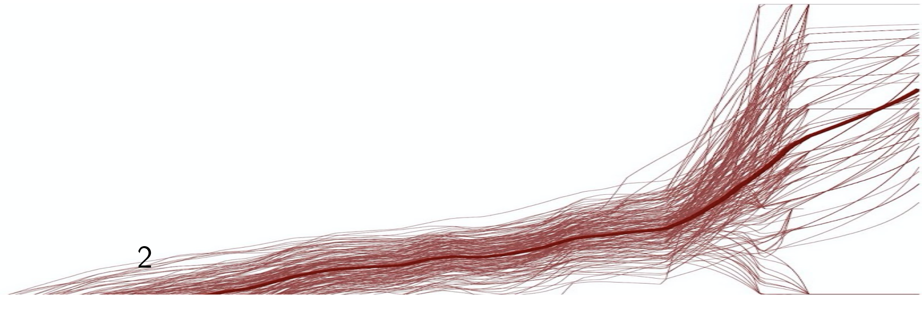
Financial Econometrics
Lab Sessions & Homeworks

Pr. Gaëlle Le Fol
UNIVERSITE PARIS - DAUPHINE
Master 203 - Master 104

September 2024

Table des matières



Preparing the lab classes

Lab classes provide students with the opportunity to apply and investigate theoretical and conceptual knowledge, to develop and code a range of techniques and approaches, and to improve skills in analysing, and interpreting data. However, if you come to the lab class unprepared, the only thing that it will provide you with will be the opportunity to copy code lines.

To make the most of these lab sessions, you must :

- Read through the Lab class handout and check that you know hwo to use the commands/functions/tools you will need all along the exercises of the day.
- Revise the part of the course that the lab class is exploring.
- Check with your lab class information to see if you are expected to do any preparatory work before the class.

By failing to prepare, you are preparing to fail
Benjamin Franklin

Preparation time is never wasted time
Anonymous

Catching up with Gretl : preparatory work

This is a catch-up class that will help you getting use to Gretl and revising the notions seen in chapters 1 (Introduction) to 4 (Simple linear regression with Gretl). Before starting this catch-up class,

1. If Gretl is not installed, download and install it from http ://gretl.sourceforge.net/ ;
2. Make sure that you know the 4 first chapters ;
3. Read "A (very) short introduction to Gretl using scripts"
4. Read the statement of the the catchup lab and find out what are the equations, the statistical methods and tests that will be useful ;
5. find out what are the commands and functions that will be needed using the help command or menu.

This is a catch-up class that will help you getting use to Gretl (or R) and revising the notions seen in chapters 1 (Introduction) to 4 (Simple linear regression with Gretl). Before starting this catch-up class,

1. make sure that you know the 4 chapters before coming to the lab class ;
2. revise the functions and commands that will be needed :
   (a) What does each command and function ?
   (b) What is the syntax to use it ?
3. read the statement of the lab class and find out what are the equations that will be useful ;
4. explore the commands and functions that will be needed and are listed in the table below.

Gretl Programming		
Commands	Functions	
append	$coeff	abs
dataset	$df	cov
genr	$ess	delete
gnuplot	$ncoeff	diag
lags	$nobs	inv
ols	$rsq	log
open	$stderr	mean
outfile	$uhat	nobs
pvalue	$yhat	pvalue
print		rows
printf		sqrt
rename		sum
scatters		var
setobs		
smpl		
store		
summary	Types	
workdir	matrix	scalar

R Programming		
abline	length	read_excel
abs	lines	read.table
as.matrix	lm	save
as.vector	mean	setwd
attach	merge	sqrt
cbind	ncol	sum
coef	nrow	summary
data.frame	plot	t.test
colnames	plot.ts	var
diag	pt	view
	Packages	
ggplot2	lmtest	readxl
tseries	zoo	

Preparatory work - Analysis of the link between the Dow Jones and the EurostoXX50

Daily return data for the CAC40, the EuroStoXX50, the Dow Jones and the S&P500 are stored in an Excel file called "DataPrepWK.xlsx" or in two separated files called "EuroStoxx2019.csv" and "DowJones2019.csv". We are trying to figure out if the evolution of the EurostoXX50 index depends on the evolutions of the Dow Jones.

