Econometric Reviews, 30(1):109–127, 2011

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2011.520571

Taylor & Francis
Taylor & Francis Group

THE RELATION OF DIFFERENT CONCEPTS OF CAUSALITY USED IN TIME SERIES AND MICROECONOMETRICS

Michael Lechner

Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, St. Gallen, Switzerland

□ *Granger and Sims noncausality (GSNC), a concept frequently applied in time series econometrics, is compared to noncausality based on concepts popular in microeconometrics, program evaluation, and epidemiology literature (potential outcome noncausality or PONC). GSNC is defined as a set of restrictions on joint distributions of random variables with observable sample counterparts, whereas PONC combines restrictions on partially unobservable variables (potential outcomes) with different identifying assumptions that relate potential outcome variables to their observable counterparts. Based on the Robins' dynamic model of potential outcomes, we find that in general neither of the concepts implies each other without further (untestable) assumptions. However, the identifying assumptions associated with the sequential selection of the observables link these concepts such that GSNC implies PONC, and vice versa.*

Keywords Dynamic treatments; Granger causality; Potential outcome model; Rubin causality; Robins causality; Sims causality.

JEL Classification C21; C22; C23.

1. INTRODUCTION

One of econometricians' most important tasks of is to uncover causal relations between economic variables and distinguish them from associational relationships, also called spurious correlations. Only causal relations are useful for policy advice, because they contain the reaction of the economic variables of interest to policy interventions. Following classical economic theorists, like Marshall, or, more recently but in the same spirit, Hicks (1979), it is the effect of the ceteris paribus intervention that is of interest.

Address correspondence to Michael Lechner, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbüelstrasse 14, St. Gallen CH-9000, Switzerland; E-mail: Michael.Lechner@unisg.ch; Website: www.sew.unisg.ch/lechner

In the research program of the Cowles commission,[1] it was already clear that their interpretation of causality does imply such "counterfactual" variations embedded in the system of simultaneous equations that was the workhorse of those days. Apparently, to have the power to analyze counterfactual situations (e.g., a world with and without a particular policy), untestable, "identifying" assumptions have to be invoked. After the research program of the Cowles commission at least partly failed, one of the reactions of econometricians to that failure was to lessen the restrictions that were implied by the structure of the linear simultaneous equations model (Heckman, 2000, gives an excellent account of these developments). In time series econometrics, a more prediction-based approach to causality was developed that was largely attributed to Granger (1969) and Sims (1972). We will call this approach Granger–Sims causality below. In microeconometrics, the so-called potential outcome approach was "imported" from statistics (biometrics) and adapted to its needs. Its formalization is frequently attributed to Rubin (1974). However, there are important aspects and differences between these approaches that are not yet fully understood. This is partly so, because time-series concepts are obviously dynamic in nature, whereas for a long time the work of microeconometricians have naturally been based on a static framework.

To be more precise, the concept used in time series econometrics is based on work of Wiener (1956), Granger (1969), and Sims (1972) (e.g., see the review article by Geweke, 1984). Their basic idea is that (non) causality is very similar to, if not the same as, (non) predictability. Therefore, they consider one variable not to cause another variable if the current value of the causing variable does not help predict future values of that variable. This statement is conditional on the information set available at each point in time. This concept is, in principle (technically), applicable if one cross-sectional unit (e.g., a country) is observed for a sufficiently long time.[2]

The alternative concept popular in microeconometrics, particularly and most explicitly in the program evaluation literature (e.g., Heckman et al., 1999), is based on the idea that the relevant comparison is between different states of the world, each of which relates to a value of the causing variable. In the absence of a causal relationship, the realized outcomes would be the same even if those potential states of the world were true. To relate this concept of different states of the world to data,

[1]For an overview of the work by the Cowles commission, see for example Christ (1994).

[2]Note that we do not attempt to analyze the relation of the approach of the Cowles commission concerning causality (see, for example, Haavelmo, 1943, or Simon, 1953, 1954) to the subsequent developments in econometrics, as this has already been done, for example, by Cooley and LeRoy (1985), for macroeconometrics, and by Heckman (2000), for microeconometrics. Heckman (2000) contains also an excellent account of the relation of causality to ceteris paribus intervention as was seen by the very early economic theorists.

it is necessary to observe different sample units in different states. Then, so-called identifying assumptions are employed to relate the observed data to the distribution of the potential outcome variables, so that causal effects can be inferred from the "real world" that is reflected in the data. The statistical formulation of the resulting inference problem was probably due to Neyman (1923) and was extended and popularized by Rubin (1974). Robins (1986) first suggested dynamic versions of the potential outcome approach. In principle, for this approach to be technically applicable, there is no need to use time series variation in the data as long as there is enough cross-sectional variation.

Apparently, there is nothing specific to these concepts such that they may only be applied in micro- or time series econometrics. They are based on different general principles that may be applied to all types of data. For example, Adams et al. (2003) use the predictive approach to analyze micro data, whereas Angrist and Kuersteiner (2004, 2008) apply the potential outcome approach in a time series context. In particular, when the data have a time dimension as well as a cross-sectional dimension, both approaches may be applicable. In this case, the dynamic approach to potential outcomes provides a useful framework to compare both concepts on an equal footing. It addresses not only heterogeneity issues that are a key concern in microeconometrics, but also dynamics that is a common feature of time series econometrics.

Explicit comparisons of these two concepts of causality are limited. Heckman (2000) in his historical account of causality in econometrics does not attempt a formal comparison of these causality concepts. Holland (1986), in his overview of causality in different fields, briefly analyzes Granger causality in a static model of potential outcomes and shows an equivalence of the two concepts under a randomization condition. The exchange between Granger (1986) and Holland (1986), which was part of the discussion of the Holland (1986) article, does not really clarify the distinguishing features either. Robins et al. (1999a) (informally) note the relationship of predictive noncausality to noncausality based on dynamic potential outcome models. This relation is also noted and discussed by Angrist and Kuersteiner (2004, 2008) who develop formal test procedures for noncausality in a time series context that are motivated by the potential outcome approach. In an attempt to broaden the understanding of the causal concepts that underlie the predictive concept of causality, Robins (2003) formally relates Granger noncausality to the concept of the faithfulness analysis of causation by Spirtes et al. (1993).[3] White (2006)

[3]Faithfulness analysis uses directed acyclical graphs to formalize its assumptions and causal relations. Details on directed acyclical graphs in causal analysis can be found, for example, in Pearl (2000).

addresses the related topic of estimating the effects of single interventions with time series data.[4]

This article formally analyzes the relation between the two concepts, so that their differences are clearly explicable. In doing so, we use the nonparametric dynamic model of potential outcomes to analyze the differences between predictive (Granger–Sims) noncausality and noncausality defined by potential outcomes. We find that, in general, neither of the concepts implies the other without further assumptions (of course, this feature of Granger–Sims causality has already been previously observed in the comparisons of it to causality as defined by the Cowles commission; see, e.g., Cooley and LeRoy, 1985; Simon, 1953, 1954, and others). However, the identifying assumptions associated with the sequential selection of the observables link these concepts. Once they are added, noncausality based on the Granger–Sims definition implies noncausality based on the Robins dynamic potential outcome version, and vice versa. Thus, if such assumptions were valid, both approaches would allow tests for zero causal effects. Moreover, the results of these tests can be interpreted using the differing intuitions on which these concepts are based.[5]

The article proceeds as follows. Section 2 presents the concepts of (non) causality based on observable variables. Section 3 presents the causal model based on potential outcomes in its dynamic form and discusses identifying assumptions. Section 4 relates those concepts to each other, and Section 5 concludes.

