Building Test assets via Machine Learning

汇报人：赵伟皓

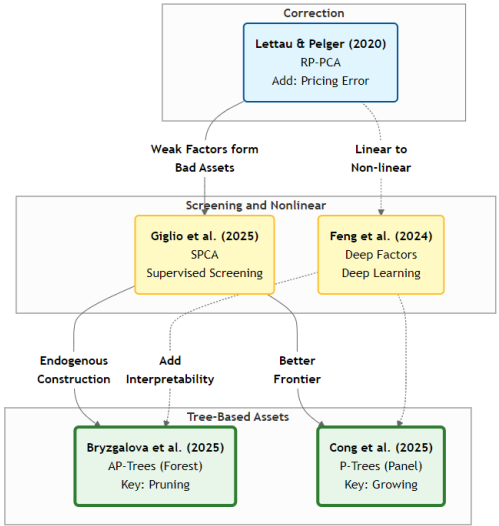
武汉大学金融系

2026.01.27

Building Test asset as a New Paradigm

1. **Challenges in Current Asset Pricing Research**
   - **ML Objectives**: The goal of standard ML is prediction accuracy (min MSE).
   - **Optimizing individual stocks is not feasible**:
     - Thousands of stocks(High noise and difficulty in finding the inverse of $\hat{\Sigma}$).
   - **Limitations of Test Assets**: Single/Double sorting ignores high-dimensional nonlinearity and asymmetric interactions between features.
2. **Theme**:
   - Using machine learning (especially decision trees) to reconstruct test assets.
   - From simple sorting to non-linear interaction, capturing complex pricing information.

Solutions for pricing high-dimensional charateristics and assets



AP-Trees and P-Trees

Comparision	**Bryzgalova et al. (2025)**	**Cong et al. (2025)**
Model	AP-Trees	P-Trees
Method	**Pruning**	**Growing**
Criterion	Global SDF Spanning	Global Split Criteria
Objective	Maximize sharpe ratio	Maximize sharpe ratio
Implemention	Pruning the branches that are ineffective for pricing	Splitting at each step to pursue utility maximization

总结

1. **实施方式：Pruning vs. Growing**
   - Bryzgalova 的方法像是在修剪盆景，把多余的树叶剪掉，剩下最具信息含量的（SDF）；
   - Cong 的方法像是盖楼，每一层都要确保稳固且有价值（Sharpe Ratio），才会盖下一层。
2. **结论**
   - 传统的 Fama-French 2x3 或 5x5 排序已经无法适应高维特征场景；
   - 三重排序虽然可以捕捉交互关系，但受维度限制，无法以用于多特征情景；
   - 基于树构建测试资产的方法确保了捕捉大数据时代下的复杂市场规律。
3. **拓展：”Grow-then-Prune” Method**
   - Grow: 利用 Random P-Tree 的随机性，克服贪婪算法的短视，生成一个足够大的“候选有效前沿组合森林”。
   - Prune: 利用 AP-Tree 的正则化剪枝思想，从这个池子中通过全局夏普比率最大化，挑选出最稀疏、最有效的组合。