

MS&E 448: Trading Strategies based on Explained Cross-Sectional Returns

Amay Aggarwal, Yu Xin Ang, Varin Satish Nair, Greg Soh, Chor Seng Tan

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

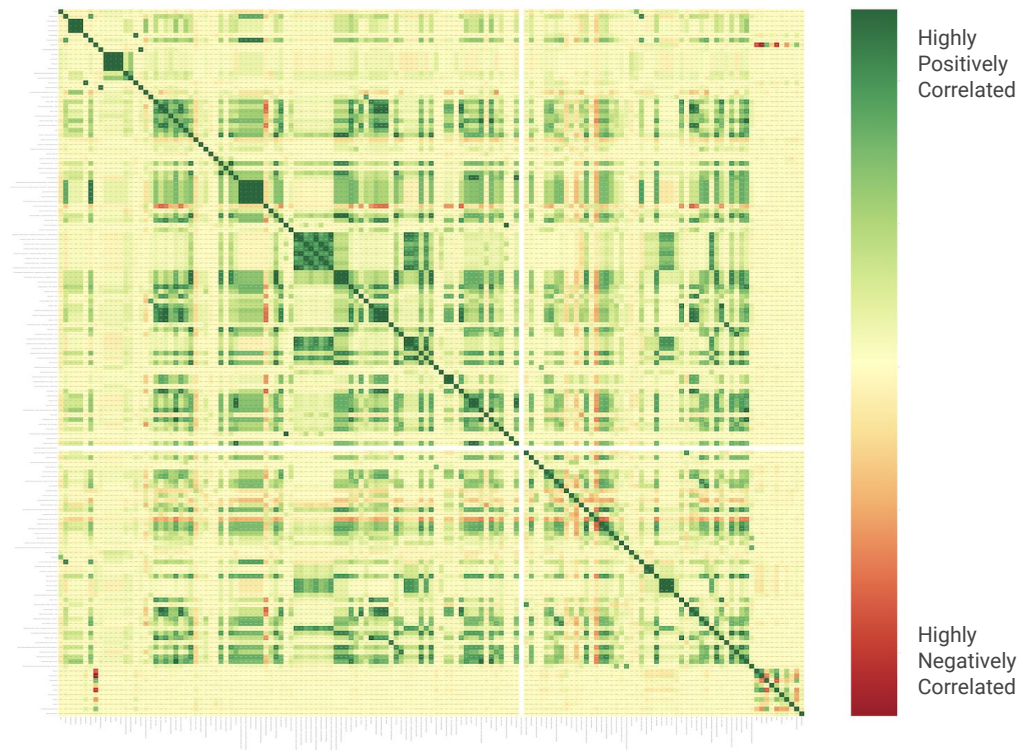
- Cross-section of expected returns should be explained by exposure to systematic risk factors
- Factor investing harvests these risk premia through exposure to factors
- Finding the “**right**” **factors** has become the central question of asset pricing
- Harvey et al. (2016): 300 published candidate 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
- Training set: 2011 to 2016, Validation set: 2017, Test set: 2018.

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

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

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

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

- α_i : intercept of asset i
- $f_{i,t}$: factor variables
- $\beta' = (\beta_1, \dots, \beta_K)'$: 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:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

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 (%)		
	ALL	INDUSTRY	COMPANY
Training	5.5904	5.5754	5.5325
Validation	5.3061	5.3386	5.7908
Test	5.5150	5.5549	6.4716
Ridge regression (Test)	5.5143	5.5527	6.4708

$$Return = \beta_1 \cdot SMA3 + \beta_2 \cdot ROE + \beta_3 \cdot RTV + \beta_M \cdot S\&P$$

$$\begin{aligned} \beta_1 &= 0.9794 & \beta_3 &= 6.9738e^{-6} \\ \beta_2 &= -0.1127 & \beta_M &= 1.132 \end{aligned}$$