1. Look at the Excel file described above and check the data formats : What is name of the spreadsheet where the data are stored ? What is the line from which we are loading the data ? What is the type of data ? How many observations do we have in each file ? What are the beginning and ending dates ? What is the frequency of the observations ?

File	DowJones	EuroStoxx50
Spreadsheet		
Cell NB		
Data type		
Begin date		
End date		
Freq.		

2. Start a new script file. by writing some comments (#) to describe your script file. Propose some code to clear all the existing data and variables from Gretl memory (command **clear**). Declare a working directory (command **workdir**) "/Users/Yourname/.../myworkingdir" and save the script as "GretlCatchUp".
3. Load the Dow Jones series (command **open**) into Gretl (or R), define the date format (command **setobs**), rename the price as `PDJ` (command **rename**) and save the file (command **store**) as a Gretl or an R database, called "DailyDowJones2019".
4. Do the same for the Eurostoxx series.
5. From the smalest data base, append (command **append**) the data of the other data base. What happen if you do the reverse (from the largest data base, append the other data base) ?
6. Compute the return of both time series using "(-1)" after the name of the variable you want to lag by one observation or the command **lags**. Save the series into a Gretl (or R) database called "Indexes2019".
7. Compute some summary statistics of the returns (command **summary**).

8. Do a scatter plot of the two indexes. Give the evolution of the two indices over the entire sample period (commands **gnuplot** or **scatters**).
9. Convert the daily series into monthly series (command **dataset**), delete the missing observation (if any) (command **smpl**) and save the (new) monthly database. We call it "MIndexes".

R 9. Load the Excel File "MIndexes_for_R.xlsx".

10. Do the same Graph as in Figure 1 and save it as a pdf file format.

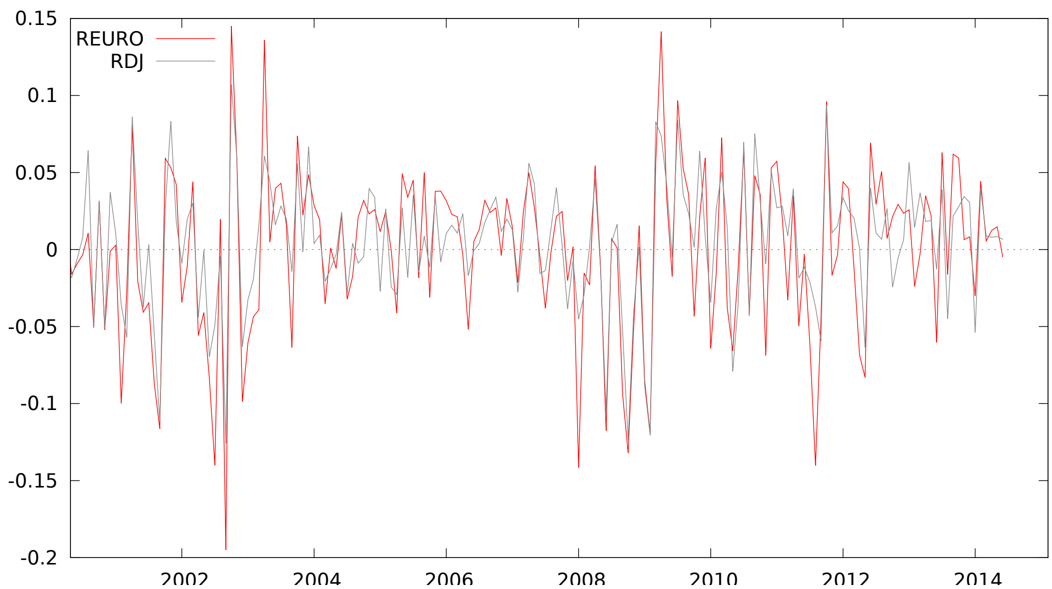


FIGURE 1 – Evolution of the Dow-Jones and the EurostoXX 50, between Jan. 2000 and Dec. 2014.