2. CAUSALITY BASED ON OBSERVABLE OUTCOMES: WIENER–GRANGER–SIMS NONCAUSALITY

Let us define two stochastic processes $D = \{D_t\}$ and $Y = \{Y_t\}$ that may not necessarily be stationary. The data available consist of a random sample $(d_{0i}, d_{1i}, \ldots, d_{Ti}, y_{0i}, y_{1i}, \ldots, y_{Ti})$ coming from independent and identical draws $(i = 1, \ldots, N)$ from the random variables within some time window of those processes $(D_0, D_1, \ldots, D_T, Y_0, Y_1, \ldots, Y_T)$. The question is whether the factors described by D are causing changes in the variable Y. We define the terminology calling Y the outcome variable (measuring the effect) and

[4]White (2006) calls these interventions *natural experiments*. He uses a technically highly sophisticated framework that is appropriate for his discussion but neither necessary nor helpful to support the ideas of this article.

[5]As already mentioned, the literature based on comparing the ceteris paribus approach to causality (based on untestable structural assumptions in simultaneous linear models) used by the Cowles commission to the Sims–Granger approach (e.g., Cooley and LeRoy, 1985) is related, as it is to some extent similar to the potential outcome approach. One of the major differences is that the latter is nonparametric and allows arbitrary effect heterogeneity and avoids explicit modelling of a large set of causal relations simultaneously. Therefore, the formal analysis of Cooley and LeRoy (1985) does not carry over to this case.

D the causing or treatment variable. The latter term is common in the biometric and econometric evaluation literature.

In its original article Granger (1969, p. 428) explains his concept of causation as "We say that D_t is causing Y_{t+1} if we are better able *to predict* Y_{t+1} using all available information than if the same information without D_t had been used" (notation adjusted; italics added). He distinguishes between instantaneous causality, when the value of Y_{t+1} can better be predicted with the value of D_t given the history of D_t than without it, and the case when it takes some periods until the effect manifests itself in the outcome variables. With a similar concept in mind, Sims (1972, p. 545) explains that "... if causality runs from D to Y only, future values of D in the regression [of Y on D and perhaps other 'exogenous' variables] should have zero coefficients." Furthermore, they also pointed out that a cause must precede any effect of it. Initially, the formalization of these concepts used linear predictors.[6] In this context, Hosoya (1977) showed the equivalence of those two concepts (see also Florens and Mouchart, 1985).

Chamberlain (1982), Florens and Mouchart (1982), and Engle et al. (1983) strengthened the conditions by basing the definitions on properties of conditional distribution functions instead of conditional means. This has the added virtue that the definitions become relevant for all types of economic variables, whether they are related by a linear conditional mean or not. In this article, we adopt this specification as well. To condense notation, the history from period 1 to t of D and Y is denoted by $\underline{D}_t = (D_1, \ldots, D_t)$ and $\underline{Y}_t = (Y_1, \ldots, Y_t)$. The initial conditions are collected in $A_0 = (D_0, Y_0)$. Furthermore, letting small letters denote specific values of the random variables, Definition 1 formally defines the concept of predictive noncausality.

Definition 1 (GNSC: Granger–Sims Noncausality). $\underline{D}_t$ does not GS-cause Y_{t+1}, if and only if $Y_{t+1} \coprod \underline{D}_t \mid \underline{Y}_t = \underline{y}_t, A_0 = a_0; \forall \underline{y}_t; \forall a_0; \forall t = 1, \ldots, T-1$.[7]

Note that we slightly deviate from the Chamberlain (1982) notation and condition directly on the random variables of the first period observed in the data (initial conditions), as in Engle et al. (1983).[8] We do this for the

[6] In those times, econometrics was almost entirely concerned with the estimation of linear relations of continuous variables.

[7] $A \coprod (B_1, B_2) \mid C = c$ means that A and the elements of B are jointly independent conditional on C taking a value of c (i.e., Dawid, 1979). Denoting the cumulative distribution function (cdf) of D conditional on E evaluated at d and e as $F_{D \mid E}(d, e)$, this statement is equivalent to $F_{A,B_1,B_2 \mid C}(a, b_1, b_2, c) = F_{A \mid C}(a, c) F_{B_1,B_2 \mid C}(b_1, b_2, c)$, $\forall a, b_1, b_2$.

[8] Engle et al. (1983) discuss related, but not identical concepts of strict exogeneity. In that their discussion focuses on likelihood functions and the role of their parameters in efficient and consistent estimation, it does not lend itself directly to the desired comparison of different concepts of causality.

sake of notational simplicity in the comparison of the concepts of causality later on. Similarly, further delays of cause and effect may be introduced, but they are an unnecessary complication for the purpose of this article.[9]

Sims (1972) proposed an alternative, but similar definition of noncausality, which in its independence version proposed by Chamberlain (1982), is given by $(Y_T, \ldots, Y_{t+1}) \perp\!\!\!\perp D_t \mid \underline{Y}_t, A_0$. It is a direct implication of Definition 1 (but not vice versa). Although, it has some intuitive appeal in that there is an absence of correlation between current intervention and future outcomes given past outcomes, there exists an ambiguity about the causal implication from not conditioning on past interventions D (which is equivalent to assuming the independence of D_t, but not of $\underline{D}_t$). Whereas in this article we focus on the (full) effect of D on Y, the Sims definition only seems to capture part of that. In particular, this is case when the time horizon is finite, as will be assumed here. The lagged effects of the intervention may be 'absorbed' in the conditioning set.

