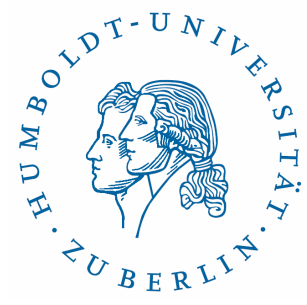
Replication of the Fama-French Three-Factor Model and Empirical Study of the S&P 500

Xin Tan
Daria Skidnova
Jeff Giddens

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://case.hu-berlin.de



Structure

Introduction
Theory and Design
- The Fama-French Three-Factor Model
- Data Preparation

Replicating the Three-Factor Model
- Batch Regression
- Formatting the Results
- Comparison with the Original Fama-French Results

Applying the Fama-French Model to the S&P 500
- Data Preparation
- Fama-French in 2010-2017
- Trend Analysis in 1980-2015
- Varying Sample Sizes
- Correlation of The factors and Top 20 and Bottom 20 returns

Conclusion

Purpose of the Study

- Replicate the Fama-French Three-Factor model (1993)
- Apply the model to S&P 500 data from 2010-2017
- Assess the explanatory power for the Fama-French Three-Factor model over five year periods from 1980-2015

Introduction

Theory and Design
- The Fama-French Three-Factor Model
- Data Preparation

Replicating the Three-Factor Model

Applying the Fama-French Model to the S&P 500

Conclusion

Theory and Design

- The Capital Asset Pricing Model:

$$R_i - R_F = \beta_M \cdot (R_M - R_F)$$

- The Fama-French Three-Factor Model:

$$R_i - R_F = \beta_M \cdot (R_M - R_F) + \beta_S \cdot SMB + \beta_V \cdot HML$$

  - $R_I - R_F$ is an excess return of the stock or portfolio
  - β_M is the sensitivity coefficient
  - SMB (Small Minus Big) is a size factor (in terms of market capitalization)
  - HML (High Minus Low) is a value factor (in terms of book-to-value ratio)

Data Preparation for Model Replication

Data:[1]

- Fama/French 3 Factors (monthly)
- 25 Portfolios Formed on Size and Book-to-Market (5x5) (Value-Weighted)

[1] source: Fama French Homepage

Fama-French 3 Factors

Date	Mkt-RF	SMB	HML	RF
192607	2.96	-2.30	-2.87	0.22
192608	2.64	-1.40	4.19	0.25
...				
199112	10.84	-2.22	-4.01	0.38

Table 1: Fama/French 3 Factors from July 1926 to December 1991 based on monthly return

Source: CRSP firms listed on the NYSE, AMEX, and NASDAQ

Calculation of SMB and HML Factors

$$
\begin{aligned}
SMB =& \frac{1}{3}(SmallValue + SmallNeutral + SmallGrowth) - \\
& \frac{1}{3}(BigValue + BigNeutral + BigGrowth) \\
HML =& \frac{1}{2}(SmallValue + BigValue) - \frac{1}{2}(SmallGrowth + BigGrowth)
\end{aligned}
$$

using 6 value-weight portfolios formed on size on book-to-market:

Small Growth	Small Neutral	Small Value
Big Growth	Big Neutral	Big Value

25 Value-Weighted Portfolios

	Low	2	3	4	High
Small	SMALL LoBM	ME1 BM2	ME1 BM3	ME1 BM4	SMALL HiBM
2	ME2 BM1	ME2 BM2	ME2 BM3	ME2 BM4	ME2 BM5
3	ME3 BM1	ME3 BM2	ME3 BM3	ME3 BM4	ME3 BM5
4	ME4 BM1	ME4 BM2	ME4 BM3	ME4 BM4	ME4 BM5
Big	BIG LoBM	ME5 BM2	ME5 BM3	ME5 BM4	BIG HiBM

Table 2: Structure of the 25 Value-Weighted Portfolios

- High BM (Book to Market) stocks are value stocks
- Big ME (Market Equity) stocks are large companies by market capitalization
- Each value is that portfolio's monthly percentage return

Introduction

Theory and Design

Replicating the Three-Factor Model
- Batch Regression
- Formatting the Results
- Comparison with the Original Fama-French Results

Applying the Fama-French Model to the S&P 500

Conclusion

Batch Regression

- Use multivariate regression

```
# batch regressing 25 portfolios
results = list()
# Data starts from the 2nd col of P25
for (i in 1:(ncol(P25) - 1)) {
    rirf = unlist (P25[, i + 1]) - rf
       y = lm (rirf ~ rmrf + smb + hml)
  results[[i]] = summary (y)
}
```


Formatting the Results

- Save all coefficients of regression results as vectors

```
for(i in 1:(ncol(P25)-1)) {
        betas = cbind(betas,results[[i]]$coefficients
          [,1])
  std.errors = cbind(std.errors,results[[i]]$sigma)
    t.values = cbind(t.values, results[[i]]$
      coefficients[,3])
  R.squareds = cbind(R.squareds, results[[i]]$adj.r.
    squared)
}
```


Replicating Table 6 (1993)

