Review of Stata par la pratique : statistiques, graphiques et éléments de programmation, by Éric Cahuzac and Christophe Bontemps

Antoine Terracol
EQUIPPE, Universités de Lille
CES, Université Paris 1
Paris, France
terracol@univ-paris1.fr

Abstract. This article reviews *Stata par la pratique*, by Éric Cahuzac and Christophe Bontemps.

Keywords: gn0041, French, learning Stata, Cahuzac, Bontemps, splp

1 Introduction

With *Stata par la pratique : statistiques, graphiques et éléments de programmation*, Éric Cahuzac and Christophe Bontemps (2008) provide the French-speaking community with the opportunity to become familiar with Stata in their first language. To the best of my knowledge, the only other book about Stata available in French is Bocquier (1998). The spectacular evolution of Stata over the last ten years makes Cahuzac and Bontemps' book particularly welcome.

The book is intended for readers already familiar with the statistical methods used throughout. The theoretical elements are thus limited to what is strictly necessary. The authors focus instead on the implementations of these methods using Stata. Researchers and graduate (or advanced undergraduate) students will undoubtedly find in this book a useful introductory guide, as well as numerous examples of how to handle often encountered problems.

The book is organized into eight chapters, starting with a general introduction (chapter 1) before turning to data management (chapter 2) and descriptive analysis (chapter 3). Modeling and inference are then described in chapter 4, and Stata's powerful graphical commands are described in chapter 5. Chapter 6 is dedicated to the best ways to produce publication-quality tables. Chapter 7 offers a gentle introduction to programming, and chapter 8 provides various examples of short programs solving real-world problems.

2   Content

In their foreword, the authors state that the book's eight chapters are meant to be independent, each focusing on a different class of tasks that can be performed with Stata. However, it would be wise for the beginner to follow the order of the book because the complexity of the problems dealt with increases through the book (although chapters 4 and 5 could be inverted without problems).

The first chapter begins by providing an overview of Stata windows (Viewer, Browser, Do-file Editor, etc.) and of the most useful buttons. Readers are then introduced to the different types of files that a Stata user will have to deal with. They are also shown how to update Stata and install new commands. A few useful tips on how to use Stata efficiently and create aliases are also described. This chapter is less satisfactory than the rest of the book. Novice readers would need more information to fully understand the commands described in the rest of the book and to be able to easily understand how to use other commands. In particular, the chapter lacks a general explanation of the structure of Stata commands (for example, that a comma separates the main arguments from the command's options).[1] Moreover, a guide on how to read a syntax diagram in Stata's help files would also have been useful to help the reader distinguish between "optional" and "compulsory" options. The description of do-files, which appears in chapter 7, would have benefited from being moved to chapter 1. The novice reader might think that do-files, being described in one of the final chapters, are for more advanced users and that the command line is the best way for the user to interact with Stata. The same comments can be made about the description of Stata Markup and Control Language and log files, currently described in chapter 7.

The second chapter shows the reader how to open a dataset, either in Stata format or in ASCII. It then explains how to describe, manage, and modify data (including appending and merging datasets, as well as collapsing and reshaping data). The chapter covers all the necessary commands in data management and more. Commands are organized by theme (loading and saving data, selecting subsets, describing the dataset, etc.), and a short description of each command, sometimes with example output, is provided. The reader is then introduced to concepts useful for slightly more advanced use, concepts such as macros, scalars, and matrices in a pedagogic manner.[2] Simple examples are also provided. A few editorial details can hinder readability. For example, the relational operator `==` is used for the first time on page 11, but a table showing the list of operators does not appear until page 14.

Chapter 3 takes the reader through descriptive statistics, statistical tests, and data analysis commands. The reader is first introduced to univariate data-exploration commands, the most often used normality tests, and association tests between variables: correlations, χ^2 tests, comparison of distributions, etc. The second part of the chap-

---

1. An option is used for the first time on page 9, but the fact that a comma separates the options from the rest of the command is not stated until page 125.
2. In their comparison of scalars and macros, the authors fail to mention that scalars have a higher numerical precision than macros (see [U] **18.5 Scalars and matrices**) and that scalars should be preferred to macros for storing intermediate results.

ter focuses on variance analysis and multivariate data analysis: analysis of variance and covariance, principal component analysis, factor analysis, multidimensional scaling, cluster analysis, etc. Whereas the first part of the chapter sometimes appears to be a catalog of commands, the second part goes into much greater detail and offers interesting details on the various data analysis methods presented, as well as detailed analysis of example outputs. The content of chapter 3 perfectly reflects the title of the book, because the reader can actually learn Stata *par la pratique* (which roughly translated means "[learning] Stata by using it") by following a step-by-step analysis and reading the detailed interpretation of the results.