11. Compute the sum (function `sum`),of
    (a) the Dow Jones return and its square (use "^2"),
    (b) the EurostoXX50 return and its square,
    (c) the EurostoXX50-Dow Jones cross product,

    and print the results (command **print** or **printf**).
12. From the previous sums, compute the
    (a) empirical means, variances and covariance (control that you did good by using the functions `mean`, `var`, `cov`),
    (b) estimated parameters,
    (c) associated standard errors (using the function `sqrt`), and print the results.
13. Estimate the simple regression model (command **ols**), and check the values calculated in the previous question. Save the regression table into a file (command **outfile ... end outfile**) called "IndexRegResults.txt".
14. Test the hypothesis that the coefficient associated with the constant is zero.

(a) Compute the test statistics (using the `abs` function).

(b) Using the `pvalue` function or the **pvalue** command, get the pvalue of the test and interpret the result. Can we use the fonction $pvalue instead ? Do it if the answer is yes and explain otherwise.

(R b) Using the `pt` command, get the pvalue of the test and interpret the result. Can we use the `t.test` command ? Do it if the answer is yes and explain otherwise.

(c) Verify the results of the previous question in the simple regression table.

15. Do again questions 12 to 13 in matrix format and use the `matrix`, and `scalar` declarations as well as the functions `diag`, `inv`, `rows`, `sqrt`.

Lab session 1 : CAPM estimations and tests

This Lab class corresponds to chapters 1 (Introduction) to 4 (Simple linear regression with Gretl) + chapters 7 (heteroscedasticity and serial correlation) to end (OLS : adaptation and beyond) appendices. Before coming to the Lab class,

1. make sure that you know the corresponding chapters and appendices ;
2. revise the functions and commands that will be needed :
   (a) What does each command and function ?
   (b) What is the syntax to use it ?
3. read the statement of the lab class and find out what are the equations that will be useful ;
4. explore the "new" commands and functions that will be needed and are listed in the table below.

Gretl Programming	
Commands	Functions
corr	corr
loop	diff
loop ... foreach	log
setobs	mcorr
set	meanc
summary	ones
	sd
	sdc
	transp
Types	
list	series

R Programming	
cbind	legend
cor	log
cor.test	print
detach	rbind
dev.off	sapply
diff	setNames
for {...}	stat.desc
lapply	
Packages & Libraries	
ggplot2	pastecs
zoo	

The objective of this lab session is to determine whether the monthly returns of a particular stock can be explained by the variation of the market risk premium in line with the Capital Asset Pricing Model (CAPM) using the simple linear regression model. We will first estimate the CAPM model using Ordinary Least Squares (OLS). We will then test for the main OLS hypotheses and discuss the implications. We will also run some statistical inference tests. Finally, we will propose some corrections and/or some alternative estimators when needed.

Exercise 1 : Estimating the CAPM model using OLS

You have access to US monthly stock prices (in bps) for Ford (FORD), GE Aerospace (GE), Microsoft (MSFT), and Oracle (ORACLE), the S&P500 index (SANDP) and the 3-Month Treasury Bill Market Rate, from January, 2002 to February, 2018.

$$ER_{i,t} = r_f + \beta_i E(R_{M,t} - rf)$$

1. Look at the database in Excel before loading it into Gretl (Python or R). Keep the names of the variables as in the Excel file.
2. Do a quick research on these stocks, what are the characteristics of these companies/stocks ? What about the S&P500 index ?
3. Load the data you need into Gretl (Python or R).
4. Construct r_f the risk free rate.
5. Build the variables of the excess return to the risk free rate for the market and for the stocks.
6. Propose two interesting graphes of the excess return of the stock and the market.
7. Estimate the simple linear regression model that comes from the CAPM (Capital Asset Pricing Model).
   (a) What should we get for β_0 and β_1 ? Explain.
   (b) What do you think of the quality of the regression ?
   (c) Conclusion.

