ANNUAL REVIEWS

Annual Review of Economics

The Econometric Model for Causal Policy Analysis

James J. Heckman[1] and Rodrigo Pinto[2]

[1]Department of Economics, University of Chicago, Chicago, Illinois, USA; email: jjh@uchicago.edu

[2]Department of Economics, University of California, Los Angeles, California, USA; email: rodrig@econ.ucla.edu

Annu. Rev. Econ. 2022. 14:893–923

First published as a Review in Advance on June 29, 2022

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-051520-015456

JEL codes: C10, C18

Keywords

policy analysis, econometric models, causality, identification, causal calculus, directed acyclic graphs, treatment effects

Abstract

This article discusses the econometric model of causal policy analysis and two alternative frameworks that are popular in statistics and computer science. By employing the alternative frameworks uncritically, economists ignore the substantial advantages of an econometric approach, and this results in less informative analyses of economic policy. We show that the econometric approach to causality enables economists to characterize and analyze a wider range of policy problems than is allowed by alternative approaches.

Erratum >

1. INTRODUCTION

Causal analysis is a key component of good policy evaluation. It examines the factors that lead to outcomes and the role played by the policy itself. It quantifies policy impacts. It provides information about the mechanisms that produce outcomes to help understand their operation, improve them, and determine if there are any other mechanisms that could be used to create outcomes. It gives a framework for collating evidence across studies. To provide good policy advice, it uses all the information available. It explores all possible counterfactual scenarios. It is grounded in thought experiments that show what could happen if certain outcomes are altered. Good policy analysis is science in this sense. Credible scenarios are created, analyzed, and tested using real-world data.

Economic analysis is based on models and thought experiments. These are used routinely by researchers who have been trained in economic theory and the hard sciences. Recently, computer scientists and statisticians have attempted to replace the causal framework used by economists. These alternative frameworks are best understood as approximations to economic causal analysis. They use novel languages and unconventional methodologies to formalize the notion of causality introduced by early economists such as Frisch (1930) and Haavelmo (1943).

The alternative frameworks exhibit the hallmark of their recent birth: Concepts are frequently imprecisely defined or incapable of addressing the entire body of causal inferences envisioned by the original founders of causal thought. Three distinct causal tasks of defining a causal model, establishing its identification, and empirically estimating it from the data are frequently intertwined. It is unsurprising that critical notions of causation are frequently shrouded in mystery and misunderstanding.

The current situation would be of little concern if applied economists continued to use and extend the fundamental notions of econometric causality. Unfortunately, this is not the case. Econometricians and applied economists increasingly emulate causal inference that neglects the groundwork in their own field, to the detriment of rigorous causal policy analysis.

Our goal is to improve the theory and practice of economic policy analysis by reacquainting economists with their own rich econometric legacy. We clarify the drawbacks of two recent causal frameworks that attempt to recast previous notions of econometric causality. We examine the well-known Neyman–Cox–Rubin–Holland model (Neyman 1923, Cox 1958, Rubin 1974, Holland 1986, Angrist et al. 1996), which is motivated by the notion of randomized control trials (RCTs), as an ideal framework for causality. We also investigate a less popular causal framework advocated by computer scientists, called do-calculus (DoC) (Pearl 1995, 2012), which makes extensive use of directed acyclic graphs (DAGs).[1]

Each of these approaches is well suited to investigating a subset of problems in causal inference. These approaches do, however, have significant limitations when applied to the wide variety of problems that economists face. Relying exclusively on these techniques limits important policy analyses. For instance, the Neyman–Rubin model (henceforth referred to as NR) does not incorporate unobservable variables. This limitation hinders the development of central concepts of policy evaluation. We demonstrate that the NR framework offers limited capabilities for assessing causal models outside of the narrow domain of the treatment-control paradigm. On the other hand, the DoC is incompatible with several well-established identification techniques commonly used in econometric evaluations. These frameworks hinder their users from accessing helpful components of the econometric tool set, including social interactions, peer effects, and general equilibrium theory.

[1] Readers are referred to Spiegler (2020) for a range of applications of DAGs in economics.

We adopt the Generalized Roy model as the leading example in our discussion. We contrast the versatility of the econometric approach to causality with the NR and the DoC approaches. We also discuss the class of inference problems for which these recent methodologies perform well.

The article is organized as follows. Section 2 defines causality. Section 3 presents the econometric model. Section 4 shows its versatility and describes various identification approaches in the Generalized Roy model. Section 5 examines how the NR causal model approximates the econometric model. Section 6 investigates how the DoC of Pearl (2009a) approximates the econometric model. Section 7 summarizes.

2. ECONOMETRIC CAUSALITY

The modern understanding of causal inference dates back to the 1930 lectures of Ragnar Frisch, who described causality as a thought experiment in which the researcher hypothetically manipulates one or more inputs affecting an output (Frisch 1930). According to Frisch, causality is not a physical property of the natural world, but rather an exercise of abstract manipulation of inputs causing an output.[2]

The notion of causality stems from two simple yet fundamental concepts: autonomy and directionality (Frisch 1995). Autonomy means that input-output relationships are invariant for different input manipulations.[3] It is also called structural (Hurwicz 1962). This view postulates that the map $g : X \to Y$ between input X and output Y remains stable as X ranges in its domain. Causality is directional. The map $g : X \to Y$ establishes a relationship in which X causes Y. Inverting the map g may generate a stable relationship of Y on X. However, such a relationship is usually devoid of causal meaning.

Consider a simple example based on the following autonomous equation:

$$Y = g(T, U). \tag{1.}$$

Equation 1 describes a causal model in which the input X consists of two variables (T, U) that cause an outcome of interest Y. In the classical setup of an RCT, U stands for an unobserved variable, while T is a treatment indicator that takes value $T = 1$ for treatment and $T = 0$ for control.

The causal effect of T on Y is described by the thought experiment that manipulates the values of the treatment variable T without affecting the variable U. This intervention is operationalized by the potential (or counterfactual) outcome generated by fixing the input T to a value t. In our model, the potential outcome is given by $Y(t) = g(t, U); t \in \{0, 1\}$, and the causal effect of the binary treatment T on the outcome Y is $Y(1) - Y(0)$. This simple example clarifies that causality does not require a probabilistic model. The definition of a causal effect precedes (and is conceptually distinct) from its identification or estimation.

2.1. Adding a Probabilistic Structure

Incorporating a probabilistic structure confers empirical relevance and improves interpretability. Let (U, T, Y) denote random variables in a probability space $(\mathcal{I}, \mathcal{F}, P)$. Each element $i \in \mathcal{I}$ represents an economic agent, and (U_i, T_i, Y_i) denotes the realized variables associated to agent i. Thus, the counterfactual outcome of agent i when the treatment T is fixed to the value $t \in \{0, 1\}$ is $Y_i(t) = g(t, U_i)$, and the treatment effect for agent i is given by $Y_i(1) - Y_i(0)$. The average treatment effect (ATE) is the expected value of treatment effects across the entire population $\mathcal{I}$,

[2] In his words, "causality is in the mind."

[3] Frisch (1995) describes autonomous equations as deterministic functions that are "invariant" to changes in their arguments. Hurwicz (1962) prefers the term "structural" to denote autonomous equations.

$ATE = E_{\mathcal{I}}(Y_i(1) - Y_i(0))$, which can also be expressed as the integral of the difference of counterfactuals over the support of the unobserved variable U. In this notation, we write

$$ATE = \int_{u \in \text{supp}(U)} E(Y(1) - Y(0)|U = u)dF(u) = \int_{u \in \text{supp}(U)} (g(1,u) - g(0,u))dF(u), \qquad 2.$$

where $F(u) = P(U \leq u)$ denotes the cumulative distribution function of variable U.

Variables T and U are called external (or exogenous) if they are not caused by any variable within the causal model. In this case, T and U are statistically independent—that is, $T \perp\!\!\!\perp U$—and the counterfactual outcome mean $E(Y(t))$ can be evaluated by the conditional outcome expectation $E(Y|T = t)$; we have

$$E(Y(t)) = E(g(t,U)) = E(g(t,U)|T = t) = E(Y|T = t); t \in \{0,1\}, \qquad 3.$$

where the second equality is due to $T \perp\!\!\!\perp U$. ATE is identified by the difference in means between treated and control participants.

Econometrics textbooks often discuss causality as a property of an estimator, usually ordinary least squares (OLS). This approach reverses the logic of causality. It also generates confusion, since the OLS model is described by statistical assumptions that are void of causality.[4] The causal content of the OLS model can only be assessed using causal notions such as fixing and counterfactuals, which are not part of the conventional statistical tool set.

The OLS model arises by assuming that function g in Equation 2 is linear, that is,

$$Y = \alpha + \beta T + U. \qquad 4.$$

The exogeneity of T and U implies that U and T do not cause each other. In this case, the linear assumption in Equation 4 imposes a constant treatment effect across individuals: $Y_i(1) - Y_i(0) = \beta$ for all $i \in \mathcal{I}$. Under exogeneity, U and T are statistically independent, and the OLS estimator for β is an unbiased estimator for the average treatment effect of T on Y. Thus, it is often said that the causal interpretation of the OLS estimator stems from the independence between T and U. This assessment blurs the logic of causality by conflating the definition of causal effects with their estimation from data. In fact, the question of whether the OLS estimator is biased is a statistical assessment that is separate from the causal inquiry of whether β in Equation 4 is the causal effect of T on Y. **Table 1** is useful in clarifying this difference.

Table 1 describes three causal relationships between U and T. Random variables ϵ_T, ϵ_U denote mutually statistically independent error terms that are not observed by the analyst. In Model 1, T and U are jointly caused by an unobserved confounding variable V. Parameter β in Equation 4 is still the causal effect of T on Y. However, U and V are not statistically independent, and the OLS estimator is biased. Model 2 differs from Model 1 in that U causes T. These models, however, are statistically equivalent in that T and U are not independent and β remains the causal effect of T on Y. In Model 3, treatment T causes U, and the parameter β in Equation 4 is not the causal effect of T on Y anymore.

Table 1 Examples of causal relationships between T and U

Model 1	Model 2	Model 3
$T = f_T(\epsilon_T, V)$	$T = f_T(\epsilon_T, U)$	$T = f_T(\epsilon_T)$
$U = f_U(\epsilon_U, V)$	$U = f_U(\epsilon_U)$	$U = f_U(\epsilon_U, T)$

[4]For an example of how confusing this concept is to statisticians, readers are referred to Pratt & Schlaifer (1984).

2.2. Why Do We Need a Causal Framework?

The mismatch between statistical and causal analyses can be traced back to the fact that statistics lacks directionality. In statistics, two random variables X, Y can be fully described by their joint distribution. Causality requires additional information that assigns the direction of the causal relationship between X and Y. This gap has been filled by causal frameworks that offer additional tools to investigate causal inquiries. These frameworks should enable the analyst to clearly distinguish three primary tasks of causal inference:

1. The first task is to precisely define a causal model that is grounded in scientific theory.
2. The second task regards the identification of causal parameters. The framework offers mathematical tools that enable the analyst to manipulate causal inputs and to investigate the identification of causal parameters.
3. The third task is the estimation and testing of identified causal parameters. This task lies within the realm of statistical analysis.

Furthermore, causal frameworks should be sufficiently adaptable to the wide range of policy questions investigated by economists. Examples of policy questions include the evaluation of models such as mediation analysis, simultaneous equation models, models of agent interactions, and collation of evidence across studies.

We make the case that recently developed causal frameworks lack the malleability needed to address the vast range of policy concerns investigated by economists. These frameworks embody Marschak's maxim (Heckman 2008a): They are extremely useful for addressing a specialized subset of policy questions faced by the economist but lack the necessary flexibility to address a wider range of problems. We now discuss these causal frameworks in greater detail.

3. ECONOMETRIC CAUSAL FRAMEWORK

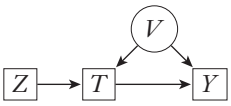
A causal model $\mathbb{M} : \mathcal{T} \to \mathbb{P}(\mathcal{T})$ is a system of structural equations defined by a mapping between a set of variables $\mathcal{T}$ and its power set. Structural equations are defined to be invariant maps from arguments to outcomes: Equations remain the same for any variations of arguments. Each variable K in a variable set $\mathcal{T}$ is mapped into a (potentially empty) subset $\mathbb{M}(K)$ contained in $\mathcal{T}$. Variables in $\mathbb{M}(K) \subset \mathcal{T}$ are said to directly cause $K \in \mathcal{T}$. Each variable $K \in \mathcal{T}$ is associated with an unobserved error term $\epsilon_K \in \mathcal{E}$. The structural equation of a variable $K \in \mathcal{T}$ is given by $K = f_K(\mathbb{M}(K), \epsilon_K)$.

The variable set $\mathcal{T}$ comprises random variables (or random vectors) that may be observed or unobserved by the analyst. Error terms in $\mathcal{E}$ are assumed to be mutually independent and externally (or exogenously) specified.[5] The error terms are not caused by any variable in $\mathcal{T}$. Similarly, a variable K is called exogenous when it is not caused by a variable in $\mathcal{T}$—that is, $\mathbb{M}(K) = \varnothing$ and $K = f_K(\epsilon_K)$. All variables are defined on a common probability space $(\mathcal{I}, \mathcal{F}, P)$.

We adopt the Generalized Roy model as our leading example. The model is a cornerstone in the literature of policy evaluation (Heckman & Taber 2008). The original model (Roy 1951) investigated the occupational choice of an economic agent that decides between two economic sectors based on the perceived difference of income between them. The model has been generalized to address a broad range of choice incentives and policy interventions that affect agents' decisions (Abbring & Heckman 2007; Heckman & Vytlacil 2007a,b). Those generalizations include psychological costs, price variations, tuition policies, and unobserved assessments of

[5]The independence among error terms comes without loss of generality, as any dependence structure could be modeled via other unobserved variables in $\mathcal{T}$.

