GROUP WORK PROJECT # _1_ MScFE 652: PORTFOLIO MANAGEMENT
GROUP NUMBER: __5756__

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Giorgio Greto	UK	giorgiogreto@yahoo.com	
Raniel Barte Taripe	Philippines	nelia.taripe@yahoo.com	
Mian Muhammad Osman Bin Saif	Pakistan	osmansaif@hotmail.com	

Remember: Any group members who did **not contribute to the project should be given all zero (0) points for the collaboration grade on the GWP submission page.*

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).	
Team member 1	Giorgio Greto
Team member 2	Raniel Barte Taripe
Team member 3	Mian Muhammad Osman Bin Saif

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.

Portfolio Management, Group Work Project 1

May 3, 2024

1 Step 1

1.1 a. Define each of the 5 factors in the Fama-French 5 model

1.2 b. For each factor, explain how it helps to explain returns

2 Step 2

Download daily data from this site for a timeframe of 3 years

Below, we include all Python package that will be used.

2.1 a. Import, structure, and graph the daily factor returns

We begin by importing the `csv` file into a `pandas DataFrame`.

We set our timeframe of observation on the 3-year period that goes from March 1st, 2021 to February 29th, 2024.

The time series for each of the Fama-French (FF) factors in the timeframe considered are visualised below.

Some additional statistics regarding the time series of the 5 FF factors:

2.2 b. Collect and compute correlations of the changes in the factor returns.

As a preliminary to the analysis of their returns, we inspect the daily correlations of the five factors themselves:

From the above we can see that almost all of the factors, with the exception of the risk free rate RF, are heavily correlated, either positively or negatively.

The daily factor returns, expressed as percent change with respect to the previous day, are derived as:

Above,

- we dropped `NaN` values stemming from the first line (that lacks previous reference data) of the DataFrame, and
- we replaced $\frac{0}{0} =$ NaN, mainly appearing in the RF column when both previous and current daily entries are $= 0$, with 0.

Plotting the time series of returns yields:

At first glance, the series look fairly stationary.

One can however observe imperfections in most of the graphs, in the form of gaps within plottings (for instance, in the plot for RF at about 2022-04). These are $-\infty$ or ∞ values in the time series originated from dividing from a previous daily value of 0. We correct this by setting infinity values to an arbitrarily large number, ± 1000.

Now we can finally compute the correlation matrix between the five factor returns.

The series of first differences of factor returns are uncorrelated, with just one exception of weak correlation between RMW and CMA.

However, for subsequent steps we will rely on the FF3/5 factors themselves, not on their first differences. This is because the latter do not lend themselves to be easily interpreted.

2.3 c. Collect economic data of your choice during that 3-year period

As a proxy for risk-free rate, we download from *Yahoo! Finance* data tracking the ^IRX index, which is based on yields from the 13-week US Treasury bills.

As could be expected, the graph below shows that this interest rates index increases in value following worldwide inflation due to - disruptions in global production and supply of goods and services after the Covid pandemic, and - sanctions to Russia which increased costs for raw materials.

3 Step 3

Find the betas of factors in the Fama-French 3 model.

3.1 a. Run both Least Squares and robust regressions on the data, and describe the train-test split.

3.1.1 Least Squares regression

For convenience, as a first step we add the 13-weeks Treasury bill index ^IRX to the `pandas` DataFrame of FF3 factors for the timeframe considered, then we plot the data.

Plots above show graphically how the FF3 factors are correlated among each other, but we can anticipate the dependent variable ^IRX, the Treasury bill rates is largely uncorrelated, i.e. independent from them.

Histograms on the main diagonal of the pairplot show distributions of the variables over the selected 3-year timeframe. The distribution histogram for ^IRX shows that the dependent variable in our incoming analysis is clearly not normally distributed.

Thus from the correlation and histogram plots above, we have just learned that - the dependent and independent variables of the incoming linear regression analysis are not in a linear relationship with each other - the three factors adopted as independent variables are fairly correlated with one another - the dependent variable ^IRX is not normally distributed.

These observations lead us to anticipate that the linear regression of ^IRX from the three factors of the Fama-French model will not be successful.

We see that the ordinary least squares (OLS) linear regression of ^IRX from the FF3 factors yields high to very high p-values for the estimates of the coefficients β_i, $i = 1, 2, 3$. This is true not only inside the testing period, but for the training period as well.

This means that none of the factors can explain the variation of the dependent variable ^IRX. Another indication that the OLS regression just performed is not significant is given by the very low value of the R^2 statistic which is practically ~ 0.

Below, for this model we plot the partial regression lines of the dependent variable, ^IDX, against each of the three independent variables of the FF3 model. The lines are obtained by plotting residuals of ^IRX against residuals of each factor, after having removed the effect of the after factors.

3.1.2 Robust regression

We will now regress the ^IRX index against the FF3 factors using a robust regression model(*M-Estimation*), in order to gain a better appreciation of the influence of outlying data points over the analysis.

We will leave the train-test split of the dataset at a 75/25 ratio.

