# Data Science Methods in Finance – Group Exam

**Group 2:**

Freek Linssen - SNR: 2015267 - ANR: U471688 - f.c.w.linssen@tilburguniversity.edu

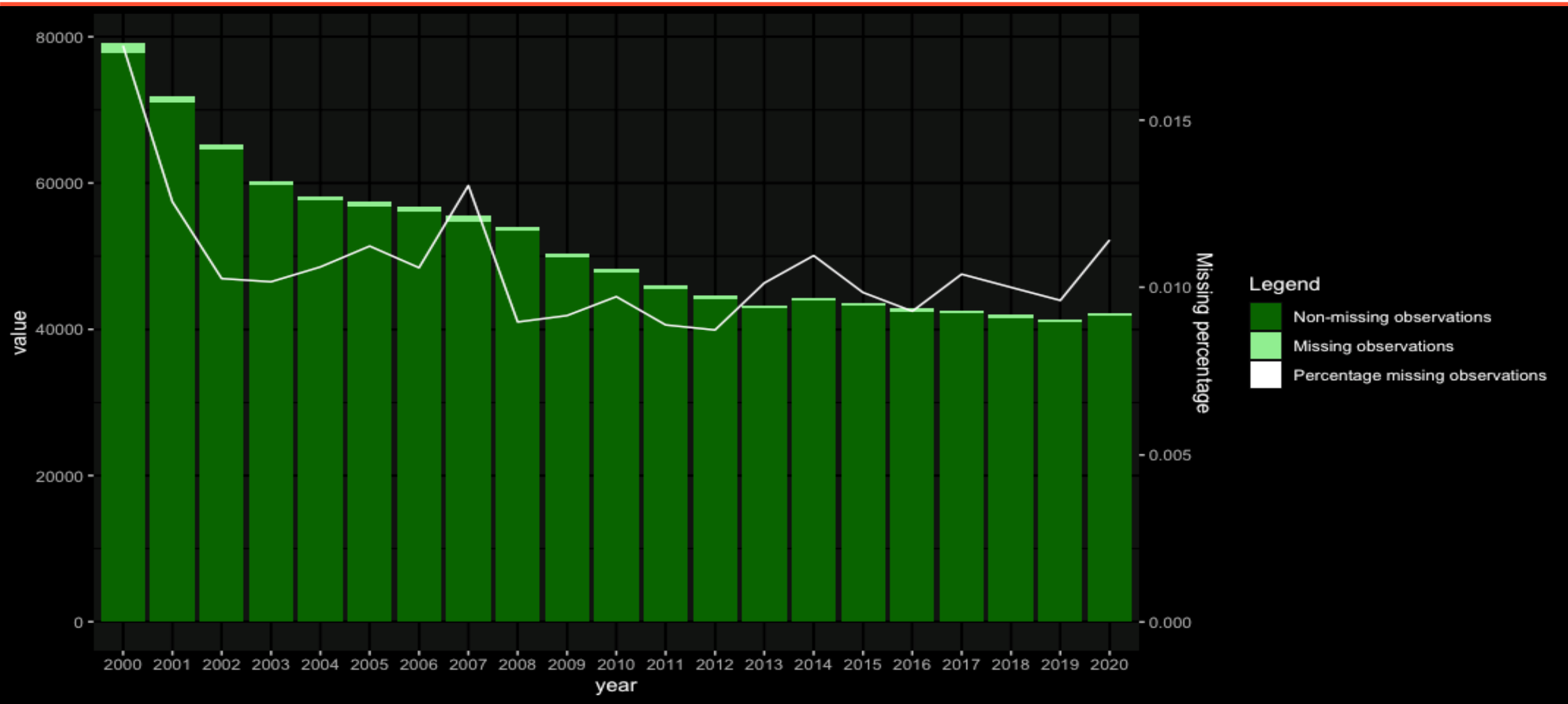
Willem van der Meiden - SNR: 2014236 - ANR: U800370 - w.f.m.vdrmeiden@tilburguniversity.edu

Sven Meijers - SNR: 2002736 - ANR: U407995 - s.j.p.meijers@tilburguniversity.edu

Zeger Vugts - SNR: 2008041 - ANR: U671613 - z.f.vugts@tilburguniversity.edu

## *Question 1*

*Number of stocks with NA return values per year*



## *Question 1*

*Summary Statistics*

# NOTES:

- Tables of summary statistics of either per month or per year.
- Returns are in percentage points
- Market Capitalization as a log-function
- FULL being the summary statistics of the complete sample

| Month | Mean return (% points) | Standard Deviation (% points) | Mean log Market Cap | Standard Deviation Log Market cap |
|---|---|---|---|---|
| 1 | 3.400 | 24.677 | 12.766 | 2.159 |
| 2 | 0.367 | 19.949 | 12.780 | 2.164 |
| 3 | 0.862 | 19.625 | 12.775 | 2.160 |
| 4 | 2.432 | 20.206 | 12.802 | 2.160 |
| 5 | 1.049 | 19.313 | 12.802 | 2.154 |
| 6 | 1.112 | 18.648 | 12.819 | 2.151 |
| 7 | -0.190 | 17.116 | 12.820 | 2.162 |
| 8 | 0.371 | 17.801 | 12.808 | 2.163 |
| 9 | -1.053 | 17.293 | 12.803 | 2.165 |
| 10 | 0.227 | 18.951 | 12.810 | 2.179 |
| 11 | 2.144 | 21.511 | 12.810 | 2.191 |
| 12 | 1.784 | 18.452 | 12.834 | 2.192 |