Table 2 Representations of the Generalized Roy model

	Variable map	Structural equation	DAG	LMC
Z	$\mathbb{M}(Z) = \varnothing$	$Z = f_Z(\epsilon_Z)$		$Z \perp\!\!\!\perp V \mid \varnothing$
V	$\mathbb{M}(V) = \varnothing$	$V = f_V(\epsilon_V)$		$V \perp\!\!\!\perp Z \mid \varnothing$
T	$\mathbb{M}(T) = \{Z, V\}$	$T = f_T(Z, V, \epsilon_T)$		$T \perp\!\!\!\perp \varnothing \mid (Z, V)$
Y	$\mathbb{M}(Y) = \{T, V\}$	$Y = f_Y(T, V, \epsilon_Y)$		$Y \perp\!\!\!\perp Z \mid (T, V)$

Abbreviations: DAG, directed acyclic graph; LMC, local Markov condition.

choice benefits. The model has been widely applied to examine the causal effects of a variety of choice settings[6] and policy interventions.[7]

The simplest representation of the Generalized Roy model comprises four random variables $\mathcal{T} = \{Z, V, T, Y\}$, where Z is an instrumental variable (IV) that causes an outcome Y only through its effects on a treatment choice T. The variable V denotes an exogenous and unobserved confounding variable that causes both T and Y.[8] In the context of the Generalized Roy model, Z stands for an external policy vector. The confounding variable V may include the subjective evaluation of choice benefits among heterogeneous agents. The confounder is a source of selection bias as it induces a covariation between choice T and outcome Y that is not due to the causal effect of T on Y. **Table 2** displays four equivalent representations of the Generalized Roy model.

The first column of **Table 2** lists the variables of the Roy model. The second column describes the causal model as a mapping of the variable set. The third column displays the corresponding structural equations. The fourth column displays the model as a DAG, where arrows denote causal relationships, circles denote unobserved variables, and squares denote observed variables.[9]

The last column of **Table 2** describes a property called the local Markov condition (LMC).[10] Some notation is necessary to state the condition. The language of Bayesian networks uses the term parents of K for the variables that directly cause K, that is, $\mathbb{M}(K)$. The children of K comprise the variables directly caused by K, namely, $\mathbb{C}(K) = \{J \in \mathcal{T};\ K \in \mathbb{M}(J)\}$. The descendants of a variable K, $\mathbb{D}(K)$, include all the variables that are directly or indirectly caused by K. These include all the subsequent iterations of the children of K.[11] A causal model is recursive (acyclic) if no variable is a descendant of itself. The LMC is a property of recursive models stating that a variable is independent of its nondescendants conditioned on its parents, that is,

$$\text{LMC:} \quad K \perp\!\!\!\perp (\mathcal{T} \setminus \mathbb{D}(K)) \mid \mathbb{M}(K). \tag{5.}$$

For instance, outcome Y has no descendants, and its parents are $\{V, T\}$. Thus, its LMC is $Y \perp\!\!\!\perp Z \mid (T, V)$, as listed in the bottom row of **Table 2**. Z has no parents, and its descendants are T, Y. The

[6]For instance, Heckman & Vytlacil (2007a) investigate multiple variations of the original model, Heckman et al. (2008) extend the model for ordered and unordered choice models, and Heckman & Pinto (2018) and Lee & Salanié (2018) investigate the case of unordered multiple choice models with multi-valued treatments. Abbring & Heckman (2007) consider dynamic discrete choice models in this framework.

[7]For example, the model has been used to evaluate schooling choices, migration decisions, criminal behaviors, neighborhood choices, and early childhood interventions.

[8]Choice T may be binary, discrete, or continuous, and the confounder variable V can denote a random vector of arbitrary dimension.

[9]We refer the readers to Spiegler (2020) and Lauritzen (1996) for information on DAGs and Bayesian networks.

[10]Kiiveri et al. (1984) and Pearl (1988) provide further information on the local Markov condition.

[11]Notationally, for any subset $\tilde{\mathcal{T}} \subset \mathcal{T}$, let $\mathbb{C}(\tilde{\mathcal{T}})$ be the union of the children of all the variables in $\tilde{\mathcal{T}}$, that is, $\mathbb{C}(\tilde{\mathcal{T}}) = \cup_{K \in \tilde{\mathcal{T}}} \mathbb{C}(K)$. The descendants of K is the smallest set $\mathbb{D}(K) \subset \mathcal{T}$ that contains the children of K, $\mathbb{C}(K) \subset \mathbb{D}(K)$, and its own children, $\mathbb{C}(\mathbb{D}(K)) = \mathbb{D}(K)$.

set of LMCs for all variables in $\mathcal{T}$ fully characterizes the causal model. Additional independence relationships may be generated by the graphoid axioms of Dawid (1976). Graphoid axioms consist of six rules that apply for any disjoint sets of variables $X, W, Z, Y \subseteq \mathcal{T}$:

1. Symmetry: $X \perp\!\!\!\perp Y \mid Z \Rightarrow Y \perp\!\!\!\perp X \mid Z.$
2. Decomposition: $X \perp\!\!\!\perp (W, Y) \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid Z.$
3. Weak union: $X \perp\!\!\!\perp (W, Y) \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid (W, Z).$
4. Contraction: $X \perp\!\!\!\perp W \mid (Y, Z)$ and $X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp (W, Y) \mid Z.$
5. Intersection: $X \perp\!\!\!\perp W \mid (Y, Z)$ and $X \perp\!\!\!\perp Y \mid (W, Z) \Rightarrow X \perp\!\!\!\perp (W, Y) \mid Z.$
6. Redundancy: $X \perp\!\!\!\perp Y | Z \Rightarrow X \perp\!\!\!\perp Y | Z.$

It is worth noting that the merits of DAGs and Bayesian networks are not limited to causal inference and probabilistic analysis. A clever and original use of Bayesian networks is proposed by Spiegler (2020), who employs DAGs to model causal misconceptions in decision making and demonstrates their far-reaching implications for economic behavior.

3.1. Counterfactual Approaches

We obtain a counterfactual outcome by the hypothetical (external) manipulation of the targeted variable that causes the outcome of interest. This causal manipulation is accomplished through the use of the fix operator, introduced in the seminal work of Haavelmo (1943). In the Roy model, the counterfactual outcome $Y(t)$ is obtained by fixing the T-input of the outcome equation to a value $t \in \text{supp}(T)$ so that $Y(t) = f_Y(t, V, \epsilon_Y)$.

Fixing only affects the outcome equation. It substitutes the treatment random variable T by the treatment value t. It does not eliminate the treatment variable T from the causal model nor modifies the choice equation $T = f_T(Z, V, \epsilon_T)$.

The do operator of Pearl (1995, 2012) is similar to fixing as it substitutes all T-inputs from the structural equations of the variables directly caused by T. The do operator differs from fixing as it deletes ("shuts down") the structural equation for the treatment variable T, which effectively suppresses the random variable T from the causal model.

Neither the fix nor the do operator is well defined in statistics. These are causal operators that affect only the distribution of the descendants of the variable being fixed. In contrast, statistical conditioning affects the distribution of all variables that are not statistically independent of the conditioning variable. Fixing T in the Roy model affects the outcome Y but does not impact the confounder V or the instrument Z, which remain statistically independent. Conditioning on T, on the other hand, alters the distributions of Z and V, which are no longer statistically independent.

Heckman & Pinto (2015) develop a causal framework that expresses the ill-defined causal operations of fixing or doing in terms of standard statistical conditioning. They distinguish the empirical model that generates observable data from a hypothetical model that is used to formulate the thought experiments involving the manipulation of inputs that determine causality.

The hypothetical model formalizes Frisch's insights on causality. It is an abstract model that shares the same structural equations and the same distributions of error terms as the empirical model. It differs from the empirical model in that it appends a hypothetical variable $\tilde{T}$ which replaces the T-input of variables directly caused by T. The hypothetical variable captures the causal notion of an exogenous manipulation of the treatment. The hypothetical model translates the causal operation of fixing T into the statistical operation of conditioning on $\tilde{T}$.

We illustrate the hypothetical model using the Generalized Roy model. For notational clarity, we use $\mathbb{M}_{\text{e}}$ for the empirical (original) model, $\mathbb{M}_{\text{fix}}$ for the model that applies the fix operator, $\mathbb{M}_{\text{do}}$ for the model using the do operator, and $\mathbb{M}_{\text{h}}$ for the hypothetical model. We also use the subscripts e, fix, do, and h for the probability distributions and expectations associated with each model. **Table 3** displays the Roy model for each of these settings.

The second column of **Table 3** presents the original empirical model. The third and fourth columns present the models generated by the fix and the do operators, respectively. Both models replace the T-input of the outcome equation by a value $t \in \text{supp}(T)$. The main difference between these models is that fix retains the treatment variable, whereas do suppresses it. The hypothetical model is displayed in the last column of **Table 3**. It replaces the T-input of the outcome equation with an external hypothetical variable $\tilde{T}$.

All models share the same distributions of error terms ϵ_Z, ϵ_T, ϵ_V, and ϵ_Y. Therefore, the joint distribution of nondescendant T, that is, (V, Z), is the same across the models. The structural equation for the counterfactual outcome $Y(t)$ in the fix or do model depends only on V and ϵ_Y, and thus the models share the same distribution of $Y(t)$.

The hypothetical variable $\tilde{T}$ enables us to circumvent the necessity of introducing a causal operator. The variable has no parents and, according to the LMC of Equation 5, it is independent of all its nondescendants, $\tilde{T} \perp\!\!\!\perp (T, V, Z)$. In particular, $\tilde{T} \perp\!\!\!\perp T$ always holds for any hypothetical model. $\tilde{T}$ is also statistically independent of error terms, as $\epsilon_{\tilde{T}} \perp\!\!\!\perp (\epsilon_Z, \epsilon_T, \epsilon_V, \epsilon_Y)$. The counterfactual outcome is obtained by simply conditioning on $\tilde{T}$. In summary, we have that

$$\Big(Y \mid \tilde{T} = t\Big)_{\mathbb{M}_{\text{h}}} \overset{d}{=} \Big(Y(t)\Big)_{\mathbb{M}_{\text{fix}}} \overset{d}{=} \Big(Y(t)\Big)_{\mathbb{M}_{\text{do}}}. \tag{6.}$$

It is also the case that Equation 6 holds when conditioned on any variable K that is a nondescendant variable of $\tilde{T}$:

$$\Big(Y \mid \tilde{T} = t, K\Big)_{\mathbb{M}_{\text{h}}} \overset{d}{=} \Big(Y(t) \mid K\Big)_{\mathbb{M}_{\text{fix}}}, \tag{7.}$$

$$\Big(Y \mid \tilde{T} = t, \{K \setminus \{T\}\}\Big)_{\mathbb{M}_{\text{h}}} \overset{d}{=} \Big(Y(t) \mid \{K \setminus \{T\}\}\Big)_{\mathbb{M}_{\text{do}}}. \tag{8.}$$

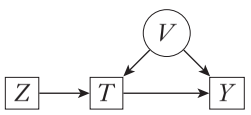
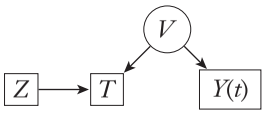
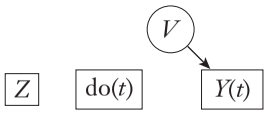
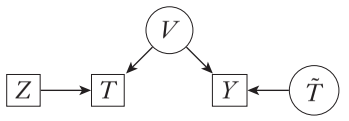
To clarify this, let T be an indicator of college graduation and let Y denote adult income. Treatment-on-the-treated (TT) is the average causal effect of college on income by those who choose to go to college ($T = 1$), which is commonly described as $TT = E_{\text{fix}}(Y(1) - Y(0)|T = 1)$ using the fix operator. The parameter is equivalently described as $TT = E_{\text{h}}(Y \mid \tilde{T} = 1, T = 1) - E_{\text{h}}(Y \mid \tilde{T} = 0, T = 1)$ using the hypothetical model. The do operator excludes the treatment variable T, which poses a serious challenge in defining the TT parameter. Shpitser & Pearl (2009) solve this issue by appending special structure to the counterfactual model.

Equation 6 suggests that the way that counterfactuals are expressed is of little relevance in the study of causality. That assessment is misleading. Small differences in characterizing counterfactuals have significant consequences for the machinery used to identify causal effects. Section 6 illustrates the difference between the identification approach using the DoC developed around the do operator and the identification approach using the hypothetical model framework.

3.2. Identification of the Counterfactual Outcome

Counterfactuals are said to be identified if they can be expressed in terms of the observed data generated by the empirical model $\mathbb{M}_{\text{e}}$. This task requires us to connect the probability distribution (or expectation) of counterfactual variables with the observed distributions of the empirical model. There are several ways to connect counterfactual variables with the empirical model.

