REMARKS ON THE ANALYSIS OF CAUSAL RELATIONSHIPS IN POPULATION RESEARCH*

ROBERT MOFFITT

The problem of determining cause and effect is one of the oldest in the social sciences, where laboratory experimentation is generally not possible. This article provides a perspective on the analysis of causal relationships in population research that draws upon recent discussions of this issue in the field of economics. Within economics, thinking about causal estimation has shifted dramatically in the past decade toward a more pessimistic reading of what is possible and a retreat in the ambitiousness of claims of causal determination. In this article, the framework that underlies this conclusion is presented, the central identification problem is discussed in detail, and examples from the field of population research are given. Some of the more important aspects of this framework are related to the problem of the variability of causal effects for different individuals; the relationships among structural forms, reduced forms, and knowledge of mechanisms; the problem of internal versus external validity and the related issue of extrapolation; and the importance of theory and outside evidence.

The problem of determining cause and effect is one of the oldest in the social sciences. However, discussions of the issue have not played a large role in population research. More than many other social science disciplines, population has a long history of noncausal descriptive analysis. The estimation of aggregate vital rates, the construction and estimation of life tables, and the description of population dynamics are just three examples. The long attention paid to these kinds of topics has led to considerable sophistication in the area and constitutes a major contribution to social science knowledge. In recent years, however, population as a discipline has developed an interest in many other issues that have an explicit causal dimension. This is a different focus from that of the older work and should naturally be expected to require a period of exploration and knowledge building before it reaches a mature state.

In this article, I bring the perspective of an economist to the study of causal questions and present a relatively nontechnical report on recent discussions of causal modeling in the field of economics. Economics has a long tradition of studying causal questions, but only in recent years has the issue received widespread direct methodological attention within large areas of the discipline. There is by no means a settled and accepted set of principles for addressing causal questions in economics or even for the proper relative role of causal and noncausal research, for there are deep differences within the discipline in viewpoint, both philosophically and at the practical level, about how to go about conducting research. I bring the perspective of only one economist to this issue, but one that is shared by many within the discipline.[1]

---

*Robert Moffitt, Department of Economics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218; E-mail: moffitt@jhu.edu. A previous version of this article, titled "Causal and Non-Causal Research in Population: An Economist's Perspective," was presented at the 2003 annual meeting of the Population Association of America, May 1–3, Minneapolis. The author thanks Chris Bachrach, Greg Duncan, Tom Fricke, Geoffrey McNicoll, Herb Smith, and two anonymous referees for their helpful discussions and comments.

1. For other expositions of causal modeling in economics that are more comprehensive and technical than the one in this article, see Heckman (2000) and Heckman and Vytlacil (forthcoming). See Winship and Morgan

Here, I attempt to provide support for several points: (1) assessing causal effects requires a strong theoretical framework specifying a particular causal channel in a particular context and that provides a theoretical basis for specifying determining factors that can be credibly argued to be exogenous to the outcomes of interest;[2] (2) applying this standard yields a set of several different types of determining factors that have been used in the economics literature, each with its own strengths and weaknesses; (3) there is an important trade-off between the validity of a particular estimated causal effect and its generalizability, for pursuit of the former often leads to the loss of the latter, and vice versa; and (4) the pursuit of knowledge in light of the necessarily more modest claims of causal determination that follow from this approach requires synthesis and reconciliation studies that are based on a variety of different approaches.

A NOTE ON NONCAUSAL ANALYSIS

There are large areas of research on population that do not aim to estimate causal effects. The construction of vital rates, life-table analysis, and indirect estimation are three examples (original data collection, in and of itself, is another). Indeed, the core of population research once consisted of these types of analyses. Those areas continue to thrive within the field. However, the analysis of causal questions has nevertheless been growing within the population research community, and it is these areas that this article addresses.

It is still sometimes argued, however, that analyses that appear to be closely linked to causal analysis are not so linked, that the analysis being presented is merely "descriptive," and that no causal claims are being made for the correlations under examination or the regression coefficients being estimated. On the one hand, it is clear that descriptive analysis and the presentation of conditional and unconditional correlations between variables has a role in population research, as it does in research in all social science disciplines. Indeed, it could even be argued that descriptive analysis is currently undervalued and is disadvantaged in the competition for journal space, given the rising emphasis on the estimation of causal effects. Nevertheless, there is a danger in proceeding too far in descriptive analyses that use regression or other methods for estimating conditional means, for they are easily misinterpreted as reflecting causal relationships. Indeed, it is difficult for most researchers even to examine a table of regression coefficients without speculating on the possible causal mechanisms that could underlie them. The choice of variables to include in a regression is almost always guided by some preliminary causal theory or hypothesized mechanism or set of mechanisms as well. Caution should therefore be exercised in presenting statistical models that purport merely to describe, not "explain," variables that are subject to human choice.

THE ECONOMIC FRAMEWORK FOR CAUSAL ANALYSIS

The modern formulation of the problem of causal analysis is based on the fundamental notion of a counterfactual for an individual, state, country, or other unit (henceforth, the term *individual* is used for convenience even though the unit of analysis can be anything).[3] Every individual i has two possible, or potential, outcomes, Y_{1i} and Y_{0i}, where Y_{1i}

(1999) for a good exposition that is designed for sociologists; Moffitt (2003) for an abbreviated and less-technical version of this article; and Smith (2003) and Fricke (2003) for other perspectives.

2. The concepts of exogeneity and endogeneity are defined later.

3. This section is titled the "economic framework for causal analysis" even though much of the basic framework comes from work in statistics (see citations presented later for the origins of the ideas). However, the framework used here relies heavily on the use of regression equations and explicit representation of error terms in these equations, and most statisticians do not use this framework. See Holland (1986) and Rosenbaum (1995) for two examples, among many, of how statisticians formulate the problem. See Heckman (2004) for a critique of the approach used by statisticians.

is the outcome if the individual experiences a particular event or takes a particular action and Y_{0i} is the outcome if she or he does not experience it, all else held fixed. For example, Y_{1i} may be the birth weight of a child if the mother has smoked during pregnancy and Y_{0i} the birth weight if the mother has not smoked during pregnancy. The difference between the two is the causal effect of the event or action on Y for individual i. Only one of the outcomes is observed in the data because an individual cannot do two things at once.[4] The other, unobserved, outcome is called the counterfactual for that individual.

In statistics, this model is often associated with the work of Rubin (1974) and is generally called the Rubin causal model in that literature, but it is often said to have originated with the work of Neyman (1923, 1935). In economics, the model is often attributed to Heckman (1978), who formalized the simultaneous discrete endogenous variable model. But the conditional logit model of McFadden (1974) has the same counterfactual structure in implicit form as did much earlier work in economics (e.g., Quandt 1972; Roy 1951), and the economic model of simultaneous equations, which dates from the 1940s (Haavelmo 1943, 1944), was implicitly a model of counterfactuals as well. In addition, even the simple linear regression model, when interpreted as a causal relationship, is implicitly a model of counterfactuals if the coefficient on X is interpreted as the difference that would occur in Y if a particular individual were exogenously given one more unit of X.