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

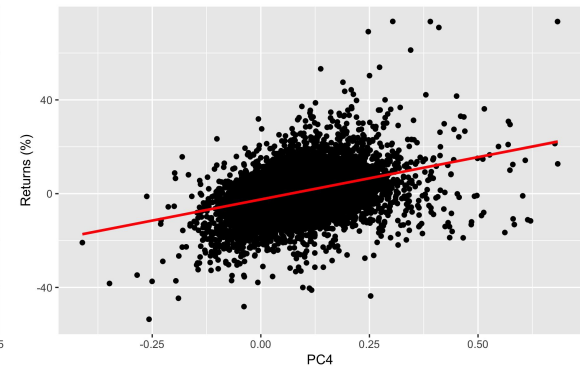
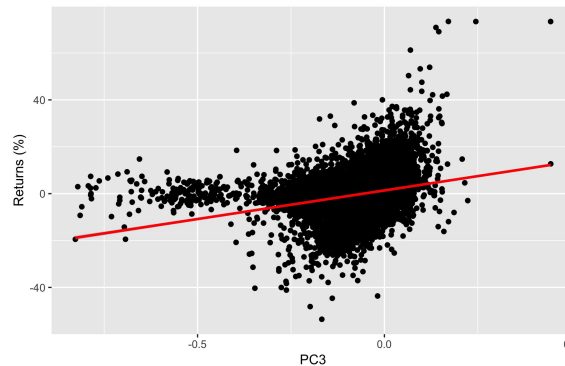
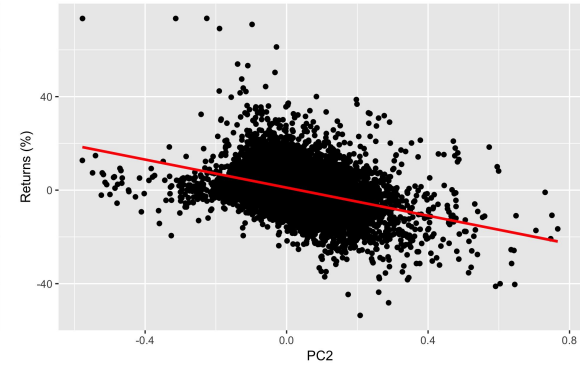
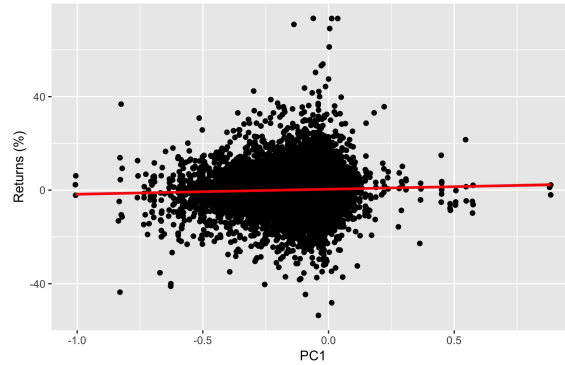
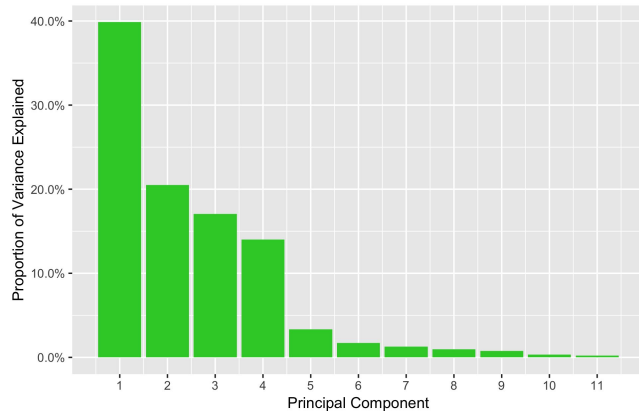
Factor Selection using PCA

- 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



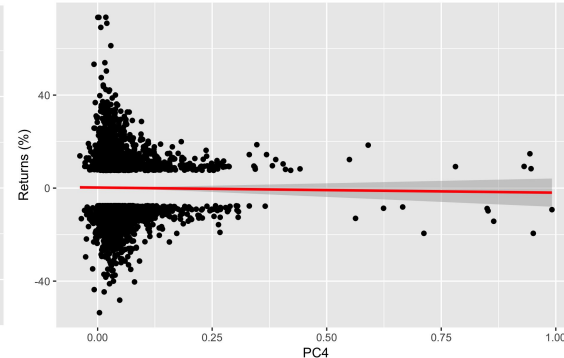
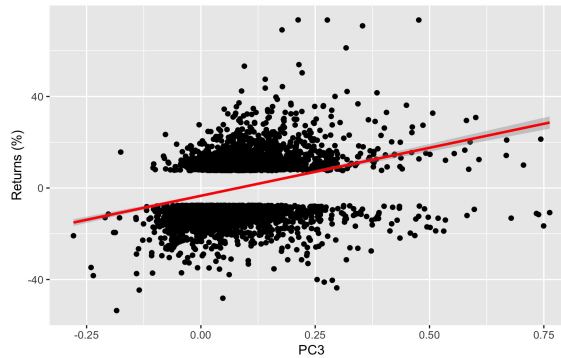
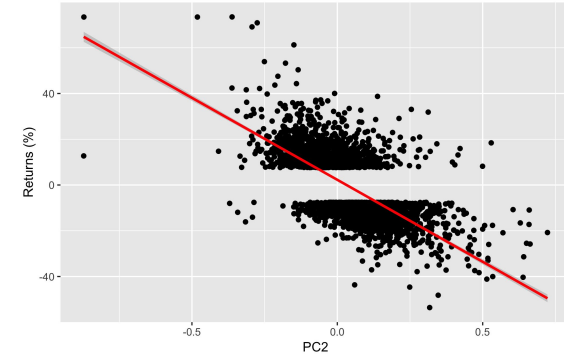
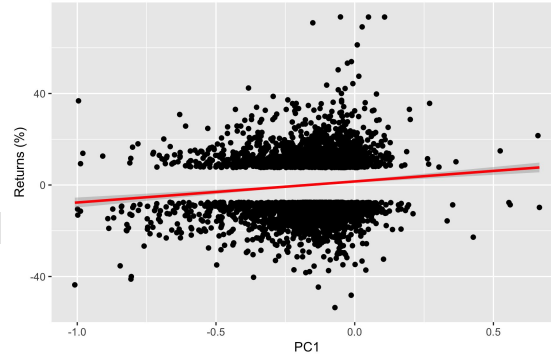
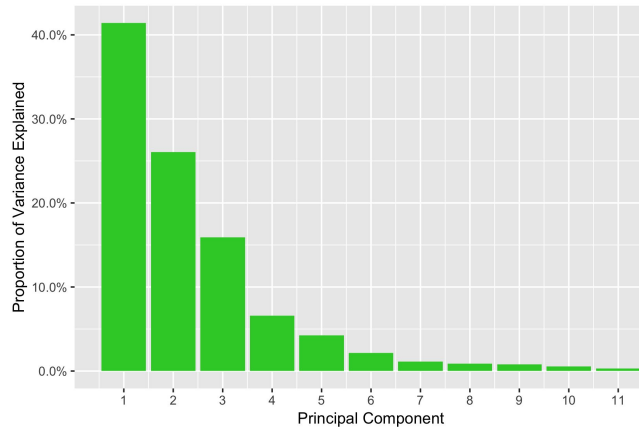
Results: When including all observations, excess returns seem correlated with small caps and positive momentum

