Part 1: Counterfactual Causality and Empirical Research in the Social Sciences

Chapter 1

Introduction

Did mandatory busing programs in the 1970s increase the school achievement of disadvantaged minority youth? If so, how much of a gain was achieved? Does obtaining a college degree increase an individual's labor market earnings? If so, is this particular effect large relative to the earnings gains that could be achieved only through on-the-job training? Did the use of a butterfly ballot in some Florida counties in the 2000 presidential election cost Al Gore votes? If so, was the number of miscast votes sufficiently large to have altered the election outcome?

At their core, these types of questions are simple cause-and-effect questions of the form, Does X cause Y? If X causes Y, how large is the effect of X on Y? Is the size of this effect large relative to the effects of other causes of Y?

Simple cause-and-effect questions are the motivation for much empirical work in the social sciences, even though definitive answers to cause-and-effect questions may not always be possible to formulate given the constraints that social scientists face in collecting data. Even so, there is reason for optimism about our current and future abilities to effectively address cause-and-effect questions. In the past three decades, a counterfactual model of causality has been developed, and a unified framework for the prosecution of causal questions is now available. With this book, we aim to convince more social scientists to apply this model to the core empirical questions of the social sciences.

In this introductory chapter, we provide a skeletal precis of the main features of the counterfactual model. We then offer a brief and selective history of causal analysis in quantitatively oriented observational social science. We develop some background on the examples that we will draw on throughout the book, concluding with an introduction to graphical causal models that also provides a roadmap to the remaining chapters.

1.1 The Counterfactual Model for Observational Data Analysis

With its origins in early work on experimental design by Neyman (1990 [1923], 1935), Fisher (1935), Cochran and Cox (1950), Kempthorne (1952), and Cox (1958), the counterfactual model for causal analysis of observational data was formalized in a series of papers by Donald Rubin (1974, 1977, 1978, 1980a, 1981, 1986, 1990). In the statistics tradition, the model is often referred to as the potential outcomes framework, with reference to potential yields from Neyman's work in agricultural statistics (see Gelman and Meng 2004; Rubin 2005). The counterfactual model also has roots in the economics literature (Roy 1951; Quandt 1972), with important subsequent work by James Heckman (see Heckman 1974, 1978, 1979, 1989, 1992, 2000), Charles Manski (1995, 2003), and others. Here, the model is also frequently referred to as the potential outcomes framework. The model is now dominant in both statistics and economics, and it is being used with increasing frequency in sociology, psychology, and political science.

A counterfactual account of causation also exists in philosophy, which began with the seminal 1973 article of David Lewis, titled "Causation."[1] It is related to the counterfactual model for observational data analysis that we will present in this book, but the philosophical version, as implied by the title of Lewis' original article, aims to be a general model of causality. As noted by the philosopher James Woodward in his 2003 book, *Making Things Happen: A Theory of Causal Explanation*, the counterfactual approach to causality championed by Lewis and his students has not been influenced to any substantial degree by the potential outcomes version of counterfactual modeling that we will present in this book. However, Woodward attempts to bring the potential outcomes literature into dialogue with philosophical models of causality, in part by augmenting the important recent work of the computer scientist Judea Pearl. We will also use Pearl's work extensively in our presentation, drawing on his 2000 book, *Causality: Models, Reasoning, and Inference*. We will discuss the broader philosophical literature in Chapters 8 and 10, as it does have some implications for social science practice and the pursuit of explanation more generally.

[1] In this tradition, causality is defined with reference to counterfactual dependence (or, as is sometimes written, the "ancestral" to counterfactual dependence). Accordingly, and at the risk of a great deal of oversimplification, the counterfactual account in philosophy maintains that it is proper to declare that, for events c and e, c causes e if (1) c and e both occur and (2) if c had not occurred and all else remained the same, then e would not have occurred. The primary challenge of the approach is to define the counterfactual scenario in which c does not occur (which Lewis did by imagining a limited "divergence miracle" that prevents c from occurring in a closest possible hypothetical world where all else is the same except that c does not occur). The approach differs substantially from the regularity-based theories of causality that dominated metaphysics through the 1960s, based on relations of entailment from covering law models. For a recent collection of essays in philosophy on counterfactuals and causation, see Collins, Hall, and Paul (2004).

