**GROUP HOMEWORK 1 – Linear Regression and the CAPM**

**Assignments**

1. Download the total return index for all constituents of your sector index at the monthly frequency; select a subsect of the equities continuously available for at least 10 years, minimum 10 equities; you have the possibility of deciding the trade-off between the number of equities and the total sample conditional on satisfying the above-mentioned minimum for years and number of equities; also download the overall equity market index and an appropriate short-term interest rate.

We initiate the process by defining a monthly date-time index using the pd.date\_range function. Starting from November 1st, 2009, to October 31st, 2023. We import data from an Excel file named 'Data\_Banks\_EuroStoxx600\_Euribor.xlsx'. This includes information on the STOXX EUROPE 600 Total Return Index, 3-month EURIBOR rates, and the banks constituents. The chosen banks are extracted, and their names are cleaned for consistency.

1. Compute the returns for all your equity time series (stocks and index) and scale them to percentages (i.e., 1 should correspond to 1%); adopt the appropriate transformation for the interest rate; compute the excess returns; provide a set of scatter plots for the equities vs the market index; are the plots suggesting the possible presence of linear relationships? Are there cases where the relation is not clear?

With the help of Excel, we arrange the data in descending order to achieve the desired scale. Following this, we calculate the monthly returns for both the market (STOXX EUROPE 600) and the selected banks, converting them into percentages and adjusting for log differences.

Finally, excess returns are calculated by subtracting the risk-free rate (monthly EURIBOR rate) from the returns of both the market and the selected banks, creating erBanks and erMkt arrays for further analysis.

The ex.OLS function is used to perform Ordinary Least Squares (OLS) regression on excess market returns (erMkt) and excess bank returns (erBanks). The regression results are stored in df\_quants.

Then, we've created scatter plots comparing the market index with individual equity time series. These plots are saved as images.

* *If the scatter plots generally show a clear upward or downward trend, it may indicate a linear relationship. This could suggest that changes in the market index are associated with proportional changes in the equity.*
* *If the points in the scatter plots are scattered randomly without a clear pattern, it might suggest a lack of linear relationship. Points forming clusters or a dispersed pattern may indicate non-linearity or other complex relationships.*
* *Check for outliers or patterns that do not follow a straight line. Presence of curves, bends, or irregular shapes may suggest non-linear relationships.*
* *Compare multiple scatter plots to see if there's consistency or if certain equities behave differently from the overall trend.*
* *Since you've mentioned computing excess returns, analyze whether the excess returns exhibit linear relationships with the market index.*

1. Run the linear regression for the CAPM estimation, select the quantities of interest (DO NOT report all regression outputs in the pdf document) and report them in a table, commenting then on the results and on the fit of the model. Are there statistically significant intercepts? What about the betas? Now, for each month, compute the average return across the equities’ excess returns; this corresponds to the evaluation of the returns of an equally weighted portfolio; estimate again the CAPM and compare the possible differences in terms of alpha, beta and R-squared. What do you observe? Are you capable of providing an explanation for this result? If you believe it is useful, you might accompany your answers with equations supporting your view.

*Alpha (α) represents the excess return of an asset or portfolio beyond what would be predicted by the CAPM.*

* *Positive alpha: The asset/portfolio is outperforming the expected return given its beta. It suggests the asset is providing higher returns than the market.*
* *Negative alpha: The asset/portfolio is underperforming the expected return. It implies that the asset is not providing returns commensurate with its beta.*

*Beta (β) measures the sensitivity of an asset's returns to changes in the market returns. It indicates the asset's systematic risk.*

* *Beta = 1: The asset tends to move in line with the market.*
* *Beta > 1: The asset is more volatile than the market.*
* *Beta < 1: The asset is less volatile than the market.*
* *Negative beta: The asset tends to move inversely to the market.*

*R-squared (R²) measures the proportion of the variance in the dependent variable (asset returns) that is explained by changes in the independent variable (market returns).*

* *R-squared = 1: The model perfectly explains the variability in the asset's returns based on market movements.*
* *R-squared close to 1: A high percentage of the variability in the asset's returns is explained by the market.*
* *R-squared close to 0: The model does not effectively explain the variability, indicating that other factors may influence the asset's returns.*

1. Perform a battery of diagnostic tests (excluding the Chow test) on the linear regression reporting the test outcomes in a table, commenting on the results, establishing if the linear model is correctly specified, and if standard inference is appropriate. For the last element, when needed, proceed to the evaluation of robust standard errors and to detect if this produces changes on the significance of model parameters.
2. Now introduce additional explanatory variables to your linear model. Download at least three factors from the Kenneth French website. Read carefully the details for the factors you download and, if needed, report the transformation you might have applied to the downloaded data. Run the multifactor model and comment on the following, you are free to support your comments with proper plots: are the betas of the Market changed in size and significance? What about the alphas? Are there factors statistically significant? Is the model fit improved? Now consider the correlations across the residuals of the CAPM model and the correlation across the residuals of the multifactor model: are there differences in their level and/or in their distribution?
3. Now proceed to the evaluation of structural breaks. Detect for each linear model if there exist a structural break by means of the Chow test. Determine if the identified break dates are concordant across equities.
4. As an alternative to the Chow test, re-estimate the CAPM model by resorting to a rolling window approach. Use a window of size 5 years and re-estimate the CAPM for very window in your sample by moving the estimation sample by one month at a time. Report the estimated parameters (alpha and beta, as well as the r-squared) in a plot, together with their confidence intervals and comment on the stability and significance over time. Do you note relevant changes in the parameters around the beak dates identified in point 6? Or around dates when specific events took place? Comment on your findings.