Financial Econometrics

Financial Econometrics

Module Introduction and Overview

Contents

1 Introduction to the Module

Welcome to the module on Financial Econometrics. The first objective of this module is to introduce the main econometric methods and techniques used in the analysis of issues related to finance. A module with the title *Financial Econometrics* assumes that such a field exists. However, as this quote reveals, this is far from true:

> What is … financial econometrics? This simple question does not have a simple answer. The boundary of such an interdisciplinary area is always moot and any attempt to give a formal definition is unlikely to be successful. Broadly speaking, financial econometrics [aims] to study quantitative problems arising from finance. It uses statistical techniques and economic theory to address a variety of problems from finance. These include building financial models, estimation and inferences of financial models, volatility estimation, risk management, testing financial economics theory, capital asset pricing, derivative pricing, portfolio allocation, risk-adjusted returns, simulating financial systems, hedging strategies, among others (Fan, 2004: 1).

In this module we define financial econometrics as 'the application of statistical techniques to problems in finance'. Although econometrics is often associated with analysing economics problems such as economic growth, consumption and investment, the applications in the areas of finance have grown rapidly in the last few decades. Your textbook by Chris Brooks, *Introductory Econometrics for Finance*, lists the following examples:

1. Testing whether financial markets are weak-form informationally efficient.
2. Testing whether the capital asset pricing model or arbitrage pricing theory represent superior models for the determination of returns on risky assets.
3. Measuring and forecasting the volatility of bond returns.
4. Explaining the determinants of bond credit ratings used by the ratings agencies.
5. Modelling long-term relationships between prices and exchange rates.
6. Determining the optimal hedge ratio for a spot position in oil.
7. Testing technical trading rules to determine which makes the most money.
8. Testing the hypothesis that earnings or dividend announcements have no effect on stock prices.
9. Testing whether spot or futures markets react more rapidly to news.
10. Forecasting the correlation between the returns to the stock indices of two countries.

The above list does not include all the possibilities, and you might think of many other topics that could be added to the list.

If financial econometrics is simply the application of econometrics to finance issues, does this mean that econometric tools you have studied in previous courses are the same as those used in this module? A simple answer to this question is *yes*. Many of the concepts that you have encountered in your other studies such as regression and hypothesis testing are highly relevant for this module. In fact, all the topics introduced in this module will require that you have a deep understanding of these concepts. However, the emphasis and the set of problems dealt with in finance issues are different from the economic problems you have encountered in previous courses. To start with, the nature of the data in finance issues is very different. Financial data are observed at a much higher frequency (in some instances minute-by-minute frequency). For macroeconomic data, we consider ourselves lucky if we are able to observe data on a monthly basis. Furthermore, recorded financial data such as stock market prices are those at which the transaction took place. There is no possibility for measurement error. This is in contrast to macroeconomic data, which are revised regularly.

Also the properties of financial series differ. For instance, the module *Econometric Analysis and Applications* analyses whether a series has a unit root, and devises methods to estimate models when the variables are integrated of order one. In financial econometrics these issues are not a major concern. Although we observe prices most of the time, financial econometrics mainly deals with asset or portfolio returns. Since returns are stationary, most of the methods used in this module also apply to stationary series.

This may imply that models of financial returns are much easier to deal with. However, this is not the case. The analysis of financial data brings its own challenges. As you will see in Unit 1, financial returns possess some common properties that need to be incorporated in econometric models. For instance, returns of assets such as stocks and bonds exhibit time-varying volatility. This requires introducing new models and estimation techniques to model time varying volatility. Not only that, financial returns can exhibit asymmetry in volatility, which requires further modification of existing models. Furthermore, financial data are not normally distributed. As you have seen in your previous studies, the assumption of normality has been central for estimation and hypothesis testing. Unfortunately, even in finance applications, existing econometric techniques still find it difficult to deal with models that assume non-normal distribution.

2 The Module Authors

Bassam Fattouh graduated in Economics from the American University of Beirut in 1995. Following this, he obtained his Masters degree and PhD from the School of Oriental and African Studies, University of London, in 1999. He is a Professor in Finance and Management and Academic Director for the MSc in International Management for the Middle East and North Africa at the Department for Financial and Management Studies, SOAS. He is also currently Senior Research Fellow and Director of the Oil and Middle East

Programme at the Oxford Institute for Energy Studies at the University of Oxford. He has published in leading economic journals, including the *Journal of Development Economics, Economics Letters, Economic Inquiry, Macroeconomic Dynamics* and *Empirical Economics*. His research interests are mainly in the areas of finance and growth, capital structure and applied non-linear econometric modelling, as well as oil pricing systems.

Jonathan Simms is a tutor for CeFiMS, and has taught at University of Manchester, University of Durham and University of London. Dr Simms has contributed to development of various CeFiMS modules including *Econometric Principles & Data Analysis*; *Econometric Analysis & Applications*; *Risk Management: Principles & Applications*; *Financial Engineering*; *Introduction to Valuation*; *Introduction to Law and to Finance*; *Public Financial Management: Reporting and Audit*; *Corporate and Investment Banking*; and *Banking Strategy*.

3 Study Materials

This module mainly uses one textbook:

Chris Brooks (2019) *Introductory Econometrics for Finance*. 4th Edition. Cambridge UK: Cambridge University Press.

This textbook has been chosen because it is extremely clear, it contains a large number of examples and covers a lot of different topics. It is a useful textbook to refresh your memory of some basic concepts studied in previous courses (especially Chapters 2, 3 and 4). The textbook has a very useful companion website with rich resources for students including practice questions, solutions to end of chapter questions, data in Excel files, and R codes. The link for the companion resources can be found at:

https://www.cambridge.org/gb/academic/subjects/economics/finance/introductory-econometrics-finance-4th-edition?format=PB

Although the textbook covers a lot of subject areas, in some units you may need to rely more heavily on the Study Guide and suggested readings.

The units in the module will closely follow the presentation in the textbook. However, for some of the units, this is not feasible either because the chapter does not cover the topic at all, or covers it in a superficial way. In such cases you may find that the Study Guide is more demanding than the material presented in the textbook, because the Study Guide analyses the issues using mathematics (though at a relatively basic level). This is necessary to gain a deeper understanding of the issues being considered.

Throughout this module it is essential that you do all the readings and solve all the exercises. In this module each idea builds on the previous ones in a logical fashion, and it is important that each idea is clear to you before you move on. You should therefore take special care not to fall behind with your schedule of studies.

🖥 Software

R

This module will use R. This is a widely used programming environment for data analysis and graphics. You will use this software to do the exercises in the units. The results presented in the units are also from R.

R is free software, released under the GNU General Public License (R Core Team, 2019). Instructions for downloading R, and a brief introduction on how to use it, are provided below.

The best advice on using R is to stay focused on the subject that is being studied in each unit, and to do the exercises for the unit; this will reinforce your understanding and also develop your confidence in using data and R.

The units include all of the R commands that are required to complete the examples and exercises. However, as you become more confident in using R, or if you are already familiar with R, you may find that you develop your own R commands.

4   Module Overview

Unit 1   Statistical Properties of Financial Returns

1.1   Introduction
1.2   Calculation of Asset Returns
1.3   Stylised Facts about Financial Returns
1.4   Distribution of Asset Returns
1.5   Time Dependency
1.6   Linear Dependency across Asset Returns

Unit 2   Matrix Algebra, Regression and Applications in Finance

2.1   Introduction
2.2   Matrix Algebra: Some Basic Concepts and Applications
2.3   OLS Regression Using Matrix Algebra
2.4   Applications to Finance

Unit 3   Maximum Likelihood Estimation

3.1   Introduction
3.2   The Maximum Likelihood Function: Some Basic Ideas and Examples
3.3   The Maximum Likelihood Method: Mathematical Derivation
3.4   The Information Matrix
3.5   Usefulness and Limitations of the Maximum Likelihood Estimator
3.6   Hypothesis Testing

Unit 4   Univariate Time Series and Applications to Finance

4.1   Introduction
4.2   The Lag Operator
4.3   Some Key Concepts

Unit 5   Modelling Volatility – Conditional Heteroscedastic Models

Unit 6   Modelling Volatility and Correlations – Multivariate GARCH Models

Unit 7   Vector Autoregressive Models

Unit 8   Limited Dependent Variable Models

The objective of the module is to extend your knowledge and equip you with methods and techniques that allow you to analyse finance-related issues.

This module starts by illustrating how to measure financial returns, the main variable that we try to model in financial applications. There are various definitions of returns, and Unit 1 illustrates how to compute the various types of returns. After defining financial returns, Unit 1 presents some stylised facts about the properties of financial returns. These include volatility clustering, asymmetric volatility and non-normality. Unit 1 then introduces various measures of moments of the distribution of financial returns, and how these can be computed for samples of financial returns. The material covered in this unit sets the scene for the rest of the module and thus it is important that you make yourself familiar with these concepts.

Unit 2 provides a brief introduction to the main principles of matrix algebra. The modules *Econometric Principles and Data Analysis* and *Econometric Analysis and Applications* develop the basic regression concepts and statistical tools without referring to matrix algebra. This is essential to develop understanding of the basic concepts involved in regression analysis. However, in most theoretical and practical applications the researcher often deals with multivariate relations. As you will discover in this unit, the simplest way to tackle these multivariate relations is to switch to matrix notation. Matrix algebra eliminates the need to use summation signs and subscripts and helps present the results in a simple way. In some of the units of this module, it will be very difficult to present the proofs and results without using matrix notation. Although matrix notation simplifies the presentation of the results, the fact remains that you may be learning a new language. Learning a new language can be exciting but it is also challenging. To help you to understand and apply matrix algebra, we use matrix algebra in some financial applications, namely the multi-factor models and portfolio theory.

Unit 3 provides a brief review of the maximum likelihood estimation method. In your previous study of econometrics, it is probable that the least squares (LS) method was used to derive the estimates of the model's parameters and for hypothesis testing. Least squares is just one of many estimation techniques available for econometricians. In Units 4, 5, 6 and 8 of this module, you will encounter models such as GARCH, ARMA and binary choice models that can't be estimated by least squares. Instead, econometricians rely on maximum likelihood estimation, which is a flexible technique, more general than OLS and, under fairly general conditions, yields consistent and efficient estimates of the parameters. However, like any estimation technique, maximum likelihood is based on a particular underlying philosophy and principles. In Unit 3 you will be introduced to these principles and how these can be applied to derive estimates of the parameters and test hypotheses about the estimated parameters. This is one of the most challenging units, but hopefully, by using various examples, you will be able to gain a deep understanding of how the estimation method works, and you will be able to identify its strengths and weaknesses.

Unit 4 presents univariate time series models. In these types of models, a series is modelled in terms of its own past values and some disturbance terms

(also known as shocks). Univariate time series models were introduced in the module *Econometric Analysis and Applications*. These models are different from the structural models you have studied in other courses, in the sense that these models are *atheoretical* – that is, they are not based on any underlying theoretical frameworks but are data driven. These models are the first building blocks for estimating financial returns and help illustrate some of the key properties of financial returns. The aim of this unit is to introduce these models, such as the autoregressive model (AR), the moving average model (MA), and a combination of these two (ARMA models).

Unit 5 presents some of the econometric methods used for modelling and forecasting volatility of asset returns. Volatility models have attracted the attention of academics and practitioners, and are widely used in many areas of finance, including models of value-at-risk, option pricing, and portfolio allocation. One of the stylised facts about asset returns is that the variance of the error terms is not equal at every point in time, and hence the error terms are said to suffer from heteroscedasticity. Thus, in modelling financial returns, one should consider approaches that relax the assumption of homoscedasticity. ARCH (autoregressive conditional heteroscedastic) and GARCH (generalised autoregressive conditional heteroscedastic) models do exactly that. They relax the assumption of constant variance and exploit the heteroscedasticity feature to model the variance of returns over time. As you will study in this unit, GARCH models are also flexible enough to allow us to incorporate asymmetry in the volatility of financial asset returns.

Unit 6 extends the GARCH model from the univariate to the multivariate setting. This proves to be an important extension because it allows researchers and financial analysts to model time-varying conditional covariance and correlation between the returns of various financial assets. This technique opens the way for many financial applications such as dynamic asset pricing models, portfolio selection, dynamic hedging, value-at-risk, and volatility transmission between assets and markets. Multivariate GARCH models also help researchers to model some of the features of asset returns, such as correlation clustering.

Unit 7 presents vector autoregressive (VAR) models, which can be thought of as generalisations to the univariate time series models. VAR models represent an improvement over univariate time series models because they allow variables to be modelled not only in terms of their own lags and their own shocks, but also in terms of the lags of other variables. This provides greater flexibility and allows us to examine the dynamic interactions between a set of variables. VAR models have become very popular in the econometrics literature and are widely used in the areas of macroeconomics and finance. The tools which have developed around VAR, such as impulse response analysis, Granger causality and variance decompositions (all discussed in this unit) have become central to understanding the interaction among variables. VAR models have also been used extensively for forecasting purposes, where these models have exhibited a better performance than structural models, especially in out-of-sample forecasting.