A basic point in this literature is that the causal effect of the event on Y cannot be estimated without some type of assumption or restriction, even in principle, because of the inherent unobservability of the counterfactual. A cross-sectional regression coefficient on X is necessarily estimated by comparing the values of Y for different individuals who have different values of X, not by comparing different values of Y that a single person would have if his or her X were to vary. Because individuals with different values of X may differ in unobservable ways, the differences in their values of Y may not accurately estimate the differences in the values of Y that would occur for the same individual if her or his X were to change. Assuming that there are no unobservables for individuals with different values of X—in other words, assuming that the counterfactual for an individual can be accurately estimated from the Y values of other individuals—is a simple example of an "identifying" assumption. If this assumption is not made, and it is allowed that individuals may differ in unobserved ways, then some other assumption must be made to "identify" the model (i.e., to estimate the true causal effect of X on Y).

The literature on causal modeling emphasizes that the estimation of a single causal effect always requires a minimum of one such assumption or restriction. Furthermore, and perhaps more important, such a single minimum assumption cannot be formally tested. With a single cross section of data, the assumption for a linear regression equation that there are no unobservables for individuals with different values of X cannot be tested without making some additional assumption that would permit the estimation of the causal effect in question without making the assumption that there are no such unobservables, and it is only by determining the sensitivity of the estimated coefficient to the relaxation of the assumption in question that it may be tested. Consequently, minimal identifying assumptions must be justified or rationalized on the basis of a priori argument, outside evidence, intuition, theory, or some other means outside the model.[5] While the necessity to make these types of arguments may at first seem dismaying, it can also be argued that

4. One may have observed woman i to have made different choices in the past, and one could compare the outcomes under those choices to those made currently. However, because there can be no guarantee that other things are held fixed over time, doing so may not properly measure the true causal effect. Inherently, an individual cannot do two things at the exact same point in time.

5. Additional data collection can, in some circumstances, turn these minimal identifying assumptions into "overidentifying" assumptions that can be tested. See Moffitt (1991) for a nontechnical discussion.

they are what social science is all about: using one's comprehensive knowledge of society to formulate theories of how social forces work, making informed judgments about these theories, and debating with other social scientists what the most supportable assumptions are. The need to make arguments on these types of grounds is partly the basis for the statement by Heckman (2000:91) that "there is no mechanical algorithm for producing a set of 'assumption free' facts or causal estimates based on those facts."

Most of the important causal questions in population research and economic research arise when the variable in question whose effect on Y is the object of interest is endogenous, to use economic parlance, or when the self-selection is "nonignorable," to use the language of Rubin (1978). This is equivalent to saying that those with different values of the variable may differ in unobserved ways, which make an ordinary least-squares or unadjusted regression coefficient an incorrect estimate of the causal effect for an individual. In this circumstance, some identifying assumption must be made. In economics, while it is understood that there are a wide variety of alternative restrictions available (in fact, an infinite number), the most common method in practice is to use what are known as exclusion restrictions.[6] The linear regression model illustrates the issue most easily:

$$Y_i = \alpha + \beta_i T_i + \gamma X_i + \varepsilon_i \quad (1)$$

$$T_i = \delta + \theta X_i + \phi Z_i + \upsilon_i, \quad (2)$$

where Y_i is individual i's outcome variable, T_i is a dummy variable equal to 1 if the individual experiences the event in question and 0 if not, X_i is an exogenous control variable, Z_i is an exogenous control variable that affects the probability of experiencing the event but does not directly affect Y_i, and ε_i and υ_i are error terms.[7] The effect of the event on the outcome (e.g., the effect of smoking on birth weight) is denoted by β_i and is allowed to vary across individuals, a key part of the modern causal model. This "heterogeneity" of response is assumed to be randomly distributed across the population and is inherently unobservable at the individual level.

Least-squares estimation of Eq. (1) will yield biased and inconsistent estimates of that effect if T_i is "endogenous," which is defined as occurring when T_i is related to either ε_i or β_i. For example, bias will result if mothers who smoke would have had lower-birth-weight children than nonsmoking mothers even if they had not smoked because they have other unobservable characteristics (ε_i) that are correlated with low birth weight; and bias will result if mothers who smoke have different unobserved effects of smoking (β_i) than nonsmoking mothers. The estimates of the effect are unbiased and consistent only if T_i is "exogenous," meaning that T_i and ε_i, and T_i and β_i, are independent.[8] In the field of economics today, it is presumed that self-selection and endogeneity are virtually always present to some degree and that the burden of proof is on the investigator to demonstrate

---

6. An alternative approach is one that seeks only to bound causal effects, such as that developed by Manski (1995) (although exclusion restrictions can be used within that framework as well). Another approach is by means of structural modeling, which typically imposes some statistical or theoretical restriction on the data and allows causal effects to be estimated on the condition that those restrictions are true.

7. The nonlinearity of Eq. (2) arising from the dichotomous nature of T_i is ignored because it does not affect most of the important issues of identification. In practice, however, a probit or logit specification is typically used. In addition, it is important to note that a continuous T_i could just as easily be used, and all the issues discussed in this article would equivalently apply. The dichotomous nature of T_i affects only the particular method of estimation, not the fundamental causal-inference approach.

8. The requirement that T_i be independent of β_i may seem unusual, but it is actually part of the usual condition. Suppose that $\beta_i = \beta + \eta_i$, where β is the mean effect in the population and η_i is the individual's deviation from that mean. Then, substituting this equation into Eq. (1) produces a composite error term equal to $\varepsilon_i + T_i\eta_i$. Thus, the requirement for exogeneity is that T_i be independent of this error term, which is fulfilled if it is independent of both ε_i and η_i. Note, too, that in most linear models, only the lack of correlation is needed, but I use the stronger condition of independence here for more generality.

that it is not present if he or she wishes to ignore the problem and apply least squares to Eq. (1). There is no important difference in economic and demographic processes that should make the presumption any different in the latter, as long as individual choice behavior is involved in the determination of T_i.

The role of an exclusion restriction in obtaining an estimate of the effect of T_i in the presence of self-selection is illustrated by Eq. (2), which assumes that T_i is partly affected by some variable Z_i that is not in Eq. (1), that is, that is excluded from Eq. (1). The variable Z_i is sometimes called an "instrument" by analogy with the estimation technique of two-stage least-squares or instrumental variables, even though other estimation methods are available.[9] In the smoking example, a possible Z_i is the price of cigarettes, assuming that the price varies cross sectionally. It should affect the likelihood of being a smoker (i.e., it should appear in Eq. (2)), and there is no reason to expect it to affect birth weight directly. However, if birth weight is, by chance, higher or lower in areas where cigarette prices are higher or lower for other reasons, this Z_i will be invalid and will produce biased estimates of the effect of T_i on Y_i. This example illustrates the two requirements for a satisfactory Z_i—that it affect T_i ("relevance") and that it be independent of ε_i and β_i except through its effect on T_i ("validity").[10]

Eq. (2) is the key additional equation in the economic formulation of the problem because in that approach a conceptualization of how and why variation in T_i arises is essential. Theory almost always plays an important role in this conceptualization. As a practical matter, this approach makes the choice of Z_i, and whether any valid Z_i is available at all, the critical issue in the estimation of causal relationships because the choice of Z_i must be made within some type of conceptual framework. Categories and types of Z_i are presented later.