Table 3 Generalized Roy model: approaches to generating counterfactuals

	Empirical models			Hypothetical model
Variables	**Empirical model ($\mathbb{M}_e$)**	**Fixing T at t ($\mathbb{M}_{fix}$)**	**Do-ing do(t) ($\mathbb{M}_{do}$)**	**Hypothetical model ($\mathbb{M}_h$)**
	Structural equations			
V:	$V = f_V(\epsilon_V)$	$V = f_V(\epsilon_V)$	$V = f_V(\epsilon_V)$	$V = f_V(\epsilon_V)$
Z:	$Z = f_Z(\epsilon_Z)$	$Z = f_Z(\epsilon_Z)$	$Z = f_Z(\epsilon_Z)$	$Z = f_Z(\epsilon_Z)$
T:	$T = f_T(Z, V, \epsilon_T)$	$T = f_T(Z, V, \epsilon_T)$	$\text{do}(T = t)$	$T = f_T(Z, V, \epsilon_T)$
Y:	$Y = f_Y(T, V, \epsilon_Y)$	$Y(t) = f_Y(t, V, \epsilon_Y)$	$Y(t) = f_Y(t, V, \epsilon_Y)$	$Y = f_Y(\tilde{T}, V, \epsilon_Y)$
$\tilde{T}$:	(not defined for the model)	(not defined for the model)	(not defined for the model)	$\tilde{T} = f_{\tilde{T}}(\epsilon_{\tilde{T}})$
	Directed acyclic graphs			
				
	Local Markov conditions			
V:	$V \perp\!\!\!\perp Z$	$V \perp\!\!\!\perp Z$	$V \perp\!\!\!\perp Z$	$V \perp\!\!\!\perp (Z, \tilde{T})$
Z:	$Z \perp\!\!\!\perp V$	$Z \perp\!\!\!\perp (V, Y(t))$	$Z \perp\!\!\!\perp (V, Y(t))$	$Z \perp\!\!\!\perp (V, Y, \tilde{T})$
T:	$T \perp\!\!\!\perp \varnothing \mid (Z, V)$	$T \perp\!\!\!\perp Y(t) \mid (Z, V)$	(not defined for the model)	$T \perp\!\!\!\perp (\tilde{T}, Y) \mid (Z, V)$
Y:	$Y \perp\!\!\!\perp Z \mid (T, V)$	$Y(t) \perp\!\!\!\perp (Z, T) \mid V$	$Y(t) \perp\!\!\!\perp Z \mid V$	$Y \perp\!\!\!\perp (Z, T) \mid (\tilde{T}, V)$
$\tilde{T}$:	(not defined for the model)	(not defined for the model)	(not defined for the model)	$\tilde{T} \perp\!\!\!\perp (T, V, Z)$
	Factorial decomposition of the joint probability distributions			
	$P_e(Y, T, V, Z) =$ $P_e(Y \mid T, V) P_e(T \mid Z, V) P_e(V) P_e(Z)$	$P_{fix}(Y(t), T, V, Z) =$ $P_{fix}(Y(t) \mid V) P_{fix}(T \mid V, Z) P_{fix}(V) P_{fix}(Z)$	$P_{do}(Y(t), V, Z) =$ $P_{do}(Y(t) \mid V) P_{do}(V) P_{do}(Z)$	$P_h(Z, V, T, \tilde{T}, Y) =$ $P_h(Y \mid \tilde{T}, V) P_h(T \mid Z, V) P_h(V) P_h(Z) P_h(\tilde{T})$

Subscript "do" denotes the model that employs the do operator. Subscript "e" denotes empirical (original) model. Subscript "fix" denotes the model that uses the fix operator—that is, treatment T is fixed to t. Subscript "h" denotes the hypothetical model.

We first consider the fix operator of model $\mathbb{M}_{\text{fix}}$ in **Table 3**. The LMC of $Y(t)$ in $\mathbb{M}_{\text{fix}}$ implies that

$$Y(t) \perp\!\!\!\perp T | V. \tag{9.}$$

Equation 9 states that the counterfactual outcome $Y(t)$ is independent of the treatment variable T conditional on the confounding variable V. This relationship is an example of a matching condition. It plays a primary role in the identification of treatment effects, as it enables us to connect the counterfactual outcome $Y(t)$ in $\mathbb{M}_{\text{fix}}$ with the empirical model $\mathbb{M}_{\text{e}}$:

$$\begin{aligned} P_{\text{fix}}\big(Y(t) \mid V\big) &= P_{\text{fix}}\big(Y(t) \mid V, T = t\big), && 10.\\ &= P_{\text{fix}}\left(\sum_{t \in \text{supp}(T)} \mathbf{1}[T = t] Y(t) \mid V, T = t\right), && 11.\\ &= P_{\text{fix}}\left(\sum_{t \in \text{supp}(T)} \mathbf{1}[T = t] f_Y(t, V, \epsilon_Y) \mid V, T = t\right), && 12.\\ &= P_{\text{fix}}\big(f_Y(T, V, \epsilon_Y) \mid V, T = t\big), && 13.\\ &= P_{\text{e}}\big(Y \mid V, T = t\big). && 14. \end{aligned}$$

Equations 10–14 use the structural equations to express the probability distribution of the counterfactual outcome $Y(t)$ in $\mathbb{M}_{\text{fix}}$ with the distribution of the outcome Y in empirical model $\mathbb{M}_{\text{e}}$. Equation 10 is a consequence of the matching condition in Equation 9. Equations 10–13 apply the definitions of the structural equations. Equation 14 uses the fact that variables T, V, ϵ_Y share the same distribution in both models $\mathbb{M}_{\text{fix}}$ and $\mathbb{M}_{\text{e}}$.

The hypothetical model $\mathbb{M}_{\text{h}}$ offers criteria that enable us to connect the counterfactual and empirical distributions in a systematic manner. For any disjoint set of variables $Y, W \in \mathcal{T}$ and any values $t, t' \in \text{supp}(T)$, we have that[12]

$$Y \perp\!\!\!\perp \widetilde{T} \mid (T, W) \Rightarrow P_{\text{h}}\Big(Y \mid \widetilde{T} = t, T = t', W\Big) = P_{\text{h}}(Y \mid T = t', W) = P_{\text{e}}(Y \mid T = t', W), \tag{15.}$$

$$Y \perp\!\!\!\perp T \mid (\widetilde{T}, W) \Rightarrow P_{\text{h}}\Big(Y \mid \widetilde{T} = t, T = t', W\Big) = P_{\text{h}}(Y \mid \widetilde{T} = t, W) = P_{\text{e}}(Y \mid T = t, W). \tag{16.}$$

Equations 15 and 16 state that we can switch from the hypothetical to the empirical model whenever the independence relationships in Equation 15, $Y \perp\!\!\!\perp \widetilde{T} \mid (T, W)$, or Equation 16, $Y \perp\!\!\!\perp T \mid (\widetilde{T}, W)$, hold.[13] The LMC of Y in $\mathbb{M}_{\text{h}}$ generates the following matching condition:

$$Y \perp\!\!\!\perp T | (\tilde{T}, V). \tag{17.}$$

Thus, according to Equation 16, we have that $P_{\text{h}}(Y \mid \widetilde{T} = t, V) = P_{\text{e}}(Y \mid T = t, V)$.

The hypothetical framework gives a systematic approach for connecting hypothetical and empirical models. The framework employs additional structure beyond what is obtained from fixing that might not be justified in the simple case of the Roy model. Section 6 explores more elaborate models where the additional complexity of the hypothetical framework is warranted.

[12] Readers are referred to Heckman & Pinto (2015) for a proof. The criteria in Equations 15 and 16 would still hold if the values $t, t' \in \text{supp}(T)$ were replaced by subsets $\mathcal{A}, \mathcal{A}' \subset \text{supp}(T)$, respectively.

[13] Heckman & Pinto (2015) offer a discussion of the connection between empirical and hypothetical models.

The do operator does not generate matching conditions such as Equation 9 or 17 because the treatment T is absent. Instead, the DoC of Pearl (2009a) checks for matching conditions using a DAG-based analysis called the back-door criterion (Pearl 1993). The method employs special jargon that is obscure for most economists. The criterion is part of the DoC, which consists of a set of DAG-oriented techniques that enables us to systematically examine the identification of causal effects. The method is general in the sense that it applies to any DAG, but it is limited in the sense that it does not accept identifying assumptions outside the DAG terminology. We discuss the DoC machinery and its benefits and limitations in Section 6. For now, we use the back-door criterion to introduce some of the DoC nomenclature.

Let G be the original DAG that represents the empirical model, and let $G_{\underline{T}}$ be the DAG that suppresses the arrows departing from T. In the DoC, the matching condition is expressed by the statement "V d-separates Y and T in the DAG $G_{\underline{T}}$," where the d-separation is a DAG criterion that checks for conditional independence among variables. Namely, let U be a path of arrows that connects variables T and Y in a DAG G regardless of the arrows' directions. A collider C in path U is a variable that has two arrows pointing at it (inverted fork). A variable V in the path U is said to block T and Y in the DAG G if it is not a collider (nor a descendant of a collider). T and Y are said to be d-separated by a set of variables V if V d-separates all paths from T to Y. The back-door criterion holds for confounder V in the Roy model of **Table 3**.

The counterfactual models $\mathbb{M}_{\text{fix}}$, $\mathbb{M}_{\text{h}}$, and $\mathbb{M}_{\text{do}}$ employ distinct techniques to generate the same conclusion: The identification of the counterfactual outcome requires analysts to control for the confounding variable V. In summary, we have that

$$P_{\text{fix}}(Y(t) \mid V) = P_{\text{h}}(Y \mid \widetilde{T} = t, V) = P_{\text{do}}(Y(t)|V) = P_{\text{e}}(Y \mid T = t, V). \qquad 18.$$

If V were observed, we would be able to evaluate the expected value of the counterfactual outcome expectation, $E_{\text{h}}(Y \mid \widetilde{T} = t)$, by integrating the observed expectation $E_{\text{e}}(Y|T = t, V)$ over the support of V. The econometric literature provides a rich menu of strategies to control for the confounding variable V. We discuss some of these strategies in the next section.

4. SOME IDENTIFICATION STRATEGIES

Section 3.1 explains that the identification of causal effects in the Generalized Roy model hinges on controlling for the confounding variable V. Unfortunately, V is not observed, and the Generalized Roy model is not identified without additional assumptions.

The literature on econometric policy evaluation offers a vast menu of assumptions to aid in identifying the causal effect of T on Y. Our discussion is unavoidably brief due to space constraints; we mention only a limited share of the identification strategies of this literature.

4.1. Matching on Observables

A popular method for identifying treatment effects is to assume that a set of observed pretreatment variables X suffices to control for the confounding variable V.[14] Otherwise stated, it assumes that the observed variable X is a balancing score for the confounding variable V. This assumption is often referred to as matching on observables, unconfoundness, ignorability, or exogeneity assumption (Rosenbaum & Rubin 1983, Heckman et al. 1998, Imbens 2004).

Figure 1 presents the empirical and hypothetical models corresponding to matching on observables. The LMC of T in the hypothetical model implies that $Y \perp\!\!\!\perp T \mid (\tilde{T}, X)$ or, equivalently,

[14]By pretreatment variables X we mean variables that are not descendants of the treatment variable T.

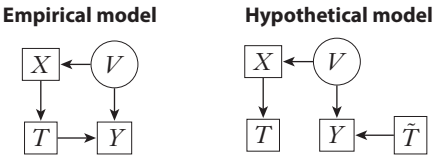


Figure 1

Matching model: empirical and hypothetical causal models.

$Y(t) \perp\!\!\!\perp T \mid X$. This matching condition enables us to identify the counterfactual outcome by conditioning on X: $P_{\mathrm{h}}(Y \mid \tilde{T} = t, X = x) = P_{\mathrm{fix}}(Y(t) \mid X = x) = P_{\mathrm{e}}(Y \mid T = t, X = x)$. The common support (overlap) assumption—$0 < P_{\mathrm{e}}(T = t \mid X = x) < 1; t \in \mathrm{supp}(T)$—enables us to identify the expected value of the counterfactual outcome as $E_{\mathrm{fix}}(Y(t)) = \int E_{\mathrm{e}}(Y \mid T = t, X = x) dF_{\mathrm{e},X}(x)$.[15]

Matching on observables solves the problem of selection bias induced by unobserved variables V via conditioning on the observed variables X. The assumption is justified in the case of RCTs, where X denotes the pretreatment variables used in the randomization protocol. The assumption is rather strong in observational studies and is often criticized (Heckman & Navarro 2004). It assumes that potential bias generated by confounding variables can be ignored when controlling for observed pretreatment variables (Heckman 2008b). Matching on observables does not require using the IV to identify causal effects. Instead, it solves the problem of selection bias by assuming that the bias does not exist or is of minor concern. Other strategies are more appealing.

4.2. Instrumental Variables

The simplest identifying assumption in IV models is that the functional form of outcome and choice equations is linear. The causal effect of T on Y is then identified by the covariance ratio $\mathrm{cov}(Y, Z)/\mathrm{cov}(T, Z)$ and can be estimated by the two-stage least squares (2SLS) regression (Theil 1953).

The linearity assumption imposes the undesirable feature of constant treatment effects, which rules out the possibility of essential heterogeneity across agents (Heckman et al. 2006). The local instrumental variable (LIV) model of Heckman & Vytlacil (1999, 2005) addresses this problem.

The LIV model considers a binary treatment $T \in \{0, 1\}$ determined by a latent threshold-crossing equation that is separable in the instrument Z and the confounder V, that is, $T = \mathbf{1}[\zeta(Z) \geq \phi(V)]$. This separability assumption arises from economic choice theory and enables the authors to rewrite the choice equation as

$$T = \mathbf{1}\big[P(Z) \geq U\big], \tag{19.}$$

where $P(Z) = P_{\mathrm{e}}(T = 1 \mid Z)$ is the propensity score and the unobserved variable U is given by $U = F_{\mathrm{e},\phi(V)}(\phi(V))$, where $F_{\mathrm{e},\phi(V)}$ is the cumulative distribution function of $\phi(V)$ and the subscript "e" denotes the empirical model. U has a uniform distribution for absolutely continuous $\phi(V)$, that is, $U \sim \mathrm{unif}[0, 1]$.