| Year | Mean return (% points) | Standard Deviation (% points) | Mean log Market Cap | Standard Deviation Log Market cap |
|---|---|---|---|---|
| 2000 | -0.824 | 26.622 | 12.025 | 2.112 |
| 2001 | 2.559 | 27.541 | 11.917 | 2.166 |
| 2002 | -0.824 | 21.354 | 11.958 | 2.166 |
| 2003 | 4.994 | 18.948 | 12.234 | 2.072 |
| 2004 | 1.788 | 15.214 | 12.659 | 1.958 |
| 2005 | 0.411 | 13.464 | 12.789 | 1.953 |
| 2006 | 1.373 | 13.038 | 12.935 | 1.933 |
| 2007 | -0.486 | 13.415 | 12.988 | 1.961 |
| 2008 | -4.416 | 20.015 | 12.616 | 2.076 |
| 2009 | 4.793 | 27.109 | 12.407 | 2.120 |
| 2010 | 2.337 | 15.997 | 12.809 | 2.050 |
| 2011 | -0.611 | 14.905 | 12.970 | 2.096 |
| 2012 | 1.633 | 14.613 | 13.027 | 2.100 |
| 2013 | 3.232 | 13.340 | 13.293 | 2.091 |
| 2014 | 0.396 | 13.921 | 13.440 | 2.063 |
| 2015 | -0.446 | 17.972 | 13.417 | 2.128 |
| 2016 | 1.581 | 17.406 | 13.349 | 2.159 |
| 2017 | 1.366 | 15.384 | 13.506 | 2.174 |
| 2018 | -1.280 | 16.268 | 13.549 | 2.204 |
| 2019 | 1.800 | 20.978 | 13.476 | 2.260 |
| 2020 | 3.427 | 28.618 | 13.473 | 2.250 |
| FULL | 1.042 | 19.598 | 12.802 | 2.167 |

## *Question 2*

*Cumulative Value weighted Portfolio and Equally weighted Portfolio return against Market Factor*

# NOTES:

- **Value Weighted portfolio**
  - Weighted relative to the market capitalization of the firm → same as market return!

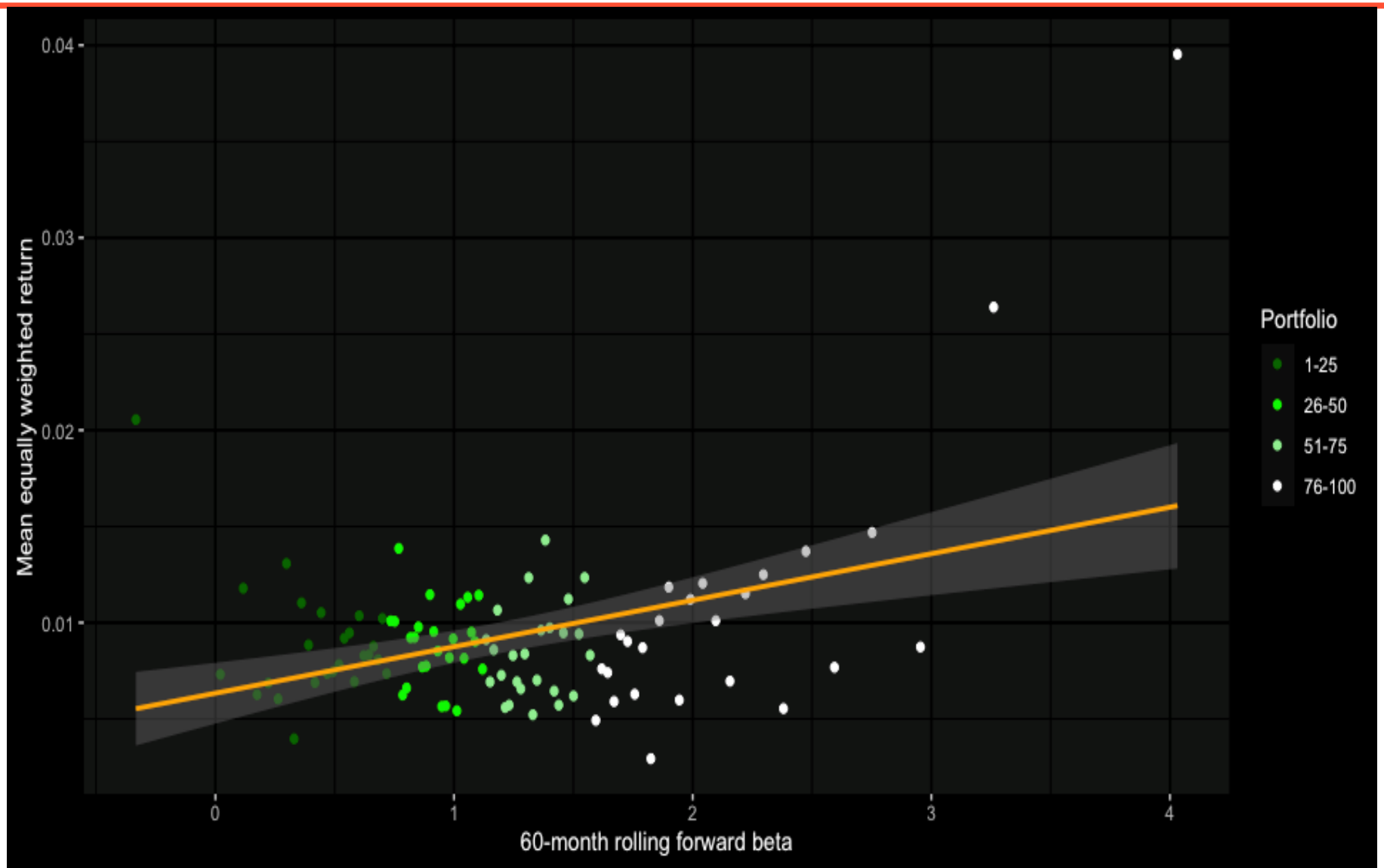
- **Equally Weighted portfolio**
  - Each stock same weight → Proves smaller firms have bigger returns

