Outline

● Introduction to cross-sectional returns trading strategy
● Our data universe
● Modeling techniques and results
  ○ Stepwise Regression
  ○ Principal Component Analysis (PCA)
  ○ Using historical returns
  ○ Neural Network (for validation of factors)
● Portfolio construction and results
● Conclusion and discussion
Overview
Introduction to Cross-Sectional Returns

- Cross-section of expected returns should be explained by exposure to systematic risk factors
- Factor investing harvests these risk premia through exposure to factors
- Finding the “right” factors has become the central question of asset pricing
- Harvey et al. (2016): 300 published candidate factors have predictive power for the cross-section of expected returns.
- Cochrane (2011): “factor zoo” leads to the questions of which risk factors are really important and which factors are subsumed by others.
Our Data Universe

- Focus on constituents of S&P 500 Index
- Collected company specific data/characteristics (e.g. market capitalization, beta, price-to-earnings ratio etc.) from 2011 to 2018 using WRDS, and also macroeconomic data from the same time period.
- Avoid survivorship bias by updating the constituents annually
Correlation of the factor universe

- 181 factors initially
- Clusters of factors are highly correlated
- Many are absolute values that are meaningless without context (e.g. earnings of $1 million for a $1 billion company is different than that for a $100 billion company)
- Some are factors that we decided to exclude in cross-sectional analysis (e.g. company location and dividend currency)
Decided to reduce to a list of well-researched factors

<table>
<thead>
<tr>
<th>Systematic Factors</th>
<th>Excess returns for companies that have:</th>
<th>Captured By</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Low prices relative to their fundamental value</td>
<td>Book Value to Market Value, CF Margin, CFO to Total Liabilities</td>
</tr>
<tr>
<td>Low Size</td>
<td>Smaller market capitalisation</td>
<td>Market Cap</td>
</tr>
<tr>
<td>Momentum</td>
<td>Stronger past performance</td>
<td>Relative returns (Lagged 1 Period, SMA of 3/6 months)</td>
</tr>
<tr>
<td>Low Volatility</td>
<td>Lower than average volatility / idiosyncratic risk</td>
<td>Relative Trading Volume</td>
</tr>
<tr>
<td>Dividend Yield</td>
<td>Higher-than-average dividend yields</td>
<td>Dividend Yield</td>
</tr>
<tr>
<td>Quality</td>
<td>Low debt, stable earnings growth</td>
<td>Earnings to Price, Returns on Equity, Financial Leverage</td>
</tr>
</tbody>
</table>
Stepwise Linear Regression
Factor Selection using Stepwise Linear Regression

Step 1: Regress stock returns on market returns to get residuals $\varepsilon_{i,t}$.

$$r_{i,t} = \alpha_i + \beta_{m,i} r_{m,t} + \varepsilon_{i,t}$$

Step 2: Regress residuals (net of market) on normalized factors to look at magnitude of significant factors at 95% confidence interval to select factors

$$\varepsilon_{i,t} = \alpha'_i + \beta'f_{i,t} + v_{i,t}$$

- $\alpha_i$: intercept of asset $i$
- $f_{i,t}$: factor variables
- $\beta' = (\beta_1, \ldots, \beta_K)'$: factor loadings for $k$ factors
Factor Selection using Stepwise Linear Regression

Step 3: Rerun linear regression on unnormalized values of selected factors and obtain the beta coefficients.

Step 4: Use the factors and beta coefficients to predict returns on validation and test set. The accuracy of returns (in %) is measured by RMSE:

\[ RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Predicted_i - Actual_i)^2} \]
Factor Selection using Stepwise Linear Regression

### Significant factors:

- **All**: Excess Monthly Return:
  - Momentum (SMA3)
  - Return On Equity (ROE)
  - Relative Trading Volume (RTV)
- **Group by Industry**:
  - Momentum (SMA3)
  - Return On Equity (ROE)
  - Relative Trading Volume (RTV)
  - BV/MV
- **Group by Company**:
  - Momentum (SMA3)
  - Return On Equity (ROE)
  - Earnings to Price
  - BV/MV