Unit 8 deals with models in which the dependent variable *ie* the variable that needs to be explained by a set of determinants, is in fact a dummy variable. There are many cases where these models can be useful. For instance, financial analysts may be interested as to why some firms list on the stock market while others don't; why some firms issue dividends while others don't; and why some firms decide to raise external finance while others don't. In all these examples, what we observe is whether a firm lists or not, issues dividends or not, or raises external finance or not. Thus, the relevant dependent variable is a dummy variable that takes the value of 1 if the event occurs, and zero if the event does not occur. Such models, known as limited dependent variable models, raise a set of estimation issues that are different from the ones you have encountered so far. The purpose of this unit is to introduce you to limited dependent models and discuss how these models can be applied to finance issues.

5 Learning Outcomes

After you complete your study of this module, you will be able to:

- define and compute measures of financial returns
- discuss the stylised statistical properties of asset returns
- formulate models using matrix notation
- derive the OLS estimators using matrix algebra
- explain the principles of maximum likelihood estimation
- derive the maximum likelihood estimators and discuss their properties
- use maximum likelihood estimation, and apply the hypothesis tests available under maximum likelihood estimation
- analyse, estimate and forecast using autoregressive (AR), moving average (MA), and autoregressive-moving average (ARMA) models
- apply the Box–Jenkins approach to time series models
- model and forecast volatility using autoregressive conditional heteroscedastic (ARCH) models
- estimate, interpret and forecast with generalised autoregressive conditional heteroscedastic (GARCH) models
- construct, estimate and interpret multivariate GARCH models
- test for spill-over of volatility between assets
- use vector autoregressive (VAR) models to analyse and interpret interaction between financial variables, including impulse response analysis
- undertake tests of hypotheses and Granger causality in a VAR framework
- formulate, estimate and interpret limited dependent variable models, including logit and probit models
- discuss models with multinomial linear dependent variables.

6 R

R is an implementation of the object-oriented mathematical programming language S. It is developed by statisticians around the world and is free software, released under the GNU General Public License. Syntactically and functionally it is very similar (if not identical) to S+, the popular statistics package.

R is much more flexible than most software used by econometricians because it is a modern mathematical programming language, not just a program that does regressions and tests. The S language is the de facto standard for statistical science. Since most users have a statistical background, the jargon used by R experts sometimes differs from what an econometrician (especially a beginning econometrician) may expect. Code written for R can be run on many computational platforms with or without a graphical user interface, and R comes standard with some of the most flexible and powerful graphics routines available anywhere. And of course, R is completely free for any use.

(extracted from Farnsworth, 2008)

Because the R software is a programming language and not just an econometrics program, some of the functions we will be interested in are available through libraries (sometimes called packages) obtained from the R website http://www.r-project.org/.

The data files required are available on the Virtual Learning Environment (VLE).

Starting up

The R software must be installed on your system. If it is not, follow the installation instructions appropriate to the operating system (OS). Installation is especially straightforward for Windows users.

To install R

Go to the website of R http://www.r-project.org/.

There, choose your preferred CRAN mirror (*eg* http://cran.ma.imperial.ac.uk/) and click on the link referring to your OS in the box "Download and Install R". Note that these units are written with the version "Windows", but Mac OS and Linux versions are also downloadable. (CRAN mirrors are provided in locations around the world to reduce download time and internet traffic.)

On the next page, click on "base" and you will be redirected to a page where there will be a link to download the latest version of R. At the top will be written something like "Download R 3.6.0 for Windows" (or whatever is the most up-to-date version). Click on this link. Updated versions of R are made available over time, but the commands explained in the module files should be similar.

Save the file "R-3.6.0-win.exe" (or whatever is the latest version of the file) in any location, for example on your Desktop. Double click on this file to start the installation, and follow the instructions. You will be asked to choose the

place where you want the file to be, and you can also choose to have a shortcut to R in the Start Menu and a Desktop icon. Choose the standard installation.

To get started with R

To open R, double click on the desktop icon or on the shortcut. The Command Window opens.

To use R, you can either type a command in the Command Window, or run a pre-written program.

1. To type a command in the command window

    You just type it on the line starting with **>** and press the key Enter on your keyboard. Note that R is case sensitive. In the unit file, if it says "type in" or "can be performed with", this means "Type in the following command in the command window and then press Enter".

    The unit files include all of the R commands required to work on the examples and exercises. To save time, you can copy and paste the required coding from the unit files into R. Note that if multiple lines of code are pasted into R in one go, they will be executed in sequence immediately, without the need to press Enter.

2. Every time you see the sign **>**, it means that R is ready for a new command. When you do not see it or when you see the sign **+**, it means that R is still working on a previous command or is stuck because your command has not been properly written (for example, a bracket may be missing at the end of the previous command). In this case, press the red button STOP in the toolbar to stop the process.

3. To run a pre-written program

    In the Menu bar go to File, then choose "New script". A new Editor Script Window opens.

    You can type all commands you wish in this file, one command per line. To run them on R, select with the mouse the commands you wish to run, and then simultaneously press the keys CTRL and the key R (CTRL + R) (or alternatively go to Edit in the Menu bar and choose "Run line or selection").

    Save your program by choosing "Save as" in File in the Menu bar (the Menu bar specific to the Editor, that is to say, the one that is available when you are in the Editor Window, and not in the Command Window). All script files have the extension ".R". Then you can close it and open it from the File menu in the Menu bar.

    In your program file you can write things that are not commands (for example, notes or explanations); but to indicate to R that these are not commands, put the symbol # in front of each non-command line.

To get help at any time, go to Help in the Menu bar. Alternatively, there are several methods of obtaining help in R. You may type in alternatively the following commands:

```
?qt

help(qt)

help.start()

help.search("covariance")
```

The last command is for a search on the term covariance, for example. Please note the use of the quotation marks " ". If you choose to develop R commands in Word, for example, and then copy the commands to R, please make sure you use the " " symbols, and not the Word symbols " ".

Preceding a command with a question mark, or writing a command as an argument in help(), gives a description of its usage and functionality. The help.start() function brings up a menu of help options, and help.search() searches the help files for the word or phrase given as an argument. Many times, though, the best help available can be found by a search online. The help tools above only search through the R functions that belong to packages already installed on your computer. However, often users have the following type of question: "Does R have a function to do …". Users do not know if functionality exists because the corresponding package is not installed on their computer, and therefore resort to the R website. To search the R website for functions and references, use:

```
RSiteSearch("scatter plot")
```

This command example searches the R website for 'scatter plot' functions.

The results from the search should appear in your web browser.

By default, R has a couple of excellent free manuals in PDF format. If you are not already familiar with using R you are advised to read *An Introduction to R*. To access the manual, click Help | Manuals and the list of available documents will be shown.

To quit R, in the Menu bar go to File, then choose "Exit". You are asked "Save workspace image"? Click on No.

In case R is installed in another language, to change the language of the Menu bar into English, in the Menu bar choose Edit (second from the left in the Menu bar) and then choose the last option "Preferences". A new window opens. In the top right of it, there is an empty box for "Language for menus and messages". Type in this box "English" and click OK. You will have to exit R and start it again.

You may want to specify a working directory for your R session, so that R chooses data files (and scripts) directly from this directory; then there is no need to specify the full path to access the files. The command getwd() returns an absolute filename representing the current working directory of the R process. The command setwd() helps you specify the working directory. For example, to set the working directory to "C:\Documents and Settings\my files\R SOAS", you would type in:

```
setwd("C:/Documents and Settings/my files/R SOAS")
```

This avoids the need to put the command scripts and the data files within the R directory on your computer. **Note that, in this command, the slash / is used**.

Alternatively, you can set the working directory using File (on the Menu bar), then Change dir… and then browse through the folders on your device to select the folder you want to use as the working directory for that session.

To save your work, go to File | Save Workspace… Provide a filename for the workspace, and it will be saved with the extension .RData. The file will be saved in your current working directory, but you can browse to another location if you prefer.

A saved workspace can be loaded with File | Load Workspace… R will display files with the .Rdata extension in the current working directory, but again, you can browse to other locations

Every time you start R, before running any command or program, it is best to clean up any object that may remain in memory.

During an R session, objects are created and stored by name. The command objects(), or the command ls(), can be used to display the names of the objects which are currently stored within R. The command rm() removes a data object whose name should be given as an argument.

To remove all objects, type in:

```
rm(list=ls())
```

Alternatively, use Misc | Remove all objects.

Removing all objects is useful if, for example, you work on one exercise and dataset, then you want to work on another data set and exercise in the same R session. If you do not remove the first set of objects before starting on the second question, your workspace will contain objects you created for both questions, which could be confusing. Remember to save a workspace before removing all the objects!

Installing and loading packages

The basic setup for R provides considerable statistical and mathematical functionality.

However, you will also need to use some of the more specialised economet-ric functions that have been written for R, and made available in packages. The module units will indicate when you need to use a particular function, and the unit will also indicate which package is required.

To use a function from a package, the relevant package must be installed on your machine. To install a package go to Packages | Install package(s)… You will be asked to select a CRAN mirror for this session. Once the mirror has been selected, you can choose from the list of available packages to install. Some of the packages (for example dynlm) take a few minutes to install, but installation of a package is a one-off procedure, and once the package is installed, you will not need to install it again.

To use a package in an R session, *the package must be loaded in that session*. Loading a package takes considerably less time than installing it. To load a package, go to Packages | Load package… You will see a list of packages that you have previously installed, and you can select the package you want to load.

To summarise, to access a package it must first be installed on your device. And to use the package in an R session, the package must be loaded in that session.

If you try to use a function but you have not loaded the relevant package in your R session, you will get an error message saying R could not find the function.

Reading data from text files and creating zoo objects

The data for the examples and exercises in the module units is provided in tab-delimited text files. R can read data from files in other formats, but tab-delimited text files are very simple, they can be created and opened in many applications, and they are robust to updates.

To work on the data in R, it must first be read from the text file into what is called a data frame, using a read.table command. A data frame is a matrix or table that contains the data in R. Once the data is in a data frame, it is then possible to use R to perform various operations, such as plotting the data, and estimating models.

However, many models in finance specifically focus on the time element of the observations. Furthermore, financial data can have irregular dates. For example, if you are using daily data, there may be no observations for weekends and bank holidays. For these reasons we have mainly used 'zoo' objects to handle the data in R, rather than data tables or regularly dated time series objects. The zoo package provides an infrastructure for regular and irregular time series (Z's Ordered Observations – hence 'zoo'), and is maintained by Achim Zeileis (Zeileis and Grothendieck, 2005). The zoo objects are indexed by the dates of the observations, which allows us to perform operations that specifically relate to the timing of the data.

The commands for reading the data from the text files and creating the zoo objects are provided in the units. These commands assume that the data text files are located in the current working directory.

R outputs

The outputs that you obtain can be viewed in R, and can also be copied to word processing applications. The tables of outputs in the module units were obtained in R and formatted for production. Some of the formatted tables of output were produced with the stargazer package (Hlavac, 2018). Advice on how to use stargazer is also provided in the units, where relevant. However, please note that it is not possible to use stargazer on the outputs of all of the packages used in this module.

7 Assessment

Your performance on each module is assessed through two written assignments and one examination. The assignments are written after Unit 4 and Unit 8 of the module session. Please see the VLE for submission deadlines. The examination is taken at a local examination centre in September/October.

Preparing for assignments and exams

There is good advice on preparing for assignments and exams and writing them in Chapter 8 of *Studying at a Distance* by Christine Talbot. We recommend that you follow this advice.

The examinations you will sit are designed to evaluate your knowledge and skills in the subjects you have studied: they are not designed to trick you. If you have studied the module thoroughly, you will pass the exam.

Understanding assessment questions

Examination and assignment questions are set to test your knowledge and skills. Sometimes a question will contain more than one part, each part testing a different aspect of your skills and knowledge. You need to spot the key words to know what is being asked of you. Here we categorise the types of things that are asked for in assignments and exams, and the words used. All the examples are from the Centre for Financial and Management Studies examination papers and assignment questions.

Definitions

Some questions mainly require you to show that you have learned some concepts, by setting out their precise meanings. Such questions are likely to be preliminary and be supplemented by more analytical questions. Generally, 'Pass marks' are awarded if the answer only contains definitions. They will contain words such as:

- Describe
- Define
- Examine
- Distinguish between
- Compare
- Contrast
- Write notes on
- Outline
- What is meant by
- List

Reasoning

Other questions are designed to test your reasoning, by explaining cause and effect. Convincing explanations generally carry additional marks to basic definitions. They will include words such as:

- Interpret
- Explain
- What conditions influence
- What are the consequences of
- What are the implications of

Judgement

Others ask you to make a judgement, perhaps of a policy or of a course of action. They will include words like:

- Evaluate
- Critically examine
- Assess
- Do you agree that
- To what extent does

Calculation

Sometimes, you are asked to make a calculation, using a specified technique, where the question begins:

- Use indifference curve analysis to
- Using any economic model you know
- Calculate the standard deviation
- Test whether

It is most likely that questions that ask you to make a calculation will also ask for an application of the result, or an interpretation.