## *Question 3*

*Security Market Line of 100 CAPM Beta portfolios (without winsorizing)*

# NOTES:

- 100 portfolios
  - Sorted from lowest to highest Beta
- Small Positive Slope of Security Market Line
  - The higher the Beta, the riskier the portfolio (higher distance between portfolios!)
- Some extreme observations!

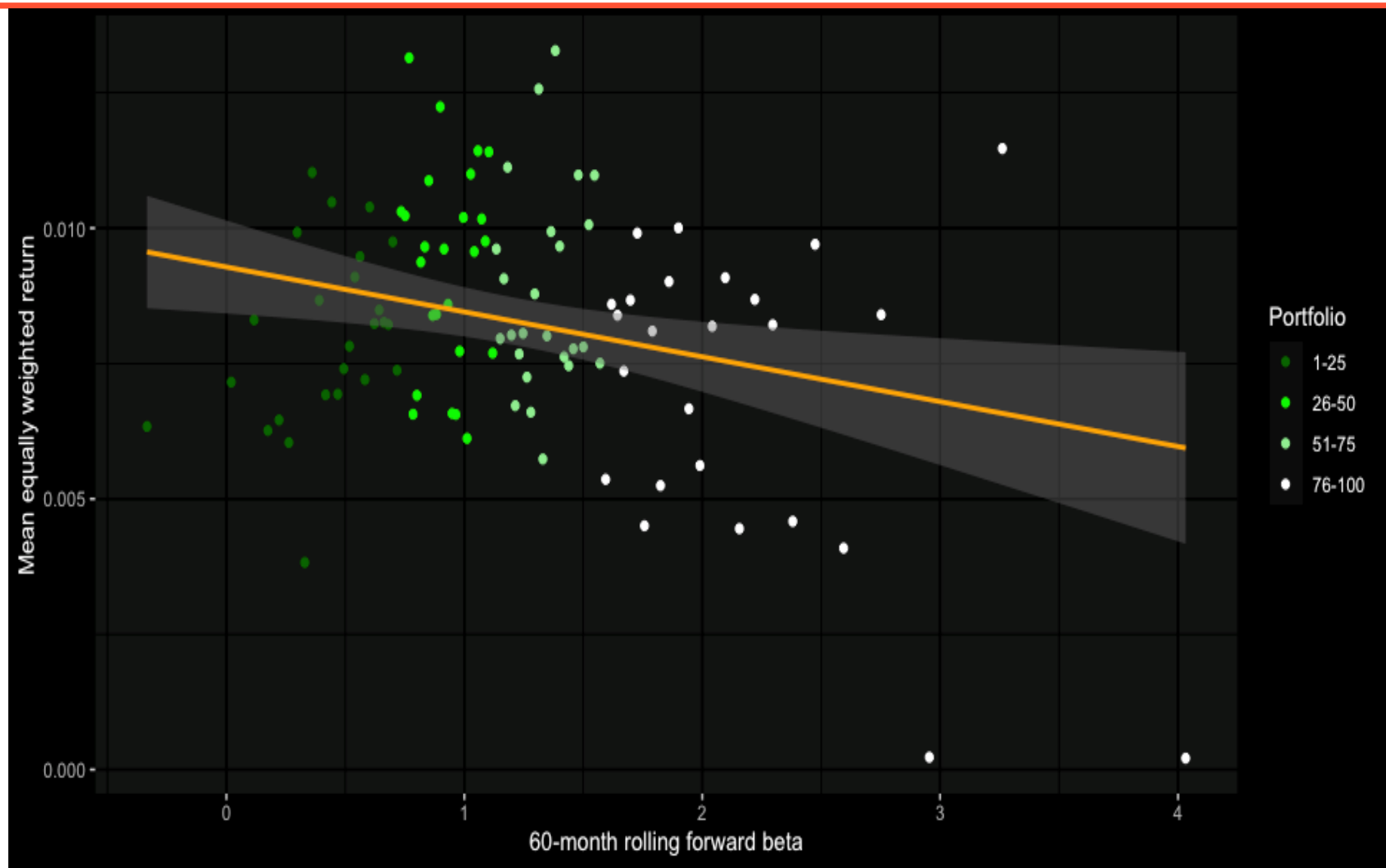


## *Question 3*

*Security Market Line of 100 CAPM Beta portfolios (with winsorizing)*

# NOTES:

- 100 portfolios
  - Sorted from lowest to highest Beta
- Winsorized at 5%-level
  - Eliminate extreme observations
  - More reliable sample
- CAPM does not hold!
  - Higher Beta means riskier
  - Higher distance between highest 25 portfolios
  - Does not translate into higher return!