- Resize the output to 5x5 format as in the Fama-French paper

```
resize = function(x) {
              df = data.frame (matrix(x, nrow = 5,
                byrow = TRUE))
  colnames(df) = c("Low", "2", "3", "4", "High")
  rownames(df) = c("Small", "2", "3", "4", "Big")
  return(df)
}
```


b	LOW	2	3	4	HIGH
SMALL	1.03	0.97	0.94	0.89	0.95
2	1.10	1.02	0.96	0.97	1.07
3	1.10	1.02	0.97	0.97	1.06
4	1.06	1.07	1.04	1.03	1.15
BIG	0.96	1.02	0.96	1.01	1.03

t(b)	LOW	2	3	4	HIGH
SMALL	39.23	50.60	58.42	57.99	57.76
2	53.20	58.56	59.98	62.77	63.25
3	59.68	56.81	53.35	58.93	51.14
4	57.16	52.61	50.34	51.30	46.30
BIG	57.20	56.98	42.80	55.04	37.70

s	LOW	2	3	4	HIGH
SMALL	1.40	1.27	1.16	1.10	1.19
2	1.00	0.94	0.83	0.71	0.85
3	0.70	0.63	0.54	0.45	0.65
4	0.30	0.27	0.25	0.22	0.36
BIG	(0.20)	(0.19)	(0.27)	(0.19)	(0.04)

t(s)	LOW	2	3	4	HIGH
SMALL	35.61	44.82	48.65	48.10	48.63
2	32.62	36.36	34.80	30.78	33.82
3	25.53	23.41	20.04	18.46	21.03
4	10.92	8.75	8.06	7.49	9.64
BIG	(8.10)	(7.08)	(7.99)	(6.91)	(1.05)

h	LOW	2	3	4	HIGH
SMALL	(0.30)	0.08	0.27	0.38	0.62
2	(0.48)	0.03	0.23	0.47	0.70
3	(0.43)	0.04	0.31	0.50	0.71
4	(0.44)	0.03	0.30	0.56	0.74
BIG	(0.44)	(0.02)	0.20	0.56	0.76

t(h)	LOW	2	3	4	HIGH
SMALL	(6.77)	2.43	9.92	14.93	22.43
2	(13.93)	0.88	8.73	18.32	24.74
3	(14.04)	1.39	10.27	18.28	20.34
4	(14.24)	0.79	8.77	16.68	17.79
BIG	(15.96)	(0.68)	5.25	18.41	16.65

Table 3: Results of the Fama-French Three-Factor Model

Table 6

Regressions of excess stock and bond returns (in percent) on the excess market return ($RM-RF$) and the mimicking returns for the size (SMB) and book-to-market equity (HML) factors: July 1963 to December 1991, 342 months.*

$$R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + e(t)$$

Dependent variable: Excess returns on 25 stock portfolios formed on size and book-to-market equity

Size quintile	Book-to-market equity (BE/ME) quintiles									
	Low	2	3	4	High	Low	2	3	4	High
	b					$t(b)$				
Small	1.04	1.02	0.95	0.91	0.96	39.37	51.80	60.44	59.73	57.89
2	1.11	1.06	1.00	0.97	1.09	52.49	61.18	55.88	61.54	65.52
3	1.12	1.02	0.98	0.97	1.09	56.88	53.17	50.78	54.38	52.52
4	1.07	1.08	1.04	1.05	1.18	53.94	53.51	51.21	47.09	46.10
Big	0.96	1.02	0.98	0.99	1.06	60.93	56.76	46.57	53.87	38.61
	s					$t(s)$				
Small	1.46	1.26	1.19	1.17	1.23	37.92	44.11	52.03	52.85	50.97
2	1.00	0.98	0.88	0.73	0.89	32.73	38.79	34.03	31.66	36.78
3	0.76	0.65	0.60	0.48	0.66	26.40	23.39	21.23	18.62	21.91
4	0.37	0.33	0.29	0.24	0.41	12.73	11.11	9.81	7.38	11.01
Big	− 0.17	− 0.12	− 0.23	− 0.17	− 0.05	− 7.18	− 4.51	− 7.58	− 6.27	− 1.18
	h					$t(h)$				
Small	− 0.29	0.08	0.26	0.40	0.62	− 6.47	2.35	9.66	15.53	22.24
2	− 0.52	0.01	0.26	0.46	0.70	− 14.57	0.41	8.56	17.24	24.80
3	− 0.38	− 0.00	0.32	0.51	0.68	− 11.26	− 0.05	9.75	16.88	19.39
4	− 0.42	0.04	0.30	0.56	0.74	− 12.51	1.04	8.83	14.84	17.09
Big	− 0.46	0.00	0.21	0.57	0.76	− 17.03	0.09	5.80	18.34	16.24

Table 4: Results of the Fama-French Three-Factor Model[2]

[2]Source: Fama and French 1993b, Table 6

Introduction

Theory and Design

Replicating the Three-Factor Model

Applying the Fama-French Model to the S&P 500
- Data Preparation
- Fama-French in 2010-2017
- Trend Analysis in 1980-2015
- Varying Sample Sizes
- Correlation of The factors and Top 20 and Bottom 20 returns

Conclusion

Data Preparation

- Download stocks information

Symbol	Name	Sector	Abbv.
AAP	Advance Auto Parts	Consumer Discretionary	CD
MO	Altria Group Inc	Consumer Staples	CS
APC	Anadarko Petroleum Corp	Energy	E
AMG	Affiliated Managers Group Inc	Financials	FI
ABT	Abbott Laboratories	Health Care	H
MMM	3M Company	Industrials	I
ACN	Accenture plc	Information Technology	IT
APD	Air Products & Chemicals Inc	Materials	M
ARE	Alexandria Real Estate Equities Inc	Real Estate	RE
CTL	CenturyLink Inc	Telecommunication Services	T
AES	AES Corp	Utilities	U

Table 5: Summary information, including sector designation

Retrieving S&P500 Data

- Download SP 500 adjusted stock prices

```
stocks=BatchGetSymbols(tickers = companies$tickers,
                       first.date = "2010-01-01",
                        last.date = "2017-12-31")

# Select the good tickers
good.tickers = stocks$df.control$ticker [stocks$df.
  control$threshold.decision == "KEEP"]
```

- stocks is a list that contains 2 dataframes:
  - df.control contains descriptive information
  - df.tickers contains the downloaded price data

Creating a Data Frame of S&P 500 Prices

- Use the dates of 3M as the date column of the data frame
- Merge price data of stocks into the data frame

```
SP500.data = data.frame(date = stocks$df.tickers[
  stocks$df.tickers$ticker == "MMM", "ref.date"])

for(i in 1:length(good.tickers)){
    X = data.frame(
    stocks$df.tickers[stocks$df.tickers$ticker ==
      good.tickers[i], c("ref.date", "price.adjusted
      ")])
    colnames(X) = c("date", as.character(good.
      tickers[i]))
    SP500.data = merge.data.frame(SP500.data, X, by
      = "date", all.x = TRUE)
}
```