Figure 2 displays the empirical and hypothetical models of LIV. Note that $Y \perp\!\!\!\perp T \mid (\tilde{T}, U)$ holds.[16] Thus, U is a matching variable and plays the role of the confounding variable V in Equation 17. U is called a balancing score for V, which means that U is a surjective function of V

[15] Readers are referred to Heckman et al. (1998) and Imbens (2004) for estimation methods invoking matching on observables.

[16] The LMC of T implies $Y \perp\!\!\!\perp T \mid (Z, \tilde{T}, U)$, and the LMC of Z implies $Y \perp\!\!\!\perp Z \mid (U, \tilde{T})$, which implies $Y \perp\!\!\!\perp T \mid (\tilde{T}, U)$ by contraction.

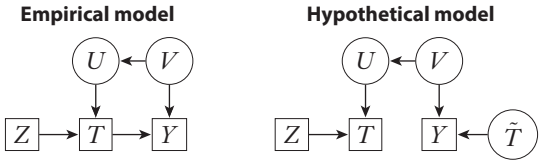


Figure 2

Empirical and hypothetical causal models for local instrumental variables.

that preserves the independence relationship $Y \perp\!\!\!\perp T \mid (\tilde{T}, V) \Rightarrow Y \perp\!\!\!\perp T \mid (\tilde{T}, U)$.[17] Controlling for U identifies counterfactual outcomes in the same fashion that controlling for V in Equation 18 does; that is, we have $P_{\mathrm{h}}(Y \mid \widetilde{T} = t, U) = P_{\mathrm{fix}}(Y(t) \mid U) = P_{\mathrm{e}}(Y \mid T = t, U)$.

Heckman & Vytlacil (1999) show that the separability assumption enables analysts to identify counterfactual outcomes conditioned on $U = u$ by differentiating the outcome with respect to the propensity score $P(Z)$ at value $u \in (0, 1)$:

$$E_{\mathrm{h}}(Y \mid \tilde{T} = 1, U = u) - E_{\mathrm{h}}(Y \mid \tilde{T} = 0, U = u) = E_{\mathrm{fix}}(Y(1) - Y(0) \mid U = u) = \left. \frac{\partial E_e(Y \mid P(Z))}{\partial P(Z)} \right|_{P(Z)=u}. \quad 20.$$

This identification strategy requires sufficient variation of the propensity score $P(Z)$ around $u \in [0, 1]$. If $P(Z)$ has full support, the average treatment effect can be evaluated by $ATE \equiv E_{\mathrm{h}}(Y \mid \tilde{T} = 1) - E_{\mathrm{h}}(Y \mid \tilde{T} = 0) = \int_0^1 \big(E_{\mathrm{h}}(Y \mid T = 1, U = u) - E_{\mathrm{h}}(Y \mid T = 0, U = u)\big) du$. For categorical instruments, the discrete counterpart of Equation 20 states that for any two values $z, z' \in \mathrm{supp}(Z)$ such that $P(z') = u' > u = P(z)$, we have that

$$\frac{E_{\mathrm{e}}(Y \mid Z = z') - E_{\mathrm{e}}(Y \mid Z = z)}{P_{\mathrm{e}}(T = 1 \mid Z = z') - P_{\mathrm{e}}(T = 1 \mid Z = z)} = \frac{\int_u^{u'} E_{\mathrm{fix}}(Y(1) - Y(0) \mid U = u) du}{u' - u} \quad 21.$$

$$= E_{\mathrm{fix}}(Y(1) - Y(0) \mid u \le U \le u'). \quad 22.$$

Equation 21 explains that the local average treatment effect (LATE) of Imbens & Angrist (1994) (left-hand side) identifies the expected value of the counterfactual outcome $E_{\mathrm{fix}}(Y(1) - Y(0) \mid U = u)$ over the interval of U defined by the propensity scores $P(z)$ and $P(z')$ in Equation 22.[18] The method of RCTs can be understood as a particular case of the LIV model in which the IV values z_0, z_1 induce full compliance: $P(z_0) = 0$ and $P(z_1) = 1$. In this case, Equations 21 and 22 identify the average treatment effect:

$$\frac{E_{\mathrm{e}}(Y \mid Z = z_1) - E_{\mathrm{e}}(Y \mid Z = z_0)}{1 - 0} = E_{\mathrm{fix}}(Y(1) - Y(0) \mid 0 \le U \le 1) = ATE.$$

4.3. Stratification

The stratification method is useful for controlling the confounding effects of V in the case of discrete instruments ($Z \in \{z_1, \ldots, z_{N_Z}\}$) and multiple choices ($T \in \{t_1, \ldots, t_{N_T}\}$). The method employs a response vector $\mathbf{S}$ that stands for the N_Z-dimensional random vector of counterfactual choices

[17] The term balancing score was introduced by Rosenbaum & Rubin (1983).

[18] Heckman et al. (2008) investigate the relationship between LIV and LATE in greater detail. Mogstad & Torgovitsky (2018) use functional form assumptions to extrapolate the evaluation of LATE parameters beyond the U-interval defined by propensity scores.

across all Z values:

$$\mathbf{S} = [T(z_1), \ldots, T(z_{N_Z})]'. \tag{23.}$$

The values that the response vector $\mathbf{S}$ takes are called response-types.[19] In the case of a binary instrument $Z \in \{z_0, z_1\}$ and a binary treatment $T \in \{0, 1\}$, the response vector $\mathbf{S} = [T(z_0), T(z_1)]'$ takes four possible response-types, $\text{supp}(\mathbf{S}) = \{(0,0), (0,1), (1,0), (1,1)\}$, which Imbens & Angrist (1994) term always-takers, compliers, never-takers, and deniers, respectively.

The response vector $\mathbf{S}$ is a balancing score for V in the same fashion that U in Equation 19 is. $\mathbf{S}$ is a function of V because each counterfactual choice $T(z) = f_T(z, V)$ is a function of V itself. Moreover, choice T can be expressed as $T = \left[\mathbf{1}[Z = z_1], \ldots, \mathbf{1}[Z = z_N]\right] \cdot \mathbf{S}$. Thus, given $\mathbf{S}$, the choice T is a function of only Z, which is independent of V and $Y(t)$. Therefore, $Y(t) \perp\!\!\!\perp T | \mathbf{S}$ holds.

Heckman & Pinto (2018) show that the response vector $\mathbf{S}$ generates a special partition of the support of V that renders T statistically independent of V within each partition set. They present the necessary and sufficient conditions for the identification of counterfactual outcomes, LATE-type parameters, and response-type probabilities.

The main difficulty in identifying causal parameters is that the number of possible response-types usually exceeds the number of linear restrictions generated by the observed data. Indeed, while the number of response-types grows exponentially in N_Z, observed data grow linearly in N_Z. Identification requires choice restrictions that systematically reduce the number of admissible response-types. Such restrictions can be obtained through monotonicity conditions or by applying revealed preference analysis (Pinto 2016, Heckman & Pinto 2018, Buchinsky & Pinto 2021).

4.4. Matching on Unobservables

Matching on unobservables is a version of matching that uses observed data to control for an unobserved variable Q that has two properties: (*a*) Q is a balancing score for the confounder V, and (*b*) Q may be measured with error by the observed variables M. **Figure 3** displays the empirical and hypothetical models corresponding to matching on unobservables. The matching condition $Y \perp\!\!\!\perp T | (\tilde{T}, Q)$ holds, and Q is a balancing score for V.

The identification strategy is to explore the relationship between Q and its measurements M in order to evaluate a proxy for Q that enables us to control for V. Formally, the method explores the structural equation $M = f_M(Q, \epsilon_M)$ to estimate Q.

Matching on unobservables has long been used in the economics of education (see, e.g., Duncan & Goldberger 1973, Goldberger 1972). The method is called the latent variable approach by Heckman & Robb (1985a). This literature offers several possibilities for estimating Q (Aakvik et al. 1999, 2005; Carneiro et al. 2001, 2003; Cunha et al. 2005). We refer readers to Cunha et al. (2010) and Schennach (2020) for discussion of nonparametric identification and estimation of Q.

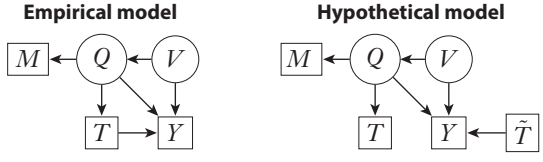


Figure 3

Matching on proxied unobservables: empirical and hypothetical causal models.

[19]The concept of a response variable was developed by Robins (1986) and further studied by several researchers. Frangakis & Rubin (2002) use the term principal strata, whereas Balke & Pearl (1993) use the term response variables.

An example in this setting consists of the evaluation of college returns, where T denotes college graduation, Y denotes earnings, and Q stands for unobserved abilities such as cognition and personality traits such as conscientiousness, neuroticism, and agreeableness. These abilities are not directly observed but are measured with error by M via psychological surveys or test scores. A common parametric approach evaluates a factor model described by structural equations $M = f_M(Q, \epsilon_M)$, where Q denotes a latent factor. A proxy for Q is estimated as the factor score of the model (see, e.g., Heckman et al. 2013).

4.5. Control Function

The control function approach was introduced by Heckman & Robb (1985b), building on earlier work by Telser (1964), and has been examined by several authors (Ahn & Powell 1993, Powell 1994, Blundell & Powell 2003, Wooldridge 2015). The approach includes a broad class of techniques that control for endogeneity issues generated by unobserved variables in the outcome equation. A simple example of a control function approach that relies on parametric assumptions is Heckman's sample selection correction (Heckman 1979). For a broader survey of identification of nonlinear models with endogenous variables, readers are referred to Matzkin (1994, 2007).

We illustrate the control function approach using a version of the Roy model with binary choices that relies on the additive separability of the outcome equation. We can express the observed outcome Y in terms of counterfactual outcomes $Y(t) = f_Y(t, V, \epsilon_Y)$; $t \in \{0, 1\}$ as

$$Y = Y(0) \cdot \mathbf{1}[T = 0] + Y(1) \cdot \mathbf{1}[T = 1]. \tag{24.}$$

It is convenient to decompose each counterfactual $Y(t)$; $t \in \{0, 1\}$ into its mean, μ_t, and its deviation from the mean, U_t. Our model is given by

$$Y_1 = \mu_1 + U_1, \tag{25.}$$

$$Y_0 = \mu_0 + U_0, \tag{26.}$$

$$T = \mathbf{1}[P(Z) \geq U], \tag{27.}$$

where Equation 27 is due to the separability assumption on the choice equation that renders $U \sim$ unif[0, 1]. The unobserved confounding variable U causes U_1, U_0. All unobserved variables are statistically independent of the instrument Z—that is, $(U_1, U_0, U) \perp\!\!\!\perp Z$. Moreover, the unconditional expectation of U_0 and U_1 is zero. We can express the expected value of the outcome Y given the IV Z and the choice $T \in \{0, 1\}$ as

$$E_e(Y \mid Z, T = 0) = \mu_0 + E_e(U_0 | Z = z, T = 0), \tag{28.}$$

$$= \mu_0 + E_e(U_0 | Z = z, P(Z) < U), \tag{29.}$$

$$= \mu_0 + E_e(U_0 | P(Z) < U), \tag{30.}$$

$$= \mu_0 + K_0(P(Z)), \tag{31.}$$

where Equation 28 uses the additivity assumption of the outcome equation in Equation 26. Equation 29 uses the fact that the event $T = 0$ is equivalent to the event $P(Z) < U$. Equation 30 uses the independence relationship $Z \perp\!\!\!\perp U_0$. Equation 31 expresses the expected value of the endogenous error term U_0 as a control function of the propensity score $K_0(P(Z)) = E_e(U_0 | P(Z) < U)$. The equation states that the endogeneity problem of error term U_0 can be addressed by evaluating the outcome Y as a function of the propensity score (i.e., the probability of selection).

The value of the control function $K_0(P(Z))$ at $P(Z) = 0$ is $K_0(0) = E_e(U_0|0 < U) = E_e(U_0) = 0$. Thus, we can estimate the mean μ_0 of the counterfactual outcome Y_0 by a two-step procedure that first estimates the propensity scores $P(Z)$ and then evaluates the outcome Y conditioned on $T = 0$ as a function of the propensity scores. For instance, let the control function $K_0(P(Z))$ be approximated by the local polynomial of degree N $K_0(P) = \sum_{k=1}^{N} \beta_{k,0} P^k$, and let P_i be the estimated propensity score for agent i. We can then estimate the following outcome equation:

$$Y_i = \alpha_0 + \sum_{k=1}^{n} \beta_{k,0} P_i^k + \epsilon_i, \text{ for all } i \text{ such that } T_i = 0. \tag{32.}$$

The counterfactual mean μ_0 is obtained by setting the propensity scores to zero and is estimated by the parameter α_0 in Equation 32. We can apply the same rationale in Equations 28–31 to the case of $T = 1$. The resulting control function is given by

$$E_e(Y \mid Z, T = 1) = \mu_0 + K_1(P(Z)), \text{ where } K_1(P(Z)) = E_e(U_1|P(Z) \geq U). \tag{33.}$$

In the case of $T = 1$, we seek to evaluate the control function $K_1(P(Z))$ at the value $P(Z) = 1$, that is, $K_1(1) = E_e(U_1|1 \geq U) = E_e(U_1) = 0$. We can then estimate the following outcome equation:

$$Y_i = \alpha_1 + \sum_{k=1}^{n} \beta_{k,1} P_i^k + \epsilon_i, \text{ for all } i \text{ such that } T_i = 1. \tag{34.}$$

The counterfactual mean μ_1 is obtained by setting the propensity scores to 1 and is estimated by the linear combination $\alpha_1 + \sum_{k=1}^{N} \beta_{k,1}$ in Equation 34.

