Essence of the Cross Section
Seyfi

Paul Huebner

Stockholm School of Economics

Discussion
PhD Nordic Finance Workshop
May 2024

Which characteristics describe the cross-section of expected stock returns?

- "Classic approach": sort stocks based on characteristics (from theory, sometimes), then see if sort lines up with subsequent returns
- With many predictors (which we do have), multivariate sorts become infeasible

$\Rightarrow$ **Need for methods capable of dealing w/ large complexity & high dimensionality**

Which characteristics describe the cross-section of expected stock returns?

- "Classic approach": sort stocks based on characteristics (from theory, sometimes), then see if sort lines up with subsequent returns
- With many predictors (which we do have), multivariate sorts become infeasible

⇒ **Need for methods capable of dealing w/ large complexity & high dimensionality**

- This paper: find the characteristics that vary across stocks with high vs low average returns in the past
  - method separates spread in predictors lining up with returns from spurious variation
  - this is not just momentum, it's about learning the mapping between characteristics and returns from past data

How it works

1. *Performance sort:* At each $t - i$, sort stocks into quintiles based on realized returns ($\forall i > 0$ in last 10 years)

2. *Dimension reduction:* Within each quintile, compute the portfolio average for each (of many) characteristics

3. *Prediction:* Going forward, sort stock into high expected return portfolio if its characteristics profile is "closest" to that of stocks with high past returns

4. *Outcomes:* Compare out-of-sample performance of high vs low expected return portfolios

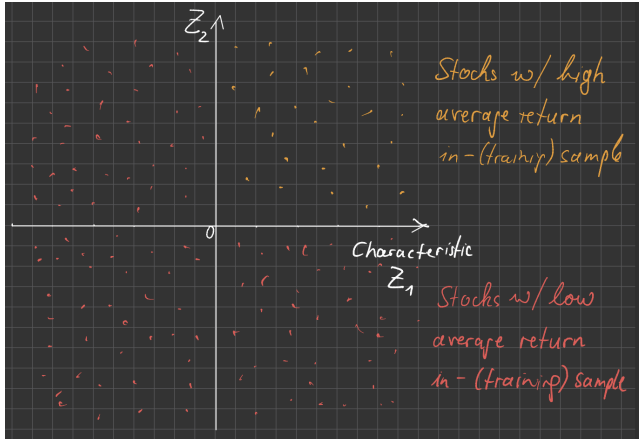
Results

- Works remarkably well!
  - High-low expected return portfolios generate out-of-sample monthly α of 1.42%
  - Annualized SR of 1.81

Findings:

- How many characteristics matter? 15 (80) explain 50% of difference between low and high return portfolios
- Which characteristics matter? Variations of momentum and idiosyncratic volatility
- Performance weaker after 2000

Example 1 - Setup



Data Generating Process with *interacting predictors*:

- Both predictors $Z_1 > 0$ & $Z_2 > 0 \Rightarrow$ high expected return
- Otherwise $\Rightarrow$ low expected return

Example 1 - Classic Methods

"Classic" methods univariately sort stocks by characteristic Z_1 and Z_2

- Doing this independently for each predictor misses their *interaction*
- Large Z_1 or Z_2 independently creates a spread in expected returns
- Fails to perfectly separate orange vs 3x red, instead orange + red vs 2x red
  - Accuracy: 75%

Example 1 - Classic Methods

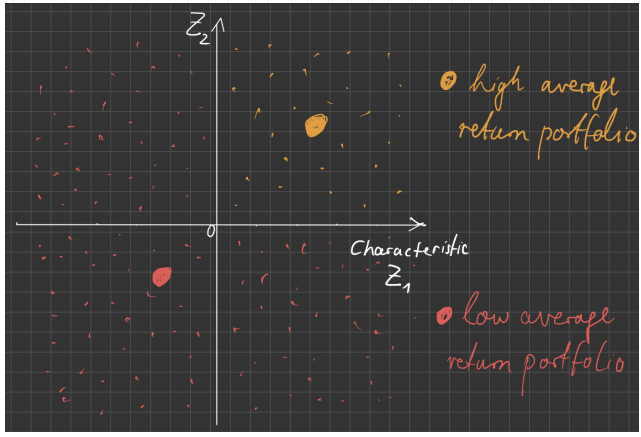
"Classic" methods univariately sort stocks by characteristic Z_1 and Z_2

- Doing this independently for each predictor misses their *interaction*
- Large Z_1 or Z_2 independently creates a spread in expected returns
- Fails to perfectly separate orange vs 3x red, instead orange + red vs 2x red
  - Accuracy: 75%

Double-sorting recovers the true data-generating process (in this example)