Convert Daily Data to Monthly Return

- Convert downloaded daily data to monthly price data series into XTS series
- quantmod::monthlyReturn()requires non-NA daily prices in xts format

```
1   Stock.Prices.Daily = xts(Stock.Prices.Daily[,-1],
      order.by = as.POSIXct(Stock.Prices.Daily$date))
2
3   Stock.Prices.Daily = Stock.Prices.Daily[!is.na(
      Stock.Prices.Daily), ]
4
5 Stock.Prices.Monthly = monthlyReturn(Stock.Prices.
    Daily)
```


Data Cleaning

- **Removing stocks with NAs** in the series ensures that remaining stocks have same number of observations
- **Remove NAs in each series** results in a smaller sample size

Price data with NAs in the middle would result in inaccurate monthly returns.

(BHY Brighthouse Financial Inc. removed for 2015-2017 runs)

Here we face choices:

1. Remove all columns with NAs, then all remaining stocks could have the regression in the same period, i.e. with the same number of observations. (2010-2017)
2. Dynamically frame the data based on the available non-NA data points, but then some stocks in the regression analysis will have fewer observations. (1980-2015 every 5 years case)

Fama-French in 2010-2017

```
SP500.data = read.csv("Data/SP500_price.adjusted_
  2010-2017.csv")
SP500.data$date = as.Date(SP500.data$date)

# ...

Results = list()
for(i in 1:ncol(Stock.Prices.Monthly)) {
  RiRF = Stock.Prices.Monthly[,i] - FF$RF
  Regression = lm(RiRF ~ FF$Mkt.RF + FF$SMB + FF$HML
    )
  Results[[i]] = summary(Regression)
}
```


Boxplot of the Regression Results

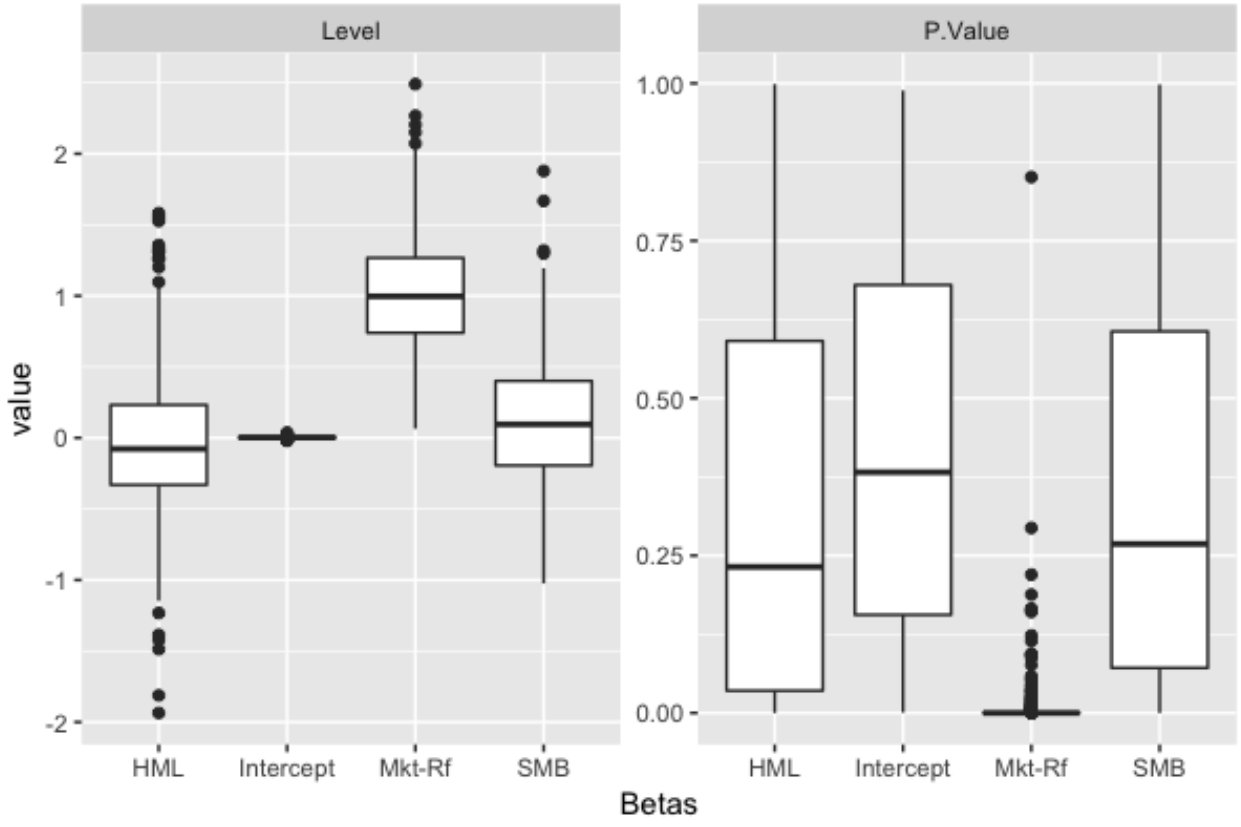


Figure 1: Regression results of S&P 500 companies from 2010-2017

Goodness of Fit

	Ticker	Name	Sector	R^2
201	HON	Honeywell Int'l Inc.	Industrials	0.76
227	IVZ	Invesco Ltd.	Financials	0.73
31	AMG	Affiliated Managers Group Inc	Financials	0.71
334	PRU	Prudential Financial	Financials	0.71
388	TROW	T. Rowe Price Group	Financials	0.71
60	BEN	Franklin Resources	Financials	0.71
386	TMK	Torchmark Corp.	Financials	0.70
270	MET	MetLife Inc.	Financials	0.70
321	PFG	Principal Financial Group	Financials	0.69
283	MS	Morgan Stanley	Financials	0.66

