# Causal Inference in Sociological Research

Markus Gangl

Department of Sociology, University of Wisconsin, Madison, Wisconsin 53706-1320; email: mgangl@ssc.wisc.edu



Annu. Rev. Sociol. 2010. 36:21–47

First published online as a Review in Advance on April 5, 2010

The *Annual Review of Sociology* is online at soc.annualreviews.org

This article's doi: 10.1146/annurev.soc.012809.102702



0360-0572/10/0811-0021$20.00

## Key Words

counterfactual model, treatment effects, identification, endogeneity, unobserved heterogeneity, nonparametric estimation

## Abstract

Originating in econometrics and statistics, the counterfactual model provides a natural framework for clarifying the requirements for valid causal inference in the social sciences. This article presents the basic potential outcomes model and discusses the main approaches to identification in social science research. It then addresses approaches to the statistical estimation of treatment effects either under unconfoundedness or in the presence of unmeasured heterogeneity. As an update to Winship & Morgan's (1999) earlier review, the article summarizes the more recent literature that is characterized by a broader range of estimands of interest, a renewed interest in exploiting experimental and quasi-experimental designs, and important progress in the areas of semi- and nonparametric estimation of treatment effects, difference-in-differences estimation, and instrumental variable estimation. The review concludes by highlighting implications of the recent econometric and statistical literature for sociological research practice.

> Avoiding causal language when causality is the real subject of our investigation either renders the research irrelevant or permits it to be undisciplined by the rules of scientific inference.... Rather we should draw causal inferences where they seem appropriate but also provide the reader with the best and most honest estimate of the uncertainty of that inference. (King et al. 1994, p. 76)

## 1. INTRODUCTION

Empirical research makes three main contributions to the sociological enterprise. It describes the structure of populations, social relations, and processes; it suggests interesting areas for theoretical development through exploration of novel observations; and it provides tools to assess the validity of theoretical claims and implications. Testing the empirical content of hypotheses is the purview of causal inference, i.e., the attempt to recover causal parameters that describe social processes of interest from empirically observed data (Heckman 2000).

Originating in the statistical (e.g., Holland 1986; Rosenbaum 2002; Rubin 2005, 2006) and econometrics literature (see Heckman 2000, 2001, 2005; Heckman & Vytlacil 2007a,b; Manski 1995, 2007), the counterfactual, Rubin, or potential outcomes model of causality has, over the past three decades, become the standard conceptual tool to unify the notion of causality, to understand the identification problem at the heart of causal inference, and to assess the utility of alternative estimation techniques (Sobel 2005). In this spirit, I first present the essentials of the potential outcomes framework and its application to prototypical identification strategies in sociology. The main part of the review is then devoted to the statistical estimation of causal effects under alternative data-generating scenarios, distinguishing in particular between cases in which the assumption of (conditional) unconfoundedness of treatment may be maintained and cases in which relevant confounders of treatment are unobserved. In the concluding section, I discuss implications for sociological research practice, notably with respect to the central role of the identification problem, causal inference in the presence of social interactions, and the causal status of key analytical concepts in sociology.

As an update to Winship & Morgan's (1999) earlier paper, this review is intended specifically to reflect the past decade's renewed interest in alternative estimands, identification through experimental, quasi-experimental, and observational designs, as well as progress in the areas of semi- and nonparametric estimation methods, difference-in-differences, and fixed-effects models and the reappraisal of instrumental variables (IV) estimation methods. For the most part, this review is restricted to summarizing the main results of a vast and growing interdisciplinary literature, although I provide references to both the technical literature and applications of sociological interest throughout; as more detailed and applied introductions, Morgan & Winship (2007), Firebaugh (2008), Angrist & Pischke (2009), Blundell & Dias (2009), Imbens & Rubin (2010), and Wooldridge (2002) are especially useful.

## 2. THE COUNTERFACTUAL FRAMEWORK

The core of the counterfactual model of causal inference is the notion of potential outcomes $Y_D$ associated with a range of causal states $d \in D$. For each unit of observation $i$, it is the realized causal state $D = d$ that determines which specific $Y_{D=d}$ is actually experienced by $i$, and can hence be in principle observed by an empirical researcher. All potential outcomes $Y_{D\neq d}$ associated with any other causal state $D \neq d$ do not materialize and are thus hypothetical (counterfactual) at the level of $i$. However, the precise task of causal inference is to learn whether the fact that unit $i$ experienced treatment (event, exposure, condition) $D = d$ instead of some other treatment $D \neq d$ implied a difference in outcomes $Y_D$. The respective comparison is counterfactual by necessity because it involves a comparison between an observable event $(Y_{D=d}, D = d)$ with one or more events $(Y_{D\neq d}, D \neq d)$ that are unobservable in

principle. Causal inference, in other words, is equivalent to answering the "what if" question about the expected change in outcomes $Y$ if unit $i$ had experienced event $D \neq d$ instead of $D = d$.

Without loss of generality, the essential logic of the potential outcomes framework is best seen for the canonical case of a dichotomous event $D \in \{0,1\}$. Here the key building block of any causal analysis is the unit (treatment) effect

$$\Delta_i \equiv (Y_i \mid D_i = 1) - (Y_i \mid D_i = 0) \qquad 1.$$

that describes the difference in outcome $Y_i$ as unit $i$ either experiences the event $D = 1$ or does not (i.e., $D = 0$). Clearly, just one of these different outcomes will be realized by any one agent $i$, and hence eventually observed by the empirical sociologist. The unit effects $\Delta_i$ or any derived quantity are therefore unobservable in principle but need to be estimated from empirically observable data under auxiliary identifying assumptions. In this, the pairwise comparison in Equation 1 underscores the completely nonparametric nature of the potential outcomes framework so that any assumption on functional form—linearity, for example—is properly identified as an assumption imposed for convenience, parsimony, or tractability during the statistical analysis.

With the unit effect as its building block, causal inference in the counterfactual tradition most naturally proceeds according to the logic of an "effects-of-causes" analysis (Holland 1986), which aims to identify and estimate the effect of a specific manipulation $D$ that defines the causal parameter of interest, while relegating the much more challenging goal of addressing the relative role of alternative causes in explaining outcomes to a secondary issue. In part, these priorities reflect epistemic pragmatism, as it may not be realistic to secure simultaneous identification of multiple effects in any given population of interest. Besides, this concern is increasing in relevance with essential heterogeneity of treatment effects where the unit effects become the structurally invariant parameters of interest, whereas any derived quantity that can be estimated under practically feasible identification conditions has to be seen as a population parameter.

## 2.1. The Average Treatment Effect

Another way of putting this is that, for example, the coefficient for $D$ in a regression of $Y$ can (at best) be seen as one particular summary statistic of the distribution of unit effects $\Delta_i$ in the population of interest. With heterogeneity of treatment effects, there is no single number that would provide the causal effect of $D$ on $Y$, much as any single quantity may describe specific features of a distribution $F(.)$, but not characterize it completely. Determining which causal parameter(s) to focus on in any given study thus becomes a first theoretical issue to be settled by the analyst.

Conventionally, many analyses continue to focus on the mean impact of treatment $D$, i.e., are interested in estimating the average treatment effect (ATE) of $D$ on $Y$. In the above notation, and in the canonical case of a binary treatment indicator, the ATE is defined as

$$\Delta_{ATE} \equiv E(\Delta_i) = E[(Y_i \mid D_i = 1) - (Y_i \mid D_i = 0)]. \qquad 2.$$

It is often also of interest to estimate the ATE within interesting subpopulations defined by observable covariates $X$. In this case, the conditional ATE for group $X = x$ becomes

$$\Delta_{CATE(x)} \equiv E(\Delta_i \mid X_i = x) = E[(Y_i \mid D_i = 1) - (Y_i \mid D_i = 0)] \mid X_i = x. \qquad 3.$$

Furthermore, the overall and the subgroup-specific ATE parameters are related by

$$\Delta_{ATE} = \frac{1}{N_k} \sum_k \Pr(X = x_k) \cdot \Delta_{CATE(x_k)}, \qquad 4.$$

i.e., the overall ATE is the weighted average of the $N_k$ subgroup-specific CATE parameters, where the weights correspond to the population share of each subgroup $k$.[1]

[1] In practice, empirical researchers would of course be estimating the sample analogs to these parameters. To keep the discussion focused on the key conceptual issues, I do not

## 2.2. Alternative Estimands

Often, however, the average treatment effect on the treated (ATT), defined as

$$\Delta_{ATT} \equiv E(\Delta_i \mid D_i = 1) = E[(Y_i \mid D_i = 1) - (Y_i \mid D_i = 0)] \mid D_i = 1, \quad 5.$$

may be both easier to identify and theoretically more informative as it describes the impact of treatment $D$ only among those units $i$ who were actually exposed to it. Since evaluating the impact of actually experienced conditions $D$ on outcomes $Y$ constitutes a (in sociology, plausibly the) cornerstone of causal analysis (Heckman 2005), the ATT is often likely to be the parameter of real interest—whether it describes the difference a training program makes for participants' subsequent career prospects (Dehejia & Wahba 2002, Smith & Todd 2005), the effect of divorce on child development for those children actually experiencing the separation of their parents (Ní Bhrolcháin 2001), or the impact of events such as illness or unemployment on the career trajectories of those workers experiencing these events (Brand & Xie 2007, Brand 2006, Gangl 2006). By perfect analogy to Section 2.1, conditional ATT parameters may be defined for subgroups with observable covariates $X = x$ (see Morgan 2001, Brand 2006 for applications), whereas it might be of interest in other applications to examine the (conditional) average treatment effect on the untreated (ATU), which can be defined by conditioning the expectation in Equation 5 on $D_i = 0$ instead of $D_i = 1$. Closely related to the ATT parameter is the local average treatment effect (LATE; see Imbens & Angrist 1994, Angrist et al. 1996) and the marginal treatment effect (MTE; see Heckman & Vytlacil 2005, 2007b), both of which are discussed in the context of IV estimation in Section 5.4 below.

Also, substantive theory might suggest examining treatment impacts on criteria other than mean outcomes or their equivalent, the average treatment effect. In that case, the quantile treatment effect (QTE)

$$\Delta_q \equiv F_q^{-1}(Y_i \mid D_i = 1) - F_q^{-1}(Y_i \mid D_i = 0) \quad 6.$$

or its close cousin, the quantile treatment effect on the treated (QTT), is useful for examining differences in outcome distributions at different quantiles $q$, e.g., the median, the quartiles, or the deciles of the distribution. Gangl's (2004) analysis of the relationship between unemployment benefits and workers' postunemployment wages is one example where theory suggests an impact on the lower tail rather than merely on the mean of the wage distribution, and Bitler et al. (2006) have estimated quantile treatment effects in evaluating the distributional impact of welfare reform. Koenker (2005) provides a canonical overview of quantile regression methods, and recent econometric work has begun to develop respective instrumental variables estimators (e.g., Chernozhukov & Hansen 2006).

