Psychological Methods
2010, Vol. 15, No. 1, 47–55

1082-989X/10/$12.00 DOI: 10.1037/a0018538

An Economist's Perspective on Shadish (2010) and West and Thoemmes (2010)

Guido W. Imbens
Harvard University

In Shadish (2010) and West and Thoemmes (2010), the authors contrasted 2 approaches to causality. The first originated in the psychology literature and is associated with work by Campbell (e.g., Shadish, Cook, & Campbell, 2002), and the second has its roots in the statistics literature and is associated with work by Rubin (e.g., Rubin, 2006). In this article, I discuss some of the issues raised by Shadish and by West and Thoemmes. I focus mostly on the impact the 2 approaches have had on research in a 3rd field, economics. In economics, the ideas of both Campbell and Rubin have been very influential, with some of the methods they developed now routinely taught in graduate programs and routinely used in empirical work and other methods receiving much less attention. At the same time, economists have added to the understanding of these methods and through these extensions have further improved researchers' ability to draw causal inferences in observational studies.

Keywords: causality, potential outcomes, Rubin causal model, Campbell causal model, regression discontinuity, instrumental variables

In "Campbell and Rubin: A Primer and Comparison of Their Approaches to Causal Inference in Field Settings" by William Shadish (2010) and "Campbell's and Rubin's Perspectives on Causal Inference" by Stephen G. West and Felix Thoemmes (2010), the authors contrast two approaches to causality. The first originated in the psychology literature and is associated with work by Campbell (e.g., Shadish, Cook, & Campbell, 2002), and the second has its roots in the statistics literature and is associated with work by Rubin (e.g., Rubin, 2006). In this article, I discuss some of the issues raised by Shadish and by West and Thoemmes. Both articles have been enjoyable to read and have given me an opportunity to reflect on the different emphases given in different disciplines to aspects of causal analyses. On the one hand are the goals (causal inferences are very similar in the different fields), whereas on the other hand, the researchers have made different choices based on the constraints and traditions in the fields. I focus mostly on the impact the two approaches have had on research in a third field, economics. In economics, the ideas of both Campbell and Rubin have been very influential, with some of the methods they developed now routinely taught in graduate programs and routinely used in empirical work and other methods receiving much less attention. At the same time, economists have added to the understanding of these methods and through this have further improved researchers' ability to draw causal inferences in observational studies.

Financial support for this research was generously provided through National Science Foundation Grants SES 0452590 and 0820361.

Correspondence concerning this article should be addressed to Guido W. Imbens, Department of Economics, Harvard University, M-24 Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138. E-mail: imbens@harvard.edu

In this article, I also attempt to summarize the way economists think about causality in general and why and how, traditionally and in current practice, economists approach causal questions, taking to heart some of the ideas developed by Campbell and Rubin. Although few researchers in economics would view their work as following directly either Campbell's or Rubin's approach to causality, their ideas have found a wide following precisely because they resonated with the core tenets of economics and were viewed as helpful in understanding the basic questions economists ask. In economics, these approaches have not been viewed as competing as much as one might surmise from Shadish (2010) and West and Thoemmes (2010), with the basic framework often following along the lines of Rubin's approach and with many of the design features in line with Campbell's work. Ultimately, the literature on causality is a remarkable success story of interdisciplinary research, with a common conceptual framework and terminology, influenced by research in various disciplines.

As a side note, it is interesting to contrast the profound impact the approaches of Campbell and Rubin have had on empirical work in economics with that of the graphical approach. Although the graphical approach to causality has been around for more than 2 decades (Pearl, 1995, 2009a; Spirtes, Glymour, & Scheines, 2001), it has had virtually no impact on practice in economics. Whereas Pearl (2009b) appears to see this as a lack of openmindedness in economics, the fast and widespread adoption of aspects of Campbell's work and Rubin's approach suggests the willingness of economists to adopt new methods, as long as the benefits are transparent. My personal view is that the proponents of the graphical approach, unlike Campbell and Rubin, have not demonstrated convincingly to economists that adopting (part) of their framework offers sufficient benefits relative to the framework currently used by economists.

Economics and Structural Models

Let me first make some general comments on the way economists think about causal questions. These are not necessarily original comments. See, for example, Manski (1999) for a similar perspective. Unlike biostatisticians, who often start from the perspective of a randomized clinical trial, economists start with the notion that individuals receive the treatments they received because they choose to. Economists study how (economic) agents, which may be individuals or firms, make optimal choices, that is, how they optimally allocate their resources—time, financial, or otherwise. In making these choices, these agents take into account both the constraints they face (the scarcity of their resources) and their preferences, that is, the utility they would derive from outcomes of their decisions. The goal of such analyses is often to infer the preferences of agents in order to predict what would happen if the constraints the agents face were changed. Examples of such changes include imposing taxes on transactions or expanding the set of choices. Underlying this approach is the notion that the preferences are relatively stable and specifically that they do not change in response to changes in the constraints.