## *Question 4*

*Fama MacBeth regression on Oil Betas*

# NOTES:

- T-statistic Oil Beta = 0.7606
  - Highly statistically insignificant
- Using a 60-month rolling forward Oil Beta, resulting in a sample period of 2005 – 2012, and a Fama MacBeth Two stage regression model, we can conclude that the Oil Beta does not predict stock returns!

**Two stage Fama-McBeth regression**

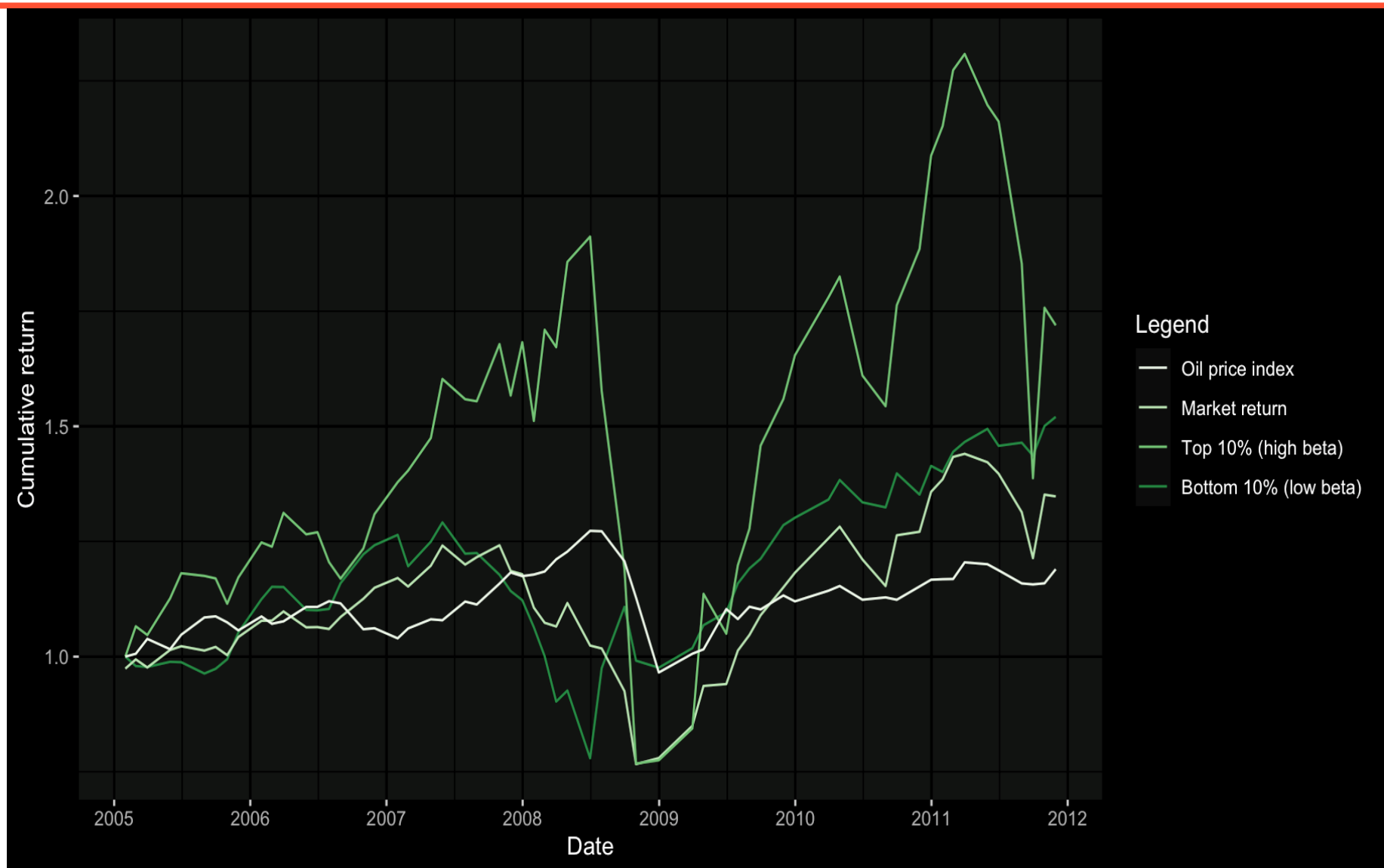
| | *Dependent variable:* |
|---|---|
| | Firm Excess Return |
| Oil Beta | 0.013 |
| | (0.018) |
| Constant | -0.001 |
| | (0.005) |
| Observations | 145,824 |
| $R^2$ | 0.215 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

## *Question 5*

Cumulative Value Weighted return – Top 10% and Bottom 10% Oil Beta Portfolios against Oil Index

# NOTES:

- Higher Oil Beta's result in higher, but riskier, returns depending on the oil price movement
- Lower Oil Beta's experience lower returns and less co-movement with oil prices!

