
Preface

This book grew out of lectures notes written for a one-semester junior statistics course offered to the undergraduate students majoring in the Department of Operations Research and Financial Engineering at Princeton University. Tidbits of the history of this course will shed light on the nature and spirit of the book.

The purpose of the course is to introduce the students to modern data analysis with an emphasis on a domain of application that is of interest to most of them: financial engineering. The prerequisites for this course are minimal, however it is fair to say that all of the students have already taken a basic introductory statistics course. Thus the 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.

Because of my background, the course is 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 *S-Plus* computer environment. For this reason an introduction to *S* and *S-Plus* (reproduced in appendix) is offered at the beginning of the course each semester, and many lectures are sprinkled with the *S* commands needed to perform the analyses discussed in class. The first two incarnations of this course were using *S-Plus* on Unix platforms and not all the students were able to cope with the steep learning curve. Moreover, the two textbooks used for the class did not seem to be of very much help to the students. So I decided

to prepare lecture notes focused on the material covered in class, and to switch to Windows in order to work with a friendlier implementation of S.

The present manuscript is a polished version of the class notes. It is divided into three parts. Part I, *Exploratory Data Analysis*, reviews the most commonly used methods of statistical data exploration. Part II, *Regression*, introduces the students to modern regression with an emphasis on robustness and non-parametric techniques. Part III, *Time Series and State Space Models*, is concerned with the theories of time series and of state space models.

Contents

Part I is a patchwork of many exploratory data analysis techniques. It begins with a discussion of various methods of density estimation, including histograms and kernel density estimators. Since the emphasis of the course is on financial applications, the notion of *heavy tail* is immediately showcased with examples. A good part of the first chapter is concerned with the practical estimation of heavy tailed distributions, their detection, their estimation and their simulation. We use the statistical concept of percentile to introduce the notion of value-at-risk so important in the financial industry, and we demonstrate its use on a couple of illustrative examples. The second chapter is concerned with multivariate distributions and the various concepts of dependence. We study the classical correlation coefficients, but we also spend a good amount of time understanding the notion of copula, and the important role it plays when the marginal distributions have heavy tails. As in the univariate case, we learn how to detect unusual dependencies, to estimate them, and to simulate them. We also give a complete discussion of principal component analysis and illustrate its power on two applications to 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 non-parametric ones. Chapter 3 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 4 is concerned with nonparametric regression. We compare the properties of data smoothers for univariate data, and we analyze in detail the multivariate kernel regression and density estimation for intermediate values of the dimension. For large values of the dimension we consider projection pursuit. To illustrate, we analyze energy forward curves and intra-day tick data on S&P 500 futures contracts. The last part of this chapter is devoted to a demonstration of the use of semi-parametric and nonparametric methods in option pricing. We review the derivation of the classical Black-Scholes pricing formula, we illustrate its shortcomings, and we walk the reader through the implementation of modern regression techniques as pricing alternatives. The actual implementations are done on liquid S&P 500 futures option data.

The first chapter of Part III is devoted to the classical linear models for time series, and to the idiosyncrasies of the *S-Plus* objects and methods needed to fit them. We discuss auto regressive and moving-average models, and we give examples of their use in practice. The main application of the material of this chapter is concerned with 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 booming market of weather derivatives. We give practical examples to illustrate the use of the statistical techniques introduced in this chapter to the pricing of these new financial instruments.

In the following two chapters, we turn to the analysis of partially observed state space systems. Chapter 6 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 7 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. We give several applications of this material, including to the estimation of stochastic volatility and commodity convenience yield.

Each chapter contains a problem section. Most problems are of a financial nature. They are preceded with symbols (E), (S), and/or (T) to indicate if they are of an empirical, simulation, and/or theoretical nature. Each chapter ends with a section called Notes & Complements that includes bibliographic references which can be used by readers interested in acquiring a deeper understanding of the topics of that chapter. The book ends with two appendices as a suite of indexes. Appendix A contains the introductory session set up to initiate the students to *S-Plus* at the beginning of the course, and Appendix B gives information on how to download the library EVANESCE and the home-grown functions used in the text, as well as the data sets used in the text and in the problems.

The code together with the data used in the text can be downloaded from the author web page at the URL:

<http://www.princeton.edu/~rcarmona/safd/>

This web page will be updated regularly, and corrections, complements, new data sets, updates, etc., will be posted frequently.

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 their encouragement helped me persevere, and over the years, figure out the recipe for the form and the content of the course. I feel guilty to have used them as guinea pigs, but I am glad that the process finally converged. My Chairman Erhan Çinlar trusted me with this course, and gave me total freedom to reshape it. What seemed like foolishness to some, may have been great insight with what needed to be done. I am grateful for his confidence and his relentless encouragements.

My interest in computational statistics was sparked over twenty 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 first part of the book would not have been possible without Julia Morrison's contribution. 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. I also want to thank Yacine Ait-Sahalia for enlightening discussions on nonparametric asset pricing. I am also indebted to Cliona Golden for a superb job proofreading an early version of the manuscript. Finally, 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 sure am proud of it.

René Carmona
Princeton, N.J.

November 4, 2003

Statistical Analysis of Financial Data in S-Plus

Carmona, R.

2004, XVI, 455 p. 138 illus., Hardcover

ISBN: 978-0-387-20286-0