I wish to highlight two important aspects to the type of questions asked by economists. First, the questions economists ask are typically causal in nature. They concern the causal effects of interventions. Second, these questions often involve treatments that have not been seen before. For example, they may concern the effect of a tax in a particular market where there was none before rather than simply changing the tax rate within a range of values observed in the past. It is the causal aspect of the questions that makes the work of Campbell and Rubin and the discussions in Shadish (2010) and West and Thoemmes (2010) so relevant for economists, although it should be kept in mind that answers to the causal questions alone are not necessarily sufficient to answer the questions raised by economists. Next, I give two examples of this conceptual framework to help frame the discussion of the approaches of Campbell and Rubin.

Supply and Demand Systems

A canonical example of the econometric approach is the analysis of supply and demand functions, discussed in many textbooks (e.g., Amemiya, 1985; Wooldridge, 2002). Suppose consumers have preferences concerning two products: apples and oranges. This means they can compare bundles of products (e.g., they know whether they prefer two apples and three oranges vs. three apples and two oranges). Consumers have a fixed budget for their fruit expenditure. The combination of a budget and preferences leads to a demand function that describes how many apples and oranges consumers are willing to buy at given prices. Under plausible assumptions, this demand function would be decreasing in the price of oranges: The higher the price, the fewer oranges people will be willing to buy. Orange suppliers face a different problem. They incur costs in bringing oranges to the market. These costs may increase as a function of the quantity of oranges they sell. As a result, they would be willing to sell more oranges if the price were higher. The supply function describes how much the suppliers are willing to sell, as a function of the price. The suppliers and consumers meet in the market. Through a mysterious process, they figure out at what price the quantity of oranges consumers demand equals the quantity of oranges the suppliers are willing to sell. These are known as the equilibrium price and equilibrium quantity. The consumers and suppliers then trade at that equilibrium price for oranges.

Within this framework, one can do remarkable things given knowledge of the supply and demand function. For example, one can predict what would happen if a tax were imposed in this market. If for each orange sold the consumer would have to pay a tax, τ, one can infer what the new equilibrium price would be and what the new quantity traded would be. This is possible even though there may never have been a tax in this market. This remarkable ability to predict the effect of interventions never seen in the past comes from the economic structure imposed on the problem. For example, implicit in what I described is the assumption that consumers do not care what fraction of the price they pay for oranges goes to suppliers and what fraction goes to the tax authorities.

To implement this framework, one would need to estimate both the demand and supply functions. These functions are causal in the Rubin causal model sense of representing potential outcomes. Estimating them is intrinsically a difficult problem and is, in fact, the original problem that led to the establishment of the field of econometrics. Suppose one has data from different markets, for example, the same market on different days. What makes it difficult is that simple comparisons of markets where the price was low with markets where the price was high do not directly tell one about either the demand or the supply function. Observed (equilibrium) prices and quantities represent intersections between supply and demand functions, and it is not clear whether a change in price between two different markets represents a shift along the demand function or a shift along the supply function. The technical term is that price is *endogenous*.

What is the point of this example? It illustrates that the starting point in many economic analyses of causal effects is not to look for comparisons of units with different values for the cause that look similar in terms of observed covariates. From the beginning, economists tend to be concerned that the fact that units have different values for the cause implies that they are different. Whereas in statistics the starting point is often a randomized experiment and observational studies are analyzed by making them more closely resemble randomized experiments, economists see the presence of differences in the value of the treatment or cause as evidence of incomparability of units. Price differences are not the result of random events; they are at least partly the result of the preferences and constraints faced by consumers, and as a result, they do not directly allow the estimation of demand function. The traditional econometric solution to problems of this type is to build a model that explicitly accounts for the decisions made by agents that led to the receipt of the treatment received.

Discrete Choice Models

The second example is drawn from a series of book chapters by McFadden (1973, 1981, 1982, 1984), part of the body of work for which he eventually was awarded the Nobel prize in economics. The eventual problem McFadden was interested in was predicting the effect that introducing the Bay Area Rapid Transport system in the San Francisco area would have on commuter behavior. To be more specific, I focus on the demand for the new transportation mode. To predict this demand, he had available, among other data sources, observations of individuals' commuting choices prior to the availability of the Bay Area Rapid Transport system as well as observations of the constraints the individuals faced, including prices and time costs of these choices.

McFadden (1973, 1981, 1982, 1984) set out to build a behavioral model of commuter choice. To make it more specific, suppose that the data available contain information on three currently available choices, commuting by car, by bus, or by train. The first component of this model is a utility function, with individuals having preferences over characteristics of the commuting choices. Such characteristics may include how long it takes to commute by car, bus, or train and how expensive these options are. These characteristics vary by choice and possibly also by individual (depending on where one lives, the time cost of commuting by car will vary). Here, the important assumption is that one can think of a transportation choice as a bundle of attributes. Individuals do not actually care about the label *car, train,* or *bus*; they care about attributes of these choices, such as the amount of time it takes to go from home to work, the amount of money it takes, and possibly other attributes. This type of model is referred to as a *hedonic model*. The ability to reduce a choice to a (small) set of attributes is what fundamentally enables one to predict the demand for the new travel option; a new option, such as a new rapid transit system, is viewed merely as a new combination of commuting time and travel costs, and understanding how individuals value these characteristics is sufficient for predicting the demand for this new travel mode. Individuals also face budget constraints based on income and wealth. Individuals are then assumed to choose the commuting option with the highest utility. With such a model, one can predict what would happen if a new commuting option with specific characteristics would be offered.

