CONTENTS

Chapter 1

Questions about *Questions*

"I checked it very thoroughly," said the computer, "and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you've never actually known what the question is."

Douglas Adams, *The Hitchhiker's Guide to the Galaxy*

This chapter briefly discusses the basis for a successful research project. Like the biblical story of Exodus, a research agenda can be organized around four questions. We call these frequently asked questions (FAQs), because they should be. The FAQs ask about the relationship of interest, the ideal experiment, the identification strategy, and the mode of inference.

In the beginning, we should ask, *What is the causal relationship of interest?* Although purely descriptive research has an important role to play, we believe that the most interesting research in social science is about questions of cause and effect, such as the effect of class size on children's test scores, discussed in chapters 2 and 6. A causal relationship is useful for making predictions about the consequences of changing circumstances or policies; it tells us what would happen in alternative (or "counterfactual") worlds. For example, as part of a research agenda investigating human productive capacity—what labor economists call human capital—we have both investigated the causal effect of schooling on wages (Card, 1999, surveys research in this area). The causal effect of schooling on wages is the increment to wages an individual would receive if he or she got more schooling. A range of studies suggest the causal effect of a college degree is about 40 percent higher wages on average, quite a payoff. The causal effect of schooling on wages is useful for predicting the earnings consequences of, say, changing the costs of attending college, or strengthening compulsory attendance laws. This relation is also of theoretical interest since it can be derived from an economic model.

As labor economists, we're most likely to study causal effects in samples of workers, but the unit of observation in causal research need not be an individual human being. Causal questions can be asked about firms or, for that matter, countries. Take, for example, Acemoglu, Johnson, and Robinson's (2001) research on the effect of colonial institutions on economic growth. This study is concerned with whether countries that inherited more democratic institutions from their colonial rulers later enjoyed higher economic growth as a consequence. The answer to this question has implications for our understanding of history and for the consequences of contemporary development policy. Today, we might wonder whether newly forming democratic institutions are important for economic development in Iraq and Afghanistan. The case for democracy is far from clear-cut; at the moment, China is enjoying robust economic growth without the benefit of complete political freedom, while much of Latin America has democratized without a big growth payoff.

The second research FAQ is concerned with *the experiment that could ideally be used to capture the causal effect of interest*. In the case of schooling and wages, for example, we can imagine offering potential dropouts a reward for finishing school, and then studying the consequences. In fact, Angrist and Lavy (2008) have run just such an experiment. Although their study looked at short-term effects such as college enrollment, a longer-term follow-up might well look at wages. In the case of political institutions, we might like to go back in time and randomly assign different government structures in former colonies on their independence day (an experiment that is more likely to be made into a movie than to get funded by the National Science Foundation).

Ideal experiments are most often hypothetical. Still, hypothetical experiments are worth contemplating because they help us pick fruitful research topics. We'll support this claim by asking you to picture yourself as a researcher with no budget constraint and no Human Subjects Committee policing your inquiry for social correctness: something like a well-funded Stanley Milgram, the psychologist who did pathbreaking work on the response to authority in the 1960s using highly controversial experimental designs that would likely cost him his job today.

Seeking to understand the response to authority, Milgram (1963) showed he could convince experimental subjects to administer painful electric shocks to pitifully protesting victims (the shocks were fake and the victims were actors). This turned out to be controversial as well as clever: some psychologists claimed that the subjects who administered shocks were psychologically harmed by the experiment. Still, Milgram's study illustrates the point that there are many experiments we can think about, even if some are better left on the drawing board.[1] If you can't devise an experiment that answers your question in a world where anything goes, then the odds of generating useful results with a modest budget and nonexperimental survey data seem pretty slim. The description of an ideal experiment also helps you formulate causal questions precisely. The mechanics of an ideal experiment highlight the forces you'd like to manipulate and the factors you'd like to hold constant.

Research questions that cannot be answered by any experiment are FUQs: fundamentally unidentified questions. What exactly does a FUQ look like? At first blush, questions about the causal effect of race or gender seem good candidates because these things are hard to manipulate in isolation ("imagine your chromosomes were switched at birth"). On the other hand, the issue economists care most about in the realm of race and sex, labor market discrimination, turns on whether someone treats you differently because they *believe* you to be black or white, male or female. The notion of a counterfactual world where men are perceived as women or vice versa has a long history and does not require Douglas Adams-style outlandishness to entertain (Rosalind disguised

[1] Milgram was later played by the actor William Shatner in a TV special, an honor that no economist has yet received, though Angrist is still hopeful.

