Financial Econometrics

Introduction to R

Maximilian Rohrer

Disclaimer: This problem set shall give you a broad overview over tasks commonly performed in R. The problem set is designed to be fairly comprehensive. You will find code very close to the solution to each task in the code of the video lecture. In addition, if a certain task is too hard, jump over it and look at the solution when published.

Task 1: First steps

- Open R-Studio
- Set the working directory to the folder you want to work in
- Intall the “plm” package and load it for the current session
- Declare two variables with a name of your choosing that consist of 100 randomly generated numbers each. (Hint: `rnorm(...)` generates normally distributed random variables)
- Check if both variables you declared are in working memory (Hint: `ls()` might help)
- Make a scatterplot with one variable on the x-axis and the other one on the y-axis (Hint: `plot(x = , y = )` makes a plot)
- Delete both variables and check if they are still in working memory

Task 2: Data objects and structures

- Make a vector holding the numbers from 1:100
- Construct a dataframe, in which column one is called “digits” and contains the previous vector, column two is called “year” which holds the number 1843 for each entry, and the last column is called “text” which contains some characters of your choosing.

Task 3: Import, export, and save data

- Read the \texttt{factor_returns.csv} file into R (use `read.csv(...)`). The file contains the daily returns of the market and the three Fama-French factors. Note that you first have to download the file and save in your working directory (Check out wikipedia if you want to know more about the three Fama-French factors)
- Investigate the data (Hint: use either `str()` or `summary()`)
- Save the data to your working directory (Hint: use `save()`)
- Write the data as a “.txt” file to your working directory. Use “\t” (tab) as a separating parameter (Hint: use `write.csv(..., sep = "`

`t")`. (Hint: either search for an example online or type `?write.csv` for explanations.) Open the resulting file in any text reader to see if it is tab delimited.

- Check if the two files created in the previous tasks are in your working directory (Hint: use `list.files()`)

Task 4: Manipulation

- Extract smb (column 3) for all days after end of 1999. Note, it is not strictly necessary to transform the dates into dates class. It is actually easier, to use the dates as numerics for this assignment and use logical operators, such as > and <.
- Apply following tasks to the series extracted: compute mean, standard deviation, and summary, and display a histogram

Task 5: Regression

- Run a regression with `mkt.rf` on the left-hand side (dependent variable) and `smb, hml, rf` on the right-hand side (independent variable). Only use data after 2000 for this regression. There is no need to understand the output yet, we will cover regression in the next three weeks.

Task 6: Loops [optional]

- Run the regression separately for each individual year from 2000 to 2005. Approach: (1) Write a loop that iterates over the sequence from 2000 to 2005 (Hint: Use `2000:2005`), (2) For each iteration, run the regression. (Hint: use dates as numerics and logical operators such as > and <), and (3) print the results to the console (Hint: use `print()`)

Task 7: Tidyverse [optional]

- Take the factor returns, (1) construct the market return (mkt.rf + rf), (2) limit to pre 2000, (3) order smb in decreasing order, and (4) limit the data.frame to hml. Only use tidyverse syntax.