

# Preface

This book is a concise guide for applied researchers in economics and finance to learn basic econometrics and use Stata with examples using typical datasets analyzed in economics. Readers should be familiar with applied statistics at the level of a simple linear regression (ordinary least squares, or OLS) model and its algebraic representation, equivalent to the level of an undergraduate statistics/econometrics course sequence.<sup>1</sup> The book also uses some multivariate calculus (partial derivatives) and linear algebra.

I presume that the reader is familiar with Stata's windowed interface and with the basics of data input, data transformation, and descriptive statistics. Readers should consult the appropriate *Getting Started with Stata* manual if review is needed. Meanwhile, readers already comfortable interacting with Stata should feel free to skip to chapter 4, where the discussion of econometrics begins in earnest.

In any research project, a great deal of the effort is involved with the preparation of the data specified as part of an econometric model. While the primary focus of the book is placed upon applied econometric practice, we must consider the considerable challenges that many researchers face in moving from their original data sources to the form needed in an econometric model—or even that needed to provide appropriate tabulations and graphs for the project. Accordingly, Chapter 2 focuses on the details of data management and several tools available in Stata to ensure that the appropriate transformations are accomplished accurately and efficiently. If you are familiar with these aspects of Stata usage, you should feel free to skim this material, perhaps returning to it to refresh your understanding of Stata usage. Likewise, Chapter 3 is devoted to a discussion of the organization of economic and financial data, and the Stata commands needed to reorganize data among the several forms of organization (cross section, time series, pooled, panel/longitudinal, etc.) If you are eager to begin with the econometrics of linear regression, skim this chapter, noting its content for future reference.

Chapter 4 begins the econometric content of the book and presents the most widely used tool for econometric analysis: the multiple linear regression model applied to continuous variables. The chapter also discusses how to interpret and present regression estimates and discusses the logic of hypothesis tests and linear and nonlinear restrictions. The last section of the chapter considers residuals, predicted values, and marginal effects.

Applying the regression model depends on some assumptions that real datasets often violate. Chapter 5 discusses how the crucial zero-conditional-mean assumption of the errors may be violated in the presence of specification error. The chapter also

---

1. Two excellent texts at this level are Wooldridge (2006) and Stock and Watson (2006).

discusses statistical and graphical techniques for detecting specification error. Chapter 6 discusses other assumptions that may be violated, such as the assumption of independent and identically distributed (i.i.d.) errors, and presents the generalized linear regression model. It also explains how to diagnose and correct the two most important departures from i.i.d., heteroskedasticity and serial correlation.

Chapter 7 discusses using indicator variables or dummy variables in the linear regression models containing both quantitative and qualitative factors, models with interaction effects, and models of structural change.

Many regression models in applied economics violate the zero-conditional-mean assumption of the errors because they simultaneously determine the response variable and one or more regressors or because of measurement error in the regressors. No matter the cause, OLS techniques will no longer generate unbiased and consistent estimates, so you must use instrumental-variables (IV) techniques instead. Chapter 8 presents the IV estimator and its generalized method-of-moments counterpart along with tests for determining the need for IV techniques.

Chapter 9 applies models to panel or longitudinal data that have both cross-sectional and time-series dimensions. Extensions of the regression model allow you to take advantage of the rich information in panel data, accounting for the heterogeneity in both panel unit and time dimensions.

Many econometric applications model categorical and limited dependent variables: a binary outcome, such as a purchase decision, or a constrained response such as the amount spent, which combines the decision whether to purchase with the decision of how much to spend, conditional on purchasing. Because linear regression techniques are generally not appropriate for modeling these outcomes, chapter 10 presents several limited-dependent-variable estimators available in Stata.

The appendices discuss techniques for importing external data into Stata and explain basic Stata programming. Although you can use Stata without doing any programming, learning how to program in Stata can help you save a lot of time and effort. You should also learn to generate reproducible results by using do-files that you can document, archive, and rerun. Following Stata's guidelines will make your do-files shorter and easier to maintain and modify.