as Ganymede fools everyone in Shakespeare's *As You Like It*). The idea of changing race is similarly near-fetched: in *The Human Stain*, Philip Roth imagines the world of Coleman Silk, a black literature professor who passes as white in professional life. Labor economists imagine this sort of thing all the time. Sometimes we even construct such scenarios for the advancement of science, as in audit studies involving fake job applicants and résumés.[2]

A little imagination goes a long way when it comes to research design, but imagination cannot solve every problem. Suppose that we are interested in whether children do better in school by virtue of having started school a little older. Maybe the 7-year-old brain is better prepared for learning than the 6-year-old brain. This question has a policy angle coming from the fact that, in an effort to boost test scores, some school districts are now imposing older start ages (Deming and Dynarski, 2008). To assess the effects of delayed school entry on learning, we could randomly select some kids to start first grade at age 7, while others start at age 6, as is still typical. We are interested in whether those held back learn more in school, as evidenced by their elementary school test scores. To be concrete, let's look at test scores in first grade.

The problem with this question—the effects of start age on first grade test scores—is that the group that started school at age 7 is . . . older. And older kids tend to do better on tests, a pure maturation effect. Now, it might seem we can fix this by holding age constant instead of grade. Suppose we wait to test those who started at age 6 until second grade and test those who started at age 7 in first grade, so that everybody is tested at age 7. But the first group has spent more time in school, a fact that raises achievement if school is worth anything. There is no way to disentangle the effect of start age on learning from maturation and time-in-school effects as long as kids are still in school. The problem here is that for students, start age

[2] A recent example is Bertrand and Mullainathan (2004), who compared employers' reponses to résumés with blacker-sounding and whiter-sounding first names, such as Lakisha and Emily (though Fryer and Levitt, 2004, note that names may carry information about socioeconomic status as well as race.)

equals current age minus time in school. This deterministic link disappears in a sample of adults, so we can investigate pure start-age effects on adult outcomes, such as earnings or highest grade completed (as in Black, Devereux, and Salvanes, 2008). But the effect of start age on elementary school test scores is impossible to interpret even in a randomized trial, and therefore, in a word, FUQed.

The third and fourth research FAQs are concerned with the nuts-and-bolts elements that produce a specific study. Question number 3 asks, *What is your identification strategy?* Angrist and Krueger (1999) used the term *identification strategy* to describe the manner in which a researcher uses observational data (i.e., data not generated by a randomized trial) to approximate a real experiment. Returning to the schooling example, Angrist and Krueger (1991) used the interaction between compulsory attendance laws in American states and students' season of birth as a natural experiment to estimate the causal effects of finishing high school on wages (season of birth affects the degree to which high school students are constrained by laws allowing them to drop out after their 16th birthday). Chapters 3–6 are primarily concerned with conceptual frameworks for identification strategies.

Although a focus on credible identification strategies is emblematic of modern empirical work, the juxtaposition of ideal and natural experiments has a long history in econometrics. Here is our econometrics forefather, Trygve Haavelmo (1944, p. 14), appealing for more explicit discussion of both kinds of experimental designs:

> A design of experiments (a prescription of what the physicists call a "crucial experiment") is an essential appendix to any quantitative theory. And we usually have some such experiment in mind when we construct the theories, although—unfortunately—most economists do not describe their design of experiments explicitly. If they did, they would see that the experiments they have in mind may be grouped into two different classes, namely, (1) experiments that *we should like to make* to see if certain real economic phenomena—when artificially isolated from "other influences"—would verify certain

> hypotheses, and (2) the stream of experiments that Nature is steadily turning out from her own enormous laboratory, and which we merely watch as passive observers. In both cases the aim of the theory is the same, to become master of the happenings of real life.

The fourth research FAQ borrows language from Rubin (1991): *What is your mode of statistical inference?* The answer to this question describes the population to be studied, the sample to be used, and the assumptions made when constructing standard errors. Sometimes inference is straightforward, as when you use census microdata samples to study the American population. Often inference is more complex, however, especially with data that are clustered or grouped. The last chapter covers practical problems that arise once you've answered question number 4. Although inference issues are rarely very exciting, and often quite technical, the ultimate success of even a well-conceived and conceptually exciting project turns on the details of statistical inference. This sometimes dispiriting fact inspired the following econometrics haiku, penned by Keisuke Hirano after completing his thesis:

T-stat looks too good
Try clustered standard errors—
Significance gone

As should be clear from the above discussion, the four research FAQs are part of a process of project development. The following chapters are concerned mostly with the econometric questions that come up after you've answered the research FAQs—in other words, issues that arise once your research agenda has been set. Before turning to the nuts and bolts of empirical work, however, we begin with a more detailed explanation of why randomized trials give us our benchmark.

INDEX