- 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)

	PC1	PC2	PC3	PC4
BVOverMV	-0.056	0.632	-0.251	0.723
Market Cap	-0.095	-0.51	-0.842	0.147
EarningsToPrice	0.006	-0.021	0.004	0.005
DivYield	-0.006	0.014	-0.007	0.002
ROE	0.002	-0.01	-0.001	-0.002
FinLeverage	0.001	-0.002	0.002	-0.001
PtoCFRatio	0.005	-0.001	-0.001	-0.003
CashFlowMargin	-0.983	0.001	0.108	-0.018
CFOtoTotalLiab	-0.141	-0.059	0.036	-0.089
TradingVolume_relative	0	-0.002	-0.005	0.002
ExcessMonthlyReturn_SMA3	0.036	-0.58	0.464	0.669

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

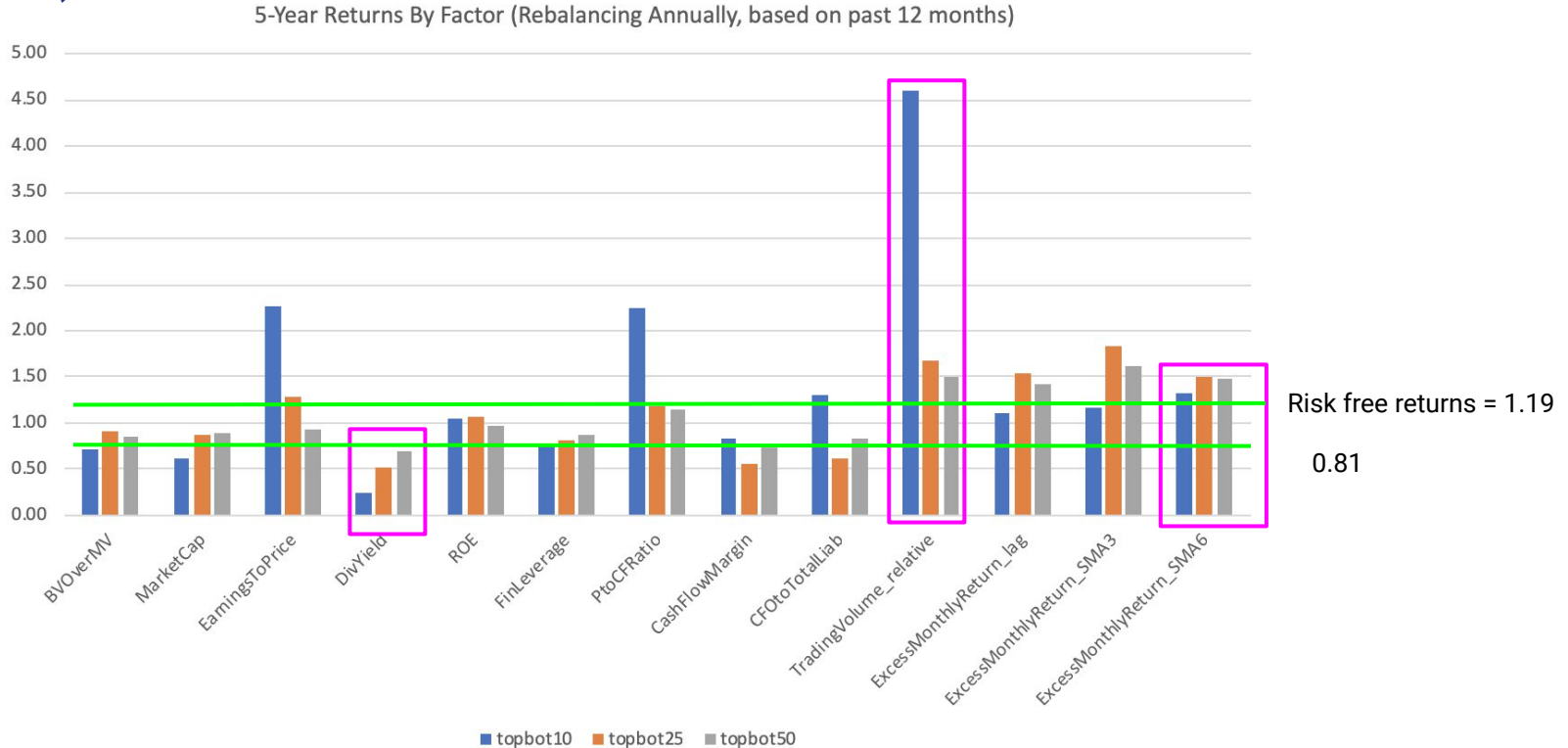
	PC1	PC2	PC3	PC4
BVOverMV	-0.122	0.441	0.867	0.136
MarketCapInverse	-0.030	-0.082	-0.107	0.988
EarningsToPrice	0.018	-0.028	-0.002	0.029
DivYield	-0.004	0.015	0.002	0.002
ROE	0.007	-0.005	-0.004	0.002
FinLeverageInverse	0.001	-0.002	-0.002	-0.004
PtoCFRatioInverse	0.002	0.001	-0.003	0.008
CashFlowMargin	-0.969	-0.134	-0.028	-0.055
CFOtoTotalLiab	-0.184	-0.087	-0.144	0.029
TradingVolume_relative	-0.002	0.000	-0.001	-0.004
ExcessMonthlyReturn_SMA3	0.107	-0.879	0.464	-0.020

