PART I

ORDINARY LEAST SQUARES

Econometrics concerns the analysis of data describing economic phenomena. Economic data come almost exclusively from nonexperimental sources. Social scientists generally must accept the conditions under which their subjects act and the responses occur. These researchers cannot specify or choose the level of a stimulus and then record the outcome. They can just observe the natural experiments that take place.

For example, many economists have studied the influence of monetary policy on macroeconomic conditions, yet the effects of actions by central banks continue to be widely debated. If a central bank could experiment with monetary policy over repeated trials under identical conditions, economists might be able to isolate the effects of policy more accurately. This would remove some of the controversies.

However, no one can turn back the clock to try various policies under essentially the same conditions. Each time a central bank contemplates an action, it faces a new set of conditions. The actors and technologies have all changed. The social, economic, and political orders are different. To learn about one aspect of the economic world, one must take into account many others. To apply past experience effectively, one must take into account similarities and differences between the past, present, and future.

In the simplest experimental setting, a researcher can repeat an experiment under two predetermined settings of a single stimulus to measure the effect of the change in stimulus. By holding everything else constant, one isolates a particular effect of interest. Consider the situation of an economist who wants to measure the effect of gender or race on earnings. It is ludicrous to imagine the economist changing the race or sex of a large group of otherwise identical individuals in order to observe the change in their earnings. Instead, one must examine the variation in earnings

observable across a heterogeneous mixture of working adults with different levels of education, different native abilities, and different levels of work experience, as well as different genders and races. Untangling race or gender variation in earnings is a difficult research goal.

In the next four chapters we introduce the method of ordinary least-squares linear regression. This is the primary way economists try to isolate such variation as the variation in earnings associated with gender and the variation in prices associated with the discount rate of the central bank.

Our introduction will focus on the properties of ordinary least-squares regression that hold no matter what data are studied with this tool. Hence, researchers can rely upon these properties in every setting. In addition, these properties have analogues in the statistical analysis that follows in Part II. We will use the analogies to organize concepts and to emphasize differences between statistical properties and those we are about to discuss in Part I.

We will use several concepts from linear algebra that the reader should have encountered before: linear vector space, linear dependence, basis, dimension, inner product, length, and orthogonality.[1] This list will look technical and threatening to some eyes. Those who have found linear algebra difficult may discover that the material in these chapters helps to make these concepts easier. We will use these concepts and the associated matrix notation, including matrix multiplication, transposes, and inverses, to express solutions to systems of linear equations.

[1] For reference, see Appendix C.