Note that this is different in a subtle way from simply estimating the causal effect of, say, a one dollar increase in the gasoline price on the probability of commuting by car. One could easily imagine, even if it would be difficult to implement such an experiment, a randomized experiment in which a random subset of individuals were faced with a one dollar increase in the gasoline price, leading to unbiased and credible estimates of the effect of a one dollar increase in the gasoline price on the probability of commuting by car. However, such an experiment would not in itself be informative about the demand for a currently nonexisting travel mode; in fact, no experiment would directly be informative about such a question without the type of structural model that McFadden (1973, 1981, 1982, 1984) developed.

At the same time, causal effects are at the core of these models. If the models are accurate representations of the way the world works, they capture causal effects of various types. At the same time, those who develop such models are much more ambitious, hoping that the models will be useful for predicting the effect of policies (such as the introduction of a new commuter option) that have not been in operation yet. To do so, such models must obviously rely on assumptions beyond those required for drawing causal inferences. To draw the causal inferences, it would be helpful to have data from settings where prices were randomly assigned. In the absence of such data, the problem of drawing causal inferences again becomes a more complicated one, similar to the discussion in the section on supply and demand.

The Influence of the Rubin Causal Model on Economics

There are two key ingredients of the Rubin causal model as I view it (and having worked with Rubin extensively over the years, including working on the forthcoming textbook discussion, Imbens & Rubin, 2010, I have had a fair amount of exposure to this model). The two key components are (a) the notion of potential outcomes and the definition of causal effects solely in terms of these potential outcomes and (b) the assignment mechanism, where the probability of assignment is expressed in terms of potential outcomes and covariates. This is quite different from the traditional approach in econometrics, where typically only the joint or conditional distribution of the observed variables was modeled. I discuss these two aspects separately. First, I discuss the role of potential outcomes, how the potential outcomes approach has had a long but somewhat uneven history in economics, and how it has recently been embraced as the preferred setup for analyzing causal questions. Second, I discuss two special cases of the assignment mechanism and how they are viewed in the economics literature.

Potential Outcomes

Potential outcomes represent, at the unit level, the outcomes that would occur if a particular level of the treatment were applied to a particular unit. Typically, an assumption similar to the stable unit type value assumption (Rubin, 2006) is made to ensure that with a binary treatment, a pair of potential outcomes is sufficient to describe all possible outcomes for each unit. A comparison between potential outcomes for a particular unit then has a causal interpretation. This requires no assumptions about the assignment mechanism or about what is observed. The causal effects are defined at the unit level without regard to how the treatments are assigned.

Historically, this potential outcome approach is very much in the spirit of the way economists think about causality, although that is not apparent in the current econometrics textbooks. Economists think of a demand function describing quantities demanded as a function of prices. This is exactly the same as the potential outcome notion, other than the minor generalization that the treatment—price in this case—is continuous. One of the early econometricians, Haavelmo, clearly had potential outcomes in mind in his description of the consumption function, which describes the expenditure on consumption of, say, food (denoted by u) as a function of income (denoted by r). Haavelmo (1943, chap. 40) wrote,

> if all consumers in the society were repeatedly furnished with the total income, or purchasing power, r per year, they would on average or normally spend $\bar{u} = \alpha \times r + \beta$. The amount actually spent would be given by $u = \alpha \times r + \beta + x$, where x is a certain random residual with mean value $= 0$, irrespective of the value given to r. (p. 4)

This regression function, in old fashioned notation with β the intercept, α the slope coefficient, and x the residual, is not simply a conditional expectation. There is a notion that income, r, takes on a particular value but that it could have been assigned a different value, and as a result a different level of consumption, u, would have been observed. The regression function describes the different values u would have assumed had total income been different. Of interest, this clear distinction between potential outcomes and realized outcomes became blurred in much of the econometrics literature that followed Haavelmo's work.

More recently, economists have moved toward adopting—or perhaps it would be more accurate to say they have returned to—the potential outcome framework more explicitly and more systematically, in particular in the context of the study of the evaluation of social programs. In the recent literature on causality, this is now the preferred setup. See the textbook discussions in Angrist and Pischke (2008), Caliendo (2006), M.-J. Lee (2005), and Wooldridge (2002), and see the surveys in Heckman, Lalonde, and Smith (2000) and Imbens and Wooldridge (2009). Adopting this framework has allowed researchers to be more explicit about the nature and degree of heterogeneity allowed for and to articulate critical assumptions more clearly. See, for example, the change from Heckman and Robb (1985), who did not explicitly distinguish between potential and realized outcomes, to Heckman (1990), who explicitly adopted the Rubin potential outcome framework.

