New York Stock Exchange Risk and Return Prediction

Presented by:

Guanzhong (Frank) Chen

Zhe (Peter) Huang

Shiyu Liu

Cangcheng Tang

Jiayuan (Justin) Tian

Data Introduction

- **Fundamentals:**
  - The quantitative factors extracted from financial statements at the end of each fiscal year
  - Time range: 2012 to 2016
- **Prices/Adjusted Prices:**
  - Time series data: Open, close, high, low price and volume for each trade day
  - Adjusted stock splitting
  - Dates from 2010 to 2016
- **Securities**
  - Company information
  - Will take industry factor from here

BROWN

Goal & Design

- **Independent Variables:**
  - Build factors as independent variables
  - Reflect the company financial condition and the stock risk
- **Logic**
  - Financial statement is released approximately 3 months after each fiscal year.
  - Then include three-month return rate as an additional factor when financial statement is released.
  - Use these factors to predict the stock return rate for the next 3 months and the standard deviation

BROWN

Factors calculated based on financial statements at fiscal year end

Prediction: 4-6 month relative return rate and standard deviation as target variable

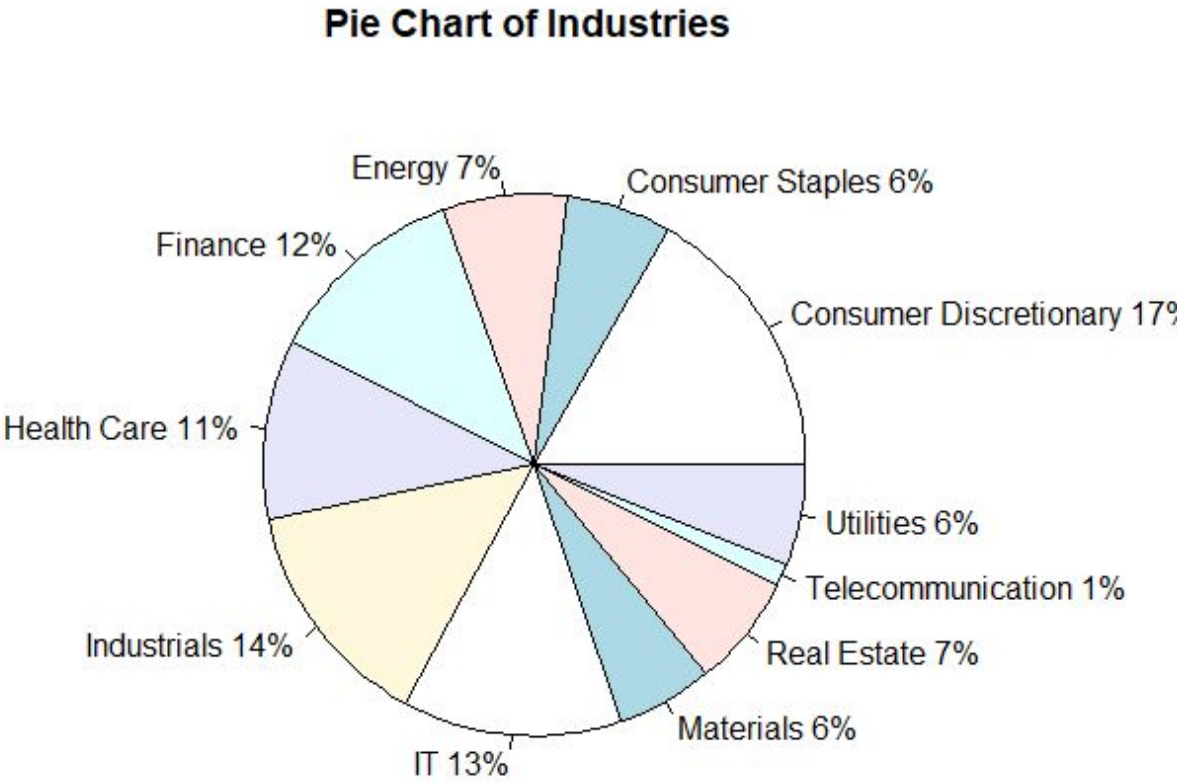
Factors calculated using time series stock price.