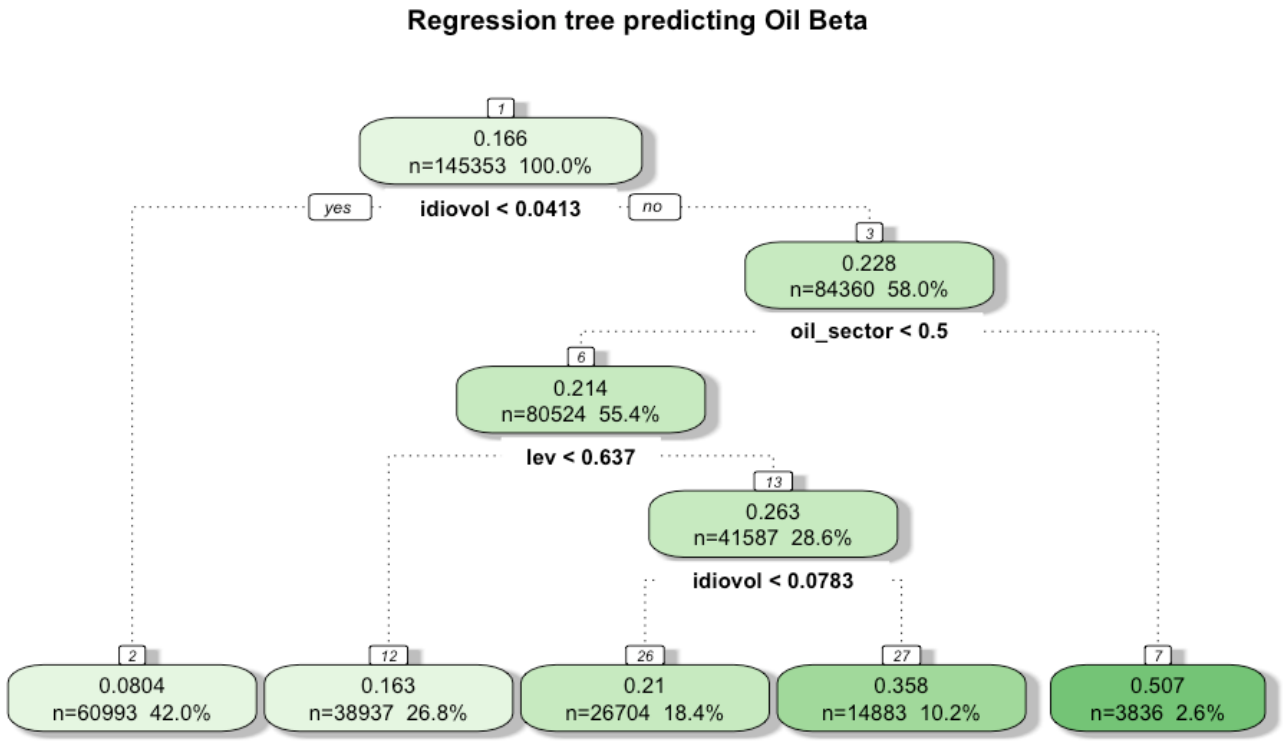


## *Question 6*

*Predicting Oil Beta's using a Regression Tree*

# NOTES:

- If the Firm-Specific Volatility of a firm is lower than 4.127 percentage points (monthly basis), oil beta is 0.0804 (1 percentage point change in il return changes firm return by 0.08 percentage points
- If the Firm-Specific Volatility is higher than 4.127 percentage points and the Firm is in an oil-related sector, oil beta is 0.5070
- If Firm-Specific Volatility is higher than 4.127 percentage points, the Firm is not in an oil-related sector, and leverage lower than 0.637, oil beta is 0.1626
- If Firm-Specific Volatility is higher than 4.127 percentage points, the Firm is not in an oil-related sector, and leverage higher than 0.637, oil beta is 0.2097 if the Firm-Specific Volatility lower than 7.83 percentage points and 0.3583 if the Firm-Specific Volatility is higher than 7.83 percentage points.




*Reasonable Cutoff = 20% of observations NA*
*Remaining NA = cross-sectional median*

Legend:
- Idiovol: Firm-specific volatility (Firm-specific riskiness)
- Oil-sector: Sectors directly related to oil ( 1 if in the oil & gas industry, air transport, and consumer goods)
- Lev = Leverage Ratio (Debt/Equity)
- n = amount of sample predicted to have that value