Advice

Other questions ask you to provide advice in a particular situation. This applies to law questions and to policy papers where advice is asked in relation to a policy problem. Your advice should be based on relevant law, principles and evidence of what actions are likely to be effective. The questions may begin:

- Advise
- Provide advice on
- Explain how you would advise

Critique

In many cases the question will include the word 'critically'. This means that you are expected to look at the question from at least two points of view, offering a critique of each view and your judgement. You are expected to be critical of what you have read.

The questions may begin:

- Critically analyse
- Critically consider
- Critically assess
- Critically discuss the argument that

Examine by argument

Questions that begin with 'discuss' are similar – they ask you to examine by argument, to debate and give reasons for and against a variety of options, for example

- Discuss the advantages and disadvantages of
- Discuss this statement
- Discuss the view that
- Discuss the arguments and debates concerning

The grading scheme: Assignments

The assignment questions contain fairly detailed guidance about what is required. All assignments are marked using marking guidelines. When you receive your grade it is accompanied by comments on your paper, including advice about how you might improve, and any clarifications about matters you may not have understood. These comments are designed to help you master the subject and to improve your skills as you progress through your programme.

Postgraduate assignment marking criteria

The marking criteria for your programme draws upon these minimum core criteria, which are applicable to the assessment of all assignments:

- understanding of the subject
- utilisation of proper academic [or other] style (e.g. citation of references, or use of proper legal style for court reports, etc.)
- relevance of material selected and of the arguments proposed
- planning and organisation
- logical coherence
- critical evaluation
- comprehensiveness of research
- evidence of synthesis
- innovation/creativity/originality.

The language used must be of a sufficient standard to permit assessment of these.

The guidelines below reflect the standards of work expected at postgraduate level. All assessed work is marked by your Tutor or a member of academic staff, and a sample is then moderated by another member of academic staff. Any assignment may be made available to the external examiner(s).

80+ (Distinction). A mark of 80+ will fulfil the following criteria:

- very significant ability to plan, organise and execute independently a research project or coursework assignment
- very significant ability to evaluate literature and theory critically and make informed judgements
- very high levels of creativity, originality and independence of thought
- very significant ability to evaluate critically existing methodologies and suggest new approaches to current research or professional practice
- very significant ability to analyse data critically
- outstanding levels of accuracy, technical competence, organisation, expression.

70–79 (Distinction). A mark in the range 70–79 will fulfil the following criteria:

- significant ability to plan, organise and execute independently a research project or coursework assignment

- clear evidence of wide and relevant reading, referencing and an engagement with the conceptual issues
- capacity to develop a sophisticated and intelligent argument
- rigorous use and a sophisticated understanding of relevant source materials, balancing appropriately between factual detail and key theoretical issues. Materials are evaluated directly and their assumptions and arguments challenged and/or appraised
- correct referencing
- significant ability to analyse data critically
- original thinking and a willingness to take risks.

60–69 (Merit). A mark in the 60–69 range will fulfil the following criteria:

- ability to plan, organise and execute independently a research project or coursework assignment
- strong evidence of critical insight and thinking
- a detailed understanding of the major factual and/or theoretical issues and directly engages with the relevant literature on the topic
- clear evidence of planning and appropriate choice of sources and methodology with correct referencing
- ability to analyse data critically
- capacity to develop a focussed and clear argument and articulate clearly and convincingly a sustained train of logical thought.

50–59 (Pass). A mark in the range 50–59 will fulfil the following criteria:

- ability to plan, organise and execute a research project or coursework assignment
- a reasonable understanding of the major factual and/or theoretical issues involved
- evidence of some knowledge of the literature with correct referencing
- ability to analyse data
- examples of a clear train of thought or argument
- the text is introduced and concludes appropriately.

40–49 (Fail). A Fail will be awarded in cases in which there is:

- limited ability to plan, organise and execute a research project or coursework assignment
- some awareness and understanding of the literature and of factual or theoretical issues, but with little development
- limited ability to analyse data
- incomplete referencing
- limited ability to present a clear and coherent argument.

20–39 (Fail). A Fail will be awarded in cases in which there is:

- very limited ability to plan, organise and execute a research project or coursework assignment

- failure to develop a coherent argument that relates to the research project or assignment
- no engagement with the relevant literature or demonstrable knowledge of the key issues
- incomplete referencing
- clear conceptual or factual errors or misunderstandings
- only fragmentary evidence of critical thought or data analysis.

0–19 (Fail). A Fail will be awarded in cases in which there is:

- no demonstrable ability to plan, organise and execute a research project or coursework assignment
- little or no knowledge or understanding related to the research project or assignment
- little or no knowledge of the relevant literature
- major errors in referencing
- no evidence of critical thought or data analysis
- incoherent argument.

The grading scheme: Examinations

The written examinations are 'unseen' (you will only see the paper in the exam centre) and written by hand, over a three-hour period. We advise that you practise writing exams in these conditions as part of your examination preparation, as it is not something you would normally do.

You are not allowed to take in books or notes to the exam room. This means that you need to revise thoroughly in preparation for each exam. This is especially important if you have completed the module in the early part of the year, or in a previous year.

Details of the general definitions of what is expected in order to obtain a particular grade are shown below. These guidelines take account of the fact that examination conditions are less conducive to polished work than the conditions in which you write your assignments. Note that as the criteria of each grade rises, it accumulates the elements of the grade below. Assignments awarded better marks will therefore have become comprehensive in both their depth of core skills and advanced skills.

Postgraduate unseen written examinations marking criteria

80+ (Distinction). A mark of 80+ will fulfil the following criteria:

- very significant ability to evaluate literature and theory critically and make informed judgements
- very high levels of creativity, originality and independence of thought
- outstanding levels of accuracy, technical competence, organisation, expression
- outstanding ability of synthesis under exam pressure.

70–79 (Distinction). A mark in the 70–79 range will fulfil the following criteria:

- clear evidence of wide and relevant reading and an engagement with the conceptual issues
- develops a sophisticated and intelligent argument
- rigorous use and a sophisticated understanding of relevant source materials, balancing appropriately between factual detail and key theoretical issues
- direct evaluation of materials and their assumptions and arguments challenged and/or appraised;
- original thinking and a willingness to take risks
- significant ability of synthesis under exam pressure.

60–69 (Merit). A mark in the 60–69 range will fulfil the following criteria:

- strong evidence of critical insight and critical thinking
- a detailed understanding of the major factual and/or theoretical issues and directly engages with the relevant literature on the topic
- develops a focussed and clear argument and articulates clearly and convincingly a sustained train of logical thought
- clear evidence of planning and appropriate choice of sources and methodology, and ability of synthesis under exam pressure.

50–59 (Pass). A mark in the 50–59 range will fulfil the following criteria:

- a reasonable understanding of the major factual and/or theoretical issues involved
- evidence of planning and selection from appropriate sources
- some demonstrable knowledge of the literature
- the text shows, in places, examples of a clear train of thought or argument
- the text is introduced and concludes appropriately.

40–49 (Fail). A Fail will be awarded in cases in which:

- there is some awareness and understanding of the factual or theoretical issues, but with little development
- misunderstandings are evident
- there is some evidence of planning, although irrelevant/unrelated material or arguments are included.

20–39 (Fail). A Fail will be awarded in cases which:

- fail to answer the question or to develop an argument that relates to the question set
- do not engage with the relevant literature or demonstrate a knowledge of the key issues
- contain clear conceptual or factual errors or misunderstandings.

0–19 (Fail). A Fail will be awarded in cases which:

- show no knowledge or understanding related to the question set

- show no evidence of critical thought or analysis
- contain short answers and incoherent argument.

[2015–16: Learning & Teaching Quality Committee]

Specimen exam papers

CeFiMS does not provide past papers or model answers to papers. Modules are continuously updated, and past papers will not be a reliable guide to current and future examinations. The specimen exam paper is designed to be relevant and to reflect the exam that will be set on this module.

Your final examination will have the same structure and style and the range of question will be comparable to those in the Specimen Exam. The number of questions will be the same, but the wording and the requirements of each question will be different.

Good luck on your final examination.

Further information

Online you will find documentation and information on each year's examination registration and administration process. If you still have questions, both academics and administrators are available to answer queries.

The Regulations are also available at www.cefims.ac.uk/regulations/, setting out the rules by which exams are governed.

References

Brooks C (2019) *Introductory Econometrics for Finance.* 4th Edition. Cambridge: Cambridge University Press.

Fan J (2004) *An Introduction to Financial Econometrics*. Department of Operation Research and Financial Engineering, Princeton, NJ: Princeton University, November.

Farnsworth GV (2008) *Econometrics in R*. [Online]. October 26. Available from: http://cran.r-project.org [Accessed 30 April 2020]

Hlavac M (2018) *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2.2. Available from: https://CRAN.R-project.org/package=stargazer [Accessed 30 April 2020]

R Core Team (2019) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/ [Accessed 30 April 2020]

Zeileis A and G Grothendieck (2005) 'zoo: S3 Infrastructure for Regular and Irregular Time Series', *Journal of Statistical Software*, 14 (6), 1–27. Available from: http://doi.org/10.18637/jss.v014.i06 [Accessed 30 April 2020]

UNIVERSITY OF LONDON

CENTRE FOR FINANCIAL AND MANAGEMENT STUDIES

MSc Examination
for External Students

91 DFMM459

FINANCE (BANKING)
FINANCE (ECONOMIC POLICY)
FINANCE (FINANCIAL SECTOR MANAGEMENT)
FINANCE (QUANTITATIVE FINANCE)

Financial Econometrics

Specimen Examination

*This is a specimen examination paper designed to show you the type of examination you will have at the end of the year for **Financial Econometrics**. The number of questions and the structure of the examination will be the same but the wording and the requirements of each question will be different. Best wishes for success in your final examination.*

The examination must be completed in **THREE** hours.

Answer **THREE** questions.

The examiners give equal weight to each question; therefore, you are advised to distribute your time approximately equally between three questions.

PLEASE TURN OVER

Answer **THREE** questions.

1.   *Answer **ALL** parts of the question.*

Consider the following GARCH(1,1) model

$$r_t = \mu + u_t \qquad u_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

where r_t is a daily stock return series.

a)   Explain carefully how you would estimate this model.

b)   Interpret the estimated coefficients of the model.

c)   Explain the difference between the conditional variance and the unconditional variance. Calculate the unconditional variance for the model. Under what conditions will unconditional variance be stationary?

d)   Describe two extensions to the original GARCH model. What additional characteristics of financial data might they be able to capture?

2.   *Answer **ALL** parts of the question.*

Suppose that you are interested in modelling the correlation between the returns on the stock of the International Airlines Group (owner of British Airways) and the returns on crude oil.

a)   Write down a constant correlation problem for this model.

b)   Explain carefully how you would estimate the model in part (a).

c)   Discuss the values of the coefficient estimates that you would expect.

d)   Discuss two alternative approaches to correlation modelling. What are the strengths and weaknesses of the multivariate GARCH model compared to the alternatives you propose?

3. *Answer **ALL** parts of the question.*

a) Explain why the linear model is inadequate as a specification for limited dependent variable estimation.

b) How does the logit model overcome the problem identified in part (a)?

c) Describe the intuition behind the maximum likelihood estimation technique used for a limited dependent variable model.

d) How can we measure whether an estimated logit model fits the data well or not?

4. *Answer **ALL** parts of the question.*

a) Explain the underlying philosophy behind maximum likelihood estimation.

b) OLS and maximum likelihood are used to estimate the parameters of a standard linear regression model. Will they give the same estimates? Explain your answer.

c) Discuss the main advantages and limitations of the maximum likelihood method.

d) Describe one hypothesis testing procedure that is available under the maximum likelihood estimation.

5. *Answer **ALL** parts of the question.*

a) Discuss three stylised features of financial data.

b) Can any of these features be modelled using linear time series models? Explain your answer.

c) Explain the moments of a random variable. How can you estimate these in empirical applications?

d) Explain carefully the Jarque–Bera test, stating clearly the null and alternative hypothesis. In the case of financial data, do you expect the Jarque–Bera test to accept or reject the null hypothesis? Explain your answer.

PLEASE TURN OVER

6. *Answer **ALL** parts of the question.*

Consider the following three models, which a researcher suggests might be a reasonable model of stock market returns.

$$y_t = y_{t-1} + u_t \tag{1}$$

$$y_t = 0.5\, y_{t-1} + u_t \tag{2}$$

$$y_t = 0.8\, u_{t-1} + u_t \tag{3}$$

a) What classes of models are these examples of?

b) What would the autocorrelation function for each of these processes look like? (Don't calculate the acf; simply consider the shape.)

c) Describe the steps that Box and Jenkins suggested should be involved in constructing the above models.

d) What procedure may be used to estimate the parameter in equation (3)? Explain how such a procedure works and why OLS is not appropriate.