Economists working in this area have ultimately found this component of the Rubin causal model useful for a variety of reasons. Some of these are the same as those that motivated Rubin to articulate the framework in the first place. In addition, in my view the framework appeals to economists because it is close to the general framework used in economic theory, in which the notion of alternative outcomes that would result from changes in exposures is familiar.

Unconfoundedness

The second key component of the Rubin causal model is the assignment mechanism. Here the probability of a particular set of assignments is described in terms of covariates and potential outcomes. A leading case, perhaps the most important case in practice, is the case where the assignment mechanism is free of dependence on the potential outcomes. Rubin (1990) referred to this as *unconfoundedness*. In addition to unconfoundedness, it is typically assumed that the probability of assignment to any treatment level is strictly between zero and one so that each unit can ex ante be exposed to any treatment. In economics, different terms for closely related assumptions are being used. The notion is sometimes referred to as *exogeneity* or as *selection on observables* (see Imbens, 2004, for a review). Neither of these terms was originally formulated in terms of potential outcomes, and in fact, they are often used without precise definitions. The term *selection on observables* was introduced by Heckman and Robb (1985), in the context of a selection model.

Unconfoundedness implies that comparisons of outcomes for units that differ in terms of treatment status but are homogeneous in terms of observed covariates have a causal interpretation. In other words, if we find a pair of units with the same covariate values, one treated and one control, then the difference in outcomes is unbiased for the average effect of the treatment for units with those values of the covariates.

Let me make some comments on this, because the assumption has generated much more controversy in economics than one might expect. This is partly because the assumption that units that look alike in terms of observed characteristics but that are in different treatment regimes are directly comparable is often suspect. If these units look alike in terms of background characteristics, but they made different choices, it must be because they are, in fact, different in terms of unobserved characteristics. In other words, if they were the same in terms of all relevant characteristics, why would they make different choices? The underlying concern among economists is that such an assumption may be difficult to reconcile with optimal behavior by individuals. Let me make this more specific. Suppose one is analyzing data from a voluntary job-training program for currently unemployed women. One has in mind a program that is offered to all unemployed women who meet certain eligibility criteria but that is not required to keep benefits. One might think that eligible women would assess the benefits of this program partly in terms of the increased chances of finding a job and that they would compare those benefits to the costs, including any direct financial cost, the travel cost, potential expenses for child care during the training, and the time cost. If, for a particular eligible woman, the expected benefits from participation exceeded the expected costs, then she would choose to participate. Suppose one has data on a number of eligible women for whom this program was designed. We know their labor market activity for the past 24 months, their age, their prior educational achievement, their marital status, and the number and age of their children. Consider now a particular woman who is 33 years old, has a high school degree, is unmarried, has two kids (ages 2 and 4), and worked for 6 months in the preceding 24 months. Suppose this woman chose to participate in the training program. Unconfoundedness suggests that to estimate the effect of the program on the labor market status for this particular woman, one should look for a woman with the exact same characteristics who choose not to go through the program. The question is then whether such a comparison would be credible as a causal effect. A standard response from an economist would be that precisely by choosing to enroll in the program, the first woman showed herself to be different from the second woman. We may call this different motivation or something else, but an economist would be concerned that by making this choice, she revealed herself to be different. Moreover, if the decision was optimal, in the sense of being based on a comparison of expected costs and benefits, it must be that the differences between the two women are related to differences in the expected utilities from their actions. Thus, the comparison would be suspect as a causal one. See Manski (1999) for a similar argument.

Why do I nevertheless view this as a very important assumption in many settings? First of all, in many settings it appears plausible that, at the very least, aligning units in terms of observed background characteristics improves the credibility of the comparison. Even if one is concerned with differences in motivation between the enrolled women and those in the control group, there is no reason to believe that one systematically improves the credibility of one's inferences by comparing women who are different in terms of observed characteristics. Specifically, there is nothing in the data that suggests one improves balance in terms of motivation by comparing a 33-year-old woman who chose to enroll with, say, a 25-year-old woman who did not. At the same time, the only alternative to comparing women with the same observed characteristics is to compare women who differ in terms of observed characteristics. Now this becomes a somewhat subtle issue. In some analyses, one does make deliberate comparisons between units that differ in observed pretreatment variables in order to improve balance on unobserved characteristics. Instrumental variables are a leading example of such a method. Those are exceptional situations though. In most cases, it would appear that making units comparable in terms of observed pretreatment variables improves rather than diminishes the credibility of the comparison. Note that I am not claiming that adjusting for observed covariates is sufficient for making credible causal statements. Rather, I am making the much weaker claim that doing so generally improves things, even if it is not sufficient.

Now one needs to be careful with even this weaker claim. It is theoretically possible that conditioning on a covariate introduces bias that would otherwise not be there. In other words, it is possible that unconfoundedness holds without conditioning on a covariate, X_i, but is not conditional on X_i, even though the covariate is not affected by the treatment. Again, instrumental variables are an example of such a case. However, outside of that setting, which is typically easily recognized, this appears to be largely a theoretical possibility. Note that this is different from the setting studied by Meehl (1971), where the concern was about adjusting for variables that themselves may be affected by the treatment, an issue also raised by Rosenbaum (1984).

