Preface

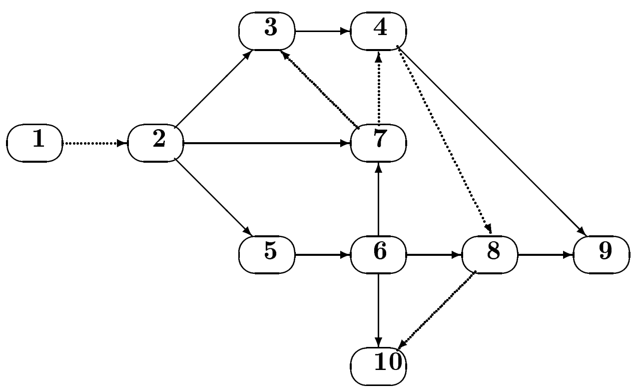
Among many exciting developments in statistics over the last two decades, nonlinear time series and data-analytic nonparametric methods have greatly advanced along seemingly unrelated paths. In spite of the fact that the application of nonparametric techniques in time series can be traced back to the 1940s at least, there still exists healthy and justified skepticism about the capability of nonparametric methods in time series analysis. As enthusiastic explorers of the modern nonparametric toolkit, we feel obliged to assemble together in one place the newly developed relevant techniques. The aim of this book is to advocate those modern nonparametric techniques that have proven useful for analyzing real time series data, and to provoke further research in both methodology and theory for nonparametric time series analysis.

Modern computers and the information age bring us opportunities with challenges. Technological inventions have led to the explosion in data collection (e.g., daily grocery sales, stock market trading, microarray data). The Internet makes big data warehouses readily accessible. Although classic parametric models, which postulate global structures for underlying systems, are still very useful, large data sets prompt the search for more refined structures, which leads to better understanding and approximations of the real world. Beyond postulated parametric models, there are infinite other possibilities. Nonparametric techniques provide useful exploratory tools for this venture, including the suggestion of new parametric models and the validation of existing ones.

In this book, we present an up-to-date picture of techniques for analyzing time series data. Although we have tried to maintain a good balance among methodology, theory, and numerical illustration, our primary goal is to present a comprehensive and self-contained account for each of the key methodologies. For practical relevant time series models, we aim for exposure with definition, probability properties (if possible), statistical inference methods, and numerical examples with real data sets. We also indicate where to find our (only our!) favorite computing codes to implement these statistical methods. When soliciting real-data examples, we attempt to maintain a good balance among different disciplines, although our personal interests in quantitative finance, risk management, and biology can be easily seen. It is our hope that readers can apply these techniques to their own data sets.

We trust that the book will be of interest to those coming to the area for the first time and to readers more familiar with the field. Application-oriented time series analysts will also find this book useful, as it focuses on methodology and includes several case studies with real data sets. We believe that nonparametric methods must go hand-in-hand with parametric methods in applications. In particular, parametric models provide explanatory power and concise descriptions of the underlying dynamics, which, when used sensibly, is an advantage over nonparametric models. For this reason, we have also provided a compact view of the parametric methods for both linear and selected nonlinear time series models. This will also give new comers sufficient information on the essence of the more classical approaches. We hope that this book will reflect the power of the integration of nonparametric and parametric approaches in analyzing time series data. The book has been prepared for a broad readership—the prerequisites are merely sound basic courses in probability and statistics. Although advanced mathematics has provided valuable insights into nonlinear time series, the methodological power of both nonparametric and parametric approaches can be understood without sophisticated technical details. Due to the innate nature of the subject, it is inevitable that we occasionally appeal to more advanced mathematics; such sections are marked with a "*". Most technical arguments are collected in a "Complements" section at the end of each chapter, but key ideas are left within the body of the text.

The introduction in Chapter 1 sets the scene for the book. Chapter 2 deals with basic probabilistic properties of time series processes. The highlights include strict stationarity via ergodic Markov chains (§2.1) and mixing properties (§2.6). We also provide a generic central limit theorem for kernel-based nonparametric regression estimation for α-mixing processes. A compact view of linear ARMA models is given in Chapter 3, including Gaussian MLE (§3.3), model selection criteria (§3.4), and linear forecasting with ARIMA models (§3.7). Chapter 4 introduces three types of parametric nonlinear models. An introduction on threshold models that emphasizes developments after Tong (1990) is provided. ARCH and GARCH models are presented in detail, as they are less exposed in statistical literature. The chapter concludes with a brief account of bilinear models. Chapter 5 introduces the nonparametric kernel density estimation. This is arguably the simplest problem for understanding nonparametric techniques. The relation between "localization" for nonparametric problems and "whitening" for time series data is elucidated in §5.3. Applications of nonparametric techniques for estimating time trends and univariate autoregressive functions can be found in Chapter 6. The ideas in Chapter 5 and §6.3 provide a foundation for the nonparametric techniques introduced in the rest of the book. Chapter 7 introduces spectral density estimation and nonparametric procedures for testing whether a series is white noise. Various high-order autoregressive models are highlighted in Chapter 8. In particular, techniques for estimating nonparametric functions in FAR models are introduced in §8.3. The additive autoregressive model is exposed in §8.5, and methods for estimating conditional variance or volatility functions are detailed in §8.7. Chapter 9 outlines approaches to testing a parametric family of models against a family of structured nonparametric models. The wide applicability of the generalized likelihood ratio test is emphasized. Chapter 10 deals with nonlinear prediction. It highlights the features that distinguish nonlinear prediction from linear prediction. It also introduces nonparametric estimation for conditional predictive distribution functions and conditional minimum volume predictive intervals.