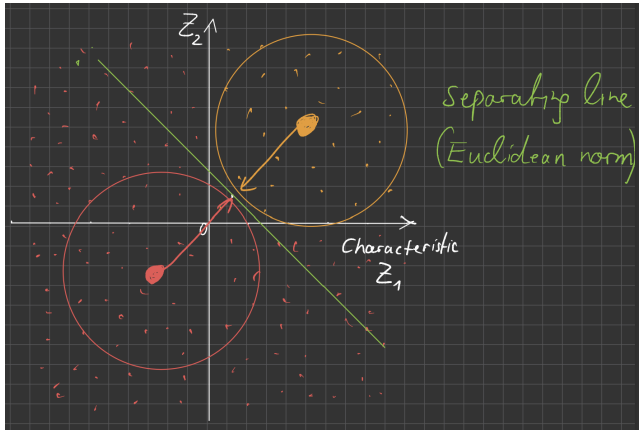
- impractical N-sort with high-dimensional data
- irrelevant characteristics generally affect sorting procedures

Example 1 - New Method



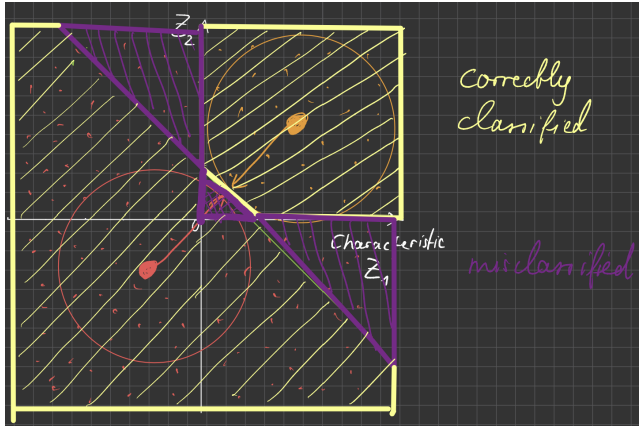
1. *Performance sort:* Separate stocks into the orange and red groups
2. *Dimension reduction:* Compute the average of Z_1 and Z_2 for each group

Example 1 - New Method



3 *Prediction:* Stocks below (above) the separating hyperplane minimize the Euclidean norm by being assigned to the low (high) return portfolio

Example 1 - New Method



- After some geometry, you find Accuracy $= \frac{237}{288} \approx 82.3\%$

Example 1 - Put together

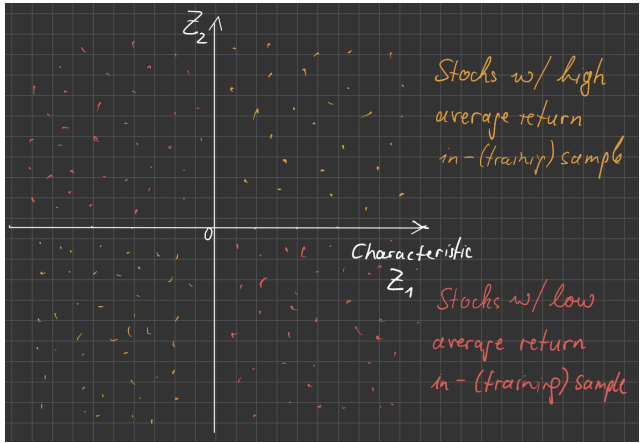
- New method is meaningfully better than univariate sorts (82.3% vs 75%)
- Worse than double-sorting (100%)
- Probably worse than ML techniques
  - e.g., decision trees and random forests
  - arbitrarily deep neural networks can approximate any function
- Method (mostly) successful in this example

Example 1 - Put together

- New method is meaningfully better than univariate sorts (82.3% vs 75%)
- Worse than double-sorting (100%)
- Probably worse than ML techniques
  - e.g., decision trees and random forests
  - arbitrarily deep neural networks can approximate any function
- Method (mostly) successful in this example

Broader questions: can you formalize when your method works well? What assumptions about the data generating process are needed?

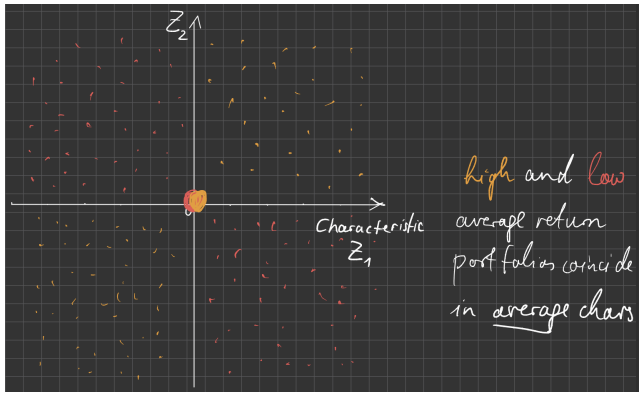
Example 2 - Setup



Data Generating Process:

- Both predictors $Z_1 > 0$ & $Z_2 > 0$ **OR** $Z_1 < 0$ & $Z_2 < 0 \Rightarrow$ high expected return
- Otherwise $\Rightarrow$ low expected return

Example 2 - Classic & New Methods



- Every stock is equally far from the two portfolios (50% Accuracy $=$ random guess)
- Same for univariate sorts, but double-sorting still works perfectly

Example 2 - Many Roads ...