Using Historical Returns

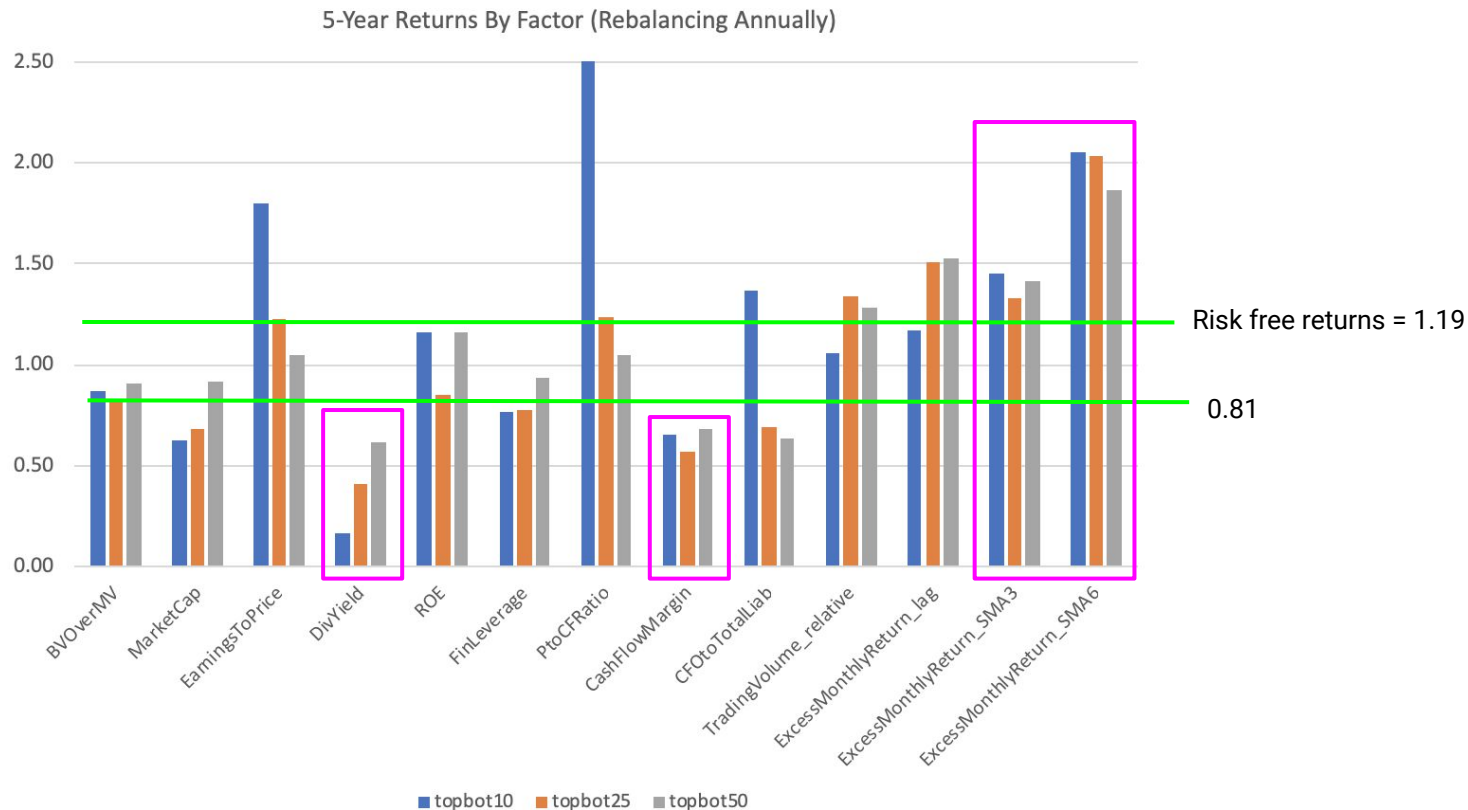
Factor Selection using Historical Returns

- 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)

