MGMT 638

Session 5

Kerry Back
Fall 2025

Agenda

1. Price and marketcap filters
2. Standardizing predictors (features)
3. Composite ranks
4. Random forests
5. Standardizing returns
6. Safe minus risky strategy

Penny Stocks

- Must filter out low-price stocks. Infeasible for equally weighted portfolios and distort portfolio returns.
- Use CQA Competition filter: $5.60?
- Example. Use data from last class. Tell Claude to read it and to drop all rows with close $\leq$ price threshold. Tell Claude to sort into quintiles each month on momentum and lagret and to compute the average return within each of the 25 groups each month.

Marketcap Filters

- If we were managing a fund, we would specify in our prospective the universe of stocks in which we would invest.
- We would likely include some marketcap filter. For example, large cap = S&P 500, large and midcap = Russell 1000, smallcap = Russell 2000, etc.
- It is hard for large funds to trade microcaps, so they will exclude them even if the prospectus does not specify.

Standardizing Features

- Many models, like linear regression, are highly affected by outliers.

- Standardizing predictive variables is common. For example, subtract mean and divide by standard deviation. This is called a 'z-score.'

- Ranking 1 through n is another way to standardize. If n varies, then can scale to 0 to 1 or (Quality Minus Junk), compute z-scores of ranks.

- We should do this each month rather than for the full sample.

- Important: For 'good' features (like football scores), rank from low to high. For 'bad' features (like golf scores), rank from high to low. Then a high rank is always a good value.
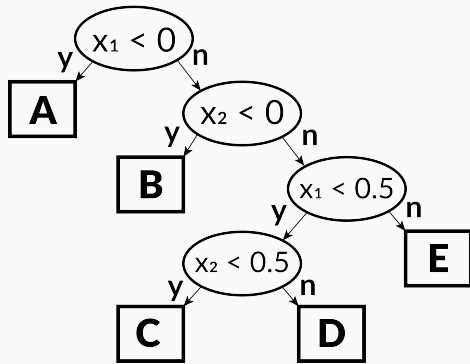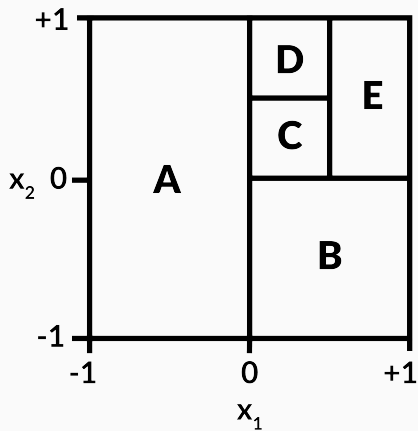
Composite Ranks

The Quality Minus Junk paper computes quality as follows.

1. Compute z-scores of features each month.
2. Group features into three categories: profitability, growth, and safety.
3. Within each category, add z-scores of features in that category each month. Compute the z-score of the sum each month.
4. Add the three category z-scores together each month. Compute the z-score of the sum each month.

Decision Trees

- Start with the estimate $\hat{y} = \bar{y}$ for all observations.
- Split into two subsets based on one variable and one threshold. All observations below the threshold go into one group. All above go into another.
- Prediction for each group is the group mean of the target variable. Calculate MSE (mean squared error) over both groups.
- Choice of variable and threshold on which to split was based on minimizing MSE.
- Split each subset into further subsets and continue.

Small Example with Data

- Tell Claude to use only the data for a single month. Let's use Jan, 2018.
- Tell Claude to fit a decision tree with a maximum depth of three to predict the return from momentum, lagret, and book-to-market.
- Tell Claude to generate an image of the fitted tree.

Random Forests

- A forest is multiple trees.
- For any observation - old or new - each tree makes a prediction.
- Average the predictions to get the final prediction.
- A random forest is created by generating random datasets and fitting a tree to each.
- A random dataset is generated by randomly drawing rows from the original dataset.

Standardizing Returns

- In most situations, it is uncommon to standardize the dependent variable.
- But if we want to pick stocks, we don't need to forecast returns. We just need to forecast ranks. Ranks are probably easier to forecast.
- So we could rank returns and compute z-scores of ranks each month to form our target variable.

Full Example: Fitting

- Tell Claude to compute ranks of stocks each month on (i) momentum in ascending order, (ii) lagret in descending order, and (iii) book-to-market in ascending order.
- Tell Claude to compute ranks of stocks each month on return in ascending order.
- Tell Claude to compute z-scores of the ranks each month.
- Tell Claude to fit a random forest using data through 2015 to predict the z-score of return ranks from the z-scores of the feature ranks. Tell Claude to use 100 trees and a maximum depth of 4.

Full Example: Predicting and Evaluating

- Tell Claude to use the fitted model to predict the target variable for each stock each month from 2016 on.
- Tell Claude to sort stocks into deciles each month based on the predicted values.
- Tell Claude to compute the average return of each group each month.
- Tell Claude to compute the mean , standard deviation, and mean-to-standard-deviation ratio of each decile return series.

Safe Minus Risky

- PDF: *Safe Minus Risky*, Kapadia, Ostdiek, Weston, and Zekhnini, 2019
- Video: *Safe Minus Risky*, Kapadia, Ostdiek, Weston, and Zekhnini, 2019
- NotebookLM: Safe Minus Risky

https://learn-investments.rice-business.org/capm/
sml-industries