MULTIVARIATE LINEAR REGRESSION: FAMA FRENCH 3 FACTOR MODEL

Azim Kachwalla,Bin Lu,Leah Chance,Siwei Chen,Tamanna Sinha

PROBLEM STATEMENT

- **Our first hypothesis** is to check if for the given data set the fama french three factor model is accurate in describing the stock returns.
- Since our data spans from 1960 to 2019 we want to check whether some of the coefficients to change in the fama french model as there is a change in the market and **our second hypothesis** checks for this assumption.

FAMA FRENCH MODEL

In asset pricing and portfolio management the Fama–French three-factor model is a model designed to describe stock returns. It expands on the capital asset pricing model by adding size risk and value risk factors to the market risk factors.
The formula for the Fama French Model is given by:

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

Where:

- R_{it} is the total return of a stock or portfolio, i at time t;
- R_{ft} is the risk free rate of return at time t;
- R_{Mt} is the total market portfolio return at time tl
- R_{it} - R_{ft} is expected excess return;
- R_{Mt} - R_{ft} is the excess return on the market portfolio (index);
- SMB_t is the size premium (small minus big); and
- HML_t is the value premium (high minus low).
- $\beta_{1,2,3}$ refer to the factor coefficients

3 FACTORS OF FAMA FRENCH

The Fama and French model has three factors: size of firms, book-to-market values and excess return on the market. In other words, the three factors used are SMB (small minus big), HML (high minus low) and the portfolio's return less the risk free rate of return. SMB accounts for publicly traded companies with small market caps that generate higher returns, while HML accounts for value stocks with high book-to-market ratios that generate higher returns in comparison to the market.
Here, book to market ratio is a financial ratio used to compare a company's current market price to its book value.

HOW IS THE MODEL USEFUL TO INVESTORS?

Fama and French highlighted that investors must be able to ride out the extra short-term volatility and periodic underperformance that could occur in a short time. Investors with a long-term time horizon of 15 years or more will be rewarded for losses suffered in the short term. Using thousands of random stock portfolios, Fama and French conducted studies to test their model and found that when size and value factors are combined with the beta factor, they could then explain as much as 95% of the return in a diversified stock portfolio.

Given the ability to explain 95% of a portfolio’s return versus the market as a whole, investors can construct a portfolio in which they receive an average expected return according to the relative risks they assume in their portfolios. The main factors driving expected returns are sensitivity to the market, sensitivity to size, and sensitivity to value stocks, as measured by the book-to-market ratio. Any additional average expected return may be attributed to unpriced or unsystematic risk.

DATA DESCRIPTION

We downloaded the data for the three factors from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

A. DESCRIPTION FOR PORTFOLIOS FORMED ON SIZE

The portfolios are constructed at the end of June. The annual returns are from January to December.
The portfolios are constructed at the end of each June using the June market equity and NYSE breakpoints.The portfolios for July of year t to June of t+1 include all NYSE, AMEX, and NASDAQ stocks for which we have market equity data for June of t.

In our project we have considered Decile portfolios. Here, decile is a quantitative method of splitting up a set of ranked data into 10 equally large subsections. For example, say it is the beginning of July 1991 and we have 1000 stocks. We are going to rank all the stocks on cash flow to price. The top 100 stocks get put in the top decile (Hi 10) and the bottom 100 stocks get put into the bottom decile (Lo 10). The rest of the stocks are ordered accordingly. (Lo 10 < Dec 2 < Dec 3 < Dec 4 < Dec 5 < Dec 6 < Dec7 < Dec 8 <Dec 9 < Hi 10)

These form the 10 portfolios we will be fitting the Fama French Model to.

B. DESCRIPTION OF THE FAMA FRENCH FACTORS

SMB (Small Minus Big) is the average return on the three small portfolios minus the average return on the three big portfolios,

SMB= 1/3 (Small Value + Small Neutral + SmallGrowth)
- 1/3 (Big Value + Big Neutral + Big Growth).

