

Publishers' page

Publishers' page

Publishers' page

Publishers' page

## Preface

This book is designed as an easily accessible and useful practical guide to getting started with the statistical software tool R for applications in Finance & Economics. There are various open source or commercial software packages available for statistical computing. R is not just used for statistical modelling in business analysis, socio-economic analysis etc., but is also gaining popularity among students and researchers in academia in the fields of Economics and Finance. R is currently one of the most popular software packages available. Its popularity is partly driven from the fact that it is a free resource for users and its great versatility in a variety of applications. The unique strength of R is that it has a multidiscipline base with packages being supplied from a wide spectrum of disciplinary contributors from science, mathematics, computing and operational research, plus a broad spectrum of the social sciences.

The design of this book has been greatly influenced by introductory courses in R that have been run by the authors both at Edith Cowan University in Western Australia and in the Edinburgh Business School at the University of Edinburgh in Scotland. In these courses it was found that beginners have considerable difficulty in setting up R and Rstudio, in the importation of data, and in the writing of code. The book addresses these issues in a simple to understand, comprehensive way that the more advanced books available on R do not. Yet these courses were run for postgraduate students and university staff members, but most seemed to struggle at first.

This book is specifically targeted towards early career researchers or research students who have some exposure to basic statistical methods used in quantitative research. Although the first author of the book has over eight years of computer programming experience and over four years of experi-

ence with R, the text is designed in such a way that people with even limited knowledge of computer programming can benefit. The second author's experience with computing stretches back to working with mainframes, in the late 1970's, at the University of Edinburgh in Scotland, before switching to personal computers in the 1980's at the University of Western Australia. The bulk of his more recent research experience in finance and econometrics was with menu-driven econometrics programs. He did not have exposure to Matlab and subsequently R until he was in his sixties. This reinforced his awareness of the importance of appropriate data import, formatting, and manipulation. The intention is to make this book useful for people with computing experience as little as recording a macro in MS-Excel up to writing codes in C++/MATLAB/R/S or other programming languages.

"R in Finance and Economics-A Beginner's Guide" provides an introduction to the statistical software R and its application with an empirical approach in finance and economics. We believe that the methods and procedures outlined in the book are likely to be of use to researchers from various fields beyond the ones of focus here. The major objective of the text is to get the reader started with R by making the reader familiar with the concepts and techniques required. We do not develop any R package via this book but use various packages which are available and are useful for research to sustain the applied nature of this text. The easy to follow text, provides examples and applications of econometric and statistical analysis techniques to central methods in economics and finance.

The book cover topics ranging from introduction to R environment to advanced topics like multivariate Vine Copulas. Chapter-1 provides an introduction to R and overview of RStudio interface which is the primary R GUI/IDE used in the book. Chapter-2 discusses some of the most common data types and structures used in statistical analysis using R followed by various methods to import, export and pre-process external data in Chapter-3. Chapter-4 provides a brief overview of some important programming concepts including program control flow and creating functions. Almost all empirical problems in Finance and Economics require a robust preliminary data exploration with summary/descriptive statistics and exploratory plots which is discussed in Chapter-6. R is gaining popularity among data analytics community not just because of its statistical prowess but also due to advanced data visualisation capabilities. Chapter-7 discusses various methods available to create various types of graphs in R including a short introduction to the *ggplot2* package.

Chapter-7 and chapter-8 present Regression Analysis; one of the most

widely used tools in quantitative methods. Chapter-7 provides an overview of linear regression (OLS) and Quantile Regression along with an empirical example illustrating the use of linear regression to evaluate multifactor models in finance. Panel Regression and Logistic Regression are two other statistical tools used in empirical finance and economics along with OLS and Quantile Regression. We discuss these two regression tools in Chapter-8 and provide an illustrative example on how to use Panel Regression in applied economic problems. After regression analysis, chapter-9 provides an overview of widely used Autogressive and conditional autoregressive time series models in Financial econometrics. We also discuss an example on modelling and forecasting Value at Risk (VaR) using GARCH model in R.

Chapter-9 introduces Extreme Value Theory (EVT), which is used in risk quantification in finance and actuarial studies. We discuss EVT for financial risk and EVT for bivariate tail dependence along with an empirical example. The last chapter, chapter-10 of the book focusses on multivariate dependence using Copulas. Chapter-10 presents some widely used bivariate copulas and also introduces the recently developed multivariate Vine-Copulas, which are more flexible than bivariate copulas. The example in chapter-10 demonstrate how to estimate portfolio VaR using Vine-Copulas.

