Econometric causality: The central role of thought experiments

James Heckman [a],[*],[1], Rodrigo Pinto [b],[1]

[a] The University of Chicago, Department of Economics, 1126 E. 59 th St., Chicago, IL 60637, United States of America
[b] University of South Florida, Department of Economics, Arts and Sciences Multidisciplinary Complex, Tampa, FL 33620, United States of America

ARTICLE INFO

ABSTRACT

This paper examines the econometric causal model and the interpretation of empirical evidence based on thought experiments that was developed by Ragnar Frisch and Trygve Haavelmo. We compare the econometric causal model with two currently popular causal frameworks: the Neyman–Rubin causal model and the Do-Calculus. The Neyman–Rubin causal model is based on the language of potential outcomes and was largely developed by statisticians. Instead of being based on thought experiments, it takes statistical experiments as its foundation. The Do-Calculus, developed by Judea Pearl and co-authors, relies on Directed Acyclic Graphs (DAGs) and is a popular causal framework in computer science and applied mathematics. We make the case that economists who uncritically use these frameworks often discard the substantial benefits of the econometric causal model to the detriment of more informative analyses. We illustrate the versatility and capabilities of the econometric framework using causal models developed in economics.

1. Introduction

Sound economic and policy analysis is causal analysis. It analyzes the factors that produce outcomes and the role of various factors and policies in doing so. It quantifies policy impacts. It elucidates the mechanisms producing outcomes in order to understand how they operate, how they can be transported to different environments, how programs might be improved, and which, if any, alternative mechanisms might be used to generate desired outcomes. It organizes evidence in interpretable frameworks. It uses all available information to give good policy advice and explicitly recognizes any limitations of data or models.

Good economic science systematically explores possible counterfactual worlds. It is grounded in thought experiments — what might happen if determinants of outcomes are changed. Credible hypotheticals are developed, analyzed, and tested with real-world data.

Models and thought experiments are central to economic analysis. Persons trained in economic theory or in the natural sciences routinely use them. Statisticians and computer scientists have only recently come to grips with the causal questions that have long been investigated by econometric pioneers such as Ragnar Frisch and Trygve Haavelmo. As a result, private languages and procedures designed to do part of what rigorous econometric models do have been developed without manifesting an understanding of the full corpus of econometric theory, often refusing to cite it and reinventing portions of it.

These private languages bear the marks of their recent birth: concepts are often not precisely defined, and the conceptually distinct issues of definition of counterfactuals, their identification, and their estimation are often tangled together. In some fields heavily influenced by statistics, certain estimation techniques are claimed to be central for the definition or identification of counterfactuals when, in fact, they are only devices for recovering counterfactuals from data.

The current state of affairs would be of little concern if applied economists continued to draw on and extend the standard econometric model. Sadly, this is not the case. Many econometricians and applied economists now emulate what they read in statistics and computer science journals. They have forgotten or never learned their own field's foundational work to the detriment of rigorous causal analysis and testing among alternatives.

This paper discusses econometric causal analysis and recently developed causal models in fields outside economics. Our goal is to enhance the theory and practice of economic policy analysis by testing and synthesizing evidence, as well as interpreting it. This involves acquainting economists with a rich econometric legacy and situating recently advocated causal frameworks within the broader context of the econometric model.[2]

The topic is broad and our paper is necessarily brief. We discuss some main points and illustrate them with analyses of a few prototypical economic models used to address policy problems and interpret evidence. It is impossible to convey here all of the insights of rigorous econometrics developed in the past 90 years.

This paper unfolds in the following way. We first define the notion of causality within a model. The concept is simple, but requires thought processes outside of statistics that are, nonetheless, quite intuitive. We discuss four distinct classes of policy problems that are addressed in econometric causal analyses. Some of them are either ignored or only partly addressed in the recent non-economic causal literatures. We demonstrate the conceptual clarity of the econometric approach and contrast it with that of rival approaches.

In particular, we consider two causal frameworks often advocated by statisticians and computer scientists. The first is the Neyman–Rubin model (1923; 1958; 1974; 1986; 1996), "NR" henceforward. It uses notions developed in rigorous econometrics but goes only part way toward implementing the full set of tools in the econometric approach to causal analysis and the interpretation of empirical evidence. It has important limitations for posing or analyzing routine policy problems outside a narrow "treatment-control" paradigm. It ignores the simultaneous equations model – a major achievement of econometrics – and replaces it with a litany of "confounding biases" readily addressed in rigorous econometrics. We also consider an approach to counterfactuals developed in computer science ("*do-calculus*", Pearl, 2012), henceforth "DoC", that relies critically on directed acyclic graphs (DAGs) and statistical conditional independence relationships. We demonstrate its limited capacity to address many important economic questions and address important empirical problems or to utilize many standard econometric estimation and identification tools.

Each of the recent approaches holds value for limited classes of problems. However, they have severe limitations when applied to the broad array of problems economists routinely confront. The danger lies in the sole reliance on these tools, which eliminates serious consideration of important policy and interpretation questions. We highlight the flexibility and adaptability of the econometric approach to causality, contrasting it with the limitations of other causal frameworks.

For instance, the NR approach does not readily accommodate unobservables and restrictions on empirical relationships produced by economic theory, which are important components of the econometric toolkit. Social interactions, peer effects, and general equilibrium theory fall outside its purview, and are currently considered frontier-topics in those fields, despite the existence of well-designed econometric tools that address these issues. These are all standard problems addressed in structural econometrics.

Similarly, the DoC approach cannot deal with the functional restrictions and covariance information routinely used in econometrics. It cannot accommodate assumptions such as monotonicity and separability restrictions, which are essential components of modern instrumental variable analysis. The prototypical Generalized Roy model cannot be identified with DoC, although it and more general models can be identified using standard econometric tools.

This paper builds on our previous work in several ways. Heckman (2008a) and Heckman and Pinto (2015) discuss econometric causality but are less explicit than this paper in establishing links between formal econometric models and competing approaches. We clarify the distinctions between the "do" operator of Pearl (2009a) and the "fix" operator of Haavelmo (1943) and exposit much more clearly why causality is such a difficult concept for statisticians.[3] We introduce a new hypothetical model that uses probabilistic tools to analyze causal models without the artifices required in competing approaches. For example, we disentangle the "SUTVA" assumption of the Neyman–Rubin model into an autonomy (structural invariance) assumption and an absence-of-general-equilibrium-effects assumption.[4] We provide concrete examples of the limits and benefits of alternative causal frameworks.

This paper is organized as follows. Section 2 defines causality and discusses the tasks of causal inference. Section 3 presents the econometric model. Section 4 shows its versatility and describes various identification approaches in the Generalized Roy model. Section 5 examines the Neyman–Rubin causal model and contrasts it with the econometric approach. Section 6 investigates the *Do-Calculus* of Pearl (2009a). Section 7 examines non-recursive models that are ruled out in the NR approach. Section 8 summarizes the paper.

---

[2] In this paper we focus on policy analysis but our message applies to a broader class of problems. The models developed in this paper also apply to the tasks of hypothesis testing, statistical inference, and synthesis of empirical evidence into interpretable evidence. Formulating meaningful alternatives is central to power analysis or Bayesian tests among alternatives.

[3] We note that Haavelmo (1943, 1944) never uses the term "fix" in his analyses. However, he introduces the notion of thought experiments in his 1943 and 1944 papers, referring to them as idealized experiments. In his 1943 paper, he discusses the "hypothetical splitting" of the real economic world into separate spheres of action. His example is a Keynesian consumption and investment function where he separates–hypothetically–consumer and producer actions in the context and investment are jointly determined. Haavelmo was a student of Ragnar Frisch, who defined the term econometrics and laid the foundations of econometric causal policy analysis in two foundational studies (Frisch, 1930, 1933). Haavelmo (1944) refers extensively to Frisch's work and later essays on policy evaluation. He is credited with formalizing Frisch's distinction between hypothetical worlds (models) and empirical data. See also Bjerkholt and Dupont (2010) and Frisch (1933).

[4] The do-calculus explicitly uses autonomous structural relationships (Pearl, 2009a).

2. Causality as a thought experiment

A formal definition of causality relies on a modification of the same thought process used to define relationships mapping inputs X, that may contain unobserved terms, to outcomes Y using a stable map g:

$$g : X \to Y \tag{1}$$

A map is **stable** if changing its arguments over the domain of X preserves the map. Another way to express this is $Y = g(X)$, where g may be a multi-valued correspondence. An elementary version of (1) is the linear model:

$$Y = \alpha + \beta X. \tag{2}$$

In this example, stability means that α and β do not change when X is changed. This invariance property is termed *autonomy* of relationships by Frisch (1938). It is a cornerstone of causal analysis.[5] Typical examples of autonomous relationships in economics are production functions or demand equations.

A second fundamental concept in causality is directionality. The map g states that X causes Y. Inverting this map (when possible) may produce a stable relationship, but it is, in general, not causal.

The range of Y is a set of *potential outcomes* associated with X over its domain. The map g may be either a function or a correspondence. For example, our analysis is applicable to settings such as Nash games with multiple equilibria.[6] *Counterfactual outcomes* $Y(x)$ refer to the potential values that Y takes across different values of X. The key idea in causality is the notion captured in Alfred Marshall's phrase, "*ceteris paribus*" — all other else being equal.[7] Comparisons of Y for different values of X – all other factors the same – are defined as *causal effects*. They are conceptual thought experiments. This definition is used explicitly in the econometric approach regardless of what is observed, the statistical properties of X and Y, the specification of functional forms for g, or how X is manipulated in any thought experiment. The Roy model (1951) is an early example of a model of two potential outcomes associated with the income that the same person would earn in different jobs. We use a generalization of it as an example prototypical model throughout this paper.

Issues of identification and estimation are important for making the concept of causality empirically operational, but not for defining it. However, these auxiliary issues are sometimes assumed to be paramount in defining causality in the recent non-economic literatures. For example, in an influential exposition of the Neyman–Rubin model, Holland (1986) insists that causal effects are only defined for experimental manipulations of X. However, issues of definition and estimation are fruitfully distinguished and are the hallmark of the econometric approach. To make our discussion more concrete, an example from the standard toolkit of empirical economics is helpful.

2.1. Regression: Conditional expectation or thought experiment?

Consider the standard workhorse of empirical economics.[8] Anticipating empirical applications, we add the distinction between observed and unobserved variables that is strictly not required for the definition of causal parameters. Consider the regression of Y on T where (Y, T) are observed and U denotes a variable that is not observed by the analyst:

$$Y = T\beta + U. \tag{3}$$

In terms of (1), $X = (T, U)$. If X is a vector of all possible causes of Y, (1) is an *all-causes* model and accommodates stochastic shocks. Coupled with stability, such a model is convenient for transporting (1) to environments where different levels of T are at play (forecasting) or in combining and summarizing evidence from different studies where T varies (research synthesis).

A major source of confusion about causal models is that (3) is often defined by statisticians as a model for describing *statistical* relationships between Y and T (see e.g., Pratt and Schlaifer, 1984; Holland, 1997). Doing so uses standard statistical tools to define empirical relationships. Note that if conditional expectations exist, $E(Y \mid T = t) = t\beta + E(U \mid T = t)$. In this approach, the statistical model could be equivalently defined as $U = Y - T\beta$.

The empirical association between T and Y operates through two channels: β and $E(U \mid T = t)$, unless T is mean independent of U. This approach introduces considerations about the properties of random variables that are unnecessary for defining causality in contrast to just defining an empirical regularity.