HML (High Minus Low) is the average return on the two value portfolios minus the average return on the two growth portfolios,

HML= 1/2 (Small Value + Big Value)- 1/2 (Small Growth + Big Growth).

Rm-Rf, the excess return on the market, value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t, good shares and price data at the beginning of t, and good return data for t minus the one-month Treasury bill rate.

MISSING VALUES:

We test our hypothesis with missing values also. For this we randomly replace 20% data in the portfolio and fama model files with missing value (NA), and do the multivariate regression for them again.
The results obtained are not very different from the results obtained without the missing values. This is due to the fact that our dataset was large enough to begin with, even after removing 20% of our data set we were still left with more than 500 sample points. In case our dataset was not so large, we could use interpolation to find the missing values between any 2 data points.
The results for linear regression are shown alongside the results for the first hypothesis.

HYPOTHESIS A:

For the first hypothesis we wanted to test whether the fama french model is an accurate fit for the portfolio returns. This is given by:

$$H_0 : \alpha_i = 0 \ VS \ H_1 : \alpha_i \neq 0$$

Here α is the intercept term. If we do not reject the null hypothesis, it indicates that the Fama French Model is a good fit for prediction and if we reject the null hypothesis, it means that the model is not a good fit for the data

Given below are the results for fitting the fama french model a few portfolios.
We compare the effects of Missing values on the coefficients of Fama French Model.
The table gives us the coefficients for the factors along with the intercept. The 99% confidence intervals along with the p values for the respective coefficients are provided as well.
We plot the coefficients along with their confidence interval to provide a visual representation.
OLS Multiple Linear Regression also makes the assumption that the residuals(errors) are normally distributed.
We test the normality of the errors through a qqplot.

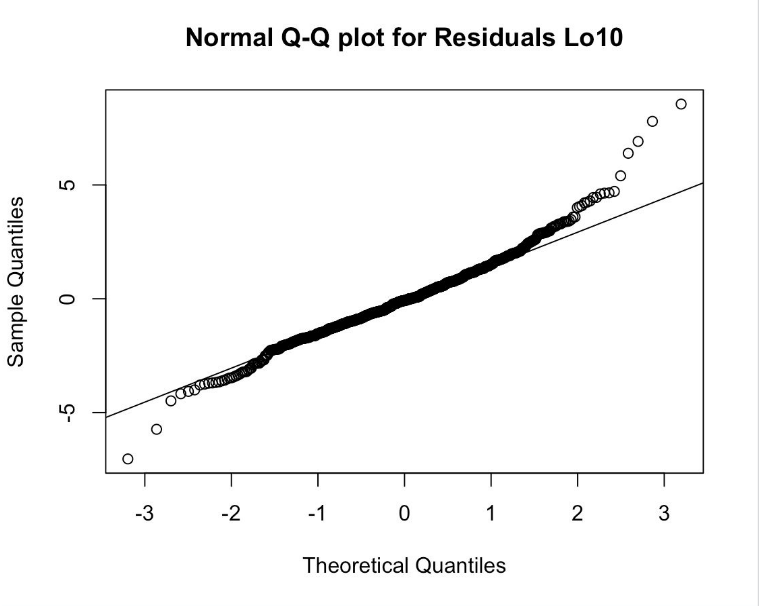
We can clearly see that $H_0 : \alpha_i = 0$ is true for nearly most of the portfolios considered.