The graph shown above - comparing regressions yielded by the ordinary least squares method against the robust Huber method - reveals results with marked differences in variance inside the training sample, but almost identical in testing. Robust regression mainly dampens the deleterious effects of outliers, and this shows in the above results, which however also exclude that the low quality of the regression is due to the impact of outliers.

But it was already evident that the bad fitting of the analysis is due to the short rates tracked by ^IRX being independent from the Fama-French factors, which are tailored to track market data, not macroeconomic.

Below, for this model we plot the partial regression lines of the dependent variable, ^IRX, against each of the three independent variables of the FF3 model. The lines are obtained by plotting residuals of ^IRX against residuals of each factor, after having removed the effect of the other factors.

The first series of figures below depicts partial regressions based on the Huber norm, while the second series is based on the bisquare norm.

3.2 b. Provide summaries of coefficients and metrics for the model

We provide a summary of the analysis below.

Refer to the equation below for naming parameters in the regression:

$$\text{IRX} = \alpha + \beta_0\,(\text{Mkt-RF}) + \beta_1\,(\text{SMB}) + \beta_2\,(\text{HML})$$

Summary of coefficients, with associated p-value in subsequent row.

Model	α	β_0	β_1	β_2
OLS	1.7562	-0.0516	0.0466	-0.1565

Model	α	β_0	β_1	β_2
OLS p-value	0.000	0.481	0.708	0.048
Robust Huber	1.7554	-0.0522	0.0457	-0.1565
Huber p-value	0.000	-0.479	0.716	0.048
Robust Bisquared	1.6869	-0.0560	0.0530	-0.1612
Bisq. p-value	0.000	-0.482	0.695	0.06

We can appreciate from the table above that for all models, only a weakly significant dependence of IRX from factor HML can be theorised. All other factors are not statistically significant.

The table omits to report the standard deviation for the coefficients, in order to avoid data cluttering on parameters that in any case are not meaningful.

Summary of metrics for the OLS regression (values for the robust regressions are probably equal or very close, which would presumably be why no statistical metrics are available for them inside the `statsmodels` package).

Model	R^2	adj R^2	**JarqueBera**
OLS	0.007	0.002	74.664

R^2-based statistics close to 0 indicates that almost none of the variance in the dependent variable can be explained by the exogenous factors. The Jarque-Bera test result is far from 0, which means the dependent variable is not normally distributed.

4 Step 4

Find the beta factors in the FF5 model

4.1 a. Run both Least Squares and robust regressions on the data, and describe the train-test split.

4.1.1 Least Squares regression

Throughout this step, we follow the same procedure we have followed in step 3 in order to perform linear regressions on ^IRX. Only, this time we base the analysis on the full five factors of the Fama-French (FF5) model, instead of just three factors as before.

Factors added will be RMW - Robust Minus Weak - and CMA - Conservative Minus Aggressive.

In terms of coding, these two factors were already included inside the `df_daily` DataFrame we created, so all we have to do is to this time keep them instead of discarding them as done in the previous step, when they were not needed.

Plots above show how the FF5 factors are correlated among each other, but we can anticipate the dependent variable ^IRX, the Treasury bill rates is largely uncorrelated, i.e. independent from them.

Histograms on the main diagonal of the pairplot show distributions of the variables over the selected 3-year timeframe. The distribution histogram for ^IRX shows that the dependent variable in our incoming analysis is clearly not normally distributed.

Thus, similarly to what noticed in step 3, from the correlation and histogram plots above, we have just learned that - the dependent and independent variables of the incoming linear regression analysis are not in a linear relationship with each other - the five factors adopted as independent variables are fairly correlated with one another - the dependent variable ^IRX is not normally distributed.

These observations lead us to anticipate that the linear regression of ^IRX from the five FF5 factors will not be successful.

We keep the 0.75-0.25 split ratio between training and testing inside the data set.

In agreement with our pessimistic forecast, we observe that p-value statistics and metrics for the OLS FF5 regression model above are extremely poor.

Graphically, above, the training set shows no improvement here for FF5 model, with respect to what we found in the previous step for the FF3-based regression.

Let's produce results over the testing set.

Qualitatively, no improvement can be observed over the testing OLS performed in the previous step.

Again we can see that the ordinary least squares (OLS) linear regression of ^IRX from the FF5 factors yields high to very high p-values for the estimates of the coefficients β_i, $i = 1, \dots, 5$. This is true not only inside the testing period, but for the training period as well.

This means that none of the factors can explain the variation of the dependent variable ^IRX. Another indication that the OLS regression just performed is not significant is given by the very low value of the R^2 statistic which is practically ~ 0.

Visually, the predictions over the testing period exhibit reduced variance with respect to those in the training period.

Below, for this model we plot the partial regression lines of the dependent variable, ^IDX, against each of the five independent variables of the FF5 model. The lines are obtained by plotting residuals of ^IRX against residuals of each factor, after having removed the effect of the after factors.

We see above that regression lines are nearly horizontal, a sign that independent and dependent variables are independent and thus uncorrelated.

