

1 Quantitative methods in finance - second semester - examination

Prepared and based on the lectures of Eric Vansteenberghe

Link to the source codes and the data useful for this examination

2 Introduction

The students will have to work during the session on the 9th of February 2017 on coding (in python and/or R as they please). Work will be individual. In case the student has no computer or its Spyder or RStudio environments are not working she will have to submit her work in paper form, otherwise submission is expected at the end in e-mail format to the facilitator.

You don't need to reply to all questions to have 100% of the points, your final grade is out of 6 and if you sum all the points below you could get 13 points, rather take time to comment your code and results, and chose which exercises you want to focus on.

You can, for each exercise use either R or python.

You have to submit all the codes you developed by 20:30 at the following e-mail adress:

`eric.vansteenberghe@supaero.org`

Each minute of delay will be a discount to your final grade.

2.1 Some remarks about the code you will submit

- If you have to perform the same actions several times, leverage on functions and loops
- Before performing a regression, apply econometric tests on the variables (some help can be found here)
- In your code, use decipherable variable names and try to avoid hard coding values (typical example: define a variable `nbobs = len(df)` instead of using the actual value (let's say 1748) of your DataFrame in your code
- Include comments in your code

3 PageRank - A Google algorithm (4 points)

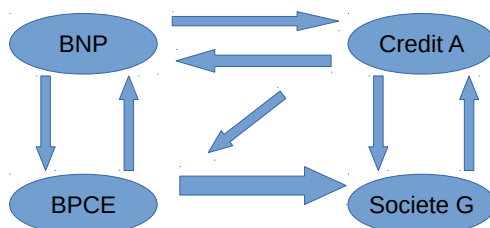
We present the simple model as in Larry Page and Sergey Brin, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Published in Computer Networks 1998. This core algorithm is at the heart of the search engine they created, Google.

You will be guided to some variables definition and will be asked to create a simple version of the PageRank algorithm.

3.1 See a Page as a bank

For illustration purpose, think instead of a web page as of a bank. Each bank is lending money to another bank, hence they are linked. A central and first question we want to answer as a financial supervisor is: "who are the 'central' banks of our network?"

Each bank points to another bank by lending, each web page points to another web page with hyperlinks.



In this visual example, the thickness of the arrow represent how much a bank lend to another. Here we can see that some bank do not lend to some others. We would like to be able to know from this set up which bank is the most "important" one to the network.

3.2 Variables definitions

- N banks (or node)
- p_i the Total Liabilities of bank i
- $p_{i,j}$ payment due from bank i to bank j
- Π the relative liabilities matrix

$$\Pi_{i,j} = \begin{cases} \frac{p_{i,j}}{p_i}, & \text{if } \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

3.3 Naive implementation of PageRank

In their first description of the PageRank algorithm, we can translate the PageRank for a bank PR_i as:

$$PR_i = (1 - d) + d \sum_{j \neq i} \frac{PR_j}{p_j} \quad (1)$$

with $d \in [0, 1]$ a damping factor.

Per design, we have $\sum_j PR_j = 1$

We see that there is some sort of circle in defining the PR 's. Hence we need to proceed iteratively.

- **Question 1**

- set $d = 0.85$, $\text{MaxIter} = 100$, $N = 4$, and $\epsilon = 10^{-5}$
- define r_0 your initial guess for the vector of PR_j , $j \in [1, N]$. $r_0 = [\frac{1}{N}, \dots, \frac{1}{N}]$

Then we need to compute the relative liabilities matrix Π .

- **Question 2**

- Import the price data on banks from the file `2016_10_12_bank_prices.csv` and select only the prices of the following 4 french banks, compute their daily returns:
 - * BNPPARIBAS
 - * CREDITAGRICOLESA
 - * BPCE
 - * SOCIETEGENERALESA
- Compute the correlation matrix, compute (1- correlation matrix) and consider that this is your relative liabilities matrix Π
 - * This is a very poor assumption, it is assuming that the market would react on one bank according to how much they lent to one another, but this will do for our exercise.
- Then compute r_1 as: $r_1 = (1 - d) + d \times \Pi \times r_0$
- You then need to create a loop and stop your loop either once $(r_1 - r_0)^2 < \epsilon$ or once we have reach the maximum number of iterations MaxIter .
- List the bank and the PageRank score you find.

