Determinants of Stock Returns

Yichun Ren

Prior Work

Capital Asset Pricing Model (CAPM)--William F. Sharpe 1964

- Predict the returns for assets with the risk of the assets and cost of capital.
- The goal of the CAPM formula is to evaluate whether a stock is fairly valued when its risk and the time value of money are compared to its expected return.
- Formula[1]:

$$ER_i = R_f + \beta_i(ER_m - R_f)$$

where:

ER_i = expected return of investment

R_f = risk-free rate

β_i = beta of the investment

$(ER_m - R_f)$ = market risk premium

The calculation for beta is as follows:

$$\text{Beta coefficient}(\beta) = \frac{\text{Covariance}(R_e, R_m)}{\text{Variance}(R_m)}$$

where:

R_e = the return on an individual stock

R_m = the return on the overall market

Covariance = how changes in a stock's returns are related to changes in the market's returns

Variance = how far the market's data points spread out from their average value

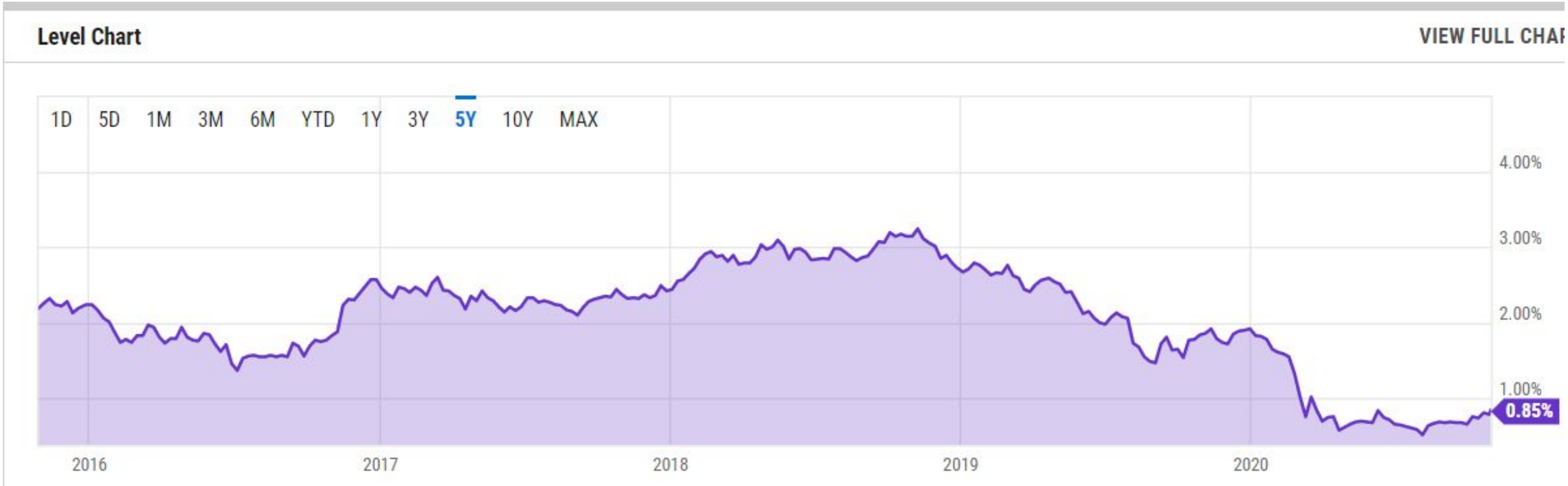
CAPM

CONS

- Risk Free Rate
- Return on the Market (Rm)
- Ability to Borrow at a Risk-Free Rate
- Determination of Project Proxy Beta

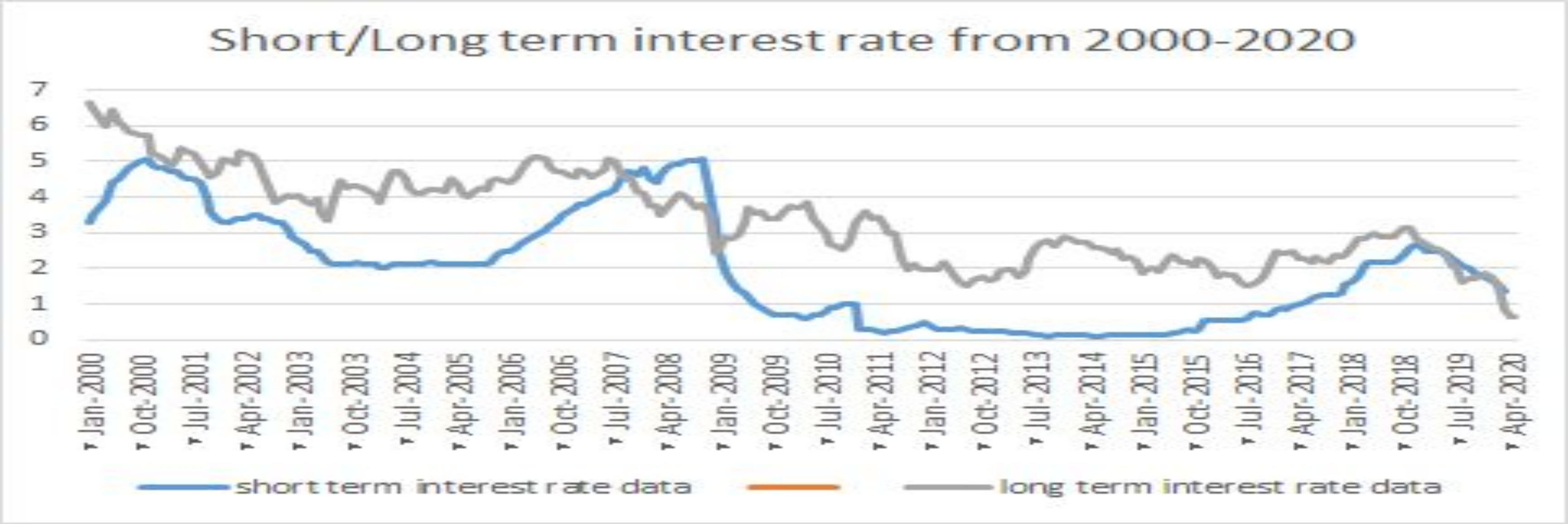
In Finance

Risk Free Rate:



In the Market

May-2019	Jun-201	Jul-201	Aug-20	Sep-20	Oct-201	Nov-20	Dec-20	Jan-202	Feb-20	Mar-20	Apr-20	May-2020
2.44	2.3	2.22	2.06	2.03	1.88	1.77	1.76	1.65	1.59	1.35		0.17
2.4	2.07	2.06	1.63	1.7	1.71	1.81	1.86	1.76	1.5	0.87	0.66	0.67



Prior Work

Fama and French Three Factor Model--Eugene Fama, Kenneth French 1993

- Adding size risk and value risk factors to the market risk factor in CAPM
- Improvement: adjusts downward for observed small-cap and value stock out-performance
- Formula[2]:

$$R_{it} - R_{ft} = \alpha_{it} + \beta_1(R_{Mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \epsilon_{it}$$

where:

R_{it} = total return of a stock or portfolio i at time t

R_{ft} = risk free rate of return at time t

R_{Mt} = total market portfolio return at time t

$R_{it} - R_{ft}$ = expected excess return

$R_{Mt} - R_{ft}$ = excess return on the market portfolio (index)

SMB_t = size premium (small minus big)

HML_t = value premium (high minus low)

$\beta_{1,2,3}$ = factor coefficients

Prior Work

Fama and French Five Factor Model----Eugene Fama, Kenneth French 2015

- New Factors: Profitability and Investment
- Outperforms at capturing the size, value, profitability, and investment patterns in average stock returns
- Formula[3]:

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it}$$

Where:

R_{it} is the return in month t of one of the portfolios

R_{Ft} is the riskfree rate

Rm-Rf is the return spread between the capitalization weighted stock market and cash.

SMB is the return spread of small minus large stocks (i.e. the size effect).

HML is the return spread of cheap minus expensive stocks (i.e. the value effect).

RMW is the return spread of the most profitable firms minus the least profitable.

CMA is the return spread of firms that invest conservatively minus aggressively (AQR, 2014).

Fama French Model

CONs

- The momentum effect
- No theoretical model
- Not clear whether HML and SMB capture risk

Can we find a better model?

Six Candidate Factors

- MKT: Market Portfolio Return
- SMB: Small Minus Big
- HML: High Minus Low
- RMW: Robust Minus Weak
- CMA: Conservative Minus Aggressive
- MOM: Monthly Momentum

Our Regression Model:

r - rf = beta0 + beta1*x1 + beta2*x2 + ... + eps

Note: n = 675, rf: risk free return, r:return of one portfolio[a]

Is there a linear relationship? Yes!

Check **Correlation**: the strength of the linear association between the excess return of the portfolio(dependent variables) and each factor(independent variables)[b]

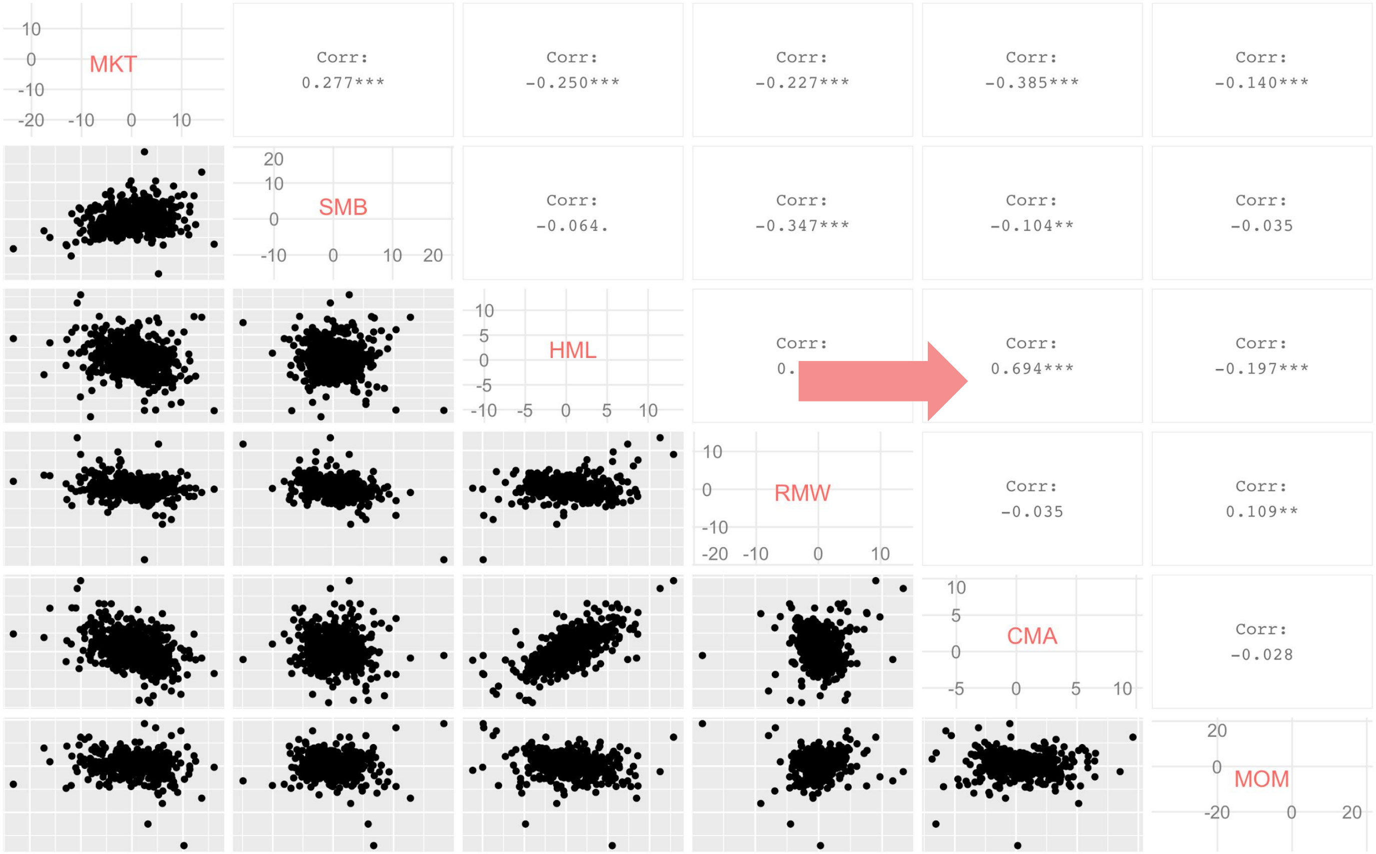
- MKT: 0.8724
- SMB: 0.6662
- HML: -0.3705
- RMW: -0.3737
- CMA: -0.4397
- MOM: -0.1013

1/2/3-25:

- corr(X,Z) = 1 mean perfect positive linear association
- corr(X,Z) = –1 means perfect negative linear association
- corr(X,Z) = 0 means no linear association

How about Multicollinearity Problem?

