1 taking a look at your data

Before setting off into econometrics, it's always worth to study the data. In this session, we calculate sample statistics, create histograms, fit distributions and plot extensively – all for the sake of better understanding the dataset at hand.

1. Open RStudio, create a new R script and save it in a dedicated folder. Set your working directory accordingly (*Session → Set Working Directory → To Source File Location*).

2. Load the following R packages: `ggplot2` (for plotting), `xts` (for dealing with time series objects), `fBasics` (for detailed summary statistics) and `tseries` (for testing the normality of returns). We will need them at some time during this session.

3. Load `s1_data.txt` into RStudio. This dataset contains the following variables:

name	description	source
SP500	S&P500 index	R. Shiller's website
SPDIV	dividend per "share" of the S&P500 index, annualized	
USRF	3-Month Treasury Bill, secondary market rate	St. Louis Fed
MSCIE	MSCI Europe index, in USD	Datastream
GOLD	Price of troy ounce of gold on LME, in USD	
SMIUSD	Swiss Market index, in USD	
BORD	Liv-ex Bordeaux 500 index, in USD	

4. Transform your data so as to get a time series object (using the `xts` package). Explore the properties of your time series.

5. Plot the index values of *SP500* and *SMIUSD* together. We see that this is not informative since the series do not start at the same date and have different scales.

6. Compute the logarithmic returns for *SP500*, *SMIUSD* and *MSCIE*. These are the three indices we will deal with during the remainder of the session.

7. Let us look at the period since 1990-01. Subsample the dataset accordingly.

   Henceforth we will only deal with returns. Also, we pretend that dividends do not exist for the time being.

8. Risk-free rates are usually (not on Kenneth French's website though) quoted in percent per annum: the number "0.23" in December 2015 thus (roughly) means that you could have earned 0.23% over the next 12 month, which is tantamount to say that in December alone you could have cashed in only 0.23%/12=0.02%. Transform the US risk-free rate into monthly values.

   Hint: you should see value of 0.3533 in January 2006

9. Compute the Sharpe ratios of the three risky assets. Define the Sharpe ratio of series i as:

$$SR(r_i) = \frac{\mathbb{E}[r_i - r_f]}{\sigma[r_i]}$$

where r_f is the risk-free rate. Interpreting it as a risk-adjusted return, which of the three assets is the best investment?

10. Calculate the correlation matrix of the returns of *SP500, SMIUSD, MSCIE*, and *rf*.

11. The *SP500* return of 0.6% per month that an American investor could have earned since 1990 is something like 8% per year, after compounding. Can we be sure – say, 99% sure – that this is not due to luck and that the true expected value of *SP500* monthly return is not zero? Run a simple t-test assuming that returns are not autocorrelated at any lag.

12. Can we be sure – say, 99% sure – that the market has on average outperformed the T-Bills since 1990? In other words, is the probability of rejecting the null hypothesis

$$H_0 : \mu_{SP500} = \mu_{rf} \tag{1}$$

lower than 1% given the evidence since 1990? Perform a paired sample t-test.

13. Look at the descriptive statistics of the three risky assets. How much could you have lost in a single month from an investment in each? What is the probability of losing this much or more if returns admitted a normal distribution with mean and variance equal to the sample estimates since 1990?

14. Study the skewness and kurtosis of the three series and identify the most normally distributed one. Plot a histogram (choose the number of bins as to maximize visual attractiveness) of this series and test the hypothesis that the underlying DGP[1] admits a Gaussian density.

15. Go on and create a Q-Q plot of the returns of *SMIUSD*.

16. Create a scatter plot of *smiusd* versus *sp500*, with *smiusd* on the x-axis. Identify the years when the Americans were doing alright, but the Swiss were not. Identify the reverse case.

[1] data-generating process

homework

1. Load `s1_data.txt` into RStudio.

2. Calculate the monthly logarithmic returns of *SP500, SMIUSD, GOLD, BORD* and *MSCIE*. Convert the risk-free rate to monthly values. Subsample your data to start in January 2004. Report the mean logarithmic return for *GOLD* and *SMIUSD*. (**2 points**)

3. In separate graphs, plot the time series of *BORD* index and of *BORD* returns, using `ggplot2`. (**1 points**)

4. Test the null hypothesis that the average monthly return of *BORD* is significantly different from zero at the 5% level. Report the t-stat and whether you reject or cannot reject the null. (**2 points**)

5. Calculate the 95% confidence interval around the mean value of *GOLD*. Can you be 95% sure that *GOLD* has been a better (in terms of average return) investment than the risk-free rate? (**2 points**)

6. Calculate the correlation matrix of the five risky assets' returns. With which assets are *BORD* and *GOLD* most correlated, respectively? (**1 point**)

7. Among the five risky assets, report the asset with the lowest excess kurtosis and the corresponding excess kurtosis coefficient. Use `ggplot2` to plot a histogram for this asset's returns, fitting a normal distribution using the sample mean and variance. Is the histogram of the returns approximately normal? (**2 point**)

8. Report which of the five risky assets is farthest from a normal distribution and report its Jarque-Bera test statistic. (**1 point**)

9. Calculate the expected return of an equally-weighted portfolio of all five risky assets and its Sharpe ratio. Report the former in percent per month and the latter in fractions of 1. (**1 point**)

   Hint: The function `rowMeans()` *can be useful for computing means by row.*

10. Test if the above average return is statistically different from zero by calculating a 95% confidence interval around the expected return. Report the interval and the inference conclusion. (**2 points**)

11. In this question, we reconsider the whole sample (1962:01 - 2017:12). Micro-studies are often published about whether markets are doing better under a Democratic or a Republican president. Let's conduct one of our own. **(6 points)**

    (a) The file `potus_by_party.txt` has one single dummy variable taking a value of 1 if the President was a Democrat in a particular year, and a value of 0 otherwise. Import this variable.

(b) Then, we should not forget about dividends, since they are an important part of the actual return. At time t you buy the index at a price listed in *SP500*, but your actual return at time $t+1$ is the value of the index at $t+1$ plus the dividend (both in USD) collected between these dates. Create a new series by summing the dividends and the index levels. Calculate the total return by dividing this new series by lagged prices.

Hints: (1) the average total return over 1962:01 - 2017:12 should be 0.7904% per month.
(2) The function `lag()` *can be used to lag a series.*
(3) When dealing with missing data (NAs), the expression `na.rm=TRUE` *can be useful.*

(c) Now, select only those dates when a Democrat president was in power and calculate the average total return. **(2 points)**

Hint: In order to restrict your data based on a criterion such as a dummy, one option in R is to use an expression such as `mydata[mydummy==1]`*.*

(d) Repeat the same exercise for the months under a Republican president. **(2 points)**

(e) Use the `t.test` function of R to perform a two-sample t-test (not a paired one), comparing the two means:

$$H_0 : \mu_D = \mu_{GOP} \tag{2}$$

Report and interpret the results of this test. Are the two means significantly different at the 5% level? **(2 points)**