The interdependence of the chapters is depicted above, where solid directed lines indicate prerequisites and dotted lines indicate weak associations. For lengthy chapters, the dependence among sections is not very strong. For example, the sections in Chapter 4 are fairly independent, and so are those in Chapter 8 (except that §8.4 depends on §8.3, and §8.7 depends on the rest). They can be read independently. Chapter 5 and §6.3 provide a useful background for nonparametric techniques. With an understanding of this material, readers can jump directly to sections in Chapters 8 and 9. For readers who wish to obtain an overall impression of the book, we suggest reading Chapter 1, §2.1, §2.2, Chapter 3, §4.1, §4.2, Chapter 5,

§6.3, §8.3, §8.5, §8.7, §9.1, §9.2, §9.4, §9.5 and §10.1. These core materials may serve as the text for a graduate course on nonlinear time series.

Although the scope of the book is wide, we have not achieved completeness. The nonparametric methods are mostly centered around kernel/local polynomial based smoothing. Nonparametric hypothesis testing with structured nonparametric alternatives is mainly confined to the generalized likelihood ratio test. In fact, many techniques that are introduced in this book have not been formally explored mathematically. State-space models are only mentioned briefly within the discussion on bilinear models and stochastic volatility models. Multivariate time series analysis is untouched. Another noticeable gap is the lack of exposure of the variety of parametric nonlinear time series models listed in Chapter 3 of Tong (1990). This is undoubtedly a shortcoming. In spite of the important initial progress, we feel that the methods and theory of statistical inference for some of those models are not as well-established as, for example, ARCH/GARCH models or threshold models. Their potential applications should be further explored.

Extensive effort was expended in the composition of the reference list, which, together with the bibliographical notes, should guide readers to a wealth of available materials. Although our reference list is long, it merely reflects our immediate interests. Many important papers that do not fit our presentation have been omitted. Other omissions and discrepancies are inevitable. We apologize for their occurrence.

Although we both share the responsibility for the whole book, Jianqing Fan was the lead author for Chapters 1 and 5–9 and Qiwei Yao for Chapters 2–4 and 10.

Many people have been of great help to our work on this book. In particular, we would like to thank Hong-Zhi An, Peter Bickel, Peter Brockwell, Yuzhi Cai, Zongwu Cai, Kung-Sik Chan, Cees Diks, Rainer Dahlhaus, Liudas Giraitis, Peter Hall, Wai-Keung Li, Jianzhong Lin, Heng Peng, Liang Peng, Stathis Paparoditis, Wolfgang Polonik, John Rice, Peter Robinson, Richard Smith, Howell Tong, Yingcun Xia, Chongqi Zhang, Wenyang Zhang, and anonymous reviewers. Thanks also go to *Biometrika* for permission to reproduce Figure 6.10, to Blackwell Publishers Ltd. for permission to reproduce Figures 8.8, 8.15, 8.16, to *Journal of American Statistical Association* for permission to reproduce Figures 8.2 – 8.5, 9.1, 9.2, 9.5, and 10.4 – 10.12, and to World Scientific Publishing Co, Inc. for permission to reproduce Figures 10.2 and 10.3.

Jianqing Fan's research was partially supported by the National Science Foundation and National Institutes of Health of the USA and the Research Grant Council of the Hong Kong Special Administrative Region. Qiwei Yao's work was partially supported by the Engineering and Physical Sciences Research Council and the Biotechnology and Biological Sciences Research Council of the UK. This book was written while Jianqing Fan was employed by the University of California at Los Angeles, the University of

North Carolina at Chapel Hill, and the Chinese University of Hong Kong, and while Qiwei Yao was employed by the University of Kent at Canterbury and the London School of Economics and Political Science. We acknowledge the generous support and inspiration of our colleagues. Last but not least, we would like to take this opportunity to express our gratitude to all our collaborators for their friendly and stimulating collaboration. Many of their ideas and efforts have been reflected in this book.

December 2002

Jianqing Fan
Qiwei Yao

Contents