STATISTICAL METHODS WITH APPLICATION TO FINANCE

a.y. 2022-2023

Introduction

R. PAPPADÀ
rpappada@units.it

Department of Economics, Business, Mathematics and Statistics "B. de Finetti"
University of Trieste

February 27, 2023

Summary

General Info

The core of this course is the use of basic statistical techniques and tools to deal with data analysis in the framework of economics and finance, with emphasis on financial risk.
It will consists of

- frontal lectures ($\approx$ 35h)
- R practical sessions ($\approx$ 10h)

Note that
$\rightarrow$ Lecture recordings will be available to the course team in MS Teams

$\rightarrow$ Any changes that occur in the timetable will be communicated via Moodle

Lecturer: Roberta Pappadà

Office: room 2.18, 2nd floor, Via Valerio 4/1
E-mail: rpappada@units.it

Lessons timetable & office hours

Academic year: 2022/2023, II semester

Course: STATISTICAL METHODS WITH APPLICATION TO FINANCE [R. PAPPADA']

	Monday	Tuesday	Wednesday	Thursday	Friday
12:00	STATISTICAL METHODS WITH APPLICATION TO FINANCE PAPPADA' ROBERTA Aula 3_B [Edificio D - Econ.] 12:00 - 14:00				
13:00					
14:00		STATISTICAL METHODS WITH APPLICATION TO FINANCE PAPPADA' ROBERTA Aula 3_B [Edificio D - Econ.] 14:00 - 16:00	STAT. METHODS WITH APPL. FIN. PAPPADA' ROBERTA Aula 3_B [Edificio D - Econ.] 14:00-15:00		
15:00					
16:00					

Office hours:

Tuesday, 11:00 - 12:30
Thursday, 16:00 - 17:00
Any change will be published on www.deams.units.it

Office hours can also be scheduled through teams by appointment.

Course calendar

→ 2nd Semester's lessons: from February 27 to May 26, 2023.

→ Easter Holidays: from April 7 to April 11, 2023.

Wednesday classes will not be held every week; in March there will be no lessons on the following dates

- 1st March 2023
- 15th March 2023
- 29th March 2023

Course calendar

→ 2nd Semester's lessons: from February 27 to May 26, 2023.

→ Easter Holidays: from April 7 to April 11, 2023.

Wednesday classes will not be held every week; in March there will be no lessons on the following dates

1st March 2023

15th March 2023

29th March 2023

Prerequisites

The *Statistics* course is a prerequisite for Financial Econometrics and Statistical Methods with Applications to Finance: it is assumed that the student is at least somewhat familiar with the basics of probability and statistics: exploratory data analysis, random variables and distribution functions, and the fundamentals of statistical inference.

No prior knowledge of computer programming is required for attending the `R` Lab sessions.

Course Objectives

The course aims to provide students with an introduction to well-established *statistical methods and models* in the financial and economic context, with a focus on time series data.

The `R` statistical software will be used in hands-on sessions for data manipulation, exploratory data analysis, time series modelling and forecasting.

We start by looking at stock prices and asset returns, which are then used for discussing more advanced topics such as time series models and forecasting techniques, thus bringing together finance, statistics and computer programming.

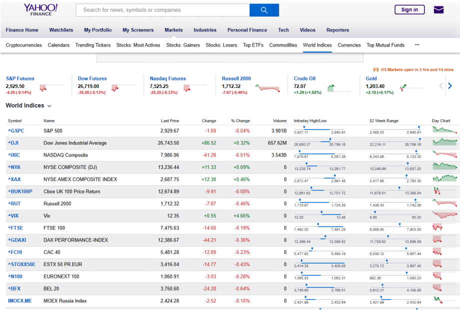
Summary

Introduction

- Statistics plays a key role in finance; examples are the empirical analysis of financial data, the random walk hypothesis, portfolio theory, risk management, etc.
- Recently, statistics has becoming crucial to deal with big-data and complex data structures
- Economic and financial data are collected systematically for thousands of variables in many countries and the literature of statistical models is huge

Financial data

In many of the problems of interest in finance, the starting point is the analysis of *price series*.



