Empirical Finance for finance and banking

Robert Sollis

Empirical Finance

for finance and banking

Robert Sollis

WILEY
A John Wiley & Sons, Ltd., Publication

This edition first published 2012

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Sollis, R.
Empirical finance : for finance and banking / Robert Sollis.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-51289-0
1. Finance. 2. Banks and banking. 3. Investments. I. Title.
HG173.S65 2011
332 - dc23

2011034175

A catalogue record for this book is available from the British Library.

Set in 10/13pt Garamond by Laserwords Private Limited, Chennai, India
Printed in Great Britain by TJ International, Padstow, Cornwall

To Clare and Joseph.

Preface

This book focuses on the use of empirical methods in finance. It is intended for students taking Master's degrees in finance and banking, MBA students and early-stage researchers in these areas. Students and researchers in other related areas, such as economics, might also find it to be useful. After the first three introductory chapters, which outline the structure of the book and review econometric and statistical techniques, the remaining chapters are self-contained discussions of a selection of key topics in finance and banking where empirical methods play an important role. Understanding these topics and methods is important for students interested in working as analysts and researchers in financial institutions.

It is becoming increasingly common that students with relatively non-technical undergraduate degrees switch to finance and banking at the Master's degree level. Therefore, while it is impossible to avoid equations in a book on this subject, I have been selective in the amount of mathematics used in the main text and have tried to write in an "easy-to-read" style. Further technical details on relevant mathematical proofs and the statistical techniques used can be found in the many books and journal articles referenced throughout. The choice of topics covered has been influenced by my own university teaching experience at both undergraduate and postgraduate levels in econometrics, finance and statistics. I have also included some material from practical courses on time series forecasting that I have taught for the Royal Statistical Society to professional economists and statisticians. Note, however, that this book is not intended to be an exhaustive text and there are many interesting topics in finance that are not discussed here.

The 2007–2009 global financial crisis highlighted many weaknesses in orthodox financial modelling and risk management, and some of these weaknesses are discussed in this book. However, I do not attempt to develop any new or improved techniques. Some have already been proposed and no doubt many more will be proposed in White Papers, academic journals and books over the next few years. This book focuses only on existing techniques.

I would like to thank Steve Hardman and Nicole Burnett at Wiley for their patience, support and enthusiasm for this project, and Paul Stringer for his diligent copy-editing. I would also like to thank attendees of my Royal Statistical Society workshops as well as colleagues and students at Newcastle University for helpful comments on much of the material in this book. Most of all, I would like to thank my family for their invaluable support.

Contents

Chapter 1

Introduction

1.1 Subject Matter and Structure

Chapters 2 and 3 of this book review the main econometric and statistical techniques employed in empirical finance. Chapters 4 to 9 are self-contained discussions of a selection of key topics in finance where empirical methods play an important role. At the end of all the following chapters there are five test questions to help clarify the topics discussed. The questions vary between purely empirical questions and essay-type questions. The empirical questions for Chapters 2 and 3 utilise simulated data; the empirical questions for_Chapters 4 to 9 require real-world data. As well as the empirical end of chapter questions, in the penultimate section of the main text for Chapters 4 to 9, selected empirical examples are presented that demonstrate many of the empirical methods discussed. Relevant computer programs and a guide to answering the end of chapter questions will be placed on the companion website: **www.wileyeurope.com/college/sollis**. The computer software MATLAB® is used for all calculations in this book.[1] Statistical tables required for some of the empirical examples are included in an appendix at the end of the book.

Chapter 2 introduces the statistical concepts of random variables and random processes before moving on to discuss **time series models** – an important family of statistical models used in many different subject areas including finance. The parameters of a time series model for a financial asset's price or return can be estimated using sample data on its values over time (**time series data**). The estimated model

[1]MATLAB® is a MathWorks product (The MathWorks Inc.: see www.mathworks.com/products/matlab/).

provides important information on the how the price or return randomly evolves over time. Time series models can also be used to compute forecasts of financial variables, which might then be used for financial trading.