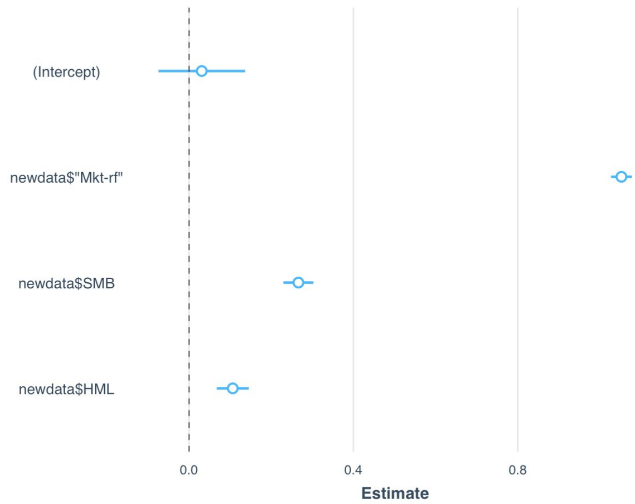
Results for Lo 10

	Lo10	Lo10 MissingValues
(Intercept)	-0.05	-0.04
	[-0.22,0.12],0.46	[-0.34,0.25],0.71
newdata$"Mkt-rf"	0.90 ***	0.89 ***
	[0.86,0.94],0.00	[0.82,0.95],0.00
newdata$SMB	1.22 ***	1.30 ***
	[1.16,1.28],0.00	[1.21,1.40],0.00
newdata$HML	0.27 ***	0.22 ***
	[0.20,0.33],0.00	[0.12,0.32],0.00
N	717	245
R2	0.92	0.94

*** p < 0.001; ** p < 0.01; * p < 0.05.

Column names: names, Lo10, Lo10 MissingValues



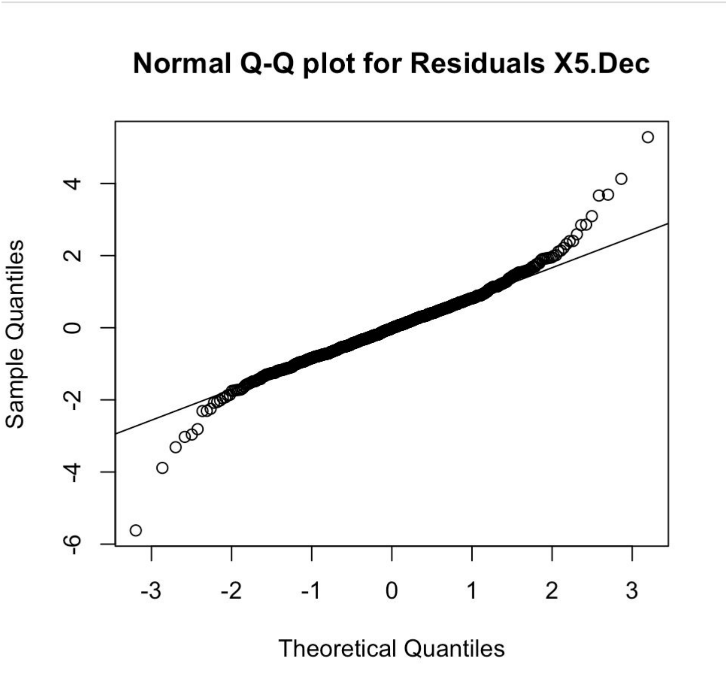


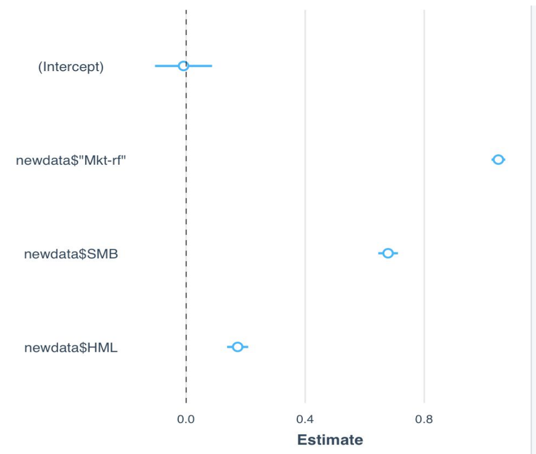
Results for X5.Dec

	X5.Dec	X5.Dec MissingValues
(Intercept)	-0.01	0.02
	[-0.10,0.09],0.81	[-0.15,0.19],0.76
newdata$"Mkt-rf"	1.05 ***	1.06 ***
	[1.03,1.07],0.00	[1.02,1.10],0.00
newdata$SMB	0.68 ***	0.65 ***
	[0.65,0.71],0.00	[0.59,0.70],0.00
newdata$HML	0.17 ***	0.19 ***
	[0.14,0.21],0.00	[0.14,0.25],0.00
N	717	244
R2	0.97	0.97