### RMSE of predicted returns (%)

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>INDUSTRY</th>
<th>COMPANY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td>5.5904</td>
<td>5.5754</td>
<td>5.5325</td>
</tr>
<tr>
<td><strong>Validation</strong></td>
<td>5.3061</td>
<td>5.3386</td>
<td>5.7908</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td>5.5150</td>
<td>5.5549</td>
<td>6.4716</td>
</tr>
<tr>
<td><strong>Ridge regression (Test)</strong></td>
<td>5.5143</td>
<td>5.5527</td>
<td>6.4708</td>
</tr>
</tbody>
</table>

*Return = \( \beta_1 \cdot SMA3 + \beta_2 \cdot ROE + \beta_3 \cdot RTV + \beta_M \cdot S&P \)*

\[
\beta_1 = 0.9794 \\
\beta_2 = -0.1127 \\
\beta_M = 1.132 \\
\beta_3 = 6.9738 \times 10^{-6}
\]
Principal Component Analysis
Factor Selection using PCA

- Dimension reduction technique used to explain the majority of information in a sample covariance matrix
- PCs are constructed and ordered so first PC explains the largest portion of the sample cov matrix
- PCs are orthogonal and normalized to have unit lengths
- Aim: Find PC’s that explain the largest variance in sample covariance matrix, and find correlations of these PCs with excess returns
Factor Selection using PCA

- Conduct PCA for
  - All observations
  - Top + bottom decile

- Include 10 normalized factors in sample covariance matrix
  - BV/MV, Market Cap, Earnings/Price, Div Yield, ROE, FinLeverage, P/CF, Cash Flow Margin, CFO/Total Liab, Relative Trading Volume, Momentum represented by lagged (1 Period) excess return / SMA 3 month excess returns / SMA 6 months excess returns
Results: When including all observations, excess returns are correlated with small caps and positive momentum

- First 4 PCs explain >85% of variance
- Excess returns correlated with negatively with PC2, and positively with PC3 and PC4
Results: When including all observations, excess returns seem correlated with small caps and positive momentum

- Negatively correlated with PC2 (growth stocks + positive momentum do better)
- Positively correlated with PC3 (small caps do better)
- Positively correlated with PC4 (value stocks + positive momentum do better)

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>BVOverMV</td>
<td>-0.056</td>
<td>0.632</td>
<td>-0.251</td>
<td>0.723</td>
</tr>
<tr>
<td>Market Cap</td>
<td>-0.095</td>
<td>-0.51</td>
<td>-0.842</td>
<td>0.147</td>
</tr>
<tr>
<td>EarningsToPrice</td>
<td>0.006</td>
<td>-0.021</td>
<td>0.004</td>
<td>0.005</td>
</tr>
<tr>
<td>DivYield</td>
<td>-0.006</td>
<td>0.014</td>
<td>-0.007</td>
<td>0.002</td>
</tr>
<tr>
<td>ROE</td>
<td>0.002</td>
<td>-0.01</td>
<td>-0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td>FinLeverage</td>
<td>0.001</td>
<td>-0.002</td>
<td>0.002</td>
<td>-0.001</td>
</tr>
<tr>
<td>PtoCFRatio</td>
<td>0.005</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.003</td>
</tr>
<tr>
<td>CashFlowMargin</td>
<td>-0.983</td>
<td>0.001</td>
<td>0.108</td>
<td>-0.018</td>
</tr>
<tr>
<td>CFOtoTotalLiab</td>
<td>-0.141</td>
<td>-0.059</td>
<td>0.036</td>
<td>-0.089</td>
</tr>
<tr>
<td>TradingVolume_relative</td>
<td>0</td>
<td>-0.002</td>
<td>-0.005</td>
<td>0.002</td>
</tr>
<tr>
<td>ExcessMonthlyReturn_SMA3</td>
<td>0.036</td>
<td>-0.58</td>
<td>0.464</td>
<td>0.669</td>
</tr>
</tbody>
</table>
Results: For top/bottom decile, there is a clearer trend of excess returns with positive momentum and value stocks

- First 4 PCs explain >85% of variance
- Excess returns is positively correlated with PC1 and PC3, and negatively correlated with PC2
Results: For top/bottom decile, there is a clearer trend of excess returns with positive momentum and value stocks

- Positively correlated with PC1: Low Cash Flow Margin
- Strongly negatively correlated with PC2: Positive momentum
- Positively correlated with PC3: Value Stocks

