Multiple Regression Assignment

Fama-French Factor Model

Submitted by: Manasi Shukla

Company Choosen : IBM

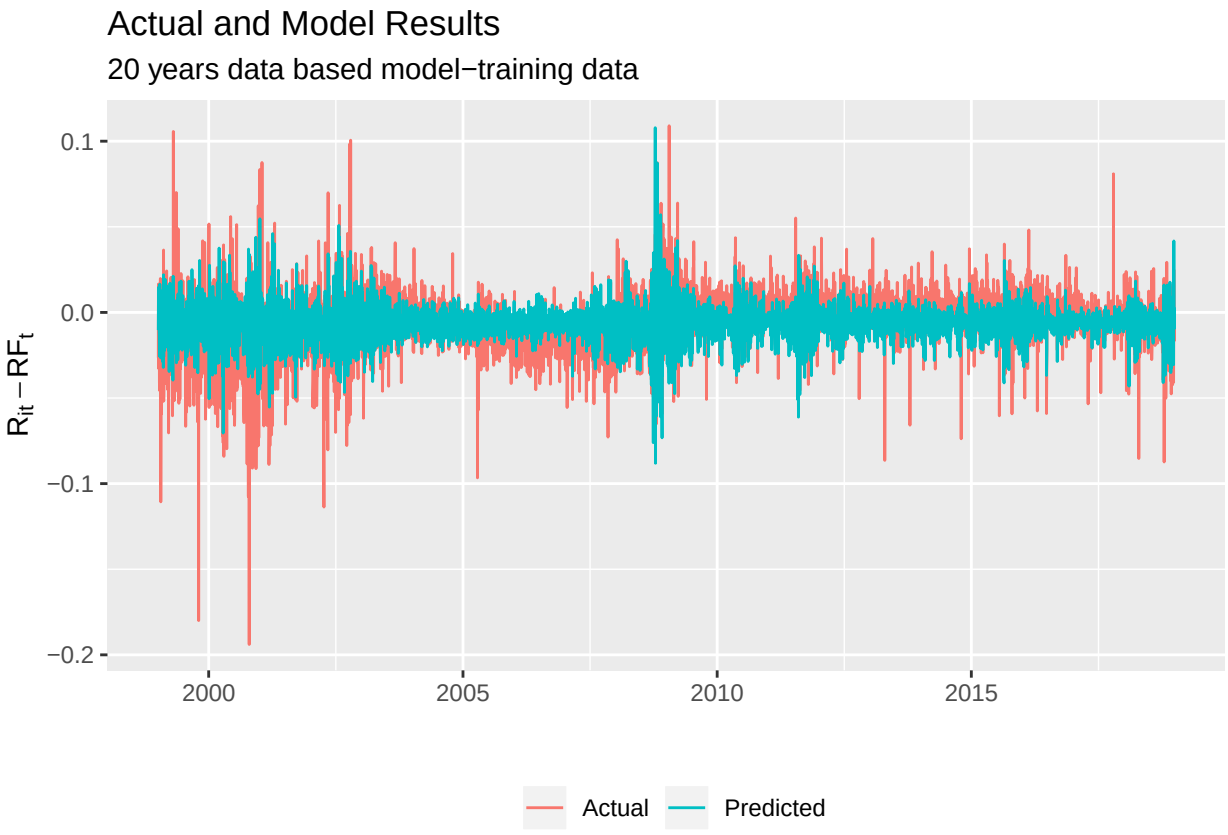
1. Model trained using 20 years of Historical Data