A second reason for viewing unconfoundedness as an important assumption is that even despite the earlier argument, there is nothing in the assumption per se that is inconsistent with optimizing behavior. Suppose that the two women in the example faced different costs to attend the job training. They may live in different neighborhoods so that, given the public transportation schedule, it would take more time for one to attend the training than for the other. If the outcome of interest is an indicator for being employed at some fixed time after the program, the optimal decision for the two women may well be different even if expected labor market outcomes are identical. The argument here relies on a difference between the outcome of interest for the researcher (the labor market outcome in this case, say, earnings) and the objective function of the economic agent (net benefits of attending the training, which subtracts the cost of attending from the earnings).

Ultimately this setup has generated a large literature in economics. Part of this literature is applied, with major contributions by Card and Sullivan (1988), Lalonde (1986), and Dehejia and Wahba (1999). In addition, a substantial theoretical literature has emerged, including extensions to the methodology of matching (e.g., Abadie & Imbens, 2006).

Instrumental Variables

The recent literature on instrumental variables in econometrics is one of the success stories of interdisciplinary research. It combined the traditional work on instrumental variables in econometrics (see textbook discussions in Amemiya, 1985, and Wooldridge, 2002), which on its own had had little impact outside of economics, with Rubin's potential outcome framework in Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996). This led to a surge of interest in instrumental variables methods outside of economics. Shadish (2010) and West and Thoemmes (2010) discussed this in more detail, but the benefits of the Rubin causal model are clear: The assumptions are more transparent and separated from functional form considerations, causal estimands are clearly defined, and the implications of heterogeneity become clear. For a detailed discussion of the modern econometric literature, see the recent survey by Imbens and Wooldridge (2009).

The Influence of Campbell's Work on Economics

Campbell's influence on empirical work in economics is more recent, although much of his work goes back considerably further in time than Rubin's. Reading Shadish (2010) and West and Thoemmes (2010), it is easy to see how in recent years much work in economics has been directly influenced by Campbell's ideas. First, and this is the most direct and visible contribution of Campbell's that has had a major influence in economics, regression discontinuity designs are now well established and are routinely and widely used in economics. Second, Campbell's ideas concerning threats to internal validity are widely used. Although explorations of threats to internal validity are not as wide ranging as they are in Campbell's approach, economists routinely use the economic models that motivate particular estimation strategies to explore why estimates may not have hoped for causal interpretations. Third, and part of this is recent, there is much debate about external validity in causal inference settings in economics. The part of Campbell's approach that has not received as much attention in economics as the other components is the stress on design, as discussed in Shadish and West and Thoemmes. Economists have traditionally taken their data as given. Although recently there have been more studies in which researchers have collected their own data and even carried out their own experiments, the overwhelming majority of studies analyze previously collected data, taking as given the measures recorded and the missing data patterns realized. Little attention is paid in published articles or in the teaching in graduate programs to the collection of data and to the choices faced by researchers collecting their own data that may affect the quality of the data and the study. Now I discuss the first three of these issues in more detail.

Regression Discontinuity Designs

Regression discontinuity designs, discussed in both Shadish (2010) and West and Thoemmes (2010), have a long and interesting history. As Cook (2008) discussed eloquently, subsequent to the original discovery of these designs by Thistlewaite and Campbell (1960), they have been reinvented multiple times in multiple disciplines. In economics, they never made much of an impact until very recently. Although Goldberger (1972a, 1972b) did some important work analyzing these methods in the 1970s, this did not generate much attention at the time. It was only in the late 1990s and in the early part of the decade starting in 2000 that economists got interested in these methods. At this time, there was an influential movement in empirical work in economics toward clear design-based strategies, including exploiting both randomized and natural experiments. This movement was led by Card, Krueger, and Angrist (e.g., Angrist, 1990; Angrist & Krueger, 1991; Card, 1990; Card & Krueger, 1994) and led to a receptive audience for regression discontinuity designs.

Early influential applications of regression discontinuity designs include van der Klaauw (2002, 2008), Black (1999), and D. Lee (2001). Since then, there has been a proliferation of both applications and theoretical work on regression discontinuity design methods in economics. A recent special issue of the *Journal of Econometrics* was entirely devoted to regression discontinuity design applications (Imbens & Lemieux, 2008). For a more comprehensive review of regression discontinuity design applications in economics, see the recent survey by D. Lee and Lemieux (2009).

Let me briefly discuss two of these applications, partly to show their richness and partly to show why such designs now have such great appeal in economics. First, DiNardo and Lee (2004) studied the effect of unions on workers' wages. This is an old question in economics: Does unionization lead to higher wages for workers, or are firms with high wages more likely to end up being unionized? The literature on the union wage effect is long and not settled. See Lewis (1986) for an early review. DiNardo and Lee exploited the rules for establishing unions through firm-level elections by selecting a sample of firms that had had elections in which the vote was close to the threshold required for establishing a union. Within this set of firms, they compared subsequent average wages for those firms where the vote implied that a union would be established with average wages for firms where the vote implied no union would be established, and they obtained credible estimates of the causal effect of unions on wages for firms where workers were close to indifferent about whether to be unionized.

