The Econometric Model for Causal Policy Analysis

James Heckman and Rodrigo Pinto*

Revised April 21, 2022

Abstract

This paper 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, resulting 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 alternative approaches.

Key words: Policy Analysis, econometric models, Causality, Identification, Causal Calculus, Directed Acyclic Graphs, Simultaneous Treatment Effects

JEL codes: C10, C18

*James Heckman is the Henry Schultz Distinguished Service Professor of Economics and Public Policy at the University of Chicago; and Director of the Center for Economics of Human Development. Rodrigo Pinto is an Assistant Professor at the Economics Department at the University of California at Los Angeles (UCLA). This research was supported in part by NIH grant NICHD R37HD065072. The views expressed in this paper 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.

Rodrigo Pinto
University of California at Los Angeles
Department of Economics
315 Portola Plaza, Room 8385
Los Angeles, CA 90095
(310) 825-0849
rodrig@econ.ucla.edu

James Heckman
The University of Chicago
Department of Economics
1126 E. 59th St.
Chicago, IL 60637
(773) 702-0634
jjh@uchicago.edu

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. To provide good policy advice, it uses all information available. It explores all possible counterfactual scenarios. It's grounded in thought experiments, which 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. They 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 the economic causal analysis. They use proprietary languages and unconventional methodologies to formalize the notion of causality introduced by early economists such as Haavelmo (1943) and Frisch (1930).

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 though. The three distinct causal tasks of defining a causal model, establishing its identification and the empirical estimation from data are frequently intertwined. It's 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 isn't 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 (1923; 1958; 1974; 1986; 1996) which is motivated by the notion of randomized control trials, as an ideal framework for causality. We also investigate a less popular causal framework advocated by computer scientists called "*do-calculus*" (Pearl, 1995, 2012), which makes extensive use of directed acyclic graphs (DAGs).[1]

Each of these approaches is well suited to investigate a subset of problems in causal inference. They 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 (hence 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 do-calculus is incompatible with several well-established identification techniques commonly used in econometric evaluations. These frameworks hinder the access of their users to critical components of the econometric toolset, including social interactions, peer effects, and general equilibrium theory.

We adopt the Generalized Roy Model as the leading example in our discussions. We contrast the versatility of the econometric approach to causality with the NR and the do-calculus. We also discuss the class of inference problems for which these recent methodologies perform well.

This paper is organized as following. 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 Neyman-Rubin causal model approximates the econo-

[1]See Spiegler (2020) for a range of clever applications of DAGs in economics.

metric model. Section 6 investigates how the do-calculus of Pearl (2009a) approximates the econometric model. Section 7 summarizes the paper.

2 Econometric Causality

The modern understanding of causal inference dates back to the 1930 lectures of Ragnar Frisch, who describes causality as a thought experiment in which the researcher *hypothetically* manipulates one or more inputs affecting an output. 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, 1938). Autonomy means that the input-output *relationship* is invariant for different input manipulations.[3] It is also called structural (Hurwicz, 1962). It postulates that the map $g : X \to Y$ between inputs 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 void of causal meaning.

To fix ideas, 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 a randomized control trial, U stands for an unobserved variable while T is a treatment indicator that takes value $T = 1$ for treated and $T = 0$ for control.

The causal effect of T on Y is described by the thought experiment that manipulates

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

[3] (Frisch, 1938) describes autonomous equations as deterministic functions that are "invariant" to changes in their arguments (Frisch, 1938). Hurwicz (1962) prefers the term "structural" to denote autonomous equations.

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 its estimation.

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 is expected value of treatment effects across the entire population $\mathcal{I}$, $ATE = E_{\mathcal{I}}(Y_i(1) - Y_i(0))$, which can also be expressed as the integral of the difference of counterfactuals over the support of the unobserved variable U :

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

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 $E(Y|T = t)$:

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

where the second equality is due to $T \perp\!\!\!\perp U$. ATE is identified by the difference in means between treated and control participants and can be estimated by a linear regression model.

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 toolset.

The OLS model arises by assuming that function g is linear:

$$Y = \alpha + \beta T + U. \tag{4}$$

The exogeneity of T and U implies that U and T do not cause each other. In this case, the linear assumption (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 whether the OLS estimator is biased is a statistical assessment that is separate from causal inquiry of whether β in (4) is the causal effect of T on Y. Table 1 is useful in clarifying the difference.

Table 1: **Examples of Causal Relationships between T and U**

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

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 the an unobserved confounding variable V. Parameter β in (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 since 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. Treatment T causes U in Model 3 and the

[4]For an example of how confusing this concept is to statisticians, see Pratt and Schlaifer (1984).

parameter β in (4) is not the cause effect of T on Y anymore.

2.1 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, which is grounded in scientific theory.

2. The second task regards the identification of causal parameters. The framework must offer mathematical tools that enable the analysis to manipulate causal concepts and investigate the identification of causal parameters.

3. The last 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 and model of agent interactions.

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 the 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: we may change the arguments, but the equations are unchanged. 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 $\epsilon_K \in \mathcal{E}$ that shares the same dimension of K. 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 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 $(\mathcal{I}, \mathcal{F}, P)$.

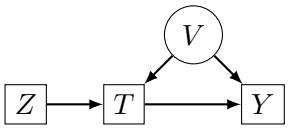
We adopt the Generalized Roy model as our leading example. The model is a cornerstone in the literature of policy evaluation (Heckman and Taber, 2008). The original model (1951) investigated the occupation choice of an economic agent that decides between two economic sectors based on the perceived difference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

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

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 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 external policy vector. The confounding variable V accounts for 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.

Table 2: **Representations of the Generalized Roy Model**

	Variable Map	**Structural Eq.**	**DAG**	**LMC**
Z	$\mathbb{M}(Z) = \varnothing$	$Z = f_Z(\epsilon_Z)$		$Z \perp\!\!\!\perp V \vert \varnothing$
V	$\mathbb{M}(V) = \varnothing$	$V = f_V(\epsilon_V)$		$V \perp\!\!\!\perp Z \vert \varnothing$
T	$\mathbb{M}(T) = \{Z, V\}$	$T = f_T(Z, V, \epsilon_T)$		$T \perp\!\!\!\perp \varnothing \vert (Z, V)$
Y	$\mathbb{M}(Y) = \{T, V\}$	$Y = f_Y(T, V, \epsilon_Y)$		$Y \perp\!\!\!\perp Z \vert (T, V)$

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 Directed Acyclic Graph (DAG), where arrows denote causal relationships, circles denote unobserved variables, and squares denote observed variables.[9]

The last representation of Table 2 employs a property called the Local Markov Condition

[6]For instance, Heckman and Vytlacil (2007a) investigate multiple variations of the original model, Heckman et al. (2008) extend the model for ordered choice models and Heckman and Pinto (2018) and Lee and Salanié (2018) investigate the case of unordered multiple choice models with multi-valued treatments. Abbring and 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 behavior, 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 to Spiegler (2020) and Lauritzen (1996) for information on DAGs and Bayesian Networks.

(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 $\mathbb{M}(K)$. Children of K comprises the variables directly caused by K, namely, $\mathbb{Ch}(K) = \{J \in \mathcal{T}; K \in \mathbb{M}(J)\}$. The descendants of a variable K, $\mathbb{D}(K)$, include all 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 non-descendants conditioned on its parents.

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

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 last row of Table 2. Z has no parents and its descendants are T, Y. The set of LMC for all variables in $\mathcal{T}$ fully characterizes the causal model. Additional independence relationships may be generated by the Graphoid Axioms[12] of Dawid (1976) or though graphical method such as the d-separation criteria of Geiger et al. (1990).

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 can be found in Spiegler (2020). He employs DAGs to model causal misconceptions in

[10] See Kiiveri et al. (1984); Pearl (1988) for further information on the Local Markov Condition.

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

[12] Dawid (1976) defines Graphoid Axioms consist of six rules that apply for any disjoint sets of variables $X, W, Z, Y \subseteq \mathcal{T}$:

(A)Symmetry:	$X \perp\!\!\!\perp Y \mid Z \Rightarrow Y \perp\!\!\!\perp X \mid Z.$
(B)Decomposition:	$X \perp\!\!\!\perp (W, Y) \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid Z.$
(C)Weak Union:	$X \perp\!\!\!\perp (W, Y) \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid (W, Z).$
(D)Contraction:	$X \perp\!\!\!\perp W \mid (Y, Z)$ and $X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp (W, Y) \mid Z.$
(E)Intersection:	$X \perp\!\!\!\perp W \mid (Y, Z)$ and $X \perp\!\!\!\perp Y \mid (W, Z) \Rightarrow X \perp\!\!\!\perp (W, Y) \mid Z.$
(F)Redundancy:	$X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid Z.$

decision-making and demonstrates 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 by 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 structural equations of the variables directly caused by T. The do-operator differs from fixing by deleting ("shutting down") the structural equation for the treatment variable T, which effectively suppresses the random variable T from the causal model.

Neither *fix* nor the *do* operator are 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 and Pinto (2015) develop a causal framework that expresses the ill-defined causal operations of *fixing* or *doing* into 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 determining

causality.