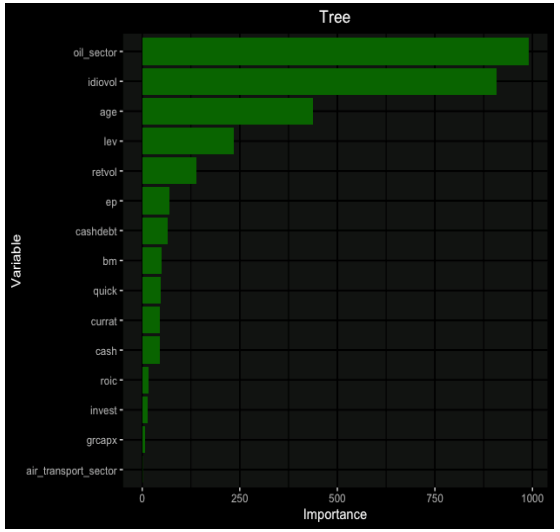
## *Question 7*

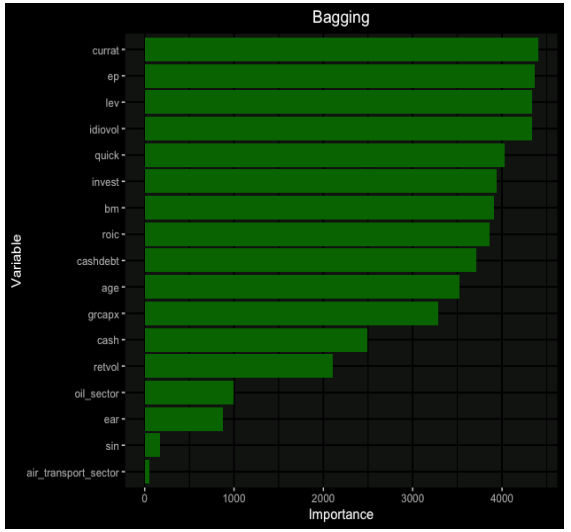
*Predict if a stock is in the least or most sensitive Oil Beta portfolio using a Classical Tree, Bagging, Random Forest, and Lasso*

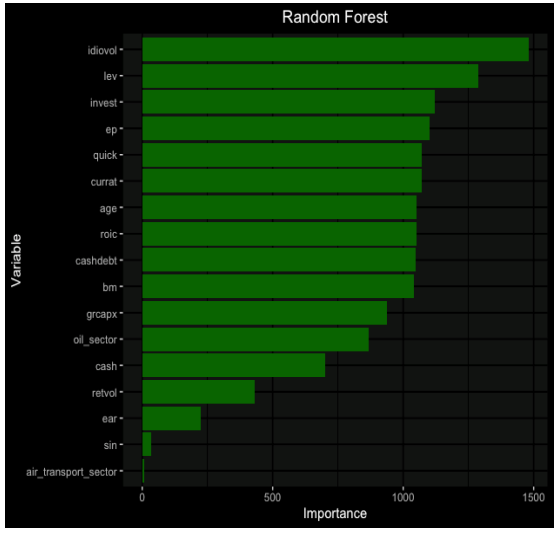
| Model | Classical Tree | Bagging | Random Forest | Lasso |
|---|---|---|---|---|
| Error Rate | 28,39% | 26,86% | 24,91% | 30,05% |
| Assumptions | Pruned to 5 terminal nodes (lowest CV error rate) | Number of bags leading to lowest error rate = 200 | Number of trees leading to lowest error rate = 200 | Alpha = 0.1, Lambda = 0.00396 |

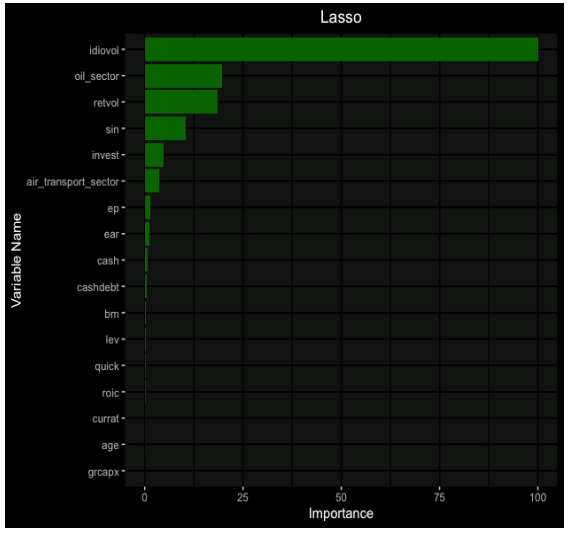
***How is the error distributed?***
***Let's look at the Confusion Matrix on the next slide.***

## Variable Importance Measures per model



## *Question 7*

*Predict if a stock is in the least or most sensitive Oil Beta portfolio using a Classical Tree, Bagging, Random Forest, and Lasso*

# NOTES:

- YES
  - Stock is in the least sensitive to oil prices portfolio (i.e. lowest oil beta)
- NO
  - Stock is in the most sensitive to oil prices portfolio (i.e. highest oil beta)

Interpretation:
The models have more false "Yes" than false "No"! In other words, the models are more biased towards predicting a stock to be in the least sensitive to oil price portfolio.