Chapter 3 reviews regression modelling. Regression models are the workhorse statistical models for empirical analysis in finance. They can be used to analyse the empirical support for financial theories; they can be employed to compute forecasts of financial variables; they can be used to understand and quantify the risks associated with particular financial assets and combinations of assets. Chapter 3 also covers modelling and forecasting **volatility**. In finance the variance and standard deviation of the return from investing in an asset are used as measures of the asset's risk. The standard deviation is also called the volatility of the return. Accurately modelling and forecasting volatility is particularly important for successful risk management.

Chapter 4 focuses on modern portfolio theory (MPT). MPT has its origins in a seminal journal article on portfolio selection by Markowitz (1952). This article had a significant impact on both financial research and practice. Markowitz (1952) shows formally how the expected return and the variance of the return for a portfolio can be computed as a function of the expected returns, variances and covariances of the returns for the constituent assets. He argues that when constructing portfolios, investors should focus on the performance of different portfolios of assets as measured by combinations of expected portfolio return and portfolio risk, and they should choose from diversified portfolios that maximise expected return for a given level of risk. MPT leads to analytical formulas for optimal asset allocation. The empirical examples in Chapter 4 demonstrate the use of these formulas.

Chapter 5 focuses on asset pricing models and factor models. The capital asset pricing model (CAPM), developed by Sharpe (1964), Lintner (1965) and Mossin (1966), is a theoretical model of the relationship between asset returns and risk under a set of simplifying assumptions. The CAPM is an attractive theoretical model; it is parsimonious and has a clear set of predictions that can be empirically evaluated using statistical techniques; estimated versions of the CAPM can be used for a range of practical tasks. Factor models are a type of asset pricing model that are more general than the CAPM and appear to have greater explanatory power than the CAPM when applied to stock market data. They are widely used by financial institutions to help understand how particular macroeconomic risks affect stock returns. The empirical examples in Chapter 5 include using stock market data to statistically evaluate the empirical support for the CAPM, and estimating and interpreting a factor model.

Chapter 6 focuses on the efficiency of financial markets. In finance the label **efficient market** refers to a financial market (e.g. a stock market, the foreign exchange market) that is informationally efficient, in the sense that current asset prices reflect all relevant information. The **efficient market hypothesis** (EMH) formalised by Fama (1970) states that financial markets are efficient in this way. Chapter 6 begins with a discussion of market efficiency and explains how the EMH can be empirically tested. The EMH implies that **abnormal profits** cannot be generated by trading on forecasts of asset prices or returns computed using historical sample data on prices or returns. Chapter 6 discusses the ability of some particular empirical forecasting techniques to generate abnormal profits. The techniques considered are **econometric forecasting** and **technical analysis**. The empirical examples in Chapter 6 cover testing for stock market efficiency using data on stock market indices for several countries, and forecasting stock index returns using econometric analysis and technical analysis.

Chapter 7 focuses on modelling and forecasting exchange rates. If the foreign exchange market is informationally efficient then certain relationships between current exchange rates, interest rates, forward exchange rates and actual future exchange rates will hold (where the "forward" exchange rate is an exchange rate agreed now for delivery at a future time period). Therefore empirically testing for whether these relationships do hold provides useful empirical evidence on the efficiency of the foreign exchange market. Chapter 7 discusses these relationships and how they can be empirically tested. Chapter 7 also focuses on the relationship between exchange rates and domestic and foreign price levels. In the absence of tariffs, non-tariff barriers and transaction costs, the **law of one price** states that when converted into a common currency the price of a good in different countries should be equal. When applied to the price of a basket of goods (measured using a price index), this law is called the **purchasing power parity** (PPP) hypothesis. The empirical examples in Chapter 7 include empirically testing exchange rate parity conditions and empirically testing the PPP hypothesis. Chapter 7 also considers exchange rate forecasting.

Chapter 8 focuses on modelling and forecasting interest rates, starting with a discussion of the main types of bonds and the traditional formulas for computing bond yields and bond prices. Chapter 8 also introduces an important family of **continuous-time** interest rate models. These models assume that interest rates are time-varying random processes that can be represented using stochastic differential equations. Continuous-time interest rate models can be used to obtain analytical formulas for pricing bonds, bond derivatives, and for describing the term structure of interest rates.

