Preface

Welcome.

Time-series analysis is a relatively new branch of statistics. Most of the techniques described in this book did not exist prior to World War II, and many of the techniques date from just the last few decades. The novelty of these techniques is somewhat surprising, given the importance of forecasting in general and of predicting the future consequences of today's policy actions in particular. The explanation lies in the relative difficulty of the statistical theory for time series. When I was in graduate school, one of my econometrics professors admitted that he had switched his focus from time series when he realized he could produce three research papers a year on cross-section topics but only one paper per year on time-series topics.

Why another book on time series?

The explosion of research in recent decades has delivered a host of powerful and complex tools for time-series analysis. However, it can take a little while to become comfortable with applying these tools, even for experienced empirical researchers. And in industry, these tools sometimes are applied indiscriminately with little appreciation for their subtleties and limitations. There are several excellent books on time-series analysis at varying levels of difficulty and abstraction. But few of those books are linked to software tools that can immediately be applied to data analysis.

I wrote this book to provide a step-by-step guide to essential time-series techniques—from the incredibly simple to the quite complex—and, at the same time, to demonstrate how these techniques can be applied in the Stata statistical package.

Why Stata? There are, after all, a number of established, powerful statistical packages offering time-series tools. Interestingly, the conventions adopted by these programs for describing and analyzing time series vary widely, much more widely than the conventions used for cross-section techniques and classical hypothesis testing. Some of these packages focus primarily on time series and can be used on non-time-series questions only with a bit of difficulty. Others have to twist their time-series procedures into a form that fits the rest of the structure of their package.

I helped out in a small way when Stata was first introduced. At that time, the most frequent question posed by users (and potential users) was, "When will time series be available?" For a long time, we would tell users (completely sincerely) that these techniques would appear in the next release, in six to twelve months. However, we

repeatedly failed to deliver on this promise. Version after version appeared with many new features, but not time series. I moved on to other endeavors, remaining a Stata user but not a participant in its production. Like other users, I kept asking for time-series features—I needed them in my own research. I finally became frustrated and, using Stata's programming capabilities, cobbled together some primitive Stata functions that helped a bit.

Why the delay? Part of the reason was other, more time-critical demands on what was, at the beginning, a small company. However, I think the primary reason was StataCorp's commitment to what they call the "human-machine interface". There are lots of packages that reliably calculate estimates of time-series models. Many of them are difficult to use. They present a series of obstacles that must be overcome before you can test your hypotheses on data. Frequently, it is challenging to thoroughly examine all aspects of your data. And they make it onerous to switch directions as the data begin to reveal their structure.

Stata makes these tasks easy—at least, easy by comparison to the alternatives. I find that the facility of Stata contributes to better analyses. I attempt more, I look more deeply, because it is easy. The teams that work for me use several different packages, not just Stata, depending on the task at hand. I find that I get better, more thorough analyses from the team members using Stata. I do not think it is a coincidence.

When Stata finally gained time-series capabilities, it incorporated a design that retains the ease of use and intuitiveness that has always been the hallmark of this package. That is why I use Stata rather than any of the other candidate packages.

Despite the good design poured into Stata, time-series analysis is still tough. That is just the nature of the time-series inference task. I tend to learn new programs by picking up the manual and playing around. I certainly have learned a lot of the newer, more complex features of Stata that way. However, I do not think it is easy to learn the time-series techniques of Stata just from reading the Stata *Time-Series Reference Manual*—and it is a very well-written manual. I know—I tried. For a long time, I stuck with my old, home-brew Stata functions to avoid the task of learning something different, even after members of my staff had adopted the new Stata tools.

Writing this book provided me with the opportunity to break out of my bad habits and make the transition to Stata's powerful time-series features. And I am glad I did. Once you come up the learning curve, I think these tools will knock your socks off. They certainly lower the barrier to many ambitious types of empirical research.

I hope you are the beneficiary of my learning process. I have attempted in these pages to link theory with tools in a way that smooths the path for you. Please let me know if I have succeeded. Contact the folks at Stata Press with your feedback—good or bad—and they will pass it along to me.

Why a revised edition?

The first edition of this book was written using Stata 12. The revised edition has been updated for Stata 16. Specifically, chapter 1 includes updated discussions of Stata's interface, datasets, and commands for importing data. Stata's default random-number generator (RNG) changed from the 32-bit KISS RNG to the 64-bit Mersenne Twister RNG in Stata 14. Therefore, simulated datasets for examples in chapters 3, 5, 7, and 10 have changed. Results of these examples, and in some cases the random-number seed used for reproducibility, have been updated. Finally, chapter 11 was updated with brief overviews of time-series features that have been added since Stata 12.

Who should read this book?

Stata users trying to figure out Stata's time-series tools. You will find detailed descriptions of the tools and how to apply them combined with detailed examples and an intuitive explanation of the theory underlying each tool.