Exercise 2 : Tests in the simple linear model

You have access to US monthly prices for Ford (FORD), GE Aerospace (GE), Microsoft (MSFT), and Oracle (ORACLE), the S&P500 index (SANDP) and the 3-Month Treasury Bill Market Rate, from Jan, 2002 to February, 2018.

$$ER_{i,t} = r_f + \beta_i E(R_{M,t} - rf)$$

1. What are the necessary hypotheses of the classical linear regression model ?
2. In each case, propose a test in Gretl (Python or R).
3. What are the validated hypotheses ? What are the implications ?
4. How would you test in Gretl (Python or R) that the beta coefficient is 1 ? Can we do it ? Explain.
5. Conclusion.

Exercise 3 : Correcting the problems

In the previous exercise we did test for the OLS hypotheses.

1. We first explore how to deal with these problems, and in each case, you are asked to propose a program in Gretl (Python or R).
   (a) How would you correct for autocorrelation and/or heteroscedasticity problems ?
   (b) How would you correct for normality problems ?
   (c) How would you correct for parameter stability problems ?
   (d) How would you correct for the presence of outliers ?
   (e) How would you correct for the presence of functional form ?
2. We decide to see what we can learn from using the quantile regression technics on our CAPM model.
   (a) What is the difference between OLS and quantile regression in our case ?
   (b) For the ten deciles $(0.1, \cdots, 0.9)$, run the quantile regression associated with the CAPM.
   (c) Interpret the (new) results and compare to the OLS regression results.
3. Conclusion.

Lab session 2 : CAPM and APT models

This Lab class corresponds to chapters 1 (Introduction) to 10 (Other assumptions violation and diagnostic tests) + appendices. Before coming to the Lab class,

1. make sure that you know your course ;
2. revise the functions and commands that will be needed :
   (a) What does each command and function ?
   (b) What is the syntax to use it ?
3. read the statement of the lab class and find out what is needed and what are the equations that will be useful ;
4. explore the "new" commands and functions that will be needed and are listed in the table below.

Gretl Programming	
Commands	Functions
boxplot	$dw
chow	$dwpval
fcast	
modtest	
normtest	
omit	
qlrtest	
qqplot	
set	

R Programming	
AIC	
BIC	predict
bgtest	qqline
boxplot	qqnorm
bptest	residuals
dwtest	rstandard
fitted	rstudent
gqtest	shapiro.test
jarque.bera.test	white_test
ks.test	
Packages	
car	skedastic

The objective of this lab session is to determine whether the monthly returns of a particular stock can be explained by the variation of the market risk premium and some unexpected changes of a set of macroeconomics and financial variables. The file macro.xls contains monthly observations for the Microsoft stock price (MICROSOFT),the S&P500 Index value (SANDP), the Consumption Price Index (CPI), an Industrial Producer Price Index (INDPPI), a measure of money supply (MSUPPLY), a measure of credit consumption (CCREDIT) and one of 'credit spread (SCREDIT)', and Treasury Bills with maturities of 3 month (USTB3M) and 10 years (USTB10Y).
We consider two models the multiple linear regression model :

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \gamma_1 DI_t + \gamma_2 DC_t + \gamma_3 DP_t \\ &\quad + \gamma_4 DM_t + \gamma_5 DS_t + \gamma_6 DT_t + \varepsilon_t, \end{aligned}$$

where y and x are the stock and the S&P500 index excess returns. The unexpected changes in the economics and financial variables are DP for Producer Price Index, DI for the inflation, DT for the term structure of the interest rate (based on the Treasury Bills), DM for the money supply, DC for the credit consumption and DS for the credit spread.