Much discussion among applied analysts concerns instead the best estimation method to use. However, this is a separate issue from the choice of Z_i and is much less important. The Heckman "lambda" method, which is now generally called a "control-function" method in economics (Heckman and Navarro-Lozano 2004), instrumental variables and two-stage least squares, longitudinal data methods, and other techniques are all appropriate under different assumptions. More-parametric or less-parametric methods can be used. Two-step methods and full-information methods (like maximum likelihood) are alternative methods, each with different relative advantages and disadvantages. Linear probability methods and nonlinear methods like probit and logit have different properties. The derivation of correct standard errors is often discussed. However, these are all second-order issues. They are not relevant to the basic identification question concerning Z_i.

The presumption that β_i is different for different individuals (see Eq. (1)) is another critical aspect of the model and is the piece of the model that distinguishes it from the classic simultaneous equations model in economics, which assumed constant coefficients.[11] It seems reasonable to allow any regressor in a model to have effects that differ across individuals, for there is no reason to expect that the unobservable differences among people affect only their levels of Y through the linear, additive error term and not

---

9. The method of instrumental variables was developed within economics in its most well-known form in the simultaneous equations model about 60 years ago. It was translated into the Rubin causal model by Angrist, Imbens, and Rubin (1996), who argued that it can be best understood by using the concepts in the "intent-to-treat" literature for biomedical randomized trials.

10. This article is concerned mostly with the validity of the exclusions, not with their relevance. The most important issue concerning relevance is whether the excluded variables have a "large" impact on T_i; if they do not, they are called "weak" instruments. See Bound, Jaeger, and Baker (1995) and Stock, Wright, and Yogo (2002).

11. However, the switching regression of Quandt (1972), the conditional logit model of McFadden (1974), and the causal models of Heckman and Robb (1985) and Björklund and Moffitt (1987) all assumed random coefficients, or what is also called "heterogeneous response."

through their responsiveness to changes in determining variables as well. But ordinarily this makes no difference except in interpretation. If $\beta_i = \bar{\beta} + \eta_i$, where $\bar{\beta}$ is the mean in the population and η_i is the (zero-mean) deviation from the mean for individual i, then estimating Eq. (1) by least squares just gives a coefficient estimate on T_i of $\bar{\beta}$, as can be seen by inserting this equation for β_i into Eq. (1). Thus, the estimated coefficient must simply be interpreted as the average effect in the population. However, if there is some type of endogenous selection into $T_i = 1$ and $T_i = 0$, and if the likelihood that an individual is selected into one of the two categories is related to his or her β_i, then such a simple reinterpretation is no longer correct. It has been shown in the literature that if a consistent estimation method is employed which uses an exclusion restriction Z_i and therefore successfully eliminates selection bias, the resulting estimated coefficient on T_i is no longer generally $\bar{\beta}$. Instead, what is estimated is the average β for those who change their values of T_i in response to the variation in Z_i.

For example, as cigarette prices vary across areas, the fraction of individuals who smoke will change as some individuals who would have smoked if prices were low instead choose not to smoke because prices are higher. With this Z_i, one can estimate the average effect of smoking of these "switchers." Suppose that the variation in cigarette prices in the data induces a variation in the fraction who smoke from 30% of the population to 40%. The price variation allows the estimation of the average β for the 10% of the population who were affected by this variation. What cannot be estimated is the average β in the entire population because doing so would require having a Z_i that moved the fraction of smokers from 0% to 100%, thereby permitting the researcher to observe how Y changes as the entire population goes from not smoking to smoking, or vice versa.[12]

Another important distinction is between what are generally termed structural forms and reduced forms. Sometimes it is of sufficient interest to learn only the total effect of Z_i on Y_i without learning how that effect works through T_i. For example, suppose T_i denotes whether two sexual partners use contraception, Z_i is an indicator of how widely available contraceptive devices are in the area, and Y_i is some pregnancy or fertility outcome. It may be sufficient for the government to know how making contraceptives more available affects pregnancy and fertility outcomes in total, with the presumption that the mechanism by which that effect occurs is contraceptive usage but without needing actual knowledge of how much contraceptive usage has changed. The reduced form of Eqs. (1) and (2), obtained by substituting Eq. (2) into Eq. (1), can be estimated by least squares to yield an estimate of this net effect of Z_i on Y_i.

As the discussion of examples in the next section demonstrates, estimation of these reduced forms is common. There is a danger, however, in reduced-form estimation precisely because the mechanism by which Z_i affects Y_i is not estimated (T_i is the mechanism). A statistically significant effect of Z_i on Y_i could occur not because Z_i affects the particular T_i presumed by the analyst, but because some completely different T_i is affected, which in turn affects Y_i. Perhaps the availability of contraceptives induces less sexual activity, for example, through some change in attitudes, and not the rate of contraceptive usage conditional on sex. Alternatively, perhaps Z_i affects Y_i through multiple mechanisms (i.e., the true equation for Y_i contains multiple T_i); reduced-form estimation will not reveal what those mechanisms are. Yet another possibility is that Z_i enters Eq. (1) directly (i.e., it is not a valid instrument) and does not affect any T_i, but still has an effect on Y_i. Again, this will not be revealed by reduced-form estimation alone.

12. See Imbens and Angrist (1994) for a formal statement of this result and Heckman and Vytlacil (1999, 2001, forthcoming) for a more-comprehensive and general discussion of the types of effects that can be estimated (e.g., marginal treatment effects, average treatment effects, and effects of the treatment on the treated). Note that the statistics literature typically uses the term *compliers* for what are here called *switchers*, and it notes that problems are created if there are many *defiers*, defined as individuals who switch the "wrong" way (e.g., who start smoking when the price of cigarettes rises); see Angrist et al. (1996) for a discussion.

Nevertheless, some economists have taken the view that if the only aim of the researcher is to know the effect of Z_i on Y_i, and it does not matter what the mechanism or the channel of effect is (e.g., if the government just wants to know the effects of its policies and does not care why they arise), reduced-form estimation is not only sufficient but requires fewer assumptions than does structural estimation because one does not need to assume that Z_i is excluded from Eq. (1); one only needs to assume that Z_i is not itself endogenous. This makes unbiased estimation much easier, but at the cost of not learning as much about the social process being studied.[13]

Types of Exclusions

As I noted earlier, in most of the applications of interest, choices for excluded variables Z_i are "just identifying," that is, they are the minimum needed to identify the structural effect or, at least, the portion of it that pertains to switchers, generally because it is usually difficult to find even a single strong Z_i, much less more than one. In addition, in many, if not most, applications, there are multiple possible mechanisms (i.e., multiple T_i) by which the instrument Z_i can affect Y_i; the illustrations in the next section makes this point abundantly clear.[14] In that case, the same number of Z_i variables is needed as the number of mechanisms. When one makes the minimal identifying assumptions, the choice of Z_i must necessarily rely on assumptions that are not formally testable, implying that the need for a conceptual framework to guide the choice of Z_i is essential. Although the framework does not have to be formal, it has to be at least sufficiently worked out that an *argument* can be made for exogeneity. Outside evidence in the form of anecdotal evidence, ethnographic data, or even intuition can play a role.[15]

While there can be no hard rules on what types of variables are appropriate, especially given the breadth of possible applications, it is possible to mention some of the types that have been used in a number of different applications. One type of variable that is used frequently, if it is not indeed the most common type, is variables that are environmental or ecological in nature, measuring some aspect of the geographic area in which an individual resides. The aggregate unemployment rate in an area, average wage rate, average price of child care, or the availability of contraceptives are just a few examples (cigarette prices are another).[16] Governmental policies in the form of laws, benefit levels, tax rates, or regulations that cover a fixed geographic area are also often used. The use of these variables as Z_i is based on the assumption that the individual has little effect on or control over characteristics of the area in which he or she is located, and thus that these characteristics are at least one step removed from the individual's own personal characteristics or actions and hence likely to be exogenous to any given individual's outcomes. At the same time, the assumption is that these area characteristics should affect the individual choices of T_i.