The hypothetical model formalizes Frisch's insight on causality. It is an abstract model that shares the same structural equations and the same distributions of error terms of the empirical model. It differs from the empirical model by appending 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 $\tilde{T}$.

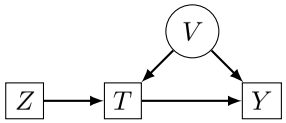
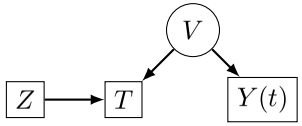
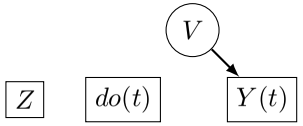
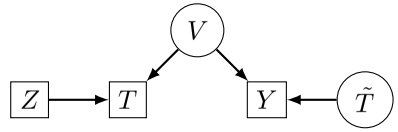
We exemplify the hypothetical model using the Generalized Roy Model. For notational clarity, we use $\mathbb{M}_e$ for the empirical (original) model, $\mathbb{M}_{fix}$ for the model that applies the *fix*-operator, $\mathbb{M}_{do}$ for the *do*-operator, and $\mathbb{M}_h$ for the hypothetical model.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

The first column of Table 3 presents the original empirical model. The second and third columns present the models generated by the *fix* and the *do* operators respectively. Both models constraint 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 while *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 $\epsilon_Z, \epsilon_T, \epsilon_V, \epsilon_Y$. Therefore the joint distribution of non-descendant 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* models 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 (5), it is independent of all its non-descendants, $\tilde{T} \perp\!\!\!\perp (T, V, Z)$. In particular, $\tilde{T} \perp\!\!\!\perp T$ always hold for any hypothetical

Table 3: Generalized Roy Model: Approaches to Generating Counterfactuals

	Empirical Models			**Hypothetical Model**
	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 : Z : T : Y : $\tilde{T}$:	$V = f_V(\epsilon_V)$ $Z = f_Z(\epsilon_Z)$ $T = f_T(Z, V, \epsilon_T)$ $Y = f_Y(T, V, \epsilon_Y)$	$V = f_V(\epsilon_V)$ $Z = f_Z(\epsilon_Z)$ $T = f_T(Z, V, \epsilon_T)$ $Y(t) = f_Y(t, V, \epsilon_Y)$	$V = f_V(\epsilon_V)$ $Z = f_Z(\epsilon_Z)$ $do(T = t)$ $Y(t) = f_Y(t, V, \epsilon_Y)$	$V = f_V(\epsilon_V)$ $Z = f_Z(\epsilon_Z)$ $T = f_T(Z, V, \epsilon_T)$ $Y = f_Y(\tilde{T}, V, \epsilon_Y)$ $\tilde{T} = f_{\tilde{T}}(\epsilon_{\tilde{T}})$
	Directed Acyclic Graphs (DAGs)			
				
	Local Markov Conditions			
V : Z : T : Y : $\tilde{T}$:	$V \perp\!\!\!\perp Z$ $Z \perp\!\!\!\perp V$ $T \perp\!\!\!\perp \varnothing \mid (Z, V)$ $Y \perp\!\!\!\perp Z \mid (T, V)$ (not defined for the model)	$V \perp\!\!\!\perp Z$ $Z \perp\!\!\!\perp (V, Y(t))$ $T \perp\!\!\!\perp Y(t) \mid (Z, V)$ $Y(t) \perp\!\!\!\perp (Z, T) \mid V$ (not defined for the model)	$V \perp\!\!\!\perp Z$ $Z \perp\!\!\!\perp (V, Y(t))$ (not defined for the model) $Y(t) \perp\!\!\!\perp Z \mid V$ (not defined for the model)	$V \perp\!\!\!\perp (Z, \tilde{T})$ $Z \perp\!\!\!\perp (V, Y, \tilde{T})$ $T \perp\!\!\!\perp (\tilde{T}, Y) \mid (Z, V)$ $Y \perp\!\!\!\perp (Z, T) \mid (\tilde{T}, V)$ $\tilde{T} \perp\!\!\!\perp (T, V, Z)$
	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 e denotes empirical (original) model. Subscript fix denotes the model that uses the fix operator, that is when treatment T is fixed to t. Subscript do denotes the model that employs the do-operator. Subscript h denotes the hypothetical model.

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:

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

It is also the case that equation (6) holds when conditioned on any variable K that is non-descendant variable of $\tilde{T}$:

$$\Big(Y \mid \tilde{T} = t, K\Big)_{\mathbb{M}_h} \stackrel{d}{=} \Big(Y(t) \mid K\Big)_{\mathbb{M}_{fix}}, \tag{7}$$

and (8)

$$\Big(Y \mid \tilde{T} = t, \{K \setminus \{T\}\}\Big)_{\mathbb{M}_h} \stackrel{d}{=} \Big(Y(t) \mid \{K \setminus \{T\}\}\Big)_{\mathbb{M}_{do}}. \tag{9}$$

To fix ideas, let T be an indicator of college graduation and Y denote adult income. Treatment-on-the-treated (TOT) is the average causal effect of college on income by those who choose to go to college ($T = 1$), which is commonly described as $TOT = E_{fix}(Y(1) - Y(0) \mid T = 1)$ using the fix operator. The parameter is equivalently described as $TOT = E_h(Y \mid \tilde{T} = 1, T = 1) - E_h(Y \mid \tilde{T} = 0, T = 1)$ using the hypothetical model. The *do* operator *excludes* the treatment variable T, which poses a serious challenge in defining the TOT parameter. Shpitser and Pearl (2009) solve this issue by adding additional 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 *do*-calculus 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 $\mathbb{M}_e$. 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 establish this connection.

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

$$Y(t) \perp\!\!\!\perp T|V. \tag{10}$$

Equation (10) 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 to connect the counterfactual outcome $Y(t)$ in $\mathbb{M}_{fix}$ with the empirical model $\mathbb{M}_e$:

$$\begin{aligned}
P_{fix}\big(Y(t) \mid V\big) &= P_{fix}\big(Y(t) \mid V, T = t\big), && (11)\\
&= P_{fix}\left(\sum_{t\in \text{supp}(T)} \mathbf{1}[T = t]Y(t) \mid V, T = t\right), && (12)\\
&= P_{fix}\left(\sum_{t\in \text{supp}(T)} \mathbf{1}[T = t]f_Y(t, V, \epsilon_Y) \mid V, T = t\right), && (13)\\
&= P_{fix}\big(f_Y(T, V, \epsilon_Y) \mid V, T = t\big), && (14)\\
&= P_e\big(Y \mid V, T = t\big). && (15)
\end{aligned}$$

Equations (11)–(15) use the structural equations to express the probability distribution of the counterfactual outcome $Y(t)$ in $\mathbb{M}_{fix}$ with the distribution of the outcome Y in empirical model $\mathbb{M}_e$. The first equation (11) is due to the matching condition (10). Equations (11)–(14) apply the definition of the structural equations. The last equation (15) uses the fact that variables T, V, ϵ_Y share the same distribution in both models $\mathbb{M}_{fix}$ and $\mathbb{M}_e$.

The hypothetical model $\mathbb{M}_h$ 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:[13]

$$Y \perp\!\!\!\perp \widetilde{T} \mid (T, W) \Rightarrow P_h\Big(Y \mid \widetilde{T} = t, T = t', W\Big) = P_h(Y \mid T = t', W) = P_e(Y \mid T = t', W). \quad (16)$$

$$Y \perp\!\!\!\perp T \mid (\widetilde{T}, W) \Rightarrow P_h\Big(Y \mid \widetilde{T} = t, T = t', W\Big) = P_h(Y \mid \widetilde{T} = t, W) = P_e(Y \mid T = t, W). \quad (17)$$

Equations (16)-(17) state that we can switch from the hypothetical to the empirical model whenever the independence relationships (16): $Y \perp\!\!\!\perp \widetilde{T} \mid (T, W)$ or (17): $Y \perp\!\!\!\perp T \mid (\widetilde{T}, W)$ hold.[14] The LMC of Y in $\mathbb{M}_h$ generates the following matching condition:

$$Y \perp\!\!\!\perp T | (\tilde{T}, V). \quad (18)$$

Thus, according to (17), we have that $P_h(Y \mid \widetilde{T} = t, V) = P_e(Y \mid T = t, V)$.

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

The *do* operator does not generate a matching conditions such as (10) or (18) because the treatment T is absent. Instead, the do-calculus 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 may sound obscure to most economists. The criterion is part of the do-calculus, 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 limited in the sense that it does not accept identifying assumptions outside the DAG terminology. We discuss the do-calculus machinery, its benefits and limitations in Section 6. For now, we use the back-door criterion to introduce some of

[13] See Heckman and Pinto (2015) for a proof. The criteria (16)–(17) still holds if the values $t, t' \in \text{supp}(T)$ were replaced by subsets $\mathcal{A}, \mathcal{A}' \subset \text{supp}(T)$ respectively.

[14] See Heckman and Pinto (2015) for a discussion on the connection between empirical and hypothetical models.

the do-calculus jargon.