2.2. Thought experiments

Another way to interpret $Y = T\beta + U$ is to hypothetically vary T and U: $(T, U) \to Y$ via $Y = T\beta + U$. This is not a statistical operation and lies outside standard statistics.[9] Economists (and other scientists) use hypothetical models (thought experiments) to analyze phenomena and explore possible relationships. These and other possible relationships are *defined* by causal operations, although they are *estimated* using statistical methods.

---

[5] Frisch (1938) defines autonomy of a function to mean functions that are "invariant" to changes in their arguments. Hurwicz (1962) prefers the term "structural" to denote autonomous equations.

[6] See e.g., Mas-Colell et al. (1995), Tamer (2003).

[7] Marshall (1961).

[8] See Haavelmo (1943) for an early discussion of the distinction made in this section.

[9] For an example of how confusing this concept is to statisticians, see Pratt and Schlaifer (1984) and Holland (1997). Holland's confusion is significant given that he was the person who formalized the "Rubin model" (1986).

To clarify these ideas, it is helpful to introduce random variables ϵ_V, ϵ_T, ϵ_U which are unobserved (by the analyst) and mutually statistically independent. They are external to the model (exogenous) and are not caused by T, U or Y.

Example 2.1. Consider four different possible causal models — all thought experiments:

Causal Model 1	Causal Model 2	Causal Model 3	Causal Model 4
$T = f_T(\epsilon_T)$	$T = f_T(\epsilon_T, \epsilon_V)$	$T = f_T(\epsilon_T, U)$	$T = f_T(\epsilon_T)$
$U = f_U(\epsilon_U)$	$U = f_U(\epsilon_U, \epsilon_V)$	$U = f_U(\epsilon_U)$	$U = f_U(\epsilon_U, T)$
$Y = T\beta + U$	$Y = T\beta + U$	$Y = T\beta + U$	$Y = T\beta + U$

In the first causal model, T does not cause U, nor does U cause T. Parameter β is the causal effect of varying T on Y for a fixed value of U. Variables T and U are statistically independent and the parameter β can be consistently estimated by OLS. In the second causal model, T does not cause U, nor does U cause T. Parameter β is still the causal effect of T on Y. However, T and U are not statistically independent because they share a common confounding variable ϵ_V and the OLS estimator of β is biased. This model is sometimes called a "common cause" model with ϵ_V being the common cause of T and U. The third causal model differs from the second model because U causes T. Although the causal relations of the second and third models differ, the causal effect of T on Y remains β. In these models, T and U are not statistically independent and the OLS estimator is generally biased.[10] The fourth model describes the case where T causes U. In this case, the OLS estimator of the parameter β does not, in general, describes the causal effect of T on Y since we need to account for the effect of T on Y that operates through U. The OLS estimator is biased and it evaluates a combination of the direct effect of T on Y and the indirect effect of T on Y via U.

Using the standard regression model as a starting point blurs the logic of this thought process. Econometrics textbooks commonly introduce causality in the context of the linear model (3). In this approach, the identification of causal effects is often reduced to a statistical property of the econometric model, namely, that causal effects can be assessed when variables T and U are uncorrelated. It gives rise to the practice of defining causal effects as conditional probability statements instead of statements about manipulating variables in a thought experiment.

In fact, OLS is based on statistical assumptions that are void of any causal interpretation. The OLS fitted value for the outcome Y conditioning on $T = t$ evaluates the conditional expectation $E(Y \mid T = t)$ instead of the counterfactual expectation $E(Y(t) \mid T = t)$, where the counterfactual outcome $Y(t)$ is the value of Y when T is externally set to a value t. The causal content of the OLS model arises only when we invoke concepts such as fixing and counterfactuals. These concepts are not part of the standard statistical toolkit. Whether or not we can identify β in a sample is an entirely separate question from defining the causal impact of T on Y.

Frisch, the founding father of modern econometric causal policy analysis, clearly understood that the study of causality is an exercise in abstract thought, and that *"Causality is in the Mind"*:

> *"...we think of a cause as something imperative which exists in the **exterior world**. In my opinion this is fundamentally **wrong**. If we strip the word cause of its animistic mystery, and leave only the part that science can accept, nothing is left except a certain way of thinking. [T]he scientific ...problem of **causality** is essentially a problem regarding our **way of thinking**, not a problem regarding the nature of the exterior world".* — *(Frisch, 1930), p. 36*

Stated differently, Frisch is saying causality is the outcome of a thought experiment, i.e., a model.

2.3. The econometric approach to causality

The econometric approach to causality develops explicit hypothetical models where inputs cause outcomes. A common context is the study of policy evaluations when economic agents choose treatments that affect economic outcomes of interest. "Treatments" are inputs (the T) which need not be restricted to binary or discrete valued variables. The mechanisms governing the choice of inputs are central to the study of the causal effect of treatment on outcomes. Identification/estimation/interpretation of empirical counterparts to the hypothetical counterfactuals require careful accounting of unobserved (by the analyst) variables (U) that cause both input choice and outcomes. Structural econometric models do just that.[11]

2.4. Four distinct policy questions

The econometric approach to causality distinguishes four distinct classes of policy problems and addresses each of them, sometimes in the same analysis.[12]

P1 *Evaluating the impacts of implemented interventions on outcomes in a given environment, including their impacts in terms of the well-being of the treated and society at large. The simplest forms of this problem are typically addressed in the non-economic literatures: does a program in place "work" in terms of policy impacts?*

---

[10] Thus, $Y(t) \perp\!\!\!\perp T \mid U$ holds for the third model but not for the second model.

[11] Caricatures sometimes made in the non-economic literatures that the choices of inputs T involve highly stylized rational choice models or perfect information are false (see, e.g., Morgan and Winship, 2015). Some hypothetical models might maintain those assumptions, but such assumptions are in no way essential to the enterprise.

[12] See Heckman (2008a).

The non-economic literatures addressing **P1** identify and estimate treatment effects (most often average treatment effects) without investigating how they arise or whether alternative programs might be better or even what "better" means. In terms of our linear equation example, it seeks to know the sign and magnitude of β. However, most economic and policy analysts seek greater generality for their findings. This leads to problem **P2**.

P2 *Understanding the mechanisms producing treatment effects and policy outcomes.*

This asks the analyst to investigate the causes of effects and is a central task of economic theory and policy analysis.[13] It embeds (3) in a model that explains how T operates (i.e., which factors explain the $Y - T$ relationship). It goes beyond the coarse description of "treatment" T to explicate the factors that produce Y. It links with **P3** and **P4** stated below to consider how alternative mechanisms generate observed outcomes and can be used to forecast policies going forward, or explain the findings of any given study in a particular environment. **P2** is also an integral part of the task of constructing alternatives to maintained hypotheses and interpreting evidence using economic models.

P3 *Forecasting the impacts (constructing counterfactual states) of interventions implemented under one environment when the intervention is applied to other environments, including their impacts in terms of well-being.*

This goes beyond **P2** to interpret why outcomes vary among environments. It replaces crude meta-analysis of treatment effects with principled explanations of mechanisms and their impacts and extrapolates mechanisms to other environments to answer **P1** in those environments.[14] Structural models are useful vehicles for summarizing evidence from multiple studies.[15] Forecasting in new environments is a traditional problem in econometrics (see, e.g., Theil, 1958; Hamilton, 2000; Chatfield, 2000). However, the truly ambitious problem addressed by policy analysts is **P4**.

P4 *Forecasting the impacts of interventions (constructing counterfactual states associated with interventions) never previously implemented to various environments, including their impacts in terms of well-being.*

This is a fundamental challenge addressed in econometric policy analysis. This problem motivated the creation of econometric causal models.[16] It is also a central feature of the scientific analysis of empirical regularities.

One impetus for the econometric structural approach was to conduct policy analysis for the post-World War II era using models fit on data from the pre-World War II Depression era. Econometric policy analysis is the vehicle for framing and addressing the likely impacts of new policies and new environments, never previously experienced. Marschak (1953) provides an insightful discussion of this task in the context of forecasting the impact of new economic policies using data collected in environments in which the proposed policies were not in place.[17] The often-cited "critique" of Lucas (1976) updates Marschak's policy analysis to stochastic environments. McFadden (1974) is a Nobel-Prize winning example of how a leading economist who successfully met this challenge in forecasting the demand for a new transportation system in the San Francisco Bay area.

The econometric approach distinguishes three tasks of econometric causal policy analysis. These tasks are described in Table 1 and are often conflated in the non-economic statistical literatures.

Table 1
Three distinct tasks in econometric causal analysis.

Task	Description	Requirements	Types of analysis
1: Model Creation	Defining the class of hypotheticals or counterfactuals by thought experiments (models)	A scientific theory: A purely mental activity	Outside Statistics; Hypothetical Worlds
2: Identification	Identifying causal parameters from hypothetical populations	Mathematical analysis of point or set identification; this is a purely mental activity	Probability Theory
3: Estimation	Estimating parameters from real data	Estimation and testing theory	Statistical Analysis

Our regression example illustrates these distinctions. Models for counterfactuals do not necessarily require any statistical analysis. Identification is a separate issue required to recover β from hypothetical model distributions of data where statistical variation is

---

[13] Holland (1986) features the narrow goal of investigating the "effects of causes" in his definition of the Neyman–Rubin model.

[14] Recent work in computer science has begun to reinvent the logic of econometric forecasting using its own colorful private language but without any fresh insights or acknowledgment of a large body of econometric thought (see, e.g., Bareinboim and Pearl, 2016).

[15] See, e.g., Bursztyn and Yang (2021), Nerlove (1967), or Mullins (2024).

[16] See Frisch (1930, 1933, 1938) and Tinbergen (1930).

[17] Knight (1921) succinctly states the problem and its solution in his enigmatic remark, *"The existence of a problem of knowledge depends on the future being different from the past, while the possibility of a solution of the problem depends on the future being like the past."* Knight meant that analysts use ingredients estimated on historical data to construct forecasts of the unknown. This is a task that involves judgments and insights about invariant mechanisms beyond straight applications of fitted statistical models.

not an issue.[18] Estimation, on the other hand, considers how to recover model parameters from empirical sampling distributions where statistical variation is a concern. Trygve Haavelmo, a student of Frisch, developed an empirically operational econometric framework for causal analysis that distinguished these three tasks ([1943](#), [1944](#)).

3. Econometric causal models

Econometric causal models are flexible frameworks that can be used to address a variety of economic policy problems that are not naturally squeezed into simple "treatment-control" frameworks. They go well beyond the narrow treatment effect literature to address the following topics listed in Table 2[19] :

Table 2
Problems addressed by econometrics.

(a)	Investigate the causes of effects, not just the effects of causes — the goal of the treatment effect literature announced by Holland (1986) in defining the "Rubin model;"
(b)	Interpret empirical relationships within economic choice and outcome frameworks;
(c)	Analyze data using *a priori* information from theory;
(d)	Account systematically for shocks, errors by agents, and measurement errors;
(e)	Analyze dynamic models;
(f)	Accommodate multiple approaches to identification beyond randomization, instrumental variables, and matching that exploit restrictions within and across equations on causal relationships produced by theory;
(g)	Exploit restrictions across equations and unobservables within and across equations to identify causal parameters;
(h)	Make forecasts in new environments;
(i)	Synthesize evidence across studies using common parameters embedded in common conceptual frameworks rather than crude statistical meta-analysis;
(j)	Make forecasts of new policies never previously implemented; and
(k)	Analyze interactions across agents within markets and also within social settings (general equilibrium and peer effects).