The final topic discussed in Chapter 8 is empirically testing the **expectations hypothesis** of the term structure of interest rates, using econometric tests developed by Campbell and Shiller (1987, 1991). This hypothesis states that the difference between interest rates for bonds with different maturities depends on investors' expectations of future short-term interest rates, where these expectations are formed rationally. If the expectations hypothesis of the term structure of interest rates is correct, the yield curve (a graph of bond yields against time to maturity) contains information on expected future short-term interest rates. Hence a vast amount of research has been undertaken on empirically testing the expectations hypothesis, some of which is discussed in Chapter 8. The empirical examples in Chapter 8 include estimating the parameters of a continuous-time interest rate model, forecasting interest rates and testing the expectations hypothesis using the econometric tests discussed.

Chapter 9 focuses on risk management, in particular the use of statistical techniques to manage **market risk. Value at Risk** (VaR) has for some time been the most popular statistical technique used by banks for managing market risk. VaR is a forecast of a lower quantile of the probability distribution for the change in the value of a portfolio over a particular horizon (e.g. one day). Hence VaR is a monetary measure of downside financial risk. Chapter 9 discusses the main approaches that can be used to compute VaR, including the popular RiskMetrics approach (see for example RiskMetrics, 1996, 2001, 2006). The statistical evaluation of VaR models using **backtesting** is also covered.

Chapter 9 also discusses the role of VaR and backtesting in international financial regulation. Under the Basel II Capital Accord, banks can use VaR to help determine the minimum capital they must hold to cope with potential losses. The performance of their VaR models in backtesting can also have an impact on this minimum capital requirement. In response to the 2007 - 2009 global financial crisis, a new Capital Accord, Basel III, has been proposed and it appears that VaR will continue to play an important role. VaR is also increasingly popular for risk management in the insurance and re-insurance industries. The empirical examples in Chapter 9 demonstrate how to compute VaR for portfolios containing assets and derivatives using the techniques discussed in the main text. An example of backtesting VaR for a portfolio of stocks is also presented.

1.2 Computer Software

I have chosen to use MATLAB® for all calculations in this book because it is widely used in practice by researchers and analysts in the financial sector. Therefore, it is helpful for students intending to pursue a career in the financial sector that they become familiar with MATLAB®. For some of the empirical examples in Chapters 4 to 9, I have included the MATLAB® program used to compute the empirical results (or a portion of the program). The programs used for all of the empirical examples in Chapters 4 to 9 will be placed on the companion website, www.wileyeurope.com/college/sollis, and further details on the programs are given in an appendix to each of those chapters. All of these programs require a registered version of MATLAB® to run. Note also that some of the programs require a registered version of one of the following **toolboxes** to run: Financial Toolbox®, Statistics Toolbox®, System Identification Toolbox®; or the Oxford MFE Toolbox.[2,3] I have tried to make the programs as simple to understand as possible and have favoured obvious ways of computing relevant estimators and test statistics rather than taking shortcuts that might be less clear (but more efficient). The programs are annotated with instructions.

MATLAB® programs have the extension ".m". A program can be opened by opening MATLAB® and then setting the "Current Directory" (or "Current Folder" depending on the version of MATLAB®) on the toolbar to the directory where the program is saved. The contents of this directory should appear in a separate window. Then double click on the program in the Current Directory and it should open in the "Editor" window. To run the program click the "Run" button on the Editor toolbar, or click "Debug" on the main toolbar and then "Run...". The output goes to the "Workspace" and any relevant graphs will appear in windows.

MATLAB® includes many different built-in functions that users can employ in their programs to simplify certain tasks. For example, the `normpdf` function, which is applied using the command `y=normpdf(X,mu,sigma)`, generates the probability density function (PDF) for a normally distributed random variable X, where mu and sigma are the mean and variance. Users can also write their own MATLAB® functions and save them as separate .m files. If a function has more than one output then when the function is applied, the outputs should be written in square brackets, separated by commas, on the left-hand-side: for instance, `[r1,r2,r3]=functionname(.)`. User-defined functions should be saved in the same directory as the program in which they are being used.