7 *Answer **ALL** parts of the question.*

a) Explain why VAR models have become popular for application in economics and finance, relative to structural models derived from some underlying theory.

b) Define carefully the following terms:

   i) the structural VAR model

   ii) the standard VAR model.

c) Describe and evaluate two methods for determining the appropriate lag lengths in VARs.

d) Explain carefully the Granger causality test in a multivariate framework.

8   *Answer **ALL** parts of the question.*

Consider the following model where the dependent variable (say stock returns) is assumed to be dependent on *k* explanatory variables (including the constant) such that

$$y_1 = \beta_1 + \beta_2 x_{2,1} + \beta_3 x_{3,1} + \beta_4 x_{4,1} + \ldots + \beta_k x_{k,1} + u_1$$
$$y_2 = \beta_1 + \beta_2 x_{2,2} + \beta_3 x_{3,2} + \beta_4 x_{4,2} + \ldots + \beta_k x_{k,2} + u_2$$
$$\vdots$$
$$y_T = \beta_1 + \beta_2 x_{2,T} + \beta_3 x_{3,T} + \beta_4 x_{4,T} + \ldots + \beta_k x_{k,T} + u_T$$

a)   Write the above model in matrix form.

b)   Derive the estimators of the parameters of the model in part (a) using the method of least squares.

c)   What is meant by multifactor models? How are these models used in finance?

d)   Consider a two factor model such that

$$\mathbf{y} = \mathbf{\alpha} + \mathbf{X\beta} + \mathbf{u}$$

where $\mathbf{y}$ is the vector of returns, $\mathbf{\alpha}$ is a vector of ones (multiplied by the constant), $\mathbf{\beta}$ is a vector of betas, and $\mathbf{X}$ is the matrix of the factors. Derive the expected value and the variance of returns. Comment on the results.

[END OF EXAMINATION]

Financial Econometrics

Unit 1 Statistical Properties of Financial Returns

Contents

Unit Overview

Unit 1 explains how to calculate returns on financial assets, and considers various stylised facts (common statistical properties) concerning financial returns. The unit then analyses the distribution of returns, and, using examples, tests whether the various returns follow the normal distribution. Following that, you will study an analysis of time dependency, considering serial correlation in returns, serial correlation in volatility and asymmetry of volatility. An important finding is that time dependency can occur at more than one level (often time dependency exists in terms of the variance of the return but not the mean), and models of financial returns should take this into account.

Learning outcomes

When you have completed your study of this unit, the readings and the exercises, you will be able to:

- define and compute the various measures of financial returns, including the simple return, gross return, multi-period returns and continuously compounded returns
- calculate the sample moments of financial returns, including the skewness and kurtosis of financial returns
- explain and discuss some of the stylised statistical properties of asset returns
- analyse and appreciate the issue of time dependency in asset returns
- analyse the linear dependence across financial assets.

📖 Reading for Unit 1

Module Reader

Eric Jondeau, Ser-Huang Poon and Michael Rockinger (2007) 'Statistical properties of financial market data'. In *Financial Modelling under Non-Gaussian Distributions*. London: Springer. pp. 7–32.

1.1 Introduction

The main purpose of this unit is to describe and analyse some of the properties of returns on financial assets. Although financial analysts often observe prices on their screens such as stock prices, commodity prices, bond prices and exchange rates, the main objective of financial econometrics is to analyse financial returns. The focus on returns has many advantages. Returns are computed as the difference between prices over a particular horizon, so financial returns are stationary. This allows us to apply many of the standard calculation methods, summary statistics, and the standard econometric techniques you have studied before. Furthermore, returns can be easily compared across assets because they are scale-free. For instance, you could compare the annual return of an investment in stocks with an investment in a bond. Finally, as you will see in this unit, by focusing on financial returns it is possible to describe some common statistical properties of asset returns. These common features can be useful in modelling the time series properties of financial returns.

This unit starts by illustrating how to measure financial returns, the main variable that we try to model in financial applications. There are various definitions of returns such as simple returns, gross returns, multi-period returns, log returns, and so on. It is important from the start to be clear on how to compute the various types of returns. It is worth stressing that although financial returns are scale free, they should always be defined with respect to a particular time interval. This will be illustrated using examples.

After defining financial returns, we present some stylised facts about the properties of financial returns. As noted by Cont and Tankov (2004: 209–10),

> After all, why should properties of corn futures be similar to those of IBM shares or the Dollar/Yen exchange rate? Nevertheless, the result of more than half a century of empirical studies on financial time series indicates that this is the case if one examines their properties from a *statistical* point of view. The seemingly random variations of asset prices *do* share some quite nontrivial statistical properties. Such properties, common across a wide range of instruments, markets and time periods are called *stylized empirical facts*.

In this unit, we focus on some of these properties, mainly the time dependency properties, volatility clustering, asymmetric volatility, non-normality and cross-correlations across assets. But before doing so, it is important to refresh your memory about the various measures of moments of the distribution of a random variable and how these can be computed for samples of financial returns. This has an additional advantage because it will allow you to learn how to derive these measures yourself.

The reading for this unit will be based on Chapter 2 of the textbook by Eric Jondeau, Ser-Huang Poon and Michael Rockinger, *Financial Modelling under Non-Gaussian Distributions*, which is reprinted in your Reader. Although this reading is extracted from an advanced econometrics textbook, it sets out the issues in a clear and an insightful way. The outline of this unit follows very closely that of the reading. However, the unit will discuss some of the issues

in more detail, and will reproduce some of its results. It is important to note that many of the issues introduced in this unit will be revisited in other units and thus one of the purposes of this unit is to set the scene for the rest of the module.

1.2 Calculation of Asset Returns

Although in financial markets we mostly observe asset prices such as share prices or commodity prices, in empirical applications we often work with returns. One major reason for dealing with returns is that while prices are non-stationary (*ie* asset prices contain a unit root), asset returns are stationary. Since the module deals heavily with analysing and estimating asset return equations, it is worth spending some time defining returns and highlighting some of stylised facts about financial returns.

1.2.1 Simple returns

There are various definitions of returns. One such definition is the simple return. Let P_t be the price of an asset at time t and let P_{t-1} be the price of the asset at time $t - 1$. Assuming that the financial asset does not pay any dividends, then the one-period (for instance, one-day, one-week, one-month or one-year) simple net return denoted as R_t is given by the following equation

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{1.1}$$

Writing

$$\frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

one can define the one-period simple gross return as

$$1 + R_t = \frac{P_t}{P_{t-1}} \tag{1.2}$$

The left-hand side of the equation is also known as the discrete compounding factor. This is the case since we could write equation (1.2) as

$$P_t = \left(1 + R_t\right) P_{t-1} \tag{1.3}$$

It is important to stress that returns should always be defined with respect to a particular time interval. For instance, a statement such as 'the investment achieved a return of 20%' is meaningless unless we specify the horizon in which this return has been achieved. Thus, the above sentence should be qualified to include the time horizon, such as 'the investment achieved a monthly return of 20%' or 'the investment achieved an annual return of 20%'.

✎ **Review Question**

Consider a one-month investment in a BMW share. You bought the stock in period $t-1$ at \$90 and sold it in period t for \$100. Calculate the simple net return and the gross return of holding the investment over this one-month period.

The one-month simple net return is

$$R_t = \frac{100 - 90}{90} = 11.11\%$$

The one-month simple gross return is given by

$$1 + R_t = \frac{100}{90} = 111.11\%$$

1.2.2 Multiperiod returns

Suppose that you hold a financial asset from period $t-k$ to t, then the multiperiod simple net return denoted as $R_t(k)$ is given by the following

$$R_t(k) = \frac{P_t - P_{t-k}}{P_{t-k}} \tag{1.4}$$

For instance, assume that you hold the financial asset for two periods from $t-2$ to t then the two-period net simple return is given by

$$R_t(2) = \frac{P_t - P_{t-2}}{P_{t-2}} = \frac{P_t}{P_{t-2}} - 1$$

Writing

$$\frac{P_t}{P_{t-2}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}}$$

the two-period simple net return can be written as

$$R_t(2) = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} - 1$$

which yields

$$R_t(2) = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} - 1 = (1 + R_t) \times (1 + R_{t-1}) - 1$$

or

$$1 + R_t(2) = (1 + R_t) \times (1 + R_{t-1})$$

Notice that the simple two-period gross return is a geometric sum of the two one-period simple gross returns. Thus, adding two simple one-period gross returns *does not* yield the two-period return.

More generally, *the k-period* gross return can be written as

$$1 + R_t(k) = (1 + R_t) \times \ldots \times (1 + R_{t-k}) \tag{1.5}$$

🖎 Review Question

Continue with the above example, but suppose now that you hold the asset for two months and in month $t - 2$ the price was \$50. Calculate the two-month net return and gross return.

The two-month net return is given by

$$R_t(2) = \frac{P_t - P_{t-2}}{P_{t-2}} = \frac{100 - 50}{50} = 100\%$$

The two-month gross return is given by

$$1 + R_t(2) = (1 + R_t) \times (1 + R_{t-1})$$

where

$$1 + R_t = \frac{100}{90} = 1.11$$

$$1 + R_{t-1} = \frac{P_{t-1}}{P_{t-2}} = \frac{90}{50} = 1.80$$

Substituting the values in the above equation (without rounding) yields

$$1 + R_t(2) = 1.11 \times 1.80 = 200\%$$

1.2.3 Portfolio return

The simple net return for a portfolio consisting of N assets, denoted as $R_{p,t}$, is just the weighted average of individual simple returns. Thus,

$$R_{p,t} = \sum_{i=1}^{N} w_i R_{i,t} \tag{1.6}$$

where w_i is the weight of asset i in the portfolio and N is the number of assets in the portfolio. This is an extremely useful property for simple returns, and thus when dealing with portfolio analysis, it is easier to calculate simple returns.

1.2.4 Log returns

In this module we will base most of our examples on continuously compounded returns. The continuously compounded one-period return (or log return) denoted as r_t is given by

$$r_t = \ln(1 + R_t) = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln P_t - \ln P_{t-1} \tag{1.7}$$

where ln is the natural log function. Another way to express the above function is as follows

$$\exp(r_t) = \frac{P_t}{P_{t-1}} \tag{1.8}$$

The left-hand side of equation (1.8) refers to the continuously compounding factor since equation (1.8) can be written as

$$P_t = \exp(r_t) P_{t-1} \tag{1.9}$$

1.2.5 Multiperiod log returns

The main advantage of using log returns is that the multiperiod return is simply the sum of one-period returns. In other words,

$$r_t(k) = \sum_{j=0}^{k-1} r_{t-j} \tag{1.10}$$

This is a very useful property, which is extremely helpful in practical applications, as you will see in the next exercise.

✎ Review Question

Table 1.1 contains monthly share prices (adjusted for splits and dividends) for Barclays Bank from December 2007 to December 2008 and the monthly log returns. The data were obtained from Yahoo! (nd accessed February 2019). Using equation (1.7), check that you can calculate the one-month log-returns. Using equation (1.10), check that you can calculate the annualised continuously compounded returns for 2008.

Table 1.1   Monthly log return, Barclays Bank, December 2007–December 2008

Date	Share price	Monthly log return
December 2007	504	
January 2008	470	−0.069843573
February 2008	477.25	0.015307768
March 2008	453	−0.052148337
April 2008	456.5	0.007696575
May 2008	375	−0.196662674
June 2008	291.5	−0.25188602
July 2008	338	0.14800589
August 2008	353	0.043422161
September 2008	326.5	−0.078038108
October 2008	178.9	−0.601602958
November 2008	169.4	−0.054564208
December 2008	153.4	−0.099213893
Annualised continuously compounded return		**−1.189527379**

To calculate the annualised continuously compounded returns for 2008, you simply add the monthly log returns to obtain −1.1895. Alternatively, you can calculate the average

monthly return (-0.09913) and then multiply it by 12 to obtain the annualised continuously compounded returns (-1.1895); in this example, this step might seem pointless (dividing by 12 observations and then multiplying by 12 months), but it is required if you do not have 12 observations.

1.2.6 Real log returns

So far we have only considered nominal returns. In some practical applications we may also be interested in real returns (*ie* nominal returns adjusted for the inflation rate). The log returns are quite useful in calculating real returns.

