Time Series: A First Course
with Bootstrap Starter

Tucker S. McElroy and Dimitris N. Politis

Contents

Preface

Time Series is a branch of statistical analysis that is mathematically intriguing but finds extremely diverse practical applications; to name just a few: engineering (electrical, mechanical, civil, etc.), medicine (biostatistics, bioinformatics, imaging, etc.), physics (acoustics, geophysics, etc.), economics, meteorology, ecology, seismology, and others.

The subject can be traced back to the 19th century when Arthur Schuster (1899) introduced the notion of the "periodogram" that was (and still is) used in order to discover hidden periodicities in physical phenomena. Even Albert Einstein took up the subject in his 1914 paper. But perhaps the modern era of Time Series Analysis can be thought to originate with Yule's (1927) paper studying the (still mysterious) sunspot numbers – see Figure 1.5 in what follows. The subject quickly flourished soon after with ground-breaking works by H. Wold, M.S. Bartlett, P. Whittle, and others. The period of the Second World War (1940-1945) was especially active with intellectual giants of the likes of Andrei Kolmogorov and Norbert Wiener working independently on different sides of the Atlantic – see Kolmogorov (1940, 1941) and Wiener (1949).

In the last 50-60 years, several influential graduate-level time series textbooks have been written. Personal favorites include (in alphabetic order): Brillinger (1981), Fan and Yao (2007), Grenander and Rosenblatt (1957), Hamilton (1994), Hannan (1970), Pourahmadi (2001), Priestley (1981), and Shumway and Stoffer (2017). By far the most popular in recent years has been the textbook by Brockwell and Davis (1991), a masterpiece that remains relevant 30 years later (denoted BD91 for short). Indeed, BD91 has since raised the bar for its clarity of presentation of the fundamental material on the statistical analysis of stationary time series. Nevertheless, it is not an easy read, and requires knowledge of measure-theoretic probability and functions of a complex variable.

One of us (DNP) has taught graduate and upper-level undergraduate classes in time series for over 25 years in four different universities. For the Ph.D. level courses, BD91 has invariably been the textbook of choice. The situation as regards undergraduate and/or M.S. level texts in time series has been less obvious; after trying at least half a dozen of texts in the classroom and being fully satisfied with none, the idea of compiling a new book came up. The challenge that we decided to undertake was to produce a text that satisfies the triptych: (i) mathematical completeness – albeit at slightly lower level than BD91, (ii) computational illustration and implementation, and (iii) conciseness

and accessibility to upper-level undergraduate and M.S. students.

With regards points (i) and (iii), through teaching the graduate courses, it became clear that most of the basic theoretical results could be presented in a mathematically convincing way without the need for measure theory or complex analysis. For instance, the Central Limit Theorem (CLT) is readily extendable from the i.i.d. setting to that of MA(q) models, leaving the extension to MA(∞) for later. Similarly, the sufficient condition for causality of an ARMA model can be shown without the technicalities of analytic functions and power series expansions.

In the 21st century, the R language has emerged as the software of choice for statistical computation. It is an open environment for computing that is easy to learn to program, has excellent graphics and interactivity, and includes numerous libraries for a variety of specific applications, including several time series libraries. More importantly, as new methods are invented, the libraries are continuously updated with new developments. Therefore, it was clear that the computational implementation – see point (ii) – must be in the R language that we have adopted.

As already mentioned, the subject of time series analysis is over 100 years old. However, the advent of computing power has radically changed its nature in the last 20-30 years, paralleling a similar evolution in all fields of statistics. Computer-intensive methods such as the bootstrap, jackknife, cross-validation, and Monte Carlo simulation have revolutionized the modern practice of statistics, but curiously have not found their way into the standard textbooks; one typically has to take a special course – and buy a specialized text – to learn about them.

To focus on the bootstrap, the case of i.i.d. and regression data was addressed in the early 1980s. Soon to follow was the research on resampling and subsampling for time series that flourished in the 1990s but is still a subject of active development – see Kreiss and Paparoditis (2020). By now, there are several distinct and well-established methods for resampling time series that are being employed in practice in all sorts of applications from climate change to econometrics.

Thus, we felt that a modern time series textbook would not be complete without covering the subject of resampling methods for time series. We could not do justice to all available methods, so we have limited our exposition to the AR-sieve bootstrap, the block bootstrap and its variations (the circular bootstrap, the stationary bootstrap, the tapered block bootstrap, etc.), the frequency domain and the time-frequency bootstrap, the linear process bootstrap, and subsampling. Before going into those topics, however, we also needed to provide a short introduction to bootstrap methods for i.i.d. and regression data.

