Preface

Like its earlier incarnation in S-Plus written over 10 years ago, this book is a polished version of the lecture notes written for a one-semester junior statistics course offered to the undergraduate students majoring in the Department of Operations Research and Financial Engineering and a core course of the Master's program of the Bendheim Center for Finance at Princeton University.

The common goal of both courses is to introduce students to modern data analysis used in the financial industry. The prerequisites are minimal, though students are expected to have already taken a basic introductory statistics course. Elementary notions of random variables, expectation, and correlation are taken for granted, and earlier exposure to statistical inference (estimation, tests, and confidence intervals) is assumed. It is also expected that the students are familiar with a minimum of linear algebra as well as vector and matrix calculus. However, all the background concepts and results necessary for the comprehension of the material presented in the book (as well as the solutions of the homework problems) are recalled before they are used or needed.

By choice, the courses are both computational and mathematical in nature. Most problems considered are formulated in a rigorous manner. Mathematical facts are motivated by applications, stated precisely, justified at an intuitive level, but essentially never proven rigorously. The emphasis is more on the relevance of concepts and on the practical use of tools, rather than on their theoretical underpinnings.

I chose to illustrate concepts, manipulate data, build models, and implement estimation and prediction procedures in the R computer environment. For this reason the text is sprinkled with the R commands needed to perform the analyses and produce the plots. The first incarnation of this text was written for S-Plus on Windows platforms. The growing presence of Mac computers in the classrooms and the ease with which Linux, Windows, and MacOS versions of R can be downloaded and installed at no cost were influential in my decision to switch from S-Plus to R. To

my surprise, the port was not as seamless as I originally expected. It took me several years to complete the transition, and in the process, an entire chapter, several sections, and a large number of examples and problems have been added to the original contents.

The text is divided into three parts. Part I, *Data Exploration, Estimation, and Simulation*, introduces heavy tail distributions and dependence concepts for multivariate data, with an emphasis on the practical applications of copulas. Part II, *Regression*, introduces the students to modern regression with an emphasis on robustness and nonparametric techniques. Part III, *Time Series and State Space Models*, is concerned with the theories of time series and state space models, including filtering applications.

CONTENTS

Part I comprises three chapters. Chapter 1 begins with a review of the classical probability distributions encountered throughout the book and presents the exploratory data analysis techniques (histograms, kernel density estimators, Q-Q plots, etc.) used to handle empirical samples. As a preparation for many analyses and problems based on random simulations, the chapter concludes with a discussion of Monte Carlo computations.

Chapter 2 is devoted to the detection, estimation, and simulation of *heavy tail* distributions already showcased in the first chapter. It contains more statements and discussions of theoretical results than most other chapters, the reason being the desire to provide insight in the estimation and simulation algorithms implemented in the `R` library `Rsafd` used in the practical applications. Illustrative examples are used to demonstrate the impact of the presence of heavy tails on the computations of measures of risk such as value at risk (also known as VaR).

The third chapter is concerned with multivariate distributions and the various concepts of dependence. We review the classical measures of correlation, demonstrate the shortcomings of the Pearson correlation coefficient, and study the notion of copula, and the important role it plays when the marginal distributions have heavy tails, both in the bivariate case and in the high dimensional case. We learn how to detect unusual dependencies, estimate and simulate them, and bring this expertise to bear on the analysis of large portfolios of financial instruments including stocks and credit derivatives. The chapter concludes with a complete discussion of principal component analysis and two applications to the fixed income markets.

Part II is concerned with regression, and it is naturally divided into two chapters: the first devoted to parametric methods and the second to nonparametric ones. Chapter 4 deals with linear models and their applications. The notion of robustness is introduced, and examples are used to illustrate the differences between least squares and least absolute deviations regressions. Applications of linear models include

polynomial and more general nonlinear regressions. We use financial examples throughout, and we analyze the term structure of interest rates in detail. Chapter 5 is concerned with nonparametric regression. We compare the properties of data smoothers for univariate data, and we analyze in detail the multivariate kernel regression. For larger dimensions, we use projection pursuit. Examples of energy forward curves and intraday S&P 500 futures tick data are given. The last part of this chapter is devoted to the use of semi-parametric and nonparametric methods in option pricing. We demonstrate the implementation of modern regression techniques as pricing alternatives to the classical Black-Scholes pricing formula.