Table 6: R^2 Values of S&P 500 Companies

Coefficients by Sector

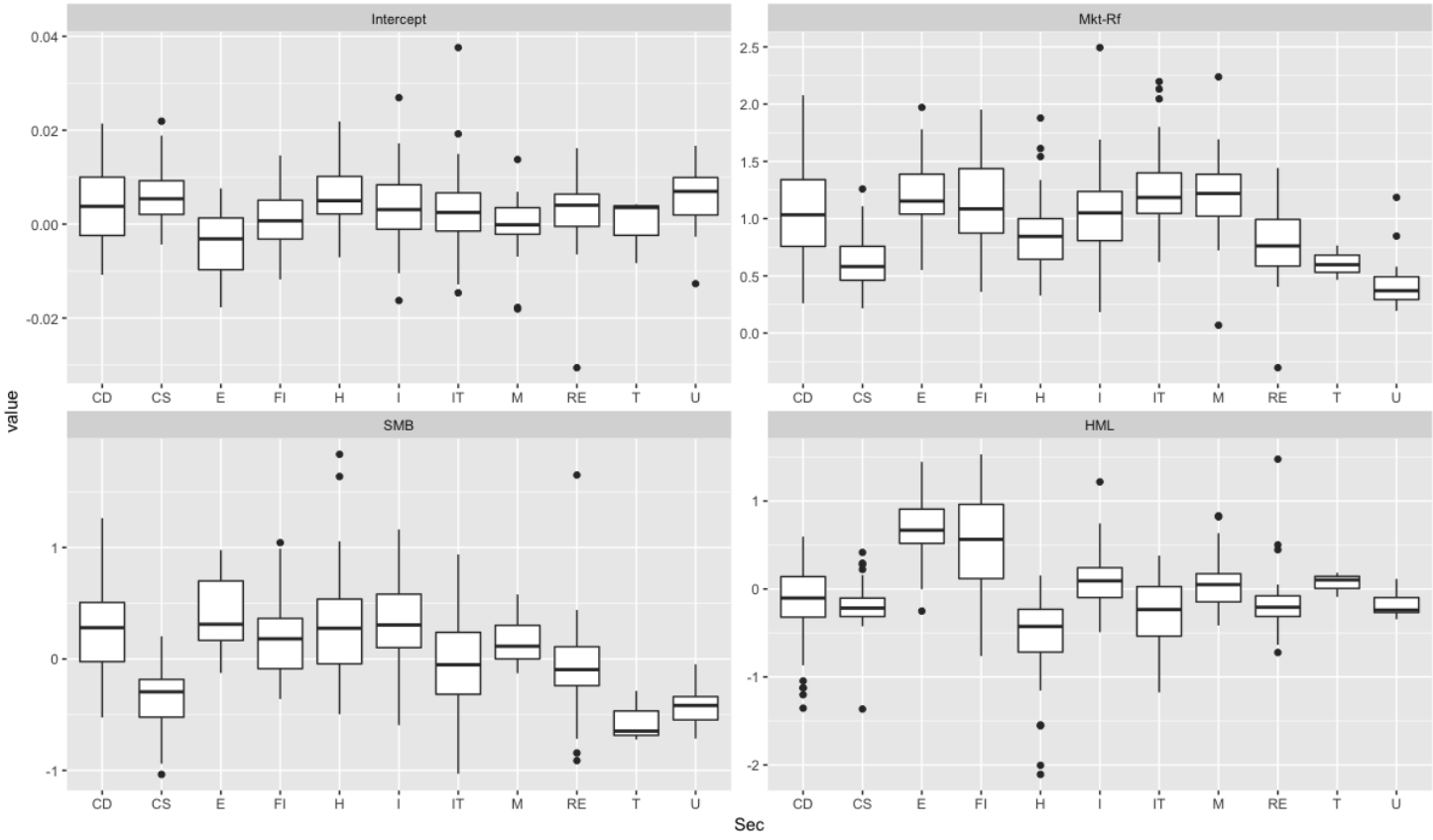


Figure 2: Coefficients by sector

P-Values

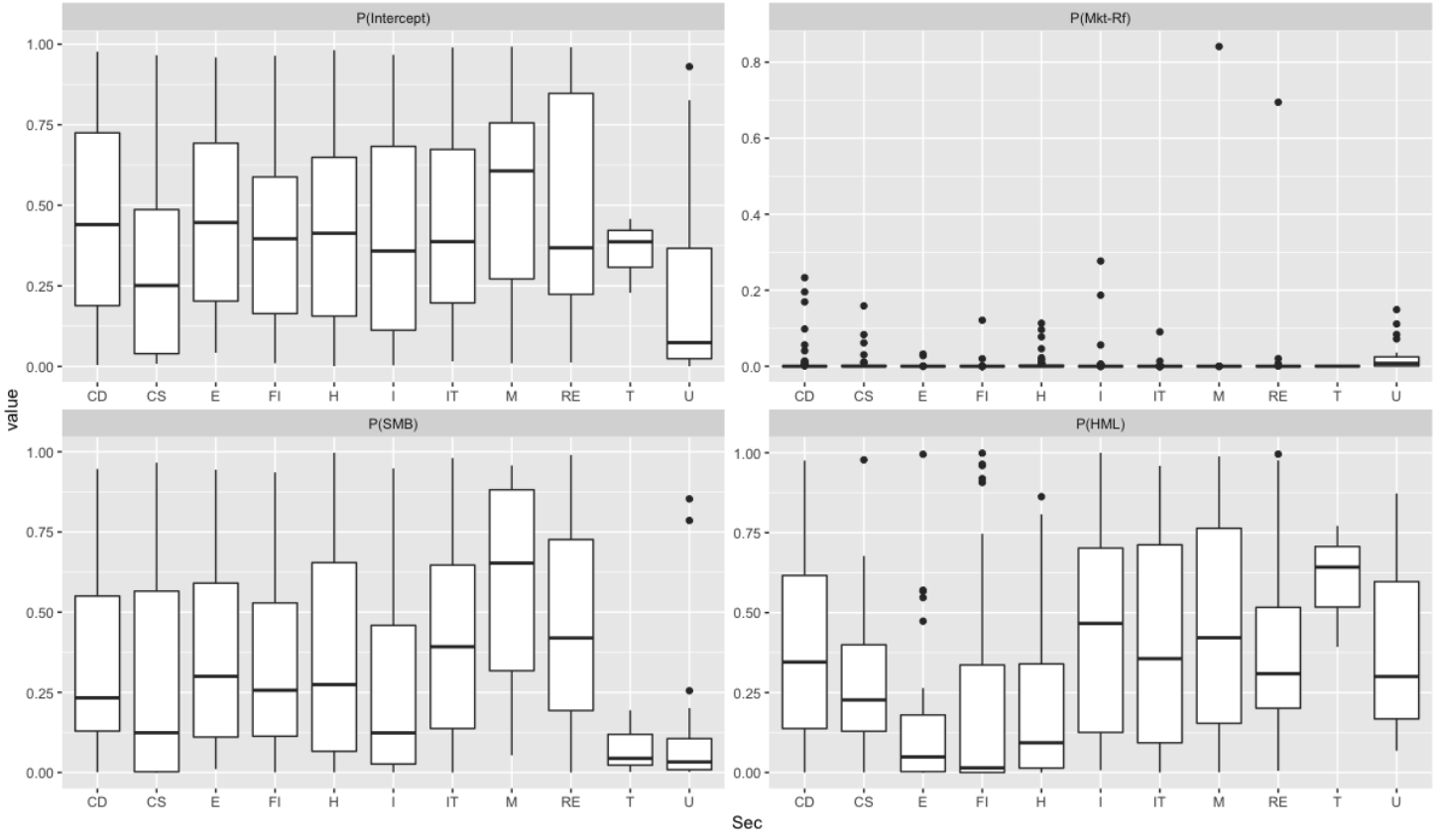


Figure 3: P-Values by sector

Trend analysis in 1980-2015

- loop over above codes to download data from 1980 - 2015, group every 5 years

```
1 library(lubridate)