[c]



Multicollinearity

Two predictor variables are collinear when they are correlated. This collinearity will complicate model estimation.

Variance Inflation Factor(VIF) measures how much the variance (the square of the estimate's standard deviation) of an estimated regression coefficient is increased because of collinearity.[4][d]

5 >

```
 fac$HML  fac$CMA  fac$MKT  fac$RMW  fac$SMB  fac$MOM
2.111725 2.244296 1.331648 1.224968 1.199767 1.108873
```

One Factor

We split the regressions into different **groups**, based on **how many factors are used**.

Then by using **combinations**, we find out that there are **63** regressions in total. [e]

We prefer using Adjusted R^2

Regressor	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6
CMA					-1.48242170	
HML			-0.88634240			
MKT	1.33567250					
MOM						-0.16248730
RMW				-1.16394230		
SMB		1.48688750				
Intercept	-0.19361850	0.17779540	0.79012280	0.82002130	0.92482990	0.62180710

Summary Statistics						
SER	3.28943926	5.01945736	6.25111341	6.24261049	6.04471584	6.69561898
Adjusted_R^2	0.76076173	0.44294195	0.13602402	0.13837282	0.19213519	0.00878383

Two Factors

Regressor	NO.16	NO.17	NO.18	NO.19	NO.20	NO.21
CMA						
HML						
MKT	1.34025857	1.27130980	1.14076340			
MOM	0.03430762			-0.09837342	-0.12540570	
RMW		-0.57587465		-1.14315405		-0.50444420
SMB			1.02650720		1.48081100	1.36139300
Intercept	-0.21852064	-0.00873585	-0.32322540	0.87900424	0.26130190	0.33831160

Summary Statistics						
SER	3.28879641	3.06038168	1.40609855	6.23375441	4.99556759	4.91792056
Adjusted_R^2	0.76085523	0.79292003	0.95628621	0.14081578	0.44823190	0.46525106

Two Factors

Regressor	NO.7	NO.8	NO.9	NO.10	NO.11	NO.12	NO.13	NO.14	NO.15
CMA	-1.18719150	-0.41050380	-1.49312560	-1.52838700	-1.26250500				
HML	-0.30183520					-0.38797924	-0.97197540	-0.83362270	-0.78738420
MKT		1.26387880				1.27347112			
MOM			-0.18234090				-0.29113730		
RMW				-1.21335700				-1.09595060	
SMB					1.39996370				1.43976450
Intercept	0.93684500	-0.04213020	1.04719930	1.25508300	0.54622620	-0.04034388	1.00732770	1.06061260	0.43254890

Summary Statistics									
SER	6.01818000	3.20370881	6.00063521	5.45133468	4.35202427	3.11736005	6.13997674	5.79138003	4.50968016
Adjusted_R^2	0.19921255	0.77306944	0.20387481	0.34295880	0.58123582	0.78513739	0.16647167	0.25843190	0.55034607

Three Factors

Regressor	NO.22	NO.23	NO.24	NO.25	NO.26	NO.27	NO.28
CMA	-0.04859790	-1.08444650	-1.38323190	-0.94410890	-0.40684023	-0.50969370	-0.41526800
HML	-0.36599947	-0.42083900	-0.14775380	-0.32512370			
MKT	1.26849558				1.26694346	1.17536670	1.06798870
MOM		-0.23260880			0.01813287		
RMW			-1.19661280			-0.63673390	
SMB				1.40242750			1.02728570
Intercept	-0.03109309	1.09768640	1.25640730	0.55850200	-0.05664389	0.19889520	-0.17007720

Summary Statistics							
SER	3.11896977	5.94783878	5.44727282	4.30502758	3.20521515	2.91753862	1.18027845
Adjusted_R^2	0.78491544	0.21782255	0.34393757	0.59023129	0.77285599	0.81179977	0.96919964

Three Factors

Regressor	NO.29	NO.30	NO.31	NO.32	NO.33	NO.34
CMA	-1.53428520	-1.27219470	-1.30936110			
HML				-0.39873557	-0.38531090	-0.39409670
MKT				1.26785675	1.20987230	1.07718790
MOM	-0.11620350	-0.14453830		-0.02909976		
RMW	-1.18899150		-0.60223370		-0.57286710	
SMB		1.39229310	1.24691530			1.02857880
Intercept	1.32643570	0.64530060	0.75153310	-0.01497251	0.14251910	-0.16779550

Summary Statistics						
SER	5.43378626	4.31283314	4.18129502	3.11747155	2.87663774	0.90787923
Adjusted_R^2	0.34718216	0.58874402	0.61344746	0.78512202	0.81703953	0.98177603

Three Factors

Regressor	NO.35	NO.36	NO.37	NO.38	NO.39	NO.40	NO.41
CMA							
HML	-0.90185510	-0.85923790	-0.77311800				
MKT				1.27803901	1.14498762	1.12947980	
MOM	-0.22334900	-0.24056130		0.05798228	0.03133577		-0.10130600
RMW	-1.04318740		-0.45823230	-0.58501474		-0.16146380	-0.48280650
SMB		1.42380800	1.32662030		1.02632076	0.99089230	1.36186720
Intercept	1.21422120	0.61598420	0.57374450	-0.04788774	-0.34594683	-0.26689120	0.39888510

Summary Statistics							
SER	5.72338171	4.40323395	4.41636740	3.05321265	1.40109732	1.36929227	4.90337075
Adjusted_R^2	0.27574361	0.57132275	0.56876172	0.79388907	0.95659662	0.95854478	0.46841052

Four Factors

Regressor	NO.42	NO.43	NO.44	NO.45	NO.46	NO.47
CMA	-0.04040124	-0.22275640	-0.04628446	-1.31268660	-0.85952060	-1.06246690
HML	-0.37995263	-0.28443550	-0.37316308	-0.22705890	-0.42486680	-0.24946130
MKT	1.26398677	1.18402590	1.07245368			
MOM	-0.02771911			-0.14509330	-0.19527300	
RMW		-0.60025240		-1.15720260		-0.56888330
SMB			1.02855549		1.39282020	1.25728120
Intercept	-0.00848574	0.19366320	-0.15898222	1.34620980	0.69611870	0.74958270