Let G be the original DAG that represents the empirical model and $G_{\underline{T}}$ be DAG that suppresses the arrows departing from T. In the do-calculus, 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 $\mathbb{M}_{fix}, \mathbb{M}_h$ and $\mathbb{M}_{do}$ 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:

$$P_{fix}(Y(t) \mid V) = P_h(Y \mid \widetilde{T} = t, V) = P_{do}(Y(t)|V) = P_e(Y \mid T = t, V). \tag{19}$$

If V were observed, we would be able to evaluate the expected value of the counterfactual outcome expectation, $E_h(Y \mid \widetilde{T} = t)$, by integrating the observed expectation $E_e(Y \mid T = t, V)$ 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 this menu 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

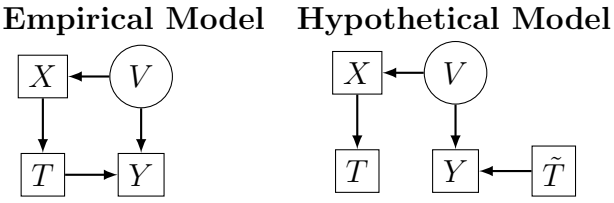
constraints. We mention only a limited share of the available identification strategies of this literature.

4.1 Matching on Observables

A popular method for identifying treatment effects is to assume that a set of observed pre-treatment variables X suffices to control for the confounding variable V. Otherwise stated, it assumes that the observed variable X is a balancing score for the confounding variable V.[15] This assumption if often referred as *matching on observables*, *unconfoundness*, *ignorability* or simply *exogeneity assumption* (Heckman et al., 1998; Imbens, 2004; Rosenbaum and Rubin, 1983).

Table 4 presents the corresponding empirical and the hypothetical models corresponding to matching on observables. The LMC of T in the hypothetical model implies that $Y \perp\!\!\!\perp \tilde{T} \mid (T, X)$, or equivalently, $Y(t) \perp\!\!\!\perp T \mid X$. This matching condition enables to identify the counterfactual outcome by conditioning on X, $P_h(Y|\tilde{T} = t, X) = P_{fix}(Y(t)|X) = P_e(Y \mid T = t, X = x)$. The common support (overlap) assumption, $0 < P_e(T = t|X) < 1; t \in \text{supp}(T)$, enables us to identify the expected value of the counterfactual outcome as $E_{fix}(Y(t)) = \int E_e(Y \mid T = t, X = x) dF_{e,X}(x)$.[16]

Table 4: Matching Model: Empirical and Hypothetical Causal Models



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

[15] By pre-treatment variables X we mean variables that are not descendants of the treatment variable T.
[16] See Heckman et al. (1998); Imbens (2004) for estimation methods invoking matching on observables.

of randomized controlled trials (RCTs), where X denotes the pre-treatment variables used in the randomization protocol. The assumption is rather strong in observational studies and often criticized (Heckman and Navarro, 2004). It assumes that potential bias generated by confounding variables can be ignored when controlling for observed pre-treatment variables (Heckman, 2008b). Matching on observables does not require using the instrumental variable to identify causal effects. Instead, it solves the problem of selection bias by assuming it does not exist or is of minor concern. The reaming strategies are more appealing.

4.2 Instrumental Variables

The simplest identifying assumption 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 $cov(Y, Z)/cov(T, Z)$ 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. It rules out the essential heterogeneity across agents (Heckman et al., 2006). The local Instrumental Variable (LIV) Model of Heckman and Vytlacil (1999, 2005) addresses this problem.

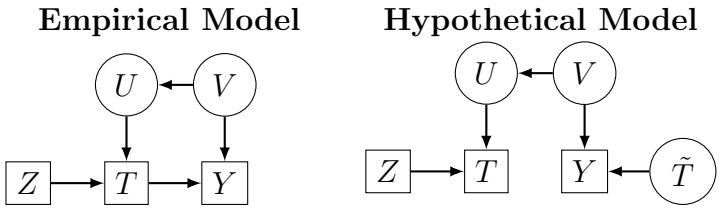
The LIV model considers a binary treatment $T \in \{0, 1\}$ determined by a latent threshold-crossing equation that is *separable* in the instrument Z and the confounder V, that is, $T = \mathbf{1}[\zeta(Z) \geq \phi(V)]$. This *separability assumption* arises from economic choice theory and enables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

$$T = \mathbf{1}\big[P(Z) \geq U\big], \tag{20}$$

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

Table 5 displays the hypothetical model of LIV where $Y \perp\!\!\!\perp T \mid (\tilde{T}, U)$ holds.[17] Thus U is a matching variable and plays role V in (18). U is called a *balancing* score for V, which means that U is a surjective function of V that preserves the independence relationship $Y \perp\!\!\!\perp T \mid (\tilde{T}, V) \Rightarrow Y \perp\!\!\!\perp T \mid (\tilde{T}, U)$.[18] Controlling for U identifies counterfactual outcomes in the same fashion that controlling for V in (19) does, that is, $P_h(Y|\widetilde{T} = t, U) = P_{fix}(Y(t)|U) = P_e(Y|T = t, U)$.

Table 5: Empirical and Hypothetical Causal Models for LIV



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

$$E_h(Y|\tilde{T} = 1, U = u) - E_h(Y|\tilde{T} = 0, U = u) = E_{fix}(Y(1) - Y(0) \mid U = u) = \frac{\partial E_e(Y|P(Z))}{\partial P(Z)}\bigg|_{P(Z)=u}. \tag{21}$$

Identification 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 $ATE \equiv E_h(Y|\tilde{T} = 1) - E_h(Y|\tilde{T} = 0) = \int_0^1 \big(E_h(Y \mid T = 1, U = u) - E_h(Y \mid T = 0, U = u)\big)du$. For categorical instruments, the discrete counterpart of equation (21) states that for any two values $z, z' \in$

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

[18] The term balancing score was introduced by Rosenbaum and Rubin (1983).

$\text{supp}(Z)$ such that $P(z') = u' > u = P(z)$ we have that:

$$\frac{E_e(Y|Z=z') - E_e(Y|Z=z)}{P_e(T=1 \mid Z=z') - P_e(T=1 \mid Z=z)} = \frac{\int_u^{u'} E_{fix}(Y(1)-Y(0)|U=u)du}{u'-u} \tag{22}$$

$$= E_{fix}(Y(1) - Y(0) \mid u \leq U \leq u'). \tag{23}$$

Equation (22) explains that the Local Average Treatment Effects (LATE) of Imbens and Angrist (1994) (left-hand side) identifies the expected value of the counterfactual outcome $E_{fix}(Y(1) - Y(0)|U = u)$ over an interval of U defined by the propensity scores $P(z)$ and $P(z')$ in (23).[19][20] The method of randomized controlled trial (RCT) can be understood as a 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 (22)–(23) identify the average treatment effect:

$$\frac{E_e(Y|Z=z_1) - E_e(Y|Z=z_0)}{1-0} = E_{fix}(Y(1) - Y(0) \mid 0 \leq U \leq 1) = ATE.$$

4.3 Stratification

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

$$\boldsymbol{S} = [T(z_1), ..., T(z_{N_Z})]'. \tag{24}$$

The values that the response vector $\boldsymbol{S}$ takes are called response-types.[21] In the case of a binary instrument $Z \in \{z_0, z_1\}$ and a binary treatment $T \in \{0, 1\}$, the response vector $\boldsymbol{S} = [T(z_0), T(z_1)]'$ takes four possible response-types $\text{supp}(\boldsymbol{S}) = \{(0,0), (0,1), (1,0), (1,1)\}$ which are termed always-takers, compliers, never-takers and deniers by Imbens and Angrist

[19] Heckman et al. (2008) investigate the relationship between LIV and LATE in greater detail.

[20] Mogstad and Torgovitsky (2018) use functional form assumptions to extrapolate the evaluation of LATE-parameters beyond the U-interval defined by propensity scores.

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

(1994).

The response vector $\boldsymbol{S}$ is a balancing score for V in the same fashion that U in (20) is. $\boldsymbol{S}$ is a function of V because counterfactual choices $T(z) = f_T(z, V); z \in \text{supp}(Z)$ are a function of V. Choice T can be expressed as $T = \left[\mathbf{1}[Z = z_1], ..., \mathbf{1}[Z = z_N]\right] \cdot \boldsymbol{S}$. Thus, given $\boldsymbol{S}$, the choice T is a function of Z, which is independent of V and $Y(t)$. Therefore $Y(t) \perp\!\!\!\perp T|\boldsymbol{S}$ holds.

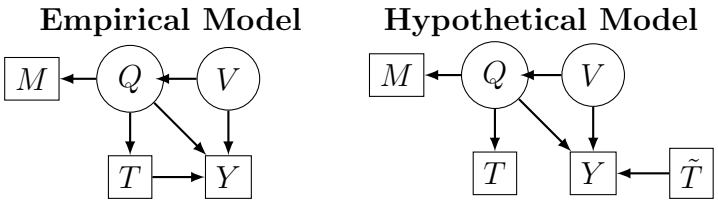
Heckman and Pinto (2018) show that the response vector $\boldsymbol{S}$ generates a special partition of the support of V that and renders T statistically independent of V within each partition set. They present the necessary and sufficient conditions for the identification of counterfactual outcomes, 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 observed data. Indeed, while the number of response-types grows exponentially in N_Z, observed data grows linearly in N_Z. Identification requires choice restrictions that systematically reduce the number of admissible response-types. Such restrictions can be obtained though monotonicity conditions or by applying revealed preference analysis (Buchinsky and Pinto, 2021; Heckman and Pinto, 2018; Pinto, 2016).