YAHOO! FINANCE

Search for news, symbols or companies

Sign in

Finance Home | Watchlists | My Portfolio | My Screeners | Markets | Industries | Personal Finance | Tech | Videos | Reporters

Cryptocurrencies | Calendars | Trending Tickers | Stocks: Most Actives | Stocks: Gainers | Stocks: Losers | Top ETFs | Commodities | World Indices | Currencies | Top Mutual Funds | ···

US Markets open in 3 hrs and 14 mins

S&P Futures 2,929.50 -4.25 (-0.14%) | Dow Futures 26,719.00 -35.00 (-0.13%) | Nasdaq Futures 7,525.25 -25.25 (-0.33%) | Russell 2000 1,712.32 -7.87 (-0.46%) | Crude Oil 72.07 +1.29 (+1.82%) | Gold 1,203.40 +2.10 (+0.17%)

World Indices

Symbol	Name	Last Price	Change	% Change	Volume	Intraday High/Low	52 Week Range	Day Chart
^GSPC	S&P 500	2,929.67	-1.08	-0.04%	3.901B	2,927.11 – 2,940.91	2,488.03 – 2,940.91	
^DJI	Dow Jones Industrial Average	26,743.50	+86.52	+0.32%	657.62M	26,680.37 – 26,769.16	22,219.11 – 26,769.16	
^IXIC	NASDAQ Composite	7,986.96	-41.28	-0.51%	3.543B	7,979.67 – 8,057.26	6,343.96 – 8,133.30	
^NYA	NYSE COMPOSITE (DJ)	13,236.44	+11.33	+0.09%	0	13,228.74 – 13,261.77	12,048.66 – 13,637.02	
^XAX	NYSE AMEX COMPOSITE INDEX	2,687.75	+12.30	+0.46%	0	2,672.47 – 2,691.48	2,417.98 – 2,790.38	
^BUK100P	Cboe UK 100 Price Return	12,674.89	-9.91	-0.08%	0	12,661.63 – 12,701.72	11,679.51 – 13,398.84	
^RUT	Russell 2000	1,712.32	-7.87	-0.46%	0	1,710.97 – 1,724.38	1,436.43 – 1,742.09	
^VIX	Vix	12.35	+0.55	+4.66%	0	12.20 – 12.48	8.56 – 50.30	
^FTSE	FTSE 100	7,475.63	-14.60	-0.19%	0	7,462.00 – 7,491.25	6,866.90 – 7,903.50	
^GDAXI	DAX PERFORMANCE-INDEX	12,386.67	-44.21	-0.36%	0	12,366.44 – 12,399.52	11,726.62 – 13,596.89	
^FCHI	CAC 40	5,481.28	-12.89	-0.23%	0	5,477.62 – 5,488.18	5,038.12 – 5,657.44	
^STOXX50E	ESTX 50 PR.EUR	3,416.04	-14.77	-0.43%	0	3,414.38 – 3,426.85	3,278.72 – 3,697.40	
^N100	EURONEXT 100	1,060.91	-3.03	-0.28%	0	1,060.53 – 1,063.31	982.36 – 1,090.20	
^BFX	BEL 20	3,750.60	-24.28	-0.64%	0	3,745.95 – 3,766.01	3,612.31 – 4,188.95	
IMOEX.ME	MOEX Russia Index	2,424.28	-2.52	-0.10%	0	2,421.99 – 2,432.64	2,421.99 – 2,432.64	

Some sources:

- Google
- Yahoo finance
- R packages
- etc.

Returns

Returns on an asset, e.g., a stock, a bond, a portfolio of stocks and bonds, are changes in price expressed as a fraction of the initial investment.

P_{t_1}: Price of asset at time t_1
P_{t_2}: Price of asset at time t_2

The net return for the holding period from t_1 to t_2 is

$$R = \frac{P_{t_2} - P_{t_1}}{P_{t_1}}$$

The return on a portfolio is a weighted average of the returns on the individual assets

Financial markets and risk

Finance is concerned with buying and selling assets, pricing them, and assessing the involved risk.
There are two main ingredients: *time* and *uncertainty*

- A stock market index shows how a specified portfolio of share prices changes over time
- Risk is related to the uncertainty in the return that we obtain from investing in stock shares

Financial markets and risk/ 2

We may model issues related to time and uncertainty within a mathematical framework, using tools from probability and statistics.

- Future returns cannot be known exactly and therefore are **random variables**
- Volatility, or the standard deviation of returns, is a common measure of risk. However, the use of volatility as a risk measure can lead to misleading conclusions!

Non-normality

How do returns on assets behave?