Exercise 1 : Estimating the multiple regression models

The APT (Arbitrage Pricing Theory) suppose that the equity returns can be explained by some unexpected changes of macroeconomics variations rather than their levels. The unexpected variation can be specified as the difference between the observed (realized) value and the expected value of the variable. Suppose that investors have naive expectations, the next period value of the variable is equal to the current value. This means that investors are expecting no change in the variable value. Unexpected changes can then be calculated as the first difference of any variable we consider, and we have

$$DX_t = X_t - X_{t-1},$$

the unexpected changes of X.

1. Load the data
2. Construct the risk free rate
3. Build the variables of the excess return to the risk free rate for the market and for the stock.
4. Compute the INF the inflation variable as the first difference in the log Consummer Price Index.
5. TS the term structure of the interest rate as the slope of the Treasury Bills yield curve.
6. Compute DI, DC, DP, DM, DS and DT.
7. What should we get for the parameters if the model is correct ?
8. Estimate the multiple regression model.
9. Is this model better than the CAPM see in LAB1 ?

Exercise 2 : Diagnostic tests for the multiple regression models

Ordinary least square estimators are BLUE only under some hypotheses.

1. What are those hypotheses, and what are the consequences if not validated on the quality of the estimators, meaning on the "B", "L", "U", "E" ?
2. What should you look at to test for :
   (a) Autocorrelation
   (b) Heteroscedasticity
   (c) Normality
   (d) Multicolinearity
   (e) Functional form
   (f) After explaining what is an outlier and a boxplot, look for the command in either Gretl (R or Python) and propose a boxplot to explore the possibility of outliers.
3. Run all the test and fill-up the summary table below, add "*", "**" and "***" for each tests, and conlude on the OLS hypotheses rejection.

Tests Name	Homoscedasticity			Non autocorrelation	
CAPM					
APT					

Tests Name	Normality			Linearity	No outliers
CAPM					
APT					

Exercise 2 : Statistical inference in the APT model

1. We want to test that the coefficients associated with the 3 largest p-values are zero. How would you do that test ? Is it wise ?
2. Do the test yourself without using any pre-programmed procedure.
3. Propose a stepwise (backward) regression procedure to select the best exogenous set of variables.

Lab session 3 : Fama - MacBeth

This Lab class corresponds to chapters 1 (Introduction) to 10 (Other assumptions violation and diagnostic tests) + appendices. Before coming to the Lab class,

1. make sure that you know your course ;
2. revise the functions and commands that will be needed :
   (a) What does each command and function ?
   (b) What is the syntax to use it ?
3. read the statement of the lab class and find out what is needed and what are the equations that will be useful ;
4. explore the "new" commands and functions that will be needed and are listed in the table below.

Gretl Programming	
Commands	Functions
append	$aic
ar1	$bic
clear	$pvalue
chow	$rsq
corr	$test
gnuplot	$vcv
nulldata	cnameset
outfile	inv
printf	mcov
reset	mnormal
restrict	normal
loop	randgen
loop foreach	rnameset
scatters	zeros
setobs	
smpl	
summary	
Types	
array	matrix
scalar	strings

R Programming	
AIC	pt
anova	rbind
BIC	residuals
bgtest	rnorm
bptest	save
cbind	sd
dwtest	seq
fitted	ts.plot
linearHypothesis	set.seed
load	vector
names	
Packages	
car	skedastic
tseries	

In finance, risk factors are key investing elements, that help explain the systematic returns in equity market. In asset pricing theories such as the capital asset pricing model or the arbitrage pricing theory, the rate of return of an asset is linear combination of observed factors included in a linear asset pricing model (for example, the Fama–French three-factor model or the Carhart 4-factor model) proxy for a linear combination of unobserved (and priced) risk factors if financial market efficiency is assumed. These risk factors may be macroeconomic (such as consumer inflation, credit consumption, or money supply) or microeconomic (such as firm size or other accounting and financial metrics of the firms).
The Fama-MacBeth two-step regression proposes one way for measuring how these risk factors explain asset or portfolio returns. The aim of the model is to determine the risk premium associated with the exposure to these risk factors.