3-month return rate as an additional factor

BROWN

Pie Chart

- Show the ratio of companies in different industry

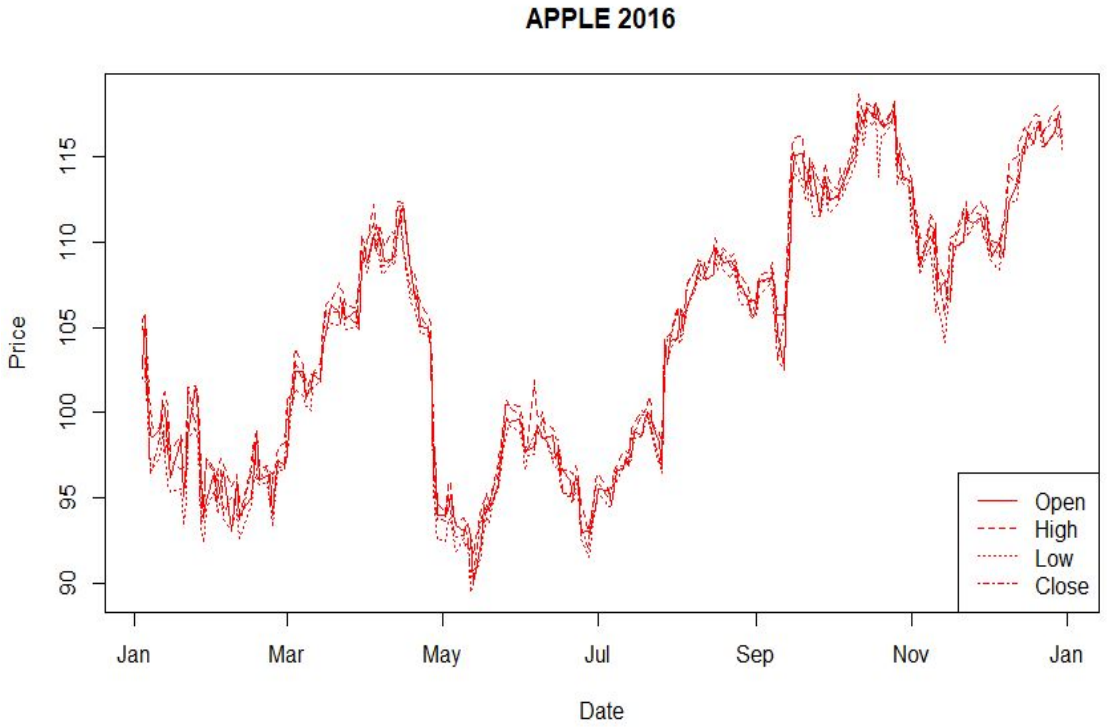


S&P 500 Trend

Show the trend of S&P 500 index from 2011 to 2016

Apple Stock Price

- Show the trend of stock price of Apple in 2016
- Show four metrics: Lows, Highs, Open, and Close

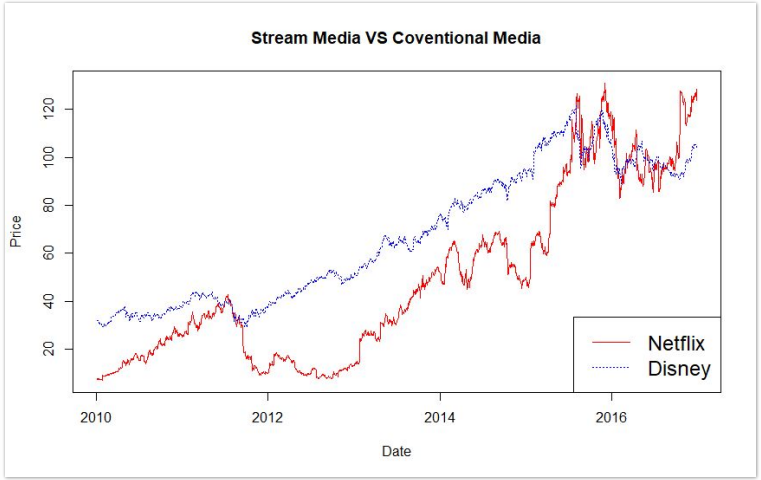


Candle Chart of Apple

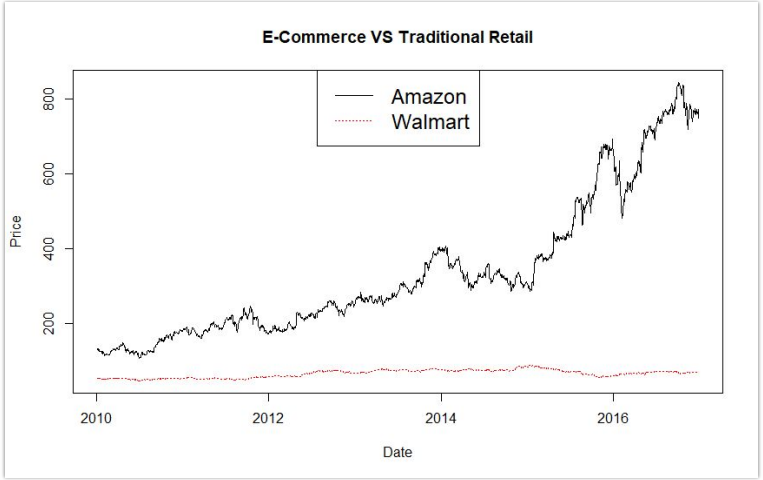
Show both the price and volume of Apple in 2016

Stock Prices Comparison

Netflix VS Disney



Amazon VS Walmart



BROWN

Factors----Source: The Barra US Equity Model

- The Barra US Equity Model (USE4) is the first in a family of models to include the latest advances in risk methodology, providing institutional investors the ability to align factor structure with their investment processes.