Calculating the real return involves two steps. In the first step, you deflate the share price by the general price level (usually the Consumer Price Index, CPI). In the second step, you calculate the return using the same methods as applied above. As an example, consider P_t the price of the share at time t and CPI_t is the consumer price at time t. The real share price is given by

$$P_t^{Real} = \frac{P_t}{CPI_t} \tag{1.11}$$

The one-period simple real return is computed as

$$R_t^{Real} = \frac{P_t^{Real} - P_{t-1}^{Real}}{P_{t-1}^{Real}} = \left(\frac{P_t}{CPI_t} - \frac{P_{t-1}}{CPI_{t-1}} \right) \div \frac{P_{t-1}}{CPI_{t-1}} = \frac{P_t}{P_{t-1}} \div \frac{CPI_t}{CPI_{t-1}} - 1 \tag{1.12}$$

The continuously compounded one-period real return denoted as r_t^{Real} is given by the following

$$r_t^{Real} = \ln\left(1 + R_t^{Real}\right) = \ln\left(\frac{P_t}{P_{t-1}} \div \frac{CPI_t}{CPI_{t-1}} \right) \tag{1.13}$$

Using the log properties, equation (1.13) can be written as

$$r_t^{Real} = \left(\ln\left(P_t\right) - \ln\left(P_{t-1}\right)\right) - \left(\ln\left(CPI_t\right) - \ln\left(CPI_{t-1}\right)\right) \tag{1.14}$$

The first term on the right-hand side is simply the log return, while the second term is the one-period continuously compounded inflation rate, π_t *ie* equation (1.14) can be written as

$$r_t^{Real} = r_t - \pi_t \tag{1.15}$$

✎ Review Question

Table 1.2 contains monthly data on the New York Stock Exchange Price Index and the monthly Consumer Price Index (CPI) for the US. Using equation (1.14), check that you can calculate the monthly real rate of return.

Table 1.2   New York Stock Exchange Price Index and CPI, December 2007–December 2008

Date	NYSE Price Index	r_t	CPI	π_t	r_t^{Real}
December 2007	9740.32		211.737		
January 2008	9126.16	−0.065129	212.495	0.0035735	−0.0687025
February 2008	8962.46	−0.0181003	212.86	0.0017162	−0.0198165
March 2008	8797.29	−0.018601	213.667	0.0037841	−0.0223851
April 2008	9299.6	0.05552767	213.997	0.0015433	0.0539844
May 2008	9401.08	0.01085319	215.044	0.0048807	0.0059725
June 2008	8660.48	−0.0820544	217.034	0.0092114	−0.0912658
July 2008	8438.64	−0.025949	218.61	0.0072353	−0.0331843
August 2008	8382.08	−0.0067251	218.576	−0.0001555	−0.0065695
September 2008	7532.8	−0.1068293	218.675	0.0004528	−0.1072821
October 2008	6061.09	−0.2173772	216.889	−0.0082009	−0.2091763
November 2008	5599.3	−0.0792481	213.263	−0.0168596	−0.0623885
December 2008	5757.05	0.0277836	211.577	−0.0079371	0.0357207
Annualised continuously compounded real return					**−0.5250928**

As can be seen from Table 1.2, the monthly real rate of return is simply the monthly log return minus the one month continuously compounded inflation. To calculate the annualised real rate of return for 2008, you can simply add the real monthly log returns.

1.2.7 Log portfolio return

The main disadvantage of using log returns is that the log return of a portfolio of assets cannot be written as the weighted average of individual log returns. In fact, the portfolio log return denoted as $r_{p,t}$ is given by

$$r_{p,t} = \ln\left(1 + R_{p,t}\right) = \ln\left(1 + \sum_{i=1}^{N} w_i R_{i,t}\right) \neq \sum_{i=1}^{N} w_i r_{i,t} \tag{1.16}$$

This is the case because the log of a sum is different from the sum of logs. In your next reading the authors claim that this problem is usually considered minor in empirical applications. This is true to some extent, especially when returns are measured over short intervals of time. In such cases,

$$r_{p,t} \approx \sum_{i=1}^{N} w_i r_{i,t} \tag{1.17}$$

However, it is not advisable to use this approximation, and when you need to construct portfolio returns, it is better to use simple returns. In this module we will be mainly examining the behaviour of asset returns over time, and not portfolio returns, so we will rely heavily on log returns.

📖 **Reading**

Please now read Section 2.1 of the chapter by Jondeau, Poon and Rockinger.

📖 **Optional Reading**

If you are unsure about how to calculate any of the above returns, perhaps at this stage it would also be useful to revise the properties of logarithms. Section 1.5.6 of your text-book, *Introductory Econometrics for Finance*, by Chris Brooks (pp. 14–16) provides a quick review of the properties of logarithms. You might also read the start of Section 2.7, Returns in financial modelling pp. 77–79 of Brooks, which covers simple returns, log returns, and log returns of a portfolio.

1.3 Stylised Facts about Financial Returns

Although different assets such as stocks, bonds or commodities behave differently and are unlikely to be affected by the same set of information or events, much empirical literature on financial time series has revealed that financial asset returns possess some common statistical properties. These properties are often referred to as stylised facts. In what follows, we choose the most important stylised facts as listed by Cont and Tankov (2004: 211).

1. **Absence of autocorrelations**: (linear) autocorrelations of asset returns are often insignificant, except for very small intraday time scales ($\cong 20$ minutes) for which microstructure effects come into play.

2. **Heavy tails**: the (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied. In particular, this excludes stable laws with infinite variance and the normal distribution. However the precise form of the tails is difficult to determine.

3. **Gain / loss asymmetry**: one observes large drawdowns in stock prices and stock index values but not equally large upward movements.

4. **Aggregational Gaussianity**: as one increases the time scale Δt over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is not the same at different time scales.

5. **Volatility clustering**: different measures of volatility display a positive autocorrelation over several days, which quantifies the fact that high-volatility events tend to cluster in time.

6. **Conditional heavy tails**: even after correcting returns for volatility clustering (*eg* via GARCH-type models), the residual time series still exhibit heavy tails. However, the tails are less heavy than in the unconditional distribution of returns.

7. **Leverage effect**: most measures of volatility of an asset are negatively correlated with the returns of that asset.

8. **Volume / volatility correlation**: trading volume is correlated with all measures of volatility.

1.4   Distribution of Asset Returns

In the rest of the module we will analyse some of these properties in detail, and discuss how different models try to incorporate these features. The next sections of this unit illustrate some of these stylised facts using data on stock market indexes. However, before doing so, it would be useful to refresh your memory about the moments of a random variable, and then show you how these can be used to illustrate the properties of financial returns.

1.4.1   Moments of a random variable

Denote by the random variable X the log return of a financial asset. You may recall from your previous studies that the cumulative distribution function for the random variable can be defined as

$$F(X) = \Pr[X \le x] = \int_{-\infty}^{x} f_X(u)\, du \qquad (1.18)$$

where f_X is the probability density function (pdf). The un-centred moments of the random variable X are defined as

$$m_k = E\left[X^k\right] = \int_{-\infty}^{+\infty} x^k f_X\, dx \ \text{ for } k = 1, 2, \ldots \qquad (1.19)$$

Although the above equations seem complex, their interpretation is quite straightforward. When $k = 1$, you obtain the first un-centred moment of the random variable, which is simply the mean of the random variable *ie*

$$m_1 = E[X] = \mu \qquad (1.20)$$

The centred first moment equals zero. When $k = 2$, we can obtain the second centred moment of the random variable, which is simply the variance *ie*

$$m_2 = E\left[X^2\right] = V(X) = \sigma^2 \qquad (1.21)$$

When $k = 3$, we obtain the skewness of the random variable, and when $k = 4$, we obtain the kurtosis of the series. The (standardised) skewness of the series, denoted as s, is defined as

$$s = Sk[X] = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right] \qquad (1.22)$$

The (standardised) kurtosis of the series, denoted as κ, is defined as

$$\kappa = Ku[X] = E\left[\left(\frac{X - \mu}{\sigma}\right)^4\right] \qquad (1.23)$$

The skewness and kurtosis of the series are important for understanding financial returns. Skewness measures the asymmetry of the distribution of financial returns. When it is positive, it indicates that large positive realisations of X are more likely. When it is negative, it indicates that large negative realisations of X are more likely. Kurtosis, on the other hand, measures the thickness of the tails of the distribution. In particular, it measures the tail thickness in relation to the normal distribution (for a normal distribution kurtosis equals 3, so excess kurtosis is measured by $\kappa - 3$). Remember from the above discussion that one of the stylised facts is that financial returns have heavy tails, and that these heavy tails persist even after correcting for volatility clustering.

1.4.2 Empirical moments

In practical application we need to consider empirical measures for the above moments. You will most likely know how to compute these moments already, but it is worth reviewing them very quickly. Consider a time series of realised asset returns r_t, $t = 1, ..., T$. The widely used measure of location is the sample mean, which is given by the following equation:

$$\bar{r} = \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} r_t \tag{1.24}$$

Variance is the most widely used measure for dispersion, and is given by the following equation

$$\hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=1}^{T} \left(r_t - \bar{r} \right)^2 \tag{1.25}$$

In financial applications the square root of the variance is often used to measure volatility. Another useful measure of dispersion is the mean absolute deviation (MAD), which is given by

$$MAD = \frac{1}{T} \sum_{t=1}^{T} \left| r_t - \bar{r} \right| \tag{1.26}$$

The sample skewness can be computed using the following equation

$$\hat{s} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{r_t - \bar{r}}{\sigma} \right)^3 \tag{1.27}$$

The sample kurtosis can be computed using the following equation

$$\hat{\kappa} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{r_t - \bar{r}}{\sigma} \right)^4 \tag{1.28}$$

The above measures are known as summary statistics. Under the assumption that financial returns are normal, we have the following *asymptotic* results

$$\sqrt{T} \left(\hat{\mu} - \mu \right) \sim N\left(0, \sigma^2 \right) \tag{1.29}$$

$$\sqrt{T}\left(\hat{\sigma}^2 - \sigma^2\right) \sim N\left(0, 2\sigma^4\right) \tag{1.30}$$

$$\sqrt{T}\,\hat{s} \sim N\left(0, 6\right) \tag{1.31}$$

$$\sqrt{T}\left(\hat{\kappa} - 3\right) \sim N\left(0, 24\right) \tag{1.32}$$

In fact, based on the results in (1.31) and (1.32), one can derive a statistic to test the hypothesis of normality, known as the Jarque–Bera test, defined as

$$JB = T\left[\frac{\hat{s}^2}{6} + \frac{\left(\hat{\kappa} - 3\right)^2}{24}\right] \tag{1.33}$$

The test is distributed asymptotically as $\chi^2\left(2\right)$ under the hypothesis that the distribution is normal. A large value of the J–B statistic implies that we can reject the null hypothesis that the returns are normally distributed.[1]

1.4.3 Example – Empirical moments

Let's now use these measures to illustrate some of the properties of financial returns. The example concerns daily, weekly and monthly data for the Standard & Poor's 500 stock price index from January 2000 to December 2019. For each of the stock price index series we calculate the corresponding one-period log returns (daily log returns for daily data, weekly log returns for weekly data, and monthly log returns for monthly data). For each of the returns series, we calculate the mean, standard deviation, skewness and kurtosis, and the J–B statistic. The histograms of the returns are shown in the Figures below.

For daily data, the mean of the daily log return is 0.016% and the standard deviation is 1.19%, which is quite high. The histogram for the daily returns is shown in Figure 1.1.

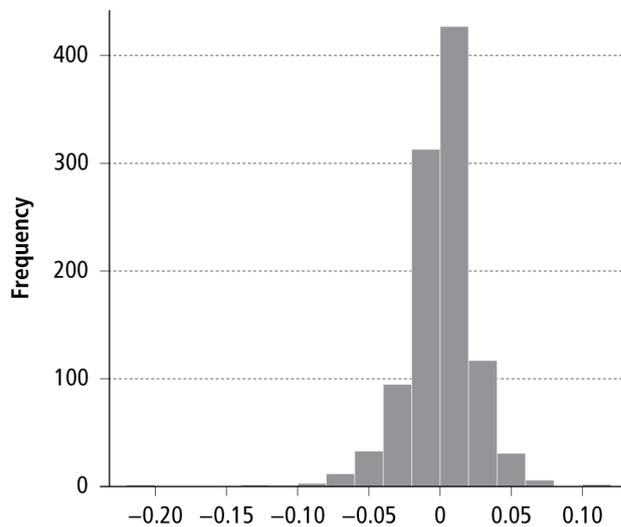**Figure 1.1 Daily log returns, S&P 500 index**

---
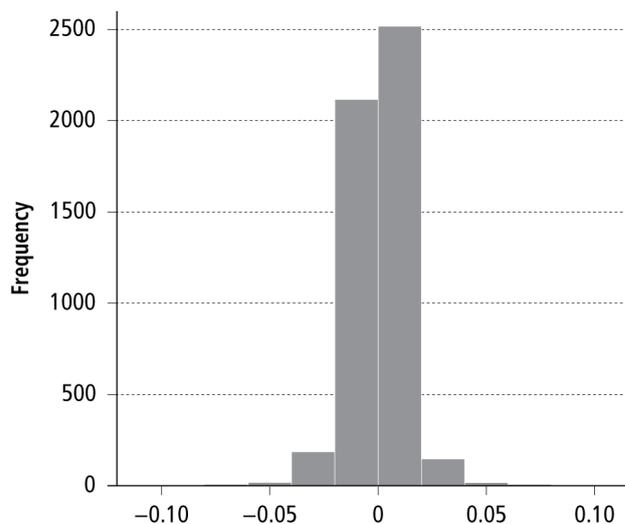
[1] It is important to stress that the J–B test applies to only large samples, as explained in your reading.