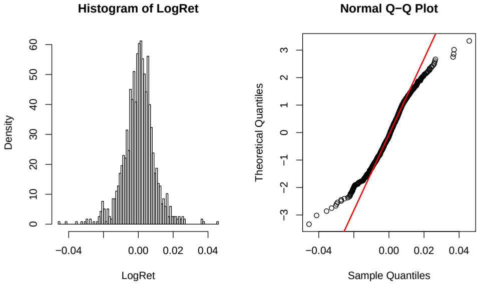


Figure 1: Histogram and QQ plot of daily log returns of the McDonald's Corporation over the period January 2010 to September 2014.

Financial markets data are not normally distributed!

For modelling financial markets data, **heavy-tailed distributions** such as the Student t-distributions are much more suitable than normal distributions

Financial data have **stylised facts** that are different from other types of data

Time series

Much of the data from financial markets are time series, that is, sequences of data sampled over time

Most analysis of financial time series is done in the **time domain**, and we will restrict our attention to this.

Data visualization: a *time series plot* is the graph displaying time series observations in chronological order. The consecutive observations are commonly regarded as equally spaced, for simplicity.

Time series/ 2

We concentrate on various types of statistical models widely used in econometrics, business forecasting, and many scientific applications.

The best-known examples of time series, which behave like *random walks*, are share prices on successive days. A model, which often gives a good approximation to such data, is

share price on day t = share price on day $(t-1)$ + random error

Simple linear models may be appropriate for many financial series, but the world is usually more complex!

Example 1: Apple stock

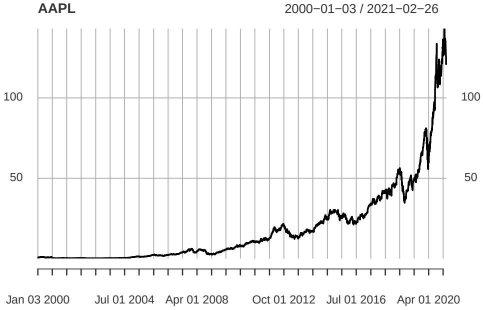


Figure 2: Closing prices of Apple stock (2000–2021)

Example 2: GNP Data

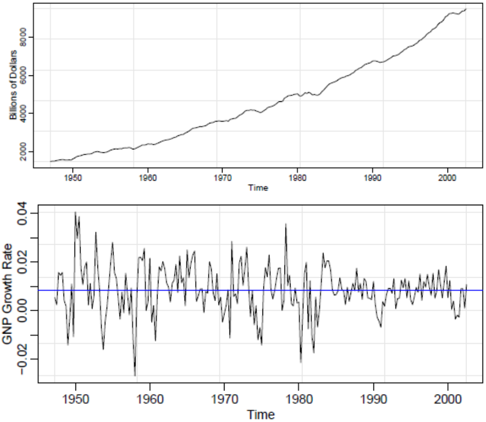


Figure 3: Quarterly U.S. Gross National Product in billions (GNP) from 1947 to 2002 and U.S. GNP quarterly growth rate.

Example 3: Johnson & Johnson quarterly earnings

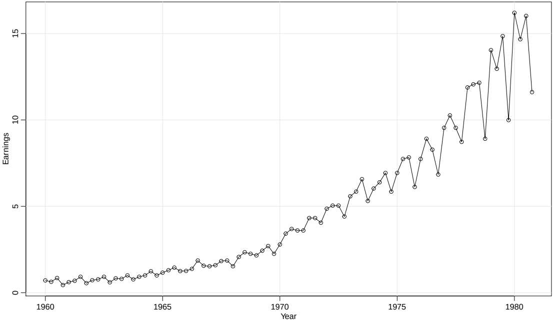


Figure 4: Johnson & Johnson quarterly earnings per share, from the first quarter of 1960 to the last quarter of 1980.

Example 4: Dow Jones Industrial Average

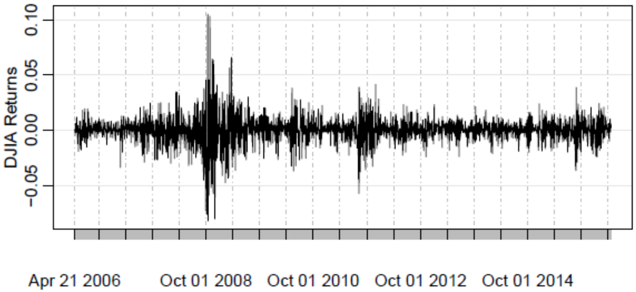


Figure 5: Daily returns of the Dow Jones Industrial Average (DJIA) from 2006 to 2016.