An important extension of this type of Z_i is that applied when panel, or longitudinal, data are available, in which case, where Z_i is taken to be the change in the environmental variable from Time 1 to Time 2 in the area in which an individual resides and where Y_i and T_i are likewise taken to be the changes in outcomes and actions over time for an individual.

---

13. This issue also bears on extrapolation and external validity, which are discussed later.

14. Rosenzweig and Wolpin (2000) presented many examples of economic theories that provide exactly such multiple mechanisms.

15. In many cases, it is possible to conduct tests for the validity of instruments by testing other implications of the theory that underlies their use. This approach is common in applications, but there is no space here to discuss it.

16. The method can also be used in time series, as illustrated by Schultz's (1985) use of the world price of butter relative to grain as the Z_i in estimating the effect of women's wages (T_i) on fertility (Y_i). Here, the use of theory is particularly strong, for one needs to connect the gender division of labor with employment in different industries to rationalize this particular instrument.

In this so-called area fixed-effects model, the change in the environmental characteristic is assumed to be exogenous to any given individual's change in outcomes, but to affect his or her change in T_i. The model can be written in simple regression form as

$$\Delta Y_i = \alpha + \beta_i \Delta T_i + \gamma \Delta X_i + \varepsilon_i \tag{3}$$

$$\Delta T_i = \delta + \theta \Delta T_i + \phi \Delta Z_i + \upsilon_i, \tag{4}$$

where Δ denotes a change in a variable from Time 1 to Time 2. In the contraceptive example, ΔY_i is the change in the outcome for individual i, ΔT_i is the change in contraceptive behavior, and ΔZ_i is the change in the availability of contraceptives in the local geographic area. The model assumes that the level Eqs. (1) and (2) contain, in their error terms, fixed effects for each area that are potentially correlated with T_i (e.g., if the greater availability of contraceptives in an area were a result, not a cause, of high contraceptive behavior in the area), fixed effects that cancel out in first differencing. The model assumes instead that the *change* in Z_i (e.g., the change in the availability of contraceptives) arose from some exogenous event that is not directly correlated with the change in Y_i. The method thus partly addresses the common critique of ecological correlations that these correlations may reflect some unobserved characteristic of the areas

Note that this model is not the same as the well-known individual fixed-effects model. If the level equations contain an individual fixed effect, it is true that Eq. (3) is the typical differenced model that is estimated. But in the individual fixed-effects model, only Eq. (3) is estimated because it is assumed that differencing alone will eliminate any bias. Here, it is presumed that there may still be bias because those individuals who change T_i may have done so because Y_i was simultaneously changing for some other reason or because there is direct reverse causality (i.e., the change in Y_i occurred "first"). The presumption that the change in T_i may be endogenous leads to the necessity to eliminate the bias, and it is the change in Z_i that serves as the instrument to do so. Another way to see this point is to recognize that the area fixed-effects model could be estimated with repeated cross-sectional data, that is, on data that are available for multiple individuals in different areas over time but where the individuals are not the same. Either by aggregating the data up to area means and then estimating Eqs. (3) and (4) on those area means or using the individual data with area and time dummy variables and using the interaction of the time dummy variables with Z_i is enough to estimate the model.[17]

There is nevertheless a variety of objections to this choice of Z_i. One objection is that the method usually requires the assumption that residential location is exogenous, which is necessary to argue that the characteristics of the area in which the individual resides are not choice variables to that individual. The second objection is that the ecological correlation problem may still be present even in the area fixed-effects model, for there may be unobservables that are changing over time that are correlated with the change in the area

17. The individual fixed-effects model for panel data—that is, the model in which Eq. (3) is estimated and Eq. (4) is ignored—has seen decreasing support among economists. Simply assuming that the change in an individual's T_i from one time to another is exogenous leaves unspecified why individual changes in T_i occur. They may very well change for endogenous reasons. In the area fixed-effects model, on the other hand, individual changes in T_i are partly ascribed to a specific and measurable variable Z_i that itself changes over time. Likewise, panel data dynamic models that make assumptions on exogeneity that are based on time patterns (e.g., allowing the temporal ordering to have a relation to causality) has seen decreasing support in economics for the same reason: there is no explanation for why time patterns are different for different individuals. However, these methods are still used in some applications (e.g., Adams et al. 2003). In addition, variants of dynamic identification are popular in the biostatistics literature, as illustrated by the work of Robins on sequential treatment structures (see Robins 1999 for a relatively nontechnical discussion). These approaches are not discussed here for reasons of space.

characteristic and with the change in individual outcomes.[18] In addition, and quite important, the area fixed-effects model, as usually specified, assumes that the response to a change in policy is immediate and permanent, yet there may be lags and adjustments in that response as individuals, families, and communities slowly adjust their norms. The third objection is that when ecological variables are several steps removed from individual choices of T_i, a certain distance is opened up between the individual outcomes Y_i and the actions T_i by relating the latter only to an aggregate, area-level variable. There are many possible mechanisms by which an ecological characteristic can affect individual actions, and one may estimate spurious relationships that reflect some other mechanism than the one specified in the model. This problem interacts with the problem of unobserved ecological correlations. In both cases, there is a trade-off between exogeneity and spurious effects with the level of the area used—for example, variables that are taken from the local neighborhood are most likely to be direct influences on individual T_i but are also most likely to be endogenous, while variables that are taken at the state or regional level are less likely to be endogenous to each individual's own actions but are further removed from the individual T_i, resulting in a greater danger of spurious correlations or incorrectly specified mechanisms. Analyses of this type are particularly in need of qualitative or other sources of information on the mechanism at the individual level to provide some reassurance that the estimated effect in question is genuine.

A second type of Z_i that has been popular in certain areas of economic research is the use of what may be called a population-segment fixed-effects model. In this model, nationwide policies are assumed to affect individuals in different demographic groups differently (e.g., married versus unmarried women or single mothers versus single childless women or even men versus women).[19] Assuming that each of these demographic groups has unobserved characteristics that are fixed over time and that the outcomes of each group would change over time by the same amount in the absence of a change in policy, any observed differences in the changes in their outcomes over time can be ascribed to the fact that they were likely affected differently by the policy.