The maximum and minimum daily returns are 11.0% and −9.5%. The daily index return has high sample kurtosis of 11.64; this indicates the sample distribution of daily returns is more peaked and has fatter tails than a normal distribution (with the same mean and variance). The daily index return is negatively skewed, with sample skewness equal to −0.23. The J–B test statistic equals 15,706.07, (Prob. value 0.0000) and strongly rejects the null hypothesis of normality.

For weekly data, the mean of the log return is 0.077% and the standard deviation is 2.40%. The maximum and minimum range is 11.4% to −20.0%. The weekly index return has a high sample kurtosis of 10.27. The weekly returns are also negatively skewed, with skewness equal to −0.88. Again, the J–B test strongly rejects the null hypothesis of normality, with calculated value 2,434.79 and Prob. value equal to 0.0000. The histogram of weekly returns is shown in Figure 1.2

Figure 1.2   Weekly log returns, S&P 500 index

Finally, we report data for monthly returns. The histogram is shown in Figure 1.3.

Figure 1.3   Monthly log returns, S&P 500 index
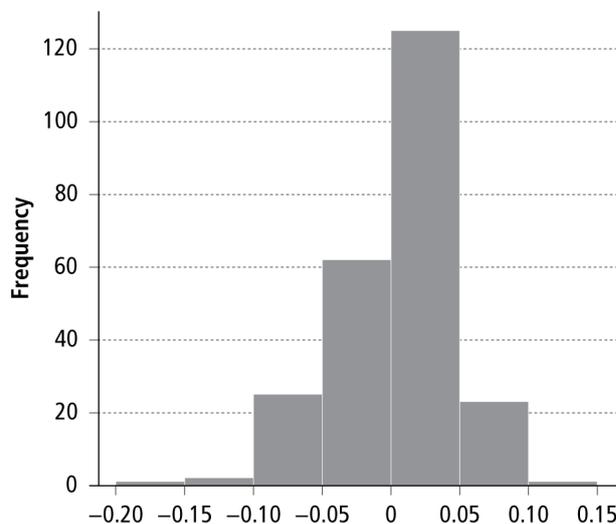
The mean monthly return is 0.352%, and the standard deviation is 4.23%. The maximum and minimum monthly return are 10.2% and −18.6%. The monthly log return has a lower kurtosis of 4.58. This is expected for monthly data. The monthly returns still exhibit negative skewness, with skewness equal to −0.79. Again, the J–B test strongly rejects the null hypothesis of normality of returns, with calculated value 49.91 (Prob. value 0.0000).

> 📖 **Reading**
>
> Please now read Section 2.2.1 to 2.2.3 from Jondeau, Poon and Rockinger.

1.5 Time Dependency

As noted above, one of the stylised facts is that autocorrelations of asset returns are often insignificant *ie* asset returns exhibit no time dependency. However, it is important to note that time dependency can occur at several levels. In what follows, we refer to three levels of dependency: Serial correlation in returns, serial correlation in squared returns, and volatility asymmetry.

1.5.1 Serial correlation in returns

Here, we are interested in testing the null hypothesis that the first p returns are not serially correlated. You may remember from your other studies that a measure of autocorrelation of returns of order j is given by the following:

$$\hat{\rho}_j = \frac{\sum_{j=t+1}^{T}\left(r_t - \overline{r}\right)\left(r_{t-j} - \overline{r}\right)}{\sum_{t=1}^{T}\left(r_t - \overline{r}\right)^2} \tag{1.34}$$

Unit 4 will use test statistics such as the Ljung–Box Q statistic to test the significance of autocorrelations, and will suggest ways to estimate models of financial returns. However, we will use this statistic in the example that follows, and it is calculated as

$$Q_p = T\left(T+2\right)\sum_{j=1}^{p}\frac{1}{T-j}\hat{\rho}_j^2 \tag{1.35}$$

It is asymptotically distributed as χ^2 with p degrees of freedom, under the null hypothesis of no correlation. As discussed in the stylised facts, autocorrelations of asset returns are often insignificant and hence there is little time dependency in asset returns. However, this stylised fact cannot be generalised. Depending on the time horizon being used, one could find weak evidence of serial correlation in some asset returns.

1.5.2 Serial correlation in volatility

To test for dependency in volatility, we need to construct models that generate time-varying volatility measures. ARCH, GARCH and their family of models do exactly that. In Unit 5, we will introduce these models as well as ways to test for serial dependence in volatility. To anticipate the discussion in

Unit 5, we could use the Ljung–Box Q statistic to test for serial correlation in squared returns and absolute returns. Most empirical evidence suggests that there is a strong evidence of serial correlations in squared returns and absolute returns, especially for daily and weekly data as shown in Table 2.4 of your reading. In other words, large returns of either sign tend to be followed by large returns of either sign or the volatility of returns tends to be serially correlated. This is often referred to in the literature as volatility clustering.

1.5.3 Volatility asymmetry

One important feature of financial returns is that volatility exhibits asymmetric behaviour. In particular, there is wide empirical evidence that volatility is more affected by negative returns than positive returns. In Unit 5 we will show you how these ARCH and GARCH models can be modified to take asymmetric volatility into account. Table 2.5 of your reading shows parameter estimates of volatility asymmetry for the various stock market indexes.

> 📖 **Reading**
>
> I would like you now to read Section 2.3 in Jondeau, Poon and Rockinger. Don't worry if you don't understand all of these equations. These will become clear in Units 4 and 5. The main lessons I want you take from this section are as follows.
>
> ✍ Make sure your notes cover these issues clearly.
>
> - Time dependency can occur at more than one level, and for financial returns time dependency often occurs at the second moment (the variance) and not the first moment (the mean);
> - Therefore it is important to construct models of time varying volatility for financial returns, and devise statistics to test for the correlation at higher moments;
> - Volatility of financial returns may exhibit asymmetric behaviour and this needs to be accounted for in empirical models.

Jondeau *et al* (2007) Section 2.3 'Time dependency'. Reproduced in the Module Reader from *Financial Modelling under Non-Gaussian Distributions*.

1.5.4 Example – Serial correlation of returns

Perhaps the best way to appreciate the issue of time dependency is to consider again the monthly log return of the S&P 500 index. The serial correlation at order 1 to 6 and the corresponding Ljung–Box Q statistic are given in Table 1.3.

Table 1.3   Monthly log returns, S&P 500 index

Lag	Autocorrelation	Ljung-Box Q statistic	Probability
1	0.076	1.3900	0.2384
2	−0.032	1.6331	0.4420
3	0.089	3.5605	0.3130
4	0.067	4.6492	0.3252
5	0.073	5.9740	0.3088
6	−0.096	8.2436	0.2208

In Unit 4, you will learn how to use such a table for diagnostic checking, but for now it is important to understand the intuition. As can be seen from this table, there is no evidence of serial correlation in the monthly log returns (the Prob. values are all greater than the 0.05 level). The Q statistic does not reject the null hypothesis of no serial correlation at the various lags.

Now let's consider the square of the returns and repeat the exercise. The results are shown in Table 1.4. As can be seen from this table, there is strong evidence of serial correlation in the squared returns. The implications of this will be studied in Unit 5.

Table 1.4   Square of monthly returns, S&P 500 index

Lag	Autocorrelation	Ljung-Box Q statistic	Probability
1	0.274	18.111	0.000
2	0.102	20.630	0.000
3	0.182	28.688	0.000
4	0.268	46.243	0.000
5	0.180	54.210	0.000
6	0.123	57.932	0.000

1.6 Linear Dependency across Asset Returns

So far we have focused on some of the stylised facts about individual series of asset returns. In this section we shift the focus towards the dependence of returns across assets. As you may recall from your other studies, the widely used measure of dependence is the correlation coefficient (also known as Pearson's correlation), which is given by the following equation

$$\rho[X,Y] = \frac{Cov[X,Y]}{\sqrt{V(X)V(Y)}} \tag{1.36}$$

where $Cov(X,Y)$ is the covariance between X and Y, $V(X)$ is the variance of X, and $V(Y)$ is the variance of Y. The correlation coefficient must lie between -1 and 1, with a zero value indicating no correlation between the two series. An estimator of the correlation coefficient is given by the following:

$$\hat{\rho} = \frac{\sum_{t=1}^{T}(X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^{T}(X_t - \bar{X})^2 \sum_{t=1}^{T}(Y_t - \bar{Y})^2}} \tag{1.37}$$

As you can see from Table 2.6 in your reading, the correlation between the various stock market indexes is positive, implying that stock indexes tend to move together. Another interesting observation is that the correlation tends to increase in turbulent times (for example, in times of crisis, the correlation between the indexes becomes more positive). However, as discussed in your reading, this finding could be a spurious outcome and driven mainly by increased volatility.

What matters for us in this module is the possibility of jointly modelling asset returns and their volatility. Unit 6 introduces the multivariate GARCH models, which are an extension of the univariate GARCH models discussed in Unit 5. As you will see in Unit 6, multivariate GARCH models provide us with a useful tool to model time-varying autocorrelation. This would allow us to identify whether there have been structural breaks in the correlation coefficient over time.

📖 **Reading**

Please now read Sections 2.4.1 and 2.4.2 of Jondeau, Poon and Rockinger.

1.6.1 Example – Linear dependence between stock market returns

The data set contains weekly prices for the Dow Jones Industrial Average, the Paris blue chip stock market index CAC, and the Frankfurt blue chip stock market index DAX, for the period January 2001 to January 2020. Table 1.5 provides summary statistics of the log weekly return for the three stock market indexes. The summary statistics were obtained using the basicStats function (in the fBasic package). The Jarque–Bera statistic was obtained using the jarque-beraTest function, (also in the fBasics package). Skewness and kurtosis were obtained using the skewness and kurtosis functions (from the moments package). As you can see, the DJI exhibits more kurtosis and skewness, compared to the CAC and DAX. We reject the null hypothesis of normality for all three indexes.

Table 1.5   Weekly log returns, DJI, CAC and DAX

	DJI	CAC	DAX
Minimum	−0.200298	−0.250504	−0.243470
Maximum	0.106977	0.124321	0.149421
Mean	0.000980	0.000008	0.000714
Median	0.002903	0.002589	0.004029
Stdev	0.022921	0.029033	0.031342
Skewness	−1.069142	−0.973264	−0.708977
Kurtosis	12.09335	9.89086	8.64527
Jarque–Bera	3617.706	2125.690	1404.592

The correlations between the weekly returns for the three indexes are shown in Table 1.6.

Table 1.6   Weekly log returns, correlations

	DJI	CAC	DAX
DJI	1.000000	0.789889	0.787434
CAC	0.789889	1.000000	0.923835
DAX	0.787434	0.923835	1.000000

R presents these correlations in a matrix form (matrix algebra is the subject of the next unit). But for now, notice that elements in the main diagonal all take the value of 1, because these measure the correlation of the returns of a particular index with itself. The off-diagonal elements measure the sample correlation across the various indexes. Interestingly, the correlation matrix shows higher correlation between the log weekly returns of the European stock indexes, relative to the correlation between each of the Paris and Frankfurt indexes and the Dow Jones.

As implied in your reading, it is highly unlikely for the correlation to remain constant throughout the entire sample. Thus it is worth estimating the time varying correlation. This will be the subject of Unit 6. But just to anticipate the discussion of Unit 6, Figures 1.4 and 1.5 show the time varying correlation between the weekly log return on DJI and CAC, and the weekly log return on CAC and DAX, using a six-month rolling window.[2]

Figure 1.4   Time-varying correlation between weekly log returns of DJI and CAC
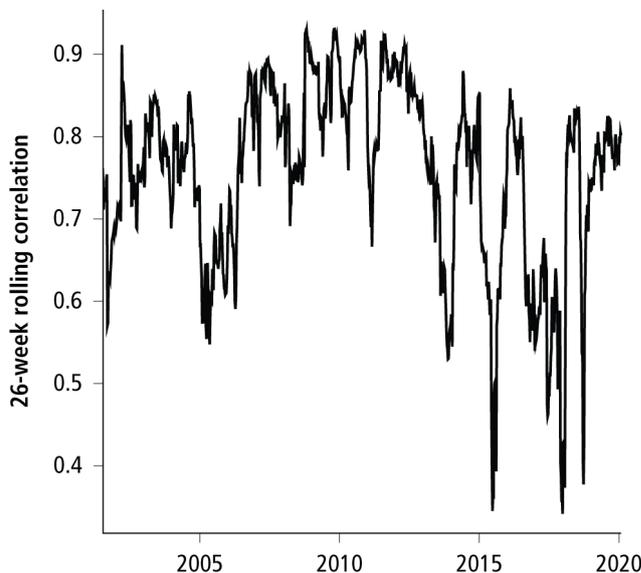
It is quite clear that the correlation coefficient exhibits very volatile behaviour. The correlation coefficient between the returns on the DJI and CAC (Figure 1.4) takes values above 0.9 and below 0.4. The correlation between the returns on the CAC and the DAX (Figure 1.5) is in a narrower range, the maximum value is just below one, and the lowest value is around 0.75. What is important to stress is that the correlation between returns is not constant and tends to vary over time.