Broadly speaking, the control function approach seeks to use observed data to estimate an unobserved variable K that corrects for the endogenous effects of the confounding variable V using the structure of the model. **Figure 4** presents a modified version of the Roy model that can assist in gaining a deeper understanding. Variable K in the DAG is not observed. It mediates the impact of V on Y. Outcome Y is statistically independent of V when conditioned on K and T. We also have that $Y(t) \perp\!\!\!\perp T|K$ holds. The technique to identify the causal effect of T on Y is based on a two-step procedure that first uses IV Z and T to identify and estimate K and then uses the matching condition $Y(t) \perp\!\!\!\perp T|K$ to identify the causal impact of K and T on Y.

As previously noted, our discussion on identification approach is necessarily brief. We can only mention a small portion of the extensive economic literature on identification techniques. Some of the essential identification strategies that are not included in this review are panel data analysis (see, e.g., Heckman & Robb 1985a, Blundell et al. 1998), difference-in-differences estimators (see Heckman et al. 1999, Bertrand et al. 2004, Abadie 2005, Athey & Imbens 2006), regression discontinuity design,[20] simultaneous equation models (see Heckman 1978; Tamer 2003;

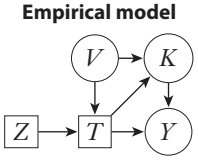


Figure 4

Control function: empirical causal model.

[20]Regression discontinuity estimators, which are versions of IV estimators, are discussed by Heckman & Vytlacil (2007b).

Quandt 1988; Matzkin 2008, 2015, 2013), and nonseparable equation models (see Matzkin 1993, Altonji & Matzkin 2005).

5. THE NEYMAN–RUBIN CAUSAL MODEL

The NR causal model was developed by Neyman (1923) and Fisher (1935) and further popularized by Cox (1958) and Holland (1986). The NR causal model is a radical departure from the causal frameworks studied in Section 3. The NR causal model is conceptually motivated by the method of RCTs. It focuses on units of analysis instead of systems of equations. Causal models are characterized by statistical independence relationships among counterfactual variables rather than structural equations.

We exemplify the method using the Generalized Roy model in **Table 2**. A unit of analysis is denoted by $i \in \mathcal{I}$, which often represents an economic agent. $T_i(z)$ is the potential treatment when the instrument Z is set to value $z \in \text{supp}(Z)$, and $Y_i(t, z)$ is the potential outcome of agent i when Z is set to value $z \in \text{supp}(Z)$ and choice T is set to $t \in \text{supp}(T)$. The NR framework characterizes the Roy model by the following IV assumptions.

1. Exclusion restriction: $Y_i(t, z) = Y_i(t, z')$ for all $z, z' \in \text{supp}(Z)$ and for all $i \in \mathcal{I}$. 35.
2. IV relevance: Z is not statistically independent of T, that is, $Z \not\perp\!\!\!\perp T$. 36.
3. Exogeneity condition: $Z \perp\!\!\!\perp (Y(t), T(z))$. 37.

The NR model is popular due to its simplicity. It suppresses the structural equations that determine the causal relationship among variable models; instead, it describes the causal model through independence relationships among the counterfactuals of the observed variables of the causal model. It is easy to verify that the structural equations of the Roy model in **Table 2** generate the IV assumptions of Equations 35–37.[21] Otherwise stated, the IV assumptions of Equations 35–37 are a consequence of the causal relations described by the structural equations of the model.

It turns out that any causal model can, in principle, be equivalently described by structural equations or by the independence relationships utilized in the NR framework (Pearl 2012). This equivalence does not imply that the NR framework offers the same capabilities as a causal framework based on structural equations. The NR framework lacks unobserved variables and structural equations, both of which are essential tools for causal inquiry.

The lack of unobserved variables precludes econometric strategies that impose restrictions on both observable and unobservable model components. It prevents the theoretical development of models and significantly restricts the ability to analyze causal models using the techniques described in Section 4.

The lack of structural equations impairs model interpretation. It complicates the assessment of causal relationships between model variables. It prevents the use of a hypothetical model that elucidates the notion of causality. The absence of structural equations is particularly troublesome when judging the plausibility of the causal assumptions that characterize a model.

In practice, the set of tractable models that can be reasonably investigated by the NR framework is restricted to a few possibilities: RCTs, matching on observables, IV models, and their many

[21] The exclusion restriction is due to the fact that Z does not directly cause Y. IV relevance is due to the fact that Z causes T. The exogeneity condition is due to $Z \perp\!\!\!\perp V$.

surrogates (see Imbens & Rubin 2015). This section illustrates some of the drawbacks of the NR framework in analyzing key economic models.

5.1. Revisiting the Local Instrumental Variable Model Under the Neyman–Rubin Framework

We examine the LIV model of Section 4.2 from the perspective of the NR framework. Consider a binary choice model $T \in \{0, 1\}$ for which the IV assumptions of Equations 35–37 hold. These assumptions are not sufficient to identify causal effects. An additional assumption that secures the identification of the LATE parameter is the monotonicity condition of Imbens & Angrist (1994). The condition states that a change in the instrument induces the agents to change their treatment choice in the same direction. Notationally, for any $z, z' \in \text{supp}(Z)$, we have that

$$T_i(z) \geq T_i(z') \ \forall i \in \mathcal{I} \ \text{ or } \ T_i(z) \leq T_i(z') \ \forall i \in \mathcal{I}. \tag{38}$$

Vytlacil (2002) shows that the monotonicity condition of Equation 38 is equivalent to the separability assumption $T = \mathbf{1}[P(Z) \geq U]$. Otherwise stated, the model generated by the monotonicity of Equation 38 and the IV assumptions of Equations 35–37 is the NR counterpart of the LIV model described in Section 4.

Although both frameworks are suitable to investigating causal effects, the LIV model explicitly displays the unobserved confounding variable U, whereas NR does not. This feature fosters further analysis that cannot be conducted in the NR framework. For instance, the unobserved variable U enables us to define the marginal treatment effect (MTE) of Heckman & Vytlacil (2005) as

$$MTE(u) = E_{\text{h}}(Y \mid \tilde{T} = 1, U = u) - E_{\text{h}}(Y \mid \tilde{T} = 0, U = u) = E_{\text{fix}}(Y(1) - Y(0) \mid U = u).$$

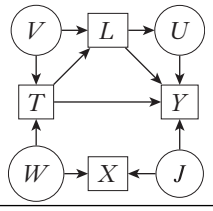
The MTE plays a primary role in characterizing a variety of causal effects that are typically sought in policy evaluations. **Table 4** shows that causal effects can be expressed as weighted averages of the MTE. This result, in turn, fosters additional literature on a variety of related questions regarding

Table 4 Causal parameters as weighted averages of the MTE

Causal parameters	MTE representation	Weights
$ATE = E(Y(1) - Y(0))$	$= \int_0^1 MTE(p) W^{ATE}(p) dp$	$W^{ATE}(p) = 1$
$TT = E(Y(1) - Y(0) \mid T = 1)$	$= \int_0^1 MTE(p) W^{TT}(p) dp$	$W^{TT}(p) = \dfrac{1 - F_P(p)}{\int_0^1 (1 - F_P(t)) dt}$
$TUT = E(Y(1) - Y(0) \mid T = t_0)$	$= \int_0^1 \Delta^{MTE}(p) W^{TUT}(p) dp$	$W^{TUT}(p) = \dfrac{F_P(p)}{\int_0^1 F_p(t) dt}$
$TSLS = \dfrac{Cov(Y, Z)}{Cov(T, Z)}$	$= \int_0^1 MTE(p) W^{TSLS}(p) dp$	$W^{TSLS}(p) = \dfrac{\int_p^1 (t - E(P)) dF_P(t)}{\int_0^1 (t - E(P))^2 dF_P(t)}$
$LATE = \dfrac{E(Y \mid Z = z_1) - E(Y \mid Z = z_0)}{P(z_1) - P(z_0)}$	$= \int_{P(z_0)}^{P(z_1)} MTE(p) W^{LATE}(p) dp$	$W^{LATE}(p) = \dfrac{1}{P(z_1) - P(z_0)}$

$F_p(p) = P(P(Z) \leq p)$ is the cumulative distribution function of the propensity score $P(Z)$. Abbreviations: ATE, average treatment effect; LATE, local average treatment effect; MTE, marginal treatment effect; TSLS, two-stage least squares; TT, treatment on the treated; TUT, treatment on the untreated.

Table 5 Hypothetical matching model

Causal model	Directed acyclic graph	Independence relationships
$V = f_V(\epsilon_V)$ $T = f_T(V, W, \epsilon_T)$ $L = f_L(T, V, \epsilon_L)$ $X = f_X(W, J, \epsilon_X)$ $Y = f_Y(T, L, U, J, \epsilon_Y)$ J, W, V, U are external variables		$Y(t) \perp\!\!\!\perp T \mid L$ $Y(t) \not\perp\!\!\!\perp T \mid X$ $Y(t) \not\perp\!\!\!\perp T \mid (X, L)$

estimation, partial identification, extrapolation, and inference of causal effects.[22] In conclusion, switching from the NR framework to structural equations allows for deeper causal analyses that unlock additional research paths.

5.2. Interpreting Matching on Observables

The most popular identification strategy in the NR framework is matching on observables, described in Section 4.1. In the NR framework, the causal model is described by the independent assumption $Y(t) \perp\!\!\!\perp T \mid X$, where X denotes preintervention variables.

Matching condition $Y(t) \perp\!\!\!\perp T \mid X$ may lead the researcher to believe that increasing the number of matching variables X always reduces the potential bias generated by unobserved confounders. This statement is known to be false (see, e.g., Greenland et al. 1999, Heckman & Navarro 2004, Pearl 2009b). However, it is rather difficult to investigate its falsity using the NR framework. The causal model of **Table 5** clarifies this point.

The causal model of **Table 5** comprises four observed variables: a treatment T, an outcome Y, a pretreatment variable X, and a posttreatment variable L. The model also contains four unobserved variables V, U, W, and J that are external.

This model generates peculiar counterfactual relationships. The matching condition $Y(t) \perp\!\!\!\perp T \mid L$ holds for the posttreatment variable L but does not hold for the pretreatment variable X. The counterfactual outcome $Y(t)$ is not statistically independent of T conditional on X, that is, $Y(t) \not\perp\!\!\!\perp T \mid X$. Moreover, adding the pretreatment variable X to the conditioning set of $Y(t) \perp\!\!\!\perp T \mid L$ invalidates the independence relationship, as $Y(t) \not\perp\!\!\!\perp T \mid (X, L)$.

The causal model of **Table 5** exemplifies the difficulty of performing causal investigation within the NR framework. The model's unusual properties stem from its peculiar causal relationships. These relationships are hard to assess using the NR framework, which lacks structural equations and unobserved variables. Moreover, the direction of the causal relationships among variables in NR is not explicitly stated.

6. THE DO-CALCULUS: ADVANTAGES AND LIMITATIONS

The DoC of Pearl (1995, 2009a, 2012) uses graph-based algorithms to examine the identification of counterfactuals in causal models represented by DAGs.[23] It employs structural equations, allows for unobserved variables, and clearly specifies the causal relationships among model variables.

[22] For examples of this literature, readers are referred to Heckman & Vytlacil (2007b), Brinch et al. (2017), Mogstad et al. (2018).

[23] For a recent book on the graphical approach to causality, readers are referred to Peters et al. (2017); for related works on causal discovery, readers may consult Hoyer et al. (2009), Glymour et al. (2014), Heckman & Pinto (2015), and Lopez-Paz et al. (2017).

The DoC is similar to the hypothetical model in the sense that both are based on structural equations. The frameworks, however, differ significantly in terms of counterfactual manipulations. DoC defines counterfactuals by "shutting down" (eliminating) the equation of the treatment choice, whereas the hypothetical model adds a hypothetical variable of treatment choice that formalizes the notion of a thought experiment. DoC employs DAG-based algorithms outside the realm of standard statistical analysis. The hypothetical model makes statistics converse with causality. Section 6.3 compares the identification techniques used in each framework.

Pearl and coauthors have made significant contributions to the theory of causality. The primary accomplishment of the DoC is that it is a complete procedure. This means that if a counterfactual outcome of a causal model defined by a DAG is identified, then it can always be assessed by the iterative application of the DoC algorithms (Huang & Valtorta 2006, Shpitser & Pearl 2006).

Despite its apparent relevance, economists rarely employ the DoC. Its major limitation is that it only applies to nonparametric models that can be fully characterized by a DAG. The DoC does not apply to equilibrium or simultaneous equation models. It does not apply to identification strategies that invoke functional forms restrictions either. DoC cannot accommodate IV assumptions such as monotonicity or separability. DoC eliminates the identification strategies discussed in Section 4. Applying the DoC to the Generalized Roy model generates the misleading claim that the Roy model is not identified.

6.1. Do-Calculus Machinery

Let G denote a DAG that represents the original causal model. Let Y, K, X, T denote disjoint variable sets in $\mathcal{T}$. In DoC notation, $T(X)$ denotes the variables in T that do not directly or indirectly cause X. The DoC uses $G_{\overline{K}}$ for the derived DAG that deletes all causal arrows arriving at K in the original DAG G. $G_{\underline{T}}$ denotes the DAG that deletes all causal arrows emerging from T. In this notation, $G_{\overline{K},\underline{T}}$ stands for the derived DAG that suppresses all arrows arriving at K and emerging from T, while $G_{\overline{K,T(X)}}$ deletes all arrows arriving at K in addition to arrows arriving at $T(X)$—namely, arriving at variables in T that are not ancestors of X.

The DoC uses three rules. Each rule combines a graphical condition and a conditional independence relation that, when satisfied, imply a probability equality.