Chamberlain (1982) suggests an alternative and stronger version of the Sims's definition that conditions on the past values of D as well as $[(Y_T, \ldots, Y_{t+1}) \perp\!\!\!\perp D_t \mid \underline{Y}_t, \underline{D}_{t-1}, A_0]$. This stronger version results in a definition, which is equivalent to the Granger definition (i.e., Chamberlain, 1982). This equivalence holds true as long as all conditioning variables are treated symmetrically, i.e., as long as they can be subsumed in Y. Using different analytical frameworks, Dufour and Tessier (1993), Florens and Fougère (1996), and Dufour and Renault (1998) show that this equivalence disappears when additional "control" variables are present, which are influenced by D but not included in Y. This non-equivalence result is also contained in Angrist and Kuersteiner (2004, 2008). Here, for the sake of brevity, we do not consider the original version of Sims (1972) explicitly. Instead, we chose the name of Granger–Sims noncausality for the relation stated in Definition 1 to give credit to both "inventors" of this type of causality. For the sake of notational simplicity, we also refrain from considering conditional versions of the two concepts of predictive causality for which the equivalence result does not hold.

Letting $F(\cdot)$ denote a cumulative distribution function and using short hand notation for the conditioning values, Definition 1 is equivalent to $F_{\underline{D}_t \mid Y_{t+1}, \underline{Y}_t, A_0}(\underline{d}_t, y_{t+1}, \underline{y}_t, a_0) = F_{\underline{D}_t \mid \underline{Y}_T, A_0}(\underline{d}_t, \underline{y}_T, a_0) = F_{\underline{D}_t \mid \underline{Y}_t, A_0}(\underline{d}_t, \underline{y}_t, a_0)$, i.e., the distribution of $\underline{D}_t$ and its elements do not depend on future outcomes conditional on the history of the process. Therefore, the joint distribution

[9]Dufour and Renault (1998) study the differences of long run causality from short run causality in a linear model by considering different lag lengths between the outcome variable and the causing and conditioning variables.

of all random variables may be written as follows:

$$\begin{aligned} F_{\underline{D}_T,\underline{Y}_T \mid A_0}(\underline{d}_T, \underline{y}_T, a_0) &= F_{\underline{D}_T \mid \underline{Y}_T, A_0}(\underline{d}_T, \underline{y}_T, a_0) F_{\underline{Y}_T \mid A_0}(\underline{y}_T, a_0) \\ &= \prod_{t=1}^{T} F_{D_t \mid \underline{D}_{t-1}, \underline{Y}_T, A_0}(\underline{d}_t, \underline{y}_T, a_0) \prod_{t=1}^{T} F_{Y_t \mid \underline{Y}_{t-1}, A_0}(\underline{y}_t, a_0) \\ &= \prod_{t=1}^{T} F_{D_t \mid \underline{D}_{t-1}, \underline{Y}_t, A_0}(\underline{d}_t, \underline{y}_t, a_0) \prod_{t=1}^{T} F_{Y_t \mid \underline{Y}_{t-1}, A_0}(\underline{y}_t, a_0). \end{aligned}$$

Furthermore, we have $F_{Y_{t+1} \mid \underline{Y}_t, \underline{D}_t, A_0}(\underline{y}_{t+1}, \underline{d}_t, a_0) = F_{Y_{t+1} \mid \underline{Y}_t, A_0}(\underline{y}_{t+1}, a_0)$ for all t. These conditions can be tested by estimating and comparing appropriate distributions using formal test procedures (e.g., Li et al., 2009). Furthermore, they have many obvious implications on sample counterparts, which can be used for testing as well.

3. CAUSAL EFFECTS DEFINED BY POTENTIAL OUTCOMES: MARSHAL–NEYMAN–RUBIN–ROBINS CAUSALITY

3.1. The Concept of Causality Based on Potential Outcomes

The approach of potential outcomes has its roots in the idea that a causal effect is a reaction of an outcome variable to a manipulation of another variable keeping other factors constant. In economics, this classical ceteris paribus condition is the cornerstone of economic analysis.[10] The factors kept constant in such an intellectual exercise are typically those not influenced by the intervention but influencing the outcomes. Typically, this is a thought experiment in that it requires imagining how the world would have developed had the specific intervention occurred or not. Therefore, additional conditions are required before the data can be used for resolving the causal question. The statistical formulation was probably based on work by Neyman (1923), Wilks (1932), and Cochran and Chambers (1965). It has been highly popularized by the works of Rubin (1974, 1977, etc.; see also the non-technical overviews contained in Heckman, 2000, or Rubin, 2005). A similar approach has been proposed in economics by Roy (1951) and already implicitly by the Cowles commission.

To simplify notation, consider a discrete intervention changing the causing variable D from d to d'. d and d' differ at least once between 1

[10]See, for example, the classical works by Marshall (1961), the Cowles Commission (e.g., Haavelmo, 1943; Simon, 1953, 1954), and others, as discussed in the historical account of causal analysis by Heckman (2000), or the extensive discussion of ceteris paribus causality provided by Hicks (1979). Heckman (2005) provides an elaborate discussion of potential outcome models and how they are embedded in economic theory.

and $T-1$. We are interested in the question whether the outcomes would change due to a change in D. As before, we presume that the cause must precede its effect. To capture the notion of a.c.p. change (i.e., the comparison of two different "states of the world"), we define the outcomes as functions of d as well as of other factors u and compare their difference for different values of d and the same value of u. We are interested in the difference between $Y(d',u)$ and $Y(d,u)$.[11]

Let us define a causal effect of $\underline{D}_t$ on Y_{t+1} given initial conditions $\theta_{t+1}(y_{t+1};\underline{d}'_t,\underline{d}_t,u_t)=F_{Y_{t+1}(\underline{d}'_t,u_t)}(y_{t+1})-F_{Y_{t+1}(\underline{d}_t,u_t)}(y_{t+1})$. First, note that here we consider the difference in distribution functions instead of the more common average or quantile effects (see, e.g., Firpo, 2007). An alternative would be to base the definition on the more general concept of the "D-parameter" introduced by Manski (1997). The D-parameter encompasses all of the effects that are based on some function $g(\cdot)$ that respect the inequality $g(v)\geq g(w)$ whenever v stochastically dominates w. Although, those effects include mean as well as quantile effects, they do not include effects, for example, on variances or other measures of spread. Thus, we stick to the most stringent definition, but the technical discussion applies to mean, quantile, and D-effects only with small and obvious changes and some additional regularity conditions (such as the existence of appropriate moments for effects based on the comparison of particular moments of the potential outcomes). Although this article does not touch on estimation issues at all, it should be pointed out that the recent advances in nonparametrically testing the equality of conditional and unconditional distributions mentioned at the end of the previous section make distribution-based definitions more attractive for applied work as well.