Econometric methodology for establishing causality is comprehensive and adaptable, as it is specifically designed to address a wide array of causal questions pertinent to economics. In contrast, alternative causal frameworks are often not conceived with the specific investigative needs of economists in mind. Consequently, these methods are typically tailored to address only a specific subset of causal questions, primarily focusing on the application of a limited range of techniques to specialized categories of problems, predominantly those within the problem class P1.

Alternative methodologies, such as the NR approach, can be highly effective in analyzing causal effects such as average treatment effects or the effect of treatment on the treated within the contexts of Randomized Controlled Trials (RCTs). However, their utility becomes markedly constrained when addressing the more complex causal questions mentioned in Table 2. This is a consequence of *Marschak's Maxim* (Heckman, 2008a) that for certain narrowly focused problems, special versions of the econometric approach are highly effective. One need not necessarily implement more general models that address a wider set of questions when addressing specific focused problems. However, such models are, by design, of limited value in addressing wider classes of problems. We now state the econometric model formally using the convenient tool of graph theory that is widely used in many branches of applied mathematics.

3.1. Causal framework

Heckman and Pinto (2015) develop a causal framework that formalizes Frisch's insight that causality is the outcome of a thought experiment and places Haavelmo's approach ([1943](#), [1944](#)) in the framework of more recent policy evaluation models. They distinguish an *empirical model* that generates the observed data from a *hypothetical model* that formalizes the thought experiments of manipulating inputs that define causality. The empirical model describes the data generating process. The model differs from the hypothetical model, which serves as an abstract thought experiment. They establish the definition and operationalization of causality within a probabilistically coherent framework, eliminating the need for special rules or procedures designed to delineate causality, a necessity often found in non-economic literature.

Some notation is useful in describing the framework. We borrow it from the literature in applied mathematics. Dawid (1979) is a major source of conditional independence relationships. Lauritzen (1996) is a concise treatment of the graph theory we use.

3.2. A causal model

A causal model $\mathbb{M}$ is a system of policy-invariant (autonomous) structural equations like (1) that characterize the mapping $\mathbb{M} : \mathcal{T} \to \mathbb{P}(\mathcal{T})$ between a set of variables $\mathcal{T}$ and its power set $\mathbb{P}(\mathcal{T})$. Elements in $\mathcal{T}$ are random variables or random vectors that may be observed or unobserved by the analyst. It is convenient to define the set $\mathcal{E} = \{\epsilon_K; K \in \mathcal{T}\}$ that contains an error term ϵ_K for each $K \in \mathcal{T}$. Error term ϵ_K shares the same dimension as K. This variable is assumed present even if there are additional unobserved

---

[18] Lewbel (2019) and Fisher (1966) are definitive treatments of identification in economics.
[19] Table 2 is only a partial list of the rich array of problems addressed by the econometric approach.

variables. Error terms are mutually independent and independent of the other externally-specified system inputs. They enable us to avoid degenerate random variables, so standard tools of probability theory can be used.

A structural equation for a variable $K \in \mathcal{T}$ is an autonomous function denoted by $f_K : (\mathbb{M}(K), \epsilon_K) \to \mathbb{R}^{|K|}$. Variables in $\mathbb{M}(K)$ are said to directly cause K. In recursive formulations, a variable cannot directly cause itself, that is, $K \notin \mathbb{M}(K)$ for all $K \in \mathcal{T}$. We relax recursivity in a later section, where we discuss simultaneous equations models in which sets of outcome variables are jointly determined.

A variable T not caused by any variable, so $\mathbb{M}(T) = \varnothing$, is called *external*. In this case, its structural function is given by $T = f_T(\epsilon_T)$. Error terms are externally-specified (i.e., exogenous). This means that error terms in set ϵ are not caused by any variable in $\mathcal{T}$. We impose, without loss of generality, that error terms are mutually statistically independent.[20] All variables are defined on a common probability space $(\mathcal{I}, \mathcal{F}, P)$, using standard notation for σ - algebras.

3.3. The generalized Roy model

We use the Generalized Roy model as our leading example of a structural model. It is a cornerstone of the literature in applied economics and policy evaluation.[21] The original Roy model of counterfactuals (1951) analyzed earnings inequality in two sectors of the economy. All persons have two potential incomes: $Y(0)$ in Sector 0 and $Y(1)$ in Sector 1. Agents choose sectors based on their perceived net benefit I. In the simplest case, the benefit is the income gain $I = Y(1) - Y(0)$. More general models allow for costs, like tuition, migration costs, and psychic costs of participation. Potential incomes $(Y(0), Y(1))$ depend on observed variables X, while benefit I may depend on $Y(1), Y(0)$, and X and an externally specified variable Z, which may be policy variables that influence participation costs. The agent's choice of sector is given by $T = \mathbf{1}[I(X, Z) > 0]$. The model has been generalized to analyze multiple sectors and dynamic discrete choices (see Abbring and Heckman, 2007; Heckman and Vytlacil, 2007a,b).

The most common form of the model is:

$$Y(1) = g_1(X) + U_1 \tag{4}$$

$$Y(0) = g_0(X) + U_0 \tag{5}$$

$$C = g_c(Z, X) + U_c \tag{6}$$

$$T = \mathbf{1}[Y(1) - Y(0) - C \geq 0]. \tag{7}$$

g_1, g_0, and g_c are autonomous functions. The variables X are observed and cause the outcome and choices. Variable Z serves as an instrumental variable. It is not an argument of the outcome equations. Variables U_1, U_0, and U_c are exogenous and unobserved variables that are statistically independent of Z, X, namely $(U_1, U_0, U_c) \perp\!\!\!\perp (X, Z)$.[22] Choice theory, as embodied in (7), helps in determining relevant variables Z which can serve as instrumental variables.

The individual level treatment effect is $Y(1) - Y(0)$. The evaluation problem arises because for each person we observe either $Y(0)$ or $Y(1)$, but not both. We observe $Y(1)$ if $T = 1$ and $Y(0)$ if $T = 0$, namely $Y = T \cdot Y(1) + (1 - T) \cdot Y(0)$.[23]

Z affects Y only through its influence on T. The typical analysis reformulates the analysis at the population level rather than at the individual level. A common parameter of interest is the average treatment effect ATE $= E(Y(1) - Y(0))$ which is the mean treatment effect across all agents. Treatment on the treated focuses on $TOT = E(Y(1) - Y(0)|T = 1)$. The probability distributions of the counterfactual outcomes $Y(t); t \in \{0, 1\}$ are sometimes investigated.

The Generalized Roy model has been extended in many ways.[24] The model is systematically ignored in the non-economic literatures, despite its intellectual priority and relevance.[25] The Generalized Roy model allows for multiple choices. It can account for subjective evaluations of the benefits of each choice by subsuming variables U_1, U_0, U_c, in an unobserved random vector V that causes both T and Y (see Heckman and Vytlacil, 2007b,a; Heckman and Pinto, 2018).

A simple yet general 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 externally-specified (exogenous) and unobserved confounding variable that causes both T and Y.[26] In the context of the Generalized Roy model, Z stands for external policy or "shifter" vectors. V is a source of selection bias as it induces covariation between choice T and outcome Y that is not due to the causal effect of the treatment T on the outcome Y. For now, we suppress the X variables for the sake of notational simplicity.

Table 3 displays four equivalent representations of the Generalized Roy model.

---

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

[21] See, e.g., Heckman and Vytlacil (2007b,a), Heckman and Taber (2008).

[22] This independence relationship may also take the form of the conditional independence $(U_1, U_0, U_c) \perp\!\!\!\perp Z|X..$

[23] This switching regression relationship was first used by Quandt (1958). See also Quandt (1988).

[24] For instance, Heckman and Vytlacil (2007a) investigate multiple versions of the original model. Heckman et al. (2008) extend the model to ordered and general unordered choice models. 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.

[25] See e.g., (Pearl, 2009a, 2012; Rubin, 1974, 1978; Holland, 1986; Imbens and Rubin, 2015).

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

Table 3
Representations of the Generalized Roy Model.

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

The first column of Table 3 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.[27] To avoid clutter, we keep the ϵ implicit.

The last representation of Table 3 uses a property called the Local Markov Condition (LMC).[28] 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 comprise 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.[29] A causal model is recursive (acyclic) if no variable is a descendant of itself, that is, $K \notin \mathbb{D}(K)$.

The LMC is a property of recursive models stating that a variable is independent of its non-descendants conditioned on its parents.

$$\textbf{LMC}: \quad K \perp\!\!\!\perp (\mathcal{T} \setminus \mathbb{D}(K)) \mid \mathbb{M}(K) \tag{8}$$

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 3. 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[30] of Dawid (1976) or through graphical methods such as the d-separation criteria of Geiger et al. (1990).

3.4. Counterfactual approaches: Formalizing Frisch's insight

Frisch's statement that "Causality is in the Mind" means that the causal analysis of treatment T relies on a thought experiment that assigns values to the treatment variable in a fashion external to the system analyzed. This hypothetical manipulation of T affects only the variables caused by T. Specifically, changing T affects its descendant Y but not its ancestors V, Z.

Frisch's thought experiment is conceptually simple. However, it is a causal operation outside the scope of statistical theory. In statistics, random variables are fully characterized by their joint distributions. This information by itself is insufficient for causal analysis as it lacks directionality — a central feature of causal models. Frisch's thought experiment uses additional information on causal direction when it partitions the variables studied into those caused by T and those that are not.

Frisch's thought experiment was formalized through the use of the "*fix*" or "*set*" operator implicit in the seminal work of Haavelmo (1943). *Counterfactual* outcomes are obtained by the hypothetical (external) manipulation of the targeted variable that causes the outcome of interest. 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 affects only the outcome equation. It substitutes the treatment random variable T by the treatment value t. It makes all descendants of T functions of the fixed value of $T = t$. It does not eliminate the equation for T from the causal model nor does it modify the choice equation $T = f_T(Z, V, \epsilon_T)$.

The *do-operator* of Pearl (1995, 2012) operates in a fashion 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

---

[27] We refer to Lauritzen (1996) for background on DAGs and Bayesian Networks.

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

[29] 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, $\mathbb{Ch}(\mathbb{D}(K)) = \mathbb{D}(K)$.

[30] Dawid (1976) defines Graphoid Axioms as consisting 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:		$X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid Z.$

.

the treatment variable T, which effectively suppresses the determining equation of random variable T from the causal model and replaces it with a fixed value $T = t$ that affects all descendent relationships. Eliminating this equation excludes the possibility of defining parameters like TOT that condition on T.

Neither *fix* nor *do* are well-defined in statistics. They are causal operators that only affect the distribution of the descendants of the variable being fixed. In contrast, statistical conditioning affects the distributions 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 causal operations of *fixing* or *doing* in a framework using standard statistical tools. 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 is an abstract model (thought experiment) that shares the same structural equations and the same distributions of error terms as the empirical model. It differs from the empirical model by appending a hypothetical variable $\widetilde{T}$ that replaces the T-input affecting descendants of T. The hypothetical variable captures the causal notion of an external manipulation of treatment. The hypothetical model operates downstream of T and translates the causal operation of fixing T into the statistical operation of conditioning on $\widetilde{T}$.