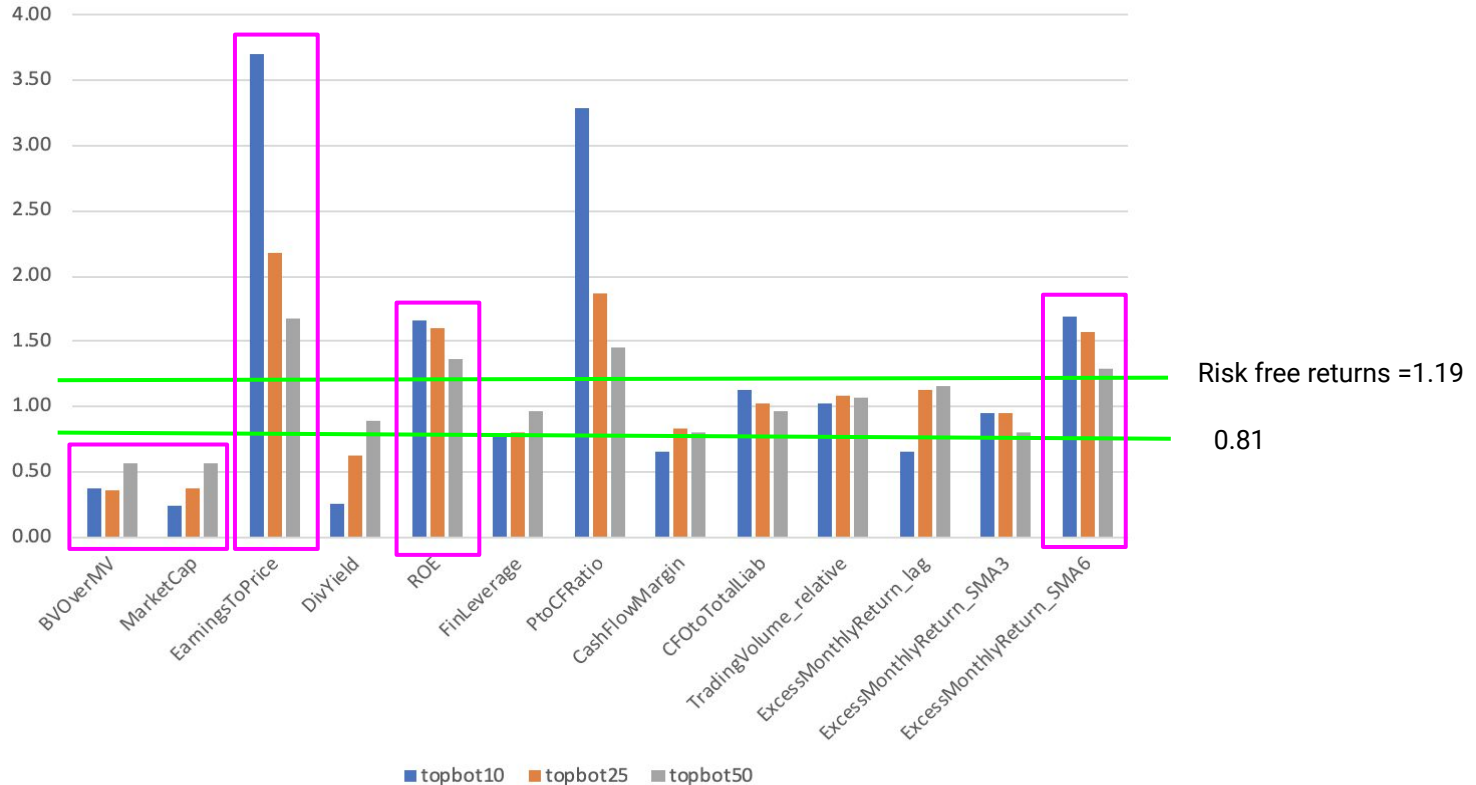


Positive Momentum, Low Div Yield, and Low CF Margin yielded returns above the risk-free rate when rebalancing annually (used previous month).

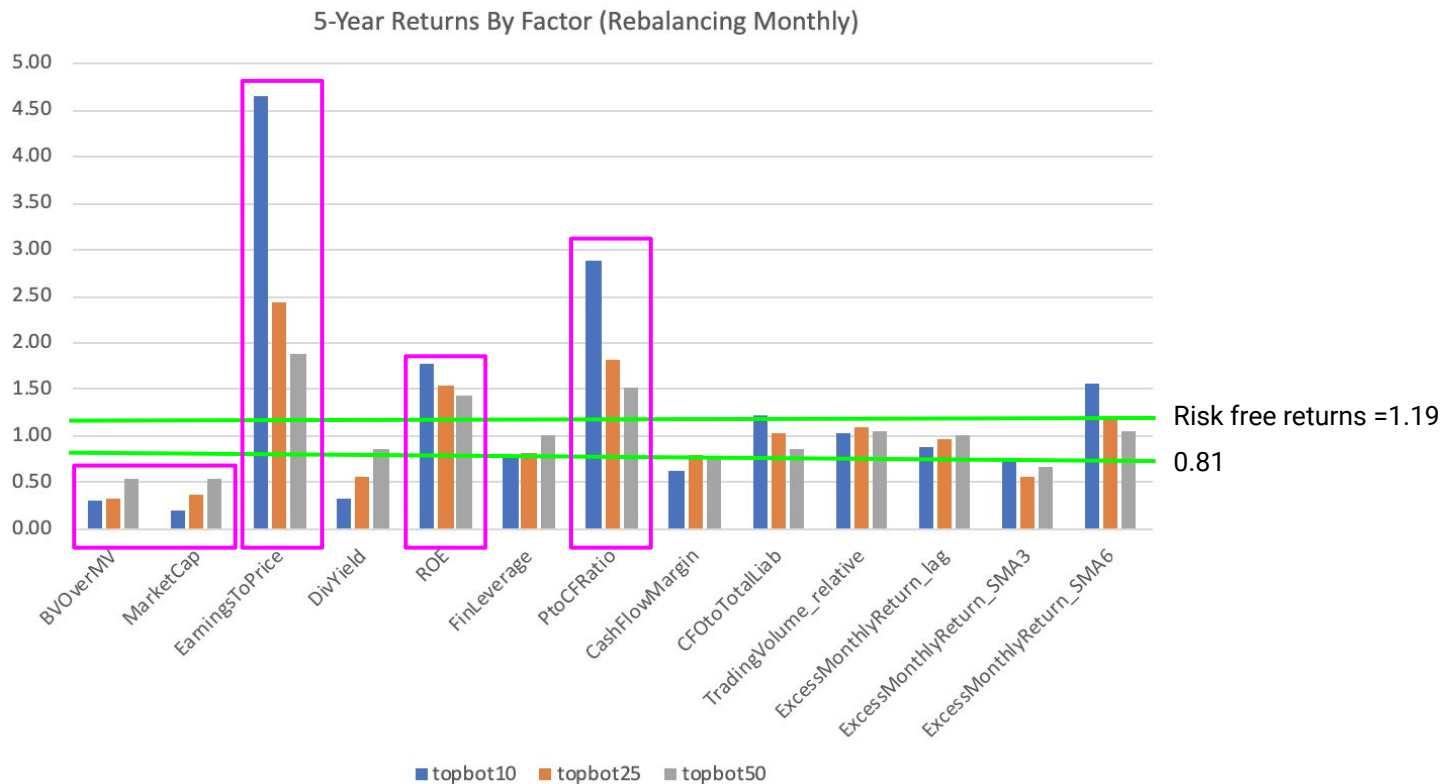


Low BV over MV, Large market cap, High E/P, High ROE, Low P/CF, Positive momentum selected when rebalancing quarterly

