Practical Session 4

Aitor Muguruza

Department of Mathematics, Imperial College London
E-mail address: AM5115@ic.ac.uk

Tuesday 14^{th} February, 2017

For this session it is strongly recommended to use R and the *lm* function in order to perform a linear regression. Moreover, the *summary* function gives all necessary information related to the fitted coefficients.

Question 1: CAPM

Choose a set of stocks of your choice that aim to represent different sectors of the economy (use at least 8 stocks) and download their daily quotes from either Yahoo Finance or Google Finance. You will then need to transform stock prices into returns, for this we will use log-returns, i.e.:

$$\mu_{t_i} = \log\left(\frac{S_{t_{i+1}}}{S_{t_i}}\right), \quad i = 1, ..., N$$

where N is the number of data samples we have.

In addition, also download an index that represents the market in your chosen economy, e.g. S&P 500. Since index and stock quote days differ it is suggested to either fill the missing days with the average of the previous and next days, or to use an index tracking Exchange-Traded Fund (ETF) as SPY instead.

1. **Clean the data:** Make sure that there are no missing values on your data.

2. **CAPM:** The CAPM is given by the following linear regression:

   $$r_{j,t} - r_{f,t} = \alpha_j + \beta_j\left(r_{mkt,t} - r_{f,t}\right) + \varepsilon_{j,t}, \quad t = 1, ...N, \ j \in \{\text{set of chosen stocks}\}$$

   where by little abuse of subscript notation we have that $r_{f,t}$, $r_{j,t}$ and $r_{mkt,t}$ represent the risk-free rate, stock j returns and market returns respectively at time t. Finally, ϵ is a noise term. In order to simplify the model we shall assume that the risk-free rate is null, hence we have:

   $$r_{j,t} = \alpha_j + \beta_j r_{mkt,t} + \varepsilon_{j,t}, \quad t = 1, ...N, \ j \in \{\text{set of chosen stocks}\}$$

   Comment the meaning of this formula with particular attention to the β coefficient.

3. **Fit the linear regression:** Fit the CAPM model by Ordinary Least Squares (OLS) to each stock. For this, you may use your chosen index as the market.

4. **Interpretation:** Interpret the meaning of the β's and conclude whether they are significant or not. For this you will need to perform the following test:

   $$H_0 : \beta = 0$$
   $$H_1 : \beta \neq 0$$

   Hint: You will find useful the general asymptotic result for a regression $y = \beta X + \epsilon$, where we get $\hat{\beta} \sim N\big(\beta,\ \sigma^2(X^T X)^{-1}\big)$ asymptotically and $\sigma^2 = Var(\epsilon)$ (of course R computes this for you, but it is important to realise where the statistics come from and more importantly the underlying assumptions).

5. **Coefficients over time:** The CAPM model is a static model in the sense that both α and β are fixed for each stock for all times. Repeat again part 3 using a rolling window of 100 days and plot the β's that you obtain. Analyse your results and conclude whether it is sensible to assume that this coefficients are constant over time.

Question 2 : Gauss-Markov assumptions

Under some assumptions we have that the OLS estimator is the Best Linear Unbiased Estimator (BLUE). Let us check if some of these conditions are satisfied in our CAPM example.

1. **Perturbations are uncorrelated:** The assumption on the residuals is that $E[\epsilon_i \epsilon_j] = 0$ for all $i \neq j$. Plot the autocorrelation function of the residuals and comment if this condition is satisfied.

2. **Homoscedasticity of perturbations:** The assumption on the residuals is that $Var(\epsilon_i) = \sigma^2$ for all i, i.e. the variance of the perturbations does not change with time. In order to analyse this, plot the residuals of each regression and comment if the volatility of the time series seems constant over time.

3. Determine whether it is sensible to assume that **Gauss-Markov** holds.

Question 3: Lagged regression

You may be tempted to use past observation on the stock or even the index to try to predict the future behaviour of the data. Perform a regression using lagged observation of the corresponding stock as well as the index and comment your findings. You may also try to add lagged squared returns or any other variable of your choice as long as it is in the past (this is indeed what hedge funds try to do with far more complex statistical models!! Linear regression is in fact one of the most simple, yet useful, machine learning devices).