Finally, one might also want to examine distributional treatment effects of the form

$$\Delta_{d.q} \equiv F_q^{-1}(\Delta_i) = F_q^{-1}[(Y_i \mid D_i = 1) - (Y_i \mid D_i = 0)], \quad 7.$$

i.e., different quantiles $q$ of the joint outcome distribution $(Y_i \mid D_i = 1, Y_i \mid D_i = 0)$. This examination inevitably requires stronger identification assumptions than are necessary for any of the other parameters discussed so far because, unlike in the case of the average treatment effect, $\Delta_{d.q}$ will, in general, not be equivalent to $\Delta_q$. Heckman et al. (1997b) and Abbring & Heckman (2007) have developed a thorough econometric framework for this purpose, achieving identification either through bounding, revealed preference, or rationality assumptions; Carneiro et al. (2003) and Cunha et al. (2006) are examples of this line of work in applications to estimate the distributions of heterogeneous returns to schooling. Using more conventional (sequential) conditional independence assumptions for identification,

distinguish between population and sample parameters in this review. Likewise, issues concerning statistical inference will not be systematically treated here due to space constraints; the interested reader is referred to Imbens (2004), Imbens & Wooldridge (2009), or Imai et al. (2008) for more comprehensive recent reviews.

Gangl (2006) provides a distributional analysis of the scar effects of unemployment, and Morgan & Todd (2008) have recently proposed a general procedure to assess the extent of variability in the distribution of treatment effects. A detailed review of the different approaches is beyond the scope of this article, however.

Although any detailed discussion is equally beyond this review, note that while the discussion so far has assumed the static case where treatment status $D$ is experienced at one point in time and outcomes $Y$ are assessed at another point in time, the overall framework is sufficiently flexible to accommodate dynamic settings. Temporal variation in effect size may easily be assessed from examining outcomes $Y$ at different time points $T = 1, \ldots, t$ after exposure to treatment, resulting in a vector of ATE (or other) parameters that needs to be estimated. Similarly, different treatment effects may be defined depending on the biographical or historical time point of treatment exposure (see Brand & Xie 2007 for an exposition of both issues). In some cases, treatment exposure has to be considered dynamic in the more general sense that past covariates and outcomes determine current treatment status, and Robins et al. (2000), Gill & Robins (2001), and Abbring & Heckman (2007) propose alternative identification and estimation strategies in this case.

## 3. IDENTIFICATION AND RESEARCH DESIGN

Given the "Fundamental Problem" (Holland 1986, p. 947) that causal effects are defined from the comparison of unobservable joint outcomes $(Y_{D=1}, Y_{D=0})$, causal inference is an effort to use observable empirical data as a valid substitute to the unobservable (counterfactual) outcome information in order to estimate the causal effect of interest. To take the case of the average treatment effect as an example, the decomposition

$$\begin{aligned}\Delta_{ATE} &\equiv E[(Y_i \mid D_i = 1) - (Y_i \mid D_i = 0)] \\ &= E(Y_{D=1}) - E(Y_{D=0}) \\ &= \Pr(D = 1) \cdot [E(Y_{D=1} \mid D = 1) \\ &\quad - E(Y_{D=0} \mid D = 1)] - (1 - \Pr(D = 1)) \\ &\quad \cdot [E(Y_{D=1} \mid D = 0) - E(Y_{D=0} \mid D = 0)] \end{aligned} \quad 8.$$

illustrates that five quantities are needed to obtain the ATE estimate (see Winship & Morgan 1999, p. 667); empirically, however, $E(Y_{D=0} \mid D = 1)$ and $E(Y_{D=1} \mid D = 0)$—the expected outcomes among treated cases if they had not been treated and the expected outcomes among nontreated units had they been treated—are unknown in principle.

Moreover, because these quantities cannot, in general, be assumed to be equal to their observable counterparts $E(Y_{D=0} \mid D = 0)$ and $E(Y_{D=1} \mid D = 1)$, the straightforward comparison of average outcomes in the two groups (sometimes called the naive estimator) will be a biased estimator of the ATE parameter. The extent of resulting bias may be expressed through the decomposition

$$\begin{aligned}\Delta_{ATE} &\equiv E(Y_{D=1} \mid D = 1) - E(Y_{D=0} \mid D = 0) \\ &\quad - [E(Y_{D=0} \mid D = 1) - E(Y_{D=0} \mid D = 0)] \\ &\quad - (1 - \Pr(D = 1)) \cdot [E(\Delta_i \mid D = 1) \\ &\quad - E(\Delta_i \mid D = 0)] \end{aligned} \quad 9.$$

(see Morgan & Winship 2007, p. 46), where the ATE is the difference in observed mean outcomes among the treated and the nontreated (first line of Equation 9), corrected for the fact that the two groups might have seen different baseline outcomes even absent the intervention $D$ (second line) and the fact that there might be a systematic difference in the average impact of treatment between treated and nontreated units, e.g., because agents maximize expected utility when choosing treatment (third line). Equation 9 thus describes selection bias, or heterogeneity, and endogeneity, or self-selection bias, as the two principal sources of bias when recovering any causal parameter from empirical data.[2]

[2]Strictly speaking, endogeneity bias is an issue only for the control group if the ATT or a related parameter is to be estimated. In this case, treatment effects on the treated will generally be identified under somewhat weaker conditions

### 3.1. Identification Fundamentals

To resolve these biases in any empirical study, valid causal inference inevitably involves substantive knowledge about actual data-generating processes, namely the outcome process

$$Y_D = f_D(X, W) + U_D \qquad 10.$$

under treatment conditions $D = \{0,1\}$ and the process of treatment assignment

$$D = g(X, Z) + V, \qquad 11.$$

both of which depend on a range of factors $X$, $W$, $Z$, $V$, and $U$, the definition and roles of which are clarified below (see also Heckman & Robb 1985, 1986). Identification of the causal effect of $D$ can be achieved via estimation of an explicit behavioral model for Equations 10 and 11, known as structural estimation in economics (see Heckman 2000, 2005; but see also Logan 1996, 1998; Logan et al. 2008 for estimable behavioral models in sociology), or careful research design. Randomized experiments in particular eliminate bias, i.e., identify average treatment effects according to Equation 9, through combining the active manipulation of treatment—rendering $g(.)$ known to the analyst—with randomized treatment assignment—rendering $E[g(.)] = 0$ and ensuring independence of $U$ and $V$—which results in balancing expected outcomes net of treatment across experimental conditions (see Cook & Campbell 1979, Shadish et al. 2002, Rossi et al. 2004, Imai et al. 2008). At the same time, even with full randomization, identification of treatment effects while eschewing a full behavioral specification of Equations 10 and 11 necessarily entails the no interference or stable unit treatment value assumption (SUTVA; see Rubin 1978, 1986). As an existence assumption about unit treatment effects, SUTVA requires outcomes $Y$ to be independent of actual treatment assignment at both the individual level and within the larger population, thus ruling out Hawthorne or John Henry effects (e.g., Shadish et al. 2002) as well as social interactions, information diffusion, norm formation, and other macro (general equilibrium) effects in the determination of outcomes (Garfinkel et al. 1992, Sobel 2006). I maintain the assumption for the present but also discuss some of its implications below.[3]

### 3.2. Identification in Observational Studies

Short of estimable behavioral models or experimental manipulation, causal inference in observational studies, the mainstay of empirical social research, is complicated by the fact that observed events $D$ reflect naturally occurring assignment processes (Equation 11) that are socially structured, that reveal agents' choices, and that, in consequence, imply a correlation between treatment assignment and expected outcomes. In the resulting ex post facto design exhibiting nonequivalent control groups and nonrandom treatment assignment (Cook & Campbell 1979), valid causal inference requires that conditioning on observable covariates $X$ is sufficient to break the association between treatment assignment and outcomes. The necessary identification assumption is

$$U \perp\!\!\!\perp V \mid X \Leftrightarrow D \perp\!\!\!\perp (Y_{D=1}, Y_{D=0}, W, U, V) \mid X \qquad 12.$$

which is variously known as conditional independence (CIA), strict ignorability, exogeneity,

(see Heckman & Robb 1985, 1986), although the practical utility of this result is probably limited. Related to that, the analytical distinction between heterogeneity and endogeneity may also blur in practice because many social science covariates may actually be understood as indexing either antecedent conditions or (perceived) costs and (expected) benefits of treatment.

[3] In addition, the presence of heterogeneity in treatment effects limits the utility of randomized experiments in principle. For one thing, explicit manipulation of treatment status naturally discards any information contained in real-world patterns of treatment assignment, whether through self-selection, dropout, or noncompliance (see Heckman & Vytlacil 2007b). Also, experimental ATE estimates may be difficult to generalize if either the study population or the administered treatments cannot be considered representative of some larger context (Smith 1990, Heckman 1992, Shadish et al. 2002), and randomization also does not ensure the identification of treatment effect distributions beyond (conditional) average treatment effects (Heckman & Vytlacil 2007b).

or selection on observables. The assumption states that observed covariates $X$ comprise a sufficient set of joint causes of $D$ and $Y$ so that, conditional on covariates, variation in $D$ is as good as randomly assigned—i.e., is unrelated to either unobserved heterogeneity (omitted variable bias) or agents' purposive choice of treatment (endogeneity). Equation 12 identifies the ATE parameter because it asserts that baseline outcomes in the treatment group can be predicted using observed covariates (the second line of Equation 9 can be computed), whereas endogeneity bias (the third line of Equation 9) is absent. Equation 12 can be considerably relaxed when repeated observations are available (see Wooldridge 2002). In this case, maintaining additive separability of error terms in the outcome process $f(.)$ is sufficient to permit the use of fixed-effects or difference-in-differences estimators (see Section 5.3) that condition on all time-constant group-specific (with repeated cross-sectional data) or individual (with panel data) unobservables $\bar{U}$ , $\bar{V}$ that may affect treatment choice or outcomes.

Being this explicit about the CIA assumption required for causal inference in ex post facto designs is not merely about knowing what to assert in a more or less cavalier fashion when proceeding with the desired interpretations of results. Rather, the counterfactual perspective embodied in Equation 12 has direct implications for how causal analysis needs to be conducted on the basis of observational data, and these implications differ considerably from the current practice of much regression analysis in sociology. First and foremost, CIA is not a primarily statistical, but rather a fundamentally theoretical statement about the joint causes $X$ of $D$ and $Y$. Guidance about which covariates $X$ to include in a regression model therefore requires input from substantive theory to determine these joint causes, and, depending on the assumed theoretical model, analysts might even differ in their assessments about which specific covariates are considered critical to identify the causal effect of interest. The fundamental point remains, however, that explicit theory is required in order to assess whether the set of empirically available covariates $X$ in any data set is sufficient to justify CIA, and hence a causal interpretation of results. Moreover, the emphasis here is on theory supplying an understanding of both the assignment process (Equation 11) and the substantive outcome process (Equation 10) in order to determine joint causes of $D$ and $Y$ that need to be conditioned on in the empirical analysis. Compared with this requirement, the usual setup in sociology papers that list alternative causes of outcomes in their "theory" sections and then proceed to "test" the relative importance of "competing hypotheses" by simultaneously including a series of observed variables in a regression specification is woefully inadequate and is eventually unlikely to identify any causal effect of interest.[4]

Although the choice of $X$ is thus fundamentally a theoretical matter, the counterfactual perspective still does offer some important guidance in this respect (see Pearl 2000 for a canonical analysis). Given that the purpose of conditioning is to equalize expected outcomes across treatment groups absent treatment (i.e., to compute the second line of Equation 9), conditioning on $X$, the joint causes of $D$ and $Y$, is all that is required to identify the causal effect of interest. Pearl's (2000) analysis in particular is explicit in that the set of conditioning factors $X$ does not need to include all potential joint causes of $D$ and $Y$, but rather a minimally sufficient set of covariates—to be determined by substantive theory—that breaks the dependency between $D$ and $Y$, and thus renders residual variation in $D$ as good as randomly assigned with respect to $Y$ (thus satisfying