1. Rule 1: If $Y \perp\!\!\!\perp T|(K,X)$ holds in $G_{\overline{K}}$, then $P(Y|\text{do}(K), T, X) = P(Y|\text{do}(K), X)$.
2. Rule 2: If $Y \perp\!\!\!\perp T|(K,X)$ holds in $G_{\overline{K},\underline{T}}$, then $P(Y|\text{do}(K), \text{do}(T), X) = P(Y|\text{do}(K), T, X)$.
3. Rule 3: If $Y \perp\!\!\!\perp T|(K,X)$ holds in $G_{\overline{K},\overline{T(X)}}$, then $P(Y|\text{do}(K), \text{do}(T), X) = P(Y|\text{do}(K), X)$.

Checking if a causal effect is identified requires the iterative use of these rules. We present several examples of how to use the DoC method below.

6.2. Using Do-Calculus to Investigate the Roy Model

The first graph of **Figure 5** presents the DAG of the original Roy model, which is denoted by G. The second graph displays the DAG $G_{\underline{Z}}$ which suppresses the arrow arising from Z. The LMC of

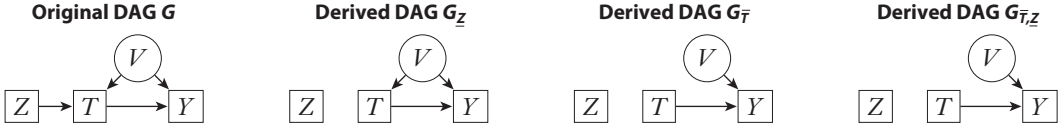


Figure 5

Using do-calculus to investigate the Roy model.

Z on DAG $G_{\underline{Z}}$ is $Z \perp\!\!\!\perp (Y, T)$. From Rule 2 of the DoC, we obtain $P(T|\text{do}(Z)) = P(T|Z)$. Summarizing, we obtain

$$G_{\underline{Z}} \Rightarrow T \perp\!\!\!\perp Z \Rightarrow \text{ by Rule 2 we have } P(T|\text{do}(Z)) = P(T|Z). \quad 39.$$

Equation 39 indicates that Z is statistically independent of T when we fix Z. In the NR framework, this refers to the exogeneity condition $T(z) \perp\!\!\!\perp Z$—namely, that the instrument Z is independent of the counterfactual choice $T(z)$. Instrument Z in DAG $G_{\underline{Z}}$ is independent of both T and Y. Thus, we can replace T by Y in Equation 39 to obtain $P(Y|\text{do}(Z)) = P(Y|Z)$. This means that conditioning on Z is equivalent to fixing Z. Indeed, the instrument Z is an external variable, and the causal operation of fixing is translated to standard statistical conditioning.

The third graph of **Figure 5** displays the DAG $G_{\overline{T}}$, which suppresses the arrow arriving at T. The LMC of Z on $G_{\overline{T}}$ implies $Z \perp\!\!\!\perp (Y, T)$. By Rule 1 of the DoC, we have that $P(Y|\text{do}(T), Z) = P(Y|\text{do}(T))$. Summarizing, we have

$$G_{\overline{T}} \Rightarrow Y \perp\!\!\!\perp Z|T \Rightarrow \text{ by Rule 1 we have } P(Y|\text{do}(T), Z) = P(Y|\text{do}(T)). \quad 40.$$

This means that Z is statistically independent of Y when we fix T. This statement refers to the exogeneity condition $Y(t) \perp\!\!\!\perp Z$ or the independence relationship $Y \perp\!\!\!\perp Z|\tilde{T}$ of the hypothetical model framework.

The last graph of **Figure 5** displays the DAG $G_{\overline{T},\underline{Z}}$, which suppresses the arrow arriving at T and arising from Z. Note that the DAGs $G_{\overline{T},\underline{Z}}$ and $G_{\overline{T}}$ are the same. To repeat for emphasis,

$$G_{\overline{T}} \Rightarrow Y \perp\!\!\!\perp Z|T \Rightarrow \text{ by Rule 1 we have } P(Y|\text{do}(T), Z) = P(Y|\text{do}(T)). \quad 41.$$

The LMC of Z in $G_{\overline{T},\underline{Z}}$ is $Z \perp\!\!\!\perp (T, Y, V)$ which implies that $Y \perp\!\!\!\perp Z|T$ holds. Using Rule 2 of the DoC, we obtain

$$G_{\overline{T},\underline{Z}} \Rightarrow Y \perp\!\!\!\perp Z|T \Rightarrow \text{by Rule 2 we have } P(Y|\text{do}(T), \text{do}(Z)) = P(Y|\text{do}(T), Z). \quad 42.$$

Combining $P(Y|\text{do}(T), Z) = P(Y|\text{do}(T))$ in Equation 41 with $P(Y|\text{do}(T), \text{do}(Z)) = P(Y|\text{do}(T), Z)$ in Equation 42, we obtain $P(Y|\text{do}(T), \text{do}(Z)) = P(Y|\text{do}(T))$. This means that the probability distribution of the outcome Y when we fix both Z and T is the same as the counterfactual outcome generated by fixing only the choice T. In the NR framework, this property refers to the exclusion restriction $Y_i(t, z) = Y_i(t, z')$ for all $z, z' \in \text{supp}(Z)$.

These statements exhaust the analysis of the Roy model that can be performed using DoC. The method describes the properties of the Roy model, but applications of its rules alone cannot deliver identification of treatment effects. The type of assumptions that would secure the identification of treatment effects in the Roy model is beyond the DAG representation.

6.3. The Front-Door Model

To make a more positive point, we apply the identification machinery of DoC to a causal model where treatment effects are identified. We investigate the front-door model of Pearl (2009a). We use this example to contrast the identification approaches of DoC with those of the hypothetical model.

The front-door model of Equations 43–46 below consists of three observed variables T, M, Y and an unobserved confounding variable V. Treatment T causes a mediator M, which in turn

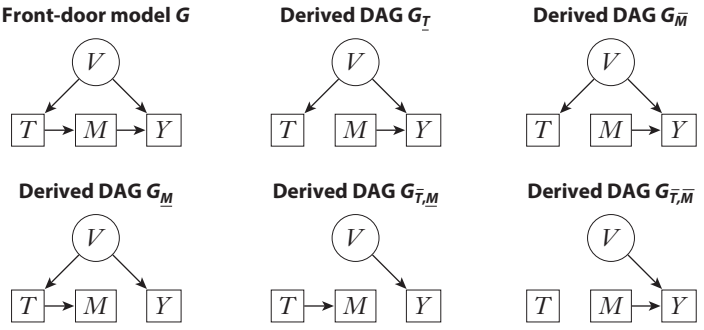


Figure 6

Using do-calculus to identify the causal effect of T on Y in the front-door model.

causes outcome Y. Confounding variable V causes T and Y but not M.[24] We have

$$V = f_V(\epsilon_V), \tag{43.}$$

$$T = f_T(V, \epsilon_T), \tag{44.}$$

$$M = f_M(T, \epsilon_M), \tag{45.}$$

$$Y = f_Y(M, V, \epsilon_Y). \tag{46.}$$

The causal effect of T on Y in the front-door model is identified. This result arises from the fact that the causal effect of T on M is not confounded by V, and therefore it is identified by standard methods. Also, conditioning on T blocks the effect of the confounder V on M. Thus, we can identify the causal effect of M on Y conditional on T. The causal effect of T on Y can be evaluated as the compound effect of T on M and M on Y.

We illustrate how to use the DoC to identify the distribution of the counterfactual outcome $P(Y(t))$ or, equivalently, $P(Y|\text{do}(T) = t)$. For the sake of notational simplicity, suppose that all variables are discrete. The DoC is cumbersome. The method requires the five derived DAGs displayed in **Figure 6**. The identification formula of the counterfactual outcome is obtained by the following sequence of steps.

1. $T \perp\!\!\!\perp M$ in $G_{\underline{T}}$ holds, thus by Rule 2 we have $P(M|\text{do}(T)) = P(M|T)$.
2. $M \perp\!\!\!\perp T$ in $G_{\overline{M}}$ holds, thus by Rule 3 we have $P(T|\text{do}(M)) = P(T)$.
3. $M \perp\!\!\!\perp Y|T$ in $G_{\underline{M}}$ holds, thus by Rule 2 we have $P(Y|T, \text{do}(M)) = P(Y|T, M)$.
4. Collecting these results, we have that

$$\therefore P(Y|\text{do}(M)) = \sum_t P(Y|T = t, \text{do}(M))P(T = t|\text{do}(M)),$$

which by the law of iterated expectations (LIE) equals

$$\sum_t P(Y|T = t, M)P(T = t).$$

5. $Y \perp\!\!\!\perp M|T$ in $G_{\overline{T},\underline{M}}$ holds, thus by Rule 2 we have $P(Y|M, \text{do}(T)) = P(Y|\text{do}(M), \text{do}(T))$.

[24] As before, the error terms $\epsilon_V, \epsilon_T, \epsilon_M, \epsilon_Y$ in the front-door model of Equations 43–46 are mutually statistically independent.

6. $Y \perp\!\!\!\perp T|M$ in $G_{\overline{T},\overline{M}}$ holds, thus by Rule 3 we have $P(Y|\text{do}(T), \text{do}(M)) = P(Y|\text{do}(M))$.
7. Collecting these results, we have that $P(Y|Z, \text{do}(T)) = P(Y|\text{do}(Z), \text{do}(T)) = P(Y|\text{do}(M))$.
8. Finally, we use the previous results to obtain the following equations:

$$\begin{aligned}
\therefore P(Y|\text{do}(T)=t) &= \sum_m P(Y|M=m, \text{do}(T)=t)P(M=m|\text{do}(T)=t) \text{ by LIE;} \\
&= \sum_m P(Y|\text{do}(M)=m, \text{do}(T)=t)P(M=m|\text{do}(T)=t) \text{ by Step 5;} \\
&= \sum_m P(Y|\text{do}(M)=m)P(M=m|\text{do}(T)=t) \text{ by Step 7;} \\
&= \sum_m \left(\sum_{T=t'} P(Y|T=t', M=m)P(T=t') \right) P(M=m|T=t) \text{ by Step 4.}
\end{aligned}$$

6.4. Reexamining the Front-Door Model Using the Hypothetical Model

We now investigate the same front-door model using the hypothetical framework. **Figure 7** displays the hypothetical model associated with the front-door model of Equations 43–46 as a DAG.

We seek to identify the counterfactual distribution of outcome $P(Y(t))$, or, equivalently, $P_h(Y|\tilde{T}=t)$. This means we seek to express $P_h(Y|\tilde{T}=t)$ in terms of the observed distribution $P_e(T, M, Y)$. Identification requires us to connect the probability distributions of the hypothetical and empirical models. To do so, we seek independence relationships in the hypothetical model that includes T and $\tilde{T}$. These are $Y \perp\!\!\!\perp \tilde{T}|(M, T)$ and $M \perp\!\!\!\perp T|\tilde{T}$.[25] It is also the case that $T \perp\!\!\!\perp \tilde{T}$ holds, as $\tilde{T}$ is exogenous and does not cause T. We can then apply the rules of Equations 15 and 16 to connect the probabilities of the empirical and hypothetical models:

$$Y \perp\!\!\!\perp \tilde{T}|(T, M) \Rightarrow \quad P_h(Y|\tilde{T}, T=t', M) = P_e(Y|T=t', M), \tag{47.}$$

$$M \perp\!\!\!\perp T|\tilde{T} \Rightarrow \quad P_h(M|\tilde{T}=t, T) = P_e(M|T=t), \tag{48.}$$

$$T \perp\!\!\!\perp \tilde{T}|T \Rightarrow \quad P_h(T=t'|\tilde{T}) = P_e(T=t'). \tag{49.}$$

The causal effect of T on Y of the front-door model is identified by

$$P_h(Y|\tilde{T}=t) = \sum_{t',m} P_h(Y|m, T=t', \tilde{T}=t)P_h(m|T=t', \tilde{T}=t)P_h(T=t'|\tilde{T}=t), \tag{50.}$$

$$= \sum_{t',m} P_e(Y|m, T=t')P_e(m|T=t)P_e(T=t'). \tag{51.}$$

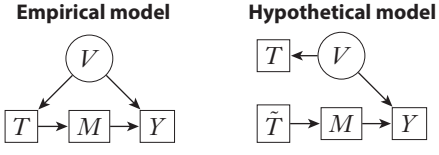


Figure 7

The empirical and hypothetical front-door models.

[25] The first independence condition is due to the LMCs $Y \perp\!\!\!\perp \tilde{T}|M$ and $(\tilde{T}, M) \perp\!\!\!\perp (T, V)$. The second one is due to the LMC of M.

Equation 50 is a sum of probabilities defined in the hypothetical model due to application of the LIE over T and M. Equation 51 replaces each of the hypothetical model probabilities with empirical model probabilities. We adopt the following short-hand notation:

$$\sum_{m,t'} P(Y|m, T = t')P(m|T = t) \text{ denotes } \sum_{m \in \text{supp}(M)} \sum_{t' \in \text{supp}(T)} P(Y|M = m, T = t')P(M = m|T = t).$$

6.5. Understanding the Hypothetical Model Identification Criteria

The identification of the counterfactual outcomes in the front-door model stems from the three independence relationships in Equations 47–49. These independence relationships comply with two general properties that enable the identification of counterfactual outcomes.

The first property is called alternate conditionals. It refers to the fact that the first relationship (Equation 47) establishes an independence regarding T conditional on $\widetilde{T}$. The second relationship (Equation 48) establishes independence of $\widetilde{T}$ conditional on T. The last one (Equation 49) cycles back, and it is an independence relationship regarding T conditional on $\widetilde{T}$. This property enables us to translate the probabilities of the hypothetical model into the probabilities of the empirical model via the connection rules of Equations 15 and 16.