- The full version can be found here: http://cslt.riit.tsinghua.edu.cn/mediawiki/images/4/47/MSCI-USE4-201109.pdf

- Presented by MSCI

BROWN

Factors

- **Current Ratio, Quick Ratio, Profit Margin, After Tax ROE**
  - Directly obtained from Fundamentals
  - Measures the Company financial health
- **Size & Nonlinear Size (Cubic)**
  - Log (Shares outstanding * Price)
  - Capture systematic risk differences between large cap and small cap assets
- **Beta**
  - $r_t - r_f = \alpha + \beta(r_m - r_f) + e_t$ (rt--stock return, rf--risk free return, rm--market return)
  - Describe the components of risk driven by market sensitivity
- **Momentum**
  - $RSTR = \sum w_t[ln(1 + r_t) - ln(1 + r_{ft})]$
  - Explains the return of stocks based on their recent relative performance

BROWN

Factors

- Residual Volatility
  - Alpha: Excess Earnings
  - Residual Std
  - Explains returns associated with high-volatility stocks that are not captured by the Beta factor
- Liquidity: Monthly, Quarterly and Annually Volume
  - $$STOM = \ln\left(\sum_{t=1}^{21}\frac{V_t}{S_t}\right) \qquad STOQ = \ln\left[\frac{1}{T}\sum_{\tau=1}^{T}\exp(STOM_\tau)\right] \qquad STOA = \ln\left[\frac{1}{T}\sum_{\tau=1}^{T}\exp(STOM_\tau)\right],$$
  - Describes return patterns to stocks based upon their relative trading activity.

BROWN

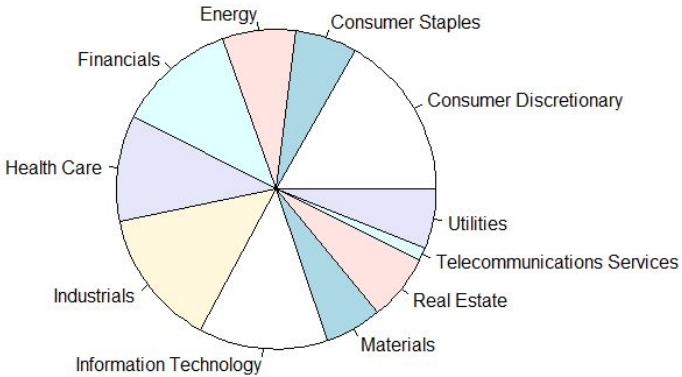
Factors

- Book-to-price $$btp = \frac{\frac{Assets - Liabilities}{OutstandingShares}}{Shareprice}$$
- Earning Yields
  - CETOP: Cash earnings-to-price ratio $\frac{EarningsPerShare}{SharePrice}$
  - ETOP: Trailing earnings-to-price ratio $\frac{NetIncome}{Size}$
- Leverage
  - Market Leverage $\frac{Size + LongtermDebt}{Size}$
  - Debt-to-asset Ratio $\frac{TotalLiabilities}{TotalAsset}$
  - Book Leverage $\frac{TotalEquity + LongtermDebt}{TotalEquity}$
- The above six factors capture the risk premium associated with the firm's business model

BROWN

Industry Factors

- Industry include 11 categories.