For example, suppose that the government has made a set of new family planning clinics available only to low-income single mothers, and the issue is how the clinics affect their childbearing. In the method under discussion here, one might compare trends over time (say, from before the introduction of the clinics to afterward) in the fertility rates of low-income single mothers with trends in the fertility rates of middle-income single mothers or of married women and take the difference to be the effect of the clinics on low-income childbearing. This type of Z_i is equivalent to the area fixed-effects model just discussed but replaces geographic areas with groups that are defined by demographic characteristic (e.g., gender, marriage, and childbearing). With panel data, the model could be estimated with Eqs. (3) and (4), with ΔZ_i replaced by a dummy variable for whether a woman is a low-income single mother versus a middle-income single mother (or a low-income married woman) and ΔT_i representing whether the woman enrolled in one of the clinics between Time 1 and Time 2; using two independent cross sections, analogous level equations of the type described previously would be estimated.

The method has many of the same strengths and weaknesses as the geographic approach. But it has much less support among empirical researchers because the assumption that the demographic characteristics that are used to separate the sample are exogenous

---

18. For example, areas with a growing availability of contraceptives may be those in which the population had already chosen to control fertility independently. In cases where Z_i is a governmental policy variable, such a bias is often said to arise because of "policy endogeneity," which occurs when areas that choose to introduce, or change, policies do so because of developments in the outcome variable in their areas that were already occurring.

19. This method is sometimes called a "difference-in-difference" method. However, all fixed-effect methods are difference-in-difference methods, so the term *population-segment fixed effects* is used here instead.

with respect to changes in policy is much more suspect than the assumption that residential location is exogenous, and the assumption that the outcomes of different demographic groups would change by the same amount over time in the absence of changes in policy is much more suspect than the assumption that changes in outcomes in different areas would be equal in the same absence.

A third type of Z_i is that based on twin or sibling models, where it is the within-family variation in T_i that is the implicit Z_i.[20] Although twin and sibling models have a long tradition in the estimation of nature-nurture models, where the common family effect is assumed to be genetic, they are more controversial when the common family effect arises from some socioeconomic family and environmental influences that have developed over time. In that case, there is less assurance that twin and sibling differences are themselves exogenous (see, e.g., Bound and Solon 1999 for arguments for why they are not). Also, using twin and sibling differences without any theoretical framework can lead to misspecification of the rest of the equation and hence to incorrect inferences (Rosenzweig and Wolpin 2000).

A fourth type of Z_i arises from a heterogeneous collection of approaches that seek instruments that have arisen from random, possibly sudden, and usually unpredictable changes in environmental or personal variables, almost always in narrow segments of the population. For example, a law may be passed in one area that affects welfare benefits for children aged 4–6 but not in an adjacent area with a similar economic environment and social composition; the 4- to 6 year olds in the two states are compared over time to assess the effect of the law on various child outcomes. This approach is similar to the area fixed-effects approach except that only two states are used and only a small part of the population is examined, and, indeed, the framework of Eqs. (3) and (4) is often used for estimation in this method as well. An example of a personal random occurrence is miscarriage, which affects the timing of fertility and has been used to estimate the effect of fertility on outcomes (see the next section). Another example is month of birth, which affects the age at which children are able to enter school according to governmental law, which, in turn, affects completed years of education and can be used to examine the effect of education on various outcomes for the narrow segment of the population whose completed education is affected by birth month (Angrist and Krueger 1991). Yet another law-related example is a sudden shift in governmental pension policy for individuals who were born after a specific calendar year t, which should cause a discontinuous change in retirement ages for cohorts who were born just before and after t, which can be used to examine the effect of retirement on various outcomes for the possibly narrow segment of the population that is affected by the change in law (Krueger and Pischke 1992). These Z_i are examples of those used in the "natural-experiment" school of economics, which seeks to find "natural" occurrences of the type of randomization that is found in laboratory experiments.[21]

Objections to some of the natural-experiment Z_i are that they are not valid (e.g., they are not truly excludable, or they are not truly exogenous). Another objection is that true, "clean" natural experiments are rare. However, the pure examples of the natural-experiment approach that use valid Z_i illustrate best a trade-off between what is called internal and external validity in the literature on randomized trials. Internal validity holds when the

20. All fixed-effects models are equivalent to instrumental variable models in which the instrument is the deviation in the regressors from their fixed-effect-specific means.

21. See Duncan, Magnuson, and Ludwig (2003) for a recent discussion of examples in the field of child development. The natural-experiment school often claims that the other three types of Z_i just discussed fall under the rubric. However, although it is true that all valid Z_i are formally equivalent to natural experiments, at least conceptually, the other three just discussed are different. In addition, the other three are all fixed-effects approaches, whereas the natural-experiment methods described here are not necessarily in that category.

treatment estimate in an experiment is unbiased for the population used in the experiment and for the particular policy or program that is tested; in other words, it holds when the experiment was actually carried out correctly with a valid randomization and with no contamination between the experimental and control groups, no biasing attrition, or related problems. External validity holds when the estimate generalizes to a larger population than that in the experiment or to a larger set of treatments (policies) of interest. In nonexperimental research, internal validity is threatened by the problem of selection bias and the associated problem of endogeneity, while external validity is just a function of how representative the population is and how wide ranging the policies are in the data. The natural-experiment examples just mentioned represent an extreme attempt to maximize internal validity (i.e., exogeneity) but that typically result in significant loss in external validity (generalizability). Variation in laws in only two states that apply only to 4- to 6 year olds may not generalize to other states or other ages, the fertility variation that is induced by miscarriage may be small in magnitude and may not generalize to fertility variations caused by other forces or to women with a low propensity to miscarry, variations in educational attainment that are induced solely by month of birth may not generalize to variations that are induced by changes in educational policy, and so on. The loss in external validity that arises in the application of the natural-experiment approach is particularly troublesome for population research, where there is a strong tradition of working at the population level and of establishing population-level relationships. The lack of generalizability that is inherent in the natural-experiment approach works very much against that tradition.

The issue of external validity is often also raised in discussions of twin and sibling models, which always necessarily require subsampling to the population of families with twins or siblings. That population may be different from the larger population and may have different responsiveness to the T_i of interest. In addition, twin and sibling models require an additional subsampling down to families in which the twins or siblings have different values of T_i (e.g., where at least one sibling has a teenage birth and another does not, one graduates from college and another does not, and so on). Even if exogeneity (internal validity) holds, the results may not generalize to the larger population of interest.

The problem of external validity is closely related to the problem of extrapolation, which means the applicability of the results to data points outside the sample actually used in the analysis. This problem arises for any Z_i in two distinct ways. First, any Z_i necessarily represents only one type of policy, environment, demographic type, or other force. The structural model written in Eqs. (1) and (2) implies, to the contrary, that there exists an effect, β_i, which is "the" effect of T_i, and is independent of the particular cause that induces T_i to change. Yet the approach to causal modeling discussed in this article cannot guarantee that any such single effect exists because the only type of effect estimated is the effect of variations in T_i that are induced by a particular Z_i. While it may be the case that a generalizable effect has been estimated, the argument that it has happened must necessarily be an assertion without any direct evidence. The question of whether there exist single, generalizable effects that are independent of the particular causes that induce them is a deep question, not easily resolved.