The property of alternate conditionals ascribes an alternating feature to the identification equation (51). The first term in Equation 51 is conditioned on $T = t'$, which refers to the first conditional T in Equation 47. The identification equation (51) sums t' over the support of T. The second term in Equation 51 is conditioned on the treatment value $T = t$, which refers to the second conditional T in Equation 48. The value t remains fixed in the summation as it is the value used to define the counterfactual $(Y|\tilde{T} = t)$. The last term in Equation 51 alternates. It is conditioned on $T = t'$, which refers to the last conditional T in Equation 49, and t' varies in the summation.

The second property of the set of independence relationships is called bridging, and it refers to the variables other than $(T, \widetilde{T})$. The first independence relationship (Equation 47) starts with the outcome Y and is conditioned on the variable M. The second relationship (Equation 48) starts with M and is conditioned on no other variable besides T or $\widetilde{T}$. We say that variable M bridges the path between Y and $(T, \widetilde{T})$, that is, $Y \rightarrowtail M \rightarrowtail (T, \widetilde{T})$.

In general terms, the bridging condition refers to a sequence of nested sets $\mathcal{T}_1 \subset ... \subset \mathcal{T}_K$ of observed variables in $\mathcal{T}$ such that the property of alternate conditionals holds. By this we mean a sequence of conditional independence relationships that starts from $Y \perp\!\!\!\perp \tilde{T}|(T, \mathcal{T}_K)$ [or $Y \perp\!\!\!\perp T|(\tilde{T}, \mathcal{T}_K)$], then evolves to $(\mathcal{T}_K \setminus \mathcal{T}_{K-1}) \perp\!\!\!\perp T|(\tilde{T}, \mathcal{T}_{K-1})$ [or $(\mathcal{T}_K \setminus \mathcal{T}_{K-1}) \perp\!\!\!\perp \tilde{T}|(T, \mathcal{T}_{K-1})$], and finally ends in $\mathcal{T}_1 \perp\!\!\!\perp T|\tilde{T}$ [or $\mathcal{T}_1 \perp\!\!\!\perp \tilde{T}|T$].

Identification is secured whenever a set of conditional independence relationships among observed variables in the hypothetical model presents the alternate conditionals and the bridging properties.

Consider the complex mediation model of **Figure 8**. The model has three observed mediating variables M_1, M_2, M_3 and three unobserved, confounding variables V_1, V_2, V_3. The following conditional independence relationships hold for the hypothetical model:

$$Y \perp\!\!\!\perp \widetilde{T}|(T, M_3, M_2, M_1), \quad 52.$$

$$M_3 \perp\!\!\!\perp T|(\widetilde{T}, M_2, M_1), \quad 53.$$

$$M_2 \perp\!\!\!\perp \widetilde{T}|(T, M_1), \quad 54.$$

$$M_1 \perp\!\!\!\perp T|\widetilde{T}. \quad 55.$$

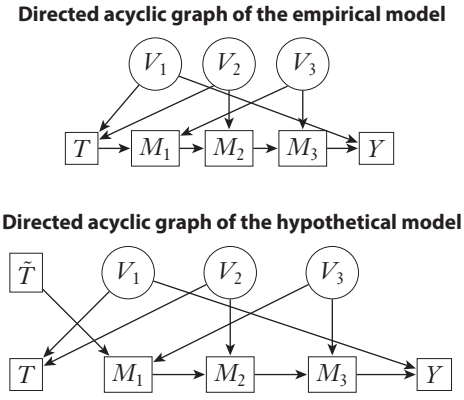


Figure 8

Using the hypothetical model to identify counterfactuals.

The set of independence relationships of Equations 52–55 is a case of alternate conditionals. The first relationship is conditioned on T, the second on $\widetilde{T}$, followed by T, and so on.

The bridging property also holds. The right-hand variable of each independence relationship gives the bridging sequence: $Y \rightarrowtail M_3 \rightarrowtail M_2 \rightarrowtail M_1 \rightarrowtail T$. The LIE and the independence relationships of Equations 52–55 enable us to express the counterfactual probability $P_b(Y|\widetilde{T})$ as

$$\begin{aligned}
&\text{hypothetical model} && P_{\mathrm{h}}(Y|\widetilde{T}=t)=\textstyle\sum_{t',m_3,m_2,m_1} A_{\mathrm{h}}\cdot B_{\mathrm{h}}\cdot C_{\mathrm{h}}\cdot D_{\mathrm{h}}\cdot E_{\mathrm{h}},\\
&\text{where} && A_{\mathrm{h}}=P_{\mathrm{h}}(Y|m_3,m_2,m_1,T=t',\widetilde{T}=t),\\
& && B_{\mathrm{h}}=P_{\mathrm{h}}(M_3=m_3|m_2,m_1,T=t',\widetilde{T}=t),\\
& && C_{\mathrm{h}}=P_{\mathrm{h}}(M_2=m_2|m_1,T=t',\widetilde{T}=t),\\
& && D_{\mathrm{h}}=P_{\mathrm{h}}(M_1=m_1|T=t',\widetilde{T}=t),\\
& && E_{\mathrm{h}}=P_{\mathrm{h}}(T=t'|\widetilde{T}=t).
\end{aligned}$$

The connection rules of Equations 15 and 16 enable us to translate the hypothetical model probabilities into empirical model probabilities. The identification equation displays the alternative pattern of values t and t' in the same fashion as the identification equation of the front-door model. We have

$$\begin{aligned}
&\text{empirical model} && P_{\mathrm{e}}(Y(t))=\textstyle\sum_{t',m_3,m_2,m_1} A_{\mathrm{e}}\cdot B_{\mathrm{e}}\cdot C_{\mathrm{e}}\cdot D_{\mathrm{e}}\cdot E_{\mathrm{e}},\\
&\text{where} && A_{\mathrm{e}}=P_{\mathrm{e}}(Y|m_3,m_2,m_1,T=t'),\\
& && B_{\mathrm{e}}=P_{\mathrm{e}}(M_3=m_3|m_2,m_1,T=t),\\
& && C_{\mathrm{e}}=P_{\mathrm{e}}(M_2=m_2|m_1,T=t'),\\
& && D_{\mathrm{e}}=P_{\mathrm{e}}(M_1=m_1|T=t),\\
& && E_{\mathrm{e}}=P_{\mathrm{e}}(T=t').
\end{aligned}$$

6.6. The Identification Expression

The DoC and hypothetical model generate equivalent but dramatically different expressions for the same identified causal effect. We illustrate this fact using the causal model of **Figure 9**, which is widely known in the DoC literature. The graph on the left depicts the observed variables of the causal model using a DAG. It states that T causes Z_1; Z_2 causes T, Z_1, Z_3; and Y is caused by T and Z_3. The bi-directed dashed arrows represent the causal path of unobserved confounding variables. The center graph in **Figure 9** presents the equivalent empirical model, which explicitly

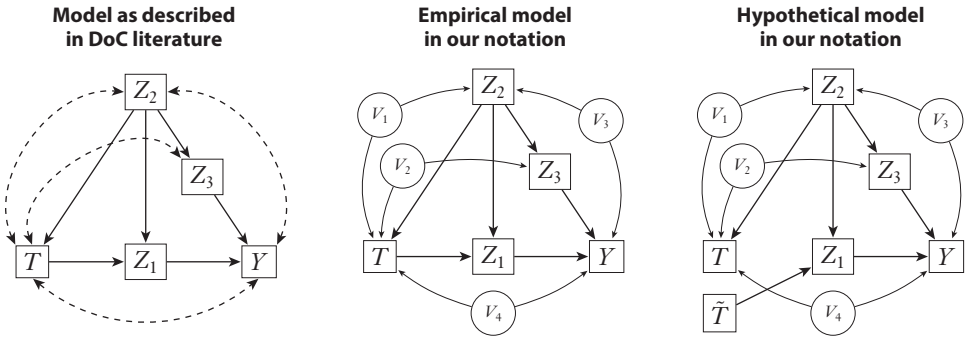


Figure 9

Example of a well-known causal model in do-calculus (DoC) literature.

displays the unobserved confounding variables V_1, V_2, V_3, V_4. The graph to the right presents the hypothetical model counterpart.

The following conditional independence relationships hold for the hypothetical model on the right side of **Figure 9**:

$$Y \perp\!\!\!\perp \tilde{T} | (T, Z_1, Z_2, Z_3), \quad 56.$$

$$Z_1 \perp\!\!\!\perp T | (\tilde{T}, Z_2, Z_3), \quad 57.$$

$$(Z_2, Z_3) \perp\!\!\!\perp \tilde{T} | T. \quad 58.$$

The independence relationships of Equations 56–58 generate the bridging sequence $Y \rightarrowtail Z_1 \rightarrowtail (Z_2, Z_3) \rightarrowtail M_1 \rightarrowtail T$, which generates the following identification expression:

$$P_{\mathrm{h}}(Y|\tilde{T} = t) = \sum_{t', z_1, z_2, z_3} P_{\mathrm{e}}(Y|z_1, z_2, z_3, T = t') P_{\mathrm{e}}(z_1|z_2, z_3, T = t) P_{\mathrm{e}}(z_2, z_3|T = t') P_{\mathrm{e}}(T = t'). \quad 59.$$

Note that the values of T in the expression above (Equation 59) alternate from t' to t. The identification expression generated by the DoC does not have this property. The DoC algorithm of Bareinboim & Pearl (2016) generates the following identification expression:

$$P(Y(t)) = \sum_{z_1, z_2, z_3} P(z_3, z_2) P(z_1|T = t, z_2) \cdot \left(\frac{\sum_{t'} P(Y, z_3|T = t', z_1, z_2) P(T = t', z_2)}{\sum_{t'} P(z_3|T = t', z_1, z_2) P(T = t', z_2)} \right) P(z_2). \quad 60.$$

The expressions shown in Equations 59 and 60 appear to be substantially different. They are, in fact, equivalent. Tikka & Karvanen (2017) have shown that the DoC expression (Equation 60) is equivalent to

$$P(Y(t)) = \sum_{z_3, z_2, z_1} P(z_1|z_2, T = t) P(z_2) \cdot \sum_{T = t'} P(Y|z_2, T = t', z_3, z_1) P(z_3|z_2, T = t') P(T = t'|z_2).$$

Moreover, the independence condition $Z_1 \perp\!\!\!\perp Z_3 | (T, Z_2)$ holds for the empirical model of **Figure 9**. We can combine this information to show that

$$
\begin{aligned}
P(Y(t)) &= \sum_{z_3,z_2,z_1} P(z_1|z_2, T=t)P(z_2) \sum_{t'} P(Y|z_3,z_1,z_2,T=t')P(z_3|z_2,T=t')P(T=t'|z_2) \\
&= \sum_{z_3,z_2,z_1} P(z_1|z_2, T=t)P(z_2) \sum_{t'} P(Y|z_2,z_3,z_1,T=t')P(z_3|z_2,T=t')\frac{P(z_2|T=t')P(T=t')}{P(z_2)} \\
&= \sum_{z_3,z_2,z_1} P(z_1|z_2, T=t) \sum_{t'} P(Y|z_2,z_3,z_1,T=t')P(z_3|z_2,T=t')P(z_2|T=t')P(T=t') \\
&= \sum_{z_3,z_2,z_1} P(z_1|z_2,z_3, T=t) \sum_{t'} P(Y|z_2,z_3,z_1,T=t')P(z_3,z_2|T=t')P(T=t') \\
&= \sum_{z_3,z_2,z_1,t'} P(Y|z_2,z_3,z_1,T=t')P(z_1|z_2,z_3,T=t)P(z_3,z_2|T=t')P(T=t').
\end{aligned}
$$

The second equation uses Bayes' theorem. The third equation cancels out $P(z_2)$. The fourth equation is due to $Z_1 \perp\!\!\!\perp Z_3|(T, Z_2)$. The last equation rearranges the terms.

Finally, the independence relationships (Equations 56–58) of the hypothetical model in **Figure 9** would also hold if we were to suppress Z_3, that is,

$$
\begin{aligned}
Y &\perp\!\!\!\perp \widetilde{T}|(T, Z_1, Z_2), \\
Z_1 &\perp\!\!\!\perp T|(\widetilde{T}, Z_2), \\
Z_2 &\perp\!\!\!\perp \widetilde{T}|T.
\end{aligned}
$$

Thus, we can also express the counterfactual outcome probability $P_{\mathrm{h}}(Y|\tilde{T}=t)$ as

$$
P_{\mathrm{h}}(Y|\tilde{T}=t) = \sum_{t',z_1,z_2} P_{\mathrm{e}}(Y|z_1,z_2,T=t')P_{\mathrm{e}}(z_1|z_2,T=t)P_{\mathrm{e}}(z_2|T=t')P_{\mathrm{e}}(T=t'). \qquad 61.
$$

7. CONCLUSION

This article presents the framework of the econometric model for causal policy analysis. We discuss the definition of causal parameters and approaches to their identification within the econometric framework. We consider two approximations to it and their relationship with the econometric approach.

The econometric model is based on clearly stated and interpretable models of behavior that adequately characterize the predictions of economic theory and allow for testing it; synthesizing evidence on it from multiple sources; constructing credible policy counterfactuals, including forecasting policy impacts in new environments; and forecasting the likely impacts of policies never previously implemented. The econometric approach delineates the definition of causal parameters and their identification as two separate tasks.

The two approximating approaches are the NR approach, rooted in the statistics of experiments, and the DoC, which originated in computer science. Both are recent developments that attempt to address some of the same problems tackled by the econometric approach. Each has important, but different, limitations. Neither has the flexibility or clarity of the econometric approach.