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: (1) Q which is a balancing score for the confounder V, and (2) Q measured with error by the observed variables M. The corresponding empirical and hypothetical causal models are displayed in Table 6. The matching condition $Y \perp\!\!\!\perp T|(\tilde{T}, Q)$ holds and Q is a balancing score for V.

Table 6: Matching on Proxied Unobservables: Empirical and Hypothetical Causal Models



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

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

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, agreeableness and agreeableness. These abilities are not directly observed, but 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.

4.5 Control Function

The control function approach was introduced in Heckman and Robb (1985b) building on earlier work by Telser (1964) and has been examined by several authors Ahn and Powell (1993); Blundell and Powell (2003); Powell (1994); 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 the Heckman's sample selection correction (1979). For a broader survey of identification of nonlinear models with endogenous variables, see Matzkin (1994, 2007).

One important class of structural equations in the control function literature are the outcomes that are *additively separable* from the error term. We illustrate the control function approach using a version of the Roy model with binary choices that relies on the additively separability of the outcome equation. We can express the observed outcome Y in terms of counterfactual outcomes $Y(t) = f_T(t, V, \epsilon_Y); t \in \{0, 1\}$ as:

$$Y = Y(0) \cdot \mathbf{1}[T = 0] + Y(1) \cdot \mathbf{1}[T = 1]. \tag{25}$$

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:

$$Y_1 = \mu_1 + U_1, \quad Y_0 = \mu_0 + U_0, \quad T = \mathbf{1}[P(Z) \geq U], \tag{26}$$

where the last equation is due to the separability assumption on the choice equation which renders $U \sim unif[0, 1]$. The unobserved confounding variable U causes U_1, U_0 and 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 instrumental variable Z and the choice $T \in \{0, 1\}$ as:

$$\begin{aligned} E_e\big(Y \mid Z, T = 0\big) &= \mu_0 + E_e\big(U_0 | Z = z, T = 0\big), && (27)\\ &= \mu_0 + E_e\big(U_0 | Z = z, P(Z) < U\big), && (28)\\ &= \mu_0 + E_e\big(U_0 | P(Z) < U\big), && (29)\\ &= \mu_0 + K_0(P(Z)). && (30) \end{aligned}$$

where the first equation uses the separability of the outcome equation in (33). The second uses the fact that the event $T = 0$ is equivalent to the event $P(Z) < U$. The third equation use the independence relationship $Z \perp\!\!\!\perp U_0$. The last equation expresses the expected value

of the endogenous error term U_0 as a control function of the propensity score $K_0(P(Z)) = E_e\big(U_0|P(Z) < U\big)$. 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.

The value of the control function $K_0(P(Z))$ at $P(Z) = 0$ is $K_0(0) = E_e\big(U_0|0 < U\big) = E_e(U_0) = 0$. 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 $P(Z)$ 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 $K_0(P) = \sum_{k=1}^{K} \beta_{k,0} P^k$ and let P_i be the estimated propensity score for agent i. We can then estimate the following outcome equation:

$$Y_i = \alpha_0 + \sum_{k=1}^{K} \beta_{k,0} P_i^k + \epsilon_i, \text{ for all } i \text{ such that } T_i = 0. \tag{31}$$

The counterfactual mean μ_0 is obtained by setting the propensity scores to zero and is estimated by the parameter α_0 in equation (31). We can apply the same rationale in equations (27)–(30) to the case of $T = 1$. The resulting control function is given by:

$$E_e\big(Y \mid Z, T = 1\big) = \mu_0 + K_1(P(Z)), \text{ where } K_1(P(Z)) = E_e\big(U_1|P(Z) \geq U\big). \tag{32}$$

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:

$$Y_i = \alpha_1 + \sum_{k=1}^{K} \beta_{k,1} P_i^k + \epsilon_i, \text{ for all } i \text{ such that } T_i = 1. \tag{33}$$

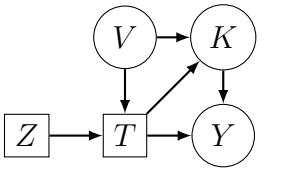
The counterfactual mean μ_1 is obtained by setting the propensity scores to one and is estimated by the linear combination $\alpha_1 + \sum_{k=1}^{K} \beta_{k,1}$ in equation (33).

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 outcome equation. Table 7 presents a modified version of the Roy model that can assist in gaining some intuition. 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. The technique to identify the causal effect of T on Y is based on a two-step procedure that first uses instrumental variable Z to evaluate the impact of T on K and then uses K to identify the joint impact of K and T on Y.

Table 7: Control Function: Empirical and Hypothetical Causal Models

Empirical Model



As previously said, our discussion on identification approach is be 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 survey are: panel data analysis,[22] difference-in-differences estimators,[23] regression discontinuity design,[24] simultaneous equation models,[25] and and non-separable equation models.[26]

5 The Neyman-Rubin (NR) Causal Model

The Neyman-Rubin (NR) causal model was developed by Fisher (1935); Neyman (1923) and further popularized by Cox (1958); 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 randomized control trials. It focuses on units of analysis instead of system of equations. Causal models are characterized by statistical independence relationships among counterfactual variables instead of structural equations.

[22] See, for instance, Heckman and Robb (1985a), Blundell et al. (1998).
[23] See Abadie (2005); Athey and Imbens (2006); Bertrand et al. (2004); Heckman et al. (1999).
[24] Regression discontinuity estimators, which are versions of IV estimators, are discussed by Heckman and Vytlacil (2007b).
[25] See Heckman (1978); Matzkin (2008, 2013, 2015); Quandt (1988); Tamer (2003).
[26] See Altonji and Matzkin (2005); Matzkin (1993).

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 characterises 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 $i \in \mathcal{I}$. (34)
2. *IV relevance:* Z is not statistically independent of T, that is $Z \not\perp\!\!\!\perp T$. (35)
3. *Exogeneity Condition:* $Z \perp\!\!\!\perp (Y(t), T(z))$. (36)

The NR model is widely popular due to its simplicity. It suppresses the structural equations that determine the causal relationship among variable models. Instead, the NR model describes the causal model through independence relationships among the counterfactual counterpart the *observed* variables of the causal model. It is easy to check that the structural equations of the Roy Model in Table 2 generate the IV-assumptions (34)–(36).[27] Otherwise stated, the IV-assumptions (34)–(36) 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 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. 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 assess-

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

ment of causal relationships between model variables. It prevents the use of a hypothetical model which elucidates the notion of causality. The absence of structural equations is particularly troublesome when judging the plausibility of 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: randomized controlled trials, matching on observables, IV models and its many surrogates (see Imbens and Rubin, 2015). This section illustrates some of the drawbacks of the NR framework in analyzing key economic models.

5.1 Revisiting the LIV Model under the NR framework

We examine the LIV model of Sction 4.2 from the lens of the NR framework. Consider a binary choice model $T \in \{0, 1\}$ for which the IV-assumptions (34)–(36) hold. These assumptions are not sufficient to identify causal effects. An additional assumption that secures the identification of the Local Average Treatment Effect (LATE) causal effects is the monotonicity condition of Imbens and Angrist (1994). The condition states that a change in the instrument induces agents to change their treatment choice towards the same direction. Notationally, for any $z, z' \in \text{supp}(Z)$, we have that:

$$T_i(z) \geq T_i(z') \ \ \forall i \in \mathcal{I} \quad \text{or} \quad T_i(z) \leq T_i(z') \ \ \forall i \in \mathcal{I} \tag{37}$$

Vytlacil (2002) shows that the monotonicity condition (37) is equivalent to the separability assumption $T = \mathbf{1}[P(Z) \geq U]$. Otherwise stated, the model generated by Monotonicity (37) and the IV-assumptions (34)–(36) is the NR-counterpart of the LIV model in Section 4.

Although both frameworks are suitable to investigate causal effects, the LIV model explicitly displays the unobserved confounding variable U while 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

Table 8: Some Causal Parameters as Weighted Average the MTE

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

and Vytlacil (2005):

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

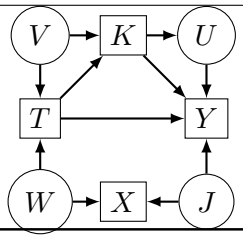
The MTE plays a primary role in characterizing a variety of causal effects that are typically sought in policy evaluations. Table 8 shows that causal effects can be expressed as weighted averages of the MTE. This result, in turn, foments additional literature on a variety of related questions regarding estimation, partial identification, extrapolation and inference of causal effects.[28] In conclusion, switching from the NR framework towards structural equations allows for deeper causal analyses that unlock additional research excursions.