1.3 Data

Virtually all of the external data used in this book has been obtained from publicly accessible sources such as Yahoo! Finance, the Federal Reserve Economic Data (FRED) database at the Federal Reserve Bank of St Louis and the Bank of England Statistics database.[4] Therefore readers should be able

[2]Financial Toolbox®, Statistics Toolbox®, and System Identification Toolbox® are MathWorks products (see www.mathworks.com/products/matlab/).

[3]For more information on the Oxford MFE Toolbox, see http://www.kevinsheppard.com/wiki/MFE_Toolbox. I would like to thank Kevin Sheppard for allowing me to use the Oxford MFE Toolbox for computing several results reported in this book.

[4]www.finance.yahoo.com; www.research.stlouisfed.org/fred2/; www.bankofengland.co.uk/statistics

to download these data and use the MATLAB® programs to obtain similar results to many of the results reported in this book.[5] Inputting data into MATLAB® programs is straightforward. For example, to input data from a single Microsoft Excel spreadsheet called "filename.xls", make sure the Excel file is saved in the same directory as the MATLAB® program being used, and then in the program, use the command `Y=xlsread('filename')`, where "filename" is the name of the Excel file without the .xls extension (this assumes the .xls file is an Excel 97-2003 Workbook). The data will be contained in the vector or matrix Y. Further details on the data used in the empirical examples are given in the appendixes to each chapter.[6]

1.4 References

Campbell, J.Y. and R.J. Shiller (1987) Cointegration and tests of present value models, *Journal of Political Economy*, 95, 1062-1088.

Campbell, J.Y. and R.J. Shiller (1991) Interest rate spreads and interest rate movements: a bird's eye view, *Review of Economic Studies*, 58, 495-514.

Fama, E.F. (1970) Efficient capital markets: a review of theory and empirical work, *Journal of Finance*, 25, 383-417.

Lintner, J. (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics*, 47, 13-37.

Markowitz, H.M. (1952) Portfolio selection, *Journal of Finance*, 7, 77-91

Mossin, J. (1966) Equilibrium in a capital asset market, *Econometrica*, 34, 768-783.

RiskMetrics: Technical Document (1996) New York: JP Morgan/Reuters.

RiskMetrics, Return to: The Evolution of a Standard (2001) New York: RiskMetrics Group.

RiskMetrics 2006 Methodology (2006) New York: RiskMetrics Group.

Sharpe, W.F. (1964) Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*, 19, 425-442.

[5] Thomson Reuters Datastream, which requires a subscription, has been used for a very small proportion of the data employed.

[6] Unfortunately, I cannot guarantee that the exact data used to compute the results in this book will still be available after the book's publication. Whilst every attempt has been made to ensure that the data, text and the MATLAB® programs are free of errors, they cannot be guaranteed to be error-free. Where possible any corrections found subsequent to the book's publication will be placed on the website www.wileyeurope.com/college/sollis. Note also that while the internet links mentioned in this book are currently working, it cannot be guaranteed that they will all continue to work after the book's publication.

Chapter 2

Random Variables and Random Processes

2.1 Introduction

A **random variable** can be formally defined as a variable that has an uncertain numerical value that is determined by the outcome of a random experiment. The uncertainty of random variables can be quantified using the concept of probability. A simple example of a random variable often used in statistics textbooks is the score from rolling a die. We know that for this random variable each of the six possible values has a probability of 1/6, hence they are equally likely. Random variables exist in all aspects of life: for example, the amount of rainfall by the end of today; the waiting time when we arrive at a bus stop; today's closing value for a stock market index.

A **random process** (or **stochastic process**) is a natural extension of the concept of a random variable, being a series of random variables ordered in a particular way. If the ordering is by time then data on the actual values of the random process are called **time series data**, or a **time series**. Many of the variables of interest in finance (e.g. the future value of an exchange rate, interest rate, stock market index) are random variables. Over a period of time these variables can be thought of as random processes, hence this chapter reviews some key econometric and statistical