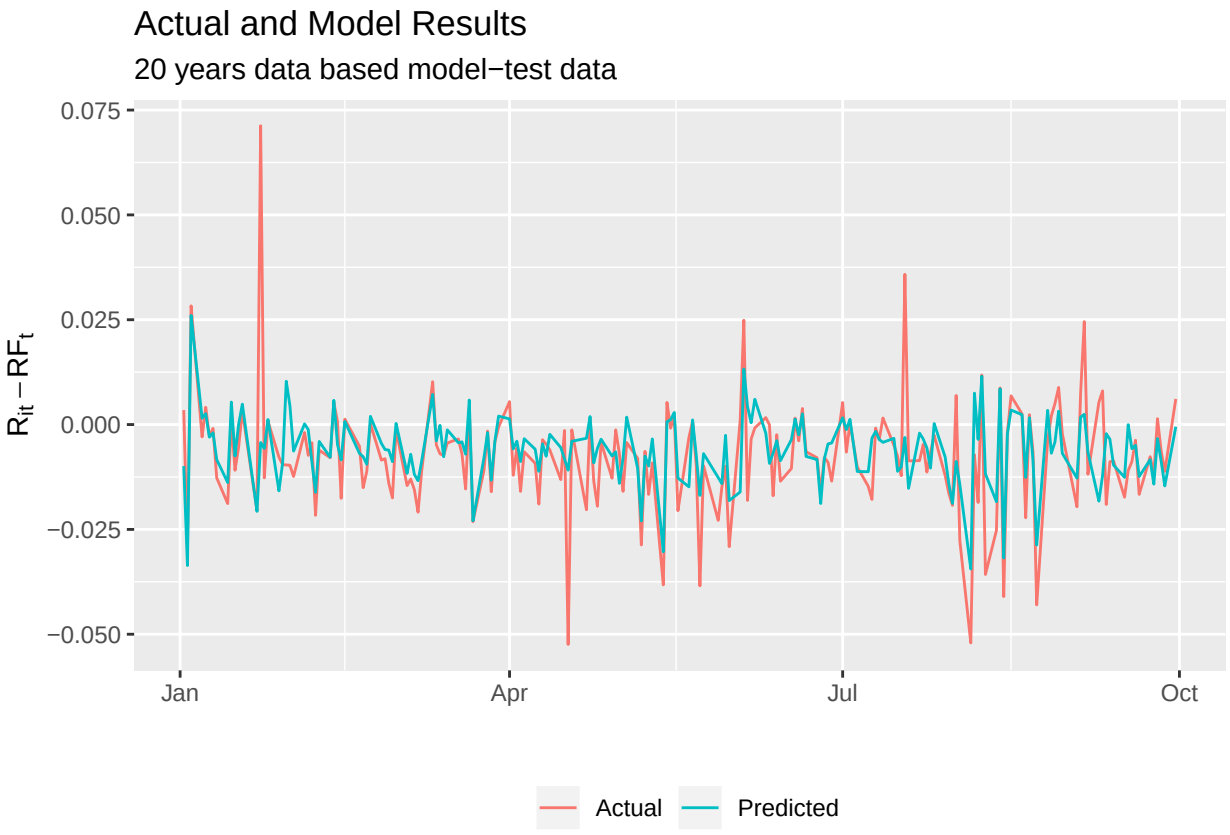
```
##
## Call:
## lm(formula = Actual ~ MktminusRF + SMB + HML, data = combined_train20)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.177918 -0.006410  0.002050  0.008184  0.096012
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0069577  0.0002123 -32.780  < 2e-16 ***
## MktminusRF   0.0089347  0.0001755  50.913  < 2e-16 ***
## SMB         -0.0022866  0.0003544  -6.452 1.21e-10 ***
## HML         -0.0029014  0.0003197  -9.077  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01504 on 5025 degrees of freedom
## Multiple R-squared:  0.3468, Adjusted R-squared:  0.3464
## F-statistic: 889.1 on 3 and 5025 DF,  p-value: < 2.2e-16
```

Observations from the above results : 1. All the p-values are statistically significant for a typical value of alpha= 1 %

2. All the t-values are outside +/- 2 range, hence statistically significant.

3. The model is able to explain only about 35 % of variance in the output.


Actual and Model Results
20 years data based model−training data
0.1
0.0
−0.1
−0.2
$R_{it} - RF_t$
2000
2005
2010
2015
Actual
Predicted



Model Performance:

```
## modelName       dataType      MAPE     MSE
## --------------  ---------  -------  ------
## 20 years data   train       2.5774   2e-04
## 20 years data   test        1.7964   1e-04
```