5.2 Interpreting Matching on Observables

The most popular identification strategy in the NR framework is matching on observables of 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 pre-intervention variables.

[28] For examples of this literature, see Brinch et al. (2017); Heckman and Vytlacil (2007b); Mogstad et al. (2018)

Table 9: Hypothetical Matching Model

Causal Model	**DAG**	**Independence Relationships**
$V = f_V(\epsilon_V)$ $T = f_T(V, W, \epsilon_T)$ $K = f_K(T, V, \epsilon_K)$ $X = f_K(W, J, \epsilon_X)$ $Y = f_Y(T, K, U, J, \epsilon_Y)$ J, W, V, U are external variables		$Y(t) \perp\!\!\!\perp T \mid K$ $Y(t) \not\perp\!\!\!\perp T \mid X$ $Y(t) \not\perp\!\!\!\perp T \mid (X, K)$

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.[29] However it is rather difficult to investigate its falsehood using the NR framework. The causal model of Table 9 clarifies this point.

The causal model Table 9 comprises four observed variables: a treatment T, an outcome Y, a pre-treatment variable X and a post-treatment variable K. The model also contains four unobserved variables V, U, W, J that are external.

This model renders unusual counterfactual relationships. The matching condition $Y(t) \perp\!\!\!\perp T \mid K$ holds for the post-treatment variable K but does not hold for the the pre-treatment 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 pre-program variable X to the conditioning set of $Y(t) \perp\!\!\!\perp T \mid K$ prevents the identification of causal effects as $Y(t) \not\perp\!\!\!\perp T \mid (X, K)$.

The causal model of Table 9 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 through the NR framework, which lacks structural equations, unobserved variables and the direction of the causal relationships among variables is not explicitly disclosed.

[29] See, for instance, Greenland et al. (1999); Heckman and Navarro (2004); Pearl (2009b).

6 The Do-Calculus: Advantages and Limitations

The *do*-calculus (DoC) of Pearl (1995, 2009a, 2012) uses graph-based algorithms to examine the identification of counterfactuals in causal models represented by DAGs.[30] It employs structural equations, allows for unobserved variables, and clearly specifies the causal relationship among model variables.

The DoC is similar to the HM in the sense that both are based on structural equations. The frameworks however differ significantly in terms of counterfactual manipulations. DoC excludes the variable being fixed from the empirical model , while HM adds a hypothetical variable that formalizes the notion of thought experiments. DoC employs DAG-based algorithms outside the realm of standard statistical analysis. HM makes statistics converse with causality. Section (6.2) compares the identification techniques of both frameworks.

Judea Pearl and his co-authors 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 algorithm (Huang and Valtorta, 2006; Shpitser and Pearl, 2006).

Despite its 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. DoC does not apply to equilibrium or simultaneous equation models. It does not apply to identification strategies that invoke functional forms restrictions either. These restrictions include IV assumptions such as monotonicity or separability conditions. This limitation marginalizes the identification strategies discussed in Section 4. Applying the DoC to the Generalized Roy model generates a non-informative output that the Roy model is not identified.

[30] For a recent book on the graphical approach to causality, see Peters et al. (2017), and for related works on causal discovery, see Glymour et al. (2014), Heckman and Pinto (2015), Hoyer et al. (2009), and Lopez-Paz et al. (2017).

6.1 DoC 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 $G_{\bar{K}}$ 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:

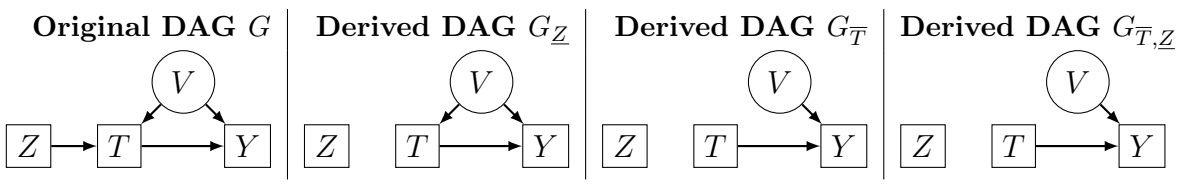
1. Rule 1: if $Y \perp\!\!\!\perp T|(K,X)$ holds in $G_{\overline{K}}$, then $P(Y|do(K),T,X) = P(Y|do(K),X)$,
2. Rule 2: if $Y \perp\!\!\!\perp T|(K,X)$ holds in $G_{\overline{K},\underline{T}}$, then $P(Y|do(K),do(T),X) = P(Y|do(K),T,X)$,
3. Rule 3: if $Y \perp\!\!\!\perp T|(K,X)$ holds in $G_{\overline{K},\overline{T(X)}}$,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

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.

6.1.1 Using Do-Calculus to Investigate the Roy Model

We illustrate the DoC mechanics by examining the Generalized Roy model.

Table 10: Using Do-Calculus to Investigate the Roy Model



The first column of Table 10 presents the DAG of the original Roy model, which is denoted by G. The second column displays the DAG $G_{\underline{Z}}$ which suppresses the arrow arising from Z. The LMC of Z on DAG $G_{\underline{Z}}$ is $Z \perp\!\!\!\perp (Y,T)$. From Rule 2 of DoC, we obtain $P(T|do(Z)) = P(T|Z)$. Summarizing:

$$G_{\underline{Z}} \Rightarrow T \perp\!\!\!\perp Z \Rightarrow \text{ by Rule 2 } P(T|do(Z)) = P(T|Z). \tag{38}$$

This says that Z is statistically independent of T when we fix Z. In the NR framework, this is 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 (38) to obtain $P(Y|do(Z)) = P(Y|Z)$. 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 column of Table 10 displays the DAG $G_{\overline{T}}$ which suppresses the arrow arriving at T. LMC of Z on $G_{\overline{T}}$ implies $Z \perp\!\!\!\perp Y$. By Rule 1 of DoC, we have that $P(Y|do(T),Z) = P(Y|do(T))$. Summarizing:

$$G_{\overline{Z}} \Rightarrow Y \perp\!\!\!\perp Z \Rightarrow \text{ by Rule 1 } P(Y|do(T),Z) = P(Y|do(T)). \tag{39}$$

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 HM framework.

The last column of Table 10 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. The LMC of Z for $G_{\overline{T}}$ implies $Z \perp\!\!\!\perp Y$. By Rule 1 of DoC, we have that $P(Y|do(T),Z) = P(Y|do(T))$. In summary:

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

This means that Z is statistically independent of Y when we fix T. This statement is the exogeneity condition $Y(t) \perp\!\!\!\perp Z$ or the independence relationship $Y \perp\!\!\!\perp Z|\tilde{T}$ of the HM framework. The LMC of Z is $Z \perp\!\!\!\perp (T,Y,V)$ which implies that $Z \perp\!\!\!\perp Z|T$ holds. Using Rule

2 of the DoC we obtain:

$$G_{\overline{T},\underline{Z}} \Rightarrow Y \perp\!\!\!\perp Z|T, \text{ so Rule 2 } P(Y|do(T), do(Z)) = P(Y|do(T), Z). \tag{41}$$

Combining $P(Y|do(T), Z) = P(Y|do(T))$ in (40) with $P(Y|do(T), do(Z)) = P(Y|do(T), Z)$ in (41) we obtain $P(Y|do(T), do(Z)) = P(Y|do(T))$. This means that the probability distribution of the outcome Y when we fix both Z, 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 analysis 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.

6.2 The Front-door Model

To make a more positive assertion, we apply the identification machinery of the 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 the DoC with the HM.

The Front-door model (42)–(45) consists of three observed variables T, M, Y and an unobserved confounding variable V. Treatment T causes a mediator M which in turn causes

outcome Y. Confounding variable V causes T, Y but not M.[31]

$$V = f_V(\epsilon_V) \tag{42}$$

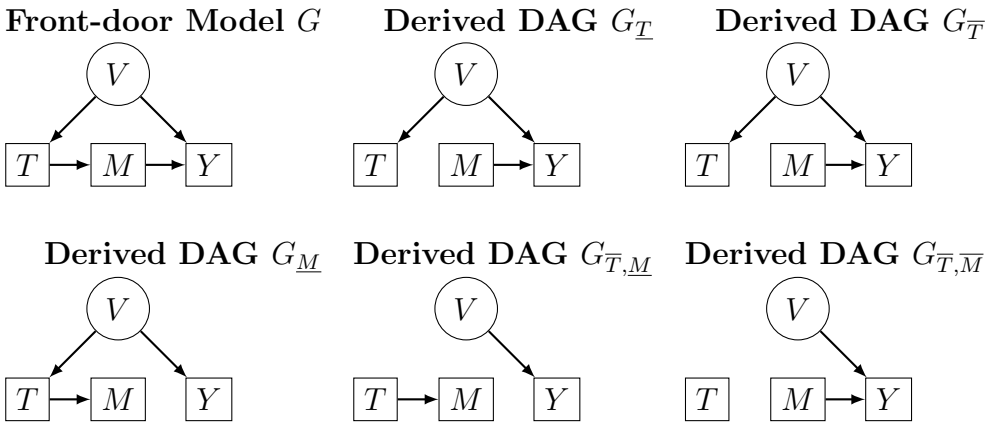
$$T = f_T(V, \epsilon_T) \tag{43}$$

$$M = f_M(T, \epsilon_M) \tag{44}$$

$$Y = f_Y(M, V, \epsilon_Y) \tag{45}$$

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.