- This sounds bad ... Why does that happen?

Example 2 - Many Roads ...

- This sounds bad ... Why does that happen?

More than one road leads to Rome!

- There is more than one unique combination of characeristics that predicts good performance
  - in the example, $Z_1 > 0$ & $Z_2 > 0$ OR $Z_1 < 0$ & $Z_2 < 0$
  - "averaging" characteristics across both loses valuable information

Example 2 - Many Roads ...

- This sounds bad ... Why does that happen?

More than one road leads to Rome!

- There is more than one unique combination of characeristics that predicts good performance
  - in the example, $Z_1 > 0$ & $Z_2 > 0$ OR $Z_1 < 0$ & $Z_2 < 0$
  - "averaging" characteristics across both loses valuable information
- **Easily addressed via a simple refinement to the dimension reduction step!**

Example 2 - Refinement

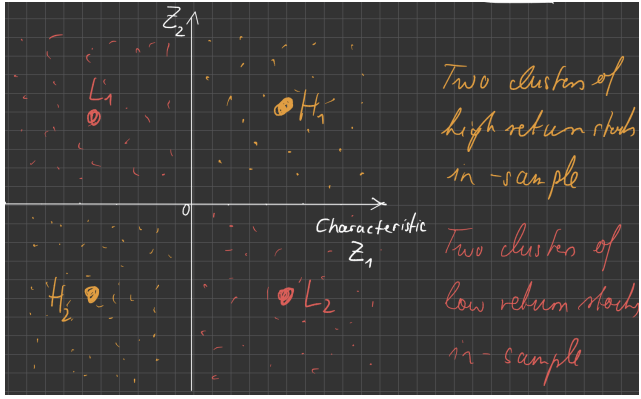
1. *Performance sort:* At each $t - i$, sort stocks into quintiles based on realized returns ($\forall i > 0$ in last 10 years)

2. *Dimension reduction:* Within each quintile, use K-means clustering to isolate different characteristics combinations and average characteristics within each (i.e., the K means)

3. *Prediction:* Going forward, sort stock into high expected return portfolio if its characteristics profile is "closest" to that of any of the different characteristics combinations of stocks with high past returns

4. *Outcomes:* Compare out-of-sample performance of high vs low expected return portfolios

Example 2 - Refinement



- Accuracy for $K = 2$ means for high and low return stocks is now 100%
- The same for Example 100 by allowing $K = 3$ means for low return stocks
- Refinement should be able to handle most (all?) difficulties faced by the proposed method

Small comments

1. 10 years of past data to get characteristics separating low & high expected return stocks
   - seems long when mapping between characteristics and expected returns is very dynamic
   - for example, it might vary across the business cycle
2. Weak performance of high-minus-low expected return stocks post-2000
   - Is there just no variation in expected returns across stocks anymore?
   - Possibly the result of many characteristics that worked in the past (e.g., momentum) working less well, and 10-year window is slow to adjust
   - Potentially different set of characteristics that matter but have not been discovered yet
3. Naturally high amount of costly trading
   - Also consistent with weaker performance in the recent era of lower trading costs

Small comments

- Methodology can cope with large (potentially infinite) number of predictors
  - So why limit yourself to characteristics that are "known" from previous literature?
  - For example, think additional "modulators" affecting the relationship between known predictors and returns, but have not been discovered because not predictive on their own
- Missing characteristics are imputed by their median values (Gu, Kelly, and Xiu, 2020)
  - Recently, better methods have emerged (e.g., Bryzgalova, Lerner, Lettau, and Pelger, 2024; Freyberger, Höppner, Neuhierl, and Weber, 2024)
- Relation to the existing literature on Big Data and dimensionality reduction
  - For example, Lettau (2023) tensor-PCA for creating portfolios that separate characteristics
  - This creates lots of variation in expected returns despite potential presence of spurious signals

Conclusion

- Interesting paper!
- Studying the characteristics of stocks with high versus low expected returns in the past, rather than the opposite direction, is intuitive and simple
- It works well in the data
- Potential to improve on the "dimensionality reduction" beyond simple averaging for cases when many roads lead to Rome

Best of luck for the job market...
... and don't forget to apply at SSE!