What did we do last Monday ?

**April, 12th**

After introducing the econometric field and the content of the course, we discussed the idea of Louis Bachelier. He postulated that returns are realizations of independent and identically distributed Normal distributions. To investigate the implication of this hypothesis, we reviewed some well-known summary statistics such as the average mean, the median, the standard deviation, the skewness and we related them to random variables. In particular, we showed that the average mean and the variance (i.e. the squared standard deviation) are unbiased estimates of independent and identically distributed normal distributions. The results can be extended to any independent and identically distributed distributions. We then discuss the hypothesis of Louis Bachelier by emphasizing that financial returns are typically not normally distributed. In fact, the Normal distribution does not assign enough probabilities to extreme events (like the black swans of N. Taleb). We ended the course by reviewing a law of large numbers (LLN). We showed that the LLN helps approximate any expectation of independent random variables as long as we have enough realizations from these random variables.

**April, 26th**

We started with a statistical refresher: An estimator aims at estimating a quantity of interest (such as the expectation or the variance) and it is a random variable (that is usually a function of many random variables). In statistics, we usually care about three properties for an estimator: unbiasedness, efficiency (we have not covered that property yet) and consistency. An unbiased estimator implies that its expectation is equal to the quantity of interest. An estimator is more efficient that another one for estimating the same quantity if it exhibits a smaller variance than the latter and a consistent estimator implies that, as the sample size increases, the estimator converges (in probability) to the quantity of interest. We discussed these statistical properties using the average of random variables that exhibit the same expectation. We also reviewed the law of large numbers and showed that it helps approximate any expectation of independent random variables as long as we have enough realizations from these random variables. After covering an empirical exercise in R on summary statistics and on risk minimization, we introduced the simple and the multiple linear regressions. We started with the simple regression, which consists of drawing a line between a dependent variable (the variable of interest) and an explanatory variable (the variable which explains linearly the values of the dependent variable).  We showed that setting the sum of the error terms to zero was not a sufficient criterion to both identify the intercept and the slope of the regression. Then, we found the estimators of the intercept and the slope coefficients by minimizing the sum of squared residuals (also called SSR). These estimators are called ordinary least squares estimators or simply OLS estimators. So, given a sample of the dependent variable y\_t and the explanatory variable x\_t, we can get estimates of the intercept and the slope coefficients by plugging our sample values into the OLS estimator formulas. At the end of the course, we introduced the coefficient of determination (also known as the R squared) which assesses by how much the explanatory variable improves the fit with respect to the sample average. We emphasized that the coefficient of determination is just the square of the empirical correlation between the dependent variable and the fitted values.  For the simple regression, the coefficient is also equal to the square of the correlation between the dependent variable and the explanatory variable.

**May, 3rd**

We first reviewed the mathematical framework of the linear regression. It consists in finding estimates of the slope and the intercept and in drawing a linear relation between the dependent and the explanatory variable. The framework only requires a criterion to find estimates of the coefficient parameters. We minimized the sum of squared residuals, which is a criterion that leads to the ordinary least squares (OLS) estimator. We showed an empirical example of the mathematical framework on Google colab using Amazon financial returns as the dependent variable of a CAPM. We ended this section by discussing the linearity assumption and how interpreting the estimates when the variables within the linear framework are transformed using the logarithm function.

We then started to focus on the statistical framework. This framework requires additional assumptions but makes our OLS estimator interpretable. We emphasized that these assumptions are needed to go beyond the observed sample and to have an idea of the possible values of our OLS estimates when we switch from one sample to another. We first focused on the first three assumptions that are particularly related to the unbiasedness property of the OLS estimator. We discuss in details the implications of the strict exogeneity assumption. Then, we covered the three last assumptions that are: white noise, homoskedasticity and that the error terms are normally distributed. We showed that the white noise assumption (no correlation between error terms) is a strong one in a time-series context. We ended the course by showing that if the strict exogeneity assumption holds (as well as the linearity and the no-collinearity assumptions) then the OLS estimators are unbiased. It means that the expected value of the OLS estimator is the true parameters of the linear regression.