Summary Statistics						
SER	3.11930913	2.86307932	0.90633213	5.41989150	4.23433449	4.15419799
Adjusted_R^2	0.78486863	0.81876016	0.98183808	0.35051654	0.60357847	0.61844136

Four Factors

Regressor	NO.48	NO.49	NO.50	NO.51	NO.52	NO.53	NO.54	NO.55	NO.56
CMA	-0.50240117	-0.41224834	-0.44934270	-1.31526160					
HML					-0.38680965	-0.40640667	-0.39315860	-0.39315860	
MKT	1.18141919	1.07053233	1.04663020		1.20916374	1.07071327	1.06635840		1.13416462
MOM	0.04032331	0.01494349		-0.11634340	-0.00404650	-0.03329877		-0.21758540	0.03825058
RMW	-0.64221955		-0.22018125	-0.57782490	-0.57221749		-0.15713100	-0.40778010	-0.16821108
SMB		1.02719116	0.97878315	1.24694410		1.02884156	0.99391470	1.32464520	0.98917649
Intercept	0.16869663	-0.18202632	-0.08069021	0.82296000	0.14583978	-0.13879571	-0.11334300	0.72411410	-0.29227246

Summary Statistics									
SER	2.91494844	1.17953331	1.09698052	4.15609183	2.87873764	0.89855158	0.85193868	4.32917184	1.36110111
Adjusted_R^2	0.81213379	0.96923851	0.97339368	0.61809338	0.81677232	0.98214857	0.98395264	0.58562212	0.95903927

Five Factors

Regressor	NO.57	NO.58	NO.59	NO.60	NO.61	NO.62
CMA	-0.22449434	-0.03680929	-0.09572733	-0.98413020	-0.44527111	
HML	-0.28183487	-0.38929345	-0.34978540	-0.33677010		-0.40298047
MKT	1.18468166	1.06719281	1.05567180		1.05017648	1.06153698
MOM	0.00489647	-0.03204075		-0.15919370	0.02282524	-0.02649109
RMW	-0.60125204		-0.17011804	-0.52381210	-0.22367548	-0.15234980
SMB		1.02881313	0.99100154	1.26094860	0.97786899	0.99517851
Intercept	0.19004398	-0.13288223	-0.09061446	0.84663410	-0.09752315	-0.09192889

Summary Statistics						
SER	2.86515145	0.89781432	0.84277629	4.10744834	1.09373063	0.84588221
Adjusted_R^2	0.81849772	0.98217786	0.98429596	0.62698085	0.97355110	0.98417999

Six Factors

Regressor	NO.63
CMA	1.05242419
HML	0.9923493
MKT	-0.36206092
MOM	-0.16484851
RMW	-0.08741040
SMB	-0.02294519
Intercept	-0.07404134

Summary Statistics	
SER	0.838403122
Adjusted_R^2	0.984458509

Sort by Adjusted R^2 [g]

TOP 10
Best Performence

49	CMA	MKT	MOM	SMB			4	0.969238512
50	CMA	MKT	RMW	SMB			4	0.973393684
61	CMA	MKT	MOM	RMW	SMB		5	0.973551097
34	HML	MKT	SMB				3	0.981776026
44	CMA	HML	MKT	SMB			4	0.981838084
53	HML	MKT	MOM	SMB			4	0.982148573
58	CMA	HML	MKT	MOM	SMB		5	0.982177855
54	HML	MKT	RMW	SMB			4	0.98395264
62	HML	MKT	MOM	RMW	SMB		5	0.984179992
59	CMA	HML	MKT	RMW	SMB		5	0.984295955
63	MKT	SMB	HML	RMW	CMA	MOM	6	0.984458509

Models mentioned in assigment:

CAPM

three-factor

five-factor

No. of Regressor	Factors Used 1	2	3	4	5	6	Number of factors	Adjusted_R^2
6	MOM						1	0.008783834
3	HML						1	0.136024019
4	RMW						1	0.138372823
19	MOM	RMW					2	0.140815784
13	HML	MOM					2	0.166471671
5	CMA						1	0.192135188
7	CMA	HML					2	0.199212547
9	CMA	MOM					2	0.203874809
23	CMA	HML	MOM				3	0.217822552
14	HML	RMW					2	0.258431896
35	HML	MOM	RMW				3	0.27574361
10	CMA	RMW					2	0.342958796
24	CMA	HML	RMW				3	0.343937571
29	CMA	MOM	RMW				3	0.347182157
45	CMA	HML	MOM	RMW			4	0.350516537
2	SMB						1	0.442941954
20	MOM	SMB					2	0.448231896
21	RMW	SMB					2	0.465251061
41	MOM	RMW	SMB				3	0.468410521
15	HML	SMB					2	0.550346068
37	HML	RMW	SMB				3	0.568761721
36	HML	MOM	SMB				3	0.571322751
11	CMA	SMB					2	0.581235816
55	HML	MOM	RMW	SMB			4	0.585622122
30	CMA	MOM	SMB				3	0.588744021
25	CMA	HML	SMB				3	0.590231294
46	CMA	HML	MOM	SMB			4	0.603578467
31	CMA	RMW	SMB				3	0.613447457
51	CMA	MOM	RMW	SMB			4	0.618093383
47	CMA	HML	RMW	SMB			4	0.618441356
60	CMA	HML	MOM	RMW	SMB		5	0.626980846
1	MKT						1	0.760761729
16	MKT	MOM					2	0.760855228
26	CMA	MKT	MOM				3	0.772855991
8	CMA	MKT					2	0.773069441
42	CMA	HML	MKT	MOM			4	0.784868628
22	CMA	HML	MKT				3	0.784915436
32	HML	MKT	MOM				3	0.78512202
12	HML	MKT					2	0.785137392
17	MKT	RMW					2	0.792920027
38	MKT	MOM	RMW				3	0.793889071
27	CMA	MKT	RMW				3	0.811799774
48	CMA	MKT	MOM	RMW			4	0.812133793
52	HML	MKT	MOM	RMW			4	0.816772319
33	HML	MKT	RMW				3	0.817039533
57	CMA	HML	MKT	MOM	RMW		5	0.818497722
43	CMA	HML	MKT	RMW			4	0.818760159
18	MKT	SMB					2	0.956286213
39	MKT	MOM	SMB				3	0.956596623
40	MKT	RMW	SMB				3	0.95854478
56	MKT	MOM	RMW	SMB			4	0.95903927
28	CMA	MKT	SMB				3	0.969199635
49	CMA	MKT	MOM	SMB			4	0.969238512
50	CMA	MKT	RMW	SMB			4	0.973393684
61	CMA	MKT	MOM	RMW	SMB		5	0.973551097
34	HML	MKT	SMB				3	0.981776026
44	CMA	HML	MKT	SMB			4	0.981838084
53	HML	MKT	MOM	SMB			4	0.982148573
58	CMA	HML	MKT	MOM	SMB		5	0.982177855
54	HML	MKT	RMW	SMB			4	0.98395264
62	HML	MKT	MOM	RMW	SMB		5	0.984179992
59	CMA	HML	MKT	RMW	SMB		5	0.984295955
63	MKT	SMB	HML	RMW	CMA	MOM	6	0.984458509