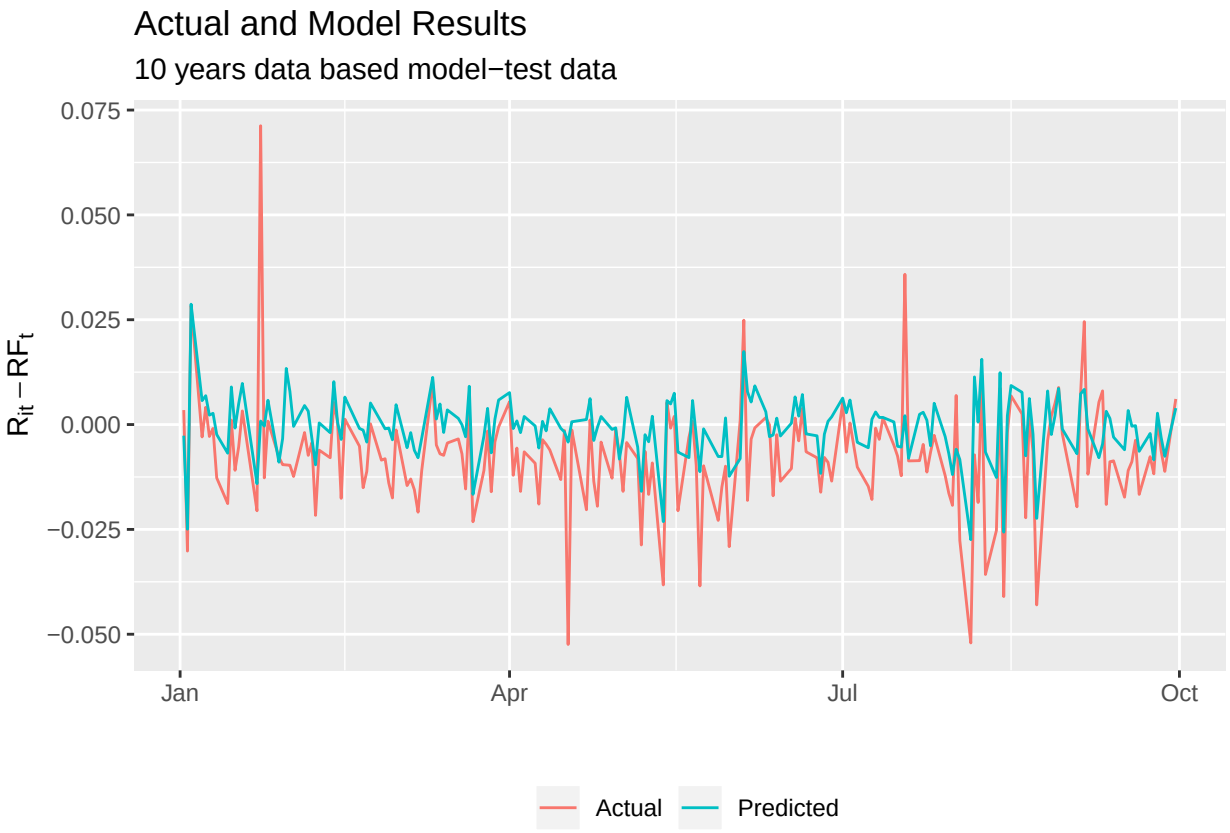
2. Model trained using 10 years of Historical Data

```
##
## Call:
## lm(formula = Actual ~ MktminusRF + SMB + HML, data = combined_train10)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.092893 -0.004274  0.000332  0.004597  0.082243
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0014768  0.0002024  -7.295 4.00e-13 ***
## MktminusRF   0.0084339  0.0002141  39.393  < 2e-16 ***
## SMB         -0.0020540  0.0003993  -5.144 2.90e-07 ***
## HML         -0.0014046  0.0003463  -4.056 5.14e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.01013 on 2510 degrees of freedom
## Multiple R-squared:  0.4098, Adjusted R-squared:  0.4091
## F-statistic:   581 on 3 and 2510 DF,  p-value: < 2.2e-16
```

Observations from the above results : 1. All the p-values are statistically significant for a typical value of alpha= 1 %

2. All the t-values are outside +/- 2 range, hence statistically significant.
3. The model is able to explain only about 41 % of variance in the output.



Model Performance:

```
## modelName       dataType      MAPE     MSE
## --------------  ---------  -------  ------
## 20 years data   train       2.5774   2e-04
## 20 years data   test        1.7964   1e-04
## 10 years data   train       1.8594   1e-04
## 10 years data   test        1.8181   1e-04
```