Notationally, we use $\mathcal{T}_e, \mathcal{E}_e, \mathbb{M}_e, P_e, E_e$ for the variable set, error terms, causal model, probability, and expectation of the empirical model, while $\mathcal{T}_h, \mathcal{E}_h, \mathbb{M}_h, P_h, E_h$ denote the counterparts in the hypothetical model. The hypothetical and empirical models are related in the following fashion: (1) the hypothetical model has an additional variable $\widetilde{T}$, $\mathcal{T}_h = \mathcal{T}_e \cup \{\widetilde{T}\}$; (2) the hypothetical variable causes all descendants of T, $\mathbb{M}_h(K) = (\mathbb{M}_e(K) \cup \{\widetilde{T}\}) \setminus \{T\}$ for all $K \in \mathbb{D}_e(T)$; (3) variable T has no descendants in the hypothetical model, that is, $\mathbb{D}_h(T) = \varnothing$; (4) all remaining causal relations stay the same, that is, $\mathbb{M}_h(K) = \mathbb{M}_e(K)$ for all $K \in \mathcal{T}_e \setminus \{\mathbb{D}_e(T) \cup \{T\}\}$.

It is useful to illustrate these ideas 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 4 displays the Roy model for each of these frameworks.

The first column of Table 4 presents the original empirical model. The second and third columns present the models generated by the *fix* and the *do* operators respectively. Both models constrain the T-input of the outcome equation by a value $t \in \mathrm{supp}(T)$. The main difference between these models is that *fix* retains the equation for treatment while *do* suppresses it. The hypothetical model is displayed in the last column of Table 4. It replaces the T-input of the outcome equation with an external hypothetical variable $\widetilde{T}$.

The first panel presents the structural equations of each approach. The second panel displays the models as DAGs. The third panel describes the independence relationships generated by each causal model, and the last panel of the table presents the factorization of the joint distribution of the model variables. We use P_e for the probability distribution of the empirical model, P_{fix} for the model generated by the fix operator, P_{do} for the *do* operator and P_h for the hypothetical model. The factorizations differ according to the number of variables and causal relations of each counterfactual model.

The empirical ($\mathbb{M}_e$), fix ($\mathbb{M}_{fix}$), and hypothetical ($\mathbb{M}_h$) models share the same distributions of error terms $\epsilon_Z, \epsilon_V, \epsilon_T, \epsilon_Y$. Therefore the joint distribution of non-descendant T, that is (V, Z), is the same across these models. The *do* model eliminates the error term ϵ_T, and the distribution of T is not defined.

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 have the same distribution of $Y(t)$. The hypothetical variable $\widetilde{T}$ enables us to circumvent the necessity of introducing a special causal operator. The variable has no parents and, according to the LMC (8), it is independent of all its non-descendants, $\widetilde{T} \perp\!\!\!\perp (T, V, Z)$. In particular, $\widetilde{T} \perp\!\!\!\perp T$ always holds for any hypothetical model $\mathbb{M}_h$. $\widetilde{T}$ is also statistically independent of error terms $\epsilon_Z, \epsilon_T, \epsilon_V, \epsilon_Y$, and the counterfactual outcome is obtained by simply conditioning on $\widetilde{T}$. In summary, we have that:

$$\left(Y \mid \widetilde{T} = t\right)_{\mathbb{M}_h} \overset{d}{=} \left(Y(t)\right)_{\mathbb{M}_{fix}} \overset{d}{=} \left(Y(t)\right)_{\mathbb{M}_{do}}. \tag{9}$$

It is also the case that Eq. (9) holds when conditioned on any variable K that is not a descendant of $\widetilde{T}$, namely, Z, V and T.
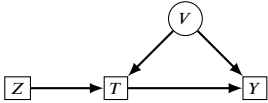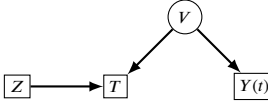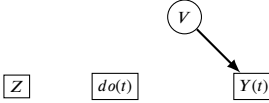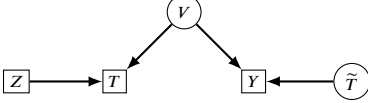
To fix ideas, let T be an indicator of college graduation and Y denote adult income. Treatment-on-the-treated (TOT) is the average causal effect of college on income for those who choose to go to college ($T = 1$), which is $TOT = E_{fix}(Y(1) - Y(0) \mid T = 1)$ using the fix operator. The parameter is equivalently described using the hypothetical model as $TOT = E_h(Y \mid \widetilde{T} = 1, T = 1) - E_h(Y \mid \widetilde{T} = 0, T = 1)$. The *do* operator *excludes* the treatment variable T, which poses a challenge in defining the TOT parameter. Shpitser and Pearl (2009) solve this problem by incorporating an additional special structure to their counterfactual model.

Eq. (9) may suggest that the way that counterfactuals are expressed is of little relevance in the study of causality. That assessment is quite misleading. Seemingly small differences in characterizing counterfactuals have significant consequences for the machinery used to identify causal effects. Section 6 illustrates the difference between an identification analysis using the *do*-calculus and an identification analysis using the hypothetical model framework. Section 5 compares identification in NR with identification in the structural model.

3.5. Identification of counterfactual outcomes

We next consider Task 2 in Table 1. 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

Table 4
Generalized Roy Model: Approaches to generating counterfactuals.

		Empirical Models			Hypothetical Model
		Empirical Model ($\mathbb{M}_e$)	Fixing T at t ($\mathbb{M}_{fix}$)	Doing $do(t)$ ($\mathbb{M}_{do}$)	Hypothetical Model ($\mathbb{M}_h$)
			Structural Equations		
V		$V = f_V(\epsilon_V)$	$V = f_V(\epsilon_V)$	$V = f_V(\epsilon_V)$	$V = f_V(\epsilon_V)$
Z		$Z = f_Z(\epsilon_Z)$	$Z = f_Z(\epsilon_Z)$	$Z = f_Z(\epsilon_Z)$	$Z = f_Z(\epsilon_Z)$
T		$T = f_T(Z,V,\epsilon_T)$	$T = f_T(Z,V,\epsilon_T)$	$do(T = t)$	$T = f_T(Z,V,\epsilon_T)$
Y		$Y = f_Y(T,V,\epsilon_Y)$	$Y(t) = f_Y(t,V,\epsilon_Y)$	$Y(t) = f_Y(t,V,\epsilon_Y)$	$Y = f_Y(\widetilde{T},V,\epsilon_Y)$
					$\widetilde{T} = f_{\widetilde{T}}(\epsilon_{\widetilde{T}})$
			Directed Acyclic Graphs (DAGs)		
			Local Markov Conditions		
V		$V \perp\!\!\!\perp Z$	$V \perp\!\!\!\perp Z$	$V \perp\!\!\!\perp Z$	$V \perp\!\!\!\perp (Z,\widetilde{T})$
Z		$Z \perp\!\!\!\perp V$	$Z \perp\!\!\!\perp (V,Y(t))$	$Z \perp\!\!\!\perp (V,Y(t))$	$Z \perp\!\!\!\perp (V,Y,\widetilde{T})$
T		$T \perp\!\!\!\perp \emptyset \mid (Z,V)$	$T \perp\!\!\!\perp Y(t)\mid(Z,V)$	(not defined for the models without $\tilde{T}$)	$T \perp\!\!\!\perp (\widetilde{T},Y)\mid(Z,V)$
Y		$Y \perp\!\!\!\perp Z\mid(T,V)$	$Y(t) \perp\!\!\!\perp (Z,T) \mid V$	$Y(t) \perp\!\!\!\perp Z\mid V$	$Y \perp\!\!\!\perp (Z,T) \mid (\widetilde{T},V)$
$\widetilde{T}$		(not defined for the models without $\tilde{T}$)	(not defined for the models without $\tilde{T}$)	(not defined for the models without $\tilde{T}$)	$\widetilde{T} \perp\!\!\!\perp (T,V,Z)$
			Factorial Decomposition of the Joint Probability Distributions		
		$P_e(Y,T,V,Z) =$	$P_{fix}(Y(t),T,V,Z) =$	$P_{do}(Y(t),V,Z) =$	$P_h(Z,V,T,\widetilde{T},Y) =$
		$P_e(Y\mid T,V)P_e(T\mid Z,V)P_e(V)P_e(Z)$	$P_{fix}(Y(t)\mid V)P_{fix}(T\mid V,Z)P_{fix}(V)P_{fix}(Z)$	$P_{do}(Y(t)\mid V)P_{do}(V)P_{do}(Z)$	$P_h(Y\mid\widetilde{T},V)P_h(T\mid Z,V)P_h(V)P_h(Z)P_h(\widetilde{T})$

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. Notice further that Z and V are externally specified so that $P_e(Z) = P_{fix}(Z) = P_{do}(Z) = P_h(Z)$ and $P_e(V) = P_{fix}(V) = P_{do}(V) = P_h(V)$.

variables with the population distributions of the empirical model. The mechanics for establishing this connection depend on which causal model is used to describe counterfactuals.

First consider the *fix* operator of model $\mathbb{M}_{fix}$ in Table 4. The LMC of $Y(t)$ in $\mathbb{M}_{fix}$ implies that:

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

Eq. (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 helps identify treatment effects as it connects the counterfactual outcome $Y(t)$ in $\mathbb{M}_{fix}$ with the empirical model $\mathbb{M}_e$:

$$P_{fix}\big(Y(t) \mid V\big) = P_{fix}\big(Y(t) \mid T - t, V\big), \tag{11}$$

$$= P_{fix}\left(\sum_{t \in \text{supp}(T)} \mathbf{1}[T = t] Y(t) \mid T - t, V \right), \tag{12}$$

$$= P_{fix}\left(\sum_{t \in \text{supp}(T)} \mathbf{1}[T = t] f_Y(t, V, \epsilon_Y) \mid T - t, V \right), \tag{13}$$

$$= P_{fix}\big(f_Y(T, V, \epsilon_Y) \mid T - t, V\big), \tag{14}$$

$$= P_e\big(Y \mid T - t, V\big). \tag{15}$$

Eqs. (11)–(15) use 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). Eqs. (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 analysts to connect the counterfactual and empirical distributions in a systematic manner. For any disjoint set of variables Y, W in $\mathcal{T}_h \setminus \{T, \widetilde{T}\}$ and any values $t, t' \in \text{supp}(T)$ we have that:[31]

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

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

Eqs. (16)–(17) state two conditions that involve independence relationships in the hypothetical model. They state that we can switch from the hypothetical to the empirical model whenever the hypothetical model yields the independence relationships (16) and (17).[32] The application of these rules is simple. For example, the LMC of Y in $\mathbb{M}_h$ of Table 4 generates the following matching condition:

$$Y \perp\!\!\!\perp T | (\widetilde{T}, V). \tag{18}$$

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 systematic approach for connecting hypothetical and empirical models. The framework employs additional structure beyond what is obtained from fixing that might not be required in analyzing the simple 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 condition such as (10) or (18) because the equation for 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 be 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.

The counterfactual models $\mathbb{M}_{fix}$, $\mathbb{M}_h$ and $\mathbb{M}_{do}$ employ distinct techniques to generate the same conclusion: that identification of the counterfactual outcome requires analysts to control for an unobserved confounding variable V. Summarizing, 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 an unobserved confounding variable V. We discuss part of this menu in the next section.

---

[31] 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.

[32] See Heckman and Pinto (2015) for further discussion of the connection between empirical and hypothetical models.

4. Econometric approaches to identification of counterfactuals in the Generalized Roy model

The Generalized Roy model is a framework for exploring the large toolkit of the econometric approach for identifying counterfactuals. We compare what is possible in the econometric approach with what can be obtained using the non-econometric paradigms. We describe several of these approaches here. We develop this discussion further in subsequent sections of this paper.

Eq. (19) states that the identification of causal effects in the Generalized Roy model hinges on controlling for the unobserved confounding variables V. A popular approach to doing so uses instrumental variables that are independent of V. It controls for V by shifting T without affecting the distribution of V. However, the IV approach using Z as an instrument does not identify interesting counterfactuals without additional assumptions.