*** p < 0.001; ** p < 0.01; * p < 0.05.

Column names: names, X5.Dec, X5.Dec MissingValues



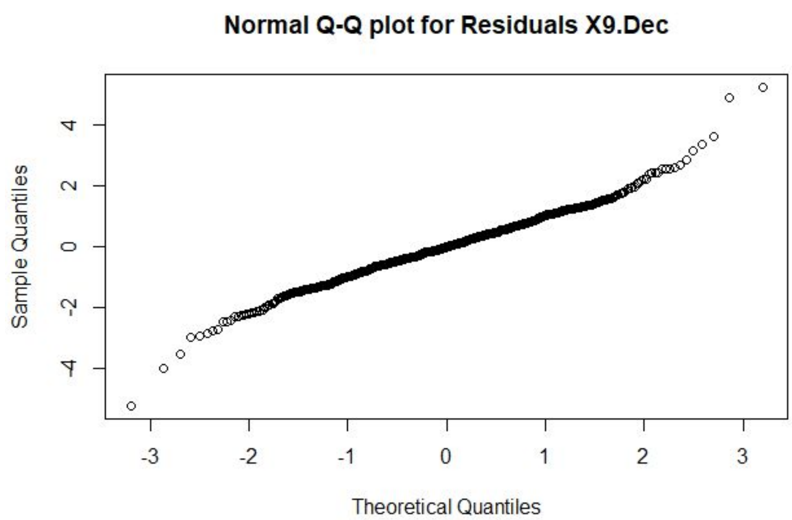


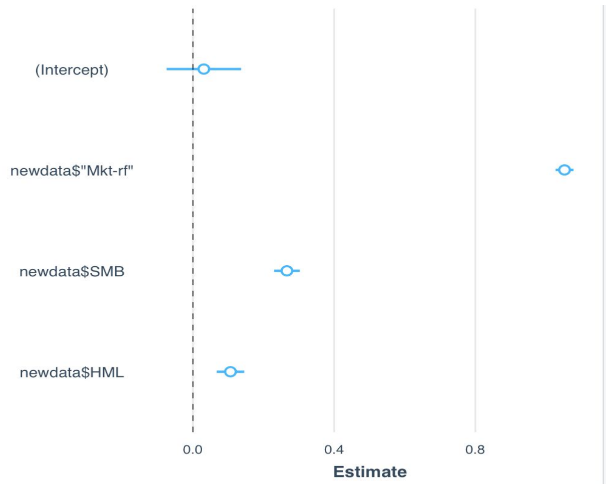
Results for X9.Dec

	X9.Dec	X9.Dec MissingValues
(Intercept)	0.03 [-0.07,0.14],0.45	0.10 [-0.07,0.26],0.12
newdata$"Mkt-rf"	1.05 *** [1.03,1.08],0.00	1.05 *** [1.01,1.09],0.00
newdata$SMB	0.27 *** [0.23,0.30],0.00	0.29 *** [0.24,0.35],0.00
newdata$HML	0.11 *** [0.07,0.15],0.00	0.12 *** [0.06,0.18],0.00
N	717	298
R2	0.95	0.96