Second, this definition is based on the difference of the distribution functions instead of on the distribution function of the differences of the potential outcomes. The reasoning behind this is as follows: (i) For the distribution function of the differences it is almost impossible to obtain consistent estimators under reasonable assumptions for any measures other than linear operators, such as averages for which the mean of the difference equals the difference of the means of the marginal distributions. Here there is no information in the data useful for nonparametric estimation of the joint distribution of the potential outcomes, because no unit can be observed in both states at the same time. Therefore, this

[11] $Y(d',u)$ and $Y(d,u)$ are called potential outcomes, because "the world cannot be in the two different states at any given time." Therefore, only $Y(d',u)$ or $Y(d,u)$ is observed if one of those two states is realized at all. For a fierce attack from the statistical point of view on such a concept of causality, see for example Dawid (2000). Despite that critique, this concept appears to be widely used in the sciences and economics, and particularly so in applied microeconometrics. For a further discussion, see the excellent exposition of the potential outcome approach by Holland (1986).

concept has (almost) never been applied in (non- or semiparametric) empirical studies.[12] (ii) Comparing marginal distributions of potential outcomes is better suited for a comparison with GSNC and is not distracted with issues irrelevant to econometric practice.

Assume that there is no u data available. Therefore, only effects averaged over some population may be estimated with the data, like, $\theta_{t+1}(y_{t+1}; \underline{d}'_t, \underline{d}_t, s_t, a_0) = E_{u_t \mid u_t \in s_t, A_0=a_0}[\theta_{t+1}(y_{t+1}; \underline{d}'_t, \underline{d}_t, u_t)] = E_{u_t \mid u_t \in s_t, A_0=a_0}\{F_{Y_{t+1}(\underline{d}'_t, u_t) \mid A_0}(y_{t+1}, a_0)\} - E_{u_t \mid u_t \in s_t, A_0=a_0}\{F_{Y_{t+1}(\underline{d}_t, u_t) \mid A_0}(y_{t+1}, a_0)\}$, where s_t denotes some population of interest defined by u_t.[13] Note that the definition takes the initial condition fully into account. However, this is not mentioned explicitly in the discussion below.

There is an issue here whether noncausality should mean that the causal effect is zero for every value of u_t (i.e., $\theta_{t+1}(y_{t+1}; \underline{d}'_t, \underline{d}_t, u_t) = 0$), or just on average for some population. The treatment effect literature places much emphasis on the fact that effects may differ in subpopulations defined by D. However, GSNC is formulated as the population as a whole, conditional on initial conditions. Therefore, we will only consider (zero) distributional effects averaged for the population, denoted by $\theta_{t+1}(y_{t+1}; \underline{d}'_t, \underline{d}_t) = 0$, to allow for a comparison that focuses on the key components of different concepts of causality. This implies that noncausality in all concepts allows for negative and positive effects at the disaggregated level as long as they wash out for the population. Finally, it should be pointed out that for notational simplicity, this notion suppresses the dependence of the effect on the initial conditions A_0.

Definition 2 (Potential Outcome Noncausality, PONC). $\underline{D}_t$ does not PO-cause Y_{t+1} if and only if $\theta_{t+1}(y_{t+1}; \underline{d}'_t, \underline{d}_t) = 0, \forall y_{t+1}, \forall d'_t \neq d_t, t = 1, \ldots, T-1$.

This notation is adapted to Granger's convention with respect to timing of cause and effect. There is a major conceptional difference to the approach presented in the previous section, namely, that in the potential outcome approach the definition of the effect and its discovery from the data are two distinct steps that are considered separately.

[12] For attempts to bound effects that are based on the joint distribution, see Heckman et al. (1997). However, their bounds turn out to be so large as to be only of very limited relevance in empirical applications.

[13] u_t may contain past values of u, but this is suppressed for notational convenience. For an overview of all the different effects discussed in the applied microeconometric literature and an attempt to put them in a unified framework, see Heckman and Vytlacil (2005). The emphasis on the effect heterogeneity in different populations that appear in many applied studies based on the potential outcome approach is not prominent in GSNC. This is probably due to their different origins and fields of application. The potential outcome approach is used frequently in fields in which cross-sectional effect heterogeneity is considered important and the data have a large cross-sectional dimension. Granger–Sims noncausality originates from the time series literature, which historically is much less concerned with heterogeneity of causal effects and frequently has to rely on only one draw from the population of interest.

Therefore, the quantity defined in Definition 2 cannot be empirically tested without further assumptions. The microeconometric literature has discussed numerous ways to identify these causal effects in the data when there are other variables available. As mentioned before, to concentrate our analysis on the key conceptional differences between the two definitions of noncausality, we consider the case without any other variables, only D and Y.

3.2. A Form of Potential Outcome Causality that can be Inferred from the Data

The first link of the observed outcome variables to the potential outcomes is the fact that potential outcomes are observed for the value of d_t that is realized in the data (d_{ti}). This is to say that the distribution of the observable outcome conditional on treatment is the same as the distribution of the potential outcome related to that treatment and conditional on it ($F_{Y_{t+1}|\underline{D}_t,A_0}(y_{t+1},\underline{d}_t,a_0) = F_{Y_{t+1}(\underline{d}_t)|\underline{D}_t,A_0}(y_{t+1},\underline{d}_t,a_0)$).[14] In the so-called treatment effect literature, this connection is rationalized by the so-called observation rule that can be stated as $Y_{t+1} = \sum_{\underline{d}_t} \underline{1}(\underline{D}_t = \underline{d}_t) Y_{t+1}(\underline{d}_t)$, where $\underline{1}(\cdot)$ denotes the indicator function which is one when the element inside the brackets is true. This 'observation rule' is closely related to Rubin's (1980) "Stable Unit Treatment Value Assumption" (SUTVA) and Robins' (1986) "Consistency Condition."

Even with the observation rule, we still cannot relate this concept of noncausality to data. For example, the observed variables can never uncover an effect like $F_{Y_{t+1}(\underline{d}'_t)|\underline{D}_t,A_0}(y_{t+1},\underline{d}_t,a_0) - F_{Y_{t+1}(\underline{d}_t)|\underline{D}_t,A_0}(y_{t+1},\underline{d}_t,a_0)$. Although the second term in the difference relates to observables (because it concerns the population that is actually observed in that state, thus $F_{Y_{t+1}(\underline{d}_t)|\underline{D}_t,A_0}(y_{t+1},\underline{d}_t,a_0) = F_{Y_{t+1}|\underline{D}_t,A_0}(y_{t+1},\underline{d}_t,a_0)$), the first one does not. Therefore, assumptions are required to relate terms like $F_{Y_{t+1}(\underline{d}'_t)|\underline{D}_t,A_0}(y_{t+1},\underline{d}_t,a_0)$ to random variables for which realizations can be found in the data, namely, elements of $(\underline{Y}_T,\underline{D}_T,A_0)$. Robins (1986, 1989, 1997), Gill and Robins (2001), and Lechner and Miquel (2005), among others, analyzed such conditions in similar dynamic causal frameworks based on potential outcomes.[15] Here, we base our account on a simplified

[14]In order to simplify notation, the dependence of outcomes and treatments on u_t is left implicit for most of this and the following sections. In such cases, u_t is integrated out with respect to some distribution, which is obvious from the specific context.