For a simple example, consider a linear model in which the structural treatment equation is $T = \alpha_0 + \alpha_1 Z + \alpha_2 V + \epsilon_T$, and the outcome function is $Y = \beta_0 + \beta_1 T + \beta_2 V + \epsilon_Y$, where α_0, α_1, α_2, β_0, β_1, β_2 are scalar parameters. In this model, the causal effect of T on Y is given by β_1 and is identified by the covariance ratio $cov(Y, Z)/cov(T, Z)$. This parameter can be estimated by a Two-Stage Least Squares (2SLS) Regression. This tool has been available to economists since the 1950s.[33]

However, the Generalized Roy model is not captured by this simple two-equation system. The causal effect, $Y(1) - Y(0)$ is, in general, a random variable and not a constant. Thus treating β_1 as a constant does not capture the essential heterogeneity of treatment effects across agents.[34] The analogue to heterogeneous β_1 is stochastically dependent on V. There are numerous approaches to identifying its distribution. We start with the use of instrumental variables in the presence of heterogeneous treatment effects and then consider alternative approaches.

4.1. Instrumental variables

Heckman and Vytlacil (1999, 2005) address the question of identifying the Roy model by assuming a separable choice equation. Their approach enables analysts to control for V and, in turn, identify counterfactual outcomes. Their local Instrumental Variable (LIV) approach considers a binary treatment $T \in \{0, 1\}$. Their *separability assumption* is motivated by economic choice theory and states that treatment is given by a latent threshold-crossing equation that includes instrument Z and the confounder V; that is, $T = \mathbf{1}[\zeta(Z) \geq \phi(V)]$. Separability enables them to rewrite the choice equation as:

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

Where the probability of treatment selection, denoted as $P(Z) = P_e(T = 1|Z)$, is typically referred to as the propensity score. The unobserved variable U is given by $U = F_{e,\phi(V)}(\phi(V))$ where $F_{e,\phi(V)}$ is the cdf of $\phi(V)$, which is monotone increasing by construction. Subscript "e" denotes that the probability distribution is constructed using the empirical model. Variable U has a uniform distribution if $\phi(V)$ is absolutely continuous; that is, $U \sim \text{unif}([0, 1])$. The structural approach uses unobservables. The Neyman–Rubin approach does not. The *do-calculus* uses them, but in a limited way. We show in Section 6 that it rules out exploiting the information used to obtain (20). This approach to unobservables precludes the use of methods that are fruitful in the econometric approach.

The hypothetical and empirical models for the Generalized Roy model that include the unobserved variable U are displayed in Table 5. The LMC of T in the hypothetical Roy model of Table 5 implies that $Y \perp\!\!\!\perp T \mid (Z, \widetilde{T}, U)$. The LMC of Z implies $Y \perp\!\!\!\perp Z \mid (U, \widetilde{T})$. These two independence relationships imply, by contraction property D, that $Y \perp\!\!\!\perp T \mid (\widetilde{T}, U)$. Following the same analysis of V as (19), $Y \perp\!\!\!\perp T \mid (\widetilde{T}, U)$ implies that:

$$P_h(Y \mid \widetilde{T} = t, U) = P_{fix}(Y(t) \mid U) = P_{do}(Y(t) \mid U). \tag{21}$$

Otherwise stated, controlling for U enables analysts to identify counterfactual outcomes in the same fashion that controlling for V does. Variable U is called a *balancing* score for V. This means that U is a surjective function of V that preserves the independence relationship $Y \perp\!\!\!\perp T \mid (\widetilde{T}, V) \Rightarrow Y \perp\!\!\!\perp T \mid (\widetilde{T}, U)$.[35]

The Local Instrumental Variable (LIV) model of Heckman and Vytlacil (1999) can be used to identify probability distributions of counterfactual outcomes conditioned on U by taking the derivative of the observed outcome with respect to the propensity score. More generally, the counterfactual expectation $E_{fix}(g(Y(t)) \mid U = u)$ for any real-valued function $g : \mathbb{R} \to \mathbb{R}$ is identified if there is sufficient variation of propensity score $P(Z)$ around the value $u \in (0, 1)$.

Identification of $E_h g(Y \mid \widetilde{T} = t, U = u)$ comes from the derivative of the expectation $(-1)^{1-t} E_e(g(Y)\mathbf{1}[T = t] \mid P(Z))$ with respect to the propensity score at the value $P(Z) = u$. In particular, it can be shown that:

$$E_h\big(Y \mid \widetilde{T} = 1, U = u\big) - E_h\big(Y \mid \widetilde{T} = 0, U = u\big) \equiv E_{fix}\big(Y(1) - Y(0) \mid U = u\big) = \left. \frac{\partial E_e(Y \mid P(Z))}{\partial P(Z)} \right|_{P(Z)=u} \tag{22}$$
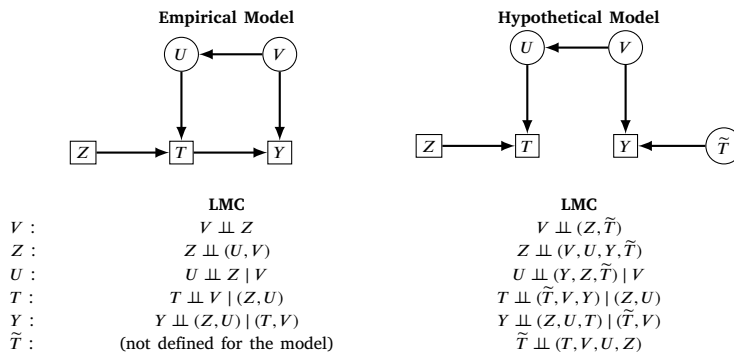
where *fix* refers to the distribution generated by fixing (which is the same as that generated by "doing") and e refers to the sample distribution. Identification requires sufficient variation of the propensity score $P(Z)$ around $u \in [0, 1]$. If $P(Z)$ has full support, the average treatment effect can be evaluated by ATE $\equiv E_h(Y \mid \widetilde{T} = 1) - E_h(Y \mid \widetilde{T} = 0) = \int_0^1 \big(E_h(Y \mid T = 1, U = u) - E_h(Y \mid T = 0, U = u)\big) du$.

---

[33] See Theil (1953, 1958, 1971), Amemiya (1985), Hansen (2022). Theil (1953) invented this method. The method is far more general and applies to nonlinear models as well.

[34] A heterogeneous treatment effect case would write $\beta_1 = (Y_1 - Y_0)$ and $\beta_0 = Y_0$.

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

Table 5
Binary Choice Roy Model: Empirical and hypothetical causal models.

	Empirical Model		**Hypothetical Model**

	LMC		**LMC**
V :	$V \perp\!\!\!\perp Z$	V :	$V \perp\!\!\!\perp (Z, \widetilde{T})$
Z :	$Z \perp\!\!\!\perp (U, V)$	Z :	$Z \perp\!\!\!\perp (V, U, Y, \widetilde{T})$
U :	$U \perp\!\!\!\perp Z \mid V$	U :	$U \perp\!\!\!\perp (Y, Z, \widetilde{T}) \mid V$
T :	$T \perp\!\!\!\perp V \mid (Z, U)$	T :	$T \perp\!\!\!\perp (\widetilde{T}, V, Y) \mid (Z, U)$
Y :	$Y \perp\!\!\!\perp (Z, U) \mid (T, V)$	Y :	$Y \perp\!\!\!\perp (Z, U, T) \mid (\widetilde{T}, V)$
$\widetilde{T}$:	(not defined for the model)	$\widetilde{T}$:	$\widetilde{T} \perp\!\!\!\perp (T, V, U, Z)$

4.2. Stratification

A recurrent theme in this section is that the identification of counterfactual outcomes hinges on controlling for the confounding variable V. The solution of the LIV model invokes the separability assumption (20), which generates a balancing score U for V. According to (23), the nonparametric point-identification of the counterfactual outcomes conditioned on $U = u$ is obtained by differentiating the outcome with respect to the propensity score $P(Z)$ at value $u \in (0, 1)$.

Eq. (22) assumes that the sample propensity score has enough variation around the value $u \in (0, 1)$. Consequently, the equation is not directly applicable to the case of discrete instruments. One approach to overcoming this limitation is to use the discrete counterpart of Eq. (22). Heckman and Vytlacil (2005) show that for any two values $z, z' \in \text{supp}(Z)$ such that $P(z') = u' > u = P(z)$ we have that:

$$\frac{E_e(Y \mid Z = z') - E_e(Y \mid Z = z)}{P_e(T = 1 \mid Z = z') - P_e(T = 1 \mid Z = z)} = \frac{\int_u^{u'} E_{fix}\left(Y(1) - Y(0) \mid U = u\right) du}{u' - u}$$
$$= E_{fix}(Y(1) - Y(0) \mid u \leq U \leq u'). \tag{23}$$

Eq. (23) states that difference of mean outcomes conditional on two instrumental values z, z' identifies the counterfactual outcome over an interval of U defined by the propensity scores $P(z)$ and $P(z')$. The equation evaluates a causal effect that depends on the values of the instrument. These effects are called Local Average Treatment Effects (LATE) by Imbens and Angrist (1994). LATE-type effects differ from causal effects such as ATE or TT, which do not depend on the IV values.[36]

A consequence of (23) is that ATE can be identified if there are two instrumental variable values z_0, z_1 such that z_0 induces full treatment nonparticipation ($P(z_0) = 0$), and z_1 induces full treatment participation ($P(z_1) = 1$):

$$E_e(Y \mid Z = z_1) - E_e(Y \mid Z = z_0) = E_{fix}\left(Y(1) - Y(0) \mid 0 \leq U \leq 1\right)$$
$$= E_h(Y \mid \widetilde{T} = 1) - E_h(Y \mid \widetilde{T} = 0) = \text{ATE}.$$

This setup is equivalent to a randomized control trial with full compliance. Mogstad and Torgovitsky (2018) use functional form assumptions to extrapolate estimates over intervals of U to point estimates.

Another approach for controlling for V exploits the discrete nature of the instrument to generate an alternative balancing score. Let instrument Z take values in the discrete set $\text{supp}(Z) = \{z_1, \ldots, z_N\}$ such that $P(z_1) < \cdots < P(z_N)$.[37] Let $T(z) = \mathbf{1}[\zeta(z) \geq \phi(V)]$ be the counterfactual choice that would occur if Z were fixed at value $z \in \{z_1, \ldots, z_N\}$. The *response vector* $S = [T(z_1), \ldots, T(z_N)]'$ is the random vector of potential choices across all Z-values.

Response vector S shares the same causal relationships of unobserved variable U in Table 5. By this we mean that S is a function of V and that the choice T can be written as function of Z and S:

$$T = \left[\mathbf{1}[Z = z_1], \ldots, \mathbf{1}[Z = z_N]\right] \cdot S.$$

Similar to U, the response vector S is a balancing score for V. The independence relationship $Y \perp\!\!\!\perp T \mid (\widetilde{T}, S)$ holds, which implies that $P_h(Y \mid \widetilde{T} = t, S) = P_e(Y \mid T = t, S)$. Heckman and Pinto (2018) show that the response vector S controls for V by generating a special partition of its support that spans the support of V and renders choice T statistically independent of V within each cell of the partition. Each column of S is a list of responses to different treatments for a person of a given V.

---

[36] Heckman et al. (2008) develop the relationship between LIV and LATE in depth.

[37] The increasing ordering of propensity scores is assumed without loss of generality.