*** p < 0.001; ** p < 0.01; * p < 0.05.

Column names: names, X9.Dec, X9.Dec MissingValues

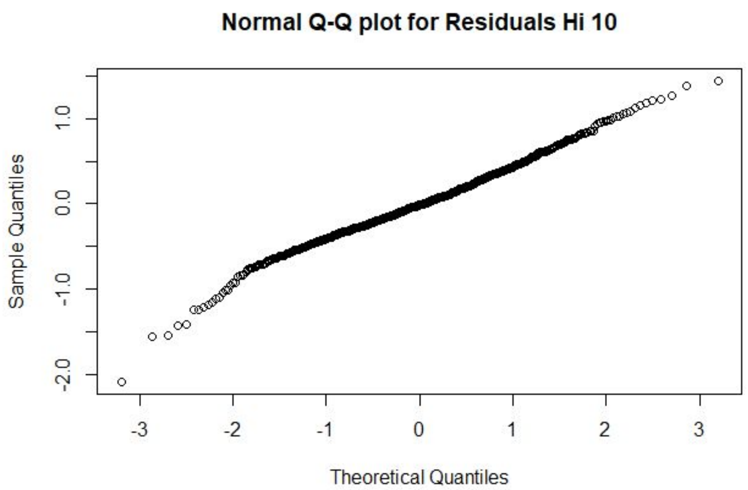


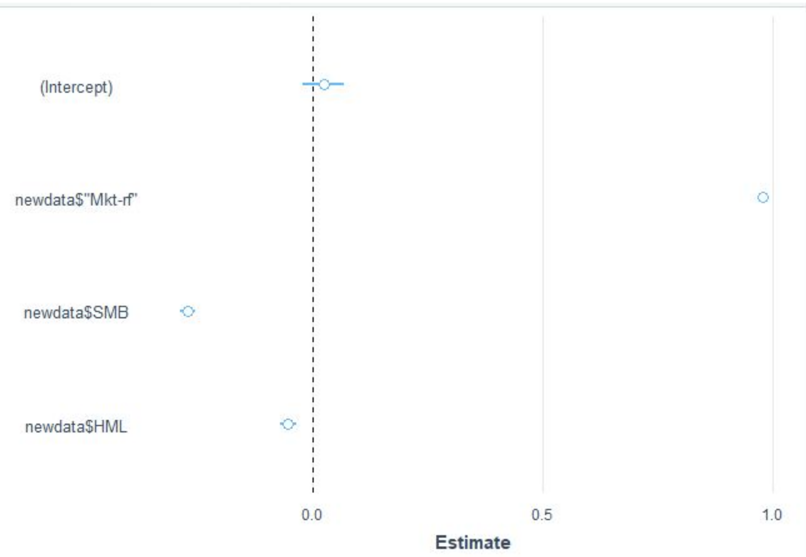


Results for Hi 10

```
                          Hi.10              Hi.10 MissingValues
                    ---------------------------------------------
  (Intercept)                0.02                   -0.03
                    [-0.02,0.07],0.19      [-0.10,0.04],0.32
  newdata$"Mkt-rf"           0.98 ***                0.99 ***
                     [0.97,0.99],0.00       [0.97,1.00],0.00
  newdata$SMB               -0.27 ***               -0.27 ***
                   [-0.29,-0.26],0.00     [-0.30,-0.25],0.00
  newdata$HML               -0.05 ***               -0.04 ***
                   [-0.07,-0.04],0.00     [-0.07,-0.01],0.00
                    ---------------------------------------------
  N                       716                    241
  R2                         0.99                   0.99
 ----------------------------------------------------------------
  *** p < 0.001; ** p < 0.01; * p < 0.05.

Column names: names, Hi.10, Hi.10 MissingValues
>
```





HYPOTHESIS B:

For the second hypothesis we want to test whether the coefficients $\beta_{i1} = \beta_{i2}$ when the data is divided into 2 time periods (t<=1990, t > 1990) for the fama french model.
The model that was fitted is:

$$y_i = \alpha_i + I_{t<=1990}\{\beta_{i1}x_{i1}\} + I_{t>1990}\{\beta_{i2}x_{i2}\} + \varepsilon_0$$

y_i = portfolio returns
$I_{t<=1990}$ = Indicator random variable that takes value 1 when t<= 1990
$I_{t>1990}$ =Indicator random variable that takes value 1 when t > 1990.
β_{i1}, β_{i2} refer to the coefficients of returns that we are checking