Table 11: Using Do-Calculus to Identify the Causal Effect of T on Y in the Front-door Model



We illustrate how to use the DoC to identify the distribution of the counterfactual outcome $P_h(Y(t))$. For the sake of notational simplicity, suppose that all variables are discrete. The do-calculus is cumbersome. The method requires the five derived DAGs displayed in Table 11. The identification formula of the counterfactual outcome is obtained by the following sequence of steps:

[31] As before, the error terms $\epsilon_V, \epsilon_T, \epsilon_M, \epsilon_Y$ in the front-door model (42)–(45) are mutually statistically independent.

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

$$\begin{aligned}\therefore P(Y|do(M)) &= \sum_t P(Y|T=t, do(M))P(T=t|do(M)) \\ &\quad \text{by Law of Iterated Expectations (L.I.E.)} \\ &= \sum_t P(Y|T=t, M)P(T=t) \text{ by steps 1,2, and 3}\end{aligned}$$

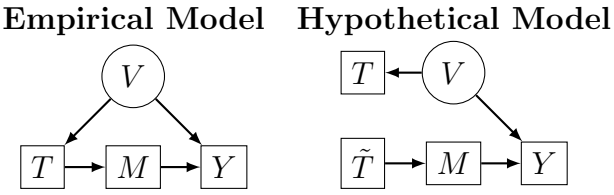
5. $Y \perp\!\!\!\perp M|T$ in $G_{\overline{T},\underline{M}}$ holds, thus by Rule 2, $P(Y|M, do(T)) = P(Y|do(M), do(T))$
6. $Y \perp\!\!\!\perp T|M$ in $G_{\overline{T},\overline{M}}$ holds, thus by Rule 3, $P(Y|do(T), do(M)) = P(Y|do(M))$
7. Collecting these results, we have that $P(Y|Z, do(T)) = P(Y|do(Z), do(T)) = P(Y|do(M))$.
8. Finally, we can use previous results to obtain the following equation:

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

6.3 Reexamining the Front Door Model using the HM

We now investigate the same front-door model using the hypothetical framework. Table 12 displays the hypothetical model associated with the Front-door model (42)–(45) as a DAG.

Table 12: The Empirical and Hypothetical Front-door Models1



We seek to identify the counterfactual outcome $P_h(Y|\tilde{T}=t)$, i.e., 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 the empirical models. To do so, we seek independence relationships in the hypothetical model that contain T and $\tilde{T}$. These are $Y \perp\!\!\!\perp \tilde{T}|(M,T)$ and $M \perp\!\!\!\perp T|\tilde{T}$ holds.[32] It is also the case $T \perp\!\!\!\perp \tilde{T}$ holds as $\tilde{T}$ is exogenous and does not cause T. We can then apply rules (16)–(17) to connect the probabilities of the empirical and hypothetical models:

$$Y \perp\!\!\!\perp \widetilde{T} \,|\, (T, M) \quad \Rightarrow \quad P_h(Y|\widetilde{T}, T=t', M) = P_e(Y|T=t', M) \tag{46}$$

$$M \perp\!\!\!\perp T \,|\, \widetilde{T} \quad \Rightarrow \quad P_h(M|\widetilde{T}=t, T) = P_e(M|T=t) \tag{47}$$

$$T \perp\!\!\!\perp \widetilde{T} \,|\, T \quad \Rightarrow \quad P_h(T=t'|\widetilde{T}) = P_e(T=t') \tag{48}$$

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

$$P_h(Y|\widetilde{T}=t) = \sum_{t',m} P_h(Y|m, T=t', \widetilde{T}=t) P_h(m|T=t', \widetilde{T}=t) P_h(T=t'|\widetilde{T}=t) \tag{49}$$

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

Equation (49) is a sum of probabilities defined in the hypothetical model due to application of the law of iterated expectation over T and M. Equation (50) replaces each of the hypothetical model probabilities with empirical model probabilities.[33]

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

[33] We adopt the following short-hand notation:

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

6.3.1 Understanding the HM Identification Criteria

The identification of the counterfactual outcomes in the Front-door Model stems from the three independence relationships in (46)–(48). 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 (46) is an independence regarding T conditional on $\widetilde{T}$. The second relationship (47) is independence on $\widetilde{T}$ conditional on T. The last one (48) cycle backs, it is an independence regarding T conditional on $\widetilde{T}$. This property enables to translate the probabilities of the hypothetical model into the probabilities of the empirical model via the connection rules (16)–(17).

The property of *alternate conditionals* ascribes an alternating feature to the identification equation (50). The first term of (50) is conditioned on $T = t'$ which refers to the first conditional T in (46). The identification equation (50) sums t' over the support of T. The second term of (50) is conditioned on the treatment value $T = t$. which refers to the second conditional T in (47). 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 (50) alternates. It is conditioned on $T = t'$ which refers to the last conditional T in (48) 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 (46) starts with the outcome Y and is conditioned on the variable M. The second relationship (47) 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, *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 start 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 $\mathcal{T}_K \setminus \mathcal{T}_{K-1}) \perp\!\!\!\perp \tilde{T}|(T, \mathcal{T}_{K-1})$

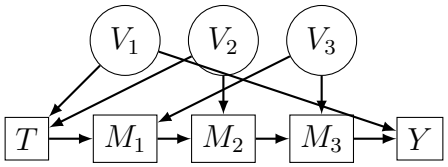
respectively) until it ends in either $\mathcal{T}_1 \perp\!\!\!\perp T|\tilde{T}$, or $\mathcal{T}_1 \perp\!\!\!\perp \tilde{T}|T$.

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.

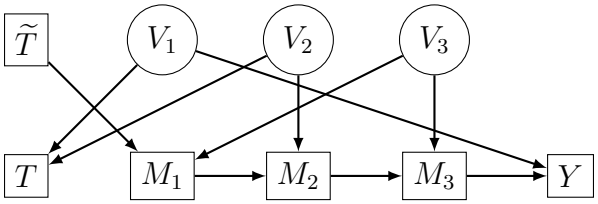
Consider a complex mediation model of Table 13. The model has three observed mediating variables M_1, M_2, M_3 and three unobserved, confounding variables V_1, V_2, V_3.

Table 13: Using the HM to Identify Counterfactuals

Directed Acyclic Graph of the Empirical Model



Directed Acyclic Graph of the Hypothetical Model



The following conditional independence relationships hold for the hypothetical model:

$$Y \perp\!\!\!\perp \widetilde{T}|(T, M_3, M_2, M_1) \tag{51}$$

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

$$M_2 \perp\!\!\!\perp \widetilde{T}|(T, M_1) \tag{53}$$

$$M_1 \perp\!\!\!\perp T|\widetilde{T} \tag{54}$$

The set of independence relationships (51)–(54) 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 law of iterated expectations and independence relationships (51)–(54) enable us to express the counterfactual probability $P_h(Y|\widetilde{T})$ as:

$$\begin{aligned}
\textbf{Hypothetical Model} \quad & P_h(Y|\widetilde{T} = t) = \sum_{t', m_3, m_2, m_1} A_h \cdot B_h \cdot C_h \cdot D_h \cdot E_h, \\
\text{where:} \quad & A_h = P_h(Y|m_3, m_2, m_1, T = t', \widetilde{T} = t) \\
& B_h = P_h(M_3 = m_3|m_2, m_1, T = t', \widetilde{T} = t) \\
& C_h = P_h(M_2 = m_2|m_1, T = t', \widetilde{T} = t) \\
& D_h = P_h(M_1 = m_1|T = t', \widetilde{T} = t) \\
& E_h = P_h(T = t'|\widetilde{T} = t)
\end{aligned}$$

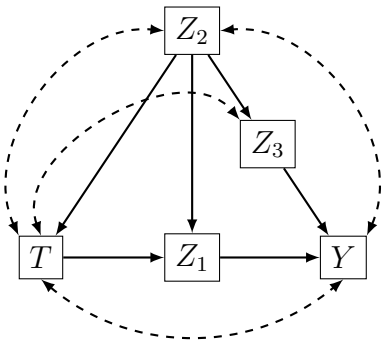
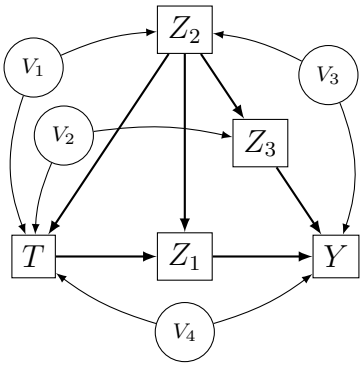
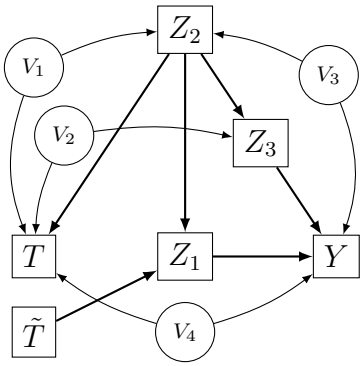
The connection rules (16)–(17) enable us to translate the hypothetical model probabilities into empirical model probabilities. The identification equation displays the alternative pattern of values t and t' in the same fashion as the identification equation of the Front-door model:

$$\begin{aligned}
\textbf{Empirical Model} \quad & P_e(Y(t)) = \sum_{t', m_3, m_2, m_1} A_e \cdot B_e \cdot C_e \cdot D_e \cdot E_e, \\
\text{where:} \quad & A_e = P_e(Y|m_3, m_2, m_1, T = t') \\
& B_e = P_e(M_3 = m_3|m_2, m_1, T = t) \\
& C_e = P_e(M_2 = m_2|m_1, T = t') \\
& D_e = P_e(M_1 = m_1|T = t) \\
& E_e = P_e(T = t')
\end{aligned}$$

