# PART 1

## 1. Introduction

Quantitative Finance is an interdisciplinary field that gels applied mathematics (probability & statistics in particular) with finance and with computing. These hands-on sessions are beneficial for you to gain a deeper understanding of the various models discussed in class.

As a group, by implementing the statistical models, you will become more proficient in programming and in handling massive (and messy) financial data. More importantly, in the process of making sense out of the results generated by your computer programmes, you will gain invaluable experience in interpreting the outcomes through the colored lens of the models. As the great statistician George Box once said, "all models are false but some models are useful." It is therefore intellectually stimulating to debate in the mind whether the model you are dealing with is wrong, or the computer code is buggy, or the data are contaminated, or all of the above.

The hands-on report after each session is to be handed in via the dropbox As for the assessment, the criteria are

1. Scientific correctness of the numbers crunched out by the computer programmes
2. Soundness and sensibility of the conclusions made
3. Organization and clarity of the report
4. Timeliness of the report submission
5. The $\mathcal{X}$ factor

As a good practice, always plot the graphs and state the sample periods. The source of the financial data must be clearly spelled out. Present the crucial and final numbers in such a way that makes it easier for the examiner to spot. It will be a bonus if some relevant financial news or insight is included in the report to add flavor to the otherwise cold and stony numbers.

The group report is to be typeset in latex with professionalism in mind. Imagine that you are a quant analyst and you are directed to write a report to your clients. Each group is to submit the report along with readable Python codes.

## 2. Dow Jones Industrial Average Index

Dow Jones Industrial Average Index, or DJIA index in short, is historically the longest equity index for the US market. It was first published on May 26, 1896. In the late 19th century, stocks were generally regarded as highly speculative and exotic securities. Relatively few people knew about stocks and even fewer people invested in them. That was the backdrop when Charles H. Dow unveiled his industrial stock average. Not surprisingly, few if any investors paid attention to this new index, for the main investment activities at that time revolved around bonds that paid coupon interests.

At the beginning, there were 12 component stocks of industrial firms. In the absence of even a hand-held calculator, Dow used a simple approach known as the price-weighted method today. He jotted down the end-of-day prices of the component stocks, added them up and then divided the sum by 12. These operations were manually manageable with pencil and paper.

The DJIA index can be easily replicated by buying an equal number of shares in each of the 12 stocks. In a way, DJIA index value is the average price in dollars of the resulting 12-stock portfolio. The first average computed was 40.94, namely, the average price of these stocks was $40.94 per share.

General Electric was one of the 12 stocks selected by Dow; the other 11 companies had long disappeared. The number of component stocks increased to 20 on October 4, 1916 in a major re-tooling of the index. Twelve years later, the list was expanded to 30, and a **divisor** was introduced to adjust for stock splits, stock distributions, and stock substitutions. Starting at the value of 16.67 in October 1928 and after 227 adjustments, the divisor is 0.129146820 as of August 13, 2012. Since then 26 more adjustments were made, and the divisor became 0.14748071991788 as of June 26, 2018 when the last survivor GE, was booted out and replaced by Walgreens Boots Alliance.

## 3. Mini-Project Task 1

You are to analyze two data sets of Dow Jones Industrial Average (DJIA) index and S&P 500 index from **Yahoo! Finance**, using all the statistical modeling tools introduced in Session 2. In addition, the following items are to be carried out.

1. Identify the 30 component stocks after the October 18, 2018 US trading session and record the end-of-day prices. Verify that the DJIA index value is indeed equal to the sum of the 30 prices divided by 0.14748071991788. Report your finding.

2. Find out more about S&P 500 and write something similar to Section 2. For example, you can talk about the history of this index and so on. What particular question is: Are the 30 stocks of DJIA index also the component stocks of S&P 500?

3. Plot the time series of the two indexes.

4. Construct the daily log return on DJIA index using the index level Do likewise for the S&P 500 index.

5. Plot the time series of log returns: $r_t = \ln P_t - \ln P_{t-1}$, where $t = 1, 2, \ldots, T$. Note that $P_0$ is simply the first index level in the sample.

6. Compute the sample mean $\widehat{\mu}$ and unbiased sample variance $\widehat{s}^2$ of log returns.

7. Compute and report the annualized average and volatility of log return in percent, assuming each year has 252 trading days. That is,

$$\text{annualized log return} := 252 \times \widehat{\mu}; \qquad \text{annualized volatility} := \sqrt{252 \times \widehat{s}^2}$$

8. Compute and report the sample skewness $\widehat{\gamma}$ and sample kurtosis $\widehat{\kappa}$ as

$$\widehat{\gamma} = \frac{\sum_{t=1}^{T}(r_t - \widehat{\mu}^2)^3}{n\widehat{\sigma}^3}; \qquad \widehat{\kappa} = \frac{\sum_{t=1}^{T}(r_t - \widehat{\mu}^2)^4}{n\widehat{\sigma}^4},$$

where the consistent variance is used here, i.e., $\widehat{\sigma}^2 = \dfrac{1}{T}\sum_{t=1}^{T}(r_t - \widehat{\mu})^2$.

