

## GROUP HOMEWORK 1 – Linear Regression and the CAPM

Instructions:

- You must provide in a compressed file the following elements: the Excel file(s) with the data you downloaded; a pdf file containing a report with the answers on the Homework questions and the role of each group member in the development of the analyses and in the report writing; the code allowing me to fully replicate the analyses included in the report, all numbers and all figures; you are free to use Markdown for Python or similar tools, conditional on providing the full working code together with the pdf document;
- Do not regurgitate class notes in the report, your answers must be motivated and all analyses properly commented; all your choices must be explained and justified.
- The grades will be given to each group, that is, the grade will be identical for all group members; the grade will be given to each group based on the report and the peer- and self-evaluation.
- Coding hint: use for loops!

**Deadline: 3 weeks after the end of lectures on Linear regression and the CAPM (for the second homework the deadline will be tailored to the exam dates).**

### Groups and reference economic sector

All groups have a reference sector of the Stoxx North America collection of indexes. You can download all relevant data by searching the corresponding *constituent list* and then downloading *all items in the list*. The reference market index is the Stoxx North America.

N.	Members	Market index		Members	Market index
1	Andretto Aresta Bissengaliyeva Montina	Technology	2	Arrighini Babiski Arruda Cottini Paciaroni	Healthcare
3	Bali Casasanta Foralosso Pellegrino	Retail	4	Capone Cortese Giorgi Roversi	Energy
5	Casotto Cilibiu Jorgji Tarenghi	Food and beverage	6	Donghi Dyussenbayeva Meinardo Tarroni	Utilities
7	Dal Pozzolo Facheris Omiraliyeva Peraro	Banks	8	Fracca Oman Pagola	Industrial Goods
9	Lai Maji Marrancone Telharaj	Financial Services	10	Gigli Lo Gatto Manzi Paganin	Insurance

### Assignments

1. Download the total return index for all constituents of your sector index at the monthly frequency; select a subset of the equities continuously available for at least 10 years, minimum 10 equities (if you do not find a sufficient number of equities pick a second sector index from those not assigned to groups – these are indices with a smaller number of constituents); 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 Market Value of all the equities you selected, the overall equity market index, and an appropriate short-term interest rate focusing on the United States market.

2. 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 time series; 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?
3. 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, take the end of month market value of the companies, and the share of each company market values compared to the sum of all the companies market value (in your selection); these are weights being all positive and summing to 1. With these weights, compute the weighted average return across the equities' excess returns; this corresponds to the evaluation of the returns of a value weighted portfolio; estimate again the CAPM and compare the possible differences in terms of alpha, beta and R-squared, including the scatter plot between the portfolio and the market. 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.
4. Perform a battery of diagnostic tests (excluding the Chow test) on the linear regression models reporting the test outcomes in one or more tables, commenting on the overall results, establishing if the linear model is correctly specified for all assets, 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.
5. 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.
6. 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 break dates identified in point 6? Or around dates when specific events took place? Comment on your findings.
7. Now introduce additional explanatory variables to your linear model. Download at least three factors: you must consider factors recovered from the Kenneth French website as well as macroeconomic variables you might collect from LSEG or from FRED; you must have at least one factor from Kenneth French website and one macroeconomic factor. Motivate the choice of the factors. 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?