All start from the basic intuitive definition of a causal effect as a ceteris paribus consequence of a policy change. However, the rules of constructing and identifying counterfactuals are very different.

The DoC invokes a special set of rules for identifying causal parameters. It relies heavily on recursive DAGs. Its rigid rules preclude the use of many traditional techniques of identification.

The NR approach eschews the benefits of structural equations and many fruitful strategies for their identification. Reflecting on its origins, it casts all policy problems into a treatment-control framework. In some versions, it conflates issues of definition with issues of identification. Its lack of reliance on structural equations makes it difficult to interpret estimates obtained from it or to analyze well-posed economic questions with it.

Economics has a rich body of theory and tools to address policy problems. Applied economists today would do well to use the impressive body of tools embodied in modern structural econometrics.

DISCLOSURE STATEMENT

This research was supported in part by NIH grant NICHD R37HD065072. The views expressed in this article are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Institutes of Health. The authors are not aware of any other affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Aakvik A, Heckman JJ, Vytlacil EJ. 1999. *Training effects on employment when the training effects are heterogeneous: an application to Norwegian vocational rehabilitation programs*. Work. Pap. 0599, Univ. Bergen, Bergen, Nor.

Aakvik A, Heckman JJ, Vytlacil EJ. 2005. Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs. *J. Econom.* 125(1–2):15–51

Abadie A. 2005. Semiparametric difference-in-differences estimators. *Rev. Econ. Stud.* 72(1):1–19

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

Ahn H, Powell J. 1993. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *J. Econom.* 58(1–2):3–29

Altonji JG, Matzkin RL. 2005. Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica* 73(4):1053–102

Angrist JD, Imbens GW, Rubin D. 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91(434):444–55

Athey S, Imbens GW. 2006. Identification and inference in nonlinear difference-in-differences models. *Econometrica* 74(2):431–97

Balke AA, Pearl J. 1993. *Nonparametric bounds on causal effects from partial compliance data*. Tech. Rep. R-199, Univ. Calif., Los Angeles

Bareinboim E, Pearl J. 2016. Causal inference and the data-fusion problem. *PNAS* 113(27):7345–52

Bertrand M, Duflo E, Mullainathan S. 2004. How much should we trust differences-in-differences estimates? *Q. J. Econ.* 119(1):249–75

Blundell R, Duncan A, Meghir C. 1998. Estimating labor supply responses using tax reforms. *Econometrica* 66(4):827–61

Blundell R, Powell J. 2003. Endogeneity in nonparametric and semiparametric regression models. In *Advances in Economics and Econometrics: Theory and Applications*, Vol. 2, ed. M Dewatripont, LP Hansen, SJ Turnovsky, pp. 312–57. Cambridge, UK: Cambridge Univ. Press

Brinch CN, Mogstad M, Wiswall M. 2017. Beyond late with a discrete instrument. *J. Political Econ.* 125(4):985–1039

Buchinsky M, Pinto R. 2021. *Using economic incentives to generate monotonicity criteria of IV models*. Unpubl. Ms., Univ. Calif., Los Angeles

Carneiro P, Hansen K, Heckman JJ. 2001. Removing the veil of ignorance in assessing the distributional impacts of social policies. *Swedish Econ. Policy Rev.* 8(2):273–301

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(2):361–422
Cox DR. 1958. *Planning of Experiments*. New York: Wiley
Cunha F, Heckman JJ, Navarro S. 2005. Separating uncertainty from heterogeneity in life cycle earnings. *Oxf. Econ. Pap.* 57(2):191–261
Cunha F, Heckman JJ, Schennach SM. 2010. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78(3):883–931
Dawid AP. 1976. Properties of diagnostic data distributions. *Biometrics* 32(3):647–58
Duncan OD, Goldberger AS. 1973. *Structural Equation Models in the Social Sciences*. New York: Seminar Press
Fisher RA. 1935. *The Design of Experiments*. London: Oliver & Boyd
Frangakis CE, Rubin D. 2002. Principal stratification in causal inference. *Biometrics* 58(1):21–29
Frisch R. 1930. *A dynamic approach to economic theory: lectures by Ragnar Frisch at Yale University*. Mimeogr., Frisch Arch., Dep. Econ., Univ. Oslo, Oslo, Nor.
Frisch R. 1995 (1938). Autonomy of economic relations: statistical versus theoretical relations in economic macrodynamics. In *The Foundations of Econometric Analysis*, ed. DF Hendry, MS Morgan, pp. 407–24. Cambridge, UK: Cambridge Univ. Press
Glymour C, Scheines R, Spirtes P. 2014. *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modeling*. New York: Academic
Goldberger AS. 1972. Structural equation methods in the social sciences. *Econometrica* 40(6):979–1001
Greenland S, Pearl J, Robins J. 1999. Causal diagrams for epidemiologic research. *Epidemiology* 10(1):37–48
Haavelmo T. 1943. The statistical implications of a system of simultaneous equations. *Econometrica* 11(1):1–12
Heckman JJ. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46(4):931–59
Heckman JJ. 1979. Sample selection bias as a specification error. *Econometrica* 47(1):153–62
Heckman JJ. 2008a. Econometric causality. *Int. Stat. Rev.* 76(1):1–27
Heckman JJ. 2008b. The principles underlying evaluation estimators with an application to matching. *Ann. Econ. Stat.* 91–92:9–73
Heckman JJ, Ichimura H, Smith J, Todd PE. 1998. Characterizing selection bias using experimental data. *Econometrica* 66(5):1017–98
Heckman JJ, LaLonde RJ, Smith JA. 1999. The economics and econometrics of active labor market programs. In *Handbook of Labor Economics*, Vol. 3A, ed. OC Ashenfelter, D Card, pp. 1865–2097. New York: North-Holland
Heckman JJ, Leamer EE, eds. 2007. *Handbook of Econometrics*, Vol. 6B. Amsterdam: Elsevier
Heckman JJ, Navarro S. 2004. Using matching, instrumental variables, and control functions to estimate economic choice models. *Rev. Econ. Stat.* 86(1):30–57
Heckman JJ, Pinto R. 2015. Causal analysis after Haavelmo. *Econom. Theory* 31(1):115–51
Heckman JJ, Pinto R. 2018. Unordered monotonicity. *Econometrica* 86(1):1–35
Heckman JJ, Pinto R, Savelyev P. 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *Am. Econ. Rev.* 103(6):2052–86
Heckman JJ, Robb R. 1985a. Alternative methods for evaluating the impact of interventions. In *Longitudinal Analysis of Labor Market Data*, Vol. 10, ed. JJ Heckman, BS Singer, pp. 156–245. Cambridge, UK: Cambridge Univ. Press
Heckman JJ, Robb R. 1985b. Alternative methods for evaluating the impact of interventions: an overview. *J. Econom.* 30(1–2):239–67
Heckman JJ, Taber C. 2008. The Roy model. In *New Palgrave Dictionary of Economics*, Vol. 7, ed. SN Durlauf, LE Blume, pp. 248–53. Basingstoke, UK: Palgrave Macmillan. 2nd ed.
Heckman JJ, Urzúa S, Vytlacil EJ. 2006. Understanding instrumental variables in models with essential heterogeneity. *Rev. Econ. Stat.* 88(3):389–432
Heckman JJ, Urzúa S, Vytlavil EJ. 2008. Instrumental variables in models with multiple outcomes: the general unordered case. *Ann. Econ. Stat.* 91–92:151–74
Heckman JJ, Vytlacil EJ. 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *PNAS* 96(8):4730–34

Heckman JJ, Vytlacil EJ. 2005. Structural equations, treatment effects and econometric policy evaluation. *Econometrica* 73(3):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 economic estimators to evaluate social programs, and to forecast their effects in new environments. See Heckman & Leamer 2007, pp. 4875–5143
Holland PW. 1986. Statistics and causal inference. *J. Am. Stat. Assoc.* 81(396):945–60
Hoyer PO, Janzing D, Mooij JM, Peters J, Schölkopf B. 2009. Nonlinear causal discovery with additive noise models. In *Advances in Neural Information Processing Systems 21*, ed. D Koller, D Schuurmans, Y Bengio, L Bottou, pp. 689–96. San Diego, CA: NIPS
Huang Y, Valtorta M. 2006. Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pp. 217–24. Arlington, VA: AUAI Press
Hurwicz L. 1962. On the structural form of interdependent systems. In *Logic, Methodology and Philosophy of Science*, ed. E Nagel, P Suppes, A Tarski, pp. 232–39. Stanford, CA: Stanford Univ. Press
Imbens GW. 2004. Nonparametric estimation of average treatment effects under exogeneity: a review. *Rev. Econ. Stat.* 86(1):4–29
Imbens GW, Angrist JD. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62(2):467–75
Imbens GW, Rubin DB. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge, UK: Cambridge Univ. Press
Kiiveri H, Speed TP, Carlin JB. 1984. Recursive causal models. *J. Aust. Math. Soc. Ser. A* (36):30–52
Lauritzen SL. 1996. *Graphical Models*. Oxford, UK: Clarendon
Lee S, Salanié B. 2018. Identifying effects of multivalued treatments. *Econometrica* 86(6):1939–63
Lopez-Paz D, Nishihara R, Chintala S, Schölkopf B, Bottou L. 2017. Discovering causal signals in images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6979–89. New York: IEEE
Matzkin RL. 1993. Nonparametric identification and estimation of polychotomous choice models. *J. Econom.* 58(1–2):137–68
Matzkin RL. 1994. Restrictions of economic theory in nonparametric methods. In *Handbook of Econometrics*, Vol. 4, ed. R Engle, D McFadden, pp. 2523–58. New York: North-Holland
Matzkin RL. 2007. Nonparametric identification. See Heckman & Leamer 2007, pp. 5307–68
Matzkin RL. 2008. Identification in nonparametric simultaneous equations models. *Econometrica* 76(5):945–78
Matzkin RL. 2013. Nonparametric identification of structural economic models. *Annu. Rev. Econ.* 5:457–86
Matzkin RL. 2015. Estimation of nonparametric models with simultaneity. *Econometrica* 83(1):1–66
Mogstad M, Santos A, Torgovitsky A. 2018. Using instrumental variables for inference about policy relevant treatment effects. *Econometrica* 86(5):1589–619
Mogstad M, Torgovitsky A. 2018. Identification and extrapolation of causal effects with instrumental variables. *Annu. Rev. Econ.* 2:577–613
Neyman J. 1923. Statistical problems in agricultural experiments. *J. R. Stat. Soc.* 2(Suppl.):107–80
Pearl J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann
Pearl J. 1993. Bayesian analysis in expert systems. Comment: graphical models, causality and intervention. *Stat. Sci.* 8(3):266–69
Pearl J. 1995. Causal diagrams for empirical research. *Biometrika* 82(4):669–88
Pearl J. 2009a. *Causality: Models, Reasoning, and Inference*. New York: Cambridge Univ. Press. 2nd ed.
Pearl J. 2009b. *Myth, confusion, and science in causal analysis.* Tech. Rep., Dep. Stat., Univ. Calif., Los Angeles
Pearl J. 2012. The do-calculus revisited. arXiv:1210.4852 [cs.AI]
Peters J, Jazzing D, Schölkopf B. 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge, MA: MIT Press
Pinto R. 2016. *Learning from noncompliance in social experiments: the case of moving to opportunity*. Unpubl. Ms., Dep. Econ., Univ. Chicago, Chicago

Powell JL. 1994. Estimation of semiparametric models. In *Handbook of Econometrics*, Vol. 4, ed. R Engle, D McFadden, pp. 2443–521. Amsterdam: Elsevier
Pratt JW, Schlaifer R. 1984. On the nature and discovery of structure. *J. Am. Stat. Assoc.* 79(385):9–33
Quandt RE. 1988. *The Econometrics of Disequilibrium*. New York: Blackwell
Robins J. 1986. A new approach to causal inference in mortality studies with a sustained exposure period: application to control of the healthy worker survivor effect. *Math. Model.* 7(9–12):1393–512
Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55
Roy A. 1951. Some thoughts on the distribution of earnings. *Oxf. Econ. Pap.* 3(2):135–46
Rubin DB. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66(5):688–701
Schennach SM. 2020. Mismeasured and unobserved variables. In *Handbook of Econometrics*, Vol. 7A, ed. S Durlauf, L Hansen, J Heckman, R Matzkin, pp. 487–565. Amsterdam: Elsevier
Shpitser I, Pearl J. 2006. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference*, pp. 1219–26. Menlo Park, CA: AAAI
Shpitser I, Pearl J. 2009. Effects of treatment on the treated: identification and generalization. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, pp. 514–21. Arlington, VA: AUAI Press
Spiegler R. 2020. Behavioral implications of causal misperceptions. *Annu. Rev. Econ.* 12:81–106
Tamer E. 2003. Incomplete simultaneous discrete response model with multiple equilibria. *Rev. Econ. Stud.* 70(1):147–65
Telser LG. 1964. Iterative estimation of a set of linear regression equations. *J. Am. Stat. Assoc.* 59(307):845–62
Theil H. 1953. *Esimation and simultaneous correlation in complete equation systems.* Mimeogr. Memo., Cent. Plan. Bur., The Hague, Neth.
Tikka S, Karvanen J. 2017. Simplifying probabilistic expressions in causal inference. *J. Mach. Learn. Res.* 18(1):1–30
Vytlacil EJ. 2002. Independence, monotonicity, and latent index models: an equivalence result. *Econometrica* 70(1):331–41
Wooldridge JM. 2015. Control function methods in applied econometrics. *J. Hum. Resour.* (50):420–45

Annual Review of Economics

Volume 14, 2022

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Economics* articles may be found at http://www.annualreviews.org/errata/economics