2 List.of.start.date = seq(as.Date("1980/1/1"), as.
    Date("2016/1/1"), "years")
3 List.of.start.date = List.of.start.date[year(List.of
    .start.date)%%5==0]
```

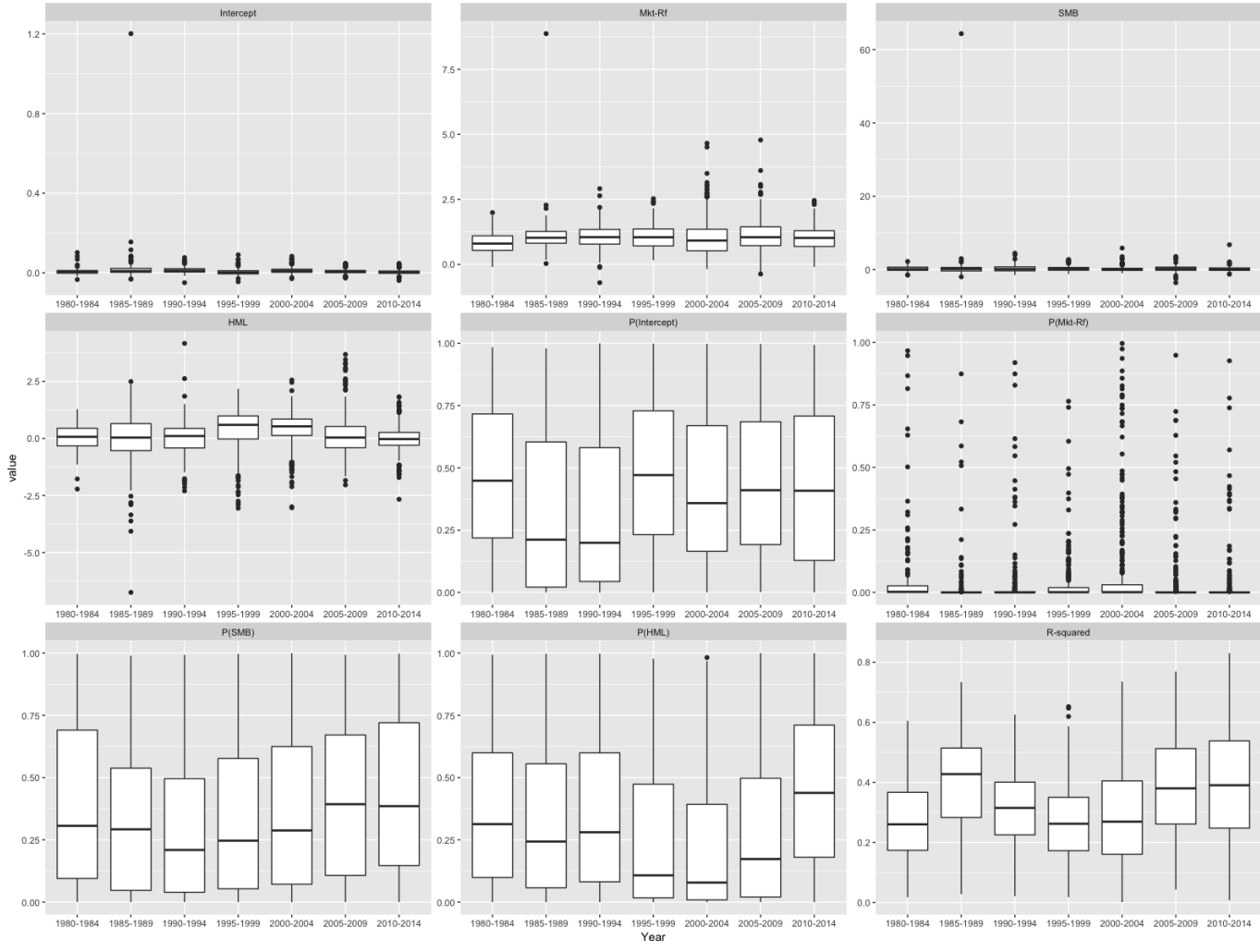



Figure 4: Trend analysis in 1980-2015 of full sample

Varying Samples

Time Period	Number of Stocks
1980-1984	170
1985-1989	229
1990-1994	271
1995-1999	345
2000-2004	394
2005-2009	432
2010-2014	459