Adjusted R^2

Low

High

Our Final Model

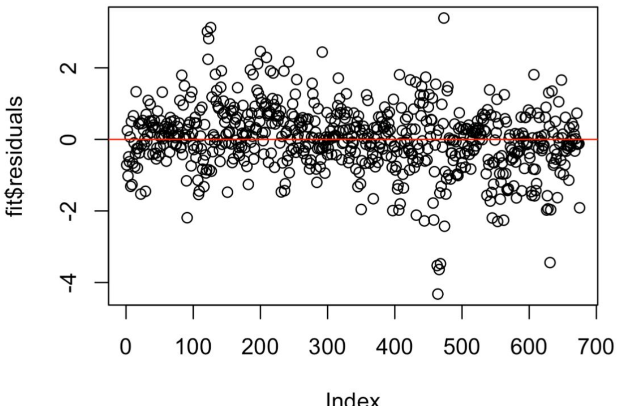
R - Rf = beta0 + beta1*MKT + beta2*SMB + beta3*HML + e

which is identical to **Fama and French Three Factor Model**

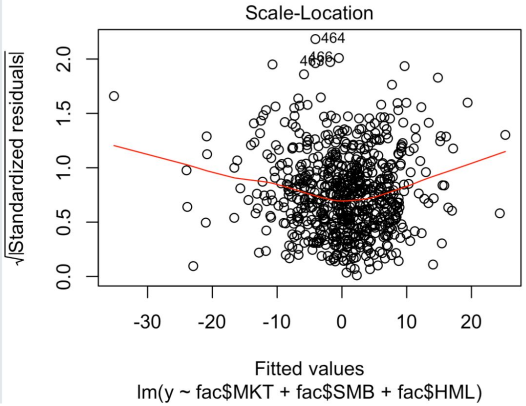
Linear Regression Assumptions

- **Independence:** the errors (residuals) are independent
- **Equal Variance (Homoscedasticity):** the error variance is the same at any set of predictor values
- **Normality:** the distribution of the errors should follow a normal distribution
- **Linearity:** the response can be written as a linear combination of the predictors

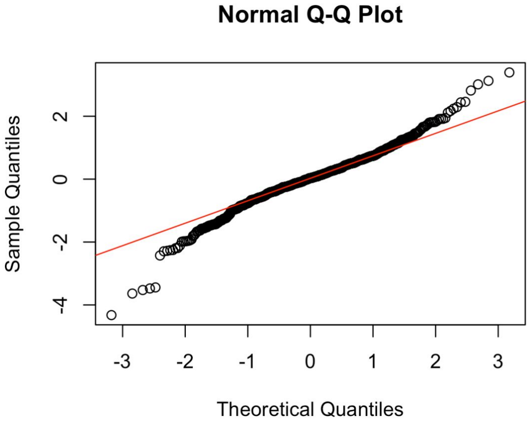
Diagnostic Plots



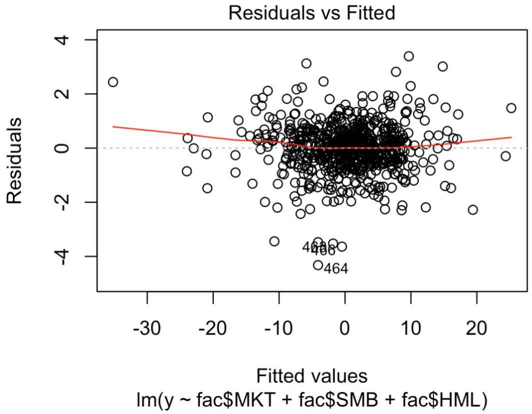
Independence



Equal Variance



Normality



Linearity

How to further improve performance?

Points with **high residual** and **high leverage** are **influential points** that need to be removed. They will skew the model fit away from the rest of the data (i.e. outliers).

Identify influential points by calculating **Cook's Distance**[5]

$$D_i = \frac{\sum_{j=1}^{n}(\hat{Y}_j - \hat{Y}_{j(i)})^2}{(p+1)\,\hat{\sigma}^2}$$

yj — the *j*th fitted response value.

yj(*i*) — the *j*th fitted response value, where the fit does not include observation *i*

p — the number of regression coefficients

σ — the estimated variance from the fit, based on all observations, i.e. Mean Squared Error

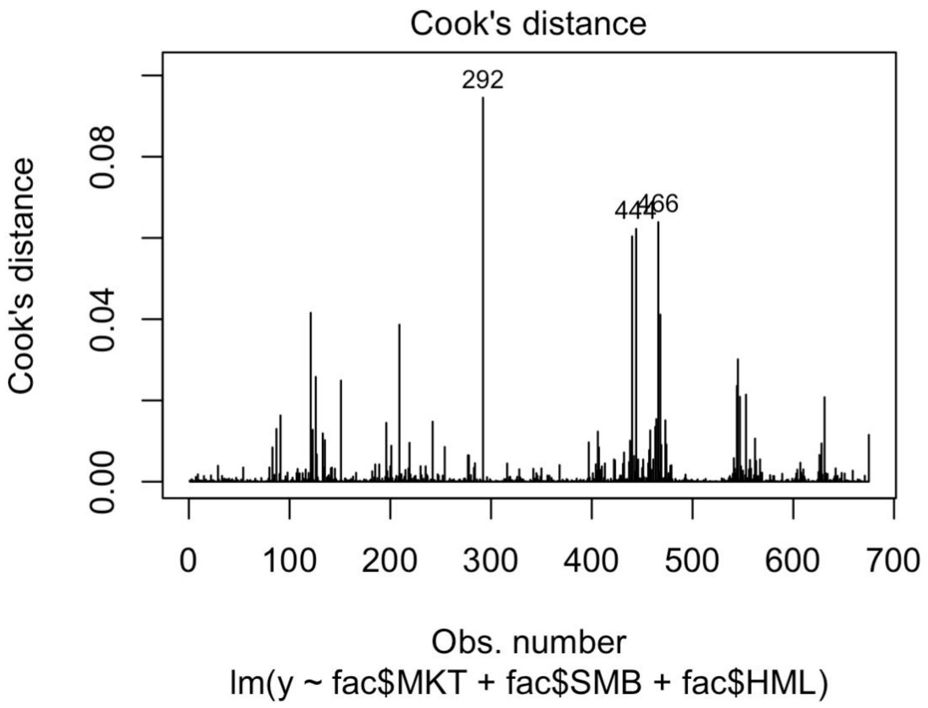
Find Influencial Points

A general rule of thumb:

If Cook's Distance of a point **> 4/n**,

the point is a possible outlier.

We find 45 outliers



Sensitive Analysis--Drop outliers

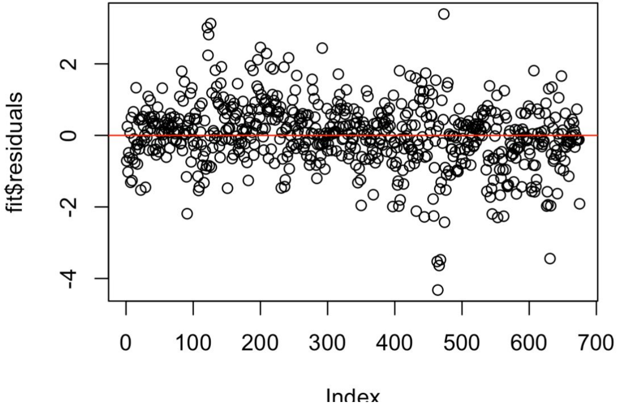
Drop the influential points and perform the same linear regression

$$R - Rf = beta0 + beta1*MKT + beta2*SMB + beta3*HML + e$$

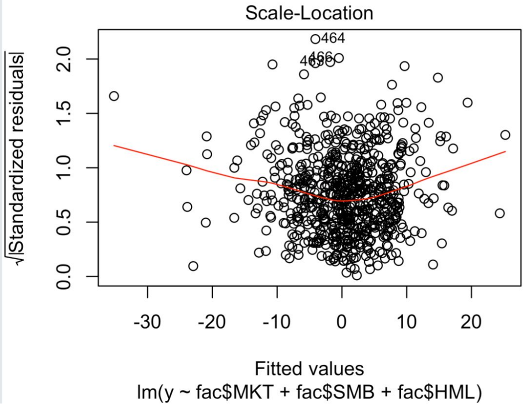
Check if there are any differences :

	Previous Model (n = 675)	Previous P-Value	New Model (n=630)	New P-Value
Intercept	-0.16780	2.95e-06	-0.146435	7.91e-07
MKT	1.07719	< 2.2e-16	1.068843	< 2.2e-16
SMB	1.02858	< 2.2e-16	1.043386	< 2.2e-16
HML	-0.39410	< 2.2e-16	-0.384322	< 2.2e-16
SER	0.9079		0.7172	
Adj R^2	0.9818		0.9859	

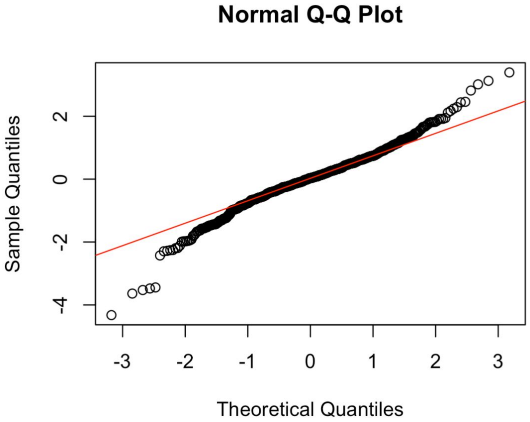
Previous Diagnostic Plots [i]



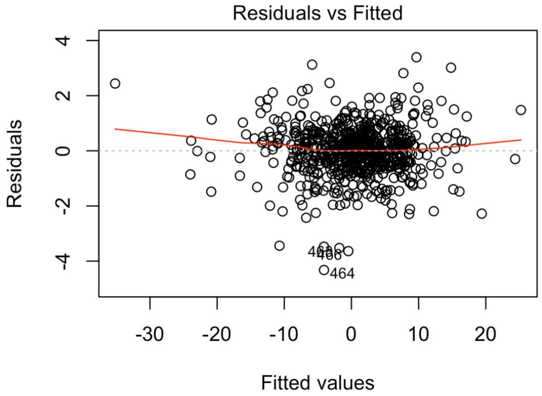
Independence



Equal Variance

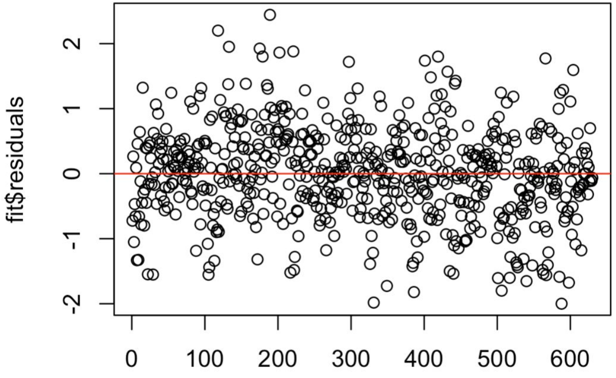


Normality

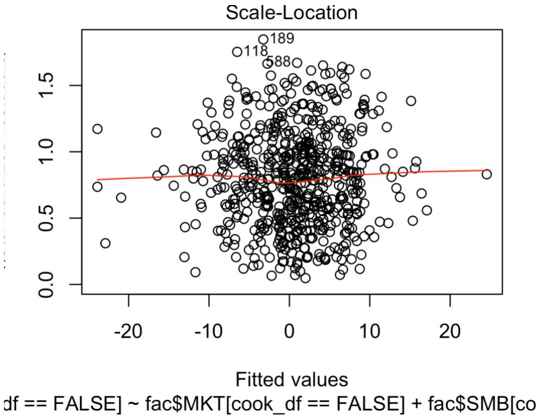


Linearity

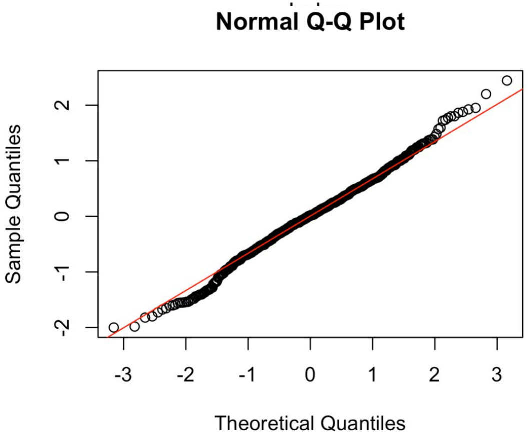
New Diagnostic Plots [j]