Acknowledgements

We are indebted to our colleagues and students for their valuable suggestions over the several years that it took us to complete this book. In particular, many

thanks are due to all students who took the MATH181E class "Introduction to Time Series" in the Winter quarters of 2015, 2017 and 2019 at the University of California, San Diego; different drafts of this book were used in the classroom, and the present volume is much improved based on the students' feedback. We are especially grateful to the MATH181E Teaching Assistants: Jie Chen, Srinjoy Das, Zexin Pan, Yiren Wang, Yunyi Zhang, Tingyi Zhu, and Nan Zou. Finally, sincere thanks are due to our editor John Kimmel for his encouragement, guidance and patience over many years.

TSM is thankful to God for providing the opportunity, motivation, and ability to complete this work. He is also grateful for the support of his wife Autumn during this project, and her patience through the seven years of weekends allocated for writing. DNP is grateful for being blessed with wonderful students and collaborators whose enthusiasm for research is contagious, and extends heartfelt thanks to his family for giving purpose and spice to his life – never a dull moment!

How the book can be used

The book is organized into the sections (of each chapter) that constitute enough material for a one credit-hour lecture. For easy reference, text is broken into Paradigms, Facts, Examples, and Remarks, as well as Definitions, Propositions, Theorems, and Corollaries. The end of each chapter has many Exercises, and typically there are Figures to give a visual augmentation of concepts.

- Paradigm: a major concept, or method of analysis
- Fact: a minor unproved assertion
- Example: a specific illustration, typically with data, of a method or mathematical result
- Remark: additional discussion, possibly with historical content, of a method, result, or example
- Definition: a formal definition of a technical concept
- Proposition: a minor technical result, with proof
- Theorem: a major technical result, with proof
- Corollary: a technical result directly following from a theorem (or proposition)
- Figure: a visual representation of data or the results of analysis
- Exercise: a technical or computational task for the pupil

Each chapter ends with an overview, which summarizes the key ideas into Concepts, and a large number of exercises, which mingle both theoretical and computational aspects.

Students should have completed a basic course in mathematical statistics, addressing in detail the concepts covered in Appendices A and B. Additional background material that may be helpful can be found in Appendices C, D, and E. The book contains some harder material that is less accessible to undergraduates; these sections are marked with a $\star$, and may be skipped without losing continuity of the material. All of Chapters 7 and 8 are starred, and likewise can be skipped. For a Master's level course, the $\star$ sections (and chapters) can be included.

Easy exercises are marked with a $\heartsuit$, and difficult exercises with a $\clubsuit$. Theoretical exercises are marked by a $\diamondsuit$, and computational exercises (requiring work in R) are marked by a $\spadesuit$.

The use of R

The statistical language R is a key feature of learning time series with this book. The intention is not merely to train students in another programming language, but more importantly to solidify theoretical concepts through algorithmic and empirical exercises. In teaching courses with R, one of us has found that today's students will resolve a computational task by internet search (consulting relevant threads) followed by download of relevant R packages; while pragmatic, this approach does not facillitate learning of a statistical method's nuts and bolts. To counteract this impulse, we have endeavored to construct R exercises that are largely independent of packages, emphasizing the construction of computer code from first principles. Whereas for applied projects at a government agency or industrial setting the use of packages is to be preferred, for educational purposes it is crucial for students to have a solid foundation in low-level algorithms. In our experience, students that only learn the higher-level nuances of package manipulation do not develop solid knowledge of the underlying statistical methods, and are therefore less able to do creative technical work.

Students and teachers new to R should download the software and study a tutorial. However, all the features of R needed for the course are taught through the Examples and Exercises of this book.

- Download instructions for R are found at: `https://www.r-project.org/`
- R can also be downloaded from one of the mirror sites (pick your nearest location) at: `http://cran.r-project.org/mirrors.html`
- Single commands can be entered at the console prompt (>). For more extensive projects one can write a *script*, which is just a text file containing lists of commands that can be highlighted and executed by a single click.

R scripts for the Examples and Figures of the text are available from the webpage `www.math.ucsd.edu/~politis/TSBOOK/Rfunctions.html`. The Exercises and Examples of Chapter 1 introduce the students to all the datasets, and give instruction as to how to upload them into their copy of R. The twelve datasets are

available from `www.math.ucsd.edu/~politis/TSBOOK/DATA/data.html`. Updates, notes and potential errata will be posted at `www.math.ucsd.edu/~politis/TSBOOK/updates.html`.