BROWN

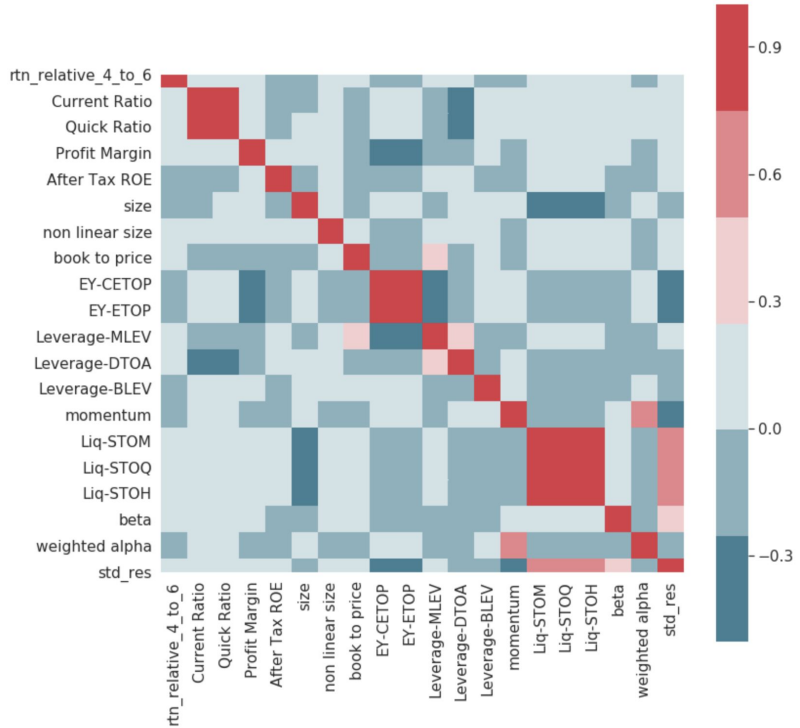
Target Variables

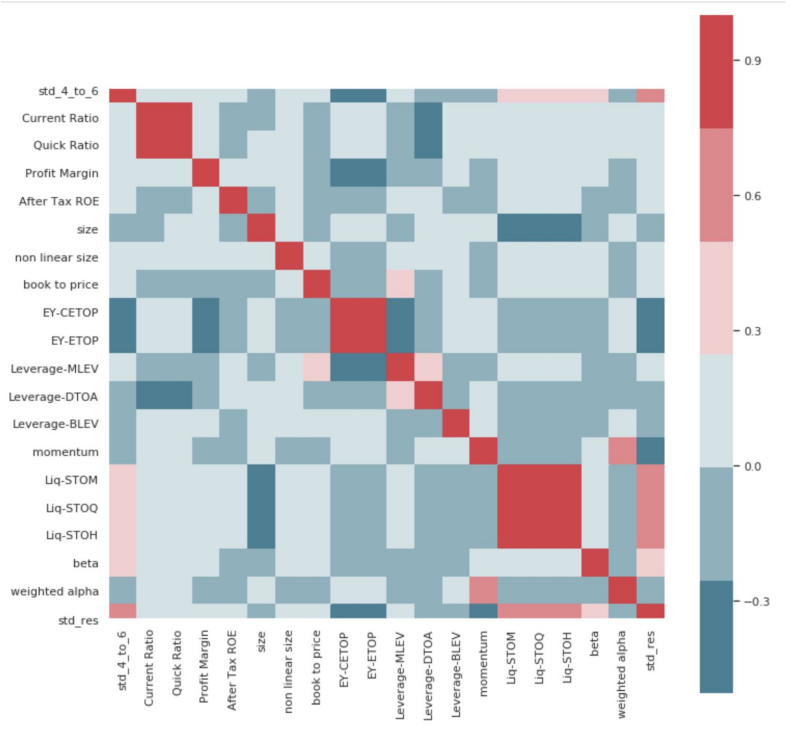
- **Profitability: Quarterly Return**
  Total return over the next three month, calculated using adjusted prices

- **Risk: Quarterly Standard Deviation**
  Standard deviation of daily return over the next three month

