

Introduction

- Definition

- Basics in statistics and mathematics

- The data

- Limits, critics & model construction

The classical linear regression model

- Presentation of the model

- Assumptions

- Properties of the OLS estimator

- Precision and standard errors

- Goodness of fit

Introduction to statistical inference

- The idea

- Distribution of the estimated parameters

- Significativity test

- Confidence interval approach

- The level of significance : choosing α

- The exact level of significance : the p-value

Simple linear regression with Gretl or R

- Presentation

- Choosing a software

- Getting started with Gretl

- Financial application

The multiple regression model

- Matrix form of the model

- OLS estimators

- Estimation of the variance of the errors

Statistical inference in the multiple regression model

- Distribution of the estimated parameters

- Testing individually the estimated parameters

- Testing simultaneously the estimated parameters

- Financial application : The APT model

- Constructing factors and excess returns

- Regression model

- Heteroskedasticity and serial correlation

- The generalised regression model

- The GLS estimators

- Heteroskedasticity of the errors

- Dealing with Heteroskedasticity

- Autocorrelation of the errors

- Dealing with Autocorrelation

- Financial application : The APT model (Ctd)

- Other assumptions violation and diagnostic tests

- Stochastic regressors and exogeneity

- Normality of the errors

- Multicollinearity

- Model selection and diagnostic tests

- Selection criteria

- Alternative to OLS

- Two stage least squares

- Maximum likelihood estimation

- Generalized Least Squares

- Quantile regression

- Appendix & References

Simple linear regression with Gretl or R

R is a package for econometric, statistical and data analysis, written in the C programming language. It is a free software.

- Very powerful software environment for statistical computing and graphics.
- A wide variety of estimators : least squares, maximum likelihood, GMM, single-equation and system methods.
- Time series methods : ARMA, GARCH, VARs and VECMs, unit-root and cointegration tests, etc.
- Data science methods
- This software can be downloaded at <https://cran.r-project.org/mirrors.html> and run under Windows, Mac/OS or Linux.
- User's guides are accessible at <https://cran.r-project.org/manuals.html>. First free book to read <https://cran.r-project.org/doc/manuals/R-intro.pdf>.

Gretl is a package for econometric analysis, written in the C programming language. It is a free, open-source software.

- Easy intuitive interface but can also use scripts.
- A wide variety of estimators : least squares, maximum likelihood, GMM, single-equation and system methods.
- Time series methods : ARMA, GARCH, VARs and VECMs, unit-root and cointegration tests, etc.
- This software can be downloaded at <http://gretl.sourceforge.net/> and run under Windows, Mac/OS or Linux.
- A user's guide is accessible at <http://www.learneconometrics.com/gretl.html>

What should you choose ?

- R is amongst the most popular languages for econometric, and statistical analysis. Not Gretl.¹
- But Gretl has an easy intuitive interface (still proposing a quite sophisticated econometric package). You do not need to practice much to use Gretl contrary to R.
- R is definitely a more comprehensive package that will require some investment and some independence toward its uses.
- Gretl or R are more than enough for this course and produce nice statistical Tables and Graphs that can easily be used in Word or LaTeX.

Be sure that you are able to use one of them before coming to the labclass. The Labclass will be conducted in Gretl but some R corrections will also be proposed.