[15]These articles are based on the so-called selection on observables assumption, which is the route followed below, although in a simplified way. Several articles by James Robins and co-authors are concerned with parametric and semiparametric estimations of this model, which thus far have been used little or not at all in econometric applications (e.g., Hernan et al., 2001; Robins, 1999; Robins et al., 1999a,b). Lechner (2009) discusses weighting and matching estimators and points to some practical issues for evaluating labor market programs. Miquel (2002) considers the case of selection on unobservables that requires more data than just the outcomes and treatments. Abbring and Heckman (2008) provide a survey over dynamic causal models.

version of the econometric dynamic treatment framework using the notation suggested by the latter authors.

Within that framework, we formulate conditions that allow us to infer some of the $\theta_{t+1}(y_{t+1}; \underline{d}'_t, \underline{d}_t)$ from the data. Without data other than the realizations from $(\underline{Y}_T, \underline{D}_T, A_0)$, the only way to achieve nonparametric point identification of an average causal effect is to assume randomization, i.e., whether unit 'i' observed in some regime d or d' is to some extent random. Of course, the specific type of randomness must be specified exactly.

Consider the weakest of such assumptions that have appeared in the literature thus far. Namely, consider the assumption that conditional on the realized history of Y and D (and A_0), the next realization of D is independent of the potential outcomes. Such a sequential randomization assumption[16] allows the units (economic agents, ...) to use the information about the past as given by $(\underline{D}_{t-1}, \underline{Y}_t)$ to select the state D_t. This randomization is conditional on the history of treatment and outcome variables. Thus, in period t different units of the population may have different probabilities to end up in d_t, depending on their past realizations of the outcome and treatment variables. This assumption is called the weak dynamic conditional independence assumption (W-DCIA) by Lechner and Miquel (2005). It resembles the conditional independence assumption (CIA), which is a prominent feature in static analysis. Holland's (1986) argument was based on CIA. Dynamic extensions of CIA were initially proposed and formally analyzed by Robins (1986).

Assumption 1 (Weak Dynamic Conditional Independence Assumption, W-DCIA)**.**

$$
\begin{aligned}
&Y_{t+1}(\underline{d}_t) \coprod D_1 \mid Y_1 = y_1, A_0 = a_0; && \forall a_0; \forall \underline{d}_t; \forall y_1; && \forall t = 1, \ldots, T-1;\\
&Y_{t+1}(\underline{d}_t) \coprod D_\tau \mid \underline{Y}_\tau = \underline{y}_\tau, && \forall a_0; \forall \underline{d}_t; \forall \underline{y}_t; && \forall t = 2, \ldots, T-1.\\
&\underline{D}_{\tau-1} = \underline{d}_{\tau-1}, A_0 = a_0; && \forall \tau = 2, \ldots, t;
\end{aligned}
$$

Lechner and Miquel (2005) show that although population treatment effects are identified based on this assumption, classical treatment on the treated effects, i.e., the effects on the population of those units subject to a specific realization of $\underline{D}_{T-1}$, are not identified. Thus, this assumption appears as a weak version of a dynamic conditional independence assumption.[17] However, it suffices for the purpose of this article because any equivalence result that can be obtained under this assumption will

[16]Due its origins in experimental evaluations, it is common in this literature to call this randomization instead of exogeneity. In fact, depending on the exact formulation of these concepts they may be either very similar or even identical (see Imbens, 2004, for further considerations on this topic).

[17]Note that Assumption 1 differs from White's (2006) DUNE assumption in that it conditions on observed past treatments and outcomes.

also hold under the assumptions that nest W-DCIA.[18] Note that in the first period there is a static version of the conditional independence assumption as used in Holland (1986).

Two more conditions are necessary to use the data together with W-DCIA to test PONC. First, it is required that realizations of the outcome variables can actually be found for all paths of interest of $\underline{D}_{T-1}$. This so-called *common support assumption* must hold conditionally on past outcomes. Second, for this notation to cover a ceteris paribus intervention, it is necessary that the potential outcomes for a specific state do not depend on the extent of the intervention. In other words, the value of $Y(d,u)$ does not depend on the fact that it is compared with $Y(d',u)$ or $Y(d'',u)$. This leads to the previously mentioned observation rule.

Property 1 (Causal Effects with Potential Outcomes Based on W-DCIA). If W-DCIA holds true, the causal effects depend on $(D_0, D_1, \ldots, D_T, Y_0, Y_1, \ldots, Y_T)$ as follows:

$$F_{Y_{t+1}(\underline{d}_t)\,|\,A_0=a_0}(y_{t+1}, a_0)$$
$$= \underset{Y_1\,|\,A_0=a_0}{E}\left[\underset{Y_2\,|\,D_1=d_1,Y_1,A_0}{E}\left[\ldots \underset{Y_t\,|\,\underline{D}_{t-1}=\underline{d}_{t-1},\underline{Y}_{t-1},A_0}{E}\left[F_{Y_{t+1}\,|\,\underline{D}_t,\underline{Y}_t,A_0}(y_{t+1},\underline{d}_t,\underline{y}_t,a_0)\right]\ldots\right]\right];$$
$$\forall \underline{d}_t; \forall a_0; \forall t.$$

$$\theta_{t+1}(y_{t+1}, \underline{d}'_t, \underline{d}_t)$$
$$= \underset{Y_1\,|\,A_0}{E}\left[\underset{Y_2\,|\,D_1=d'_1,Y_1,A_0}{E}\left[\ldots \underset{Y_t\,|\,\underline{D}_{t-1}=\underline{d}'_{t-1},\underline{Y}_{t-1},A_0}{E}\left[F_{Y_{t+1}\,|\,\underline{D}_t,\underline{Y}_t,A_0}(y_{t+1},\underline{d}'_t,\underline{y}_t,a_0)\right]\ldots\right]\right.$$
$$\left. - \underset{Y_2\,|\,D_1=d_1,Y_1,A_0}{E}\left[\ldots \underset{Y_t\,|\,\underline{D}_{t-1}=\underline{d}_{t-1},\underline{Y}_{t-1},A_0}{E}\left[F_{Y_{t+1}\,|\,\underline{D}_t,\underline{Y}_t,A_0}(y_{t+1},\underline{d}_t,\underline{y}_t,a_0)\right]\ldots\right]\right];$$
$$\forall y_{t+1}; \forall \underline{d}_t, \underline{d}'_t; \forall a_0; \forall t.$$

The proofs of these properties follow directly from the identification proofs of Robins (1986, 1989, 1997), Gill and Robins (2001), and Lechner and Miquel (2005). Therefore, they are not repeated here.