5-Year Returns By Factor (Rebalancing Quarterly)



Similarly, Low BV over MV, Large market cap, High E/P, High ROE, Low P/CF, Positive momentum selected when rebalancing monthly



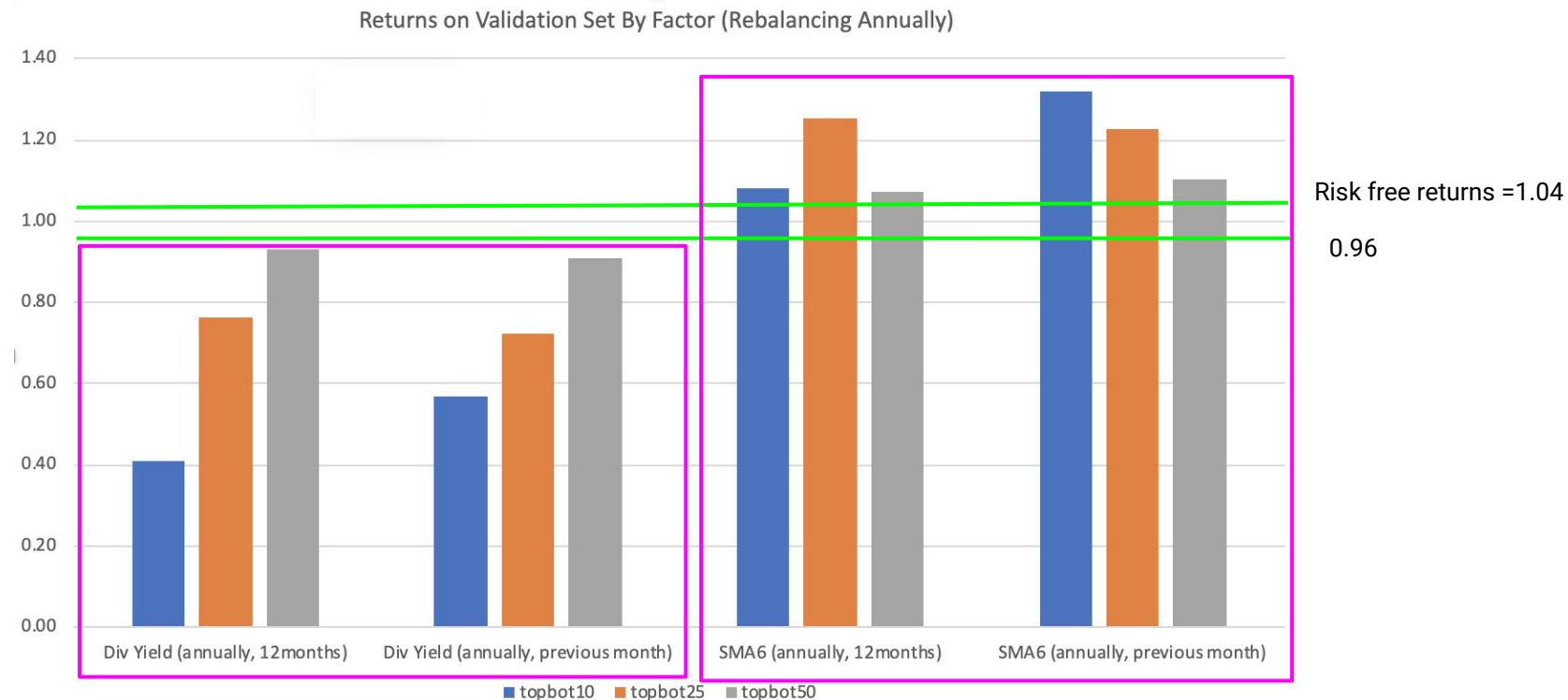
Frequency of rebalancing selects different factors. Thus, choose factor and rebalancing frequency together.