[4]Another way of putting this is that sociologists often implicitly follow the "all causes" strategy of structural modeling in economics (Heckman 2005), while ignoring that this approach requires covariates that describe strictly exogenous (structural, deep) causes of outcomes to be successful. Sociologists' typically eclectic use of both exogenous and endogenous covariates is actually more in tune with the counterfactual goal of balancing expected outcomes across treatment groups, which does not require covariates to be strictly exogenous (Heckman 2005, Heckman & Vytlacil 2007a), yet, as explained here, still requires more careful selection of covariates than evident in the prototypical empirical study in sociology.

the d-separation criterion of Pearl). One direct implication of this is that consideration of alternative causes $W$ that are relevant predictors of $Y$, but otherwise unrelated to $D$, is irrelevant if the purpose of the analysis is solely to identify the causal effect of $D$, although inclusion of $W$ may help with statistical precision of the resulting estimate. Inclusion of $W$ might even put the causal interpretation in jeopardy if $W$ is endogenous to $D$, i.e., if $W$ is an intervening mechanism that transmits the effect of $D$ on $Y$. In that case, the regression coefficient for $D$ is just the "direct" effect of $D$ on outcomes in the language of path analysis, net of its "indirect" effect via $W$, neither of which has a clear causal interpretation with treatment effect heterogeneity (Sobel 1995, 1998). From a counterfactual perspective, it is only the total effect of $D$ on $Y$ that has straightforward causal content.[5]

On a more subtle level, the counterfactual framework also implies that covariates like $Z$ that determine outcomes only through choice of treatment, i.e., variables that have no direct effect on $Y$ controlling for $D$, should not be included as conditioning factors in the empirical analysis. If $Z$ does predict treatment but not outcomes net of treatment, then $Z$ describes variation in treatment status that is exogenous with respect to outcomes, i.e., describes variation in treatment status that is as good as randomly assigned, at least conditional on covariates $X$. By (over)controlling for exogenous variation in $D$, conditioning on $Z$ might indeed destroy an excellent opportunity to identify the treatment effect. Variables like $Z$, which often result from natural experiments (see below), should instead be considered as candidate instruments in IV (see Section 5.4) or control function models (see Section 5.2).

Finally, Pearl's (2000) canonical analysis of identification conditions also warns against conditioning on collider variables (or endogenous covariates with multiple causes) in $X$ that are typically intended to proxy some unobserved factor $U$, because partial control of the causes of the collider (proxy) variable may inadvertently induce rather than alleviate dependency between treatment status and outcomes. Pearl's seminal example is the inclusion of a college admission variable in, say, a basic status attainment model: Even when (observed) high school grades and (unobserved) personality independently determine admission, controlling for observed admission decision creates a negative correlation between both factors in the sample, which, in turn, undermines the identification of the causal effect of interest even for personality traits that are otherwise thought to be unrelated to treatment but that have an independent effect on outcomes. The upshot is again that identification of causal effects with observational data requires an explicit theoretical model to determine an appropriate set of conditioning factors and a subsequent assessment of the extent to which available observable covariates fall short of the identifying information.

### 3.3. Natural Experiments

Given the challenges involved in drawing causal inferences from observational data, there has been a resurgence of interest in natural or quasi-experimental designs to identify causal effects across the social sciences in recent years (see Angrist & Krueger 2001 for an overview). Natural experiments come in different flavors, loosely comprising regression discontinuity designs, (multiple) interrupted time-series (ITS) designs, and nonequivalent control group designs in Cook & Campbell's (1979) terminology, but all share the idea of exploiting some external event or institutional condition that creates exogenous variation in the social process of interest. In ITS designs, this event typically is some form of shock to individuals or change in an institution or program, whereas in regression discontinuity designs, it is a known (deterministic or probabilistic, i.e., "fuzzy")

[5] Also, this implies that even where researchers may correctly claim that inclusion of covariates $X$ in a regression model identifies the effect of $D$ on $Y$, none of the regression coefficients for covariate $X$ possibly sustains a direct causal interpretation because all of them have at least been purged of their indirect effects on $Y$ through $D$.

institutional rule to assign or withhold treatment that creates the relevant variation (see Imbens & Lemieux 2008 for a recent review). In either case, the problem of endogenous treatment choice is circumvented, and the causal effect of a respective condition $Z$ is identified in line with Equation 12 whenever available covariates $X$ properly condition for both concomitant causes of change in $Y$ and population heterogeneity. Furthermore, longitudinal data and fixed-effects or difference-in-differences estimators may again be employed to condition on time-invariant unobservables and thus weaken the maintained CIA assumption.

While traditionally confined to program evaluation (Rossi et al. 2004; but see also Berk & Rauma 1983, Berk & de Leeuw 1999, Downey et al. 2004 for applications with broader sociological appeal), the newly increased interest in these designs has occurred in conjunction with the "natural" natural experiments school associated with Joshua Angrist, who pioneered the analysis of the role of exogenous variation in demographic (Angrist & Evans 1998), organizational (Angrist & Lavy 1999), or institutional (Angrist 1990) processes for educational attainment and labor market outcomes. A distinctive trait of the recent econometric literature is that interest is not merely with net effects of exogenous events $Z$ on outcomes—because $Z$ is (conditionally) exogenous, its net effect would result from including it as a covariate in a standard regression model—but in utilizing exogenous variation induced by $Z$ as an instrument to identify the causal effect of $D$ on outcomes. In other words, $Z$ is seen as an exogenous shock (or condition) whose impact on outcomes is fully transmitted through mechanism $D$, thereby enabling the analyst to recover the causal effect of the latter on $Y$ through instrumental variables estimation (see Section 5.4 below and the respective critical discussion in Rosenzweig & Wolpin 2000). A recent sociological application of this identification strategy is Kirk's (2009) study of the effect of residential mobility on recidivism rates using the occurrence of Hurricane Katrina as an instrument.

# 4. ESTIMATION OF TREATMENT EFFECTS UNDER UNCONFOUNDEDNESS

## 4.1. Parametric Regression Models

Regression modeling serves many purposes in empirical research in sociology, providing descriptive summaries of multivariate statistical associations, tools for prediction and classification, and methods to establish causal relationships from available data. And whereas correlation clearly is not causation, regression models do identify the causal effect of interest if treatment status and outcomes are unconfounded, i.e., when respective CIA assumptions may plausibly be defended through either properly implemented experimental protocols, successful identification of natural exogenous variation, or the availability of theoretically critical covariate information to balance expected outcomes across comparison groups.

Without going into details of any of the many variants of parametric regression models used in empirical applications (but see Fox 2008, Long 1997, Wooldridge 2006 for introductions), the general approach essentially amounts to estimating the two regression functions

$$\begin{aligned} &E(Y_{D=0} \mid X) = \alpha_0 + \beta_0'(\mathbf{x_i} - \bar{\mathbf{X}}) \\ &\text{and} \quad E(Y_{D=1} \mid X) = \alpha_1 + \beta_1'(\mathbf{x_i} - \bar{\mathbf{X}}) \end{aligned} \qquad 13.$$

for expected outcomes $Y$ in the treatment and the control group (or for the latter only if an estimate of $\Delta_{\text{ATT}}$ is sufficient; see the expositions in Imbens & Wooldridge 2009, Angrist & Pischke 2009 for the following). With covariate values $x_i$ expressed as deviations from the grand mean, the regression estimator of the average treatment effect is the difference in (in nonlinear models: appropriately transformed) intercepts, $\Delta_{ATE(R)} = \alpha_1 - \alpha_0$. Alternatively, $\Delta_{\text{REG}}$ is also given from a pooled regression model that includes the full interaction terms between treatment status $D$ and $X$ in the model specification. In the context of the linear model, the regression correction to the observed group

difference in mean outcomes is then

$$\Delta_{REG} = E(Y_{D=1}) - E(Y_{D=0}) - \left( \frac{N_{D=0}}{N_{D=0} + N_{D=1}} \cdot \beta_{D=1} + \frac{N_{D=1}}{N_{D=0} + N_{D=1}} \cdot \beta_{D=0} \right) \cdot (\bar{\mathbf{X}}_{D=1} - \bar{\mathbf{X}}_{D=0}), \quad 14.$$

so that, in perfect analogy to Equation 9 above, mean differences in outcomes get adjusted by the difference in average covariate values between groups multiplied by the regression coefficients for covariates on outcomes [the third line of Equation 9 is absent if the CIA assumption (Equation 12) holds]. In the general form of Equations 13 and 14, adjustment builds on the group size–weighted average regression coefficient vector, whereas the more typical sociological practice of estimating a main effects model

$$E(Y_{D=0} \mid X) = \alpha + \mathbf{X}\beta + \Delta D \quad 15.$$

without treatment x covariate interactions leads to a common [and in ordinary least squares (OLS): variance-weighted] coefficient vector $\beta$ that describes the dependency of conditional expected outcomes on $X$ for the pooled sample. Also note that in the context of many familiar nonlinear models used in categorical data analysis, the above focus on average treatment effects on outcomes implies that logit, probit, or similar parameter estimates need to be converted into marginal (probability) effects and averaged across the enumerated sample (see Train 2003). Assessing treatment effects in terms of the index function or related metrics, e.g., the predicted odds, implicitly redefines the quantity of interest.

## 4.2. Matching Estimators

Over the past decade, nonparametric methods for causal inference, matching methods prime among them, have become increasingly popular across the social sciences. Matching methods themselves are not new and have long been used for covariate balancing and on efficiency grounds in randomized experiments. Their practical usefulness for conditioning in observational studies was long considered questionable because the requirement to find suitable matches across potentially high-dimensional covariate vectors $X$ is bound to run into sparse data problems in typical data sets. However, this assessment has been revised in the light of Rosenbaum & Rubin's (1983b) result that matching on an adequate linear combination of covariates $X$, namely the predicted probability of treatment or the propensity score

$$\hat{p}(X_i) = \Pr(D_i = 1 \mid X_i), \quad 16.$$

is a valid substitute for matching on the full covariate vector $X$ itself. The propensity score thus reduces a high-dimensional to a one-dimensional matching problem and has become the main vehicle to implement matching estimators. Rosenbaum (2002), Rubin (2006), and Heckman et al. (1998) provide overviews of the method from statistical and econometric perspectives, whereas Smith (1997), Morgan & Harding (2006), Morgan & Winship (2007), and Gangl & DiPrete (2006) provide introductions aimed at sociology audiences. Morgan (2001), Brand (2006), Brand & Halaby (2006), Harding (2003), and Gangl (2006) are recent applications of propensity score matching in sociology.

Fundamentally, however, the identification strategy behind matching and regression is the same: Like regression methods, matching relies on the availability of a sufficiently rich set of covariates $X$ that serve to balance expected outcomes absent treatment across comparison groups and that justify the CIA assumption (Equation 12) required for causal inference. Matching estimators can effectively be understood as a nonparametric reweighing of the data, where weights correspond to some transformation of the estimated propensity score. Specifically, the ATT parameter typically addressed in matching applications can be

expressed as

$$\Delta_{ATT(M)} = \frac{1}{N_{D=1}} \sum_{i \in D=1 \cap S} \left[ Y_{1i} - \sum_{j \in D=0 \cap S} W_{i,j} Y_{0j} \right], \quad 17.$$

or the average difference in outcomes between any treated observation (with outcome $Y_{1i} = Y_i \mid D_i = 1$) and a weighted average of observationally [i.e., in terms of $\hat{p}(X)$] similar observations in the control group, computed across the common support $S$ of the distribution of the propensity score in the two groups (see Heckman et al. 1998). By analogy, an estimator for the ATU can be defined if the direction of matching is reversed, and the ATE estimate is then just the group size–weighted average of these two parameters.