1. Python is more popular for Data Science.

Figure : Data importation under gretl

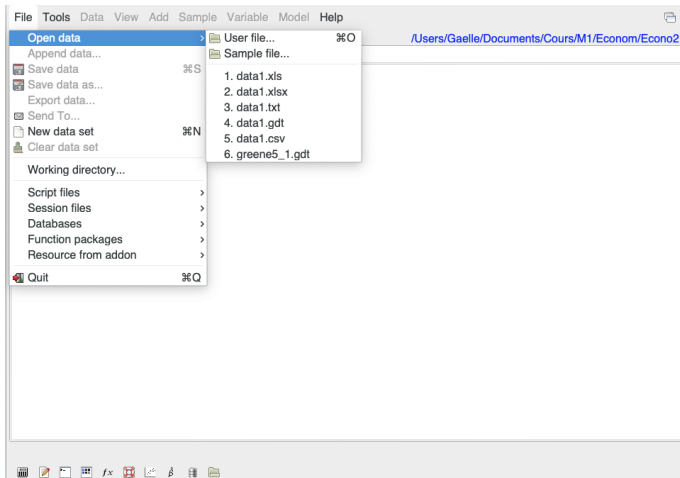


Figure : Opening data

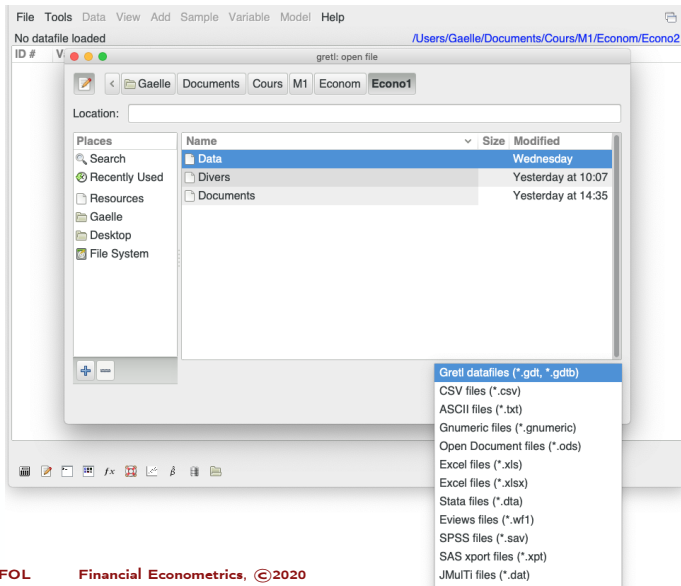
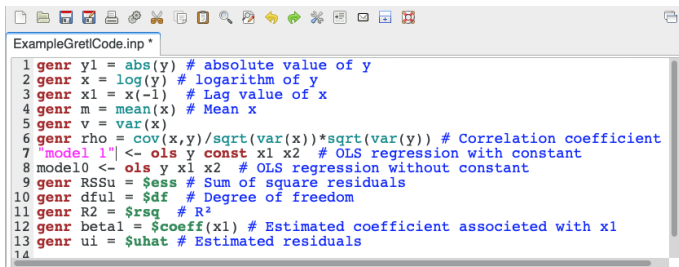


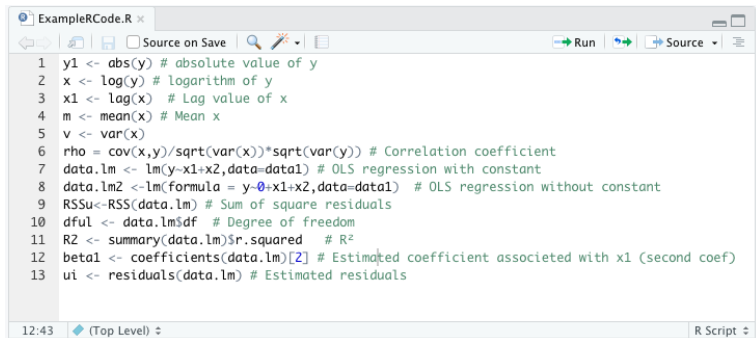
Figure : Using a script



The screenshot shows the Gretl software interface with a script editor window titled 'ExampleGretlCode.inp *'. The script contains 14 lines of code, each starting with the 'genr' command. The code calculates various statistical measures and performs OLS regressions. The variables and their descriptions are as follows:

```
1 genr y1 = abs(y) # absolute value of y
2 genr x = log(y) # logarithm of y
3 genr x1 = x(-1) # Lag value of x
4 genr m = mean(x) # Mean x
5 genr v = var(x)
6 genr rho = cov(x,y)/sqrt(var(x))*sqrt(var(y)) # Correlation coefficient
7 "model 1" <- ols y const x1 x2 # OLS regression with constant
8 model0 <- ols y x1 x2 # OLS regression without constant
9 genr RSSu = $ess # Sum of square residuals
10 genr dfu = $df # Degree of freedom
11 genr R2 = $rsq # R^2
12 genr betal = $coeff(x1) # Estimated coefficient associated with x1
13 genr ui = $uhat # Estimated residuals
14
```

Figure : equivalent script example



```
1 y1 <- abs(y) # absolute value of y
2 x <- log(y) # logarithm of y
3 x1 <- lag(x) # Lag value of x
4 m <- mean(x) # Mean x
5 v <- var(x)
6 rho = cov(x,y)/sqrt(var(x))*sqrt(var(y)) # Correlation coefficient
7 data.lm <- lm(y~x1+x2,data=data1) # OLS regression with constant
8 data.lm2 <- lm(formula = y~0+x1+x2,data=data1) # OLS regression without constant
9 RSSu<-RSS(data.lm) # Sum of square residuals
10 dfu1 <- data.lm$df # Degree of freedom
11 R2 <- summary(data.lm)$r.squared # R2
12 beta1 <- coefficients(data.lm)[2] # Estimated coefficient associated with x1 (second coef)
13 ui <- residuals(data.lm) # Estimated residuals
```