6.4 The Identification Expression

The DoC and HM can generate equivalent but dramatically different expressions for the same identified causal effect. We illustrate this fact using the causal model of Table 14, which is widely known in the DoC literature. The first column of Table 14 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 second column of the table presents the equivalent empirical model, which explicitly displays the unobserved confounding variables V_1, V_2, V_3, V_4. The last column of the table presents the hypothetical model counterpart.

Table 14: Example of a Well-known Causal Model in DoC Literature

Original Model	Empirical Model	Hypothetical Model
		

The following conditional independence relationships hold for the hypothetical model in third column of Table 14:

$$Y \perp\!\!\!\perp \widetilde{T} \mid (T, Z_1, Z_2, Z_3) \tag{55}$$

$$Z_1 \perp\!\!\!\perp T \mid (\widetilde{T}, (Z_2, Z_3)) \tag{56}$$

$$(Z_2, Z_3) \perp\!\!\!\perp \widetilde{T} \mid T \tag{57}$$

The independence relationships (55)–(57) generate the the bridging sequence $Y \rightarrowtail Z_1 \rightarrowtail (Z_2, Z_3) \rightarrowtail M_1 \rightarrowtail T$ which renders the following identification expression:

$$P_e(Y(t)) = \sum_{t', z_1, z_2, z_3} P_e(Y|z_1, z_2, z_3, T = t')P_e(z_1|z_2, z_3, T = t)P_e(z_2, z_3|T = t')P_e(T = t') \tag{58}$$

Note that the values of T in expression (58) alternate from t' to t. The identification expression generated by the DoC does not abide to this property. The DoC algorithm of

Bareinboim and Pearl (2016) returns the following identification expression:

$$P(Y(t)) = \sum_{z_1,z_2,z_3} P(z_3,z_2)P(z_1|T=t,z_2) \cdot \left(\frac{\sum_{t'} P(Y,z_3|T=t',z_1,z_2)P(T=t',z_2)}{\sum_{t'} P(z_3|T=t',z_1,z_2)P(T=t',z_2)} \right) P(z_2) \tag{59}$$

Expressions (58) and (59) appear to be substantially different. They are, in fact, equivalent. Tikka and Karvanen (2017) has shown that the DoC expression (59) 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 $Z_1 \perp\!\!\!\perp Z_3|(T,Z_1)$ holds for the empirical model of Table 14. We can combine these two information to show that:

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

The second equality uses the Bayes Theorem. The third equality cancels out $P(z_2)$. The fourth equality is due to $Z_1 \perp\!\!\!\perp Z_3|(T,Z_1)$. The last equality rearranges the terms.

Finally, the independence relationships (55)–(57) of the hypothetical model in Table 14

also hold if we were to suppress Z_3, that is:

$$Y \perp\!\!\!\perp \widetilde{T} \mid (T, Z_1, Z_2)$$
$$Z_1 \perp\!\!\!\perp T \mid (\widetilde{T}, (Z_2))$$
$$Z_2 \perp\!\!\!\perp \widetilde{T} \mid T$$

Thus, we can also express the counterfactual outcome probability $P(Y(t))$ as:

$$P_e(Y(t)) = \sum_{t',z_1,z_2} P_e(Y|z_1, z_2, T = t')P_e(z_1|z_2, T = t)P_e(z_2|T = t')P_e(T = t'). \tag{60}$$

7 Conclusion

This paper discusses the econometric model for causal policy analysis and two alternative causal frameworks that are approximations to it.

Standard statistical theory is not equipped to describe causality. Statistics lacks directionality, which is essential for causal analysis. Causal frameworks devise additional tools to investigate causal inquiries.

The econometric model for causal policy analysis stems from the insights of early economists such as Haavelmo (1943) and Frisch (1930). It employs structural equations to describe the causal relationship between model variables and employs the fixing operation to define counterfactual variables. The framework enables the analyst to clearly distinguish the three primary tasks of causal inference: the definition of the causal model, the identification of causal parameters, and the estimation from data. The econometric approach to causality is highly adaptable and has been used to draw causal inferences in a wide range of policy questions investigated by economists.

The two approximating approaches we consider are the Neyman-Rubin approach rooted in the statistics of experiments and the do-calculus that originated in computer science. Both are recent developments that attempt to address some of the same problems tackled by the

econometric approach. These frameworks share the same theoretical foundation. They all rely on the intuitive definition of a causal effect as a *ceteris paribus* consequence of a policy change. However, the frameworks differ substantially on the causal machinery used to define, construct, and identify counterfactuals.

We make the case that the recently developed causal frameworks can excel in examining particular subsets of causal inquiries, but they lack either the flexibility or the clarity that the econometric approach to causality offers.

The do-calculus only applies to recursive models that can be represented by directed acyclic graphs (DAG). Its identification method is based on the reiterative use of rules that combine DAG representations, independence conditions, and probability equalities.

The use of the do-calculus has been limited in economics. The method has a rigid structure that only applies to models that can be fully described by directed acyclic graphs. These do not include equilibrium models or simultaneous equations. The method also precludes identification strategies that invoke functional forms restrictions. These restrictions preclude most identification methods used in economics, including IV assumptions such as monotonicity or separability conditions.

The Neyman-Rubin (NR) approach has great appeal among statisticians due to its simplicity. The method is motivated by well-known "treatment-control" design of randomized controlled trials. The method does not employ structural equations nor models unobserved variables explicitly. Instead, causal relations are implicitly expressed through independent conditions among observed variables' counterfactuals. These characteristics enable a parsimonious description of simple causal models closely related to the "treatment-control" paradigm.

Social scientists trained on the NR method are seldom aware of its limitations. The lack of structural equations substantially limits the complexity of causal models that can be successfully assessed by the NR method. As the number of variables increases, it becomes increasingly complex to translate the independence relationship of the NR approach into

causal relationships among model variables. In practice, the NR approach limits the models that the analyst can successfully examine. The NR method is commonly used to describe a few relatively simple causal models. The lack of unobserved variables imposes further limitations. It prevents the analyst from investigating model properties beyond its identification and estimation.

The study of econometric causality dates back to the seminal ideas of Frisch (1930) and Haavelmo (1943). Thought these years, the economic literature has accumulated a rich body of theoretical and practical knowledge to address policy questions. Today, applied economists would do well by using the outstanding collection of tools available from their own past.

References

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (1999). Training effects on employment when the training effects are heterogeneous: An application to Norwegian vocational rehabilitation programs. University of Bergen Working Paper 0599, and University of Chicago.

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: An application to Norwegian vocational rehabilitation programs. *Journal of Econometrics 125*(1–2), 15–51.

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies 72*(1), 1–19.

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 72, pp. 5145–5303. Amsterdam: Elsevier Science B. V.

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

Altonji, J. G. and R. L. Matzkin (2005, July). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica 73*(4), 1053–1102.

Angrist, J. D., G. W. Imbens, and D. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), 444–455.

Athey, S. and G. W. Imbens (2006, March). Identification and inference in nonlinear difference-in-differences models. *Econometrica 74*(2), 431–497.

Balke, A. A. and J. Pearl (1993). Nonparametric bounds on causal effects from partial compliance data. Technical Report R-199, University of California at Los Angeles.

Bareinboim, E. and J. Pearl (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences 113*, 7345–7352.

Bertrand, M., E. Duflo, and S. Mullainathan (2004, February). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics 119*(1), 249–275.

Blundell, R., A. Duncan, and C. Meghir (1998, July). Estimating labor supply responses using tax reforms. *Econometrica 66*(4), 827–861.

Blundell, R. and J. Powell (2003). Endogeneity in nonparametric and semiparametric regression models. In L. P. H. M. Dewatripont and S. J. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Volume 2. Cambridge, UK: Cambridge University Press.

Brinch, C. N., M. Mogstad, and M. Wiswall (2017). Beyond late with a discrete instrument. *Journal of Political Economy 125*(4), 985–1039.

Buchinsky, M. and R. Pinto (2021). Using economic incentives to generate monotonicity criteria of iv models. Unpublished Manuscript, UCLA.