Table 7: Number of S&P 500 stocks in the sample in each period

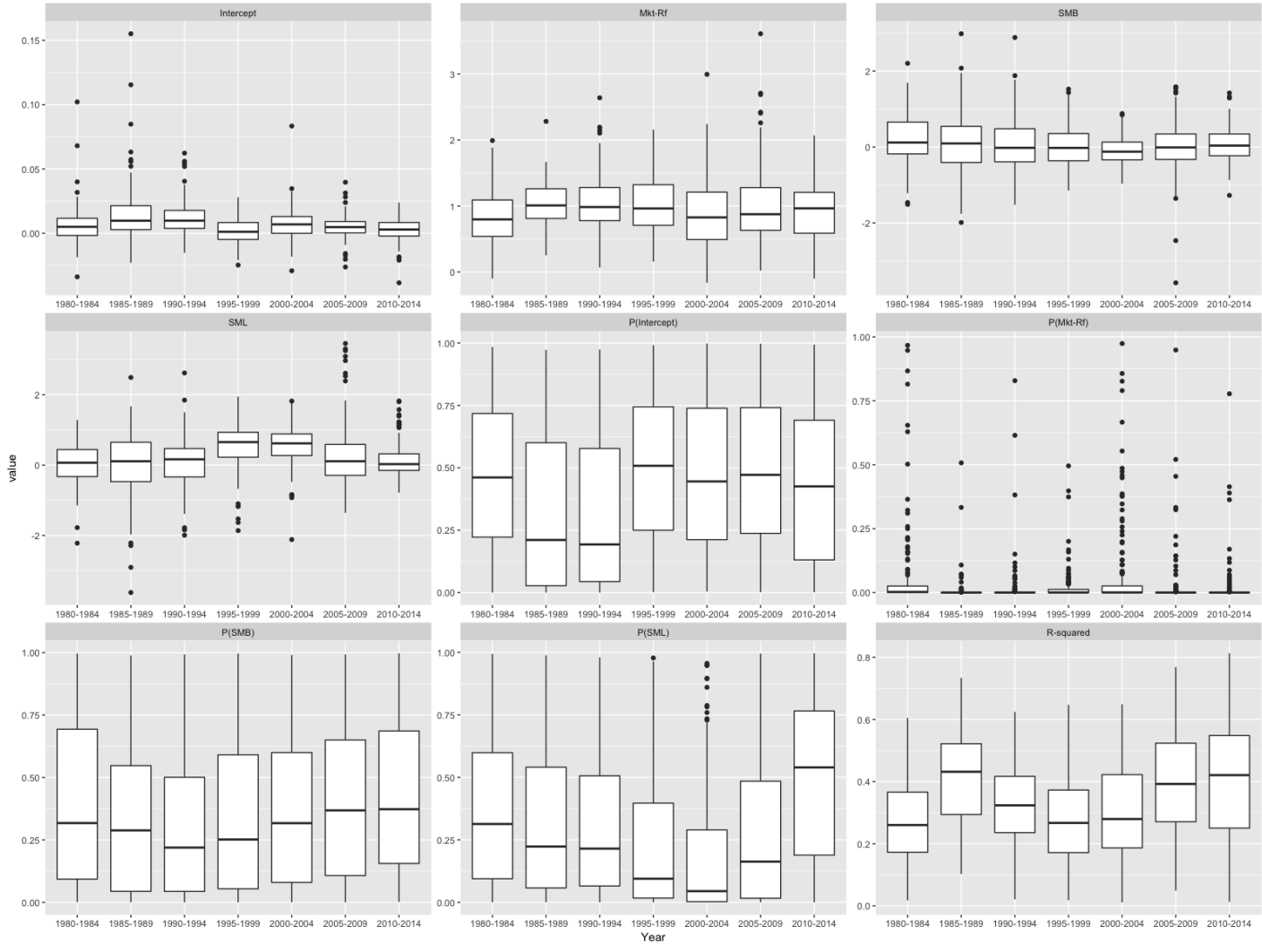


Figure 5: Trend analysis of 168 stocks listed from 1980-2015

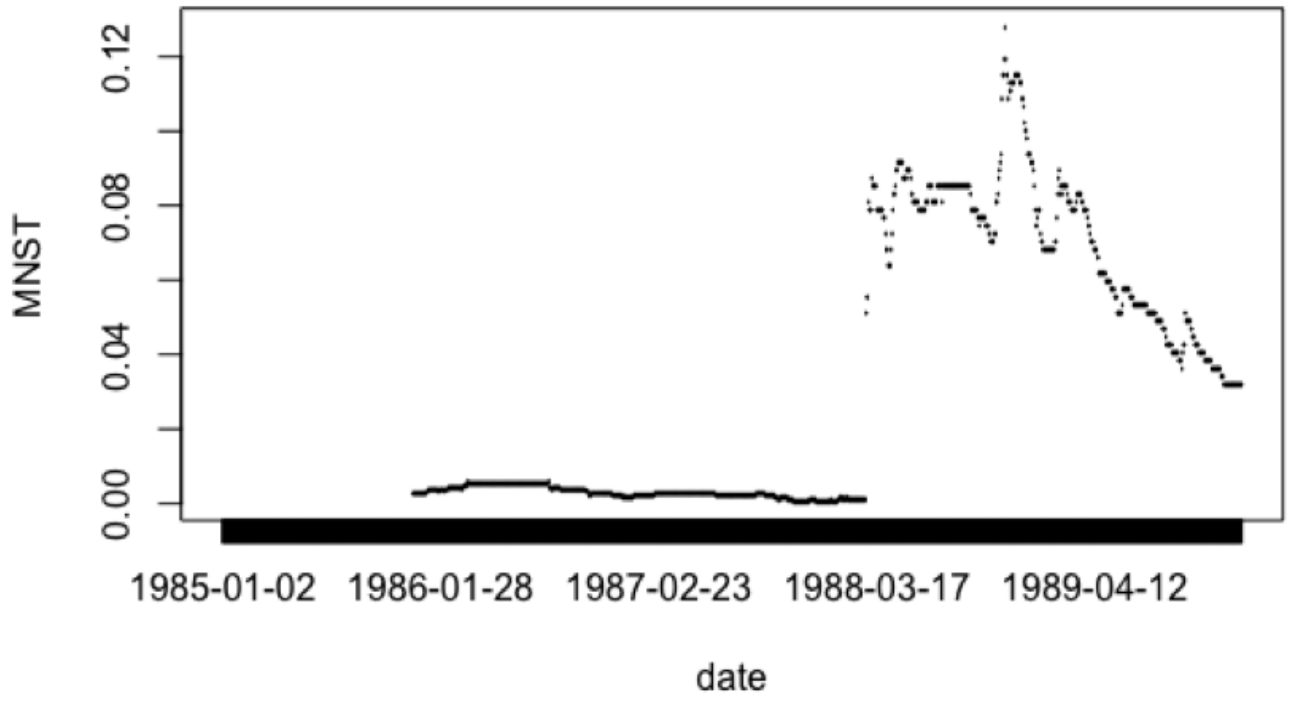