Time-series researchers considering Stata for their work. Each commercial time-series package takes a different approach to characterizing time-series data and models. Stata's unique approach offers distinct advantages that this book highlights.

Researchers who know a bit about time series but want to know more. The gestalt of time-series analysis is not immediately intuitive, even to researchers with a deep background in other statistical techniques.

Researchers who want more extensive help than the manual can provide. It is clear and well written, but, at the end of the day, it is a manual, not a tutorial.

How is this book organized?

Like Gaul, this book is divided into three parts.

Preliminaries. Preparation for reading the rest of the book.

- **Chapter 1: Just enough Stata.** A quick and easy introduction for the complete novice. Also useful if you have not used Stata for a while.
- **Chapter 2: Just enough Statistics.** A cheat sheet for the statistical knowledge assumed in later chapters.

Filtering and Forecasting. A nontechnical introduction to the basic ways to analyze and forecast time series. Lots of practical advice.

- **Chapter 3: Filtering time-series data.** A checklist of questions to answer before your analysis. The four components of a time series. Using filters to suppress the random noise and reveal the underlying structure.

Chapter 4: A first pass at forecasting. Forecast fundamentals. Filters that forecast.

Time-series models. Modern approaches to time-series models.

Chapter 5: Autocorrelated disturbances. What is autocorrelation? Regression models with autocorrelation. Testing for autocorrelation. Estimation with first-order autocorrelated data.

Chapter 6: Univariate time-series models. The general linear process. Notation conventions. The mixed autoregressive moving-average model. Stationarity and invertibility.

Chapter 7: Modeling a real-world time series: The example of U.S. gross domestic product. Getting ready to model a time series. The Box–Jenkins approach. How to specify, estimate, and test an autoregressive moving-average model. Forecasting with autoregressive integrated moving-average models. Comparing forecasts.

Chapter 8: Time-varying volatility: Autoregressive conditional heteroskedasticity and generalized autoregressive conditional heteroskedasticity models. Examples of time-varying volatility. A model of time-varying volatility. Extensions to the autoregressive conditional heteroskedasticity model.

Chapter 9: Models of multiple time series. Vector autoregressions. A vector autoregression of the U.S. macroeconomy. Cross correlations, causality, impulse–response functions, and forecast-error decompositions. Structural vector autoregressions.

Chapter 10: Models of nonstationary time series. Trends and unit roots. Cointegration. From intuition to vector error-correction models.

Chapter 11: Closing observations. Making sense of it all. What did we miss?

Ready, set, . . .

I am a reporter. I am reporting on the work of others. Work on the statistical theory of time-series processes. Work on the Stata statistical package to apply this theory. As a reporter, I must give you an unvarnished view of these topics. However, as we are frequently reminded in this postmodern world, none of us can be completely objective, try as we will. Each of us has a perspective, a slant informed by our life experiences.

Here is my slant. I was trained as an academic economist. I became a software developer to pay my way through graduate school and found I liked the challenges of good software design as much as I liked economic research. I began my postgraduate career in academics, transitioned to the Federal Reserve System, and eventually ended up in research in the financial services industry, where I have worked for a number of leading firms (some of them still in existence). I believe I have learned something valuable at each stage along the way.

For the purposes of this book, the most important experience has been to see how statistical research, good and bad, is performed in academics, the Fed, and industry. Good academic research applies cutting-edge research to thorny problems. Bad academic research gets caught up in footnotes and trivia and loses sight of real-world phenomena. The Federal Reserve produces high-quality research, frequently published in the best academic journals. A signature characteristic of research within the Fed is a deep knowledge of the institutional details that can influence statistical relationships. However, Fed research occasionally exhibits an oversimplified perspective of the workings of the financial services industry. Industry has to make decisions in real time. Accordingly, industry research has to generate answers quickly. Good industry research makes wise tactical choices and selects reasonable shortcuts around technical obstacles. Bad industry research is "quick and dirty".

Embrace the good, avoid the bad. Perhaps because the latter half of my career has been spent in industry, my personal bent is to recognize the limitations of the tools I use without becoming distressed over them. I am more interested in intuition than in proofs.

Here are three articles that sum up the approach I try to emulate:

- Diaconis, P. 1985. Theories of data analysis: From magical thinking through classical statistics. In *Exploring Data Tables, Trends, and Shapes*, ed. D. C. Hoaglin, F. Mosteller, and J. W. Tukey, 1–36. New York: Wiley.
- Ehrenberg, A. S. C. 1977. Rudiments of numeracy. *Journal of the Royal Statistical Society, Series A* 140: 277–297. https://doi.org/10.2307/2344922.
- Wainer, H. 1984. How to display data badly. *American Statistician* 38: 137–147. https://doi.org/10.2307/2683253.

Do not say I did not warn you. Now get cracking and learn some stuff.