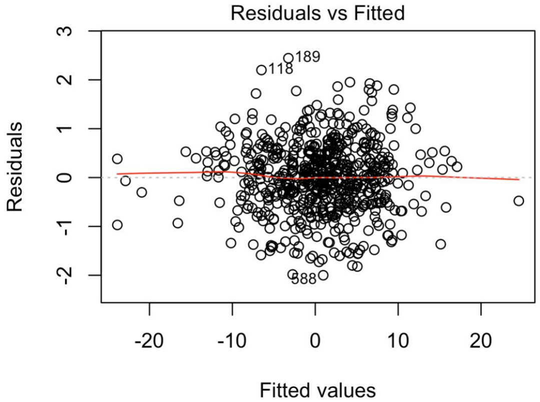
Independence



Equal Variance



Normality



Linearity

Our Final Model

R - Rf = beta0 + beta1*MKT + beta2*SMB + beta3*HML + e

R - Rf = -0.146435 + 1.068843*MKT + 1.043386*SMB + -0.384322*HML + e

Confidential Interval

****Cofidential Interval** proposes an range of pausible values for an unknown parameter.*

*The interval has an associated probability(**confidence level)** that the true parameter is in the proposed range.* [7][f]

	beta0	*beta1*	*beta2*	*beta3*
Confidental Interval (*confidence level 95%*)	(-0.20397511, -0.08889475)	(1.053943, 1.083743)	(1.022391, 1.064382)	(-0.4070082, -0.3616360)

**** by our final model: R - Rf = -0.146435 + 1.068843*MKT + 1.043386*SMB + -0.384322*HML + e***

Other portfolios

***Tested our model (for Portfolio 1) on other portfolios (Portfolio 2 to 6).*

Code for re-running [h]

	Portfolio2	Portfolio3	Portfolio4	Portfolio5	Portfolio6	Average	Variance
beta0	0.07626008	0.06726995	0.09852829	-0.05957827	-0.09726339	0.017043332	0.006321206
beta1(MTK)	0.95881156	0.98220165	0.98502192	0.9818439	1.07050029	0.995675864	0.001488811
beta2(SMB)	0.83053568	0.8686893	-0.14257982	-0.10549655	0.02359338	0.294948398	0.208290601
beta3(HML)	0.2245572	0.5674316	-0.28563429	0.28841785	0.73601555	0.306157582	0.122046585
Adjusted R^2	0.985894413	0.987005971	0.977642582	0.950335404	0.958042399	0.971784154	0.000222802

Our model also perform well on other portfolios :

- high values of Adjusted R^2 for all portfolios.
- The value of coefficient showed not much deviation. (especially beta0 and beta1)

Limitations

- Small dataset (675 rows)
- Huge time span (1963 - 2019)
- Overfitted
- Only 6 portfolios

Appendix

```
[a]rf = mean(data[,8])

r = data[,2:7]

e_r = r - rf

fac = data[,9:14]

factor_names <- c("MKT", "SMB", "HML", "RMW", "CMA", "MOM")

y = e_r[,1] #choose one portfolio to build the model

[b] round(cor(fac, y), 4)

[c] ggpairs(fac[,factor_names], axisLabels = "internal")

[d]reg = lm(y ~ fac$HML+fac$CMA+fac$MKT+fac$RMW+fac$SMB+fac$MOM)

library(car)

vif(reg)
```

Appendix

[e]factor_sets <- list()

adj_r2 <- list()

```
# One factor:
for (i in 1:6) {
  reg = lm(y~fac[,factor_names[i]])
  print(paste("Factor:",factor_names[i]))
  beta = reg$coefficients
  print(beta)
  print(paste("Residual standard error:",summary(reg)[[6]]))
  print(paste("Multiple R-squared:",summary(reg)[[8]]))
  print(paste("Adjusted R-squared:",summary(reg)[[9]]))
  cat("\n")
  factor_sets <- append(factor_sets, factor_names[i])
  adj_r2 <- append(adj_r2, summary(reg)[[9]])
}
```

```
#Two factors:
library('gtools')
coms_2 = combinations(6, 2, factor_names)
coms_2
for (i in 1:length(coms_2[,1])) {
  reg = lm(y~fac[,coms_2[i,][1]]+fac[,coms_2[i,][2]])
  cat("Factor:", coms_2[i,])
  cat("\n")
  beta = reg$coefficients
  print(beta)
  print(paste("Residual standard error:",summary(reg)[[6]]))
  print(paste("Multiple R-squared:",summary(reg)[[8]]))
  print(paste("Adjusted R-squared:",summary(reg)[[9]]))
  cat("\n")
  factor_sets <- append(factor_sets, list(coms_2[i,]))
  adj_r2 <- append(adj_r2, summary(reg)[[9]])
}
```