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

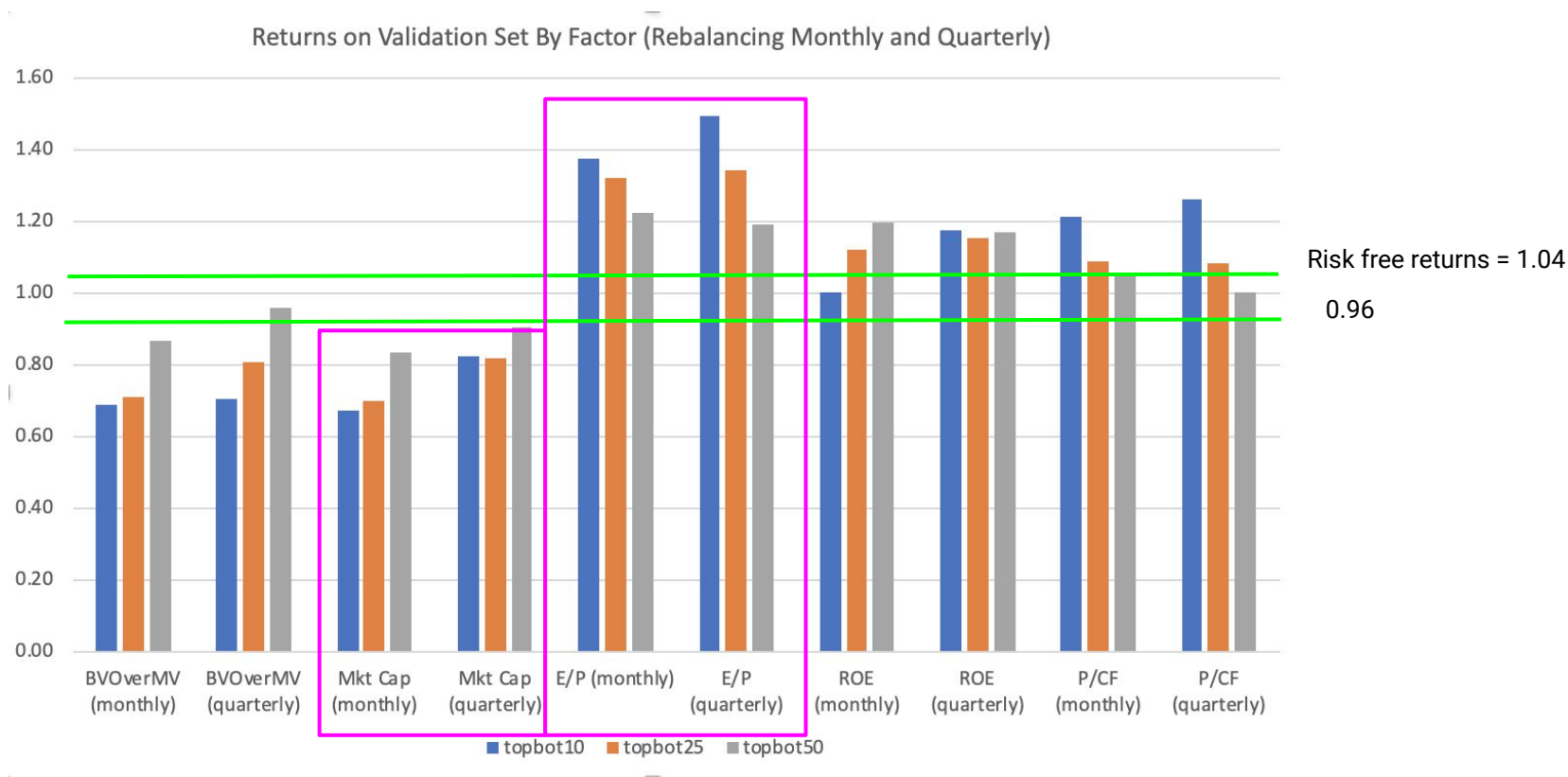
Works for rebalancing monthly/quarterly strategy

Works for rebalancing annually strategy

Validating factors - When rebalancing annually, low dividend yields and positive SMA6 continue to yield above risk-free returns



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



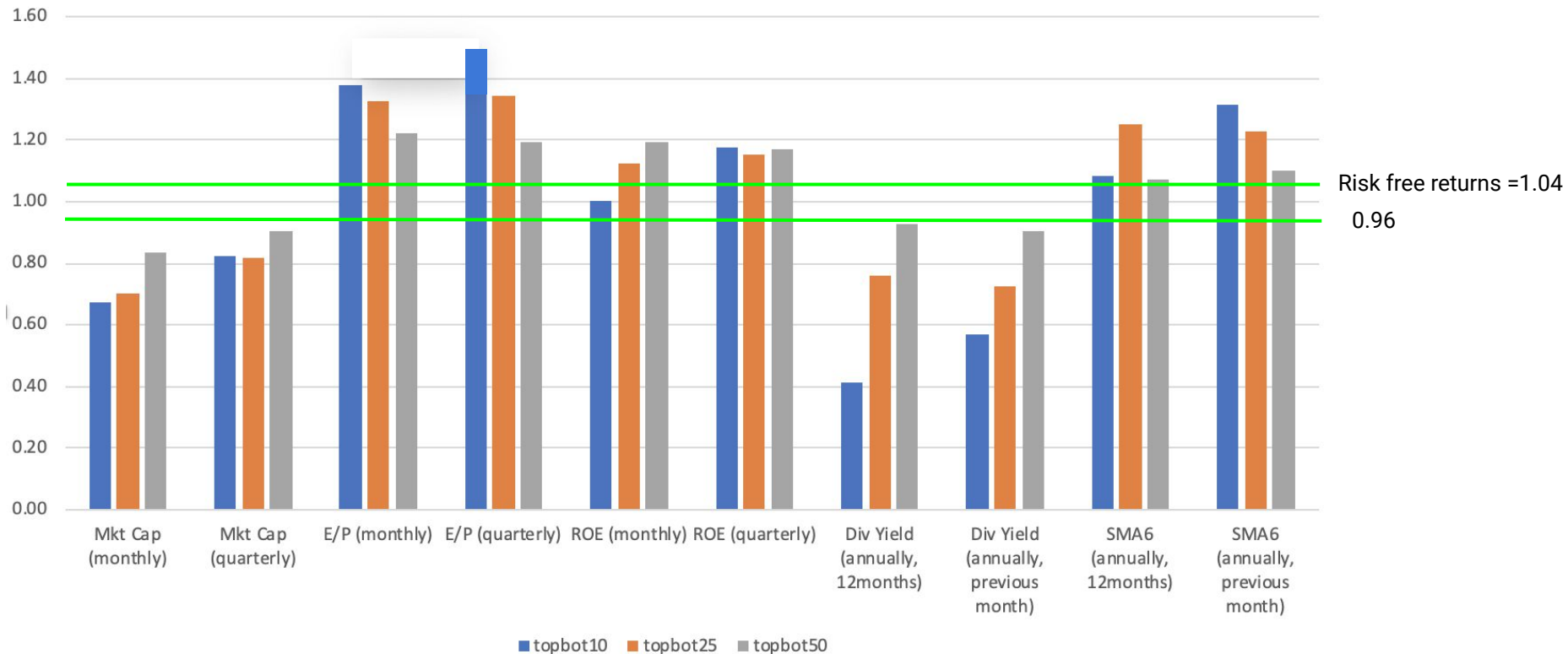
Summary of factors selection

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)