---

[2]   This involves calculating the correlation coefficients for the first six months, and then rolling the sample forward by including a new observation and dropping the first observation from the previous sample. It is an arbitrary method but an easy way to calculate the time varying correlation.

Figure 1.5   Time-varying correlation between weekly log returns for CAC and DAX
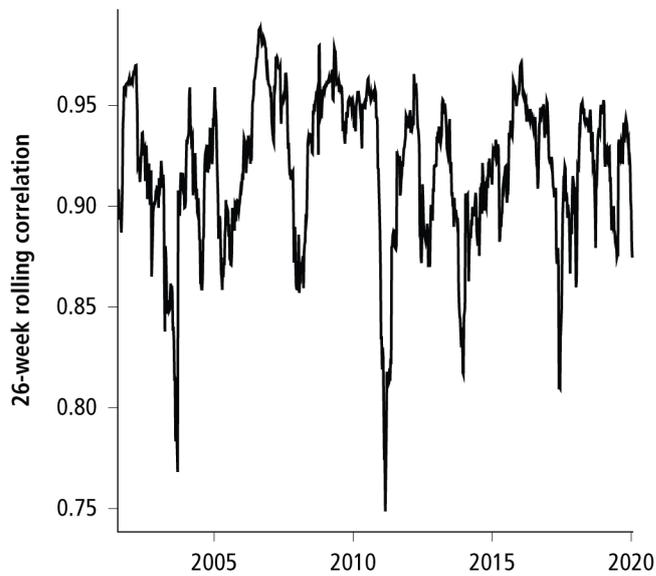
1.7   Conclusion

The main objective of this unit has been to analyse some of the properties of financial asset returns. Please check that you have achieved the Learning Outcomes listed at the start of the unit. They are repeated here, so that you can now test yourself against them. You should now be able to:

- define and compute the various measures of financial returns, including the simple return, gross return, multiperiod returns and continuously compounded returns
- calculate the sample moments of financial returns, including the skewness and kurtosis of financial returns
- explain and discuss some of the stylised statistical properties of asset returns
- analyse and appreciate the issue of time dependency in asset returns
- analyse the linear dependence across financial assets.

1.8 Exercises

Question 1

The data used in the Example in Section 1.4.3 is contained in the three text files M459_U1_SP500_daily.txt, M459_U1_SP500_weekly.txt and M459_U1_SP500_monthly.txt.

Use the data to replicate the results in Section 1.4.3.

The index data is for the S&P 500 index (Yahoo! nd accessed February 2020).

Read data from text file and create zoo object

For the monthly observations, the following commands read the data and create a zoo object for the stock index, assuming the text file is in the current working directory (the zoo function is available in the zoo package, Zeileis and Grothendieck, 2005).

```
M459_U1_SP500_monthly <-
read.table("M459_U1_SP500_monthly.txt", sep = "\t", header
= TRUE)

M459_U1_SP500_monthly_zoo <-
read.zoo(M459_U1_SP500_monthly, format = "%d/%m/%Y")

SP500 = zoo(M459_U1_SP500_monthly_zoo)
```

The first command reads the data from the text file and creates a data frame. The second command creates a zoo object in which the entries are indexed by the date of the observations. The third command creates a separate zoo object for the monthly SP500 index data, which is easier to manipulate in other commands.

In the read.table command, sep = "\t" indicates the data are separated by tabs. Within the read.zoo command, the format for the dates as they appear in the data file is specified with format = "%d/%m/%Y".

Note the use of the quotation symbol " in the R commands. The symbols " and " will not be recognised in R.

To see the data for SP500, type SP500 (followed by Enter).

To see which objects have been created, use

```
ls()
```

Saving a Workspace

To save a Workspace, use File | Save Workspace… (or use Control and s). Provide a name for the file, which will be saved as an R image with the extension .Rdata. The file will be saved in the current working directory (or you can browse to another folder).

To load a Workspace that you have saved previously, use File | Load Workspace…, and select the file to load. You can browse to another folder if the .Rdata file you want to load is not in the current working directory.

Deleting an object

To delete an object, use rm( ). For example, to remove the object SP500 you would use rm(SP500).

To remove all objects (for example, if you want to start with a new data set but you would like to keep working with the same loaded packages), go to Misc | Remove all objects. The same thing can be achieved with

```
rm(list=ls(all=TRUE))
```

Creating a histogram

To create the histogram of returns use

```
hist(diff(log(SP500)),col = "lightblue")
```

Summary statistics

The mean of the logged return for the SP500 index is computed with

```
mean(diff(log(SP500)))
```

The standard deviation of the logged return is computed with

```
sd(diff(log(SP500)))
```

The maximum value of the logged return is found with

```
max(diff(log(SP500)))
```

The minimum value of the logged return is found with

```
min(diff(log(SP500)))
```

Jarque–Bera test, skewness and kurtosis

To perform the Jarque–Bera test you will need to use the function jarqueberaTest( ) in the fBasics package (Wuertz *et al*, 2017). To calculate skewness and kurtosis, please use the functions skewness( ) and kurtosis( ) in the moments package (Komsta and Novomestky 2015). The values of skewness and kurtosis you will obtain are consistent with the value of the Jarque–Bera statistic produced by jarqueberaTest.

Please do *not* use the skewness( ) function available in the timeDate package (which loads with fBasics), and please do *not* use the kurtsosis( ) function available in the timeDate package. These functions do not produce values of skewness and kurtosis consistent with the Jarque–Bera statistic produced by jarqueberaTest.

Also, please do *not* use the jarque.test( ) function from the moments package; jarque.test works on a vector of values, not a zoo object.

To make sure you use the skewness and kurtosis functions from the moments package, load the moments package *after* you have loaded the fBasics package.

The following command performs the Jarque–Bera test using the jarqueberaTest( ) function in the fBasics package. Note the capital T.

```
jarqueberaTest(diff(log(SP500)))
```

To calculate skewness, use the skewness( ) function in the moments package

```
skewness(diff(log(SP500)))
```

To calculate kurtosis, use the kurtosis( ) function in the moments package

```
kurtosis(diff(log(SP500)))
```

Daily and weekly data for S&P 500 index

To work with the daily and weekly data, you can adapt the commands above as required.

Question 2

Using the monthly data on the S&P 500 index, compute the correlations and Ljung–Box Q statistic in Section 1.5.4 Example – Serial correlation of returns.

Autocorrelation function

The correlations for up to six lagged values of the monthly return on the S&P 500 index can be computed with the following command (and saved in dlsp500_acf). The command assumes the Workspace containing the monthly SP500 zoo object is open.

```
dlsp500_acf <- acf(coredata(diff(log(SP500))), type =
"correlation", lag.max = 6, plot = FALSE)
```

This command applies acf to the logged return, and computes the correlation for the various lags. The type could instead be set to "covariance" or "partial". The script suppresses the production of a plot of the autocorrelation function.

Note that the acf( ) function works with data that is regularly spaced. Much financial data has irregular dates. To overcome this we use coredata( ) within the acf function; this produces the autocorrelation function for the data itself, in sequence but ignoring any gaps in the dates.

Ljung–Box Q statistic

The Ljung–Box Q statistic can be computed with the Box.test function (in the stats package, which is loaded as part of R) using the following command

```
Box.test(coredata(diff(log(SP500))), lag = 1, type =
"Ljung-Box")
```

📝 Study Note

In most applications you will be able to use the Box.test function on a numeric vector, univariate time series, or zoo object without any adaption. However, for this particular monthly series the command

```
Box.test(diff(log(SP500)), lag = 1, type = "Ljung-Box")
```

produces NA values.

Including the coredata( ) specification addresses this issue. The statement

---

```
Box.test(as.numeric(diff(log(SP500))), lag = 1, type =
"Ljung-Box")
```

also works.

The Ljung–Box Q statistic can be computed for lags 2, 3, 4, 5 and 6 by varying the 'lag = ' specification in the Box.test command.

Square of monthly returns

The autocorrelation for up to six lags of the squared monthly log return for the S&P 500 index can be computed with

```
dlsp500_squared_acf <- acf(coredata(diff(log(SP500))^2),
type = "correlation", lag.max = 6, plot = FALSE)
```

The Ljung–Box Q statistic can be computed for one lag with

```
DLSP500 = as.numeric(diff(log(SP500)))

Box.test(DLSP500^2, lag = 1, type = "Ljung-Box")
```

The first command creates a numeric vector containing the logged returns. The Ljung–Box Q statistic for lags 2, 3, 4, 5 and 6 can be computed by varying the 'lag = ' specification in the Box.test command.

Question 3

Replicate the results in Section 1.6.1 Example – Linear dependence between stock market returns. The data for the three stock market indexes is in the text file M459_U1_Indexes.txt.

Reading the data and creating zoo objects

The following commands read the data from the text file (assuming the text file is located in the current working directory), create zoo objects using the zoo function (in the zoo package), and create the three weekly log returns.

```
M459_U1_Indexes <- read.table("M459_U1_Indexes.txt",
sep="\t", header = TRUE)

M459_U1_Indexes_zoo <- read.zoo(M459_U1_Indexes, format =
"%d/%m/%Y")

DJI = zoo(M459_U1_Indexes_zoo$DJI)

CAC = zoo(M459_U1_Indexes_zoo$CAC)

DAX = zoo(M459_U1_Indexes_zoo$DAX)

DLDJI = diff(log(DJI))

DLCAC = diff(log(CAC))

DLDAX = diff(log(DAX))
```

Summary statistics

The summary statistics for each of the returns could be computed individually with mean( ), min( ) *etc*. Alternatively it is possible to combine the three series and work on the group as a whole. The following command merges the three series

```
returns = merge(DLDJI,DLCAC, DLDAX)
```

The following command then produces the list of summary statistics for the three series, using the basicStats( ) function in the fBasics package

```
returnstats <- basicStats(returns)
```

The Jarque–Bera statistic (for the return on the DJI, for example) can be obtained with the jarqueberaTest( ) function in the fBasics package

```
jarqueberaTest(DLDJI)
```

Skewness can be computed with the skewness( ) function in the moments package

```
skewness(DLDJI)
```

To calculate kurtosis, use the kurtosis( ) function in the moments package

```
kurtosis(DLDJI)
```

The correlation matrix for the three returns can be obtained using the cor( ) function *applied to the merged series*,

```
cor(returns)
```

Rolling returns

The rolling correlation can be computed with the roll_corr function in the roll package (Foster 2020).

The first command below creates a zoo object for the rolling correlation for the returns on DJI and CAC, using a rolling window of width 26 observations (corresponding to 26 weeks, or half a year). The command uses the date index of the zoo object DLDJI to index the zoo object for the rolling correlation.

The first 26 observations will be lost, so the second command omits any missing observations. The final command plots the rolling correlation over time.

```
roll_DJI_CAC = zoo(roll_cor(DLDJI, DLCAC, width = 26),
order.by = index(DLDJI))

roll_DJI_CAC = na.omit(roll_DJI_CAC)

plot(roll_DJI_CAC, lwd = 2, ylab = "26-week rolling
correlation", xlab = "", xaxs = "i")
```

The specification xaxs = "i" ensures the plot extends to the edges of the plot area.

The rolling correlation for the weekly log returns on the CAC and the DAX can be computed and plotted with

```
roll_CAC_DAX = zoo(roll_cor(DLCAC, DLDAX, width = 26),
order.by = index(DLCAC))

roll_CAC_DAX = na.omit(roll_CAC_DAX)

plot(roll_CAC_DAX, lwd = 2, ylab = "26-week rolling
correlation", xlab = "", xaxs = "i")
```

Question 4

The text file M459_U1_OIL_ER.txt contains weekly data on oil prices (Energy Information Administration) and the trade-weighted exchange rate index of the US dollar against the currencies of a broad group of major US trading partners (Board of Governors of the Federal Reserve System), from January 2000 to December 2019.

a) Calculate the sample mean, standard deviation, Jarque–Bera statistic, skewness and kurtosis of weekly log returns, and produce the histogram of weekly log returns, for both the oil price and the exchange rate.

b) Compute the correlation coefficient between the weekly log return for the oil price and the exchange rate. Comment on the results.

c) Compute and plot the time varying correlation between the weekly log returns for the oil price and the exchange rate using a rolling 26-week window. Comment on the results.