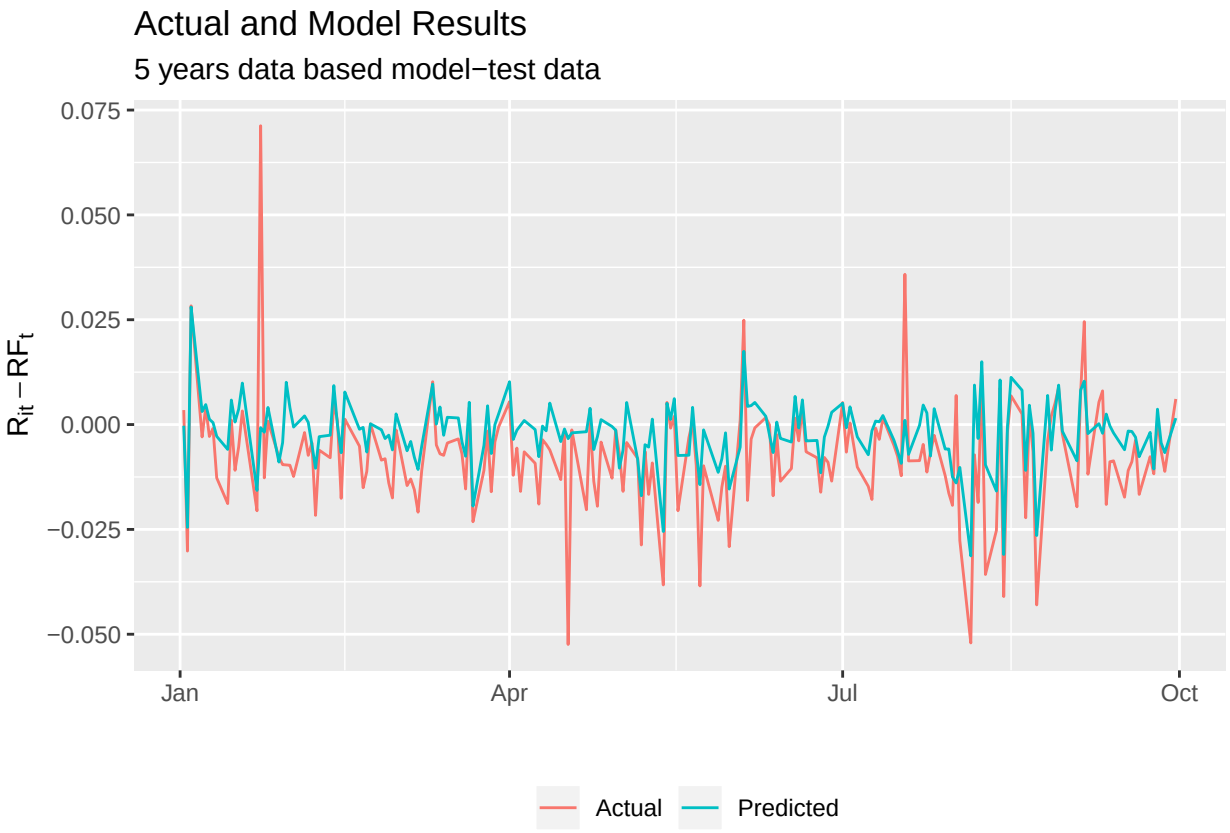
3. Model trained using 5 years of Historical Data

```
##
## Call:
## lm(formula = Actual ~ MktminusRF + SMB + HML, data = combined_train5)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.085390 -0.004136  0.000538  0.004967  0.082806
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0027824  0.0002954  -9.418  < 2e-16 ***
## MktminusRF   0.0092470  0.0003520  26.271  < 2e-16 ***
## SMB         -0.0020855  0.0005840  -3.571 0.000369 ***
## HML          0.0016446  0.0005755   2.858 0.004336 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01046 on 1252 degrees of freedom
## Multiple R-squared:  0.3554, Adjusted R-squared:  0.3538
## F-statistic: 230.1 on 3 and 1252 DF,  p-value: < 2.2e-16
```

Observations from the above results : 1. All the p-values are statistically significant for a typical value of alpha= 1 %

2. All the t-values are outside +/- 2 range, hence statistically significant.
3. The model is able to explain only about 35 % of variance in the output.



Model Performance:

modelName	dataType	MAPE	MSE
20 years data	train	2.5774	2e-04
20 years data	test	1.7964	1e-04
10 years data	train	1.8594	1e-04
10 years data	test	1.8181	1e-04
5 years data	train	1.8673	1e-04
5 years data	test	1.3039	1e-04

Recommendation: I would recommend going for 5 years based model, as it produces least error (MAPE) on the test data. My selection basis is on predictive performance as I am choosing MAPE as the Criterion. I donot belive any of these would be a good fit for explanatory model, as R squared value for all these are significantly less than 100 %.