The book is designed with due consideration for the difficulties which a beginner programmer can face. The text contains illustrative examples and code is provided in the text itself. The examples are reproducible using the open source R and Rstudio programs. The following are the key strengths of this book:

- (1) Short and Concise: The book will follow a minimalistic approach in teaching how to use R. The discussions will focus primarily on easy to follow methods and which are mostly used by beginner and intermediate users. There will be a discussion for more advanced users but that will be included in separate chapters.
- (2) Easy to follow: The contents in the book will be kept to the level of a beginner or intermediate R user so it will be easily understood also by early career researchers and research students (PhD, Masters by Research etc)
- (3) Examples: Every method explained in the book has a detailed following example with tips and tricks. The code used for the example is given in the book itself so that the user can copy it to reproduce the results (when using an ebook).

- (4) Reproducible Codes: R codes (scripts) will be provided for all the chapters in the book.
- (5) Practical applications: Research based examples are included which will include the most widely used statistical and econometric applications such as linear regression, econometrics (GARCH etc). This will enable beginners to get started and produce sound results. These examples will include applications to similar problems as addressed in the research papers by the authors of this book and thus will be current and state of the art.
- (6) Reproducible Research: The examples and case studies in the book will follow step by step procedure which will enable to reproduce them. The data is provided with the book on its companion website. **Insert url here**

To summarise, the book demonstrates the latest research methods in finance and economics with applications featuring linear regression, quantile regression, panel regression, econometrics, Extreme Value methods and copulas using a range of data sets and examples.

*A. K. Singh & D. E. Allen*



# Contents

<i>Preface</i>	i
1. Introduction	1
1.1 Why should we learn R? . . . . .	2
1.2 R-Getting Started . . . . .	3
1.2.1 Installing R on Windows . . . . .	3
1.2.2 RStudio- A better way to do run R . . . . .	4
1.2.3 Installing RStudio for Windows . . . . .	5
1.2.4 RStudio GUI/IDE . . . . .	5
1.3 Proxy Setup . . . . .	5
1.4 R Packages . . . . .	8
1.4.1 Installing Packages . . . . .	8
1.5 R core packages. . . . .	9
1.6 Task Views in R-Introduction & Installation . . . . .	11
1.7 Getting Help . . . . .	12
1.7.1 Getting help from the web . . . . .	14
1.8 Summary . . . . .	14
2. Data Objects in R	17
2.1 Introduction . . . . .	17
2.2 Data Types . . . . .	17
2.2.1 Double . . . . .	18
2.2.2 Integer . . . . .	18
2.2.3 Complex . . . . .	19
2.2.4 Logical . . . . .	19
2.2.5 Character . . . . .	19

2.2.6	Factor . . . . .	20
2.2.7	Date & Time . . . . .	20
2.2.8	Missing Data in R . . . . .	22
2.3	Data Structures in R . . . . .	22
2.3.1	Vector . . . . .	23
2.3.2	Matrices . . . . .	23
2.3.3	Arrays . . . . .	27
2.3.4	Data Frames . . . . .	27
2.3.5	Lists . . . . .	29
2.4	Summary . . . . .	31
3.	Data Handling in R . . . . .	33
3.1	Introduction . . . . .	33
3.2	Importing and Exporting Tabular Data in R . . . . .	33
3.2.1	Reading Data from a Text File . . . . .	33
3.2.2	Reading Data from CSV File . . . . .	36
3.2.3	Reading Data from Excel Files . . . . .	37
3.2.4	Reading Data from Databases . . . . .	37
3.2.5	Reading from Data Files from other Statistical Systems . . . . .	37
3.3	Data Preprocessing in R . . . . .	41
3.3.1	Extracting Data . . . . .	41
3.3.2	Combining Data Frames . . . . .	43
3.3.3	Sub-setting and Logical Data Selection . . . . .	45
3.4	Summary . . . . .	47
4.	R Programming & Control Flow . . . . .	49
4.1	Introduction . . . . .	49
4.2	Control Flow . . . . .	50
4.2.1	if-else Conditional Statements . . . . .	50
4.2.2	Using <code>switch</code> . . . . .	51
4.2.3	Loops . . . . .	52
4.3	Functions in R . . . . .	54
4.4	Summary . . . . .	56
5.	Data Exploration . . . . .	57
5.1	Introduction . . . . .	57
5.2	Summary Statistics . . . . .	57