Read data and create zoo objects

The following commands read the data from the text file (assuming the text file is located in the current working directory) and create the required zoo objects for the oil price, exchange rate, and the two logged returns (using the zoo function in the zoo package)

```
M459_U1_OIL_ER <- read.table("M459_U1_OIL_ER.txt", sep =
"\t", header = TRUE)

M459_U1_OIL_ER_zoo <- read.zoo(M459_U1_OIL_ER, format =
"%d/%m/%Y")

OIL = zoo(M459_U1_OIL_ER_zoo$OIL)

ER = zoo(M459_U1_OIL_ER_zoo$ER)

DLOIL = diff(log(OIL))

DLER = diff(log(ER))
```

Summary statistics and histogram

The mean of the return on the oil price can be computed with

```
mean(DLOIL)
```

The standard deviation can be computed with

```
sd(DLOIL)
```

The Jarque–Bera test can be performed (using the jarqueberaTest function in the fBasics package) with

```
jarqueberaTest(DLOIL)
```

Skewness can be computed using the skewness( ) function (in the moments package) with

```
skewness(DLOIL)
```

Kurtosis can be computed using the kurtosis( ) function (in the moments package) with

```
kurtosis(DLOIL)
```

The histogram of the logged oil returns can be obtained with

```
hist(DLOIL ,col = "lightblue")
```

The summary statistics and histogram for the logged return on the exchange rate can be obtained by adapting the above commands (replacing DLOIL with DLER).

Correlation matrix

The correlation between the logged oil returns and logged exchange rate returns can be obtained with the following two commands

```
returns = merge(DLOIL, DLER)

cor(returns)
```

Rolling correlation coefficient

The rolling correlation coefficient can be computed and plotted with the following commands (using the roll_cor function in the roll package)

```
roll_OIL_ER = zoo(roll_cor(DLOIL, DLER, width = 26),
order.by = index(DLOIL))

roll_OIL_ER = na.omit(roll_OIL_ER)

plot(roll_OIL_ER, lwd = 2, ylab = "26-week rolling
correlation", xlab = "", xaxs = "i")

abline (h = 0)
```

Question 5

Go to the website http://uk.finance.yahoo.com and download daily data for the FedEx Corporation stock over the years 2004 to 2019. Use the closing price.

Hint: For the Company or symbol, type 'FDX' and for the market, choose USA or New York Stock Exchange. Then choose Historical Prices. For reference, on 2 January 2004 the closing price was 67.89, on 5 January 2004, 67.95, and on 30 December 2019, 150.14.

For this exercise we are asking you to work with the *unadjusted* closing price of FedEx stock. You should be aware that in quantitative analysis it is common to use the adjusted prices. The unadjusted prices can make discrete jumps, due, say, to stock splits, which are not reflective of the underlying behaviour of the return (in a two-for-one stock split the stock price can halve in value, for example).

a) Using an Excel spreadsheet (or similar application) calculate the daily log return for the stock.

b) Using the daily log returns, compute the continuously compounded annual return for 2004 to 2019. Plot the annual return on a graph and comment on the graph.

1.9 Answers to Exercises

Question 1

The results for the empirical moments in relation to the daily, weekly and monthly log returns on the S&P 500 index are presented in Section 1.4.3, including histograms of returns, skewness, kurtosis and Jarque–Bera test.

Question 2

The autocorrelation function, and the Ljung–Box Q statistic, are computed for the monthly log returns on the S&P500 index, and the squared returns, in Section 1.5.4.

Question 3

The results for the weekly log returns on the three stock market indexes, being the Dow Jones, CAC and DAX, are presented and discussed in Section 1.6.1. This includes summary statistics, skewness, kurtosis, Jarque–Bera test and the correlation matrix. This section also includes a plot of the rolling correlation coefficient between the returns on the DJ index and the CAC index, and the rolling correlation coefficient between the CAC index and the DAX index.

Question 4

This exercise concerns weekly data on oil prices and the US trade-weighted exchange rate index.

a) The sample mean, standard deviation, skewness, kurtosis and Jarque–Bera statistic are shown for the weekly log exchange rate returns and oil returns in Table 1.7. The histogram of the weekly log exchange rate returns is displayed in Figure 1.6, and the histogram of weekly log oil returns is displayed in Figure 1.7.

Table 1.7   Weekly log returns, oil price and trade-weighted exchange rate

	Oil price	Exchange rate
Mean	0.000865	0.000114
Standard deviation	0.040586	0.005777
Skewness	−0.497754	0.236317
Kurtosis	4.916469	5.480535
Jarque–Bera	202.4905	276.8436

Figure 1.6   Weekly log returns, exchange rate
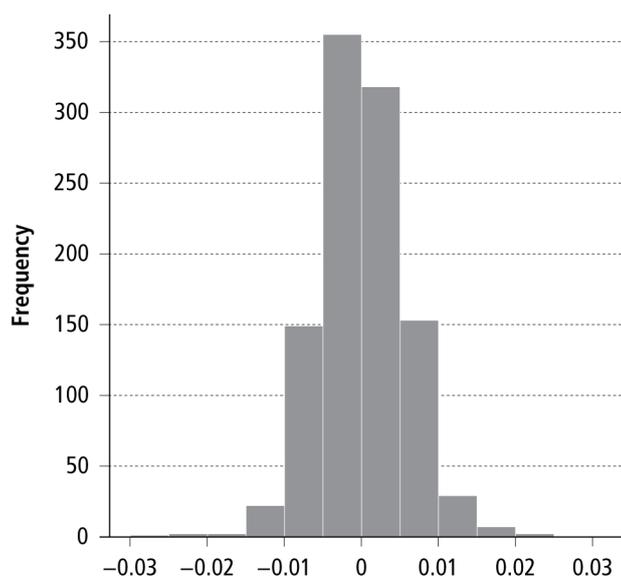
Like stock returns and commodity returns, exchange rate returns exhibit excess kurtosis. The Jarque–Bera test suggests that we can strongly reject the null hypothesis of normality.

Figure 1.7   Weekly log returns, oil price
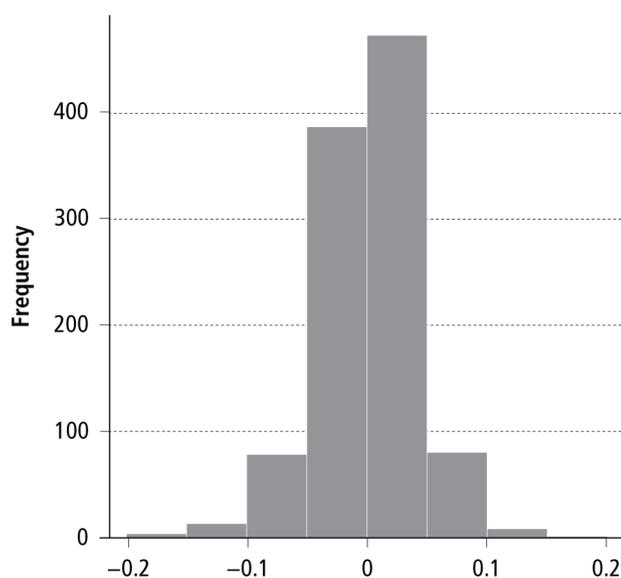
b) Table 1.8 indicates the correlation between the two series is negative *ie* a depreciation of the US currency against other currencies is associated with positive oil returns.

Table 1.8   Weekly log returns, oil price and exchange rate, correlation

	Oil return	Exchange rate return
Oil return	1	−0.235314
Exchange rate return	−0.235314	1

c) The time varying correlation is shown in Figure 1.8.

Figure 1.8   Time-varying correlation, weekly log returns, oil price and exchange rate
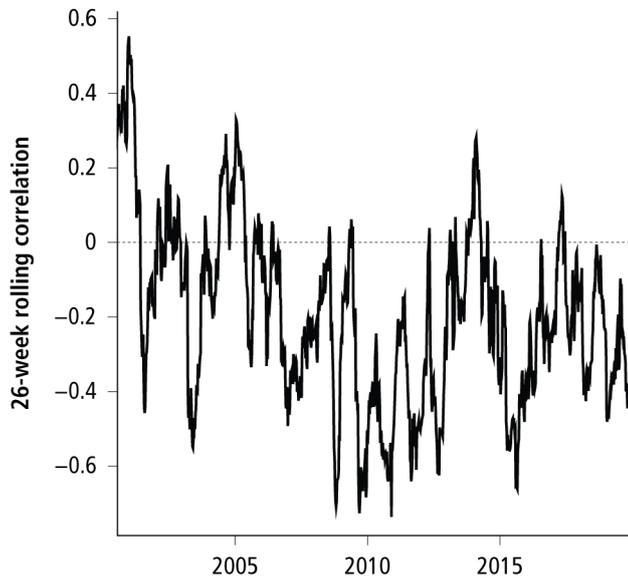

The time varying correlation shows quite clearly that the correlation is far from constant. There are periods of time when the degree of (negative) correlation increased dramatically, then disappeared, and then reappeared again. Given this behaviour, one may wonder whether the two series are correlated at all!

Question 5

This question concerns daily share prices, daily log returns and annualised returns for the FedEx Corporation (FDX).

a) See the Excel file M459_U1_FDX.xls (1997-2003 compatible) for the calculations of the daily log return. (If you have problems downloading the data, the tab-delimited text file M459_U1_FDX.txt contains the date and daily (unadjusted) closing price for FedEx, 2004–2019.)

b) See the Excel file M459_U1_FDX.xls (1997-2003 compatible) for the calculation of the annualised continuously compounded return. The annualised compounded return shown in Figure 1.9 exhibits a pattern of cyclicality.

Figure 1.9   Annualised compounded return, FedEx Corporation
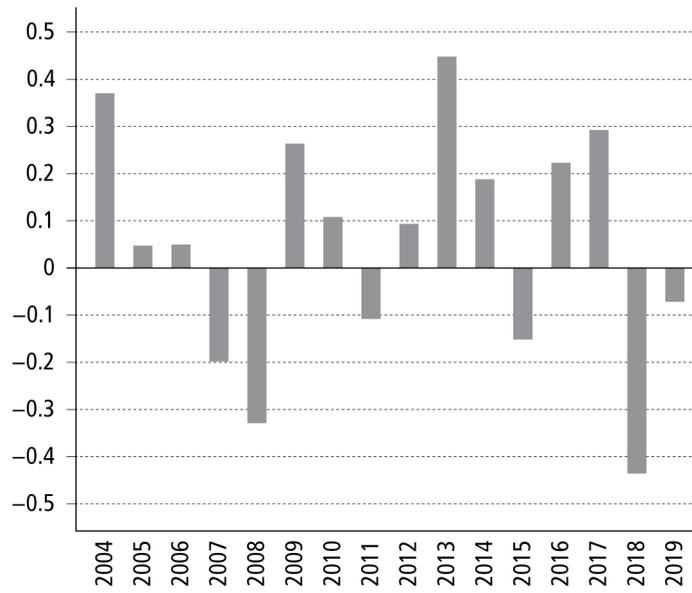
References

Board of Governors of the Federal Reserve System, US (nd) *Trade Weighted U.S. Dollar Index: Broad, Goods (DISCONTINUED) [TWEXB].* [Online]. Federal Reserve Bank of St. Louis. Available from: https://fred.stlouisfed.org/series/TWEXB [Accessed 27 February 2020]

Brooks C (2019) *Introductory Econometrics for Finance*. 4th Edition. Cambridge: Cambridge University Press.

Cont R and P Tankov (2004) *Financial Modelling with Jump Processes.* Boca Raton FL: Chapman and Hall.

Energy Information Administration (nd) [Online]. Available from: http://www.eia.gov [Accessed 27 February 2020]

Foster J (2020) *roll: Rolling Statistics*. R package version 1.1.4. Available from: https://CRAN.R-project.org/package=roll

Jondeau E, S Poon and M Rockinger (2007) *Financial Modelling under Non-Gaussian Distributions*. London: Springer Finance.

Komsta L and F Novomestky (2015) *moments: Moments, cumulants, skewness, kurtosis and related tests*. R package version 0.14. Available from: https://CRAN.R-project.org/package=moments

R Core Team (2019) *R: A Language and Environment for Statistical Computing*. [Online]. R Foundation for Statistical Computing. Vienna, Austria. Available from: https://www.R-project.org/

Wuertz D, T Setz and Y Chalabi (2017) *fBasics: Rmetrics – Markets and Basic Statistics*. R package version 3042.89. Available from: https://CRAN.R-project.org/package=fBasics

Yahoo! (nd) *Finance*. [Online]. Available from: http://finance.yahoo.com [Accessed 24 February 2020]

Zeileis A and G Grothendieck (2005) 'zoo: S3 Infrastructure for Regular and Irregular Time Series'. *Journal of Statistical Software*, 14 (6), 1–27. Available from: http://doi.org/10.18637/jss.v014.i06