Second, a narrower issue of extrapolation necessarily arises for the reason, noted earlier, that any Z_i induces variation in T_i only within a particular range (e.g., from 30% to 40% in the smoking example). Extrapolation to the rest of the population requires additional assumptions beyond those used to estimate effects on the sample used in the analysis. Other things being equal, therefore, Z_i that induce greater variation in T_i are to be preferred to those that induce lesser variation. Unfortunately, there is typically a trade-off between the strength of the argument for internal validity and the range of variation in T_i induced; this trade-off is clearest in the natural-experiment approach, which often generates only small variation in T_i. If generalizability is one of the major goals of a particular

analysis, some type of implicit weighting is needed that trades off the bias resulting from the loss of internal validity against the gains in external validity resulting from obtaining a wider spread of T_i. It is fair to say that this problem has not really been addressed in the applied research community.

In conclusion to this section, it should be clear that the search for candidate Z_i variables is typically a difficult one. Indeed, in many instances, there may be no arguably strong Z_i available at all, either in the particular data sets available or even in the real world. In this case, much current thinking in economics is to pull back and admit that the effect of interest is essentially unknowable with any level of confidence and to turn one's attention to other issues that are more tractable. This thinking has led to a change in the level of ambition of empirical work, to more-modest claims for knowledge and certainty, and to a contraction in what may be called the outer boundaries of established causal relationships.

ILLUSTRATIONS

It is impossible within normal constraints on length to do much in the way of illustrations. However, this section presents a brief discussion of three examples, the first at somewhat greater length than the other two.

The Effect of Teenage Childbearing on Child Outcomes

There is a significant literature in the United States on the causes and consequences of teenage childbearing, particularly childbearing out of wedlock (Maynard 1997; U.S. Department of Health and Human Services 1995; Wu and Wolfe 2001). It is often feared that having children too early may hurt the educational, income, and possibly marital prospects of the mother and, in turn, may disadvantage the children who were born to such mothers. One of the typical outcomes for a child is cognitive ability at, say, age 5.

The first issue that must be addressed, as always, is the definition of the counterfactual. In this case, it must be the cognitive ability of a child at age 5 if his or her mother had delayed having the child until the mother was older and not a teenager (say, at age 20 or older). The clear selectivity issue is that a simple comparison of cognitive outcomes of children who are born to teenage mothers and to nonteenage mothers may be biased because of unobserved differences between the mothers. Teenage mothers may have lower incomes; come from more-disadvantaged backgrounds; and may even, on an individual level, be disproportionately composed of women who have less ambition for conventional educational and social-betterment goals. Some of these factors can be controlled for by family background or individual background variables, but there is no doubt much that cannot be controlled for.

An issue for this counterfactual is what else is held constant when imagining the postponement of birth, and therefore what mechanism is aimed to be captured. The problem is that there are many different mechanisms. Women who postpone childbearing may have different outcomes for education, earnings, and family income and other changes in family structure (i.e., marriage or cohabitation). But, in addition, they may simply be older and more mature, which could have its own effects on parenting ability, which will, in turn, affect the child in question. Most studies in this literature that have attempted to address the endogeneity problem have not been able to determine the mechanism by which postponement affects outcomes, which makes it difficult to interpret the results.[22]

22. Note that the mechanisms referred to here are those that intervene between T_i and Y_i in Eq. (1) to generate an effect. This is different from T_i being the mechanism itself, which is the way mechanisms were discussed previously. Alternatively, one could define T_i as some intervening outcome—say, educational attainment—and then define Z_i as the postponement of birth; but then another equation would be needed for a "Z_i" to deal with the endogeneity of the postponement of birth. The conclusion should be that a larger set of equations is needed in this case.

The multiplicity of mechanisms is closely tied, in this case, to the issue of whether all causes of postponement would have the same effect (the issue discussed earlier). If the cause of postponement is an increase in the resources put into local schools, thereby improving their quality and making educational continuation more attractive, it will clearly affect educational attainment. Because education has its own effects on an individual's breadth of perspective, it could affect child rearing. If the cause of the postponement is a sudden increase in the number of jobs available in the local labor market, women may drop out of school to take jobs but postpone births because they are able to work. This cause is likely to have different effects on the woman's subsequent life experiences. Another cause of postponement could be an improvement in contraceptive technology. In this case, women could continue sexual activity without pregnancy but whether doing so would lead to additional educational or employment outcomes depends very much on the attractiveness of these two alternative uses of time and how many women would avail themselves of it. If the cause of postponement is a reduction in the supply of available sexual partners, it would have obviously different effects on future marriage rates than if the cause is one of the others mentioned. If the cause of postponement is a media campaign or national campaign to reduce teenage pregnancy, the effects could again depend on what the affected women would choose to do with their lives in the absence of having children while young. It would seem, therefore, that the simple question of what "the" effect of the postponement of birth is on child outcomes is ill posed and needs a theoretical framework within which the reason for postponement is part of the model and is tied to specific mechanisms by which the effect will arise.[23]

The search for a candidate Z_i consists of looking for a variable that induces the postponement of birth for reasons that are not related to child outcomes. Ecological or environmental variables that may affect postponement are many: labor market opportunities, the quality of education in the area, and the sex ratio are three candidates that are tied to the specific causes mentioned earlier. However, these three are unlikely to satisfy the exclusion requirement that they not affect outcomes in more direct ways, which they are likely to do. Two ecological variables that have been argued to satisfy this criterion are the local availability of contraception and state abortion policy. Both may affect teenage birth rates and, at least at first blush, do not appear to affect child outcomes directly. As with all ecological variables, there is some danger that states and areas with a different availability of contraceptives and abortion policies differ in some other unobserved way, leading to spuriously estimated effects. Examining changes in child outcomes over time in response to changes in the availability of contraceptives or changes in abortion policy must confront the issue of time-varying unobservables and lagged responses. Changes in the availability of contraceptives and abortion policy may also have effects on the rate of teenage childbearing that are not large, with the consequence that the results do not generalize to what these effects would be for larger segments of the population. As I emphasized in relation to the random-coefficient model, the estimated effects on teenage childbearing using these instruments may reflect a set of β_i that are different from those of teenagers who are affected by a different policy change.

For this question, sibling and twin models are clearly available as well and have indeed been applied (Geronimus and Korenman 1992; Grogger and Bronars 1993; Hoffman, Foster, and Furstenberg 1993), albeit for a case in which the outcome for the mother was the object of interest, rather than that of the child (see Geronimus, Korenman, and

23. An "economic" framework would be one in which women weigh the relative costs and benefits of postponing childbearing, where both costs and benefits include a long list of complex and uncertain factors involving prospects in the labor market and returns to education, the quality of education, the availability of sexual partners and the attractiveness of mates, the availability of contraceptives, and social norms in a community related to teenage childbearing, for example.

Hillemeier 1994 for a cousin comparison). The strengths and weaknesses of the approach for this application are the same as those mentioned in general. The analysis must be conducted on a relatively small subset of the population—that with at least two siblings, one of whom has had a teenage birth and one of whom has had a nonteenage birth—and it must be the case that there are no within-family, individual differences between the women that led them to make different birth-timing choices that may be correlated with the outcomes for their children. The latter assumption seems particularly strong.