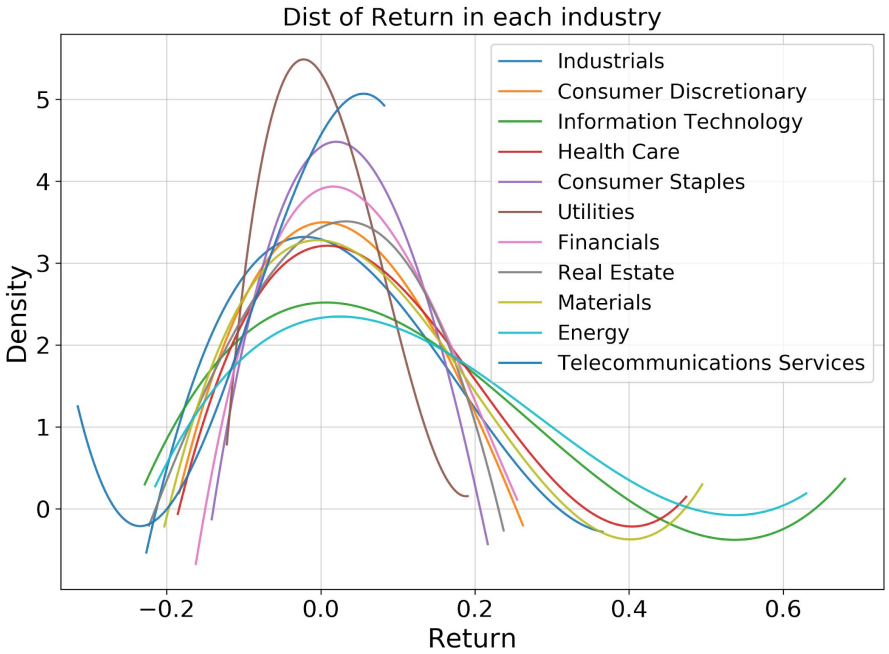
BROWN

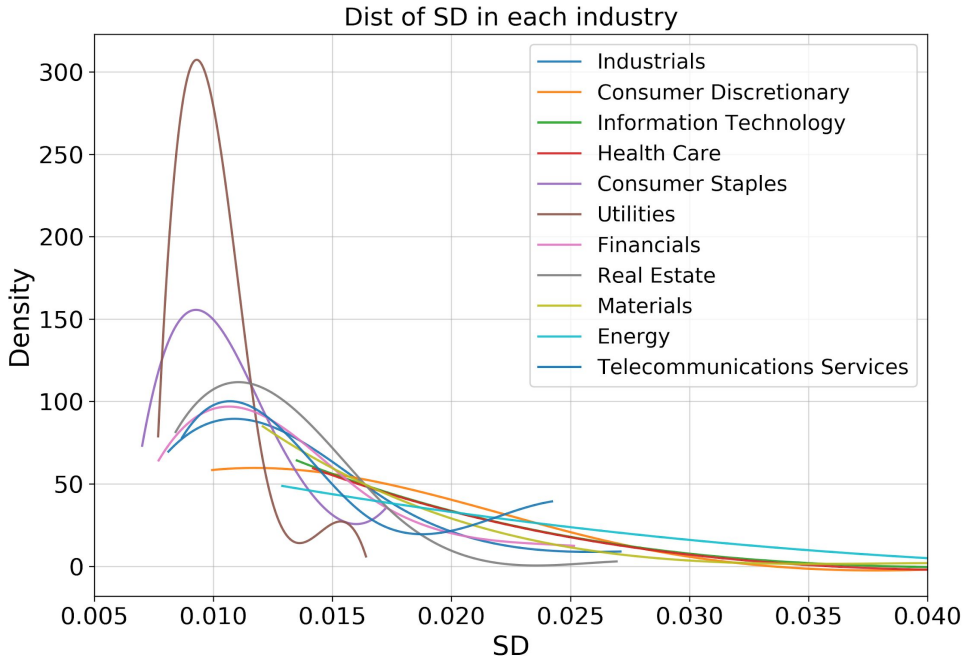
Target & Continuous Variables





Target & Categorical Variable: Industry





Feature Selection Using stepAIC()

- **Strategy**:

1. Train-test Split 0.8~0.2
2. Forward Stepwise Selection

- **Algorithm**:
  - Start with Null Model M_0
    - model_init = lm(half_year_return ~ 1, data = data_trn)
    - model = lm(half_year_return ~ ., data = data_trn)

BROWN

Feature Selection Using stepAIC()

- Algorithm:
  - For k = 0,1,…,p-1
    - Consider all p-k models that augment the predictors in M_k with one additional predictor
      - At first k = 0, we have a total of p models to choose from. Each contains only one covariate. Repeat the process so on and so forth.
    - Choose the best among these p-k models, and call it M_(k+1)
    - Best: smallest RSS

BROWN

Feature Selection Using stepAIC()

- Algorithm:
  - Select the best model from M_0 ,…, M_p using AIC

$$\mathrm{AIC} = \frac{1}{n\hat{\sigma}^2}\left(\mathrm{RSS} + 2d\hat{\sigma}^2\right)$$

```
model_forward = stepAIC(
  model_init,
  direction = "forward",
  scope = list(upper = model, lower = model_init),
  trace = 0
)
```

BROWN

AIC Feature Selection

Quarterly Return

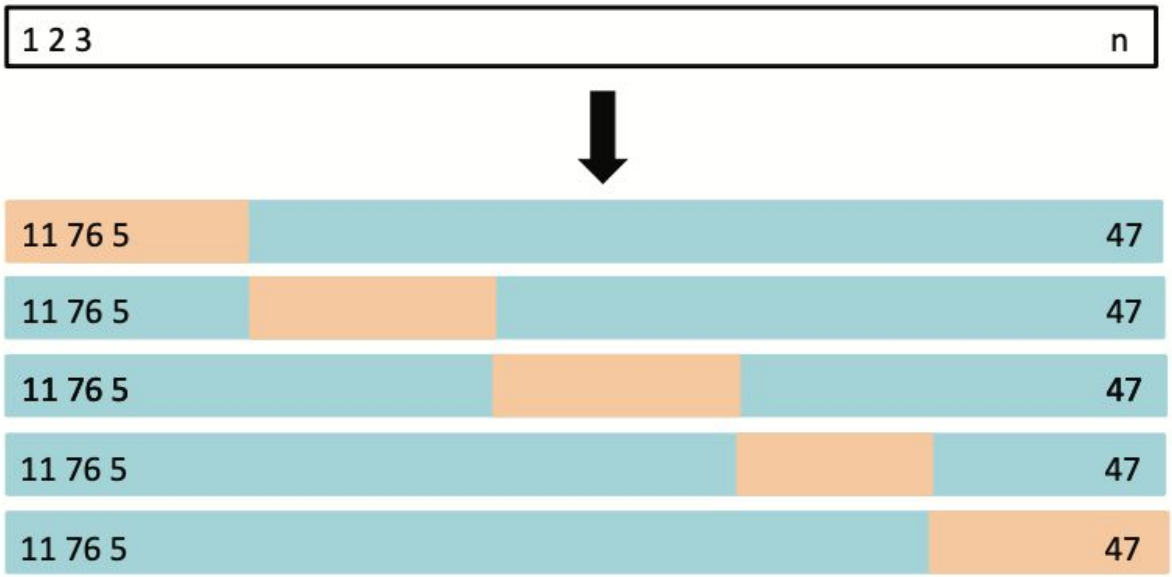
- Earning yields: cash earnings-to-price ratio,
- Industry factor,
- Non linear size,
- Size,
- Return of the first 3 month,
- After tax ROE,
- Leverage: debt-to-asset ratio,
- Beta,
- Current ratio,
- Profit margin,
- Book to price

Quarterly SD

- Residual standard deviation,
- Standard deviation over the past 3 month,
- Half-year liquidity,
- Quarterly-year liquidity,
- Beta,
- Return of the first 3 month,
- Trailing earnings-to-price ratio,
- Quick.Ratio,
- Profit.Margin,
- Momentum,
- Weighted alpha over the past 3 month

BROWN

Cross Validation



BROWN

Cross Validation

- R Packages: caret, elasticnet
- 5 folds
- Metric: root mean square error (RMSE)
- Methods
  - K-nearest neighbors (KNN)
  - Ridge
  - Elastic net (including Lasso)