*Contents*

vii

5.3	Example-Descriptive Statistics of Stock Returns . . . . .	65
5.3.1	Introduction . . . . .	65
5.3.2	Importing the Data . . . . .	66
5.3.3	Some basic plots . . . . .	77
5.4	Summary . . . . .	81
6.	Graphics in R . . . . .	83
6.1	Introduction . . . . .	83
6.2	Basic Plots in R . . . . .	84
6.3	Exporting Graphics . . . . .	97
6.4	R Graphical Parameters . . . . .	98
6.5	Introduction to ggplot2 . . . . .	103
6.5.1	Getting started with <code>qplot</code> . . . . .	103
6.5.2	Layered graphics using <code>ggplot</code> . . . . .	105
6.6	Transforming data from wide to long format . . . . .	112
6.7	Summary . . . . .	114
7.	Regression Analysis-I . . . . .	115
7.1	Introduction . . . . .	115
7.2	OLS . . . . .	115
7.3	Quantile Regression . . . . .	121
7.3.1	Estimating QR . . . . .	123
7.4	Example-Fama French Factor Model and Multiple Regression . . . . .	128
7.4.1	Introduction . . . . .	128
7.4.2	Data . . . . .	130
7.4.2.1	Data Preprocessing . . . . .	130
7.4.3	OLS Regression Analysis of the Fama-French Three Factor Model . . . . .	135
7.4.4	Quantile Analysis of the Fama-French Three Factor Model . . . . .	138
7.5	Summary . . . . .	143
8.	Regression Analysis-II . . . . .	145
8.1	Introduction . . . . .	145
8.2	Panel Data Linear Regression . . . . .	145
8.2.1	Fixed and Random effects using the <i>plm</i> package . . . . .	147
8.2.1.1	Fixed Effect Estimation . . . . .	148
8.2.1.2	Random Effect Estimation . . . . .	150

viii	<i>R in Finance &amp; Economics-A Beginner's Guide</i>	
	8.2.1.3 Panel or OLS . . . . .	151
	8.2.1.4 Fixed Effect or Random Effect . . . . .	153
8.3	Logistic Regression . . . . .	153
8.4	Example-Economic Growth and Unemployment A Panel Analysis . . . . .	156
	8.4.1 Data and Methodology . . . . .	157
	8.4.2 Data Preprocessing . . . . .	157
	8.4.3 Linear Panel Regression Analysis . . . . .	160
8.5	Summary . . . . .	166
9.	Time Series Analysis	167
9.1	Introduction . . . . .	167
9.2	Time Series-Some Properties . . . . .	168
	9.2.1 Stochastic Process . . . . .	168
	9.2.2 Stationarity . . . . .	169
	9.2.3 Autocorrelation Function (ACF) . . . . .	171
	9.2.4 White noise . . . . .	173
9.3	Autoregressive Moving Average Model (ARMA) . . . . .	174
	9.3.1 Fitting an ARMA model . . . . .	175
9.4	Volatility Modelling using GARCH . . . . .	181
	9.4.1 Fitting a GARCH(1,1) model using the <i>rugarch</i> package . . . . .	182
9.5	Example-Modelling and Forecasting daily VaR using GARCH	190
	9.5.1 Data & Methodology . . . . .	191
	9.5.2 VaR Forecasts . . . . .	193
9.6	Summary . . . . .	198
10.	Extreme Value Theory Modelling	199
10.1	Introduction . . . . .	199
10.2	EVT and Financial Risk Modelling . . . . .	199
	10.2.1 GPD and POT Method . . . . .	201
	10.2.1.1 Sample Mean Excess Plot . . . . .	202
	10.2.1.2 Estimation of GPD . . . . .	205
	10.2.1.3 VaR and Expected Shortfall . . . . .	207
10.3	EVT and Tail Dependence . . . . .	211
	10.3.1 Measures of Tail Dependence . . . . .	212
	10.3.1.1 Asymptotic Dependence - The Conventional Approach . . . . .	215

*Contents*

ix

10.3.1.2	Asymptotic Independence- An Alternative Measure of Dependence. . . . .	215
10.3.2	Estimating $\chi$ and $\bar{\chi}$ - Nonparametric Method . . .	218
10.4	Example-EVT VaR-A Dynamic Approach . . . . .	221
10.4.1	Data & Methodology . . . . .	222
10.4.2	VaR Forecasts using Dynamic EVT Model . . . . .	224
10.5	Summary . . . . .	226
11.	Introduction to Multivariate Analysis using Copulas . . . . .	227
11.1	Introduction . . . . .	227
11.2	Copula . . . . .	228
11.2.1	Types of Copula . . . . .	229
11.2.1.1	Elliptical Copulae . . . . .	229
11.2.1.2	Archimedean Copulae . . . . .	235
11.2.2	Copula Selection . . . . .	240
11.3	Multivariate Vine Copulas . . . . .	241
11.3.1	Estimating R-Vine Copulas Using R . . . . .	242
11.4	Example- Portfolio VaR Estimation Using Vine Copula . . . . .	249
11.4.1	Data & Methodology . . . . .	249
11.4.2	VaR forecasts using R . . . . .	250
11.5	Summary . . . . .	255
	<i>Bibliography</i> . . . . .	257
	<i>Index</i> . . . . .	265