A natural-experiment Z_i that has been used (although, again, for adult, rather than child, outcomes) is the occurrence of miscarriage (Hotz, McElroy, and Sanders 1997). If miscarriage is random, then reduced-form estimates are valid estimates of the effect of miscarriage; if miscarriage is also excluded from the Y equation, then the mechanism—the postponement of birth—has been correctly identified. Whether miscarriage is random is an issue of whether it is a result of biological, behavioral, or environmental factors or, even if the former, whether these biological factors are correlated with behavioral or environmental factors (which are likely to be correlated directly with maternal and child outcomes). The excludability of miscarriage from the main equation requires the assumption that having a miscarriage has no direct effects on the mother or the child born later, either economic or affective. Questions of external validity would arise if the population of mothers with miscarriages is a specialized subset of the population and if the change in the percentage of teenage births resulting from a particular miscarriage rate is small.[24]

This short review of this literature has revealed how difficult addressing causal problems in this area can be. With a large number of possible mechanisms, with the strong possibility that different causes would have different effects, and with the problems of finding excludable and relevant Z_i, progress is likely to be difficult in gaining traction on the question.

Effects of Economic Growth and of Race

Two other illustrations, even more briefly reviewed, nevertheless reveal the commonality of the themes that emerged in the example of teenage childbearing. The first is a more aggregate-level question: how economic growth affects rates of internal migration (Kelley and Williamson 1984; Kuznets 1966). Working at the aggregate level, rather than the individual level, generates additional problems of multiple unobservables (since different countries are so different) and small sample sizes (the number of countries is limited). It makes reasonable sense to consider possible exogenous shocks to economic growth and to consider the effects of these shocks on internal migration, particularly on rural-to-urban migration. However, given the complexity of society-wide, as compared with individual, processes, it is even more an issue in this case to ask whether all mechanisms by which economic growth may be stimulated would have the same effects on internal migration. On the one hand, different exogenous technological shocks, for example, introduced, say, from a foreign country, will have different effects on relative productivity in urban and rural areas and therefore different effects on internal migration. On the other hand, increases in economic growth that arise from an improved management of the economy by political and economic leaders will have different effects on urban-rural migration, depending on the mechanism by which the improvement occurs. Therefore, it could be argued that this question is likewise ill defined, given the importance of the mechanism in defining the counterfactual.

The endogeneity problem would appear to be potentially important in this case as well. A relaxation of constraints on internal migration, for example, can lead to increased productivity as individuals are capable of moving to higher-productivity locations, leading

24. Hotz, Mullin, and Sanders (1997) later conducted a sensitivity analysis (bounds analysis) to some of these potential problems.

in turn to higher economic growth. Alternatively, the government may follow policies that affect some third factor that independently affects economic growth and internal migration, inducing a spurious relationship between them.

The best conceptual framework for the analysis of this question, it would seem, is the economic model of simultaneous equations, which is somewhat different from the causal model described earlier but is an extension of it to allow full reverse causality. Furthermore, given the multiple unobservables in simple cross-sectional, cross-area, or cross-country regressions, some type of area fixed-effect regression would seem essential. Identification of the effects of the model requires exclusions from both equations, that is, changing variables that affect changes in economic growth that do not affect changes in internal migration, and vice versa.

This having been said, the study of this relationship is so complex, the data limitations are so severe, and the inherent limitations on sample size are so strong that there has to be a correspondingly modest expectation that the effects of interest can be credibly isolated. While establishing correlations among variables is important, finding exclusions that have strong statistical power and that are strongly and credibly exogenous is likely to be difficult. One may conclude in this case that more progress on the study of specific aspects of the relationship between economic growth and internal migration may be best pursued at the micro level, where specific causal effects in both directions involving specific mechanisms could be studied with somewhat more confidence, although with less generalizability. In fact, there has been a great deal of microdata work on migration and its economic and other determinants that can be interpreted as fitting into this category, although it has rarely treated endogeneity in the way described in this article by searching for a Z_i to adjust for that endogeneity.

Another, still different, causal question is the effect of race on educational attainment. The effect of race on any particular outcome variable raises issues of the definition of causal effects. Some have argued that because race cannot be randomized, it cannot be defined as a causal variable (Holland 1986). However, this is a mistaken application of the experimental analogy, and the more basic counterfactual analogy is the superior and more general one. It does make conceptual sense to imagine that, at any point in the lifetime of, say, an African American, having experienced everything she or he has experienced up to that time, her or his skin color were changed to white (this is sometimes called a gedanken, or thought, experiment). Although it is a well-defined question, it may nevertheless be unanswerable, and it may not even be the main question of interest. For example, would the individual in question move to a different neighborhood, live in a different family, and go to a different school? If not, the question is not very interesting. It may make more sense in adult life to imagine that an individual had a different skin color when visiting an employer or a landlord, for example, holding other things fixed—and this has, indeed, been an experimental variable in audit studies that have tested for racial discrimination. But for many of the questions about the effect of race on, say, educational attainment, researchers have in mind the difference between the total effects of growing up in a household of a different race, with all the observables and unobservables that go with it. Although discrimination is certainly a key object of interest, of equal interest is the determination of how the other differences affect child outcomes like education. Since there are so many myriad uncontrollable differences between individuals who grow up in households of different races, the "pure" racial effect is a causal question that probably cannot be answered with any reasonable degree of certainty.[25]

---

25. See Heckman (2004) for a critique of Holland's (1986) assertion and Smith (2003) for a different perspective on this issue. Note that in the initial question in this case—the effect of an individual's race on one of her or his outcomes—race is clearly a Z_i, not a T_i. The issues addressed in the theory section are relevant to the extent that one is interested in interpreting the reason for the difference (i.e., the mechanisms, which are the T_i).

But this merely means that descriptive work on racial differences in background and other family variables and in the quality of schools and neighborhoods should be examined in a search for explanations for racial differences in educational attainment. These other variables could explain part of the difference, and it would certainly be an interesting finding to discover that there is no racial difference in educational attainment once observables like these were controlled.

CONCLUSIONS

This review of the analysis of causal relationships illustrates the difficulties in reaching conclusions about those relationships. There are few cases in which a single, clearly superior method of addressing the traditional self-selection and endogeneity problem can be found, and most of the methods that have been used in the past are open to serious objection. This situation has led to a more-pessimistic view of the progress of the field and of the prospects for increasing knowledge in the future and signals a retreat both in what is known and even what can be known. Modesty of claims for truth is the clear lesson from this review.

The direction in which it is best to proceed in light of these difficulties is the question that will be addressed by the discipline in the future. To be sure, the solution is not to return to ignoring the problem. Indeed, given that average practice at the moment often still ignores the endogeneity problem, or addresses it in ways that are now thought to be inadequate, the best direction in which to move is to raise the standards for acceptable solutions and to be more discriminating and careful in their use. More attention to theory, to mechanisms, and to the possible threats to exclusion restrictions is needed to change the tone of research toward a more cautious and considered approach.

At the same time, there is a danger in maximizing internal validity at the expense of external validity. To do so would lead to a field consisting only of narrowly defined exercises without generalizability and to a collection of miscellaneous facts that do not add up to any general knowledge. Such an approach is intellectually constricting and does not allow for the pursuit of the larger social science questions and the explanation of society-wide trends that motivates much of social science research.