The core of the counterfactual model for observational data analysis is simple. Suppose that each individual in a population of interest can be exposed to two alternative states of a cause. Each state is characterized by a distinct set of conditions, exposure to which potentially affects an outcome of interest, such as labor market earnings or scores on a standardized mathematics test. If the outcome is earnings, the population of interest could be adults between the ages of 30 and 50, and the two states could be whether or not an individual has obtained a college degree. Alternatively, if the outcome is a mathematics test score, the population of interest could be high school seniors, and the two states could be whether or not a student has taken a course in trigonometry. In the counterfactual tradition, these alternative causal states are referred to as alternative treatments. When only two treatments are considered, they are referred to as treatment and control. Throughout this book, we will conform to this convention.

The key assumption of the counterfactual framework is that each individual in the population of interest has a potential outcome under each treatment state, even though each individual can be observed in only one treatment state at any point in time. For example, for the causal effect of having a college degree rather than only a high school degree on subsequent earnings, adults who have completed high school degrees have theoretical what-if earnings under the state "have a college degree," and adults who have completed college degrees have theoretical what-if earnings under the state "have only a high school degree." These what-if potential outcomes are counterfactual.

Formalizing this conceptualization for a two-state treatment, the potential outcomes of each individual are defined as the true values of the outcome of interest that would result from exposure to the alternative causal states. The potential outcomes of each individual i are y_i^1 and y_i^0, where the superscript 1 signifies the treatment state and the superscript 0 signifies the control state. Because both y_i^1 and y_i^0 exist in theory for each individual, an individual-level causal effect can be defined as some contrast between y_i^1 and y_i^0, usually the simple difference $y_i^1 - y_i^0$. Because it is impossible to observe both y_i^1 and y_i^0 for any individual, causal effects cannot be observed or directly calculated at the individual level.[2]

By necessity, a researcher must analyze an observed outcome variable Y that takes on values y_i for each individual i that are equal to y_i^1 for those in the treatment state and y_i^0 for those in the control state. We usually refer to those in the treatment state as the treatment group and those in the control state as the control group.[3] Accordingly, y_i^0 is an unobservable counterfactual

[2] The only generally effective strategy for estimating individual-level causal effects is a crossover design, in which individuals are exposed to two alternative treatments in succession and with enough time elapsed in between exposures such that the effects of the cause have had time to dissipate (see Rothman and Greenland 1998). Obviously, such a design can be attempted only when a researcher has control over the allocation of the treatments and only when the treatment effects are sufficiently ephemeral. These conditions rarely exist for the causal questions that concern social scientists.

[3] We assume that, for observational data analysis, an underlying causal exposure mechanism exists in the population, and thus the distribution of individuals across the treatment and

outcome for each individual i in the treatment group, and y_i^1 is an unobservable counterfactual outcome for each individual i in the control group.

In the counterfactual modeling tradition, attention is focused on estimating various average causal effects, by analysis of the values y_i, for groups of individuals defined by specific characteristics. To do so effectively, the process by which individuals of different types are exposed to the cause of interest must be modeled. Doing so involves introducing defendable assumptions that allow for the estimation of the average unobservable counterfactual values for specific groups of individuals. If the assumptions are defendable, and a suitable method for constructing an average contrast from the data is chosen, then an average difference in the values of y_i can be given a causal interpretation.

1.2 Causal Analysis and Observational Social Science

The challenges of using observational data to justify causal claims are considerable. In this section, we present a selective history of the literature on these challenges, focusing on the varied history of the usage of experimental language in observational social science. We will also consider the growth of survey research and the shift toward outcome-equation-based motivations of causal analysis that led to the widespread usage of regression estimators. Many useful discussions of these developments exist, and our presentation here is not meant to be complete.[4] We review only the literature that is relevant for explaining the connections between the counterfactual model and other traditions of quantitatively oriented analysis that are of interest to us here. We return to these issues again in Chapters 8 and 10.

1.2.1 Experimental Language in Observational Social Science

Although the word experiment has a very broad definition, in the social sciences it is most closely associated with randomized experimental designs, such as the double-blind clinical trials that have revolutionized the biomedical sciences and the routine small-scale experiments that psychology professors perform on

control states exists independently of the observation and sampling process. Accordingly, the treatment and control groups exist in the population, even though we typically observe only samples of them in the observed data. We will not require that the labels "treatment group" and "control group" refer only to the observed treatment and control groups.