As is seen in Property 1, identification is achieved by continuously reweighting the units that receive d_t towards the distributions of characteristics that describe the population of interest. By doing so, the growing number of conditioning variables and time order of variables is respected. This is called the *g*-formula by Robins (1986).

[18]To identify all usual treatment effects, Lechner and Miquel (2005) suggest a more restrictive version of the W-DCIA by imposing additional conditions on the way in which past treatments can influence past observed outcomes (strong dynamic conditional independence assumption, S-DCIA). Furthermore, if the complete treatment path is randomized in the beginning of the first period, then this assumption is stronger than W-DCIA as well.

4. RELATION BETWEEN THE DIFFERENT CONCEPTS

4.1. General Results

Note that Definition 1 summarizes the conditions that GSNC imposes on the data. Definition 2 defines PONC. Property 1 shows how the PO-causal effects depend on the data if either of the "identifying" Assumption 1 (W-DCIA) holds true. Hence, if GSNC together with these properties imply a zero causal effect ($\theta_{t+1}(y_{t+1}, \underline{d}'_t, \underline{d}_t) = 0; \forall y_{t+1}; \forall \underline{d}_t, \underline{d}'_t; \forall t$), we conclude that GSNC together with W-DCIA implies PONC. Conversely, if the restrictions $\theta_{t+1}(y_{t+1}, \underline{d}'_t, \underline{d}_t) = 0; \forall y_{t+1}; \forall \underline{d}_t, \underline{d}'_t; \forall t$ imposed on Property 1 imply Definition 1, we conclude that the combination of these assumptions with PONC implies GSNC.

However, before considering the combinations of identifying assumptions with causality definitions, we state the obvious in Lemma 1.

Lemma 1 (GSNC and PONC Only)**.**

a) *GSNC does not imply PONC.*
b) *PONC does not imply GSNC.*

This lemma is true, because PONC, without further assumptions, does not impose any restrictions on the distribution of $(D_0, D_1, \ldots, D_T, Y_0, Y_1, \ldots, Y_T)$ that are relevant to GSNC.

This result may seem trivial. However, it points to the important fact that ceteris paribus interventions, which are directly reflected in models based on contrasts of outcomes in two different states of the world, have no consequences for the data, if they are not *enriched* with further (*untestable*) assumptions. In other words, any restrictions put on the data (in the form of testable hypothesis) are silent about underlying causal effects that generated the data unless further untestable assumptions are added to relate the potential worlds required to define the effects of c.p. interventions to the data.

The following theorem shows that the sequential randomization assumption W-DCIA provides the following equivalence results for the different concepts of causality.

Theorem 1 (GSNC and PONC Combined with W-DCIA)**.** *Suppose Assumption* 1 *(W-DCIA) holds true and there is common support.*

a) *GSNC implies PONC.*
b) *PONC together with the monotonicity condition*

$$F_{Y_{t+1} \mid \underline{D}_t, \underline{Y}_t, A_0}(y_{t+1}, \underline{d}'_t, \underline{y}_t, a_0) \leq F_{Y_{t+1} \mid \underline{D}_t, \underline{Y}_t, A_0}(y_{t+1}, \underline{d}_t, \underline{y}_t, a_0), \quad \forall y_t, \forall t = 1, \ldots, T-1; \quad \textit{or}$$

$$F_{Y_{t+1}\mid \underline{D}_t,\underline{Y}_t,A_0}(y_{t+1},\underline{d}'_t,\underline{y}_t,a_0) \geq F_{Y_{t+1}\mid \underline{D}_t,\underline{Y}_t,A_0}(y_{t+1},\underline{d}_t,\underline{y}_t,a_0), \quad \forall y_t, \forall t = 1,\ldots,T-1,$$

implies GSNC.

Part a): GSNC implies that the distribution of Y_{t+1} given past outcomes does not depend on any of the past D_t, $F_{Y_{t+1}\mid \underline{D}_t=\underline{d}_t,\underline{Y}_t,A_0}(y_{t+1},\underline{d}_t,\underline{y}_t,a_0) = F_{Y_{t+1}\mid \underline{Y}_t,A_0}(y_{t+1},\underline{y}_t,a_0)$. This condition leads to an equality of the inner terms of the causal effects given in Property 1, i.e., $F_{Y_{t+1}\mid \underline{D}_t,\underline{Y}_t,A_0}(y_{t+1},\underline{d}'_t,\underline{y}_t,a_0) = F_{Y_{t+1}\mid \underline{D}_t,\underline{Y}_t,A_0}(y_{t+1},\underline{d}_t,\underline{y}_t,a_0)$. Furthermore, in that this equality holds for all values of t, the weights implied by those iterated expectations are identical as well. Therefore, GSNC implies PONC if W-DCIA holds.

Part b): The monotonicity condition restricts the underlying effect heterogeneity allowed to go together with a zero average effect. An alternative would be to define PONC not as a population average effect, but relating to all possible subpopulations (which then implies this monotonicity condition). Our approach is also somewhat less restrictive than requiring the potential outcomes to be the same with a probability of one as in Robins et al. (1999a) as they define their "sharp causal null hypothesis."

For the proof, it is important to note that W-DCIA comes with an initial condition, i.e., the problem of the first period is essentially static:

$$\theta_2(y_2;d'_1,d_1) = \underset{Y_1\mid A_0}{E}[F_{Y_2\mid D_1,Y_1,A_0}(y_2,d'_1,y_1,a_0) - F_{Y_2\mid D_1,Y_1,A_0}(y_2,d_1,y_1,a_0)] \stackrel{!}{=} 0.$$

Assuming that $F_{Y_1\mid A_0}(y_1,a_0)$ is nonzero in the support of interest (as ensured by the common support assumption), then it must hold true that $F_{Y_2\mid D_1,Y_1,A_0}(y_2,d'_1,y_1,a_0) = F_{Y_2\mid D_1,Y_1,A_0}(y_2,d_1,y_1,a_0)$. This, however, has implications for the causal effect in the next period. Consider the zero causal effect for period 3:

$$\theta_3(y_3;\underline{d}'_2,\underline{d}_2) = \underset{Y_1\mid A_0=a_0}{E}\left[\underset{Y_2\mid D_1=d'_1,Y_1,A_0=a_0}{E} F_{Y_3\mid \underline{D}_2,\underline{Y}_2,A_0}(y_3,\underline{d}'_2,\underline{y}_2,a_0) - \underset{Y_2\mid D_1=d_1,Y_1,A_0=a_0}{E} F_{Y_3\mid \underline{D}_2,\underline{Y}_2,A_0}(y_3,\underline{d}_2,\underline{y}_2,a_0)\right] \stackrel{!}{=} 0.$$