The hypothesis is as follows:

$$H_0 : \beta_{i1} = \beta_{i2} \;\; VS \;\; H_1 : \beta_{i1} \neq \beta_{i2}$$

If we do not reject the null hypothesis it means that there was no significant change in the market over the time period from 1960 to 2019. If we reject the null hypothesis then this means there has been a change in the coefficients for the period after 1990 indicating that there has been a significant change in the market

The results for fitting the fama french model described above are presented for a few portfolios.
We use linearHypothesis Function in R to test the null hypothesis

Each coefficient of the three factors has been tested. In the output of R under the column P(>F), if there is the symbol *** against the value, this indicates that there is a significant difference between the coefficients and we reject the null hypothesis, otherwise we do not reject it.

Results for Lo10

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.04266    0.06552  -0.651    0.515
xtable$MktRf1   0.90877    0.02280  39.852   < 2e-16 ***
xtable$SMB1     1.30719    0.03449  37.905   < 2e-16 ***
xtable$HML1     0.23856    0.03772   6.325 4.47e-10 ***
xtable$MktRf2   0.87235    0.02198  39.697   < 2e-16 ***
xtable$SMB2     1.14998    0.03023  38.043   < 2e-16 ***
xtable$HML2     0.27276    0.03185   8.565   < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.713 on 709 degrees of freedom
Multiple R-squared:  0.9243,    Adjusted R-squared:  0.9237
F-statistic:  1443 on 6 and 709 DF,  p-value: < 2.2e-16
```

```
Hypothesis:
xtable$SMB1 - xtable$SMB2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 +
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    710 2114.0
2    709 2079.5  1    34.484 11.757 0.0006411 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Linear hypothesis test

Hypothesis:
xtable$HML1 - xtable$HML2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 +
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    710 2080.9
2    709 2079.5  1    1.4193 0.4839 0.4869
>
```

```
Linear hypothesis test

Hypothesis:
xtable$MktRf1 - xtable$MktRf2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 +
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    710 2083.4
2    709 2079.5  1     3.921 1.3369  0.248
>
```

Results for X5.Dec

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -0.003429   0.037057  -0.093    0.926    
xtable$MktRf1  1.019289   0.012897  79.032  < 2e-16 ***
xtable$SMB1    0.678433   0.019504  34.784  < 2e-16 ***
xtable$HML1    0.138958   0.021331   6.514 1.38e-10 ***
xtable$MktRf2  1.076539   0.012428  86.619  < 2e-16 ***
xtable$SMB2    0.687163   0.017096  40.193  < 2e-16 ***
xtable$HML2    0.196373   0.018011  10.903  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9686 on 709 degrees of freedom
Multiple R-squared:  0.9695,	Adjusted R-squared:  0.9693 
F-statistic:  3758 on 6 and 709 DF,  p-value: < 2.2e-16
```

```
Linear hypothesis test

Hypothesis:
xtable$SMB1 - xtable$SMB2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 + 
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    710 665.27                           
2    709 665.17  1   0.10633 0.1133 0.7365
> 
```

```
Linear hypothesis test

Hypothesis:
xtable$HML1 - xtable$HML2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 + 
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F  Pr(>F)  
1    710 669.17                              
2    709 665.17  1    4.0003 4.2639 0.03929 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> 
```

```
Hypothesis:
xtable$MktRf1 - xtable$MktRf2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 + 
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F   Pr(>F)   
1    710 674.85                                
2    709 665.17  1    9.6883 10.327 0.001371 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> 
```