Second, consider the study of the effect of class size on children's educational outcomes by Angrist and Lavy (1999). Again, this is a substantive issue that has received much attention in various disciplines. Observational studies based on comparisons of schools with bigger and smaller classes have met with skepticism, arising from concerns that unobserved differences between such schools may invalidate the causal interpretation of such differences. Angrist and Lavy exploited a rule in Israel, known as Maimonides' rule, that establishes that class size cannot be larger than 40. In the data Angrist and Lavy analyzed, this shows up in a discontinuity in average class size, as a function of cohort size, at multiples of forty. This in turn allowed them to estimate credibly, for example, the causal effect of moving from a class size of 40 to a class size of 20 or 21.

Let me discuss briefly some of the recent theoretical contributions in econometrics to the regression discontinuity literature, reviewed in Imbens and Lemieux (2008). Hahn, Todd, and van der Klaauw (2000) studied the interpretation of the fuzzy regression discontinuity estimand in settings with heterogeneous effects. In the sharp regression discontinuity, it is clear that the estimand is the average effect for all units at the threshold. Using Rubin's potential outcome approach, in combination with the insights from the instrumental variables literature (Angrist et al., 1996), Hahn et al. established that one can interpret the fuzzy regression discontinuity estimand as an average causal effect, where the averaging is over a subset of the units close to the threshold, namely the subset for which the threshold mattered. In a more technical contribution, Porter (2003) established the optimality of local linear methods in regression discontinuity settings. A third contribution, by Imbens and Kalyanaraman (2009), focused on optimal bandwidth selection in regression discontinuity settings. Often in regression discontinuity settings, observations with values for the forcing or assignment variable that are a substantial distance away from the threshold are dropped from the analysis. Choosing the distance beyond which observations should be dropped is akin to a bandwidth selection problem in the nonparametric regression literature. Imbens and Kalyanaraman developed optimal ways of selecting this bandwidth.

In an important contribution, McCrary (2008) suggested assessing the internal validity of regression discontinuity methods by investigating the marginal distribution of the forcing variable. Specifically, the question is whether there is a discontinuity in the marginal density of the forcing variable at the threshold. Whether there is such a discontinuity is immaterial for the mechanical application of regression discontinuity methods, but it is important for the interpretation. Typically, the story behind regression discontinuity estimates is that the forcing variable is an immutable characteristic of the unit, with the threshold chosen in a way that leaves units on either side of the threshold comparable. That story is difficult to reconcile with a discontinuity in the density of the forcing variable. As McCrary argued, such a discontinuity can suggest that the value of the forcing variable is manipulated by individuals, in an attempt to influence their treatment status. As an example, consider the case in which a test score at least equal to 60 is required for admission to a particular program. One may wish to use regression discontinuity methods to evaluate the effect of the program with the discontinuity in program admission at a score of 60. However, if one were to find that there are few individuals with a test score of 59 relative to the number of individuals with a test score of 60, one may be concerned that the comparability of individuals with test scores of 59 and 60 that underlies the regression discontinuity approach may be violated through manipulation of the test score, for example, by graders knowing the importance of a test score of 60. Testing for the presence of a discontinuity in the density of the forcing variable is therefore an important way of supporting the causal conclusions of the analysis.

Threats to Internal Validity

Both Shadish (2010) and West and Thoemmes (2010) stressed the role of exploring threats to internal validity in Campbell's approach. Although Campbell's work is not always given proper credit for this, his influence in this area is extensive in the empirical economics literature. Part of this is natural. Economic theories make it natural to contemplate threats to internal validity. I have already discussed some examples of this in the context of regression discontinuity designs. Let me also discuss this in the context of analyses of observational studies based on unconfoundedness assumptions.

Consider an observational study analyzing the effect on subsequent employment status of a labor market program designed to help unemployed individuals find a job. The analysis of such programs (e.g., the Job Training Partnership Act or Greater Avenues to Independence in California) is the motivation for much of the theoretical work in econometrics in this area. If program status is not randomly assigned, an economist is by training inclined to be concerned that those who choose to seek the job search assistance are different from those who choose not to seek it. These differences may reflect motivation or other characteristics of unemployed individuals that can affect their labor market prospects irrespective of the job search assistance. This, in turn, has prompted researchers to explore evidence of such differences in motivation between participants in the program and nonparticipants. One might expect such differences to show up in differences in earlier labor market outcomes and choices between individuals who subsequently participated and individuals who subsequently did not participate in labor market programs.