The first chapter of Part III is devoted to the classical linear models for time series and to the idiosyncrasies of the `R` objects and methods included in the library `Rsafd` for the sole purpose of their analyses. We discuss autoregressive and moving-average models, and we give examples of their use in practice. The main application is the analysis of temperature data. Even if it may not appear to be much of a financial application at first, we recast this analysis in the framework of financial risk management via a thorough discussion of the market of weather derivatives. We give practical examples to illustrate the use of the statistical techniques introduced in this chapter to the control of these financial instruments.

In the following two chapters, we turn to the analysis of partially observed state space systems. Chapter 7 deals with linear models and the classical Kalman filter. For illustration purposes, we study two financial applications, one related to an extension of the CAPM model and a second dealing with the analysis of quarterly company earnings. Chapter 8 is devoted to the analysis of nonlinear time series. We first consider the natural generalizations of the linear time series models, and we provide an extensive review of the theory and the practice of the famous ARCH and GARCH models. We also consider models from continuous time finance through their discretized forms. A special section is devoted to the use of scenarios for economic modeling. We concentrate on scenarios for a stock index and the short and long interest rates. These scenarios are of crucial importance in risk management where they are used as input to large stochastic optimization programs. Finally, we revisit the theory presented in the case of partially observed linear systems, and we extend the filtering paradigm to nonlinear systems with the help of recent advances in Monte Carlo techniques and the so-called particle filters. We give several applications of this material, including the estimation of stochastic volatility and commodity convenience yield.

Each chapter contains a problem section. Most practical problems are rooted in financial applications. Each problem is preceded by one or several symbols Ⓔ, Ⓢ, and/or Ⓣ intended as hints suggesting if it is of an empirical, simulation, and/or theoretical nature. Chapters end with Notes and Complements sections that include complements and bibliographic references for the readers interested in acquiring a deeper understanding of the topics of that chapter. The book ends with an appendix and a suite of indexes. The appendix contains the text of an introductory session to `R` intended to help the reader unfamiliar with `R` get started and to a crash course on Black-Scholes option pricing theory used in several chapters.

The code and the data sets used in the text and the problems are contained in the library `Rsafd` developed as a companion to the book. It can be downloaded from the URL:

`http://www.princeton.edu/~rcarmona`

This web page will be updated regularly, with corrections, complements, new data sets, code updates, etc.

ACKNOWLEDGMENTS

First and foremost, I want to thank all the students who suffered painfully through early versions of the course and primitive drafts of the lecture notes. Their patience and encouragements helped fine-tune the course.

My interest in computational statistics was sparked over 20 years ago by two dear friends: Anestis Antoniadis and Jacques Berruyer. Time and distance pulled us apart, but what their collaboration taught me will remain with me forever. The library implementation of heavy tail estimation and two-dimensional copulas is based on original code of Julia Morrison. It was a real pleasure to work with her on the development of the `S-Plus` library `EVANESCE`. I am very grateful for this experience.

As for the `S-Plus` version of the book, I want to thank my wife Debra for tolerating my insane working habits and my three wonderful daughters Stephanie, Chelsea, and Chanel for their limitless patience and unconditional love. I may not deserve it, but I am definitely proud of it.

Princeton, NJ, USA René Carmona

http://www.springer.com/978-1-4614-8787-6

Statistical Analysis of Financial Data in R

Carmona, R.

2014, XVII, 588 p. 187 illus., 37 illus. in color.,
Hardcover
ISBN: 978-1-4614-8787-6