Results for X9.Dec

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)     0.023245   0.017426   1.334    0.183    
xtable$MktRf1   0.978728   0.006065 161.375  < 2e-16 ***
xtable$SMB1    -0.292029   0.009172 -31.839  < 2e-16 ***
xtable$HML1    -0.054003   0.010031  -5.384 9.93e-08 ***
xtable$MktRf2   0.979413   0.005845 167.578  < 2e-16 ***
xtable$SMB2    -0.258714   0.008040 -32.180  < 2e-16 ***
xtable$HML2    -0.050999   0.008470  -6.021 2.78e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4555 on 709 degrees of freedom
Multiple R-squared:  0.9882,	Adjusted R-squared:  0.9881 
F-statistic:  9929 on 6 and 709 DF,  p-value: < 2.2e-16
```

```
Hypothesis:
xtable$SMB1 - xtable$SMB2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 + 
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F   Pr(>F)   
1    710 148.64                                
2    709 147.09  1    1.5484 7.4634 0.006453 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

```
Hypothesis:
xtable$HML1 - xtable$HML2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 + 
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    710 147.10                           
2    709 147.09  1   0.01095 0.0528 0.8184
> |
```

```
Linear hypothesis test

Hypothesis:
xtable$MktRf1 - xtable$MktRf2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 + 
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    710 673.59                           
2    709 672.50  1    1.0922 1.1514 0.2836
```

Results for Hi.10

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      0.023245   0.017426   1.334    0.183    
xtable$MktRf1    0.978728   0.006065 161.375  < 2e-16 ***
xtable$SMB1     -0.292029   0.009172 -31.839  < 2e-16 ***
xtable$HML1     -0.054003   0.010031  -5.384 9.93e-08 ***
xtable$MktRf2    0.979413   0.005845 167.578  < 2e-16 ***
xtable$SMB2     -0.258714   0.008040 -32.180  < 2e-16 ***
xtable$HML2     -0.050999   0.008470  -6.021 2.78e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4555 on 709 degrees of freedom
Multiple R-squared:  0.9882,	Adjusted R-squared:  0.9881 
F-statistic:  9929 on 6 and 709 DF,  p-value: < 2.2e-16
```

```
Hypothesis:
xtable$SMB1 - xtable$SMB2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 + 
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F   Pr(>F)   
1    710 148.64                                
2    709 147.09  1    1.5484 7.4634 0.006453 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> |
```

```
Linear hypothesis test

Hypothesis:
xtable$MktRf1 - xtable$MktRf2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 + 
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    710 147.10                           
2    709 147.09  1 0.0013878 0.0067 0.9348
> |
```

```
Hypothesis:
xtable$HML1 - xtable$HML2 = 0

Model 1: restricted model
Model 2: (xtable$ri - xtable$RF1 - xtable$RF2) ~ xtable$MktRf1 + xtable$SMB1 + 
    xtable$HML1 + xtable$MktRf2 + xtable$SMB2 + xtable$HML2

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    710 147.10                           
2    709 147.09  1   0.01095 0.0528 0.8184
> |
```

CONCLUSION

1. HYPOTHESIS A - We see from the observations, the variable “Constant” refers to our α_i , is equal to zero for all our 10 portfolios. This implies that we **do not reject** the null hypothesis and conclude that the Fama French 3 factor model is a good fit for our data.

2. HYPOTHESIS B - We see from the observations, that for coefficients of each factor and the portfolios we get a different result.

3. MISSING VALUES - We see from the observations, that missing value doesn’t bring significant difference for our regression results. We got similar coefficient, R square and constant value for both linear regression with and without missing values.

NOTE

WE HAVE TESTED OUR RESULTS FOR ALL 10 PORTFOLIOS BUT WE HAVEN’T DISPLAYED THE RESULTS AS THERE ARE TOO MANY TO DISPLAY. BY RUNNING THE CODE YOU CAN GET RESULTS FOR EACH AND EVERY PORTFOLIO AND TEST