The values of S are called response-types or strata.[38] The separability assumption eliminates some of the potential response-types. An influential example is due to Imbens and Angrist (1994), who investigate the case of a binary instrument and a binary treatment. There are four possible response-types termed always-takers, compliers, never-takers, and defiers. They invoke a monotonicity condition that is equivalent to the separability assumption Vytlacil (2002). The assumption eliminates the defiers and enables the identification of treatment effects for the compliers. See Heckman and Pinto (2018) and Buchinsky and Pinto (2021) for general results on identification.

4.3. The matching assumption

A popular method for identifying treatment effects assumes that a set of observed pre-treatment variables suffice 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. This assumption is called *Matching*.[39] Another (structural) way to state this is that X spans the space of V.

Table 6 presents the empirical and the hypothetical models that justify the matching assumption. The LMC of T in the hypothetical model implies that $Y \perp\!\!\!\perp T \mid (\widetilde{T}, X)$. According to (17), we have that $P_h(Y \mid \widetilde{T} = t, X) = P_{fix}(Y(t) \mid X) = P_e(Y \mid T = t, X)$ which means that the counterfactual outcome is identified by conditioning on X. Matching variables X are assumed not to be descendants of the hypothetical variable $\widetilde{T}$. Thus, $P_h(X) = P_e(X)$ and the probability distribution of the counterfactual outcome is given by $P_{fix}(Y(t)) = \int (P_e(Y \mid T = t, X = x) dF_{e,X}(x))$. The average causal effect of a binary treatment $T \in \{0,1\}$ is evaluated by the weighted average of mean difference between the treated and not-treated participants that *match* on X, namely, ATE $= \int \Big(E_e(Y \mid T = 1, X = x) - E_e(Y \mid T = 0, X = x) \Big) dF_{e,X}(x)$.[40]

Table 6
Matching model: Empirical and hypothetical causal models.
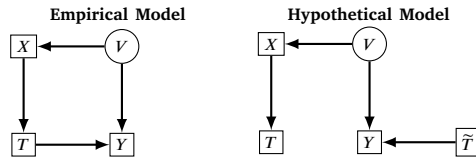
The matching assumption replaces the *unobserved* variable U of the Generalized Roy model in Table 5 by the *observed* variable X. In practice, it assumes that potential bias generated by confounding variables can be ignored when controlling for observed pre-treatment variables. Under matching, the identification of treatment effects does not require an instrumental variable nor additional assumptions such as separability. This assumption enables us to solve the problem of selection bias induced by unobserved variables V by conditioning on the observed variables X.

The matching assumption is justified in the case of randomized controlled trials (RCTs). In this case, the matching variables X denote the pre-treatment variables conditioned on in the randomization protocol. In observational studies, the matching assumption is often rather strong. It assumes that the analyst observes enough information to make all the agent's unobserved variables irrelevant (see Heckman, 2008b). Otherwise stated, matching assumes a symmetry in information between the economic agent and the econometrician.

There are several identification approaches that acknowledge the possibility of information asymmetries between the agent being studied and the econometrician: control function approaches, replacement functions, or proxy variables. These methods often differ considerably in terms of assumptions and methodology. However, they all share the same identification principle: they use observed data to evaluate a proxy variable that plays the role of a matching variable.

4.4. Matching on proxied unobservables

Matching on proxied unobservables is a version of matching that uses observed data to control for the confounding effects of V. Consider the modification of the Generalized Roy model in Table 7. The unobserved variable Q is a balancing score for the unobserved confounder V. The matching conditions of the hypothetical model, $Y \perp\!\!\!\perp T \mid (\widetilde{T}, Q)$, and its respective counterpart in the empirical model, $Y(t) \perp\!\!\!\perp T \mid Q$, hold. Variable Q has two additional properties: (1) it may cause outcome Y; and (2) it may be measured with error by the observed variable M.

A common setup where Q arises is in the evaluation of college economic returns where T denotes college graduation, Y denotes earnings, and Q denotes unobserved abilities such as cognition or conscientiousness. These abilities are not directly observed but measured with error by an observed vector of variables M, such as psychological surveys or test scores. Formally, we write

---

[38] The concept was developed by Robins (1986) and embellished in Frangakis and Rubin (2002).

[39] Heckman et al. (1998) investigate several estimation methods that invoke the matching assumption.

[40] Heckman et al. (1998) incorporate additive separability between observable and unobservable variables as well as exogeneity conditions that isolate outcomes and treatment participation into the matching framework. Additionally, they compare various types of estimation methods to show that kernel-based matching and propensity score matching have similar treatment of the variance of the resulting estimator.

$M = f_M(Q, \epsilon_M)$. The identification strategy exploits the structural function $M = f_M(Q, \epsilon_M)$ to evaluate Q, which, in turn, allows us to control for V and identify causal effects.

Matching on proxied unobservables has long been used in the economics of education (see, e.g., the essays in Goldberger and Duncan, 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., 2003; Cunha et al., 2005). Olley and Pakes (1996) is an application of this method. A common parametric approach extracts factors from psychological measurements to extract Q as a latent factor. Nonparametric factor analysis is developed in Cunha et al. (2010) and Schennach (2020). It is also possible to condition nonparametrically on Q without knowing the functional form of f_M.

Table 7
Matching on proxied unobservables: Empirical and hypothetical causal models.
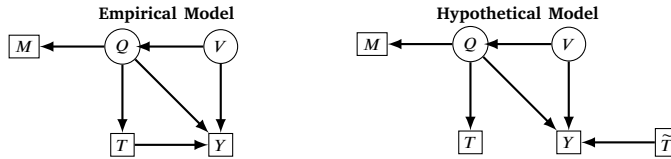
4.5. Control functions

The control function principle specifies the dependence of the relationship between observables and unobservables in a nontrivial fashion. The principle was introduced in Heckman and Robb (1985b,a) building on earlier work by Telser (1964) and later popularized by Blundell and Powell (2003). It was also applied in Carneiro et al. (2003) and Cunha et al. (2005). Heckman's sample selection correction (1979) is a control function.

We illustrate the control function principle using a version of the Generalized Roy model where V is a scalar random variable and the binary choice T is given by the *separable* equation $T = \mathbf{1}[\mu(Z) \geq V]$. Let $J = f_J(T, V, \epsilon_J)$ represents unobserved skills caused by the treatment T and the unobserved confounding variable V. In addition, let the outcome equation be *additive* in J, that is to say that the outcome Y can be written as $Y = f_Y(T, \epsilon_Y) + \psi(J)$, The model is displayed as a DAG in Table 8. The LMC of Y in the hypothetical model implies that $Y \perp\!\!\!\perp T \mid (\widetilde{T}, J)$. This means that J is a matching variable. The control function approach seeks to control for variable V by estimating the function $\psi(J)$ of the outcome equation.

Heckman and Vytlacil (2007a,b) use the assumption of separability of observables and unobservables in the choice equation and the outcome assumption of additivity to evaluate $\psi(J)$ as a function of the propensity score $P(Z)$. Similar to the LIV Model, we can use the CDF transformation to write the choice equation as $T = \mathbf{1}[P(Z) \geq F_V(V)]$, where $F_V(V) \sim \text{unip}([0, 1])$. Note that the expected value of the outcome conditional on $T = 1$ gives the *conditional* counterfactual mean:

$$E_e(Y \mid Z, T = 1) = E_{fix}(Y(1) \mid Z, T = 1) = E_h(Y \mid \widetilde{T} = 1, Z, T = 1),$$

where the first term is observed, the second term uses fixing and the last one uses the hypothetical model. Under separability and outcome additivity, we can express $E_h(Y(1) \mid \widetilde{T} = 1, Z, T = 1)$ as:

$$E_h(Y \mid \widetilde{T} = 1, Z = z, T = 1) = E_h(f_Y(\widetilde{T}, \epsilon_Y) \mid \widetilde{T} = 1) + E_h(\psi(J) \mid \widetilde{T} = 1, Z = z, T = 1),$$

$$= E_h(f_Y(1, \epsilon_Y)) + E_h(\psi(f_J(1, V, \epsilon_J)) \mid Z = z, T = 1),$$

$$\left(\text{setting } E_h(f_Y(1, \epsilon_Y)) = \alpha_1\right)$$

$$= \alpha_1 + E_h(\psi(f_J(1, V, \epsilon_J)) \mid P(z) > F_V(V)),$$

$$= \alpha_1 + E_e(\psi(f_J(1, V, \epsilon_J)) \mid P(z) > F_V(V)).$$

$$\therefore E_h(Y \mid \widetilde{T} = 1, Z, T = 1) = \alpha_1 + \underbrace{\phi_1(P(Z))}_{\text{control function}},$$

$$\text{where } \phi_1(P(Z)) = E_e(\psi(f_J(1, V, \epsilon_J)) \mid Z, T = 1).$$

where the first equality uses the additivity assumption, the second uses the fact the $\widetilde{T}$ is an external variable, the third uses the separability assumption, the fourth switches the hypothetical model into the empirical model as V, ϵ_J, Z are non-descendants of $\widetilde{T}$. The last equation gives the expectation $E_h(Y \mid \widetilde{T} = 1, Z, T = 1)$ as a function of the propensity score $P(Z)$. Control function $\phi_1(P(Z))$ can be estimated from observed data and the expected value of the counterfactual outcome can be evaluated as

$$E_h(Y(1)) = \int_0^1 \alpha_1 + \phi_1(p) \, dF_{P(Z)}(p).$$

Table 8
Control function: Empirical and hypothetical causal models.
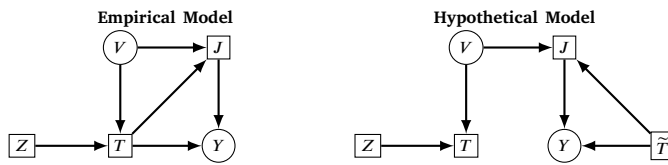
4.6. Panel data analysis and other approaches

A commonly used panel data method is **difference-in-differences** as discussed in Heckman and Robb (1985a), Blundell et al. (1998), Heckman et al. (1999), and Bertrand et al. (2004). All of the estimators previously discussed can be adapted to a panel data setting. Heckman et al. (1998) introduce difference-in-differences matching estimators to eliminate the bias in estimating treatment effects. Abadie (2005) extends this work. Separability between errors and observables is a common feature of the panel data approach in its standard application. Altonji and Matzkin (2005) and Matzkin (1993) present analyses of nonseparable panel data methods. Regression discontinuity estimators, which are versions of IV estimators, are discussed by Heckman and Vytlacil (2007b).

Table 9 summarizes some of the main identification approaches for the Generalized Roy model discussed here. The table barely scratches the surface, but gives a sense of the broad menu in the econometric approach. The essays in the *Handbooks of Econometrics* (Heckman and Leamer, 2001, 2007) give a range of other estimation approaches.