The fourth chapter focuses on modeling and estimation commands. Classical linear regression is presented first, and then the authors turn to models for limited dependent variables and count data. As in the second part of chapter 3, this chapter takes the reader through a detailed analysis of estimation commands, interpretation of output, and postestimation tools (tests, residual analysis, etc.). Probit, logit, and multinomial models are given very complete coverage that, as opposed to a simple introduction, provides the reader with the necessary tools for a thorough analysis of such data. Although the material in this chapter is presented clearly, there is a lack of information on the general structure of estimation commands and on the postestimation tools available for nearly all models.

Chapter 5, which covers graphics, avoids the pitfalls of the previous chapters and starts with a general description of the syntax of graphics commands. The reader will thus be able to easily create graphics other than those presented in this chapter. Once the syntax has been laid out clearly, the authors present the most common options for axis labels, symbols, and line patterns. The reader is then taken through univariate graphics and graphs for descriptive statistics before dwelling on bivariate (`twoway`) graphs. Finally, the authors explain how to combine graphs, how to add parametric or nonparametric regression plots, and how to use more-advanced syntax to handle axes separately. The description of each command or option is completed with an example of the resulting graph. My experience in teaching Stata has taught me that presenting the vast graphical capabilities of Stata in a concise yet complete way could not have been an easy thing. Cahuzac and Bontemps have succeeded in this difficult task.

Chapter 6 is the most original chapter of the book. It takes the reader through the various commands (often user-written) to generate tables of descriptive statistics and tables of estimation results. The authors focus mainly on tables in LaTeX format, widely used in the scientific community, but do not neglect more commercial software. This chapter will enable readers to save a lot of time, create publication-quality tables, and avoid tedious copying and pasting from their log files or other documents.

Chapter 7 covers the basics of Stata programming. The authors first give some useful general tips on how to program efficiently, and then they turn to a thorough description of local and global macros, with an emphasis on the management of alphanumeric macros. The various kinds of loops available in Stata, as well as the `bysort` prefix, are also described as a way to repeatedly run commands. Finally, the creation of do-files and of programs (strict sense) in the form of ado-files are introduced. I found the

first part of chapter 7 to be clear and to contain very valuable information for Stata beginners, although more advanced users could also learn from it. The second part is a little more disappointing. The section covering ado-files seems too short to be really useful and deserved to be extended. For example, the use of the `syntax` command is never explained, although it is used in some of the examples. Likewise, there is a paragraph on program classes (e-class and r-class) within the programming section of chapter 7, but the authors do not explain how to create such programs, nor how to save results in `e()` or `r()` macros.[3] The section on programming ado-files might not contain enough information for the reader to be able to write a usable program.[4] Chapter 7 does, however, provide the reader with a solid enough background to write do-files in an efficient and compact way, and to understand and be able to adapt and modify the examples provided in the next chapter.

Finally, chapter 8 provides numerous examples of short programs, with comments, designed to solve problems that are often encountered by practitioners. A careful reading of those programs will enable novice programmers to understand the way to tackle similar problems they will undoubtedly encounter.

3   Conclusion

With *Stata par la pratique*, Cahuzac and Bontemps aim to provide readers with the basics of Stata and with the means to learn more about the software by themselves. This goal is largely met, and the community of French-speaking Stata users now has a learning and teaching tool that was much needed. Cahuzac and Bontemps have managed to describe a wide array of Stata concepts and commands without drowning the reader in details. On the contrary, novice Stata users will find clear and concise explanations about the correct way to interpret their output. The examples in chapter 8 will show the reader how the commands discussed in the previous chapters can be made to work together efficiently in order to solve real-life problems. Finally, the authors emphasize one of Stata's greatest strengths: its use of the Internet to add new commands and features and to find support online. Of course, no book can claim perfection, and *Stata par la pratique* has its flaws in the way information is organized: Several Stata features that form a coherent group are described in different chapters. For example, `_n` and `_N` are presented in chapter 2, `bysort` is described in chapter 7, and their joint use (a particularly powerful feature of Stata) appears only in an example in chapter 8. Despite these minor flaws, French-speaking practitioners will undoubtedly find in *Stata par la pratique* a valuable addition to their bookshelves.

---

3. Results saved by previous commands are discussed earlier in the chapter.
4. Moreover, the authors fail to say that one must give the ado-file the name of the program it defines.

4   References

Bocquier, P. 1998. *L'essentiel de Stata.* Paris: Ritme Informatique - Global Design.

Cahuzac, É., and C. Bontemps. 2008. *Stata par la pratique : statistiques, graphiques et éléments de programmation.* College Station, TX: Stata Press.

About the author

Antoine Terracol is an assistant professor of economics at the Université Charles-de-Gaulle – Lille 3. His research interests include labor economics, the evaluation of public policies, and behavioral economics. He has used Stata in his researches since 1999 and has occasionally taught it.