12:43 (Top Level) R Script

Figure : Command references

Index		Gretl Command Reference					
▢ Tests		add	adf	anova	append	ar	ari
▢ Statistics		arbond	arch	arima	boxplot	break	catch
▢ Dataset		chow	coeffsum	coint	coint2	corr	corrgm
▢ Estimation		cusum	data	dataset	debug	delete	diff
▢ Graphs		difftest	discrete	dummify	duration	elif	else
▢ Programming		end	endif	endloop	eqnprint	equation	estimate
▢ Transformations		fcast	foreign	freq	function	garch	genr
▢ Printing		gmm	gnuplot	graphpg	hausman	heckit	help
▢ Prediction		hsk	hurst	if	include	info	intreg
▢ Utilities		kalman	kpss	labels	lad	lags	ldiff
		leverage	logistic	logit	logs	loop	mahal
		meantest	mle	modeltab	modprint	modtest	mpols
		negbin	nls	normtest	nulldata	ols	omit
		open	orthdev	outfile	panel	pca	pergm
		poisson	print	printf	probit	pvalue	qlrtest
		qqplot	quantreg	quit	rename	reset	restrict
		rmplot	run	runs	scatters	sdiff	set
		setinfo	setobs	setmiss	shell	smpl	spearman
		sprintf	square	sscanf	store	summary	system

Figure : Function references

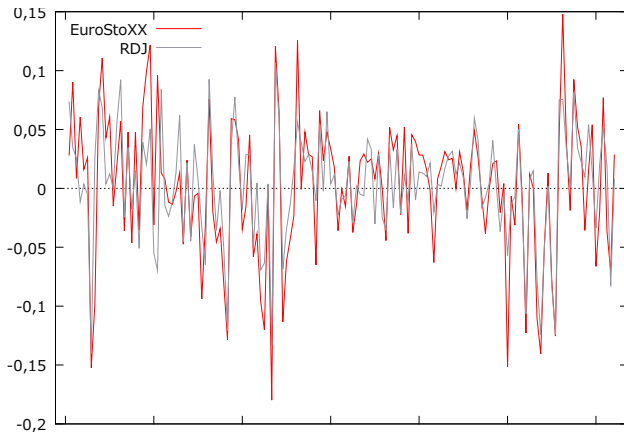
Index	Gretl Function Reference				
⊕ Accessors					
⊕ Mathematical					
⊕ Strings					
⊕ Numerical methods					
⊕ Filters					
⊕ Statistical					
⊕ Probability					
⊕ Linear algebra					
⊕ Matrix building					
⊕ Matrix shaping					
⊕ Transformations					
⊕ Data utilities					
	\$ahat	\$aic	\$bic	\$chisq	\$coeff
	\$companion	\$datatype	\$df	\$dwpval	\$ec
	\$error	\$ess	\$fstat	\$gmmcrit	\$h
	\$hausman	\$hq	\$jalpha	\$jbeta	\$jvbeta
	\$lnl	\$ncoeff	\$nobs	\$nvars	\$pd
	\$pvalue	\$rho	\$rsq	\$sample	\$sargan
	\$sigma	\$stderr	\$stopwatch	\$sysA	\$sysB
	\$sysGamma	\$T	\$t1	\$t2	\$test
	\$trsq	\$uhat	\$unit	\$vcv	\$version
	\$windows	\$xlist	\$xtxinv	\$yhat	
	abs	acos	acosh	argname	asin
	asinh	atan	atanh	bessel	BFGSmax
	bkfilt	boxcox	cdemean	cdf	cdiv
	ceil	cholesky	cmult	cnorm	colnames
	cols	corr	corrgram	cos	cosh
	cov	critical	cum	deseas	det
	diag	diff	digamma	dnorm	dsort
	dummify	eigegen	epochday	exp	fcstats
	fdjac	fft	fft1	filter	firstobs
	floor	fracdiff	gammafun	getenv	gini
	gamma	bkfilt	T	gamma	gamma

Does the Dow Jones Industrial index returns have an impact on the EurostoXX50 index return ? I

- Monthly return data are saved in worksheet Feuil3 of the files DowJones.xls and EuroStoXX50.xls.
 1. Import the first file under Gretl, define the date format and save it as a Gretl's data base.
 2. Import the second file under Gretl, define the date format and save it as a Gretl's data base.
 3. From the second file, append the data of the first file.
 4. Plot the two indexes evolutions.
 5. Make the scatter plot of REURO against RDJ.