However, because the zero causal effect from the previous period results in $F_{Y_2\mid D_1,Y_1,A_0}(y_2,d'_1,y_1,a_0) = F_{Y_2\mid D_1,Y_1,A_0}(y_2,d_1,y_1,a_0)$, the weights appearing in the difference are the same. With nonzero weights guaranteed by common support, this condition on the weights implied by PONC, W-DCIA, and monotonicity requires that $F_{Y_3\mid \underline{D}_2,\underline{Y}_2,A_0}(y_3,\underline{d}'_2,\underline{y}_2,a_0) =$

$F_{Y_3 \mid \underline{D}_2, \underline{Y}_2, A_0}(y_3, \underline{d}_2, \underline{y}_2, a_0)$. This in turn implies the equality of weights for the next period. Applying this reasoning to every period up to period T, it follows that PONC in combination with W-DCIA and monotonicity implies $F_{Y_{t+1} \mid \underline{D}_t, \underline{Y}_t, A_0}(y_{t+1}, \underline{d}'_t, \underline{y}_t, a_0) = F_{Y_{t+1} \mid \underline{D}_t, \underline{Y}_t, A_0}(y_{t+1}, \underline{d}_t, \underline{y}_t, a_0)$. This is exactly the condition for GSNC. Note that conditioning on some initial conditions as well as the definition of zero effects in all periods plays a key role in this proof.

4.2. Further Issues and Generalizations

This section takes on some issues that are related to simplifications made in this article with the purpose of clarifying the main differences between the different approaches.

The first such issue relates to additional conditioning (control) variables: All results hold true in any subset defined by variables that are not influenced by treatment variables. The previously mentioned articles describe the necessary identification results when predetermined variables are added to Assumption 1.

Another interesting type of data that might become available would be instrumental variables, i.e., variables that influence D but do not influence Y other than by changing D. In a world of heterogeneous causal effects that underlies this article, such variables identify treatment effects for a subpopulation that reacts to changes in the instruments by changes in D, the so-called compliers (Imbens and Angrist, 1994). If the instrument is discrete, individual membership in the complier population is usually unknown. Thus, since GSNC is not defined for an unobservable subpopulation, there is not much sense in comparing GSNC and PONC for that group. If, however, the instrument is continuous, every member of the population may become a complier (Heckman and Vytlacil, 2005) and a similar analysis as made in this article, adapted to the dynamic case, is appropriate.

In the comparison of GSNC and PONC, this article considered PONC for the population instead of subpopulations defined by treatment status as would be common in applied microeconometrics, and in particular in the program evaluation literature (e.g., Heckman et al., 1999). If the latter is explicitly taken into account, then for those effects that are actually identified, Lechner and Miquel's (2005) results show that the structure of the key elements in the comparison remains intact.

5. CONCLUSION

This article highlights issues of uncovering the effects of ceteris paribus interventions with econometric methods. For quite some time, ceteris

paribus interventions are typically thought of by economic theorists (like Hicks, 1979; Marshall, 1961, as examples) as comparisons of different states of the world that could have occurred. This article shows that Granger–Sims noncausality under some conditions can indeed detect the absence of such effects. The necessary additional identifying conditions required for the Granger–Sims approach to have this property are, however, not empirically testable: they have to be established from outside (theoretical) knowledge about the underlying causal structures as has, of course, already been observed at the time of the Cowles commission.[19]

In this article, we use the dynamic model of potential outcomes for formally analyzing the differences between Granger–Sims noncausality and noncausality defined by potential outcomes. In general, we find that neither of these concepts implies the other without further assumptions. However, the identifying assumptions associated with the sequential selection of the observables provide the link between these concepts. Once added, noncausality based on the Granger–Sims definition implies noncausality based on the dynamic potential outcome definition, and vice versa. Thus, if these specific untestable assumptions are plausible, then tests for zero causal effects could be based on either of the approaches.[20] Moreover, the results of those tests could be interpreted using the differing intuitions behind the different concepts.

It is worthwhile noting that our findings are unrelated to the main criticism of the Granger–Sims approach that appeared in Holland (1986) as well as in other articles. The issue is that the availability of new data may lead to additional variables entering the information set. This in turn implicitly leads to a new definition of Granger–Sims noncausality. In other words, knowing more may lead to the result that a variable previously considered a cause becomes a spurious relation. The potential outcome approach in comparison seems immune to that problem, because the identification steps are separated from the estimation steps and the available data. However, the comparison is probably not entirely fair, because in empirical practice, having new data may lead researchers to change their identifying assumptions by increasing the set of conditioning variables required for the DCIA assumptions to hold true, and thus the same phenomena as for Granger–Sims noncausality may appear.

ACKNOWLEDGMENT

I am affiliated with ZEW, Mannheim, CEPR and PSI, London, IAB, Nuremberg, and IZA, Bonn. I am thankful to Jim Heckman for

[19]"... and since the determination of the causal ordering implies identifiability, the test for spuriousness of the correlation requires additional assumptions to be made." (Simon, 1954, p. 479).

[20]Most likely, such tests are most powerful when based on full distributions instead of quantile or moments and employ recent developments in this field (e.g., Li et al., 2009).

convincing me to write down some of the issues that appear in this article. Furthermore, I am grateful to James Robins, the editor of this Journal, Esfandiar Maasoumi, and an anonymous referee for very helpful comments on earlier versions of this article. I thank Stefan Wiehler for careful proofreading. Of course, the usual disclaimer applies. I very much appreciate the previous joint work on dynamic potential outcome models with Ruth Miquel, in which we touched on a couple of issues that reappear here. The first version of the article has been written while I was visiting the Economics Department of the University of Michigan. The hospitality is appreciated.

REFERENCES

Abbring, J. H., Heckman, J. J. (2008). Dynamic Policy Analysis. *CeMMaP Working Papers CWP05/08*.

Adams, P., Hurd, M. D., McFadden, D., Merrill, A., Ribeiro, T. (2003). Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics* 112:3–46.

Angrist, J. D., Kuersteiner, G. M. (2004). Semiparametric Causality Tests Using the Policy Propensity Score. *NBER Working Paper 10975*.

Angrist, J. D., Kuersteiner, G. M. (2008). Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinominal propensity score. *IZA Discussion Paper 3606*.

Chamberlain, G. (1982). The general equivalence of granger and sims causality. *Econometrica* 50:569–581.

Christ, C. F. (1994). The Cowles Commission's contribution to econometrics at Chicago, 1939–1955. *Journal of Economic Literature* 32:30–59.

Cochran, W. J., Chambers, S. P. (1965). The planning of observational studies of human populations. *Journal of the Royal Statistical Society, Serie A* 128:234–266.

Cooley, T. F., LeRoy, S. F. (1985). Atheoretical macroeconometrics: a critique. *Journal of Monetary Economics* 16:283–308.

