics show robust predictability, unaffected by sample filtering
- 14 firm characteristics show varied predictability
- Strongest Positive effects in LIME5 group
- Strongest Negative effects in LIME1 group

STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®