Practical applications require the analyst to estimate the propensity score, typically using a logit, probit, or linear probability model for $\Pr(D = 1 \mid X)$, from which a balanced sample of comparable units from the treatment and control group is then constructed. Although estimation of treatment effects from the balanced sample requires only elementary statistical operations, the construction of matched samples is not fully standardized. Available matching algorithms include stratification, nearest-neighbor, caliper, kernel matching, and many combinations of these, which can be flexibly adapted to specific features of the data set when forming pairs of treatment and control units, respectively appropriate weights $W_{i,j}$ in Equation 17. Also, the computation of standard errors of the treatment effect estimates is not yet standardized; in practice, bootstrapping is the default, although Imbens (2004) provides analytical standard errors for a broad class of matching estimators.

These practical issues aside, the features that have drawn sociologists toward matching methods are their close alignment with the counterfactual approach to causal analysis, the absence of functional form assumptions in the outcome model, and, associated with that, the absence of extrapolation outside the range of common covariate support in estimating the treatment effect. Requiring the analyst to explicitly model treatment assignment inevitably enforces an assessment of identification, and thus separates concern for appropriate research design from estimating the outcome model of substantive interest (see Rubin 2001). Furthermore, matching estimators protect against bias due to misspecification of functional form in conventional regression modeling because the role of covariates $X$ in matching estimators is solely to balance expected outcomes absent treatment across comparison groups, whereas covariate adjustment in regression models also involves parameter estimates from the outcome model. Finally, nonparametric estimation also minimizes the impact of extrapolation across sparse areas of the multivariate covariate distribution: Since matching estimators form comparisons through pairing observationally close, if not identical, observations, matching cannot estimate a treatment effect in areas where covariate distributions do not overlap for the treatment and control group, and the resulting estimate only applies to the population defined over the common covariate support $S$.

### 4.3. Alternative Approaches and Extensions

In fact, matching and regression can easily be combined into "doubly robust" estimators that exploit the relative advantages of either approach. In that spirit, Rubin & Thomas (2000) and van der Laan & Robins (2003) have proposed estimating treatment effects from a regression of $Y$ on $D$ and $X$ in the matched sample generated from propensity score matching in order to safeguard the analysis against misspecification in the assignment or substantive model (but see more critically Freedman & Berk 2008), and Ho et al. (2007) and Imbens & Wooldridge (2009) similarly recommend matching to produce a balanced sample for subsequent regression analysis in

order to minimize extrapolation bias in the regression estimates.

At a more general level, matching and regression can even be seen as members of a larger class of inverse probability of treatment weighted (IPW; see Robins et al. 1992, Hirano et al. 2003, Wooldridge 2007) estimators of the form

$$\Delta_{IPW} = E\left[g(X_i)\left[\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1-D_i)}{1-p(X_i)}\right]\right], \quad 18.$$

where $p(X_i)$ is again the propensity score and $g(X_i)$ is a known weighting function. For $g(X_i) = 1$, Equation 18 is the matching estimator for the ATE, and $g(X_i) = p(X_i)/\Pr(D_i = 1)$ gives the matching estimator for the ATT from Equation 17. For OLS, weights $g(X_i)$ involve the variance of treatment status $D$, which explains why OLS coefficients will differ from matching parameters if treatment effects are heterogeneous: Matching averages covariate-specific treatment effects by population shares, OLS by the conditional variances of treatment status instead.

These differences aside, IPW estimators assume special significance in dynamic treatment assignment settings, i.e., in cases in which the timing of treatment exposure may be informative, and respective (observable) time-varying confounding needs to be corrected. Although a discussion of the IPW approach to these situations is beyond this review, Robins et al. (2000), Hernan et al. (2001), and van der Laan & Robins (2003) provide introductions to the estimation of so-called marginal structural models, which are built on IPW principles. Barber et al. (2004), Sampson et al. (2006), Sampson et al. (2008), and Hong & Raudenbush (2008) are recent applications of this modeling strategy. In addition, Fredriksson & Johansson (2008) and Crepon et al. (2009) discuss handling dynamic treatment assignment in the context of propensity score matching proper, and Gangl (2006) provides an example of that approach in sociology.

## 5. ESTIMATION OF TREATMENT EFFECTS IN THE PRESENCE OF UNMEASURED CONFOUNDERS

### 5.1. Sensitivity Analysis and Bounds on Treatment Effects

Causal inference becomes considerably more difficult when, as is common, the empirical data are not sufficiently rich to justify conditional independence assumptions that would permit the analyst to recover the causal parameter of interest from regression or matching. In this case, sensitivity analyses intend to understand the extent of remaining bias in the empirical estimates under plausible assumptions about the degree of unmeasured confounding in the data and a parsimonious parametric model that describes the dependency between the unobservables $D$ and $Y$. Originally developed for matching estimators by Rosenbaum & Rubin (1983a) and Rosenbaum (2002), sociological applications include Harding (2003) and DiPrete & Gangl (2004), and Frank (2000) and DiPrete & Gangl (2004) discuss related procedures for the case of linear regression and IV estimation. In each case, the interest is in identifying the level of unmeasured confounding up to which the causal effect of interest may still be considered robust, typically as judged by reaching statistical significance on an appropriate test statistic. In general, sensitivity analyses are most meaningful if the range of simulated values has empirical content, for example because the magnitude of likely confounding can usefully be bounded from external information or be communicated in terms of comparable effects of some known covariate (Imbens 2003, DiPrete & Gangl 2004).

Departing from the purely statistical character of sensitivity analysis, Manski's (1995, 2003, 2007) work on causal inference from partially identified outcome distributions offers an intellectually more rigorous approach to understand the inherent uncertainty of causal inference in the presence of unobserved confounders. Manski's key contribution is to use statistical consistency requirements as well

as behavioral assumptions rooted in economic theory to bound treatment effects of interest, i.e., to derive a range of effect estimates that is consistent with the data under maintained assumptions that are often less restrictive than those implied in conventional point estimators. Within sociology, Morgan's (2005) careful analysis of the role of student expectations for educational attainment is a prototypical example of the approach (but also see Manski et al. 1992 for a related application).

## 5.2. Control Function Models

In many respects, control function or treatment effects models take the exact opposite approach to address bias due to endogenous treatment assignment. Originally developed by Heckman (1978) as an extension to his better-known sample-selection model, control function models amount to estimating the process of treatment choice and outcomes according to Equations 10 and 11 simultaneously, which then permits the analyst to incorporate the correlation of error terms arising as a consequence of unobserved causes of $D$ and $Y$ or endogenous treatment choice. Assuming joint normality of the errors $(U, V)$ in the canonical case of a binary treatment, Heckman (1978) derived the conditional expected outcomes in the treatment and control group as

$$\begin{aligned} E(Y_1 \mid X, D = 1) &= X\beta_1 \\ &+ \frac{Cov(U_1, V)}{Var(V)} \cdot \lambda(X\gamma) \\ E(Y_0 \mid X, D = 0) &= X\beta_0 \\ &+ \frac{Cov(U_0, V)}{Var(V)} \cdot \tilde{\lambda}(X\gamma), \end{aligned} \qquad 19.$$

where bias due to endogenous selection into treatment status is captured by the second term on the right-hand side, the control function. With $\lambda = \phi(X\gamma)/\Phi(X\gamma), \tilde{\lambda} = -\phi(X\gamma)/\Phi(-X\gamma)$, $\phi(.)$ the normal density function, and $\Phi(.)$ the cumulative normal distribution, the treatment effect model can be estimated using Heckman's two-step estimator, albeit by including two correction terms into the second-stage OLS specification (see Blundell et al. 2005 for respective empirical estimates of returns to education).

As a structural model, the advantage of the treatment effect model is that alternative estimands of interest, whether the ATE, ATT, or distributional treatment effects, may be derived from the model in principle (Heckman et al. 2001). However, identification and estimation of the model rely on and may be sensitive to untestable functional form assumptions. More recent semiparametric approaches estimate the treatment effect model under a factor structure model for the latent unobserved confounders and additive separability of error terms (see Carneiro et al. 2003, Cunha et al. 2006) and thus considerably relax the assumption of joint normality in the original formulation. More reliably, identification may also be secured from the availability of an instrument $Z$ that predicts treatment status $D$, but not outcomes $Y$. Yet although reliance on instruments renders control function models and IV estimation structurally equivalent under certain conditions (Angrist 2001, Vytlacil 2002), the aim of instrumentation is quite different. In the treatment effect model, $Z$ uses variation in the truncation point of the outcome distribution to estimate the unobserved (mean of the) full outcome distribution, whereas, as discussed below, IV estimation aims to identify the average treatment effect among those respondents who were induced to change treatment status by $Z$. A further implication of this difference in perspective is that control function models would seek to employ as many plausible instruments $Z$ as possible simultaneously, whereas the LATE parameter resulting from IV estimation is most convincingly interpreted for a reasonably powerful single instrument $Z$ that corresponds to a specific manipulation of treatment status.

## 5.3. Fixed-Effects and Difference-in-Differences Estimators

Even absent a fully parametric specification of a joint model of treatment status and outcomes, confounding due to unmeasured factors may often be addressed when repeated observations

are available. Specifically, if the treatment effect is additive and the error terms additively separable in the outcome and the assignment model, the resulting panel regression model can be written as

$$Y_{it} = \alpha_i + \beta X_{it} + \gamma_i W_i + \Delta D_{it} + \lambda_t + \varepsilon_{it} \quad 20.$$

under the simplifying assumption that the parameter vector $\beta$ is the same in both the treatment and control group. In Equation 20, the subtle but all-important difference to a cross-sectional model is the inclusion of individual-specific intercepts $\alpha_i$ that capture the impact of any unobserved but temporally stable characteristic of unit $i$ on outcomes $Y_i$. This feature of the model is particularly relevant if Equation 20 is estimated in a way that permits the $\alpha_i$ parameters to be correlated with other observed covariates, which implicitly treats $\alpha_i$ as a joint cause of $D$ and $Y$ and thus addresses the potentially endogenous selection into treatment based on any temporally invariant unobservable.

This is precisely what is achieved by the fixed-effects (FE) or within estimator

$$Y_{it} - \bar{Y}_i = \beta(X_{it} - \bar{X}_i) + \Delta(D_{it} - \bar{D}_i) + \lambda_t - \bar{\lambda} + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad 21.$$

that uses within-transformed data (also known as demeaned or change score data) to estimate Equation 20 from variation in observed covariates and outcomes within observational units over time. Differencing the data as in Equation 21 eliminates the impact of any temporally stable characteristic of individual units, whether observed ($W_i$) or unobserved ($\alpha_i$), so that observed changes in outcomes ($Y_{it} - \bar{Y}_i$) only depend on changes in observed covariates $X_{it}$, changes in treatment status $D_{it}$, and time-varying idiosyncratic errors $\varepsilon_{it}$. The FE estimator, in other words, identifies the average treatment effect of $D$ on the treated if exogeneity of time-varying idiosyncratic errors $\varepsilon_{it}$, or

$$E(Y \mid \alpha_i, X_{it}, W_i, t, D_{it}) = E(Y \mid \alpha_i, X_{it}, W_i, t) \Leftrightarrow E(\varepsilon_{it} - \bar{\varepsilon}_i | D = 1) = E(\varepsilon_{it} - \bar{\varepsilon}_i | D = 0), \quad 22.$$

can be maintained (see Wooldridge 2002), which is a considerably weaker form of CIA than required for any of the methods discussed in Section 4. Moreover, Rosenbaum (1987) and Heckman & Hotz (1989) provide specification tests that use pretreatment observations or comparison groups with known causal effects to assess the validity of Equation 22.