Table 9
Some alternative approaches that identify treatment effects by controlling for V.

$$Y \perp\!\!\!\perp T \mid (\widetilde{T}, X, V), \qquad T \in \{0, 1\}$$
$$E_h(Y \mid \widetilde{T} = t, X = x) = \int E_e(Y \mid T = t, X = x, V = v) \, dF_{e,V|X=x}(v)$$

	Method Assumes	Need Instrument (Z)?	Identify Distribution of V?
Matching[a]	V, X known	No	Yes (V observed)
Control Functions[b]	V estimated, X, Z known (continuous T); Bounds on quantiles of V estimated (discrete case)	Yes (except cases where functional forms secure identification)	Yes (over support)
Factor Method[c]	Distribution of V estimated from additional measurements of V (M)	No	Yes (with auxiliary measurements over support)
IV: LATE, LIV[d]	Z, X known	Yes	Estimate intervals of quantiles of V (Heckman and Vytlacil, 1999, 2005) and conditions on them; LIV shrinks interval of quantiles of V to a point using continuous instruments and conditions on them
Stratification[e]	Z, X known	Categorical instruments give restrictions on strata (balancing scores for V)	Identify distribution of strata which places interval bounds on V and conditions on them
Longitudinal Data Methods[f]	Variety of assumptions	Covariance restrictions	Yes; in long panels can identify V
Mixing Distributions[g]	$V \perp\!\!\!\perp X$	No (intervals of V)	Yes (Mixtures)

[a] (Rosenbaum and Rubin, 1983; Heckman et al., 1998); [b] (Blundell and Powell, 2003; Heckman and Robb, 1985b,a); [c] (Carneiro et al., 2003; Heckman et al., 2018; Cunha et al., 2010); [d] See review in Heckman and Vytlacil (2007a); [e] (Heckman and Pinto, 2018; Frangakis and Rubin, 2002); [f] (Abbring and Heckman, 2007; Heckman and Robb, 1985a); [g] (Cameron and Heckman, 1998; Prakasa Rao, 1992; Heckman and Singer, 1984).

5. The Neyman–Rubin (NR) causal model

The Neyman–Rubin causal approach uses the language and framework of experimental design developed by Neyman (1923), Fisher (1935), and Cox (1958) and popularized by Holland (1986). It ignores essential aspects of the econometric approach to causality and conflates distinct concepts (e.g., SUTVA).[41] It does not define hypothetical models nor does it employ structural equations to characterize causal models. It focuses on units of analysis instead of system of equations.

---

[41] Heckman (2008a) explains that: SUTVA – Stable Unit Treatment Value Assumption – conflates two two distinct concepts regarding functional autonomy (structural invariance) and no interactions among agents.

In this approach, causal models are characterized by statistical independence relationships among counterfactual counterparts of observed variables, never precisely justified. In place of the thought experiments that characterize the econometric approach, it uses randomized controlled experiments as the foundational paradigm. The contrast between thought experiments and randomized control experiments is central to understanding the differences in the approaches.

The NR approach lacks the clarity of interpretation offered by causal models described by structural equations. It is often difficult to map the independence relationships of an NR model into the causal relationships produced by economic theory. In particular, NR makes it difficult to use economic theory to assess the credibility of assumptions about the underlying structural equations that ensure the identification of causal effects or to interpret economic data using economic models.

Another drawback is that the NR framework lacks the fundamental tools of econometric causal analysis. It does not explicitly model unobserved variables in structural models. This feature substantially limits the use of standard econometric tools. It rules out (or makes cumbersome) several fruitful econometric strategies, such as balancing bias within models using compensating variations of arguments of structural functions to keep agents at the same levels of well being,[42] and cross-equation restrictions on both observable and unobservable model components,[43] or functional form restrictions. In practice, the set of tractable identification strategies that employ the NR framework is limited to a few possibilities: randomized trials, matching, IV and its many surrogates and differences-in-differences.[44] This section illustrates the drawbacks of NR in analyzing core policy questions or in synthesizing and interpreting evidence.

5.1. The Generalized Roy model under NR

The NR framework focuses on the unit of analysis $i \in \mathcal{I}$ which usually represents an economic agent or entity. The framework describes part of the Generalized Roy model of Table 3 using two counterfactuals: $T_i(z)$ is the potential treatment when the instrument Z is externally 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 $t \in \text{supp}(T)$. Properly formulated, potential outcomes are the outputs of structural equations. NR does not explicitly characterize the treatment choice equation. It prides itself on being nonparametric, although some proponents claim that assuming linearity is an innocuous assumption, even when models are fundamentally nonlinear.[45]

The NR framework characterizes the Generalized Roy model (4)–(7) by three assumptions:

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

The exclusion restriction means that Z does not directly cause Y. Thus, we can express the counterfactual outcome as $Y_i(t)$ instead of $Y_i(t, z)$. IV relevance means that T is caused by Z. The exogeneity condition of the NR framework can be traced back to the independence relationship between Z and V of the Generalized Roy model (4)–(7). In the NR framework, the exogeneity condition is an assumption. In the Generalized Roy model, the exogeneity condition is a consequence of the causal relation among model variables. Namely, that the Z and V are external variables. LMC (8) implies that $Z \perp\!\!\!\perp V$, which, in turn, generates the exogeneity condition.

The identification of counterfactual outcomes requires additional assumptions. A popular assumption securing identification is the monotonicity condition (24) of Imbens and Angrist (1994). It states that a change in an instrument induces agents to change their treatment choice in the same direction. Notationally, for any $z, z' \in \text{supp}(Z)$, it says that:

$$T_i(z) \geq T_i(z') \ \forall i \in \mathcal{I} \quad \text{or} \quad T_i(z) \leq T_i(z') \ \forall i \in \mathcal{I}. \tag{24}$$

Vytlacil (2002) shows that the monotonicity condition (24) is equivalent to the separability assumption $T = \mathbf{1}[\zeta(Z) \geq \phi(V)]$. Otherwise stated, the NR counterpart for the Generalized Roy model separability assumption is the monotonicity condition. Each condition enables the identification of causal effects of T on Y in its respective framework. At this level, the IV models in the two frameworks are equivalent.

Model equivalence does not, however, imply that they offer the same analytical capacities. In particular, the Generalized Roy model (4)–(7) explicitly displays the unobserved confounding variable V, while NR does not. This feature enables analysts to further investigate the model and use other approaches for controlling for it. Section 4 shows that the identification of counterfactual outcomes hinges on the analysts's ability to control for the unobserved confounding variable V. Heckman and Vytlacil (2005) use the fact that U is a balancing score for V to define and identify a new parameter called the marginal treatment effect (MTE):

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

The MTE plays a primary role in generating a range of causal effects commonly sought in policy evaluations. A few of these causal parameters are presented in Table 10.

---

[42] See e.g., Rosen (1986), Ekeland et al. (2004).

[43] See, e.g., Hansen and Sargent (1982).

[44] See Imbens and Rubin, 2015.

[45] Angrist and Pischke (2009). Ekeland et al. (2004) show that nonlinearity is intrinsic to hedonic models and that linearizing it produces identification problems.

Table 10
Some causal parameters as weighted average the MTE.
Source: Heckman and Vytlacil (2005).

Causal Parameters	MTE Representation	Weights
$\text{ATE} = E_h(Y(1) - Y(0))$	$= \int_0^1 \text{MTE}(p) W^{\text{ATE}}(p) dp$	$W^{\text{ATE}}(p) = 1$
$\text{TT} = E_h(Y(1) - Y(0) \mid T = 1)$	$= \int_0^1 \text{MTE}(p) W^{\text{TT}}(p) dp$	$W^{\text{TT}}(p) = \dfrac{1 - F_{e,P}(p)}{\int_0^1 \left(1 - F_{e,P}(t)\right) dt}$
$\text{TUT} = E_h(Y(1) - Y(0) \mid T = t_0)$	$= \int_0^1 \Delta^{\text{MTE}}(p) W^{\text{TUT}}(p) dp$	$W^{\text{TUT}}(p) = \dfrac{F_{e,P}(p)}{\int_0^1 \left(1 - F_{e,P}(t)\right) dt}$
$\text{TSLS} = \dfrac{Cov(Y,Z)}{Cov(T,Z)}$	$= \int_0^1 \text{MTE}(p) W^{\text{TSLS}}(p) dp$	$W^{\text{TSLS}}(p) = \dfrac{\int_p^1 \left(t - E_e(P)\right) dF_{e,P}(t)}{\int_0^1 \left(t - E_{e,P(t)}\right)^2 dF_{e,P}(t)}$
$\text{LATE} = \dfrac{E_e(Y\mid Z = z_1) - E_e(Y\mid Z = z_0)}{P_e(z_1) - P_e(z_0)}$	$= \int_{P(z_0)}^{P(z_1)} \text{MTE}(p) W^{\text{LATE}}(p) dp$	$W^{\text{LATE}}(p) = \dfrac{1}{P_e(z_1) - P_e(z_0)}$

The analytical gain generated by switching from the NR framework to a structural equation framework is substantial. The use of structural equations facilitates a richer analysis and a deeper investigation of the properties of the Generalized Roy model. Such analyses cannot be achieved in the NR framework because it does not include unobserved variables, nor does it employ structural equations. This analytical deficiency of the NR framework limits the researcher's ability to extend causal analysis of the Generalized Roy model and other economic models.
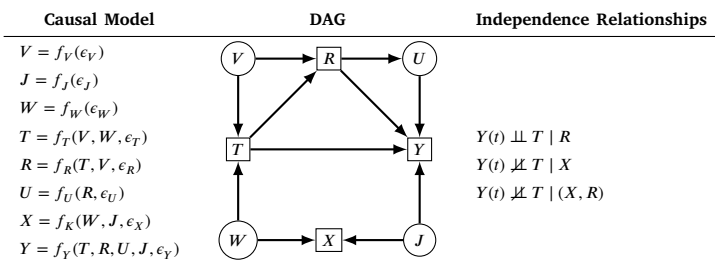
The parsimonious machinery of the NR framework is often misunderstood as endowing the Generalized Roy model with a greater level of generality. This impression is misleading as the IV model featured in the NR framework is equivalent to the Generalized Roy model described by Eqs. (4)–(7) and its monotonicity criteria is equivalent to a separability condition. Its apparent simplicity is due to its lack of explicit statements about its assumptions.

5.2. The matching model in NR

A common identification approach in NR is a *matching* assumption on observed variables. It states that treatment choice T is independent of counterfactual outcomes $Y(t)$ when conditioning on observed pre-treatment variables X, that is, $Y(t) \perp\!\!\!\perp T \mid X$.[46] Intuitively, the assumption states that pre-treatment variables X are sufficiently rich to account for all the unobserved variables that jointly influence treatment choice T and outcome Y. The assumption can be easily criticized as often being overly optimistic for the case of observational studies (Heckman, 2008b; Heckman and Navarro, 2004).

It is natural to assume that increasing the number of pre-treatment variables used for matching decreases bias from unobserved confounders. This notion however is known to be false.[47] Understanding why this assumption is incorrect can be challenging within the NR framework. Conversely, the causal model presented in Table 11 provides a straightforward example that clarifies this misconception.

Table 11
Example of a causal model with non-standard matching conditions.

Causal Model	DAG	Independence Relationships
$V = f_V(\epsilon_V)$		
$J = f_J(\epsilon_J)$		
$W = f_W(\epsilon_W)$		
$T = f_T(V, W, \epsilon_T)$		$Y(t) \perp\!\!\!\perp T \mid R$
$R = f_R(T, V, \epsilon_R)$		$Y(t) \not\perp\!\!\!\perp T \mid X$
$U = f_U(R, \epsilon_U)$		$Y(t) \not\perp\!\!\!\perp T \mid (X, R)$
$X = f_K(W, J, \epsilon_X)$		
$Y = f_Y(T, R, U, J, \epsilon_Y)$		

The causal model Table 11 consists of four observed variables: the treatment T, the outcome Y, a pre-treatment variable X and a post-treatment variable R. The model also contains four unobserved variables V, U, W, J. The causal relationships among the observed and unobserved variables renders $Y(t) \perp\!\!\!\perp T \mid R$ even though $Y(t) \not\perp\!\!\!\perp T \mid X$. The independence relationship that commonly characterizes a matching assumption applies to the post-treatment variable R, but not to the pre-treatment variable X. Moreover, adding the pre-program variable X to the conditioning set of $Y(t) \perp\!\!\!\perp T \mid R$ prevents identification because $Y(t) \not\perp\!\!\!\perp T \mid (X, R)$.