Figure 6: Monthly Return of MNST (Monster Beverage Corp)

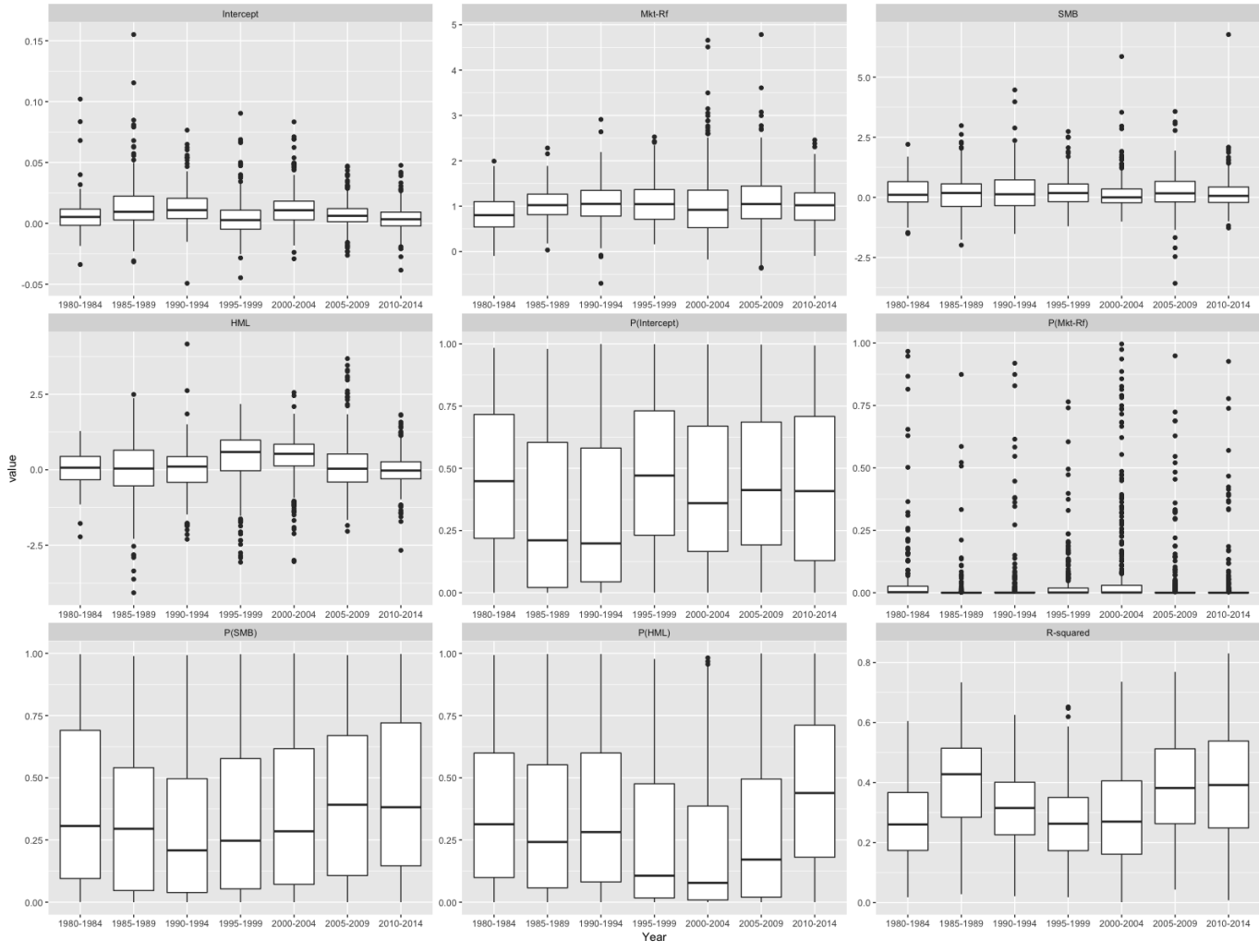


Figure 7: Trend analysis in 1980-2015 without MNST

Table 8: Top five stocks in terms of gross returns (2010 to 2017)