**May, 10th**

We started the course by reviewing the difference between a conditional distribution and an unconditional (or marginal) distribution in the linear regression context. We then discussed the statistical properties of the OLS estimator. In particular, we focused on unbiasedness, consistency and efficiency. We showed that some *linear* estimators can have a smaller variance than the OLS estimator but, in such a case, they must be biased. Using graphics, we derived the variance of our OLS estimators under the additional assumptions of white noise and no correlation (of the error terms) and we highlighted how to estimate the variance of the error terms using the residuals. By adding the normality assumption, we proved that the OLS estimators are normally distributed. It allows performing statistical tests. We introduced the one-sided and the two-sided tests. These tests consist in assuming a Null hypothesis that the true coefficient is equal to a specific value and then rejecting the hypothesis if the test statistic is too far from the realizations we expect under the Null hypothesis. This approach is motivated by the fact that the test statistic should be one realization of a standardized Normal distribution if the Null hypothesis is true.  For rejecting the hypothesis, we need to set a cut-off (or a threshold). The significant level of the test fixes this threshold. For instance, a significant level of 95% implies that we reject 5% of the tests when the Null hypothesis is true. In a two-sided test with a significant level of 95%, the corresponding threshold of the standardized Normal distribution amounts to 1.96. It means that we reject the hypothesis is the test statistic in absolute value is above 1.96.

**May, 25th**

We first reviewed the statistical part of the linear regression framework and how performing a statistical test at a specific significant level. Then, we discussed three possible ways to perform a statistical test that are 1) the standard summary statistic compared to a quantile of the distribution (i.e. the threshold like 1.96 for a 95% significant level when the statistic follows a standardized normal distribution under the Null hypothesis), 2) the confidence interval and 3) the p-value. After emphasizing the issue related to the error of type one, we discussed the fact that the significant level of a statistical test is still a hot issue in current research. Statistical tests are possible when we know the distribution of our estimator. To relax the restrictive assumption stating that the error terms are normally distributed (which makes our OLS estimators normally distributed), we introduced and illustrated the central limit theorem (CLT) for i.i.d. random variables. When the sample size is large, the CLT makes the OLS estimator normally distributed even though the error terms are not normally distributed.

**May, 25th (second course)**

We introduced the multiple linear regression framework with the omitted variable bias. To do so, we watched a short TedX video presenting a research on how social status can affect individual’s generosity. Not surprisingly, the multiple linear framework is just an extension of the simple linear one as it consists of having a linear model with more than one explanatory variable. This extension can also be decomposed into two parts: a mathematical and a statistical framework. We started with the mathematical framework and, by using linear algebra and matrix notation, we derived the analytical expression of the OLS estimators that minimize the sum of squared residuals.  We then showed that the OLS estimator formula based on matrix notations gives the same OLS estimates as those derived in the simple linear regression framework. Next, we moved to the statistical part of the multiple linear regression framework. We introduced the six statistical assumptions and emphasized that the most different assumption compared to the simple linear regression framework was the no multicollinearity hypothesis. Relying on these assumptions, we proved the unbiasedness property of the OLS estimator. Then, we showed that the OLS estimator given the explanatory variables is Normally distributed with expectation equal to the true coefficient and a variance given by (X’X)^{-1}sigma^2 when the six linear regression assumptions hold. Knowing the distribution of the OLS estimator makes statistical tests possible. We thus revisited the t-test which allows for testing the value of a coefficient. We ended the course by observing that the t-test is not sufficient to perform join tests (i.e. under the Null, we test jointly the values of at least two parameters). This motivates the introduction of the Fisher test that will be cover next week.

**May, 31th**

We started the course with a recap of the multiple linear regression (MLR) framework. The emphasis was on the difference between the mathematical part and the statistical part. We also reviewed the standard t-test to perform a hypothesis test on one parameter in the MLR framework. Then, we introduced the F-test (or Fisher test) which allows for jointly testing multiple parameter values. We discussed the F-statistic and why it is following a Fisher distribution. We also provided a rule of thumb to decide whether we reject the Null hypothesis in large sample using the F-test. To summarize this discussion, we coded the OLS estimates, the t-test and the Fisher test in R.

We then showed how to heuristically test and relax the normality assumption of the error term. We presented the QQ-plot which consists of plotting the quantile of a standard Normal distribution with respect to the quantile of the normalized residuals sorted in ascending order. Then, we explained that the OLS estimator remains normally distributed if the error terms are not normally distributed but still i.i.d. in large sample (as long as the explanatory variables are also i.i.d). This is due to one central limit theorem. In time series context, the i.i.d. assumption does not make sense so we had to rely on another central limit theorem. Among the needed assumptions for this second CLT, we have to assume that our explanatory variables are stationary. Stationarity means that the (unconditional) expectation and variance of the explanatory variable do not change over time. This precludes explanatory variables exhibiting time trends. We explained how to get rid of linear trends and exponential trends. Eventually, we discussed how to relax the no multicollinearity assumption by using penalized estimators. We ended the course by introducing the Ridge and the Lasso estimators which are used in a Big Data context.