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>BVOverMV</td>
<td>-0.122</td>
<td>0.441</td>
<td>0.867</td>
<td>0.136</td>
</tr>
<tr>
<td>MarketCapInverse</td>
<td>-0.030</td>
<td>-0.082</td>
<td>-0.107</td>
<td>0.988</td>
</tr>
<tr>
<td>EarningsToPrice</td>
<td>0.018</td>
<td>-0.028</td>
<td>-0.002</td>
<td>0.029</td>
</tr>
<tr>
<td>DivYield</td>
<td>-0.004</td>
<td>0.015</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>ROE</td>
<td>0.007</td>
<td>-0.005</td>
<td>-0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>FinLeverageInverse</td>
<td>0.001</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.004</td>
</tr>
<tr>
<td>PtoCFRatiolInverse</td>
<td>0.002</td>
<td>0.001</td>
<td>-0.003</td>
<td>0.008</td>
</tr>
<tr>
<td>CashFlowMargin</td>
<td>-0.969</td>
<td>-0.134</td>
<td>-0.028</td>
<td>-0.055</td>
</tr>
<tr>
<td>CFOtoTotalLiab</td>
<td>-0.184</td>
<td>-0.087</td>
<td>-0.144</td>
<td>0.029</td>
</tr>
<tr>
<td>TradingVolume_relative</td>
<td>-0.002</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.004</td>
</tr>
<tr>
<td>ExcessMonthlyReturn_SMA3</td>
<td>0.107</td>
<td>-0.879</td>
<td>0.464</td>
<td>-0.020</td>
</tr>
</tbody>
</table>
Using Historical Returns
Factor Selection using Historical Returns

- For each factor (e.g. size, earnings to price), rank the stocks according to that factor: long the top decile (50 stocks) and short the bottom decile to construct portfolio
  - Rebalance annually, based on average value for past 12 months
  - Rebalance annually, based on December rankings of previous year
  - Rebalance quarterly (Jan, Apr, Jul, Oct), based on previous month rankings
  - Rebalance monthly, based on previous month rankings
- Repeat for top/bottom 25, 10 stocks.
- Compare returns of the portfolio of each factor at the end of 6 years
- Selection criteria: factors that beat the risk-free rate (3.5%), not S&P 500
Positive Momentum, Low Div Yield, and Positive Momentum (SMA6) yielded returns above the risk-free rate when rebalancing annually (used 12 months)

Risk free returns = 1.19

0.81
Positive Momentum, Low Div Yield, and Low CF Margin yielded returns above the risk-free rate when rebalancing annually (used previous month).
Low BV over MV, Large market cap, High E/P, High ROE, Low P/CF, Positive momentum selected when rebalancing quarterly

Risk free returns = 1.19
0.81
Similarly, Low BV over MV, Large market cap, High E/P, High ROE, Low P/CF, Positive momentum selected when rebalancing monthly

Risk free returns = 1.19
0.81
Frequency of rebalancing selects different factors. Thus, choose factor and rebalancing frequency together.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Rebalance Annually (using last 12m)</th>
<th>Rebalance Annually (using previous month)</th>
<th>Rebalance Quarterly</th>
<th>Rebalance Monthly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Book Value over Market Value (growth)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Large Market Cap</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Earnings to Price</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Low Dividend Yield</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High ROE</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Low Financial Leverage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low P/CF</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>High CF Margin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High CFO to Total Liab</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Trading Volume</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lagged excess monthly returns (positive short-term momentum)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMA3 returns (positive momentum)</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>SMA6 returns (positive momentum)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Works for rebalancing monthly/quarterly strategy | Works for rebalancing annually strategy
Validating factors - When rebalancing annually, low dividend yields and positive SMA6 continue to yield above risk-free returns

Risk free returns = 1.04

0.96
Validating factors: For monthly/quarterly rebalancing strategies, large market caps and high E/P continue to yield above risk-free returns.
Summary of factors selection

Based on our 3 factor evaluation methods, we decided to build a portfolio based on the 5 factors with the following weights: positive momentum (50%), small market cap, high ROE, high E/P ratio, dividend yield (12.5% each)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Linear Regression</th>
<th>PCA</th>
<th>Historical Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Momentum</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>High ROE</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large Cap</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>High E/P Ratio</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Low Dividend Yield</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
Adopting a long-short portfolio strategy for top/bottom 10 rather than top/bottom 50 amplifies returns