3.4 With teleportation - bonus

In order to treat dead-end nodes, we add a teleportation distribution f . In practice, it is a vector of length N satisfying $\sum_i f_i = 1$.

We then define compute the PageRank vector r as:

$$r = (1 - d)f + d(\Pi + f d'_{out})r \quad (2)$$

With:

$$d_{out,i} = \begin{cases} 1, & \text{if } \sum_j p_{j,i} = 0 \\ 0, & \text{otherwise} \end{cases}$$

- **Question 3**

- Make a simple assumption about f : like 1% uniform chance $f = 0.01$
- Keep this question if you have time at the end of the examination period and implement an algorithm with the teleportation aspect.

4 Before and after regressing - some steps (2.5 points)

Use the files `pea_price.csv` and `pea_price_2.csv`. Compute the returns

- **Question 4**

- Keep only stocks that have no NA (all element of the column are a number).
- Compute the daily returns of the stocks.
- Plot a random selection of 7 stock returns.

- **Question 5**

- Perform an ADF test on the time series for the stock price of one of the stocks (just randomly select one stock and carry on!). Explain which pair of hypotheses you are using. Bonus: does your result change with the information criterion used to select the lag length (e.g. SIC instead of AIC)?
- Now, perform the ADF test on the time series for the *first difference* of the stock price. Specify the ADF regression and the hypotheses.

- **Question 6**

- Propose a strategy to decide which 4 stock to buy and implement it. Imagine you have 10 000 euros to invest. You use the first half of the price history to make your decision, then you buy the stocks and keep them for the remaining of the second half and sell them at the end.

Imagine I create a strategy: "Buy the stocks which price increased most", you rank the price increase of each stock from 2006 to 2011. In 2011 you spend you 10 000 euros buying the 4 stock which prices increased most, then you sell the portfolio in 2016, what is your return?

Some strategies you could chose from:

- * large cap "under priced"
- * momentum
- * mean-reversion
- * generous dividends
- * low/high volatility

- **Question 7**

- Select the 10 stock which are the most volatile (you first rank the standard deviation, then take the 10 highest).
- From those 10 stocks, Create a market index (use the price here, not the returns, to build the index):
 - its starting value should be 100
 - only accept stock in this index with no N/A
 - equally weight the stocks
 - plot this index

- Test the most volatile stock of the 10 and the market index for unit root, test their returns
- If relevant, regress the most volatile returns on your market index returns
- **Question 8**
 - Draw a random realization of returns, from a gaussian distribution with the historical mean and standard deviation of the stocks used above, or the index, as you please.
 - plot the histogram of the historical stock returns and of the random realization.

5 AR, MA and ARIMA (1 points)

Use the files `pea_price.csv` and `pea_price_2.csv`.

Consider the following MA(2)-Process

$$y_t = u_t + 0.9u_{t-1} - 0.3u_{t-2}, \quad u_t \sim i.i.d.(0, 1) \quad (3)$$

Consider the following AR(1)-Process

$$y_t = 0.9y_{t-1} + u_t, \quad u_t \sim i.i.d.(0, 1), \quad y_0 = 0, \quad t = 1, 2, 3, \dots \quad (4)$$

- **Question 9**
 - Plot the autocorrelation and partial-autocorrelation functions ($k = 0 \dots 10$) for the two processes.
 - Try to fit an ARIMA model on the stock returns of 10 stock you randomly selected from the PEA. What orders do you select, what are the value of the coefficients?
 - Project your ARIMA model for 60 days, compute how it translates in terms of price and plot it.

6 CAPM and time series (1 points)

Use the files `pea_price.csv` and `pea_price_2.csv`.

- **Question 10**
 - Present the CAPM equation, which riskless return do you propose, regress some stock returns on a market proxy you construct yourself from the PEA data set.

7 Macroeconomy (2.5 points)

We propose to use some data of the French economy based on usual variables.