[4]For a more complete synthesis of the literature on causality in observational social science, see, for sociology, Berk (1988, 2004), Bollen (1989), Goldthorpe (2000), Lieberson (1985), Lieberson and Lynn (2002), Marini and Singer (1988), Singer and Marini (1987), Sobel (1995, 1996, 2000), and Smith (1990, 2003). For economics, see Angrist and Krueger (1999), Heckman (2000, 2005), Moffitt (2003), Pratt and Schlaifer (1984), and Rosenzweig and Wolpin (2000). For political science, see Brady and Collier (2004), King, Keohane, and Verba (1994), and Mahoney and Goertz (2006).

their own students.[5] Randomized experiments have their origins in the work of statistician Ronald A. Fisher during the 1920s, which then diffused throughout various research communities via his widely read 1935 book, *The Design of Experiments.*

Statisticians David Cox and Nancy Reid (2000) offer a definition of an experiment that focuses on the investigator's deliberate control and that allows for a clear juxtaposition with an observational study:

> The word *experiment* is used in a quite precise sense to mean an investigation where the system under study is under the control of the investigator. This means that the individuals or material investigated, the nature of the treatments or manipulations under study and the measurement procedures used are all selected, in their important features at least, by the investigator.
>
> By contrast in an observational study some of these features, and in particular the allocation of individuals to treatment groups, are outside the investigator's control. (Cox and Reid 2000:1)

We will maintain this basic distinction throughout this book. We will argue in this section that the counterfactual model of causality that we introduced in the last section is valuable precisely because it helps researchers to stipulate assumptions, evaluate alternative data analysis techniques, and think carefully about the process of causal exposure. Its success is a direct result of its language of potential outcomes, which permits the analyst to conceptualize observational studies as if they were experimental designs controlled by someone other than the researcher – quite often, the subjects of the research. In this section, we offer a brief discussion of other important attempts to use experimental language in observational social science and that succeeded to varying degrees.

Samuel A. Stouffer, the sociologist and pioneering public opinion survey analyst, argued that "the progress of social science depends on the development of limited theories – of considerable but still limited generality – from which prediction can be made to new concrete instances" (Stouffer 1962[1948]:5). Stouffer argued that, when testing alternative ideas, "it is essential that we always keep in mind the model of a controlled experiment, even if in practice we may have to deviate from an ideal model" (Stouffer 1950:356). He followed this practice over his career, from his 1930 dissertation that compared experimental with case study methods of investigating attitudes, to his leadership of the team that produced *The American Soldier* during World War II (see Stouffer 1949), and in his 1955 classic *Communism, Conformity, and Civil Liberties.*

On his death, and in celebration of a posthumous collection of his essays, Stouffer was praised for his career of survey research and attendant explanatory success. The demographer Philip Hauser noted that Stouffer "had a hand

[5] The *Oxford English Dictionary* provides the scientific definition of experiment: "An action or operation undertaken in order to discover something unknown, to test a hypothesis, or establish or illustrate some known truth" and also provides source references from as early as 1362.

in major developments in virtually every aspect of the sample survey – sampling procedures, problem definition, questionnaire design, field and operating procedures, and analytic methods" (Hauser 1962:333). Arnold Rose (1962:720) declared, "Probably no sociologist was so ingenious in manipulating data statistically to determine whether one hypothesis or another could be considered as verified." And Herbert Hyman portrayed his method of tabular analysis in charming detail:

> While the vitality with which he attacked a table had to be observed in action, the characteristic strategy he employed was so calculating that one can sense it from reading the many printed examples.... Multivariate analysis for him was almost a way of life. Starting with a simple cross-tabulation, the relationship observed was elaborated by the introduction of a third variable or test factor, leading to a clarification of the original relationship.... But there was a special flavor to the way Sam handled it. With him, the love of a table was undying. Three variables weren't enough. Four, five, six, even seven variables were introduced, until that simple thing of beauty, that original little table, became one of those monstrous creatures at the first sight of which a timid student would fall out of love with our profession forever. (Hyman 1962:324-5)

Stouffer's method was to conceive of the experiment that he wished he could have conducted and then to work backwards by stratifying a sample of the population of interest into subgroups until he felt comfortable that the remaining differences in the outcome could no longer be easily attributed to systematic differences within the subgroups. He never lost sight of the population of interest, and he appears to have always regarded his straightforward conclusions as the best among plausible answers. Thus, as he said, "Though we cannot always design neat experiments when we want to, we can at least keep the experimental model in front of our eyes and behave cautiously" (Stouffer 1950:359).