| Predicted Class | True Class: Yes | True Class: No |
|---|---|---|
| Yes | Classical Tree = 10.862 (42.47%)<br>Bagging = 9.721 (38.01%)<br>Random Forest = 10.029 (39.21%)<br>Lasso = 11.410 (44.62%) | Classical Tree = 5.286 (20.67%)<br>Bagging = 3.795 (14.84%)<br>Random Forest = 3.730 (14.59%)<br>Lasso = 6.374 (24.92%) |
| No | Classical Tree = 1.974 (7.72%)<br>Bagging = 3.115 (12.18%)<br>Random Forest = 2.807 (10.98%)<br>Lasso = 1.426 (5.58%) | Classical Tree = 7.451 (29.14%)<br>Bagging = 8.942 (34.97%)<br>Random Forest = 9.007 (35.22%)<br>Lasso = 6.363 (24.88%) |

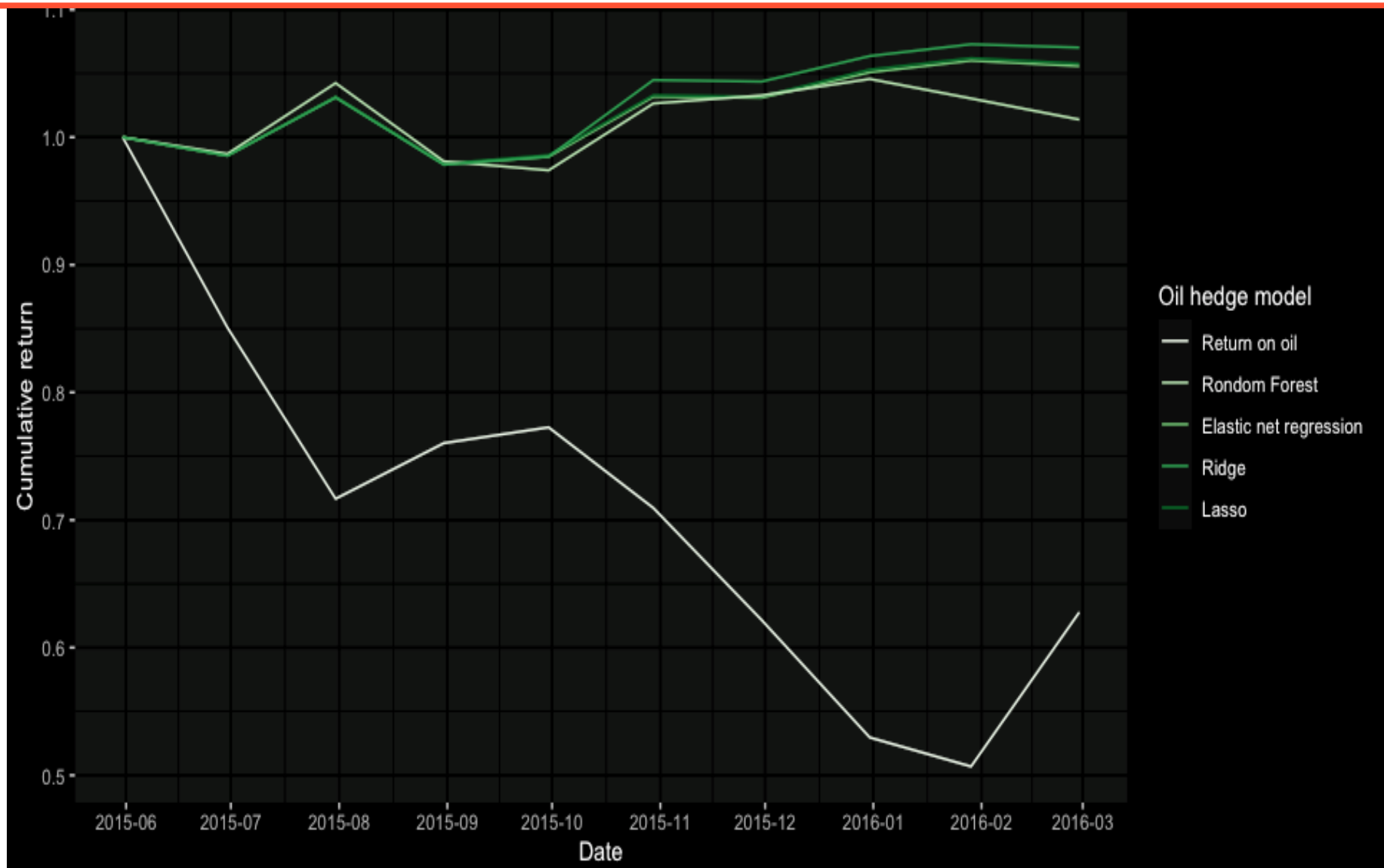
*Sample = 25573 observations (test data 2012 – 2020)*
*Total True No = 12737 observations (49,8%)*
*Total True Yes = 12836 observations (50,2%)*
*(%) behind number of observations = % of total observations*

## *Question 8*

Performance Hedge Portfolios based on different models

# NOTES:

- All models used to create the hedge portfolios are trained using two large drops in oil prices during the period 2008-06 until 2009-01 and 2014-06 until 2015-01
- All hedge portfolios do pretty well when tested on the oil price drop in the period of 2015-06 until 2016-02