BROWN

Elastic Net Regularization

- Lasso: minimizes

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p}|\beta_j|$$

- Ridge: minimizes

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p}\beta_j^2$$

- Elastic net: minimizes

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \left[\alpha \sum_{j=1}^{p}|\beta_j| + \frac{1-\alpha}{2}\sum_{j=1}^{p}\beta_j^2\right]$$

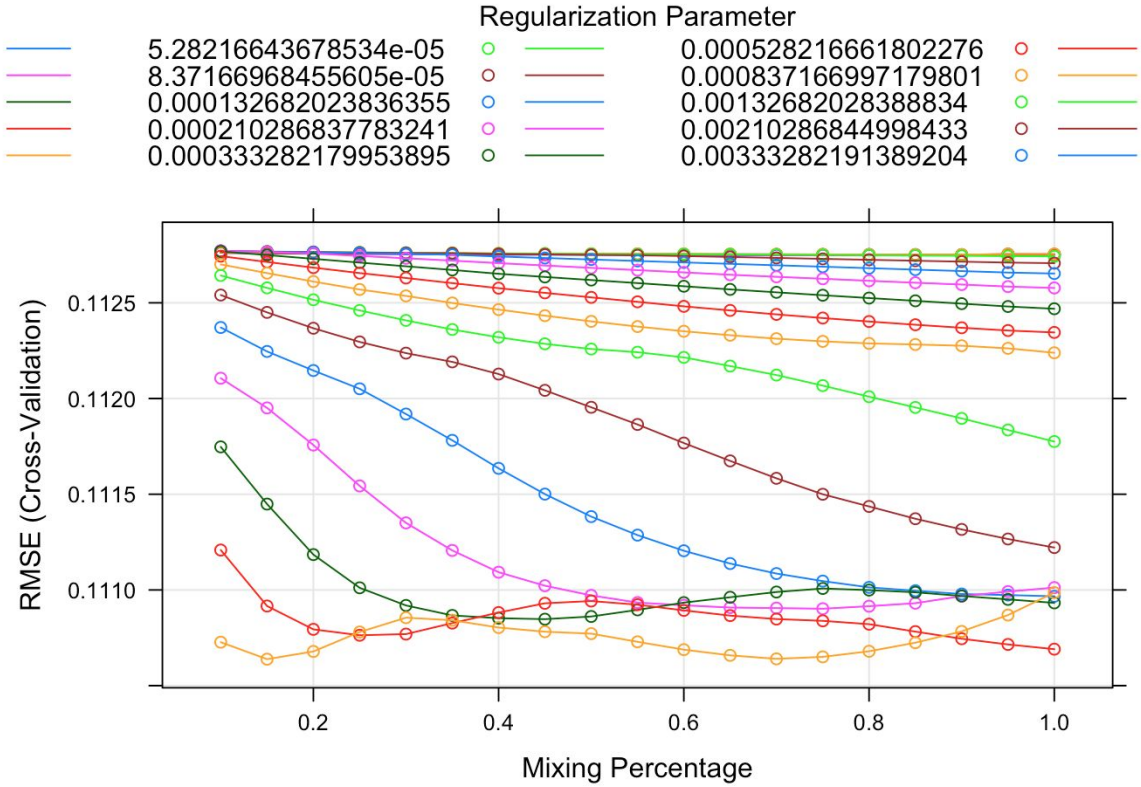
BROWN

Results - Quarterly Return

Model	Cross-validated RMSE	Test RMSE
Linear regression	0.1109	0.1001
K-nearest neighbors	0.1094	0.0995
Ridge	0.1127	0.0996
Elastic net	0.1106	0.0993