External Validity

In economics, external validity has not always received as much attention as it has in other disciplines. Researchers often pay little attention to the representativeness of the samples they use or to which populations they may represent. One aspect of external validity, however, has been the focus of much attention in the economics literature. It appears to be somewhat special to economics and does not directly translate into the categories Shadish (2010) and West and Thoemmes (2010) discussed. It is also different from the reactivity concern to validity raised in Shadish, Cook, and Campbell (2002). This is what is known in economics as the Lucas critique (Lucas, 1986), although some of the remedies Shadish et al. suggested for reactivity may apply here, including deception of the participants. The Lucas critique involves individuals' expectations about the value of the treatment to which they are exposed and how these expectations can affect their responses. Lucas (1976) studied this in the context of macroeconomic policy, but it can arise in simpler settings. Suppose a double-blind randomized experiment is conducted with a fraction, p, exposed to the active treatment and $1 - p$ exposed to a placebo, with the value of p chosen by the researcher and known to the participants in the study. Suppose now that individuals view the active treatment as a substitute for other actions they may take. To be specific, they may view exercise as a substitute for a cholesterol-reducing drug. Thus, if they were to know that they were given the active drug, they would choose not to exercise, and if they were to know that they did not receive the active drug, they would choose to exercise. With their actual treatment status unknown, individuals may estimate their probability of receiving the active drug. If this probability is above some threshold, p_0, individuals view exercise as unnecessary because they are likely to be receiving the active drug, but if this probability is below p_0, they exercise, because it is unlikely they are receiving the active treatment. If the magnitude of the treatment effect increases when the individual exercises, then the effect of the treatment varies by p, even though there is no interaction between individuals. An experiment with a small value for the fraction treated, p, will not predict the effects of a general adoption of the drug.

There has also been a recent and somewhat heated discussion in economics concerning the relative importance of internal and external validity, echoing some of the themes in Shadish (2010) and West and Thoemmes (2010). The context for this discussion in economics is the recent increase in the number of randomized experiments carried out in development economics. A group of young development economists located in Cambridge and associated with the MIT Poverty Action Lab, led by Abhijit Banerjee, Esther Duflo, and Michael Kremer, have argued that traditional development economics has by and large been unsuccessful in providing credible evidence regarding the efficacy of development programs (Banerjee & Duflo, 2008; Duflo, Glennerster, & Kremer, 2008). These researchers have forcefully argued that credible evidence, based on randomized experiments, is needed to improve practices in the developing world. This group of development economists has been successful in promoting their view, both through carrying out experiments themselves and through convincing organizations such as the World Bank to put more emphasis on randomized experiments.

The success of this group has created a backlash, with a number of senior development economists arguing against the proliferation of experiments. Among the prominent critics are Deaton (2009) and Ravallion (2009). See Imbens (2009) for a comment that is more supportive of the move toward randomization. The criticisms fall into a number of categories. First, experiments are not feasible in many important settings, and economists should not ignore questions for which experiments are not possible. The more interesting criticism is concerned with external validity. The argument goes as follows. A randomized experiment may be useful for estimating the effect of a particular treatment in a particular setting. However, it does not generalize to other settings. A well thought out structural model, however, identifies parameters that are invariant to populations and therefore can be used to predict the effects of policies in different settings and for different populations. There is typically little evidence provided that these structural models are indeed effective in both capturing the causal effects in the populations studied or in generalizing them to other settings and populations.

Conclusion

Both Rubin and Campbell have had a major influence on empirical work in economics. Rubin's causal model and especially the potential outcome framework are now the standard setup for analyzing causal questions in economics. Regression discontinuity designs are widely used, thanks to Campbell. Although it has taken a considerable amount of time for their work to make its way through the discipline of economics, its influence now is a tribute to interdisciplinary work and to the power of their ideas. The discussions of Shadish (2010) and West and Thoemmes (2010) eloquently speak to this and clarify the contributions of the different approaches.

References

Abadie, A., & Imbens, G. (2008). On the failure of the bootstrap for matching estimators. *Econometrica, 76,* 1559–1564.

Amemiya, T. (1985). *Advanced econometrics*. Cambridge, MA: Harvard University Press.

Angrist, J. (1990). Lifetime earnings and the Vietnam era draft lottery: Evidence from Social Security administrative records. *American Economic Review, 80,* 313–335.

Angrist, J., Imbens, G., & Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91,* 444–472.

Angrist, J., & Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics, 106,* 979–1014.

Angrist, J., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics, 114,* 1243.

Angrist, J., & Pischke, S. (2008). *Mostly harmless econometrics: An empiricists' companion*. Princeton, NJ: Princeton University Press.

Banerjee, A., & Duflo, E. (2008). *The experimental approach to development economics.* Unpublished manuscript, Department of Economics, Massachusetts Institute of Technology.

Black, S. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics, 114,* 577–599.

Caliendo, M. (2006). *Microeconometric evaluation of labour market policies.* Berlin, Germany: Springer-Verlag.

Card, D. (1990). The impact of the Mariel Boatlift on the Miami Labor market. *Industrial and Labor Relations Review, 43,* 245–257.

Card, D., & Krueger, A. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review, 84,* 772–784.

Card, D., & Sullivan, D. (1988). Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica, 56,* 497–530.

Cook, T. (2008). Waiting for life to arrive: A history of the regression–discontinuity design in psychology, statistics, and economics. *Journal of Econometrics, 142,* 636–654.

Deaton, A. (2009). *Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development* (NBER Working Paper 14690). Cambridge, MA: National Bureau of Economic Research.

