Financial Econometrics 2022/2023
Exercise set 2

1 Notes

- Due date: November 28$^{\text{th}}$, 2022, 23:59 (CET).
- Send your solutions through Microsoft Teams. Start a one-to-one private chat with the instructor and attach a *single* .txt file, name it as `GX_E1`, where `X` is the group number (e.g., `G4_E1.txt` is the name for group 4).
- Fill up the `Groups_and_grades` Excel file in teams to make up the groups.
- If you work in a group, please designate *one* person to send the solutions and open only *one* chat for submitting the assignment. Please *avoid* changing the groups in the following exercise sets, and please *reuse* the same chat for the following submissions.
- The document is supposed to contain, answers, codes, and possible comments, for each question. Make a .txt file with the codes and comments that runs smoothly in the R console. This means that if I copy-paste the *whole* content from your .txt file in the R console, *the code should run all the way with no errors*, see the file *Example.txt* and copy-paste its content in R. Makes sure your code runs.
- Additionally, write your student names, surnames and number(s) on the top of the document as a comment. You can work in groups of 2-5 students or individually.
- If you would like to discuss the exercises and get some feedback before the due date you can come to office 2.23 during regular office hours (please book), or ask to arrange a different time.
- Avoid sending lines and lines of code through Teams: rely on office hours to get help.
  - As the deadline approaches, I get more and more request for help: then I end up being fully booked and don't have time for all. Plan the work well ahead!
  - Be considerate, do not expect replies on the weekend/late at night. Note that the deadline is on Monday: you don't wanna work on the assignment the day before, as I won't be around for helping.
- In the tasks, name the variables exactly as indicated! so that when, e.g., I see `r` I have a reference and understand what you are trying to do. Write your code in a way that is understandable: e.g., you would name a variable that contains prices (returns) as `p` (`r`) and not `h` (`u`).
- Last updated on 2023-11-23 at 07:16:10 (UT).

! The data file `Table_txt.txt` is uploaded along with this document in file tab under the exercises stream on Teams. The data consists of 465 observations for three variables: `AAPL` are (adjusted) closing prices for AAPL (Apple Inc.), `IRX` are the risk-free rates (30-days treasury bill), annualized and expressed as percentages (check slides), `SP` are the (adjusted) closing prices of the S&P500 index (whose returns are used in the CAPM as proxies for the US market returns).

- In financial econometrics by "returns" we mean "log-returns", unless differently specified.
- Last updated on 2023-11-23 at 07:16:10 (UT).

2 Exercises

Total points: 1.1.

1. Import the data
   (i) Import the file `Table_txt.txt` with the function `read.table`, into a variable `data`. (ii) Change columns' names in `data` respectively to `AAPL`, `RF`, `MKT`.
   Hint: (i) there are missing values in the data. This means that some columns have numbers (prices) mixed with strings (for the missing values). This causes `R` to import all the columns as strings even though they mostly contain numbers. As a consequence, doing simple math operations on such columns turns unfeasible (e.g. `R` does not allow you to multiply a string by two). Check the documentation for `read.table` and use properly the optional parameter for handling missing values (besides the other you might need to specify for correctly importing the data). What is the string used to indicate missing values in the data? (ii) type `?colnames`. Note that you are not asked to replace the missing values with something else, just keep them as and where they are.
   0.1 points.

2. Excess returns.
   (i) Compute daily[1] excess returns, express them as a percentage[2]. (ii) Use the function `cbind` to combine the excess returns for AAPL and MKT into a single matrix. (ii) Use this matrix to create a data frame of excess returns, assign it to a variable `r` (i.e. call the new data frame `r`). (iv) Change columns' names to `zAAPL` and `zMKT`.
   Hint: (i) for AAPL and MKT your imported data (`data`) contains n prices (rows), for RF your data consists in n risk-free rates. When you compute log-returns you end up with $n-1$ returns which need to be subtracted to the first $n-1$ RF values. (ii) Risk-free rates are expressed on an annual basis and as a percentage (10 stands for 10%) while log-returns are based on daily prices (therefore they are not annualized) and not expressed as a percentage (0.1 stands for 10%). (iii) One-by-one, extract the price columns from `data` (use e.g. `data$column`) for `AAPL` and `MKT`, compute excess returns and store them into two new variables (e.g. `zAAPL` and `zMKT`). Then use `cbind` on these two new variables and lastly call the `data.frame` function on it.
   0.3 points.

3. Fit a linear model.
   (i) Type `?lm` and learn how to fit a linear model on the data frame `r` with an intercept (plenty of examples on Google too). Use AAPL excess returns as the dependent variable. Assign the fitted model to some variable. (ii) Call the function `summary` on the variable storing the fitted model. (iii) Read the residuals' degrees of freedom that `summary` displays and compute them manually. Store the output of `summary` into a new variable called `res` (which stands for "results"). (iv) Comment on the t-stat and p-values for the estimated parameters.
   Hint: (i) `lm` fits a model and that is, `summary(fitted_model)` computes and prints additional statistics about the model, among them the standard errors, t-statistics, p-values and residuals' standard deviation. (ii) We know from the classes how returns' degrees of freedom should be, what is happening here?
   0.2 points.

[1] To convert the risk-free rates expressed on a yearly basis to a daily basis divide by the number of periods, in this case 360. The RF rate is computed by the convention that a year has 360 days, for more check the last slides of Chapter 3.

[2] This means nothing else than e.g. using 10.5 instead of 0.105 for log-returns.

4. T-statistic and p-value.
   (i) Access the data from the `summary` call (i.e. stored into the variable `res`) and extract the estimated intercept and its standard error. (ii) Use these values to manually compute the t-statistics for the intercept (this must be the same as the one `summary` prints). (iii) Manually compute the p-value for the t-statistics (t_{stat}), the solution must match the value displayed for the intercept by `summary` in the column `Pr(>|t|)`.
   Hint: (i) To access the data in `res` you might use e.g. `res$...`. (ii) Basically you are asked to compute $\Pr(t > |t_{\text{stat}}|)$: this means "Probability that t is less that $-(t_{\text{stat}})$ *or* that t is greater than t_{stat}", where of course $\Pr(t > t_{\text{stat}}) = 1 - \Pr(t < t_{\text{stat}}) = 1 - \text{CDF}_X(t_{\text{stat}})$. What is the distribution X to be used? And what is the function to be used to evaluate the CDF, i.e. how to evaluate in `R` $\Pr(X \leq x)$ for a certain choice of X?
   0.2 points.

5. Residuals' standard error.
   (i) From `res` extract model's residuals. (ii) Make a plot of them and provide a comment on possible issues this regression might have (iii) Manually compute residuals' standard error (this must match the value `summary` prints).
   0.1 points.

6. Confidence interval for the intercept.
   (i) Compute the 95% confidence interval for the intercept. (ii) What happens if you change the confidence level at 99%? Provide a comment.
   Hint: (i) Form `res` you already extracted the values of the estimated intercept and its standard error: all you need now is to get the proper critical values from the correct distribution and compute the upper and lower bound of the acceptance region corresponding to the H_0 hypothesis intercept $= 0$.
   0.1 points.

7. Variance of the slope.
   (i) Manually compute the variance of the slope.
   Hint: (i) You need to calculate the total sum of squares for the independent variable STT_{zMKT} . What happens if you try to compute STT_{zMKT} on `r$zMKT`? In the console type `?na.omit`.
   0.1 points.