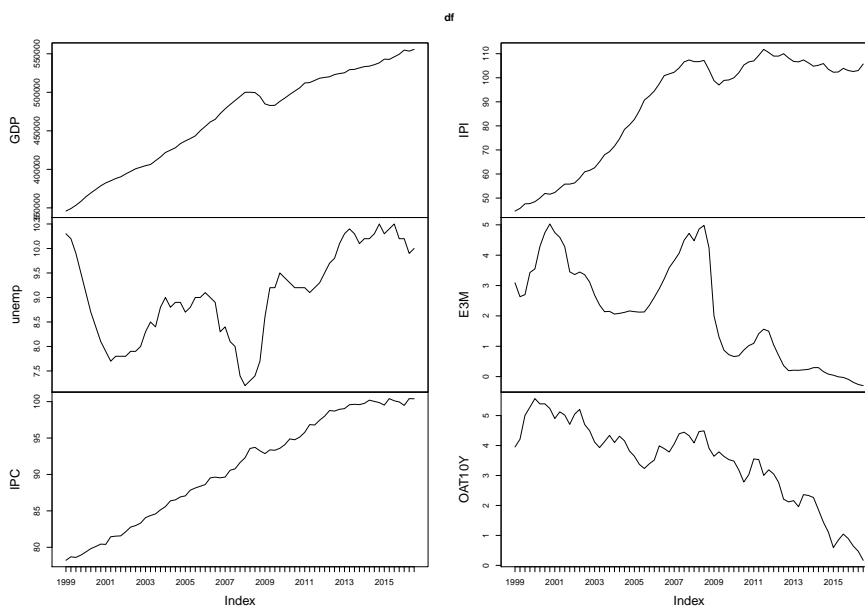
Our data are taken from INSEE and Banque de France website. Click on a series name to be redirected to the source:

- daily data
 - **E3M** the 3-month Euribor
 - **OAT10Y** French government rate at 10 year

- monthly data
 - **IPC** consumer index price
- quarterly data
 - **unemp** unemployment rate
 - **PIB** Gross Domestic Product
 - **IPI** retail real estate price index

The data can be found under the files:

- var_daily.csv
- var_monthly.csv
- var_quarterly.csv



• **Question 11**

- Take any combination of variables and perform 3 regressions.
- Now focus in the real estate price, what other variable seems to explain its evolution?
- Create a simple VAR model, please order the variables as follow:

$$\begin{bmatrix} GDP_t \\ unemp_t \\ IPC_t \\ IPI_t \\ E3M_t \\ OAT10Y_t \end{bmatrix}$$

- Use impulse response function, the magnitude of which should be chosen from the mean and standard deviation of the historical observation.
- Which variables seem to cause one another (Grange causality)?
- Bonus: build a VECM with the same variables and comment on your findings.

8 Campbell and Shiller, JPE, 1987 (2 points)

In their paper, Campbell and Shiller propose methods to test present value model such as:

$$Y_t = \theta(1 - \delta) \sum_{i=0}^{\infty} \delta^i E_t y_{t+i} + c \quad (5)$$

With:

- Y_t can be the long-term yield or the stock price
- y_t can be the one-period rate or the dividend
- δ is the discount factor
- θ is the coefficient of proportionality
- c is a constant

We propose to use the data made available by Robert Shiller on his website from his book Irrational Exuberance.

In the file ie_data.xls, used by Robert J. Shiller, and available online, one finds the following series:

- P: S&P 500 index (prices)
- D: S&P 500 dividends (annual rate)
- E: S&P 500 earnings
- CPI: US consumer price index
- GS10: yield to maturity of 10-year US Treasuries, this is referred to as the long term interest rate in his book Irrational Exuberance
- CAPE: cyclically adjusted price-earning ratio, i.e. (inflation-corrected) S&P Composite Index divided by the ten-year moving average of real earnings on the index
- **Question 12**
 - Propose some econometric tests between the different variables (e.g. between dividend and earnings). Comment your results.
 - Propose a VAR model using some of the variables and the US GDP time series. Impose a positive shock on dividends, what would be the effect on GDP your model predict (is the predicted effect positive or negative in the short term, and what about the long term)?