The Fama-MacBeth procedure is a simple two-step approach : The first step uses the exposures (characteristics) as explanatory variables in cross-sectional regressions. For example, if denote the excess returns of asset in month , then the famous Fama-French three-factor model implies the following return generating process (see also Campbell et al. 1998) :

$$R_i = \alpha + \lambda_M \beta_{i,M} + \lambda_S \beta_{i,S} + \lambda_V \beta_{i,V} + \lambda_U \beta_{i,U} + e_i \tag{1}$$

From this general model, we will simulate a no-problem model (when all the hypotheses are satisfied) that will serve as a benchmark. We will then estimate this model, test the OLS hypotheses and perform some statistical inference tests of the parameters. Second, we will simulate models that reject the classical OLS hypotheses : multicolinearity and autocorrelation. Third, we will estimate the model (by OLS), test the OLS hypotheses and perform statistical inference tests. Finally, we will compare our results.

Exercise 1 : Time-series regressions of returns - A Carhart 4-factor model

We run a set of time-series regressions to estimate the betas (exposures). We want to run the Carhart (1997) 4-factor model separately for each of the 25 portfolios. A Carhart 4-factor model regresses the portfolio returns on the excess market returns ('rmrf'), the size factor ('smb'), the value factor ('hml') and the momentum factor ('umd').
We do have access to two data bases : the first one is called `monthlyfactors.xlsx` and includes the time series of returns on all of the factors (the monthly return of the value-weighted index called 'rmrf', the monthly premium of the size factor called 'smb', the monthly premium of the book-to-market factor called 'hml', the monthly premium on winners minus loosers called 'umd'), the return on the market portfolio (rm) and the return on the risk-free asset (rf)). The second one is named `vw_sizebm_25groups.xlsx` and contains the time series of returns on 25 value-weighted portfolios formed from a large universe of stocks, two-way sorted according to their sizes and book-to-market ratios.

1. Load the data and create a joint data set. Save it as a Gretl (R or Python) dataset called DataLAB3.

2. Compute the excess return of the 25 portfolios.
3. Run the Carhart 4-factor model separately for each of the twenty-five portfolios.
4. Save the matrix of $\beta_{i,j}$ (exposures), $i = 1, \cdots, 25$ and $j =$ 'rmrf', 'smb','hml', and 'umd'.
5. Save the vector of Jensen's alpha estimates and test each of them to zero.
6. Conclusion.

Exercise 2 : Cross-sectional regressions

The second stage of the Fama-MacBeth procedure is to run a separate cross-sectional regression for each point in time.

$$R_i = \alpha + \lambda_M \beta_{i,M} + \lambda_S \beta_{i,S} + \lambda_V \beta_{i,V} + \lambda_U \beta_{i,U} + e_i$$

1. Compute the matrix of explanatory variables made of a constant plus exposures.
2. Compute the return variables (one per date) across the 25 portfolios.
3. Use a loop to construct the matrix of risk premiums $\lambda_{j,t}$.
4. Compute the averages and standard deviation of theses estimates

$$\hat{\lambda}_j = \frac{1}{T_{FMB}} \sum_{t=1}^{T_{FMB}} \hat{\lambda}_{t,j},\ j = 1, \cdots, 5$$

5. Test that these risk premiums are different from zero.
6. Conclusion.

The three golden rules of econometrics are "test, test, and test"
Robert Hendry

Doing applied work involves a synthesis of various elements. You must be clear about why you are doing it [...]. You must understand the characteristics of the data you are using and appreciate their weaknesses. You must use theory to provide a model of the process that may have generated the data. You must know the statistical methods, which rely on probability theory, to summarise the data, e.g. in estimates. You must be able to use the software [...]. You must be able to interpret the statistics or estimates in terms of your original purpose and the theory.
Ron Smith