Risk management

Econometrica, Vol. 50, No. 4 (July, 1982)

AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY WITH ESTIMATES OF THE VARIANCE OF UNITED KINGDOM INFLATION[1]

BY ROBERT F. ENGLE

Traditional econometric models assume a constant one-period forecast variance. To generalize this implausible assumption, a new class of stochastic processes called autoregressive conditional heteroscedastic (ARCH) processes are introduced in this paper. These are mean zero, serially uncorrelated processes with nonconstant variances conditional on the past, but constant unconditional variances. For such processes, the recent past gives information about the one-period forecast variance.

A regression model is then introduced with disturbances following an ARCH process. Maximum likelihood estimators are described and a simple scoring iteration formulated. Ordinary least squares maintains its optimality properties in this set-up, but maximum likelihood is more efficient. The relative efficiency is calculated and can be infinite. To test whether the disturbances follow an ARCH process, the Lagrange multiplier procedure is employed. The test is based simply on the autocorrelation of the squared OLS residuals.

This model is used to estimate the means and variances of inflation in the U.K. The ARCH effect is found to be significant and the estimated variances increase substantially during the chaotic seventies.

1. INTRODUCTION

IF A RANDOM VARIABLE y_t is drawn from the conditional density function $f(y_t \mid y_{t-1})$, the forecast of today's value based upon the past information, under standard assumptions, is simply $E(y_t \mid y_{t-1})$, which depends upon the value of the conditioning variable y_{t-1}. The variance of this one-period forecast is given by $V(y_t \mid y_{t-1})$. Such an expression recognizes that the conditional forecast variance depends upon past information and may therefore be a random variable. For conventional econometric models, however, the conditional variance does not depend upon y_{t-1}. This paper will propose a class of models where the variance does depend upon the past and will argue for their usefulness in economics. Estimation methods, tests for the presence of such models, and an empirical example will be presented.

Consider initially the first-order autoregression

$$y_t = \gamma y_{t-1} + \epsilon_t$$

where ϵ is white noise with $V(\epsilon) = \sigma^2$. The conditional mean of y_t is γy_{t-1} while the unconditional mean is zero. Clearly, the vast improvement in forecasts due to time-series models stems from the use of the conditional mean. The conditional

[1] This paper was written while the author was visiting the London School of Economics. He benefited greatly from many stimulating conversations with David Hendry and helpful suggestions by Denis Sargan and Andrew Harvey. Special thanks are due Frank Srba who carried out the computations. Further insightful comments are due to Clive Granger, Tom Rothenberg, Edmond Malinvaud, Jean-Francois Richard, Wayne Fuller, and two anonymous referees. The research was supported by NSF SOC 78-09476 and The International Centre for Economics and Related Disciplines. All errors remain the author's responsibility.

987

Nobel Prize winner Engle (1982) has determined a recent revolution in finance developing new models (ARCH Models) extended by Bollerslev (1986), that successfully capture changes in volatility

DJIA - Volatility Forecasting

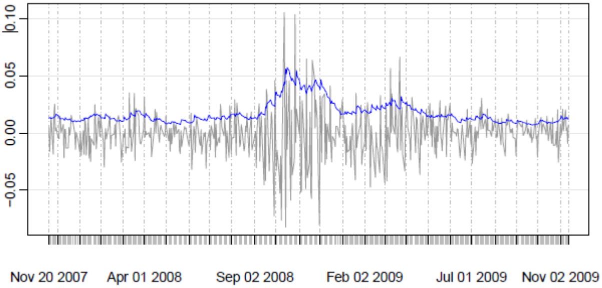


Figure 6: GARCH one-step-ahead predictions of the DJIA volatility (solid line) superimposed on part of the DJIA series including the financial crisis of 2008.

Summary

Course Content

Review of basic concepts on financial markets, prices and risk.

- Returns; log returns; multi-period returns; portfolios.
- Distributional properties of returns; review of statistical distributions and their moments; visualization of financial data. The random walk hypothesis.
- Index numbers; Laspeyres Price Index and Paasche Price Index; Stock Market Indices.