4.1.2 Robust regression

We will now regress the ^IRX index against the FF5 factors using a robust regression model(*M-Estimation*), as did in point b of step 3.

We will leave the train-test split of the dataset at a 75/25 ratio.

All factors are not significant, according to the p-values of their respective parameters.

We deliberate to proceed with analysis only for the Huber robust regression, as the bisquare results will largely be similar.

We see in the plot above that the robust regression is way off the mark, analogously to what it was in step 3. Now, in addition we can observe how the robust regression predictions have a much lower variance than int was the case for the OLS regression. This derives from the feature of dampening the influence of outliers exhibited by robust regression methods.

In the figure above, the marked difference in variance between training and testing predictions from the robust regression, clearly stands out.

The graph shown above - comparing regressions yielded by the ordinary least squares method against the robust Huber method - reveals results with marked differences in variance inside the training sample, but almost identical in testing. Robust regression mainly dampens the deleterious effects of outliers, and this shows in the above results, which however also exclude that the low quality of the regression is due to the impact of outliers.

But it was already evident that the bad fitting of the analysis is due to the short rates tracked by ^IRX being independent from the Fama-French factors, which are tailored to track market data, not macroeconomic.

Below, for this model we plot the partial regression lines of the dependent variable, ^IRX, against each of the five independent variables of the FF5 model. The lines are obtained by plotting residuals of ^IRX against residuals of each factor, after having removed the effect of the other factors.

The first series of figures below depicts partial regressions based on the Huber norm, while the second series is based on the bisquare norm.

4.2 b. Provide summaries of coefficients and metrics for the model

We provide a summary of the analysis below.

Refer to the equation below for naming parameters in the regression:

$$\text{IRX} = \alpha + \beta_0\,(\text{Mkt-RF}) + \beta_1\,(\text{SMB}) + \beta_2\,(\text{HML}) + \beta_3\,(\text{RMW}) + \beta_4\,(\text{CMA})$$

Summary of coefficients, with associated p-value in subsequent row.

Model	α	β_0	β_1	β_2	β_3	β_4
OLS	1.7584	-0.0624	0.0422	-0.1065	0.0178	-0.1297
OLS p-value	0.000	0.410	0.768	0.998	0.899	0.527
Robust Huber	1.7580	-0.0625	0.0419	-0.1074	0.0179	-0.1285
Huber p-value	0.000	-0.412	0.772	0.397	0.899	0.534
Robust Bisquared	1.6889	-0.0654	0.0469	-0.1103	0.0130	-0.1276
Bisq. p-value	0.000	-0.427	0.764	0.421	0.932	0.569

We can appreciate from the table above that for all models, there is no significant dependence of IRX from any of the five factors in FF5. The table omits to report the standard deviation for the coefficients, in order to avoid data cluttering on parameters that in any case are not meaningful.

Summary of metrics for the OLS regression.

Model	R^2	adj R^2	**JarqueBera**
OLS	0.008	-0.001	74.899

R^2-based statistics close to 0 indicates that almost none of the variance in the dependent variable can be explained by the exogenous factors. The Jarque-Bera test result is distant from 0, which means the dependent variable is not normally distributed.

5 Step 5

5.1 c. Correlation matrix of factor returns

Below, we repeat the computation of the correlation matrix between the 5 factors in the Fama-French model.

Now Visualizing the Correlation Matrix of Factor Returns

5.1.1 1. Heatmap

A heatmap is an effective way to visualize the correlation matrix using colors to represent the correlation coefficients. The color palette of the heatmap shows, warmer colors represent positive correlations and cooler colors represent negative correlations.

5.1.2 2. Pairplot

A pairplot is used to visualize pairwise relationships between different factors. The grid of scatterplots for each pair of factors, showing their relationships along with histograms for each individual factor.

5.1.3 3. Clustermap

The clustermap visually organizes similar factors into clusters based on their correlation coefficients.

5.2 d. Covariance Matrix of Factor Returns

5.3 e. Comparison of the Two matrices

Correlation and covariance matrix for factors in the Fama-French model during timeframe specified (March 2021 - February 2024).

Correlation Matrix: The correlation matrix measures the linear relationship between pairs of factors, normalized to a scale of -1 to 1. Values closer to 1 indicate a strong positive linear relationship, while values closer to -1 indicate a strong negative linear relationship. The diagonal elements are always 1, indicating perfect correlation of a factor with itself. Example: The correlation between Mkt-RF and SMB is 0.184681, suggesting a weak positive linear relationship.

Covariance Matrix: The covariance matrix measures the extent to which two factors move together, regardless of the scale of their values. Larger values indicate greater variability between the factors, while values closer to zero indicate less variability. The diagonal elements represent the variance of each factor. Example: The covariance between Mkt-RF and HML is -0.43434, indicating a negative covariance (opposite movement).

6 Step 6

Effects of CMA and RMW

7 Step 7

7.1 f. Markowitz portfolio optimization

7.2 g. Portfolio dependence from factors in FF3

7.3 h. Portfolio dependence from factors in FF5