Appendix

```
#three factors:
coms_3 = combinations(6, 3, factor_names)
coms_3
for (i in 1:length(coms_3[,1])) {
  reg = lm(y~fac[,coms_3[i,][1]]+fac[,coms_3[i,][2]]+fac[,coms_3[i,][3]])
  cat("Factor:", coms_3[i,])
  cat("\n")
  beta = reg$coefficients
  print(beta)
  print(paste("Residual standard error:",summary(reg)[[6]]))
  print(paste("Multiple R-squared:",summary(reg)[[8]]))
  print(paste("Adjusted R-squared:",summary(reg)[[9]]))
  cat("\n")
  factor_sets <- append(factor_sets, list(coms_3[i,]))
  adj_r2 <- append(adj_r2, summary(reg)[[9]])
}
```

```
#four factors:
coms_4 = combinations(6, 4, factor_names)
coms_4
for (i in 1:length(coms_4[,1])) {
  reg = lm(y~fac[,coms_4[i,][1]]+fac[,coms_4[i,][2]]+fac[,coms_4[i,][3]]+fac[,coms_4[i,][4]])
  cat("Factor:", coms_4[i,])
  cat("\n")
  beta = reg$coefficients
  print(beta)
  print(paste("Residual standard error:",summary(reg)[[6]]))
  print(paste("Multiple R-squared:",summary(reg)[[8]]))
  print(paste("Adjusted R-squared:",summary(reg)[[9]]))
  cat("\n")
  factor_sets <- append(factor_sets, list(coms_4[i,]))
  adj_r2 <- append(adj_r2, summary(reg)[[9]])
}
```

```
#five factors:
coms_5 = combinations(6, 5, factor_names)
coms_5
for (i in 1:length(coms_5[,1])) {
  reg = lm(y~fac[,coms_5[i,][1]]+fac[,coms_5[i,][2]]+fac[,coms_5[i,][3]]+fac[,coms_5[i,][4]]+fac[,coms_5[i,][5]])
  cat("Factor:", coms_5[i,])
  cat("\n")
  beta = reg$coefficients
  print(beta)
  print(paste("Residual standard error:",summary(reg)[[6]]))
  print(paste("Multiple R-squared:",summary(reg)[[8]]))
  print(paste("Adjusted R-squared:",summary(reg)[[9]]))
  cat("\n")
  factor_sets <- append(factor_sets, list(coms_5[i,]))
  adj_r2 <- append(adj_r2, summary(reg)[[9]])
}
```

```
#six factors:
reg = lm(y~fac[,1]+fac[,2]+fac[,3]+fac[,4]+fac[,5]+fac[,6])
cat("Factor:", factor_names)
beta = reg$coefficients
print(beta)
print(paste("Residual standard error:",summary(reg)[[6]]))
print(paste("Multiple R-squared:",summary(reg)[[8]]))
print(paste("Adjusted R-squared:",summary(reg)[[9]]))
factor_sets <- append(factor_sets, list(factor_names))
adj_r2 <- append(adj_r2, summary(reg)[[9]])
```

Appendix

[f]

```
#CI
beta0=fit$coefficients[1]
beta1=fit$coefficients[2]
beta2=fit$coefficients[3]
beta3=fit$coefficients[4]

beta0_sd=summary(fit)[[4]][1, 2]
beta1_sd=summary(fit)[[4]][2, 2]
beta2_sd=summary(fit)[[4]][3, 2]
beta3_sd=summary(fit)[[4]][4, 2]

c(beta0 - 1.96*beta0_sd, beta0 + 1.96*beta0_sd)
c(beta1 - 1.96*beta1_sd, beta1 + 1.96*beta1_sd)
c(beta2 - 1.96*beta2_sd, beta2 + 1.96*beta2_sd)
c(beta3 - 1.96*beta3_sd, beta3 + 1.96*beta3_sd)
```

[g]

```
#Display all the combinations factors and their corresponding Adjusted_R^2
fac_df = do.call("rbind", factor_sets)
names(fac_df)[1] = "Factors"
r2_df = do.call(rbind.data.frame, adj_r2)
names(r2_df)[1] = "Adjusted_R^2"
num = c(rep(1, 6),rep(2, 15), rep(3, 20), rep(4, 15), rep(5, 6), 6)
num
comparison_df <- cbind(fac_df, r2_df, num)
comparison_df
#R will auto fill NA cells by repeating front cells
#Our true factors sets need remove duplications.
comparison_df[order(comparison_df['Adjusted_R^2']),]
```

[h]

```
#change different portfolio to build the model
y = e_r[,6]

factor_sets <- list()
adj_r2 <- list()

fit <- lm(y ~ fac$MKT+fac$SMB+fac$HML)

cook_df = cooks.distance(fit) > 4/675
cook_df[cook_df==TRUE]

fit <- lm(y[cook_df==FALSE] ~ fac$MKT[cook_df==FALSE]+fac$SMB[cook_df==FALSE]+fac$HML[cook_df==FALSE])

beta = fit$coefficients
beta
print(paste("Adjusted R-squared:",summary(fit)[[9]]))
```

Appendix

[i]

```
fit <- lm(y ~ fac$MKT+fac$SMB+fac$HML)
summary(fit)

qqnorm(fit$residuals)
qqline(fit$residuals, col="red")
plot(fit$residuals)
abline(0, 0, col="red")
```

[j]

```
cook_df = cooks.distance(fit) > 4/675
cook_df[cook_df==TRUE]

fit <- lm(y[cook_df==FALSE] ~
fac$MKT[cook_df==FALSE]+fac$SMB[cook_df==FALSE]+fac$HML[cook_df==FALSE])
summary(fit)
print(paste("Adjusted R-squared:",summary(fit)[[9]]))

qqnorm(fit$residuals)
qqline(fit$residuals, col="red")
plot(fit$residuals)
abline(0, 0, col="red")
```

References

[1] Beta. (2020). Retrieved 30 October 2020, from https://www.investopedia.com/terms/b/beta.asp

[2]Kenton, W. (2020, August 28). Capital Asset Pricing Model (CAPM). Retrieved October 30, 2020, from https://www.investopedia.com/terms/c/capm.asp

[3]Hayes, A. (2020, September 16). Fama and French Three Factor Model Definition. Retrieved October 30, 2020, from https://www.investopedia.com/terms/f/famaandfrenchthreefactormodel.asp

[4]Fama French Five Factor Asset Pricing Model. (2020, March 03). Retrieved October 30, 2020, from https://blog.quantinsti.com/fama-french-five-factor-asset-pricing-model/

[5]Wikipedia contributors. "Variance inflation factor." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 8 Jun. 2020. Web. 30 Oct. 2020.

References

[6]Nguyenova, L. (2020, February 10). A little closer to Cook's distance. Retrieved October 30, 2020, from https://medium.com/@lymielynn/a-little-closer-to-cooks-distance-e8cc923a3250

[7]Yahoo is now a part of Verizon Media. (2020). Retrieved 30 October 2020, from https://ca.finance.yahoo.com/quote/0P0000MNMB.SI/history?p=0P0000MNMB.SI

[8] Dekking, F.M. (Frederik Michel), 1946- (2005). A modern introduction to probability and statistics : understanding why and how. Springer. ISBN 1-85233-896-2. OCLC 783259968.

Thank You