Dawid, A. P. (1979). Conditional independence in statistical theory. *The Journal of the Royal Statistical Society, Series B* 41:1–31.

Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Society* 95:407–448. With discussion.

Dufour, J. M., Tessier, D. (1993). On the relationship between impulse response analysis: innovation accounting and granger causality. *Economics Letters* 42:327–333.

Dufour, J. M., Renault, E. (1998). Short run and long run causality in time series: theory. *Econometrica* 66:1099–1125.

Engle, R., Hendry, D. F., Richard, J.-F. (1983). Exogeneity. *Econometrica* 51:277–304.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica* 75:259–276.

Florens, J. P., Mouchart, M. (1982). A note on noncausality. *Econometrica* 50:583–591.

Florens, J. P., Mouchart, M. (1985). A linear theory for noncausality. *Econometrica* 53:157–176.

Florens, J.-P., Fougère, D. (1996). Noncausality in continuous time. *Econometrica* 64:1195–1212.

Geweke, J. (1984). Inference and causality in economic time series. In: Griliches, Z., Intriligator, M. D., eds. *Handbook of Econometrics*. Vol. 2. Amsterdam: North-Holland, pp. 1102–1144.

Gill, R. D., Robins, J. M. (2001). Causal inference for complex longitudinal data: the continuous case. *The Annals of Statistics* 29:1–27.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37:424–438.

Granger, C. W. J. (1986). Comment. *Journal of the American Statistical Association* 81:967–968.

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

Heckman, J. J. (2000). Causal parameters and policy analysis in economics: a twentieth century retrospective. *Quarterly Journal of Economics* 115:45–97.

Heckman, J. J. (2005). The scientific model of causality. *Sociological Methodology* 35:1–98.

Heckman, J. J., Smith, J., Clemens, N. (1997). Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. *The Review of Economic Studies* 64:487–535.

Heckman, J. J., Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73:669–738.

Heckman, J. J., LaLonde, R. J., Smith, J. A. (1999). The economics and econometrics of active labor market programs. In: Ashenfelter, O., Card, D., eds. *Handbook of Labor Economics*. Vol. III A. Amsterdam: North-Holland, pp. 1865–2097.

Hernan, M. A., Brumback, B., Robins, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* 96:440–448.

Hicks, J. (1979). *Causality in Economics*. Oxford: Basil Blackwell.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* 81:945–970. With discussion.

Hosoya, Y. (1977). On the Granger condition for noncausality. *Econometrica* 45:1735–1736.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics* 86:4–29.

Imbens, G. W., Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* 62:446–475.

Lechner, M. (2009). Sequential causal models for the evaluation of labor market programs. *Journal of Business & Economic Statistics* 27:71–83.

Lechner, M. (2009). Matching estimation of dynamic treatment models: some practical issues. In: Millimet, D., Smith, J., Vytlacil, E., eds. *Advances in Econometrics Volume 21, Modelling and Evaluating Treatment Effects in Econometrics*, pp. 289–333.

Lechner, M., Miquel, R. (2005). Identification of Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions, Discussion Paper, Department of Economics, University of St. Gallen, 2002: Revised 2005.

Li, Q., Maasoumi, E., Racine, J. S. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics* 148(2):186–200.

Marshall, A. (1961). *Principles of Economics*. 9th Variorum ed. London: Macmillan.

Manski, C. F. (1997). Monotone treatment response. *Econometrica* 65:1311–1334.

Miquel, R. (2002). *Identification of Dynamic Treatments Effects by Instrumental Variables*. Department of Economics, University of St. Gallen.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. *Essay on Principles. Section 9: Translated in Statistical Science (with discussion)* 1990:5:465–480.

Pearl, J. (2000). *Causality – Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Rubin, D. B. (1974). Estimating Causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66:688–701.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics* 2:1–26.

Rubin, D. B. (1980). Discussion of 'randomization analysis of experimental data in the Fisher randomization test' by Basu, D. *Journal of the American Statistical Association* 75:591–593.

Rubin, D. B. (2005). Causal inference using potential outcomes: design, modelling, decisions. *Journal of the American Statistical Society* 100:322–331.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods – Application to control of the healthy worker survivor effect. *Mathematical Modelling* 7:1393–1512: With 1987 Errata to "A new approach to causal inference in mortality studies with sustained exposure periods – Application to control of the healthy worker survivor effect. *Computers and Mathematics with Applications* 14:917–921; 1987 Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods – Application to control of the healthy worker survivor effect. *Computers and Mathematics with Applications* 14:923–945; and 1987 Errata to "Addendum to 'A new approach to causal inference in mortality studies with sustained exposure periods – Application to control of the healthy worker survivor effect. *Computers and Mathematics with Applications* 18:477.

Robins, J. M. (1989). The analysis of randomized and nonrandomized aids treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest, L., Freeman, H., Mulley, A., eds. *Health Service Research Methodology: A Focus on Aids*. Washington, D.C.: Public Health Service, National Center for Health Services Research, pp. 113–159.

Robins, J. M. (1997). Causal inference from complex longitudinal data. Latent variable modeling and applications to causality. In: Berkane, M., ed. *Lecture Notes in Statistics*. Vol. 120. New York: Springer, pp. 69–117.

Robins, J. M. (1999). Association, causation, and marginal structural models. *Synthese* 121:151–179.

Robins, J. M. (2003). Commentary: general methodological considerations. *Journal of Econometrics* 112:89–106.

Robins, J. M., Greenland, S., Hu, F.-C. (1999a). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association* 94:687–700: With discussion and rejoinder, 708–712.

Robins, J. M., Rotnitzky, A., Scharfstein, D. O. (1999b). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran, M. E., Berry, D., eds. *Statistical Modeling in Epidemiology: The Environment and Clinical Trials*. New York: Springer, pp. 1–92.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3:135–146.

Simon, H. A. (1953). Causal ordering and identifiability. In: Hood, W. C., Koopmans, T. C., ed. *Studies in Econometric Method*. New York: Wiley, Chapter 3, pp. 49–74.

Simon, H. A. (1954). Spurious correlation: a causal interpretation. *Journal of the American Statistical Association* 49:467–479.

Sims, C. A. (1972). Money, income, and causality. *American Economic Review* 62:540–552.

Spirtes, P., Glymour, C., Scheines, R. (1993). *Causation, Prediction, and Search*. New York: Springer.

Wiener, N. (1956). The theory of prediction. In: Beckenham, E. F., ed. *Modern Mathematics for Engineers, Series 1*. New York: McGraw-Hill, Chapter 8.

Wilks, S. S. (1932). On the distribution of statistics in samples from a normal population of two variables with matched sampling of one variable. *Metron* 9:87–126.

White, H. (2006). Time-series estimation of the effects of natural experiments. *Journal of Econometrics* 135:527–566.