BROWN

Results - Quarterly Return - Elastic Net



BROWN

The Best Model - Quarterly Return

Type	α	λ
Elastic net	0.15	0.021

Cross-validated RMSE	Test RMSE
0.1106	0.0993

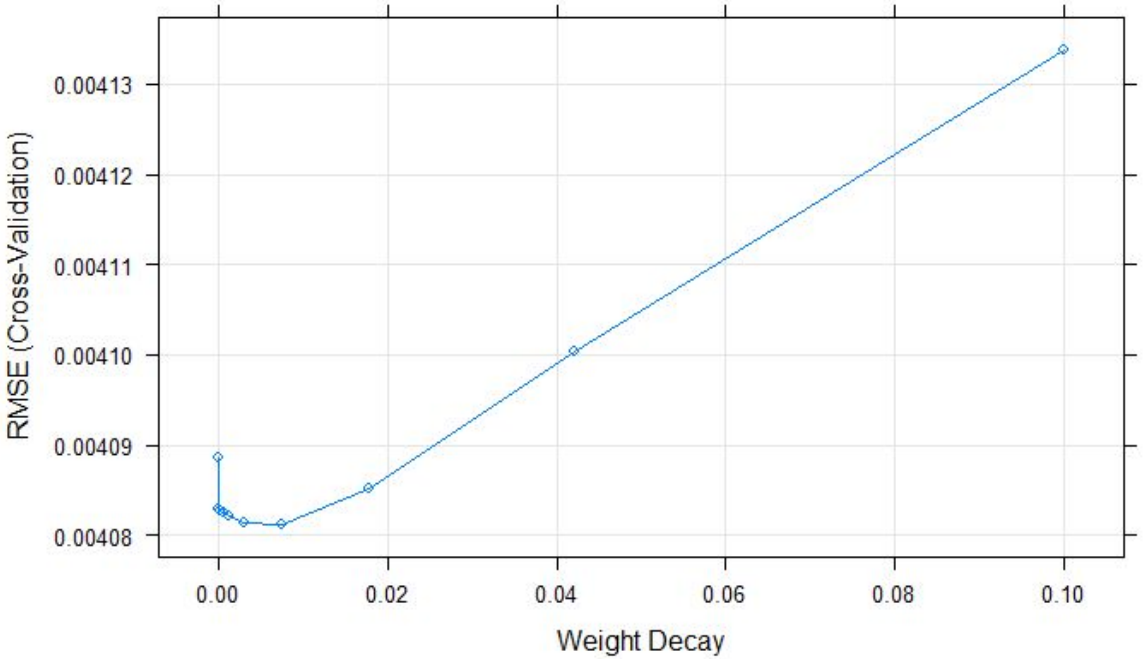
BROWN

Results - Quarterly SD

Model	Cross-validated RMSE	Test RMSE
Linear regression	0.0041	0.0042
K-nearest neighbors	0.0045	0.0042
Ridge	0.0041	0.0040
Elastic net	0.0041	0.0040

BROWN

Results - Quarterly SD - Ridge



BROWN

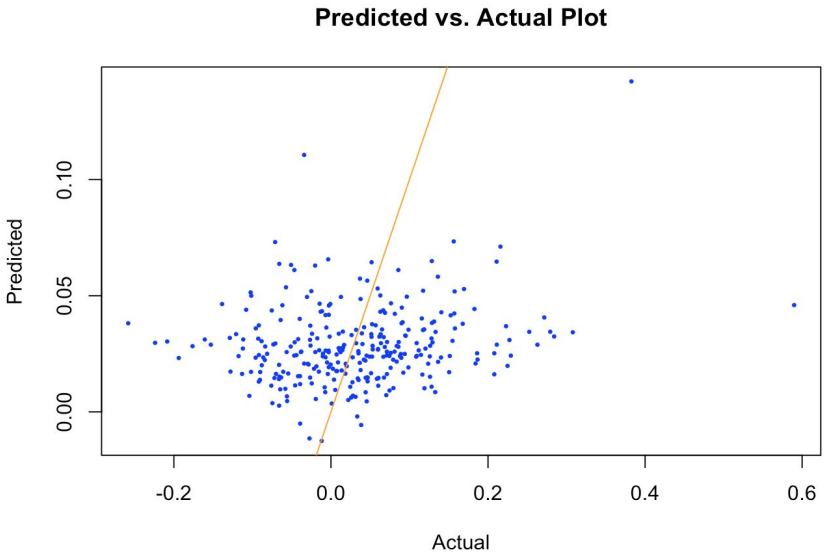
The Best Model - Quarterly SD

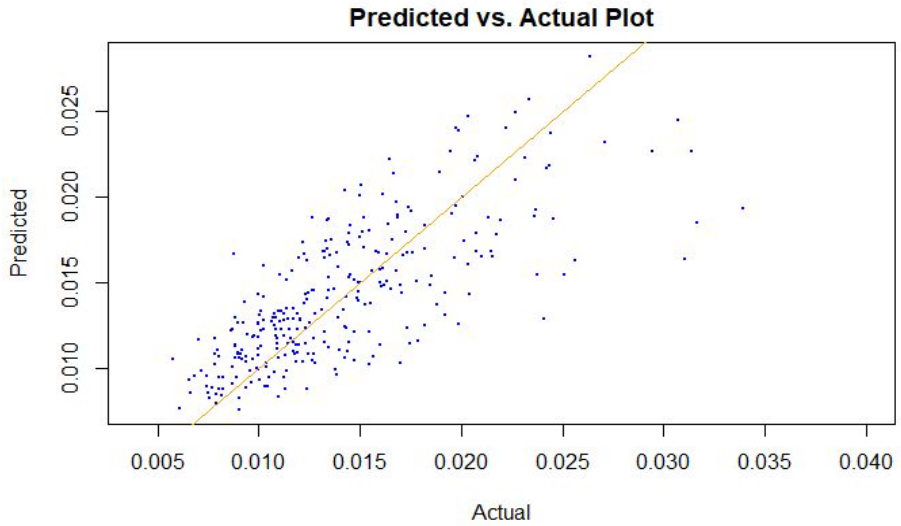
Type	λ	Cross-validated RMSE	Test RMSE
Ridge	0.0075	0.0041	0.0040