9. Compute and report the **Jarque-Bera test statistic** (JB) and make an inference on the test of normality at the 5% significance level.
$$\text{JB} := n\left(\frac{\widehat{\gamma}^2}{6} + \frac{(\widehat{\kappa}-3)^2}{24}\right) \sim \chi_2^2.$$
The hypothesis is $H_0 : \text{JB} = 0$ and the alternative hypothesis is $H_a := \text{JB} \neq 0$.

## 4. Mini-Project Task 2

This part analyzes the relationship between the log returns of DJIA (1) and S&P 500 (2) indexes. In particular, the **two-sample $t$-test** is a statistical algorithm to determine if two population means are equal. An application is to test if the log return of S&P 500 index. is superior that of the DJIA index.

1. Compute and report the **correlation** between the log returns of DJIA and S&P 500 indexes.

2. Conduct a test to examine whether the two samples have equal mean at the $\alpha = 5\%$ significance level.

   The two-sample $t$-test is defined as
$$H_0: \qquad \mu_1 = \mu_2; \qquad\qquad H_a: \qquad \mu_1 \neq \mu_2.$$
   The test statistic is
$$T = \frac{\widehat{\mu}_1 - \widehat{\mu_2}}{\sqrt{\widehat{s}_1^2/T_1 + \widehat{s}_2^2/T_2}},$$
   where $T_1$ and $T_2$ are the sample sizes, $\widehat{\mu}_1$ and $\widehat{\mu}_2$ are the sample means, and $\widehat{s}_1^2$ and $\widehat{s}_2^2$ are the unbiased sample variances. Reject the null hypothesis that the two means are equal if
$$|T| > t_{1-\alpha/2,\upsilon},$$
   where $t_\upsilon$ is the critical value of the t distribution with $\upsilon$ degrees of freedom and
$$\upsilon = \frac{\left(\widehat{s}_1^2/T_1 + \widehat{s}_2^2/T_2\right)^2}{(\widehat{s}_1^2/T_1)^2/(T_1-1) + (\widehat{s}_2^2/T_2)^2/(T_2-1)}.$$
   In general, there are three possible alternative hypotheses and rejection regions for the one-sample $t$-test

| Alternative Hypothesis | Rejection Region |
|---|---|
| $H_a : \mu_1 \neq \mu_2$ | $\lvert T\rvert > t_{1-\alpha/2,\upsilon}$ |
| $H_a : \mu_1 > \mu_2$ | $T > t_{1-\alpha,\upsilon}$ |
| $H_a : \mu_1 < \mu_2$ | $T < t_{\alpha,\upsilon}$ |

3. Conduct the $F$-test for equality of two variances at the $\alpha = 5\%$ level of significance.

   An ***F*-test** is a statistical algorithm to test if the variances of two populations are equal. This test can be a two-tailed test or a one-tailed test. The two-tailed version tests against the alternative that the variances are not equal. The one-tailed version only tests in one direction, i.e., the variance from the first population is either greater than or less than the second population variance.

The two-tailed hypothesis test is defined as

$$H_0: \qquad \sigma_1^2 = \sigma_2^2; \qquad\qquad H_a: \qquad \sigma_1^2 \neq \sigma_2^2.$$

The $F$ test statistic is

$$F = \frac{\widehat{s}_1^2}{\widehat{s}_2^2},$$

where $\widehat{s}_1^2$ and $\widehat{s}_2^2$ are the unbiased sample variances. The more this ratio deviates from 1, the stronger will be the evidence for unequal population variances.

The hypothesis that the two variances are equal is rejected if

$$F < F_{1-\alpha/2, N_1-1, N_2-1} \qquad \text{or} \qquad F > F_{\alpha/2, N_1-1, N_2-1}.$$

Here, $F_{\alpha, N_1-1, N_2-1}$ is the critical value of the $F$ distribution with $N_1 - 1$ and $N_2 - 1$ degrees of freedom at the significance level of $\alpha$.

## 5. References

- **Investopedia**
- **Yahoo! Finance**
- **CNN**'s Dow 30
- **Engineering Statistics Handbook** from the National Institute of Standards and Technology (NIST).