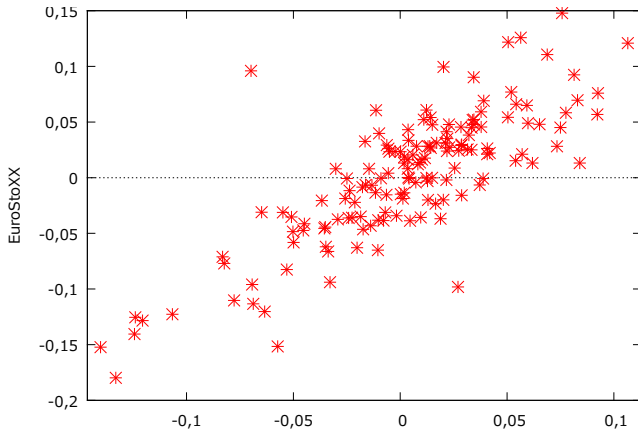
Does the Dow Jones Industrial index returns have an impact on the EurostoXX50 index return ? II

Figure : Dow Jones and EuroStoxx50 returns evolutions



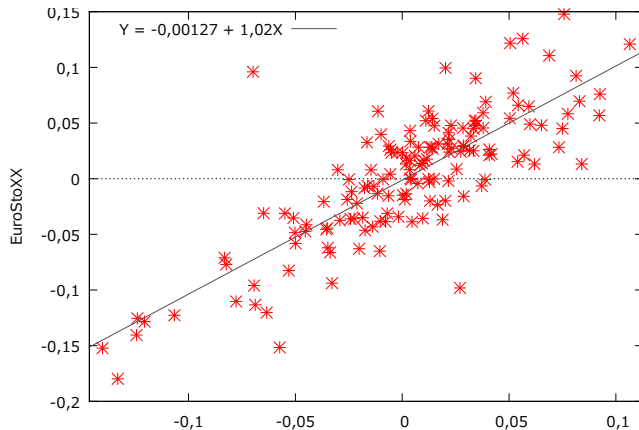
Does the Dow Jones Industrial index returns have an impact on the EurostoXX50 index return ? III

Figure : EuroStoxx50 against Dow Jones scatter plot



Does the Dow Jones Industrial index returns have an impact on the EurostoXX50 index return? IV

Figure : EuroStoxx50 against Dow Jones scatter plot



Does the Dow Jones Industrial index returns have an impact on the EurostoXX50 index return? V

- Construct the Dow Jones squared returns (RDJ2) and the cross returns Dow Jones - Eurostoxx (RDJREURO) :

```
genr RDJ2 = RDJ * RDJ
```

```
genr RDJREURO = RDJ * REURO
```

- Calculate the sum of these variables and print them

```
genr sumREURO = sum(REURO)
```

```
...
```

```
print sumRDJ sumREURO sumRDJ2 sumRDJREURO
```

Does the Dow Jones Industrial index returns have an impact on the EurostoXX50 index return? VI

Summary	<i>RDJ</i>	<i>REURO</i>	$(RDJ)^2$	$RDJ \times REURO$
Sum	0,402352	0,223013	0,319466	0,326798

We can calculate, the estimated parameters from this summary : [► Equations](#)

$$\hat{\beta}_0 =$$

$$\hat{\beta}_1 =$$

Does the Dow Jones Industrial index returns have an impact on the EurostoXX50 index return? VII

Model 1 : OLS, using observations 1998 :02–2010 :06 ($T = 149$)
Dependent variable : REURO

	Coefficient	Std. Error	t-ratio	p-value
const	−0.00126992	0.00284956	−0.4457	0.6565
RDJ	1.02455	0.0615401	16.6485	0.0000
Mean dependent var	0.001497	S.D. dependent var	0.058786	
Sum squared resid	0.177247	S.E. of regression	0.034724	
R^2	0.653443	Adjusted R^2	0.651085	
$F(1, 147)$	277.1725	P-value(F)	1.21e−35	
Log-likelihood	290.2729	Akaike criterion	−576.5458	
Schwarz criterion	−570.5379	Hannan–Quinn	−574.1049	
$\hat{\rho}$	−0.021802	Durbin–Watson	2.029049	