Business Intelligence per i Servizi Finanziari

R-PYTHON LABORATORY PROJECT

Antonio Candelieri: antonio.candelieri@unimib.it

Riccardo Perego: riccardo.perego@unimib.it

R-Python Laboratory Project

The project is focalized on **data acquisition**, **visualization**, **exploratory analysis**, **predictive analysis** and (simulated) **portfolio management**, **web applet**, by using the main packages already explained during the lab:

- R: *"tseries"*, *"zoo"*, *"xts"*, *"quantmod"*, *"dygraph", "kernlab", "e1071", "Rshiny"* …
- Python: "Pandas", "Matplotlib", "Quandl", "ScikitLearn", "Dash"

and their functions:

- R: *get.hist.quote()*, *zoo()*, *merge()*, etc.
- Python: *get_data_yahoo(), pd.DataFrame(),* etc.

Students are free to use other packages (e.g. *"dygraph"/ "Matplotlib"*) in order to demonstrate their abilities in using R/Python resources/packages.

Each student has to choose **at least 5** stocks to be analyzed. These could be chosen after a brief analysis of the financial news from the main websites and should be **representative of 3 different sectors** (e.g. energy, banks, insurance, media, technology, etc.).

Results organization and final report

Each student has to produce a *10-15 minutes presentation* (e.g. power point) to present the main results and a *full report* structured as follows

- **Data summary**: an introduction with the chosen financial instruments and motivations (e.g. a set of relevant news).
- **Descriptive analytics**: a section organized in different subsections summarizing the main results from exploratory data analysis including returns, variances and covariances at different time scale (daily, weekly, monthly).
- **Predictive analytics**: a section organized in different subsections summarizing the time-series mining approaches (e.g. price forecasting, return forecasting, etc.)
- Compute the "beta" for the chosen stocks
- **Portfolio management**: the investments decisions and the return on its own "virtual" portfolio, including transaction costs
  - **Portfolio management** via mean-variance
  - **Trading** based on forecasting
- **Web applet** for data visualization of multiple task (using tools as Rshiny and Dash)
- **Conclusions**: general comments about data, analytical process and results.

Each student has to sent both the report and the R/Python code to riccardo.perego@unimib.it

Data summary

- A short description about each selected instrument and the motivations of the choice (e.g. relevant associated news, observable trends, etc.).
- Functions for loading data from Yahoo! Finance to get price data from October 2017 to October 2018.
- Merging all data into one data object (taking care to rename the columns)
- Presenting the loaded data (through a simple chart)

Descriptive Analytics (1/3)

- Compute simple and compounded monthly returns and visualize them into a time plot
- Provide comments with respect to:
  - What common features in possible to see? (plot the CC monthly return in only one graph with different colors)
  - Is there a positive correlation for stocks from the same industrial sector? Can this be identified also graphically comparing, by eye, the graphs? (beta values, correlations, indices,…).
  - Are there some unusual returns, too large or too small? If yes, try to identify news regarding the instrument that might explain the behavior
- Create diagnostic plots of histograms and scale them to understand relative dispersion for each monthly return and comment

Descriptive Analytics (2/3)

- Create four panel diagnostic plots containing histograms (to estimate pdf), smoothed density plots, boxplots and qq-plots for each return series and comment.
  - Do the returns look normally distributed?
  - Are there any outliers in the data (use boxplots() over data merged into one object in order to scale them)?

- Compute univariate descriptive statistics (mean, variance, standard deviation, skewness, kurtosis, quantiles) for each return series and comment.
  - Which equities have highest and lowest mean returns?
  - Which have the highest and lowest standard deviation?
  - Comment on volatilities and means, do they differ in different period?
  - Which equity looks most and least normally distributed? Which returns show a distribution closest/farthest from the normal?

Descriptive Analytics (3/3)

- Compute the sample covariance matrix of the returns on your assets and comment on the direction of linear association between the asset returns.
- Which equity appear to be riskier? *(consider volatility, histograms, boxplots etc. using same bins)*
- Compute the sample correlation matrix of the returns on assets. Which assets are most highly correlated? Which are least correlated?
- Compute and plot all pair-wise scatterplots (use function pairs()).
  - Briefly comment on any relationships you see.
  - Do the series seem to move together? Are there particular period in which they move closer together?
  - Does the correlation between equities change over time?
  - Does the positive correlation seem to be greater during and after a period? Do volatilities seem to increase/decrease during this time?
  - Which instruments are most/least correlated? (*Based on the scatterplot and the correlation matrix*). If you can, explain any of this consultation by analysing any specific news.
  - Consider scatterplot of returns, does it appear that cc returns of the 2 instruments are linearly dependent? *(does the scatterplot data lie close to a straight line?)*

Predictive analytics

- Infer a forecasting model (e.g. ARIMA or SVM) to predict prices or returns for every financial instruments:
  - use ***n*** months as training set
  - use ***m*** months as test set
  - leave last ***l*** months apart…

- use (last) ***l*** months to:
  - compare forecasts with actual values (evaluating the accuracy of the forecasting model);
  - validate your forecasting-based trading strategy: suppose to have a budget of $V at the begin of the ***l*** months and decide how to invest in order to maximize return; compute your net return (you must consider transaction costs) at the end of the period according to your investment decisions .

- *n = 80*
- *m = 30*
- *l = 10*

Beta computation

- Compute the Beta measure of the volatility in comparison to the market (usually also available on Yahoo Finance) for each one of the assets you have chosen
- Provide considerations about the variations of Beta over time, focusing on the periods with high volatility
- Use the value of Beta to calculate the expected return of your portfolio and compare it with the effective return obtained

Portfolio Management

- Use Mean-Variance analysis to define and manage your portfolio: choosing the weights of the 5 stocks you have choosen

- suppose to have a budget of $V at the begin of the ***l*** months (with the same value of ***l*** previously defined in the forecasting) and decide how to invest your budget to create your portoflio.

- compute your return (you must consider transaction costs) at the end of the period according to your investment decisions

Web Applet

- Use **Rshiny/Dash** to support at least one of the following tasks:
  - Descriptive Analysis
  - Predictive Analysis
  - Portfolio Management
- The student has to present the application during the exam.
- The response speed of the application will be considered for loading an forecasting task.