Carneiro, P., K. Hansen, and J. J. Heckman (2001, Fall). Removing the veil of ignorance in assessing the distributional impacts of social policies. *Swedish Economic Policy Review 8*(2), 273–301.

Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Cox, D. R. (1958). *Planning of Experiments*. New York: Wiley.

Cunha, F., J. J. Heckman, and S. Navarro (2005, April). Separating uncertainty from heterogeneity in life cycle earnings, The 2004 Hicks Lecture. *Oxford Economic Papers 57*(2), 191–261.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, May). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Dawid, A. P. (1976). Properties of diagnostic data distributions. *Biometrics 32*(3), 647–658.

Duncan, O. D. and A. S. Goldberger (1973). Structural equation models in the social sciences. In O. D. Duncan and A. S. Goldberger (Eds.), *Social Science Research Council (États-Unis) and University of Wisconsin. Social Systems Research Institute*. New York: Seminar Press.

Fisher, R. A. (1935). *The Design of Experiments*. London: Oliver and Boyd.

Frangakis, C. E. and D. Rubin (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. Lectures at Yale University beginning September, 1930. Mimeographed, 246 pp. Frisch Archives, Department of Economics, University of Oslo.

Frisch, R. (1938). Autonomy of economic relations: Statistical versus theoretical relations in economic macrodynamics. Paper given at League of Nations. Reprinted in D.F. Hendry and M.S. Morgan (1995), *The Foundations of Econometric Analysis*, Cambridge University Press.

Geiger, D., T. Verma, and J. Pearl (1990). Identifying independence in bayesian networks. *Networks 20*(5), 507–534.

Glymour, C., R. Scheines, and P. Spirtes (2014). *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modeling*. Academic Press.

Goldberger, A. S. (1972, November). Structural equation methods in the social sciences. *Econometrica 40*(6), 979–1001.

Greenland, S., J. Pearl, and J. Robins (1999). Causal diagrams for epidemiologic research. *Epidemiology 10 1*, 37–48.

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

Heckman, J. and R. Pinto (2018). Unordered monotonicity. *Econometrica 86*, 1–35.

Heckman, J. J. (1978, July). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*(4), 931–959.

Heckman, J. J. (1979, January). Sample selection bias as a specification error. *Econometrica 47*(1), 153–162.

Heckman, J. J. (2008a, April). Econometric causality. *International Statistical Review 76*(1), 1–27.

Heckman, J. J. (2008b). The principles underlying evaluation estimators with an application to matching. *Annales d'Economie et de Statistiques 91–92*, 9–73.

Heckman, J. J., H. Ichimura, J. Smith, and P. E. Todd (1998, September). Characterizing selection bias using experimental data. *Econometrica 66*(5), 1017–1098.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999). The economics and econometrics of active labor market programs. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3A, Chapter 31, pp. 1865–2097. New York: North-Holland.

Heckman, J. J. and S. Navarro (2004, February). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics 86*(1), 30–57.

Heckman, J. J. and R. Pinto (2015). Causal analysis after Haavelmo. *Econometric Theory 31*(1), 115–151.

Heckman, J. J. and R. Robb (1985a). Alternative methods for evaluating the impact of interventions. In J. J. Heckman and B. S. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*, Volume 10, pp. 156–245. New York: Cambridge University Press.

Heckman, J. J. and R. Robb (1985b, October-November). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics 30*(1–2), 239–267.

Heckman, J. J. and C. Taber (2008). The roy model. In S. N. Durlauf and L. E. Blume (Eds.), *New Palgrave Dictionary of Economics* (2 ed.). Basingstoke, UK: Palgrave Macmillan.

Heckman, J. J., S. Urzúa, and E. J. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics 88*(3), 389–432.

Heckman, J. J., S. Urzúa, and E. Vytlavil (2008). Instrumental variables in models with multiple outcomes: the general unordered case. *Annales d'conomie et de Statistique* (91/92), 151–174.

Heckman, J. J. and E. J. Vytlacil (1999, April). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences 96*(8), 4730–4734.

Heckman, J. J. and E. J. Vytlacil (2005, May). Structural equations, treatment effects and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 70, pp. 4779–4874. Amsterdam: Elsevier B. V.

Heckman, J. J. and E. J. Vytlacil (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. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 71, pp. 4875–5143. Amsterdam: Elsevier B. V.

Holland, P. W. (1986, December). Statistics and causal inference. *Journal of the American Statistical Association 81*(396), 945–960.

Hoyer, P. O., D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf (2009). Nonlinear causal discovery with additive noise models. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (Eds.), *Advances in Neural Information Processing Systems 21*, pp. 689–696. Curran Associates, Inc.

Huang, Y. and M. Valtorta (2006). Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06, Arlington, Virginia, USA, pp. 217224. AUAI Press.

Hurwicz, L. (1962). On the structural form of interdependent systems. In E. Nagel, P. Suppes, and A. Tarski (Eds.), *Logic, Methodology and Philosophy of Science*, pp. 232–239. Stanford University Press.

Imbens, G. W. (2004, February). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics 86*(1), 4 – 29.

Imbens, G. W. and J. D. Angrist (1994, March). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Kiiveri, H., T. P. Speed, and J. B. Carlin (1984). Recursive causal models. *Journal of the Australian Mathematical Society (Series A)*, 30–52.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford, UK: Clarendon Press.

Lee, S. and B. Salanié (2018). Identifying effects of multivalued treatments. *Econometrica 86*, 1939–1963.

Lopez-Paz, D., R. Nishihara, S. Chintala, B. Schölkopf, and L. Bottou (2017). Discovering causal signals in images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6979–6989.

Matzkin, R. L. (1993, July). Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics 58*(1–2), 137–168.

Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, pp. 2523–2558. New York: North-Holland.

Matzkin, R. L. (2007). Nonparametric identification. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B. Amsterdam: Elsevier.

Matzkin, R. L. (2008). Identification in nonparametric simultaneous equations models. *Econometrica 76*(5), 945–978.

Matzkin, R. L. (2013). Nonparametric identification of structural economic models. *Annual Review of Economics 5*(1), 457–486.

Matzkin, R. L. (2015). Estimation of nonparametric models with simultaneity. *Econometrica 83*(1), 1–66.

Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using instrumental variables for inference about policy relevant treatment effects. *Econometrica 86*(5), 1589–1619.

Mogstad, M. and A. Torgovitsky (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics 2*, 577–613.

Neyman, J. (1923). Statistical problems in agricultural experiments. *Journal of the Royal Statistical Society II (Supplement)*(2), 107–180.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Pearl, J. (1993). [Bayesian analysis in expert systems]: Comment: Graphical models, causality and intervention. *Statistical Science 8*(3), 266–269.

Pearl, J. (1995, December). Causal diagrams for empirical research. *Biometrika 82*(4), 669–688.

Pearl, J. (2009a). *Causality: Models, Reasoning, and Inference* (2nd ed.). New York: Cambridge University Press.

Pearl, J. (2009b). Myth, confusion, and science in causal analysis. *Technical Report, UCLA, Department of Statistics*.

Pearl, J. (2012). The do-calculus revisited. *CoRR abs/1210.4852*.

Peters, J., D. Jazzing, and B. Schölkopf (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. Unpublished paper based on Ph.D. Thesis, University of Chicago, Department of Economics.

Powell, J. L. (1994). Estimation of semiparametric models. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, pp. 2443–2521. Amsterdam: Elsevier.

Pratt, J. W. and R. Schlaifer (1984, March). On the nature and discovery of structure. *Journal of the American Statistical Association 79*(385), 9–33.

Quandt, R. E. (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. *Mathematical Modelling 7*(9–12), 1393–1512.

Rosenbaum, P. R. and D. B. Rubin (1983, April). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Roy, A. (1951, June). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Rubin, D. B. (1974, October). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688–701.

Schennach, S. M. (2020). Mismeasured and unobserved variables. *Handbook of Econometrics 7*(A), 487–565.

Shpitser, I. and J. Pearl (2006, November). 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, AAAI-06/IAAI-06*, Proceedings of the National Conference on Artificial Intelligence, pp. 1219–1226. 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference, AAAI-06/IAAI-06 ; Conference date: 16-07-2006 Through 20-07-2006.

Shpitser, I. and J. Pearl (2009). Effects of treatment on the treated: Identification and generalization. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, UAI 2009*, Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, UAI 2009, pp. 514–521. AUAI Press.

Spiegler, R. (2020). Behavioral implications of causal misperceptions. *Annual Review of Economics 12*(1), 81–106.

Tamer, E. (2003, January). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies 70*(1), 147–165.

Telser, L. G. (1964, September). Iterative estimation of a set of linear regression equations. *Journal of the American Statistical Association 59*(307), 845–862.

Theil, H. (1953). *Esimation and Simultaneous Correlation in Complete Equation Systems*. The Hague: Central Planning Bureau. Mimeographed memorandum.

Tikka, S. and J. Karvanen (2017). Simplifying probabilistic expressions in causal inference. *Journal of Machine Learning Research 18*(1), 1–30.

Vytlacil, E. J. (2002, January). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica 70*(1), 331–341.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 420–445.