BROWN

Best Model Comparison - Quarterly SD

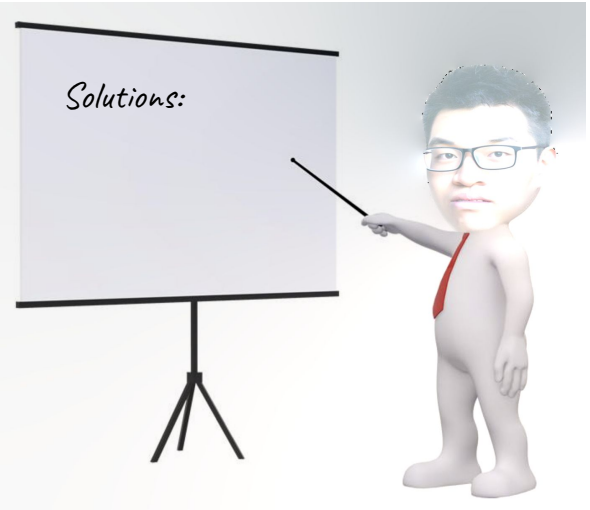
	Quarterly Return	Quarterly SD
Test R-square	0.0408	0.4941





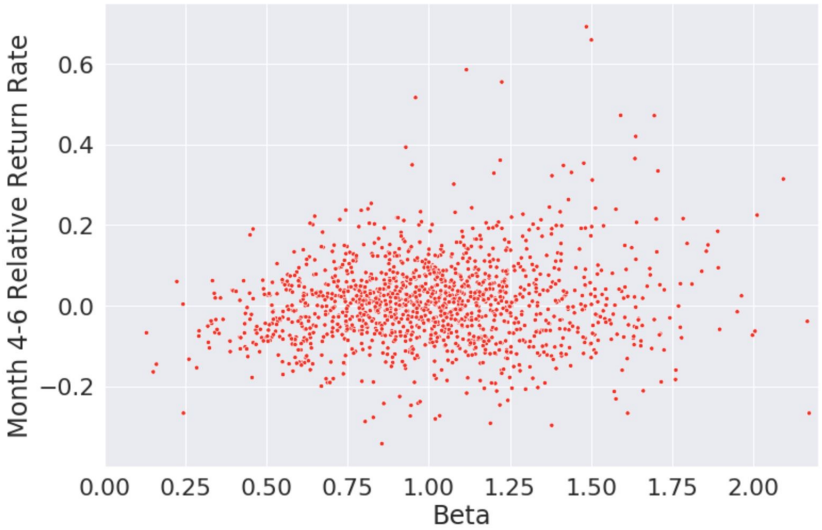
Thank you!

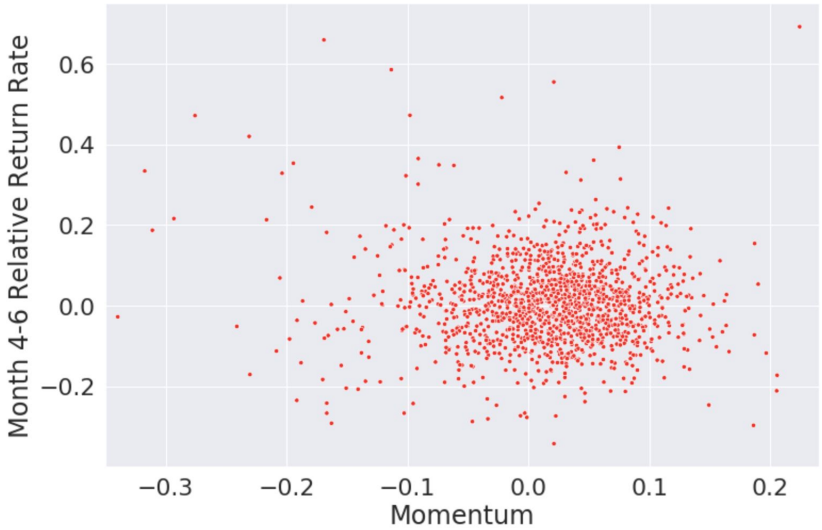
Q & A



BROWN

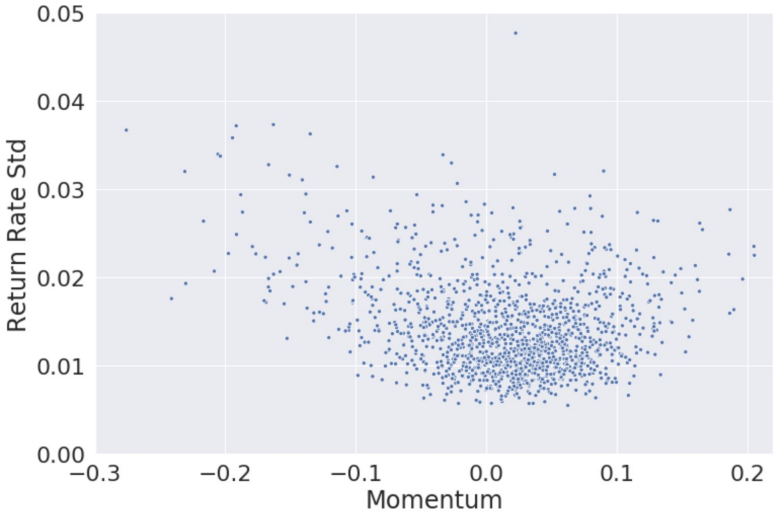
EDA--Advanced, Relative Return Rate





BROWN

EDA--Advanced, Quarterly Standard Deviation



BROWN