To achieve the correct balance requires a weighing of evidence from different studies with different strengths and weaknesses to achieve a consensus. To be sure, it is fair to rule some types of evidence completely out of court if the threats to their inference are too severe, but there still remains a larger set of studies with different mixes of internal and external validity that should be given some positive weight when drawing conclusions on what the best estimate is. This approach calls for a wide variety of evidence and for more research that attempts to reconcile the different findings of different studies and to synthesize literatures in which different approaches are taken. Areas in which a variety of approaches yield similar results are those in which synthesis should be easy, and areas in which a variety of approaches yield dissimilar results are those in which synthesis should be the most difficult. Formal theory, as well as informal evidence from ethnographic and other accounts, should also be given positive weight. Such a programmatic, considered, and synthesis approach to research questions is likely to yield the most progress in light of the impossibility that any single study or set of studies can provide the best answer.

REFERENCES

Adams, P., M. Hurd, D. McFadden, A. Merrill, and T. Ribeiro. 2003. "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths Between Health and Socioeconomic Status." *Journal of Econometrics* 112:3–56.

Angrist, J., G. Imbens, and D. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91:444–55.

Angrist, J. and A. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106:979–1014.

Björklund, A. and R. Moffitt. 1987. "Estimation of Wage Gains and Welfare Gains in Self-Selection Models." *Review of Economics and Statistics* 69:42–49.

Bound, J., D. Jaeger, and R. Baker. 1995. "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American Statistical Association* 90:443–50.

Bound, J. and G. Solon. 1999. "Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling." *Economics of Education Review* 18:169–82.

Duncan, G., K. Magnuson, and J. Ludwig. 2003. "The Endogeneity Problem in Developmental Studies." Mimeographed. Institute for Policy Research, Northwestern University, Evanston, IL.

Fricke, T. 2003. "Culture and Causality: An Anthropological Comment." *Population and Development Review* 29:470–79.

Geronimus, A. and S. Korenman. 1992. "The Socioeconomic Consequences of Teen Childbearing Reconsidered." *Quarterly Journal of Economics* 107:1187–214.

Geronimus, A., S. Korenman, and M. Hillemeier. 1994. "Does Maternal Age Adversely Affect Child Development? Evidence From Cousin Comparisons in the United States." *Population and Development Review* 20:585–609.

Grogger, J. and S. Bronars. 1993. "The Socioeconomic Consequences of Teen Childbearing: Findings From a Natural Experiment." *Family Planning Perspectives* 25:156–61.

Haavelmo, T. 1943. "The Statistical Implications of a System of Simultaneous Equations." *Econometrica* 11:1–12.

———. 1944. "The Probability Approach in Econometrics." *Econometrica* 12(Suppl.):1–115.

Heckman, J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46:931–60.

———. 2000. "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective." *Quarterly Journal of Economics* 115:45–97.

———. 2004. "The Scientific Model of Causality." Mimeographed. Department of Economics, University of Chicago.

Heckman, J. and S. Navarro-Lozano. 2004. "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models." *Review of Economics and Statistics* 86: 30–57.

Heckman, J. and R. Robb. 1985. "Alternative Models for Evaluating the Impact of Interventions." Pp. 156–245 in *Longitudinal Analysis of Labor Market Data,* edited by J. Heckman and B. Singer. Cambridge, England: Cambridge University Press.

Heckman, J. and E. Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences* 96:4730–34.

———. 2001. "Policy-Relevant Treatment Effects." *American Economic Review* 91:107–11.

———. Forthcoming. "The Econometric Evaluation of Social Programs." In *Handbook of Econometrics, Vol. 6,* edited by J. Heckman and E. Leamer. New York: Elsevier.

Hoffman, S., E.M. Foster, and F. Furstenberg. 1993. "Re-evaluating the Costs of Teenage Childbearing." *Demography* 30:1–13.

Holland, P. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81:945–60.

Hotz, V.J., S. McElroy, and S. Sanders. 1997. "The Impacts of Teenage Childbearing on the Mothers and the Consequences of Those Impacts for the Government." Pp. 55–94 in *Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy,* edited by R. Maynard. Washington, DC: Urban Institute Press.

Hotz, V.J., C. Mullin, and S. Sanders. 1997. "Bounding Causal Effects Using Data From a Contaminated Natural Experiment: Analysing the Effect of Teenage Childbearing." *Review of Economic Studies* 64:575–603.

Imbens, G. and J. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62:467–76.

Kelley, A. and J. Williamson. 1984. *What Drives Third World City Growth? A Dynamic General Equilibrium Approach*. Princeton, NJ: Princeton University Press.

Krueger, A. and J.-S. Pischke. 1992. "The Effect of Social Security on Labor Supply: A Cohort Analysis of the Notch Generation." *Journal of Labor Economics* 10:412–37.

Kuznets, S. 1966. *Modern Economic Growth*. New Haven, CT: Yale University Press.

Manski, C. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.

Maynard, R., ed. 1997. *Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy*. Washington, DC: Urban Institute.

McFadden, D. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." Pp. 105–42 in *Frontiers in Econometrics,* edited by P. Zarembka. New York: Academic Press.

Moffitt, R. 1991. "Program Evaluation With Nonexperimental Data." *Evaluation Review* 15: 291–314.

———. 2003. "Causal Analysis in Population Research: An Economist's Perspective." *Population and Development Review* 29:448–58.

Neyman, J. 1923. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles." *Roczniki Nauk Rolniczych* 10:1–51 [in Polish], English translation of Section 9 by D.M. Dabrowska and T.P. Speed (1990), *Statistical Science* 9:465–80.

———. 1935. "Statistical Problems in Agricultural Experimentation." *Supplement to the Journal of the Royal Statistical Society* 2:107–80.

Quandt, R. 1972. "Methods for Estimating Switching Regressions." *Journal of the American Statistical Association* 67:306–10.

Robins, J.M. 1999. "Association, Causation, and Marginal Structural Models. *Synthese* 121: 151–79.

Rosenbaum, P. 1995. *Observational Studies*. New York: Springer-Verlag.

Rosenzweig, M. and K. Wolpin. 2000. "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature* 38:827–74.

Roy, A. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3: 135–46.

Rubin, D. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66:688–701.

———. 1978. "Bayesian Inference for Causal Effects." *Annals of Statistics* 6:34–58.

Schultz, T.P. 1985. "Changing World Prices, Women's Wages, and the Fertility Transition: Sweden 1860–1910." *Journal of Political Economy* 93:1126–54.

Smith, H. 2003. "Some Thoughts on Causation as It Relates to Demography and Population Studies." *Population and Development Review* 29:459–69.

Stock, J., J. Wright, and M. Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business and Economic Statistics* 20:518–29.

U.S. Department of Health and Human Services. 1995. *Report to Congress on Out-of-Wedlock Childbearing*. Washington, DC: U.S. Government Printing Office.

Winship, C. and S. Morgan. 1999. "The Estimation of Causal Effects From Observational Data." *Annual Review of Sociology* 25:659–706

Wu, L. and B. Wolfe, eds. 2001. *Out of Wedlock: The Causes and Consequences of Nonmarital Fertility*. New York: Russell Sage Foundation.