Risk free returns = 1.04
0.96
Testing our portfolio: Creating portfolio based on deciles does not beat risk free rate, but selecting top/bot 10 does

Portfolio Allocation
- Positive momentum (50%)
- Large Cap (12.5%)
- High E/P (12.5%)
- Low Div Yield (12.5%)
- High ROE (12.5%)
Neural Network
Predicting excess returns using neural network

- Constructed a 7-layer neural network to predict excess returns
- Trained on dataset containing 13 firm-level factors
- Training:
  - 7 fully connected layers, backpropagation
- Hyperparameters:
  - Number of epochs: 50
  - Learning rate: $0.01 / t$ (where $t$ is the current epoch)
- Loss:
  - Training: 4.7%
  - Validation: 4.9%
Predicting excess returns using neural network

- Methods to prevent overfitting
  - Implemented dropout regularization, early stopping
  - Aimed to achieve similar loss on training set and validation set, to ensure we were not overfitting to our training data

- Reducing loss
  - Reduced the number of features by filtering out irrelevant factors
  - Normalized features (ratios) before inputting into our model
  - Future: could implement Nesterov momentum to escape local minima

- Prediction method
  - Currently using fundamental characteristic data for the current month to predict returns for the current month
  - Need to adapt this to using fundamental characteristics for the current month to predict returns for the next month
Predicting excess returns using neural network

Results

Initial sort based on predicted excess returns to construct the following portfolios:

- Long top 10 & short bottom 10
  - Annualized Return: 20.39%
- Long top 25 & short bottom 25
  - Annualized Return: 12.94%
- Long top 50 & short bottom 50
  - Annualized Return: 9.46%
Evaluating factors from neural network predicted returns

Factors desired: low dividend yield, high ROE, high E/P, high market value, high positive momentum (6 month SMA) - all agree except for dividend yield

<table>
<thead>
<tr>
<th></th>
<th>Dividend Yield/%</th>
<th>ROE</th>
<th>Earnings to Price</th>
<th>Market Value/billion $</th>
<th>Momentum (6 month SMA)/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10</td>
<td>0.41%</td>
<td>0.139252</td>
<td>0.010434</td>
<td>86.919</td>
<td>0.41%</td>
</tr>
<tr>
<td>Bottom 10</td>
<td>0.32%</td>
<td>-4.106257</td>
<td>-0.018279</td>
<td>8.102</td>
<td>-2.40%</td>
</tr>
<tr>
<td>Top 25</td>
<td>0.45%</td>
<td>0.083983</td>
<td>0.01228</td>
<td>86.648</td>
<td>0.64%</td>
</tr>
<tr>
<td>Bottom 25</td>
<td>0.37%</td>
<td>-1.672887</td>
<td>0.001149</td>
<td>9.563</td>
<td>-2.35%</td>
</tr>
<tr>
<td>Top 50</td>
<td>0.47%</td>
<td>0.066367</td>
<td>0.012473</td>
<td>76.554</td>
<td>0.49%</td>
</tr>
<tr>
<td>Bottom 50</td>
<td>0.56%</td>
<td>-0.821791</td>
<td>0.006562</td>
<td>12.854</td>
<td>-2.03%</td>
</tr>
<tr>
<td>All</td>
<td>0.50%</td>
<td>-0.057927</td>
<td>0.011955</td>
<td>40.815</td>
<td>0.12%</td>
</tr>
</tbody>
</table>
Conclusion
Conclusion and discussion

We built a zero net portfolio by longing/shorting companies based on the following factors and weights:

- Positive momentum (6 month SMA) - 50%
- Large market value - 12.5%
- High earnings to price - 12.5%
- High return on equity (ROE) - 12.5%
- Low dividend yield - 12.5%

Building portfolio by decile (top/bot 50) gave only 2.6% returns, but top/bot 10 strategy gave 19.4%. Additionally, returns on test set vary significantly by factor and the top/bot x strategy.
Conclusion and discussion

Data processing

- Splitting training/validation/test by year might affect results due to time-related market conditions
- Would be interesting to do a random split

Factor Selection

- Some factors might have not worked for topbot50, but worked well for topbot10 and 25, so we did not include them. Can build more specific portfolio construction strategies with specific factors, topbot x strategy.

Portfolio Construction

- Look at magnitude of returns when choosing and weighting factors as well
Questions?