Not all attempts to incorporate experimental language into observational social science were as well received. Most notably in sociology, F. Stuart Chapin had earlier argued explicitly for an experimental orientation to nearly all of sociological research, but while turning the definition of an experiment in a direction that agitated others. For Chapin, a valid experiment did not require that the researcher obtain control over the treatment to be evaluated, only that observation of a causal process be conducted in controlled conditions (see Chapin 1932, 1947). He thus considered what he called "ex post facto experiments" to be the solution to the inferential problems of the social sciences, and he advocated matching designs to select subsets of seemingly equivalent individuals from those who were and were not exposed to the treatment of interest. In so doing, however, he proposed to ignore the incomparable, unmatched individuals, thereby losing sight of the population that Stouffer the survey analyst always kept in the foreground.

Chapin thereby ran afoul of emergent techniques of statistical inference, and he suffered attacks from his natural allies in quantitative analysis. The

statistician Oscar Kempthorne, whose 1952 book *The Design and Analysis of Experiments* would later became a classic, dismissed Chapin's work completely. In a review of Chapin's 1947 book, *Experimental Designs in Sociological Research*, Kempthorne wrote:

> The usage of the word "experimental design" is well established by now to mean a plan for performing a comparative experiment. This implies that various treatments are actually applied by the investigator and are not just treatments that happened to have been applied to particular units for some reason, known or unknown, before the "experiment" was planned. This condition rules out practically all of the experiments and experimental designs discussed by the author. (Kempthorne 1948:491)

Chapin's colleagues in sociology were often just as unforgiving. Nathan Keyfitz (1948:260), for example, chastised Chapin for ignoring the population of interest and accused him of using terms such as "experimental design" merely to "lend the support of their prestige."

In spite of the backlash against Chapin, in the end he has a recognizable legacy in observational data analysis. The matching techniques he advocated will be discussed later in Chapter 4. They have been reborn in the new literature, in part because the population of interest has been brought back to the foreground. But there is an even more direct legacy. Many of Chapin's so-called experiments were soon taken up, elaborated, and analyzed by the psychologist Donald T. Campbell and his colleagues under the milder and more general name of "quasi-experiments."[6]

The first widely read presentation of the Campbell's perspective emerged in 1963 (see Campbell and Stanley 1966[1963]), in which quasi-experiments were discussed alongside randomized and fully controlled experimental trials, with an evaluation of their relative strengths and weaknesses in alternative settings. In the subsequent decade, Campbell's work with his colleagues moved closer toward observational research, culminating in the volume by Cook and Campbell (1979), *Quasi-Experimentation: Design & Analysis Issues for Field Settings*, wherein a whole menu of quasi-experiments was described and analyzed: from the sort of *ex post* case-control matching studies advocated by Chapin (but relabelled more generally as nonequivalent group designs) to novel proposals for regression discontinuity and interrupted time series designs (which we will discuss later in Chapter 9). For Cook and Campbell, the term quasi-experiment refers to "experiments that have treatments, outcome measures, and experimental units, but do not use random assignment to create the comparisons

[6] In his first publication on quasi-experiments, Campbell (1957) aligned himself with Stouffer's perspective on the utility of experimental language, and in particular Stouffer (1950). Chapin is treated roughly by Campbell and Stanley (1963:70), even though his *ex post facto* design is identified as "one of the most extended efforts toward quasi-experimental design."

from which treatment-caused change is inferred" (Cook and Campbell 1979:6).[7] And, rather than advocate for a reorientation of a whole discipline as Chapin had, they pitched the approach as a guide for field studies, especially program evaluation studies of controlled interventions. Nonetheless, the ideas were widely influential throughout the social sciences, as they succeeded in bringing a tamed experimental language to the foreground in a way that permitted broad assessments of the strengths and weaknesses of alternative study designs and data analysis techniques.