Ticker	Name	Sec	Return	R^2	Mkt.rf	SMB	HML
NFLX	Netflix Inc.	IT	24.13	0.04	0.96	0.41	-0.49
URI	United Rentals, Inc.	I	16.12	0.56	2.49	1.19	1.34
REGN	Regeneron	H	14.26	0.17	1.06	0.62	-1.49
STZ	Constellation Brands	CS	13.58	0.12	0.76	0.01	-0.22
AVGO	Broadcom	IT	12.52	0.23	1.04	0.05	-0.76

Table 9: Bottom five stocks in terms of gross returns (2010 to 2017)

Ticker	Name	Sec	Return	R^2	Mkt.Rf	SMB	HML
RRC	Range Resources Corp.	E	-0.67	0.16	0.68	0.62	1.03
APA	Apache Corporation	E	-0.56	0.35	1.20	0.62	0.93
MOS	The Mosaic Company	M	-0.52	0.28	1.40	-0.13	0.64
FCX	Freeport-McMoRan Inc.	M	-0.42	0.27	2.27	0.22	0.88
DVN	Devon Energy Corp.	E	-0.40	0.41	1.41	0.55	1.27

Correlation of the Factors and Top 20 and Bottom 20 returns

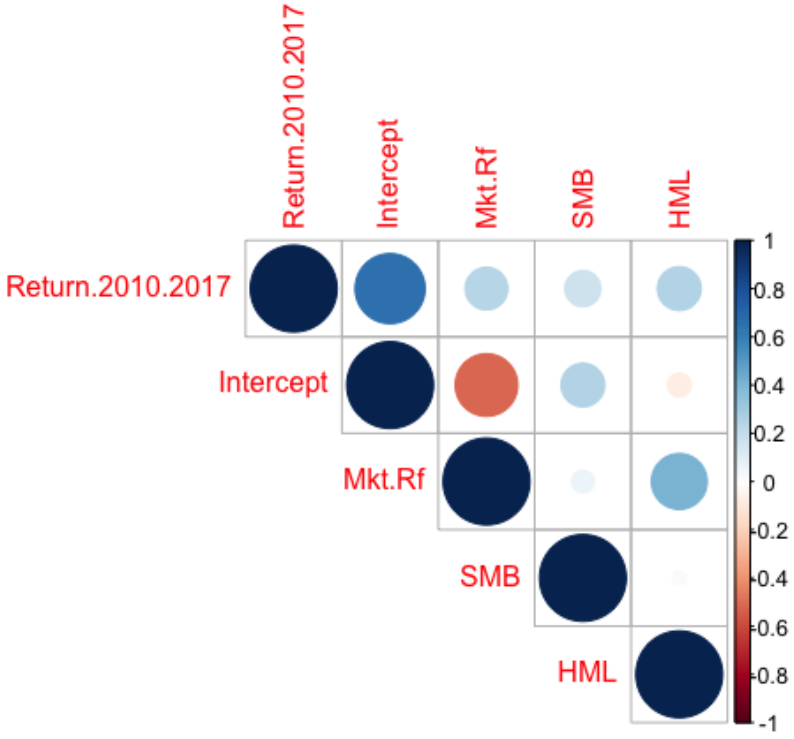


Figure 8: Top 20

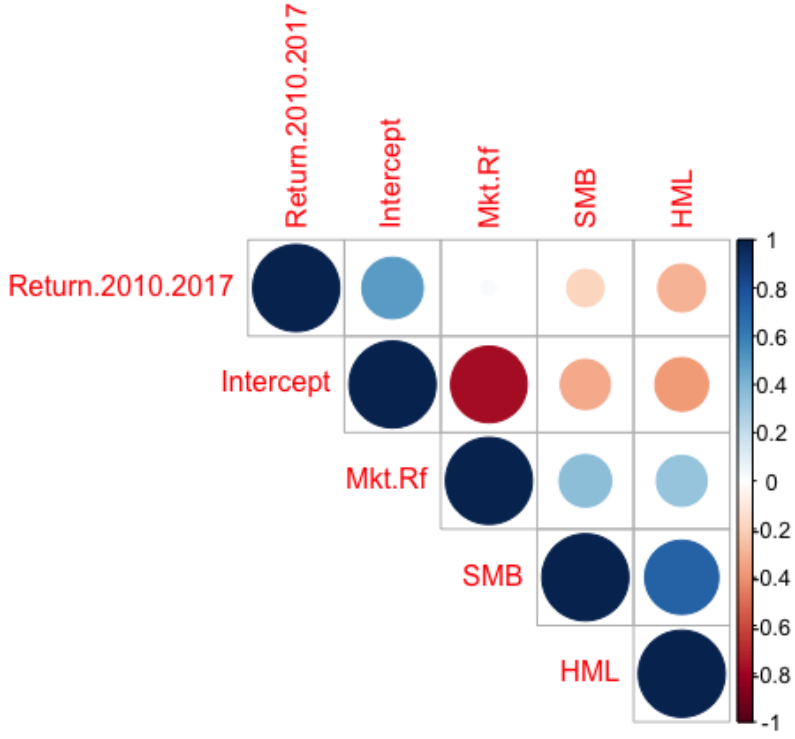


Figure 9: Bottom 20

Conclusion

1. The Fama-French Three-Factor model shows strong statistical significance in the portfolios studied by the paper, but does not explain returns of individual stocks over time as well.
2. Monthly returns of stocks are still explained by market return (CAPM) from time series perspective.
3. From 2010 to 2017, market returns are positively correlated with SMB and HML factors for the top 20 companies with the largest returns within SP 500, while the bottom 20 companies' returns show negative correlations.
4. The explanatory power of the Fama French model is more about cross-sections.

Comparison with Five-Factor Model