  R Lab Exploratory data analysis and Stylized Facts of returns

Linear models for financial time series

- Characteristics of time series data; white noise and linear time series; autoregressive and moving average models; model selection techniques
- Unit-root nonstationarity; Integrated ARMA (or ARIMA) models; introduction to seasonal models.

  R Lab Fitting ARMA models to time series data

Course Content / 2

Forecasting

- ☐ Point forecasts and prediction intervals; forecasting using ARMA models
- ☐ exponential smoothing

  R Lab Prediction of time series using linear models

GARCH and conditional volatility

- ☐ Testing for ARCH effects; the ARCH model; GARCH models and their extensions.

  R Lab Fitting ARMA-GARCH models to financial returns

Risk measures

- ☐ How to measure financial risk? Value at risk and expected shortfall. Risk aggregation.

  R Lab Estimation of risk measures and simulation.

Course Activities

The activities include

- Lessons with slides illustrating the theoretical background of the statistical methods presented in the course, as well as examples on simulated and real-world datasets
- Sessions of exercises with full solutions
- *R* Labs to focus on statistical models for financial returns, primarily for the purpose of forecasting risk
- Wocclap or Moodle Quizzes for self-evaluation (homework or live quizzes)

More on quizzes and live tests

- During the course, moodle quizzes will be scheduled, which include true/false questions, multiple choice, or problems with numeric solution
- Students wishing to participate in the Moodle quizzes will have a deadline to submit the answers (few days)
- There will some interactive quizzes during the lessons, which may serve for self-evaluation
- Students that obtain at least the 60% of correct answers in all tests can receive up to 2 additional points in the final grade (only for the summer session)

Summary

Textbooks and other resources

Slides, exercises, R source files and datasets will be available on the Moodle page of the course.

Online resources for review of statistical concepts and `R` language

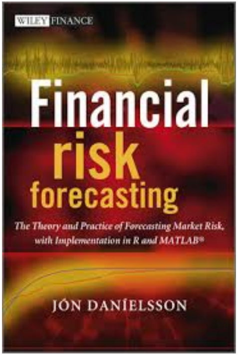
- D. M. Diez, C. D. Barr, M. Çetinkaya-Rundel, *OpenIntro Statistics* `https://leanpub.com/openintro-statistics` (get the full book PDF)
- An introduction to `R` available at `https://cran.r-project.org/`

Textbooks and other resources

Main textbooks:

J. Danielsson (2011)

Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab, Wiley

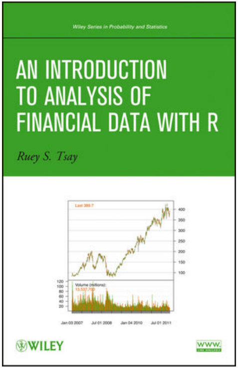


Textbooks and other resources

Main textbooks:

Tsay, Ruey S. (2013)

An Introduction to the Analysis of Financial Data with R, Wiley

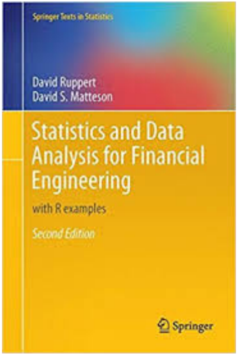


Textbooks and other resources

Further textbooks (demanding)

D. Ruppert, D.S. Matteson (2015)

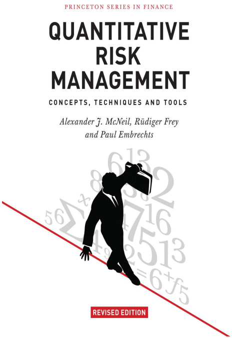
Statistics and Data Analysis for Financial Engineering with R examples
Second edition, Springer Texts in Statistics



Textbooks and other resources

Further textbooks (demanding)

J. McNeil, R. Frey, and P. Embrechts (2015) *Quantitative Risk Management: Concepts, Techniques and Tools*, Revised Edition, Princeton Series in Finance



Summary

Final exam

The final exam consists of a written test of **2 hours** with

- exercises and open questions to assess your comprehension level of all the topics covered during the course
- questions concerning the output of various statistical analysis to assess your ability to interpret results, plots, and evaluate different models in applied contexts
- Examples of past exams will be made available

Remember to register for the exam within the given deadline (**four days** prior to the exam)

Questions?