Dehejia, R., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association, 94,* 1053–1062.

DiNardo, J., & Lee, D. (2004). Economic impacts of new unionization on private sector employers: 1984–2001. *Quarterly Journal of Economics, 119,* 1383–1441.

Duflo, E., Glennerster, R., & Kremer, M. (2008). Using randomization in development economics research: A toolkit. In P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (pp. 3895–3962). Amsterdam, The Netherlands: Elsevier.

Goldberger, A. (1972a). *Selection bias in evaluating treatment effects: Some formal illustrations.* Unpublished manuscript.

Goldberger, A. (1972b). *Selection bias in evaluating treatment effects: The case of interaction.* Unpublished manuscript.

Hahn, J., Todd, P., & van der Klaauw, W. (2000). Identification and estimation of treatment effects with a regression–discontinuity design. *Econometrica, 69,* 201–209.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica, 11,* 1–12.

Heckman, J. (1990). Varieties of selection bias. *American Economic Review, Papers and Proceedings, 80,* 313–318.

Heckman, J., Lalonde, R., & Smith, J. (2000). The economics and econometrics of active labor markets programs. In A. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3, pp. 1031–1189). New York, NY: Elsevier Science.

Heckman, J., & Robb, R. (1985). Alternative methods for evaluating the impact of interventions. In J. Heckman & B. Singer (Eds.), *Longitudinal analysis of labor market data* (pp. 131–195). Cambridge, England: Cambridge University Press.

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics, 86,* 1–29.

Imbens, G. (2009). *Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009)* (NBER Working Paper 14896). Cambridge, MA: National Bureau of Economic Research.

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62,* 467–475.

Imbens, G., & Kalyanaraman, K. (2009). *Optimal bandwidth selection in regression discontinuity designs.* Unpublished manuscript, Department of Economics, Harvard University.

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142,* 615–635.

Imbens, G., & Rubin, D. (2010). *Causal inference in statistics and social sciences.* Cambridge, England: Cambridge University Press.

Imbens, G., & Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47,* 1–86.

Lalonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review, 76,* 604–620.

Lee, D. (2001). *The electoral advantage of incumbency and the voter's valuation of political experience: A regression discontinuity analysis of close elections.* Unpublished manuscript, Department of Economics, University of California, Berkeley.

Lee, D., & Lemieux, T. (2009). *Regression discontinuity designs in economics.* Unpublished manuscript.

Lee, M.-J. (2005). *Micro-econometrics for policy, program, and treatment effects.* Oxford, England: Oxford University Press.

Lewis, G. (1986). *Union relative wage effects: A survey*. Chicago, IL: University of Chicago Press.

Lucas, R. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy, 1,* 19–46.

Manski, C. (1999). Choice as an alternative to control in observational studies. *Statistical Science, 14,* 279–281.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics, 142,* 698–714.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York, NY: Academic Press.

McFadden, D. (1981). Econometrica models of probabilistic choice. In C. Manski & D. McFadden (Eds.), *Structural analysis of discrete data with econometric applications* (pp. 198–272). Cambridge, MA: MIT Press.

McFadden, D. (1982). Qualitative response models. In W. Hildenbrand (Ed.), *Advances in conometrics: Econometric Society monographs* (pp. 1–37). Cambridge, England: Cambridge University Press.

McFadden, D. (1984). Econometric analysis of qualitative response models. In Z. Griliches & M. Intriligator (Eds.), *Handbook of econometrics* (Vol. 2, pp. 1395–1457). Amsterdam, The Netherlands: North Holland.

Meehl, P. (1971). Some thoughts on the distribution of earnings. *Journal of Abnormal Psychology, 77,* 143–148.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika, 82,* 669–688.

Pearl, J. (2009a). *Causality: Models, reasoning and inference* (2nd ed.). Cambridge, England: Cambridge University Press.

Pearl, J. (2009b). *Myth, confusion, and science in causal analysis.* (Technical Report R-348). Los Angeles, CA: University of California, Los Angeles.

Porter, J. (2003). *Estimation in the regression discontinuity model.* Unpublished manuscript, Department of Economics, University of Wisconsin—Madison.

Rosenbaum, P. (1984). The consequences of adjusting for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A, 147,* 656–666.

Rubin, D. (1990). Formal modes of statistical inference for causal effects. *Journal of Statistical Planning and Inference, 25,* 279–292.

Rubin, D. (2006). *Matched sampling for causal effects*. Cambridge, England: Cambridge University Press.

Shadish, W. R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods, 15,* 3–17.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT Press.

Thistlewaite, D., & Campbell, D. (1960). Regression–discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology, 51,* 309–317.

van der Klaauw, W. (2002). A regression–discontinuity evaluation of the effect of financial aid offers on college enrollment. *International Economic Review, 43,* 1249–1287.

van der Klaauw, W. (2008). Breaking the link between poverty and low student achievement: An evaluation of Title I. *Journal of Econometrics, 142,* 731–756.

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods, 15,* 18–37.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Received June 5, 2009
Revision received October 28, 2009
Accepted November 12, 2009 ■