1.2.2 "The Age of Regression"

Even though the quasi-experiment tradition swept through the program evaluation community and gained many readers elsewhere, it lost out in both sociology and economics to equation-based motivations of observational data analysis, under the influence of a new generation of econometricians, demographers, and survey researchers who developed structural equation and path-model techniques. Many of the key methodological advances took place in the field of economics, as discussed by Goldberger (1972) and Heckman (2000), even though the biologist Sewall Wright (1925, 1934) is credited with the early development of some of the specific techniques.

In the 1960s, structural equation models spread quickly from economics throughout the social sciences, moving first to sociology via Hubert Blalock and Otis Dudley Duncan, each of whom is usually credited with introducing the techniques, respectively, via Blalock's 1964 book *Causal Inferences in Nonexperimental Research* and Duncan's 1966 article, "Path Analysis: Sociological Examples," which was published as the lead article in that year's *American Journal of Sociology*. In both presentations, caution is stressed. Blalock discusses carefully the differences between randomized experiments and observational survey research. Duncan states explicitly in his abstract that "Path analysis focuses on the problem of interpretation and does not purport to be a method for discovering causes," and he concludes his article with a long quotation from Sewall Wright attesting to the same point.

A confluence of developments then pushed structural equations toward widespread usage and then basic regression modeling toward near complete dominance of observational research in some areas of social science. In sociology, the most important impetus was the immediate substantive payoff to the techniques. *The American Occupational Structure*, which Duncan cowrote with Peter Blau and published in 1967, transformed scholarship on social stratification, offering new decompositions of the putative causal effects of parental background and individuals' own characteristics on later educational and occupational

[7] Notice that Cook and Campbell's definition of quasi-experiments here is, in fact, consistent with the definition of an experiment laid out by Cox and Reid, which we cited earlier. For that definition of an experiment, control is essential but randomization is not. The text of Cook and Campbell (1979) equivocates somewhat on these issues, but it is clear that their intent is to discuss controlled experiments in which randomization is not feasible and that they then label quasi-experiments.

attainment. Their book transformed a core subfield of the discipline of sociology, leading to major theoretical and methodological redirections of many existing lines of scholarship.[8]

In part because of this success, it appears undeniable that Blalock and Duncan became, for a time, less cautious. Blalock had already shown a predilection toward slippage. When introducing regression equations in his 1964 book, specified as $Y_i = a + bX_i + e_i$, where X is the causal variable of interest and Y is the outcome variable of interest, Blalock then states correctly and clearly:

> What if there existed a major determinant of Y, not explicitly contained in the regression equation, which was in fact correlated with some of the independent variables X_i? Clearly, it would be contributing to the error term in a manner so as to make the errors systematically related to these particular X_i. If we were in a position to bring this unknown variable into the regression equation, we would find that at least some of the regression coefficients (slopes) would be changed. This is obviously an unsatisfactory state of affairs, making it nearly impossible to state accurate scientific generalizations. (Blalock 1964:47)

But Blalock ends his book with a set of numbered conclusions, among which can be found a different characterization of the same issue. Instead, he implies that the goal of causal inference should not be sacrificed even when these sorts of assumptions are dubious:

> We shall assume that error terms are uncorrelated with each other and with any of the independent variables in a given equation . . . . In nonexperimental studies involving nonisolated systems, this kind of assumption is likely to be unrealistic. This means that disturbing influences must be explicitly brought into the model. But at some point one must stop and make the simplifying assumption that variables left out do not produce confounding influences. Otherwise, causal inferences cannot be made. (Blalock 1964:176)

Blalock then elevates regression models to high scientific status: "In causal analyses our aim is to focus on causal laws as represented by regression equations and their coefficients" (Blalock 1964:177). And he then offers the practical advice that "The method for making causal inferences may be applied to models based on a priori reasoning, or it may be used in exploratory fashion to arrive at models which give closer and closer approximations to the data" (Blalock 1964:179).

Not only are these conclusions unclear – Should the exploration-augmented model still be regarded as a causal model? – they misrepresent the first 171 pages of Blalock's own book, in which he stressed the importance of assumptions grounded in substantive theory and offered repeated discussion of the differences

[8] For example, compare the methods (and substantive motivations) in Sewell (1964), with its nonparametric table standardization techniques, to Sewell, Haller, and Portes (1969), with its path model of the entire stratification process.