It is hard to overstate the gain in identifying power provided by the beautifully simple method of FE estimation over standard cross-sectional estimators (e.g., Allison 1990, 1994; Winship & Morgan 1999; Halaby 2004). If anything, the appeal of FE methods has only been growing over the past decade as panel data have increasingly become available, and feasible FE estimators have been developed for many popular classes of regression models, including models for categorical, count, and event-history data (see Allison 2009, Baltagi 2008, Hsiao 2003, Wooldridge 2002). Sociological applications of respective FE estimators are also becoming more common (e.g., Budig & England 2001, DiPrete & McManus 2000, McManus & DiPrete 2001, Yakubovich 2006). In addition, the FE approach has also been extended to matching estimators by Heckman et al. (1997a, 1998), and respective difference-in-differences (DID) matching estimators have found extensive application in econometric evaluations of job-training programs (e.g., Dehejia & Wahba 2002, Smith & Todd 2005, Dehejia 2005), but also in recent studies of job histories in sociology (e.g., Brand 2006, Gangl 2006). Finally, Athey & Imbens (2006) have introduced a nonparametric extension of DID estimation, the so-called changes-in-changes (CIC) estimator, that relaxes the assumption of additive observation-specific error terms.

The versatility of FE estimation also extends well beyond traditional panel data. Best known in sociology is the use of sibling or twin models—i.e., within-family differenced estimation—to control for unobserved family fixed effects while capitalizing on educational differences between siblings or twins to estimate the causal effect of education on outcomes (e.g., Ashenfelter & Rouse 1998, Sieben

& de Graaf 2004), or utilizing temporal variation in family income to assess its impact for child development and educational attainment (Duncan et al. 1998, Waldfogel et al. 2002). Elwert & Christakis's (2008) study of widowhood effects is another ingenious application of a within-estimator that controls for unobserved personality traits by exploiting the difference in the bereavement effect of the death of respondents' former versus current spouse.

The FE principle also extends to repeated cross-sectional data more generally in cases in which an aggregate (group-level) event or intervention is concerned and pre- and postevent data are available, i.e., when the data structure follows a (multiple) interrupted time series design. The resulting FE estimators, more generally known as DID estimators in econometrics, can be effective tools to control for unobserved area, organization, or population segment effects, but have not seen much use in sociology so far. Excellent illustrations are Goldin & Rouse's (2000) study of the introduction of blind auditions in major U.S. symphony orchestras since the 1970s using orchestra and person fixed effects, or Rindfuss et al. (2007), who use municipality fixed effects to evaluate the impact of child care on fertility in Norway. Ruhm's (1998) analysis of the relationship between parental leave mandates and women's employment is particularly instructive because his comparison of employment trends across countries and between sexes results in a triple differenced (DDD) estimator that, in Moffitt's (2005) terminology, combines an area FE with a population segment FE approach.

However, although weaker than with cross-sectional estimators, the CIA assumption (Equation 22) that is required to identify causal effects from the FE estimator is still a strong one. In longitudinal settings, Equation 22 requires conditional independence of treatment of both past and future outcomes, thus ruling out endogenous selection into treatment based on agents' reasonably accurate predictions about treatment impact (Heckman & Robb 1985, 1986; see Wooldridge 2002, Halaby 2004 for a discussion of potential econometric solutions); in sibling, twin, or related studies, the equivalent CIA amounts to maintaining that sibling differences are as good as randomly assigned rather than are a consequence of sibling order or unmeasured differences in ability or motivation. Also, with treatment effect heterogeneity, FE estimates may have little external validity as they represent average treatment effects for the possibly quite selective population that empirically experiences the condition of interest, e.g., parents of twins or families newly entering poverty within some observation window, because only these contribute within-unit variation in $D$ to the FE estimator.

Finally, the FE estimator also rests on the assumption of common (parallel) time trends across groups or, equivalently, a unity constraint on the coefficient of the lagged or mean outcome variable (Allison 1990). An alternative to FE is to estimate a dynamic panel, known as the lagged dependent variable (LDV) or analysis of covariance model

$$Y_{it} = \alpha_i + \beta X_{it-j} + \gamma Y_{it-j} + \Delta D_{it} + \lambda_t + \varepsilon_{it}, \quad 23.$$

where past outcomes and current or past covariates enter the specification as covariates. Compared to FE, Equation 23 rests on a somewhat less restrictive identification assumption

$$E(Y \mid Y_{it-b}, X_{it-b}, t, D_{it}) = E(Y \mid Y_{it-b}, X_{it-b}, t, D_{it}), \quad 24.$$

which merely requires sequential exogeneity of treatment status given covariates and past outcomes (Wooldridge 2002). As they represent different assumptions about the substantive process under study, the LDV and FE estimators are not nested and are likely to lead to divergent estimates in practice (see Allison 1990). As Angrist & Pischke (2009) show, however, the two estimators do have a useful bracketing property in the sense that FE will overestimate a positive treatment effect if Equation 23 is the correct specification, whereas LDV will underestimate a positive treatment effect if the FE specification (Equation 20) is substantively correct. LDV

and FE can also be combined into a dynamic regression specification that allows for fixed unit effects, yet the gain in theoretical eclecticism comes at the price of considerable estimation difficulties since the within-transformed error terms are necessarily correlated with the lagged dependent variable, thus necessitating IV estimation for consistent parameter estimation (see Halaby 2004, Baltagi 2008 for further details).

### 5.4. Instrumental Variables Estimation

Finally, reliance on natural experiments may provide leverage to estimate a treatment effect in the presence of unobserved confounders. If an exogenous factor $Z$ is observed, $Z$ can naturally always be used as a covariate in a regression of $Y$ on $Z$ to determine its net effect on outcomes in a straightforward fashion. However, $Z$ is even more useful if $D$ can plausibly be considered the sole mechanism that transmits the impact of $Z$ on $Y$. In that case, $Z$ can be considered an instrument that can then be used to recover the treatment effect of $D$ even in the presence of unobserved confounders (i.e., under endogeneity of treatment status $D$) using the method of instrumental variables (IV).

More formally, an exogenous $Z$ (i.e., $(Y, D) \perp\!\!\!\perp Z$) can be considered a valid instrument for $D$ if two additional conditions are fulfilled. First, $Z$ needs to be relevant for treatment assignment, thus inducing empirical variation in treatment status that is as good as randomly assigned by virtue of exogeneity of $Z$ itself.[6] Second, potential outcomes need to be independent of $Z$ given $D$, i.e., $Z$ must not affect $Y$ once the mediating effect of $D$ is accounted for, which is the so-called exclusion restriction in IV estimation. Unlike instrument relevance for treatment status, the exclusion restriction

$$E(Y_i \mid X_i, D_i, Z = z) = E(Y_i \mid X_i, D_i, Z = z') \quad \text{for all} \quad z \neq z' \qquad 25.$$

is untestable in principle and needs to be justified theoretically in any empirical application (see Rosenzweig & Wolpin 2000). If the exclusion restriction is maintained, and ignoring covariates $X$ for simplicity of exposition, the IV estimator of the treatment effect is the Wald estimator

$$\Delta_{Wald} = \frac{\mathrm{Cov}(\mathrm{Y_i}, \mathrm{Z_i})}{\mathrm{Cov}(\mathrm{D_i}, \mathrm{Z_i})} = \frac{E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)}{E(D_i \mid Z_i = 1) - E(D_i \mid Z_i = 0)}, \qquad 26.$$

which forms the effect estimate as the ratio of the change in expected outcomes induced by $Z$ over the change in expected treatment status induced by $Z$. When covariates $X$ are present, the analogous IV estimator is given by two-stage least squares (2SLS) of $Y$ on $D$, $X$, and $Z$.

A fundamental weakness of IV estimation is that IV does not reliably identify the average treatment effect of $D$ unless the $\Delta$ is constant in the population (see Heckman 1997). However, Angrist et al. (1996; see also Imbens & Angrist 1994) note that IV is informative about treatment effects among the population whose treatment status is actually affected by the instrument $Z$, because $Z$ has either induced or prevented participation in $D$. Angrist et al. (1996) derive $\Delta_{\mathrm{IV}}$ as the weighted average of the average treatment effects in the two groups of compliers and defiers, and render the IV estimate meaningful through a monotonicity condition that assumes responsive units to either uniformly increase (no defiers) or decrease participation (no compliers), but not both. In the case of uniform compliance, the resulting local average treatment effect (LATE)

$$\Delta_{LATE} = \Delta_{IV} = \frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)} = E[Y_{1i} - Y_{0i}|D_i(Z = 1) > D_i(Z = 0)] \qquad 27.$$

gives the average treatment effect of $D$ on those units $i$ that complied with treatment $D$ because of the instrument $Z$.

However, although LATE reconciles IV with a framework of treatment effect heterogeneity, the utility of this parameter has also

[6] In traditional IV estimation, it is sufficient that $Z$ is (partially) correlated with treatment status. If $Z$ cannot be considered a cause of $D$, the LATE interpretation of IV estimates is compromised, however, and most current applications of IV search for strictly exogenous instruments $Z$ in consequence.

been questioned in principle. First, LATE is defined by the specific instrument $Z$ available to the researcher (Heckman 1997, Heckman & Vytlacil 2007b), i.e., multiple instruments $Z$ will be associated with multiple LATE parameters that describe the same causal association between $D$ and $Y$, yet without any guarantee of consistency across estimates because different instruments $Z$ will affect treatment status $D$ for different segments of the population (as one implication of this, traditional overidentification tests can be reinterpreted as tests of treatment effect heterogeneity; see Angrist & Pischke 2009). Second, although LATE is defined as the average treatment effect among compliers, the population of compliers cannot actually be identified in the sample data because LATE is based on changes in the expected exposure to treatment, not on actually observed changes of treatment status (Angrist et al. 1996).

As a partial remedy, Angrist & Imbens (1995) and Imbens & Rubin (1997) discuss methods to characterize the nature of treatment response to $Z$ and features of the complier population more succinctly. In that context, variation in LATE estimates based on alternative instruments $Z$ is indeed a desirable property of the analysis because heterogeneity in LATE parameters alerts the analyst to the fact that different valid instruments $Z$ induce heterogeneous consequences in the affected segments of the population. In general, LATE parameter estimates are most useful if $Z$ is capturing variation in the costs of, or the opportunities for participating in, $D$, making relevant policy changes an evident candidate instrument. In that line of reasoning, it is perfectly plausible that different policy changes $Z$ would affect different segments of the population and would create different impacts on outcomes and that respective heterogeneity of treatment effects would be of key substantive interest. Also, LATE is often of interest in experimental studies as it identifies the complier average causal effect (CACE), whereas the net effect of the manipulation $Z$ corresponds to the intention-to-treat (ITT) effect (e.g., Ludwig et al. 2008).

In practice, IV estimates may often be less than compelling, however, because reliance on weak instruments undermines the statistical consistency and efficiency of the estimator (see Bound et al. 1995). Nevertheless, Heckman & Vytlacil (2005, 2007b; see also Heckman et al. 2006) generalize the logic of IV estimation when deriving the marginal treatment effect (MTE) as the fundamental parameter of causal inference that describes the series of LATE parameters stemming from infinitesimally small changes in incentives or opportunities to participate in treatment. The actual estimation of MTEs requires the availability of respectively rich instruments $Z$ and results in the local instrumental variable (LIV) estimator of Heckman & Vytlacil (2005; see Heckman et al. 2006 for an empirical application). Also, Pearl's (2000) proposal of identification via the front-door criterion, i.e., via an isolated mechanism $W$ that transmits the impact of $D$ on $Y$, directly relates to IV estimation, yet identifies treatment effects from a reversal of the traditional roles of instrument and mechanism. Winship & Harding (2008) provide a first sociological application to achieve identification in age-period-cohort models.