## *Question 8*

*Summary Statistics Hedge Portfolio post 2016 (4 portfolios selected using 4 different models)*

| Model | Lasso | Ridge | Elastic Net | Random Forest | Oil Return |
|---|---|---|---|---|---|
| Mean Return | 0.667 | 0.805 | 0.645 | 0.210 | -4.309 |
| Standard Deviation Return | 3.036 | 3.273 | 3.037 | 3.598 | 13.088 |
| Min. | -5.081 | -5.102 | -5.092 | -5.882 | -15.776 |
| Max. | 4.760 | 6.096 | 4.775 | 5.620 | 5.620 |
| Mean Beta | 0.470 | 0.479 | 0.475 | 0.641 | - |
| Mean Oil Beta | -0.058 | -0.055 | -0.057 | -0.067 | - |

*Numbers are in percentage points*
*Returns are monthly returns*

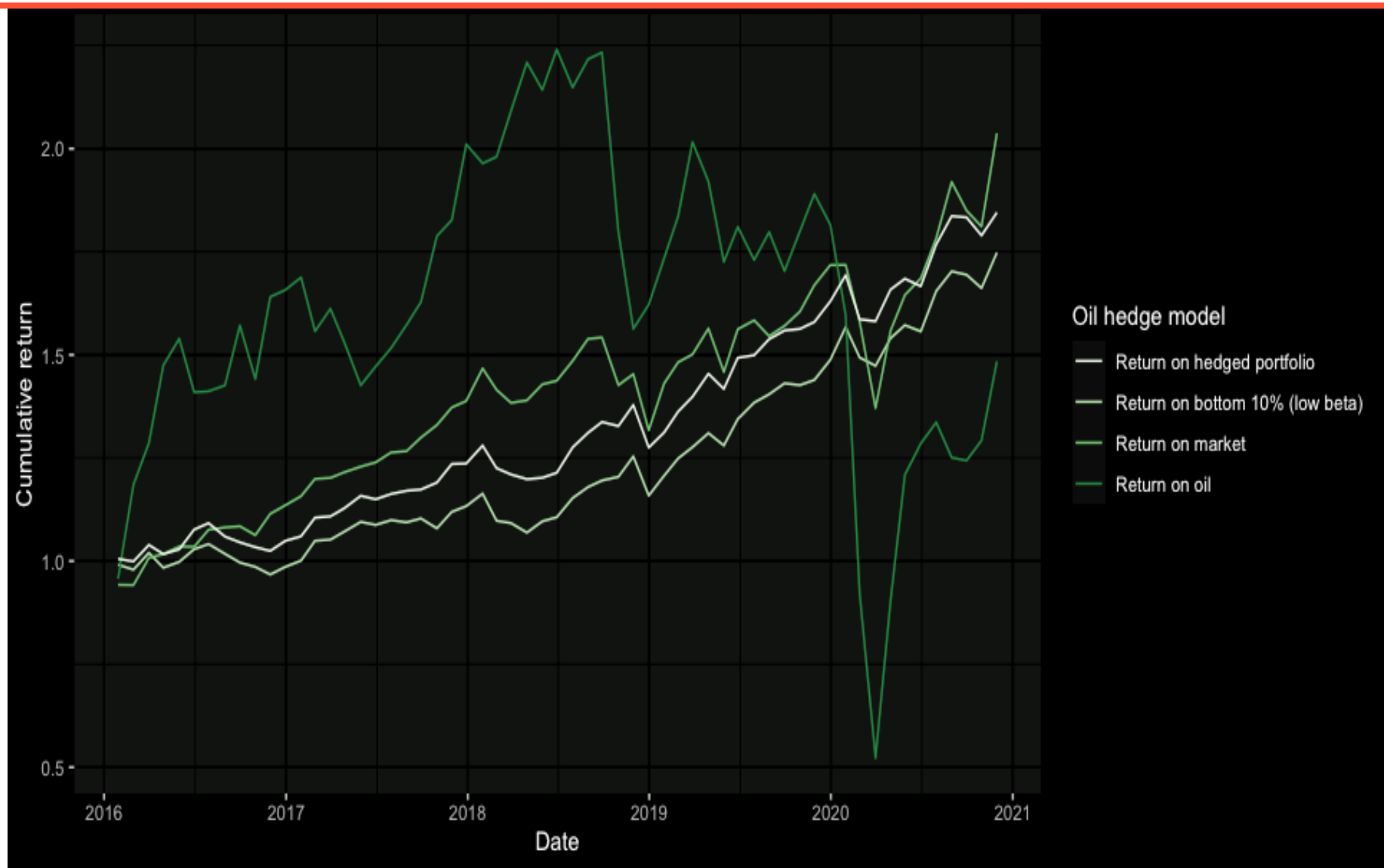
## *Question 9*

Cumulative Return Oil Hedge Portfolio and Bottom 10% Oil Beta against Market

# NOTES:

- Method used for hedged portfolio = Elastic Net Regression

- Machine learning does help at identifying stocks that have certain exposure to oil price shocks. As seen here, the hedged portfolio trained on various oil price shocks does not move with oil prices, and slightly outperforms the bottom 10% oil beta portfolio.



## *Question 10*

Optimal Portfolio for sovereign wealth funds with large oil reserves

# NOTES:

- Two large oil price drops are included in the graph
- As shown, when oil price and market price drop during these shocks, the optimal hedged portfolio remains stable or even increases, suggesting it is a working hedge.