Factor	Linear Regression	PCA	Historical Returns
Positive Momentum	✓	✓	✓
High ROE	✓		
Large Cap			✓
High E/P Ratio			✓
Low Dividend Yield			✓

Adopting a long-short portfolio strategy for top/bottom 10 rather than top/bottom 50 amplifies returns

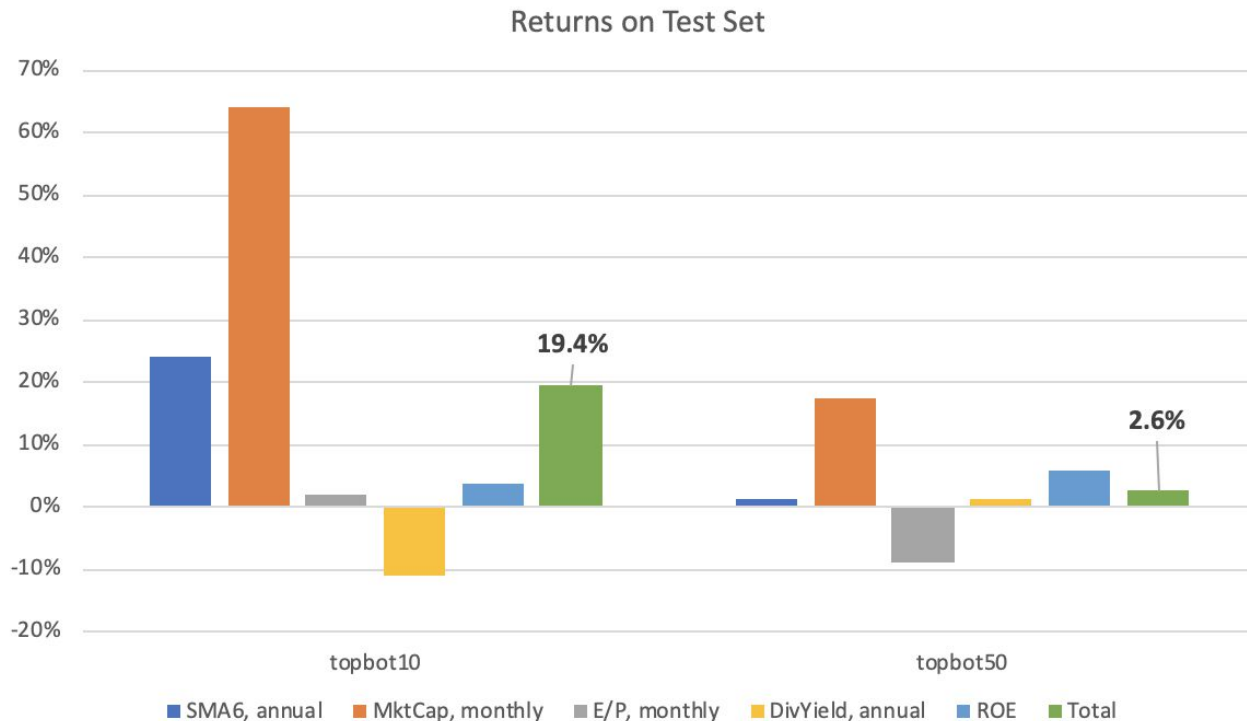
Returns On Validation Set By Selected Factors



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

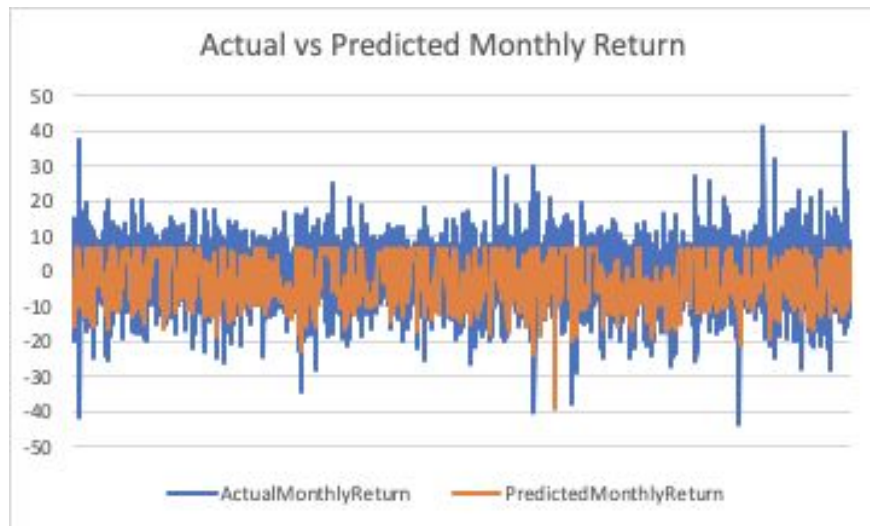
- Methods to prevent overfitting
 - Implemented dropout regularization, early stopping
 - Aimed to achieve similar loss on training set and validation set, to ensure we were not overfitting to our training data
- Reducing loss
 - Reduced the number of features by filtering out irrelevant factors
 - Normalized 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

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

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?