## 6. IMPLICATIONS FOR SOCIOLOGICAL RESEARCH PRACTICE

### 6.1. Identification Trumps Estimation

Although the benefits of randomized experiments are widely understood, the character of sociology as a population science inevitably makes observational data and regression or similar statistical methods the natural workhorses of empirical research. However, perhaps the most important lesson of the recent literature has been the sobering assessment of the very possibility of drawing causal inferences from nonexperimental data, let alone as a natural by-product of regression analysis as currently practiced in much of sociology. The identification problem at the heart of any causal inference clearly requires either an estimable

structural model of the process of interest or proper research design that enables the analyst to base causal inferences on exogenous variation in treatment status. And given the dearth of estimable structural models (but see Logan 1996, 1998; Logan et al. 2008 for exceptions) and our notoriously vaguely specified theories, an emphasis on the primacy of research design seems the more natural starting point for many, if not most, attempts to recover causal parameters in sociology.[7] In that respect, the greater reliance on (field) experiments as in Pager's (2003), Bertrand & Mullainathan's (2004), or Correll et al.'s (2007) studies of discrimination and the greater ingenuity at identifying natural experiments that help address long-standing theoretical concerns in the sociology of education (see Downey et al. 2004) or the sociology of crime (e.g., Kirk 2009) are encouraging signs that empirical research is becoming more conscious about precisely identifying theoretically relevant manipulations and the conditions under which causal inferences are justified.

That said, a greater reliance on (quasi-) experimental designs clearly complements rather than substitutes for parallel efforts to improve on causal analysis in nonexperimental settings. (Quasi-) experimental methods have their own set of problems, from issues of implementation and concomitant change to the difficulties of manipulating social environments, of finding historical events with theoretically informative implications, or the inability to detect equilibrium effects at the macro level (Cook & Campbell 1979, Garfinkel et al. 1992, Heckman 1992, Moffitt 2005). And even more simply, the evident trade-off between internal and external validity, i.e., the fact that (quasi-) experimental methods inevitably identify only selected causal parameters for specific populations and specific interventions, limits their utility as the exclusive tool of causal inference in the social sciences. Eventually, (quasi-) experimental designs will unfold their true potential only to the extent that findings are calibrated against nonexperimental data from representative surveys and integrated into comprehensive theoretical models of the process under study (Moffitt 2005). If so, our ability to draw causal inferences from observational data might improve as a by-product because any theoretical understanding gained will aid identification in nonexperimental settings by providing a model of joint causes of treatment and outcomes.

### 6.2. The Causal Program of Sociology: Class, Race, and Gender, or Estimating Treatment Effects Versus Causal Accounting

The perception that the counterfactual framework would primarily apply to the effects of policy interventions or other explicitly manipulated (or at least manipulable) treatments is perhaps the single most important impediment to its more widespread adoption in sociology. This perception is a major misunderstanding on the part of sociologists (cf. also Heckman 2005, Moffitt 2005, Sobel 1998). Whether nonmanipulable factors such as gender, race, or class affect life courses is a perfectly sensible counterfactual question to begin with, although not necessarily one that would excite most interest in the respective subfields of the discipline. With respect to gender, for example, the counterfactual "manipulation" in question is the determination of fetal gender at inception, which, moreover, is plausibly random (Rubin 1986), so that its causal effect is directly identified from the comparison of mean life-course outcomes among men and women from, for example, the same birth cohort or country. In this specific case, and ignoring SUTVA (but see Section 6.3), the main impediment to causal inference is not so much a lack of controls,

[7] Even Heckman (2005), a most ferocious advocate of the structural approach in economics, concedes this to be a defensible epistemological strategy if, as arguably applies to sociology, the primary objective of causal analysis is to recover causal parameters as they apply to actually occurring events, i.e., explanation rather than prediction and extrapolation to novel situations. In fact, this debate is not necessarily between disciplines, but runs within econometrics itself where the benefits and costs of design-based strategies have been widely discussed in recent years (see Angrist & Krueger 1999, Heckman et al. 1999, Heckman & Vytlacil 2007a,b, and Section 6.3 below).

as a lack of representative samples (see Sobel 1998).

More generally, many of sociology's core analytical categories—including class, gender, race, ethnicity, but also age, period, and cohort or adverse events such as parental divorce, illness, or job loss, the impact of which sociologists have more recently examined—identify social constraints that people are exposed to and that, in turn, shape life courses, perceptions, and values. As such, these categories describe social causes that are plausibly "assigned" exogenously of outcomes, though, with the possible exception of gender, not necessarily randomly so. Still, the estimation of causal effects is greatly facilitated in each case because accounting for population heterogeneity with respect to other exogenous constraints is sufficient to identify the causal effect of interest.[8] In other cases, however, sociological interest may be with the implications of social constraints, yet exogeneity of observed measures is only partial at best. For example, neighborhood or network effects are typically understood as describing inequality of opportunities and hence restrictions on individual action, yet both residential neighborhood and support networks are amenable to (constrained) individual choice. The resulting endogeneity needs to be taken into account when estimating respective causal effects, although in some cases mere precision about the manipulation of interest, e.g., between neighborhood effects in a developmental or in a current residential sense, may clarify the issue. With yet other potential causes of social action such as educational attainment, job change, or family formation, voluntary action is a significant element of event occurrence almost by definition, and all the concerns discussed at length in this review forcefully apply.

It is also true that the aims of causal analysis as practiced in sociology often go beyond the counterfactual goal of estimating a treatment effect convincingly and include interest in the generative mechanisms that may produce any observed "black-box" treatment effect. The important point is that, although typically not explicitly raised as an issue in standard statistical or econometric treatments, this concern is well aligned with the counterfactual perspective and is well founded in substantive terms in any case. Whereas class, race, and gender are exogenous causes of social behavior, they at the same time constitute fuzzy treatments that involve opportunities and constraints along multiple dimensions, which may each be more or less relevant for producing an effect, may reinforce each other in specific ways, or may vary in relevance over time and location. Adopting a mode of causal accounting that seeks to understand the actual causal manipulations, i.e., social mechanisms involved in bringing about a certain effect, seems a perfectly natural way to proceed in these cases.

The recent exchange on the Moving to Opportunity (MTO) experiment between Clampet-Lundquist & Massey (2008), Ludwig et al. (2008), and Sampson (2008) is an instructive example of this difference in emphasis between sociology and other disciplines: Whereas Ludwig et al. (2008) see the experimental moving vouchers program as an opportunity to exploit exogenously induced variation in residential environments to estimate neighborhood effects on various outcomes, Clampet-Lundquist & Massey (2008) rightly emphasize that MTO is far from addressing neighborhood effects proper (i.e., in the developmental sense of primary interest to sociologists), and they embark on an analysis that seeks to understand how the experimental effects observed in MTO are related to participants' exposure to specific neighborhood characteristics. In light of the above argument,

[8]This statement still leaves room for considerable disagreement about which exogenous covariates specifically to control for in order to identify an effect of interest. For example, while family fixed effects would capture the causal impact of growing up with one particular family instead of the average family, identifying the causal effect of social class would require controlling for other family characteristics, e.g., genetic, personality, or attitudinal traits of parents, that may determine both parental class and children's life-course outcomes (e.g., Freese 2008). In other words, the required controls depend on the specific manipulation of interest, and transparency in this respect is another major advantage implied in the potential outcomes framework.

both studies are pursuing complementary goals, with Ludwig et al. (2008) providing evidence on the fundamental causal parameters of interest identified by the experimental design, and Clampet-Lundquist & Massey (2008) undertaking a form of mediation analysis that intends to decompose the observed experimental effect. Although much of the current practice of mediation analysis (see MacKinnon 2008) is itself in dire need of being realigned with the potential outcomes framework, the more general point is that in many sociological applications, and especially if exogenous factors such as race, class, and gender are studied, all the concerns of causal inference will typically apply at the level of generative mechanisms that constitute the actual causal manipulations behind socially relevant attributes and conditions.

### 6.3. Causality as Social Process: SUTVA Redux

While the counterfactual model seems perfectly compatible with many aspects of sociologists' understanding and practice of causal analysis, its reliance on the stable unit treatment value assumption is much more restrictive and problematic than is commonly recognized in the discipline. In essence, SUTVA assumes that potential outcomes for any unit *i* are unaffected by treatment assignment of both unit *i* and any other member *j* of the population under study, thus conveniently guaranteeing the existence of unit effects while ruling out agent responses to treatment (non)assignment like Hawthorne or John Henry effects, but more problematically, any form of social interactions between units. If, through processes of information diffusion, norm formation, leadership, endogenous reinforcement, or competition in tournaments, social interactions are important for outcomes, unit effects are inherently undefined because, in this case, outcomes for any particular unit *i* depend on the number or distribution of treated units *j* in the population, and the standard interpretation of any parameter estimate no longer applies (see Sobel 2006 for a more extensive discussion).

Evidently, any denial of social interactions as an emergent source of social behavior goes against the very essence of sociology and the social sciences more broadly. An obvious response to potential violations of SUTVA is to move the analysis to a more aggregate level, i.e., a classroom, family, organization, or local labor market, at which SUTVA can more plausibly be maintained and estimate macro treatment effects at that level (e.g., Moffitt 2005, Morgan & Winship 2007, Smith 2003). Empirically, this corresponds to sociologists' interests in group-level rather than individual-level interventions (see Axinn & Barber 2001; Smith 2005; Hong & Raudenbush 2006, 2008; Sampson 2008), and on a larger scale this is also consistent with Coleman's (1990) theoretical position of sociology as the study of social systems. However, the more inconvenient implication is that actual generative mechanisms will be left unspecified in the analysis, and, to the extent that they are built on social interactions, cannot readily be addressed in the standard statistical framework. Most likely, progress will require either more explicitly formalized estimable interaction-based structural models as have recently been developed in economics (Durlauf 2001, Brock & Durlauf 2001), or careful model specification and identification analysis that permits the definition of appropriate group-level instrumental variables to be designed (Manski 1993, Durlauf 2002). Neither line of research has yet begun to inform work in sociology, except for Jones's (1990) little-noted reanalysis of the original Hawthorne data.

In more practical terms, a requirement to move the analysis to the aggregate level would also mean that inequality and differentiation are to some extent off limits to empirical sociology. Acknowledging this, the second response to likely violations of SUTVA is to weaken the regularity assumptions associated with any estimated treatment effect in the presence of social interactions (e.g., Sobel 2006, Morgan &

Winship 2007). That is, standard treatment effect estimates are then to be considered as local, i.e., historically and situationally contingent rather than structurally invariant in a broader sense. In other words, the estimated local treatment effect is seen as providing accurate predictions only within "small" perturbations of, or incremental change to, the current equilibrium in the social system under study. With that, a full disciplinary research program needs to incorporate replication over time and across locations to obtain a sense of (and ultimately, a theoretical model for) contextual variation in treatment effects and its structural and institutional determinants. Basically, this seems in accordance with the implicit research program that sociologists pursue, especially when emphasizing contextual, multilevel, or cross-national comparative research. Ironically, though from a different angle, this conclusion also resonates Heckman's (2005) critique of purely design-based conceptions of causal analysis by restating that causal inference eventually aims at an integrative substantive model that specifies empirically relevant conditions of treatment choice, treatment variability, and resulting treatment effects.

## 7. CONCLUSIONS

Taken at face value, the crystallization of the potential outcomes model of causality over the past 25 years has brought many a sobering assessment of possibility of causal inference from observational data in general, and the utility of sociologists' typical use of regression methods to that end in particular. On the other hand, the counterfactual perspective has provided a unified conceptual framework for causal inference across the social sciences. There has been a revival of interest in proper research design that generates or identifies exogenous variation in events or conditions of interest, and the conditions under which different statistical estimators yield valid estimates of causal effects are increasingly well understood. In particular, the availability of longitudinal data and a search for informative natural experiments should greatly aid the identification of causal effects even in the presence of unobservable confounders that otherwise plague causal inference in the social sciences. Overall, while causal inference will inevitably remain a formidable challenge, the prospects for causal analysis in sociology may thus be better than its current reputation would suggest.



## DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

I am grateful to Richard Breen, Tom DiPrete, Glenn Firebaugh, Chuck Halaby, Stephen Morgan, Herb Smith, and Chris Winship for generous and extensive comments on earlier drafts. All remaining errors, omissions, or choice of emphasis are of course my own responsibility. John Logan is to be credited with the comment that saw me through the revision.

## LITERATURE CITED

Abbring JH, Heckman JJ. 2007. Econometric evaluation of social programs, Part III: distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. See Heckman & Leamer 2007, pp. 5145–303

Allison PD. 1990. Change scores as dependent variables in regression analysis. *Sociol. Methodol.* 20:93–114

Allison PD. 1994. Using panel data to estimate the effects of events. *Sociol. Methods Res.* 23:174–99

Allison PD. 2009. *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage
Angrist JD. 1990. Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records. *Am. Econ. Rev.* 80:313–35
Angrist JD. 2001. Estimations of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *J. Bus. Econ. Stat.* 19:2–16
Angrist JD, Evans WN. 1998. Children and their parents' labor supply: evidence from exogenous variation in family size. *Am. Econ. Rev.* 88:450–77
Angrist JD, Imbens GW. 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *J. Am. Stat. Assoc.* 90:431–42
Angrist JD, Imbens GW, Rubin DB. 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91:444–55
Angrist JD, Krueger AB. 1999. Empirical strategies in labor economics. In *Handbook of Labor Economics*, ed. O Ashenfelter, D Card, 3:1277–366. Amsterdam: Elsevier
Angrist JD, Krueger AB. 2001. Instrumental variables and the search for identification: from supply and demand to natural experiments. *J. Econ. Perspect.* 15:69–85
Angrist JD, Lavy V. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Q. J. Econ.* 114:533–75
Angrist JD, Pischke J-S. 2009. *Mostly Harmless Econometrics. An Empiricist's Companion.* Princeton, NJ: Princeton Univ. Press
Ashenfelter O, Rouse C. 1998. Income, schooling, and ability: evidence from a new sample of identical twins. *Q. J. Econ.* 113:253–84
Athey S, Imbens GW. 2006. Identification and inference in nonlinear difference-in-difference models. *Econometrica* 74:431–97
Axinn WG, Barber JS. 2001. Mass education and fertility transition. *Am. Sociol. Rev.* 66:481–505
Baltagi BH. 2008. *Econometric Analysis of Panel Data*. Chicester: Wiley. 4th ed.
Barber JS, Murphy SA, Verbitsky N. 2004. Adjusting for time-varying confounding in survival analysis. *Sociol. Methodol.* 34:163–92
Berk RA, de Leeuw J. 1999. An evaluation of California's inmate classification system using a generalized regression discontinuity design. *J. Am. Stat. Assoc.* 94:1045–52
Berk RA, Rauma D. 1983. Capitalizing on nonrandom assignment to treatments: a regression-discontinuity evaluation of a crime-control program. *J. Am. Stat. Assoc.* 78:21–27
Bertrand M, Mullainathan S. 2004. Are Greg and Emily more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am. Econ. Rev.* 94:991–1013
Bitler MP, Gelbach JB, Hoynes HW. 2006. What mean impacts miss: distributional effects of welfare reform experiments. *Am. Econ. Rev.* 96:988–1012
Blundell R, Dearden L, Sianesi B. 2005. Evaluating the effects of education on earnings: models, methods and results from the national child development survey. *J. R. Stat. Soc. A* 168:473–512
Blundell R, Dias MC. 2009. Alternative approaches to evaluation in empirical microeconomics. *J. Hum. Resour.* 44:565–640
Bound J, Jaeger DA, Baker RM. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J. Am. Stat. Assoc.* 90:443–50
Brand JE. 2006. The effects of job displacement on job quality: findings from the Wisconsin longitudinal study. *Res. Soc. Stratif. Mobil.* 24:275–98
Brand JE, Halaby CN. 2006. Regression and matching estimates of the effects of elite college attendance on educational and career achievement. *Soc. Sci. Res.* 35:749–70
Brand JE, Xie Y. 2007. Identification and estimation of causal effects with time-varying treatments and time-varying outcomes. *Sociol. Methodol.* 37:393–434
Brock WA, Durlauf SN. 2001. Discrete choice with social interactions. *Rev. Econ. Stud.* 68:235–60
Budig MJ, England P. 2001. The wage penalty for motherhood. *Am. Sociol. Rev.* 66:204–25
Carneiro P, Hansen K, Heckman JJ. 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *Int. Econ. Rev.* 44:361–422

Chernozhukov V, Hansen C. 2006. Instrumental quantile regression inference for structural and treatment effect models. *J. Econom.* 132:491–525
Clampet-Lundquist S, Massey DS. 2008. Neighborhood effects on economic self-sufficiency: a reconsideration of the Moving to Opportunity experiment. *Am. J. Sociol.* 114:107–43
Coleman JS. 1990. *Foundations of Social Theory*. Cambridge, MA: Harvard Univ. Press
Cook TD, Campbell DT. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally
Correll SJ, Benard S, Paik I. 2007. Getting a job: Is there a motherhood penalty? *Am. J. Sociol.* 112:1297–338
Crepon B, Ferracci M, Jolivet G, van den Berg GJ. 2009. Active labor market policy effects in a dynamic setting. *J. Eur. Econ. Assoc.* 7:595–605
Cunha F, Heckman JJ, Navarro S. 2006. Counterfactual analysis of inequality and social mobility. In *Mobility and Inequality: Frontiers of Research in Sociology and Economics*, ed. SL Morgan, DB Grusky, GS Fields, pp. 290–348. Stanford, CA: Stanford Univ. Press
Dehejia R. 2005. Practical propensity score matching: a reply to Smith and Todd. *J. Econom.* 125:355–64
Dehejia RH, Wahba S. 2002. Propensity score-matching methods for nonexperimental causal studies. *Rev. Econ. Stat.* 84:151–61
DiPrete TA, Gangl M. 2004. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociol. Methodol.* 34:271–310
DiPrete TA, McManus PA. 2000. Family chance, employment transitions, and the welfare state: household income dynamics in the United States and Germany. *Am. Sociol. Rev.* 65:343–70
Downey D, Broh B, von Hippel P. 2004. Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *Am. Sociol. Rev.* 69:613–35
Duncan GJ, Yeung WJ, Brooks-Gunn J, Smith JR. 1998. How much does childhood poverty affect the life chances of children? *Am. Sociol. Rev.* 63:406–23
Durlauf SN. 2001. A framework for the study of individual behavior and social interactions. *Sociol. Methodol.* 31:47–87
Durlauf SN. 2002. On the empirics of social capital. *Econ. J.* 112:F459–79
Elwert F, Christakis NA. 2008. Wives and ex-wives: a new test for homogamy bias in the widowhood effect. *Demography* 45:851–73
Firebaugh G. 2008. *Seven Rules for Social Research*. Princeton, NJ: Princeton Univ. Press
Fox J. 2008. *Applied Regression Analysis and Generalized Linear Models*. Los Angeles: Sage. 2nd ed.
Frank KA. 2000. Impact of a confounding variable on a regression coefficient. *Sociol. Methods Res.* 29:147–94
Fredriksson P, Johansson P. 2008. Dynamic treatment assignment: the consequences for evaluations using observational data. *J. Bus. Econ. Stat.* 26:435–45
Freedman DA, Berk RA. 2008. Weighting regressions by propensity scores. *Eval. Rev.* 32:392–409
Freese J. 2008. Genetics and social structure. *Am. J. Sociol.* 114:S1–35
Gangl M. 2004. Welfare states and the scar effects of unemployment: a comparative analysis of the United States and West Germany. *Am. J. Sociol.* 109:1319–64
Gangl M. 2006. Scar effects of unemployment: an assessment of institutional complementarities. *Am. Sociol. Rev.* 71:986–1013
Gangl M, DiPrete TA. 2006. Kausalanalyse durch Matchingverfahren [Matching methods for the causal analysis of observational data]. In *Methoden der empirischen Sozialforschung. Sonderh. 44 Kölner Z. Soziol. Sozialpsychol.*, ed. A Diekmann, pp. 396–420. Wiesbaden: VS Verlag
Garfinkel I, Manski CF, Michalopoulos C. 1992. Micro experiments and macro effects. In *Evaluating Welfare and Training Programs*, ed. CF Manski, I Garfinkel, pp. 253–73. Cambridge, MA: Harvard Univ. Press
Gill RD, Robins JM. 2001. Causal inference for complex longitudinal data: the continuous case. *Ann. Stat.* 29:1785–811
Goldin C, Rouse C. 2000. Orchestrating impartiality: the impact of "blind" auditions on female musicians. *Am. Econ. Rev.* 90:715–41
Halaby CN. 2004. Panel models in sociological research: theory into practice. *Annu. Rev. Sociol.* 30:507–44
Harding DJ. 2003. Counterfactual models of neighborhood effects: the effect of neighborhood poverty on dropping out and teenage pregnancy. *Am. J. Sociol.* 109:676–719

Heckman JJ. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46:931–59
Heckman JJ. 1992. Randomization and social policy evaluation. In *Evaluating Welfare and Training Programs*, ed. CF Manski, I Garfinkel, pp. 201–30. Cambridge, MA: Harvard Univ. Press
Heckman JJ. 1997. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. *J. Hum. Resour.* 32:441–62
Heckman JJ. 2000. Causal parameters and policy analysis in economics: a twentieth century retrospective. *Q. J. Econ.* 115:45–97
Heckman JJ. 2001. Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *J. Polit. Econ.* 109:673–748
Heckman JJ. 2005. The scientific model of causality. *Sociol. Methodol.* 35:1–97
Heckman JJ, Hotz VJ. 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *J. Am. Stat. Assoc.* 84:862–74
Heckman JJ, Ichimura H, Todd PE. 1997a. Matching as an econometric evaluation estimator: evidence from evaluating a job training program. *Rev. Econ. Stud.* 64:605–54
Heckman JJ, Ichimura H, Todd PE. 1998. Matching as an econometric evaluation estimator. *Rev. Econ. Stud.* 65:261–94
Heckman JJ, LaLonde RJ, Smith JA. 1999. The economics and econometrics of active labor market programs. In *Handbook of Labor Economics*, ed. O Ashenfelter, D Card, 3:1865–2097. Amsterdam: Elsevier
Heckman JJ, Leamer EE. 2007. *Handbook of Econometrics Volume 6B*. Amsterdam: Elsevier
Heckman JJ, Robb R. 1985. Alternative methods for evaluating the impact of interventions. In *Longitudinal Analysis of Labor Market Data*, ed. JJ Heckman, B Singer, pp. 156–245. Cambridge, UK: Cambridge Univ. Press
Heckman JJ, Robb R. 1986. Alternative methods for solving the problem of selection bias in evaluating the impact of treatment on outcomes. In *Drawing Inferences from Self-Selected Samples*, ed. H Wainer, pp. 63–113. New York: Springer
Heckman JJ, Smith J, Clements N. 1997b. Making the most out of program evaluations and social experiments: accounting for heterogeneity in program impacts. *Rev. Econ. Stud.* 64:487–535
Heckman JJ, Tobias J, Vytlacil E. 2001. Four parameters of interest in the evaluation of social programs. *South. Econ. J.* 68:210–23
Heckman JJ, Urzua S, Vytlacil E. 2006. Understanding instrumental variables in models with essential heterogeneity. *Rev. Econom. Stat.* 88:389–432
Heckman JJ, Vytlacil E. 2005. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73:669–738
Heckman JJ, Vytlacil EJ. 2007a. Econometric evaluation of social programs, Part I: causal models, structural models and econometric policy evaluation. See Heckman & Leamer 2007, pp. 4779–874
Heckman JJ, Vytlacil EJ. 2007b. Econometric evaluation of social programs, Part II: using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. See Heckman & Leamer 2007, pp. 4875–5143
Hernán Má, Brumback B, Robins JM. 2001. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J. Am. Stat. Assoc.* 96:440–48
Hirano K, Imbens GW, Ridder G. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71:1161–89
Ho DE, Imai K, King G. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit. Anal.* 15:199–236
Holland PW. 1986. Statistics and causal inference. *J. Am. Stat. Assoc.* 81:945–60
Hong G, Raudenbush SW. 2006. Evaluating kindergarten retention policy: a case study of causal inference for multilevel observational data. *J. Am. Stat. Assoc.* 101:901–10
Hong G, Raudenbush SW. 2008. Causal inference for time-varying instructional treatments. *J. Educ. Behav. Stat.* 33:333–62
Hsiao C. 2003. *Analysis of Panel Data*. Cambridge, UK: Cambridge Univ. Press. 2nd ed.
Imai K, King G, Stuart EA. 2008. Misunderstandings between experimentalists and observationalists about causal inference. *J. R. Stat. Soc. A* 171:481–502

Imbens GW. 2003. Sensitivity to exogeneity assumptions in program evaluation. *Am. Econ. Rev.* 93:126–32
Imbens GW. 2004. Nonparametric estimation of average treatment effects under exogeneity: a review. *Rev. Econom. Stat.* 86:4–29
Imbens GW, Angrist JD. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62:467–75
Imbens GW, Lemieux T. 2008. Regression discontinuity designs: a guide to practice. *J. Econom.* 142:615–35
Imbens GW, Rubin DB. 1997. Estimating outcome distributions for compliers in instrumental variables models. *Rev. Econ. Stud.* 64:555–74
Imbens GW, Rubin DB. 2010. *Causal Inference in Statistics and the Social Sciences*. Cambridge, UK: Cambridge Univ. Press. In press
Imbens GW, Wooldridge JM. 2009. Recent developments in the econometrics of program evaluation. *J. Econ. Lit.* 47:5–86
Jones SRG. 1990. Worker interdependence and output: the Hawthorne studies reevaluated. *Am. Sociol. Rev.* 55:176–90
King G, Keohane RO, Verba S. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton Univ. Press
Kirk DS. 2009. A natural experiment on residential change and recidivism: lessons from Hurricane Katrina. *Am. Sociol. Rev.* 74:484–505
Koenker R. 2005. *Quantile Regression*. Cambridge, UK: Cambridge Univ. Press
Logan JA. 1996. Opportunity and choice in socially structured labor markets. *Am. J. Sociol.* 102:114–60
Logan JA. 1998. Estimating two-sided logit models. *Sociol. Methodol.* 28:139–73
Logan JA, Hoff PD, Newton MA. 2008. Two-sided estimation of mate preferences for similarities in age, education, and religion. *J. Am. Stat. Assoc.* 103:559–69
Long JS. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage
Ludwig J, Liebman JB, Kling JR, Duncan GJ, Katz LF, et al. 2008. What can we learn about neighborhood effects from the Moving to Opportunity experiment? *Am. J. Sociol.* 114:144–88
MacKinnon DP. 2008. *Introduction to Statistical Mediation Analysis*. New York: Erlbaum
Manski C. 1993. Identification of endogenous social effects: the reflection problem. *Rev. Econ. Stud.* 60:531–42
Manski CF. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard Univ. Press
Manski CF. 2003. *Partial Identification of Probabilities Distributions*. New York: Springer
Manski CF. 2007. *Identification for Prediction and Decision*. Cambridge, MA: Harvard Univ. Press
Manski CF, Sandefur GD, McLanahan S, Powers D. 1992. Alternative estimates of the effect of family structure during adolescence on high school graduation. *J. Am. Stat. Assoc.* 87:25–37
McManus PA, DiPrete TA. 2001. Losers and winners: the financial consequences of separation and divorce for men. *Am. Sociol. Rev.* 66:246–68
Moffitt RL. 2005. Remarks on the analysis of causal relationships in population research. *Demography* 42:91–108
Morgan SL. 2001. Counterfactuals, causal effect heterogeneity, and the Catholic school effect on learning. *Sociol. Educ.* 74:341–74
Morgan SL. 2005. *On the Edge of Commitment: Educational Attainment and Race in the United States*. Stanford, CA: Stanford Univ. Press
Morgan SL, Harding DJ. 2006. Matching estimators of causal effects. Prospects and pitfalls in theory and practice. *Sociol. Methods Res.* 35:3–60
Morgan SL, Todd JJ. 2008. A diagnostic routine for the detection of consequential heterogeneity of causal effects. *Sociol. Methodol.* 38:231–81
Morgan SL, Winship C. 2007. *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. Cambridge, UK: Cambridge Univ. Press
Ní Bhrolcháin M. 2001. 'Divorce effects' and causality in the social sciences. *Eur. Sociol. Rev.* 17:33–57
Pager D. 2003. The mark of a criminal record. *Am. J. Sociol.* 108:937–75
Pearl J. 2000. *Causality. Models, Reasoning and Inference*. Cambridge, UK: Cambridge Univ. Press
Rindfuss RR, Guilkey D, Morgan SP, Kravdal Ø, Guzzo KB. 2007. Child care availability and first-birth timing in Norway. *Demography* 44:345–72

Robins JM, Hernán Má, Brumback B. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11:550–60
Robins JM, Mark SD, Newey WK. 1992. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* 48:479–95
Rosenbaum PR. 1987. The role of a second control group in an observational study. *Stat. Sci.* 2:292–306
Rosenbaum PR. 2002. *Observational Studies*. New York: Springer. 2nd ed.
Rosenbaum PR, Rubin DB. 1983a. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J. R. Stat. Soc. B* 45:212–18
Rosenbaum PR, Rubin DB. 1983b. The central role of the propensity score in observational studies. *Biometrika* 70:41–55
Rosenzweig MR, Wolpin KI. 2000. Natural 'natural experiments' in economics. *J. Econ. Lit.* 38:827–74
Rossi PH, Lipsey MW, Freeman HE. 2004. *Evaluation: A Systematic Approach*. Thousand Oaks, CA: Sage. 7th ed.
Rubin DB. 1978. Bayesian inference for causal effects: the role of randomization. *Ann. Stat.* 6:34–58
Rubin DB. 1986. Statistics and causal inference: comment: Which ifs have causal answers? *J. Am. Stat. Assoc.* 81:961–2
Rubin DB. 2001. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Serv. Outcomes Res. Methodol.* 2:169–88
Rubin DB. 2005. Causal inference using potential outcomes: design, modeling, decisions. *J. Am. Stat. Assoc.* 100:322–31
Rubin DB. 2006. *Matched Sampling for Causal Effects*. Cambridge, UK: Cambridge Univ. Press
Rubin DB, Thomas N. 2000. Combining propensity score matching with additional adjustments for prognostic covariates. *J. Am. Stat. Assoc.* 95:573–85
Ruhm CJ. 1998. The economic consequences of parental leave mandates: lessons from Europe. *Q. J. Econ.* 113:285–317
Sampson RJ. 2008. Moving to inequality: neighborhood effects and experiments meet social structure. *Am. J. Sociol.* 114:189–231
Sampson RJ, Laub JH, Wimer C. 2006. Does marriage reduce crime? A counterfactual approach to within-individual causal effects. *Criminology* 44:465–508
Sampson RJ, Sharkey P, Raudenbush SW. 2008. Durable effects of concentrated disadvantage on verbal ability among African-American children. *Proc. Natl. Acad. Sci. USA* 105:845–52
Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin
Sieben I, de Graaf PM. 2004. Schooling or social origin? The bias in the effect of educational attainment on social orientations. *Eur. Sociol. Rev.* 20:107–22
Smith HL. 1990. Specification problems in experimental and nonexperimental social research. *Sociol. Methodol.* 20:59–91
Smith HL. 1997. Matching with multiple controls to estimate treatment effects in observational studies. *Sociol. Methodol.* 27:325–53
Smith HL. 2003. Some thoughts on causation as it relates to demography and population studies. *Popul. Dev. Rev.* 29:459–69
Smith HL. 2005. Introducing new contraceptives in rural China: a field experiment. *Ann. Am. Acad. Polit. Soc. Sci.* 599:246–71
Smith JA, Todd PE. 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? *J. Econom.* 125:305–53
Sobel ME. 1995. Causal inference in the social and behavioral sciences. In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, ed. G Arminger, CC Clogg, ME Sobel, pp. 1–38. New York: Plenum
Sobel ME. 1998. Causal inference in statistical models of the process of socioeconomic achievement. A case study. *Sociol. Methods Res.* 27:318–48
Sobel ME. 2005. Discussion: 'the scientific model of causality.' *Sociol. Methodol.* 35:99–133
Sobel ME. 2006. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *J. Am. Stat. Assoc.* 101:1398–407

Train KE. 2003. *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge Univ. Press
van der Laan MJ, Robins JM. 2003. *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer
Vytlacil EJ. 2002. Independence, monotonicity, and latent index models: an equivalence result. *Econometrica* 70:331–41
Waldfogel J, Han W-J, Brooks-Gunn J. 2002. The effects of early maternal employment on child cognitive development. *Demography* 39:369–92
Winship C, Harding DJ. 2008. A mechanism-based approach to the identification of age-period-cohort models. *Sociol. Methods Res.* 36:362–401
Winship C, Morgan SL. 1999. The estimation of causal effects from observational data. *Annu. Rev. Sociol.* 25:659–706
Wooldridge JM. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press
Wooldridge JM. 2006. *Introductory Econometrics: A Modern Approach*. Mason, OH: Thompson/South-Western. 3rd ed.
Wooldridge JM. 2007. Inverse probability weighted estimation for general missing data problems. *J. Econom.* 141:1281–301
Yakubovich V. 2006. Weak ties, information, and influence: how workers find jobs in a local Russian labor market. *Am. Sociol. Rev.* 70:408–21

Annual Review
of Sociology
Volume 36, 2010

# Contents

**Institutions and Culture**



**Formal Organizations**



**Political and Economic Sociology**



**Differentiation and Stratification**

**Individual and Society**



**Policy**



**Indexes**



**Errata**

An online log of corrections to *Annual Review of Sociology* articles may be found at http://soc.annualreviews.org/errata.shtml