The causal model of Table 11 exemplifies the difficulty of performing causal investigations within the NR framework. The unusual properties of the model stem from the particular causal relationships among its observed and unobserved variables. This model is not

---

[46] In the language of Pearl (2009a), X *d-separates* Y and T.

[47] See, for instance, Pearl (2009c), Heckman and Navarro (2004), Greenland et al. (1999).

easily analyzed within the NR framework which it lacks unobserved variables and suppresses the structural equations that clearly describe the causal relationships among variables.

5.3. Mediation models under NR: An example

Mediation models originate in the literature on path analysis and simultaneous equations.[48] They trace the impacts of interventions on outcomes through their multiple channels of operation. Identifying the causal models generated by NR assumptions is often a daunting task. The economic content of these assumptions is often far from clear. We examine several mediation models to illustrate this point and show the power of the econometric approach compared to an approach based on NR principles. Table 12 uses the econometric approach to present a general mediation model in which a treatment T causes a mediator M and an outcome Y that is caused by both T and M. V denotes a random vector that plays the role of the unobserved confounder causing T, M and Y. The counterfactual mediator when the treatment if fixed at $t \in \text{supp}(T)$ is $M(t) = f_M(t, V, \epsilon_M)$. The counterfactual outcome when the treatment is fixed at t and the mediator is fixed at $m \in \{0, 1\}$ is $Y(t, m) = f_Y(t, m, V, \epsilon_Y)$. The counterfactual outcome when we fix only T at t is $Y(t) = f_Y(t, M(t), V, \epsilon_Y)$.

Table 12
Mediation model with confounding variable.
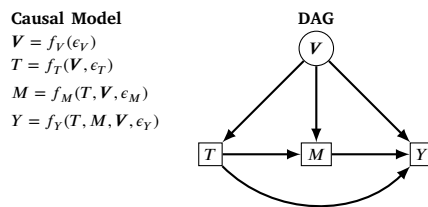
The goal of mediation models is to decompose the total effect of T on Y into an indirect effect that includes the effect of T on M and M on Y and a direct effect not mediated by M. To facilitate the discussion, let T and M denote binary variables taking values in $\{0, 1\}$. The average (total) effect of T on Y is $E_{fix}(Y(1) - Y(0))$.[49] We can also define the average direct effect of T on Y as $E_{fix}(Y(1, M) - Y(0, M)) = \sum_{m=0}^{1} E_{fix}(Y(1, m) - Y(0, m))P_e(M = m)$ and the average indirect effect as $E_{fix}(Y(T, 0) - Y(T, 1)) = \sum_{t=0}^{1} E_{fix}(Y(t, 1) - Y(t, 0))P_e(T = t)$.[50]

Table 13 displays three hypothetical models suitable for examining the total, direct and indirect effects. The first DAG corresponds to the total effect. The hypothetical variable $\widetilde{T}$ replaces the T-input of both the mediator M and the outcome Y equations. The second DAG corresponds to the indirect effect only and the hypothetical variable replaces only the T-input of the mediator equation. The last DAG corresponds to the direct effect only where the hypothetical variable $\widetilde{T}$ replaces only the T-input of outcome equation.

Table 13
Hypothetical models for the mediation model: Total, direct and indirect effects.
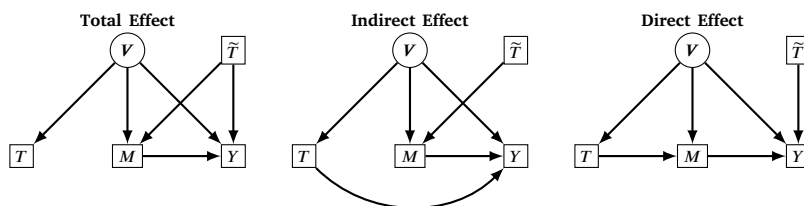
The presence of confounding variable V prevents the identification of the counterfactual means $E_{fix}(M(t))$ and $E_{fix}(Y(t, m))$. A solution to this identification problem using NR is the Sequential Ignorability (SI):[51]

$$\big(Y(t', m), M(t)\big) \perp\!\!\!\perp T, \tag{27}$$

$$Y(t', m) \perp\!\!\!\perp M(t) \mid T, \tag{28}$$

---

[48] See Wright (1921, 1934), Klein and Goldberger (1955), Bollen (1989).

[49] This is the same as $E_h(Y(1) - Y(0))$.

[50] Alternatively, we can then define the direct effect and indirect effects for a given t by (25) and (26) respectively.

$$DE(t) = E_{fix}\big(Y(1, M(t)) - Y(0, M(t))\big) = \int E_{fix}\big(Y(1, m) - Y(0, m)\big) dF_{M(t)}(m) \tag{25}$$

$$IE(t) = E_{fix}\big(Y(t, M(0)) - Y(0, M(1))\big) = \int E_{fix}(Y(t, m)) dF_{M(1)}(m) - \int E_{fix}(Y(t, m)) dF_{M(0)}(m). \tag{26}$$

[51] See Imai et al. (2010, 2011) for the properties of these assumptions. Robins (1986) uses such assumptions in his g-computation algorithm.

for any $t, t' \in \text{supp}(T)$ and $m \in \text{supp}(M)$. SI (27)–(28) enables analysts to identify counterfactual means by statistical conditioning $E_e(M(t)) = E_{fix}(M \mid T = t)$ and $E_{fix}(Y(t, m)) = E_e(Y \mid T = t, M = m)$.

SI assumptions (27)–(28) can be understood as an application of the matching condition to mediation models. Assumption (27) states that the choice T is exogenous with respect to the outcome and mediator counterfactuals. The assumption would be justified if T were randomly assigned by a RCT experiment.

The interpretation of assumption (28) is less straightforward. It states that the counterfactual mediator $M(t)$ is independent of the counterfactual outcome $Y(t, m)$ when conditioned on T. The assumption cannot be directly tested even in randomized experiments that randomize T (Imai et al., 2010). SI assumptions (27)–(28) are much more easily interpreted using structural equations. The assumptions rule out any confounding variable V, generating the model in Table 14.

Table 14
Mediation model with no confounding variables.
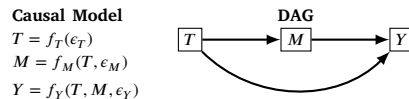
In light of a structural analysis, it can be seen that SI assumptions (27)–(28) are rather strong. They can be weakened if instrumental variables are available, as depicted in Table 15. We use the model to exemplify a case in which NR assumptions are logically possible but generate a causal model that is difficult to justify using any plausible economic argument. The structural model enables the analyst to interpret the statistical assumptions using behavioral theory.

Table 15
Mediation model with instrumental variables.
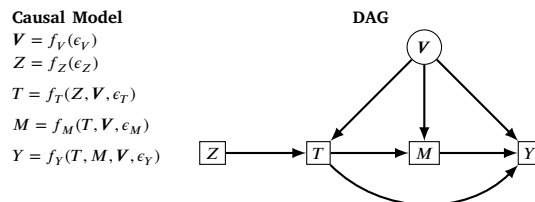
The mediation model with IV has four counterfactuals, $T(z)$, $M(t)$, $Y(t)$, $Y(t, m)$ previously defined. In the language of NR, the model would be characterized by IV exogeneity condition $Z \perp\!\!\!\perp (T(z), M(t), Y(t), Y(t, m))$. The condition holds due to the independence of Z and V.[52] Suppressing Y generates an IV model where M plays the role of the outcome.

To dig more deeply, we investigate the case of a binary instrument $Z \in \{0, 1\}$. The response vector $S_i = [T_i(0), T_i(1)]'$ denotes the vector of treatment choices that agent i would take if it were assigned to each of the instrumental values. Section 4 shows that, given S, the treatment choice T depends only on the instrument Z. The exogeneity condition states that Z is independent of the counterfactual outcome $Y(t)$. Thus

$$T \perp\!\!\!\perp Y(t) \mid S. \tag{29}$$

S is a balancing score for V.

Yamamoto (2014) uses the language of NR to identify mediation effects using instrumental variables. His solution merges SI (27)–(28) with the matching property of the response vector S in (29). He advocates an assumption that he terms the *local average causal mediation effects (LACME) assumption*:

$$(Y(t, m), M(t')) \perp\!\!\!\perp T \mid (S = [0, 1]'), \tag{30}$$

$$Y(t, m) \perp\!\!\!\perp M(t') \mid (T, S = [0, 1]'). \tag{31}$$

LACME (30)–(31) adds the response vector S as an additional conditioning variable to the SI independence relationships in (27)–(28). Assumption (30) is an extension of the matching property of S from the IV model of Table 14 to the mediator model of Table 15. Under monotonicity (24), the LACME assumption identifies the direct and indirect mediation effects for compliers.

It is easy to interpret LACME in terms of NR assumptions: assumptions (30)–(31) are a weaker version of SI (27)–(28) that incorporates the LATE analysis of Imbens and Angrist (1994). On the other hand, it is difficult to gauge how the LACME assumptions fit into the mediation model of Table 12. It is even harder to interpret the causal content of these assumptions.

Table 16 presents two DAGs that use the structural approach to clarify the causal content of LACME. The first DAG in Table 16 places the unobserved response vector S into the mediation model of Table 12. $S = [T(0), T(1)]'$ is expressed as a function of the confounding variable V because $T(z)$ is a function of V. Choice T is a function of S and Z since $T = \big[\mathbf{1}[Z = 0], \mathbf{1}[Z = 1] \big] S$. Thus, the

---

[52] Note that if we were to suppress M from the DAG of Table 15, we would obtain the empirical model of Table 4.

response vector S plays the role of a balancing score for V for choice T only. The resulting DAG does not include more information than the original model of Table 12 because S is unobserved.

Table 16
Mediation model including S and the mediation model under LACME assumption.
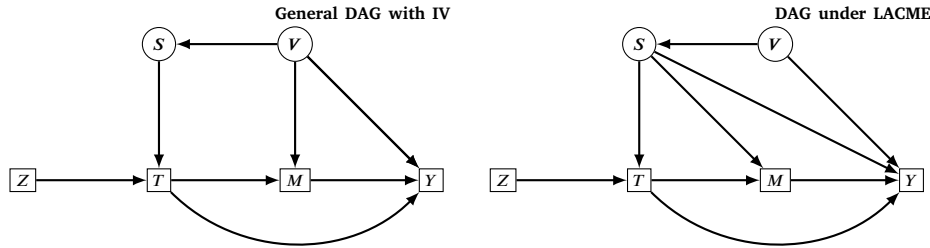
The second DAG in Table 16 displays the mediation model under LACME. According to assumption (31), the response vector S plays the role of a balancing score for T and M. In addition, LACME prevents V from jointly causing M, Y and implies that S directly causes M, Y.

LACME's foundation is rooted in statistical rather than causal considerations. Consequently, it is difficult to translate LACME into economic models with credible causal interpretations. Its underlying assumptions are devoid of economic substance, a trait commonly found in analyses within the NR framework.

5.3.1. Using structural equations to identify the mediation model with IV

Dippel et al. (2020) study the identification of causal effects for the mediation model with an instrumental variable. Their analysis illustrates the gain in clarity and interpretability when a causal model is expressed by structural equations instead of NR statistical independence relationships.

A typical empirical setting of an IV model consists of one instrument and various outcomes. A mediation model with an instrument arises when the treatment causes an intermediate outcome (the mediator), which in turn causes a final outcome. The DAG of this empirical model is presented in the first column of Table 17.

Table 17
Dissecting the mediation model.
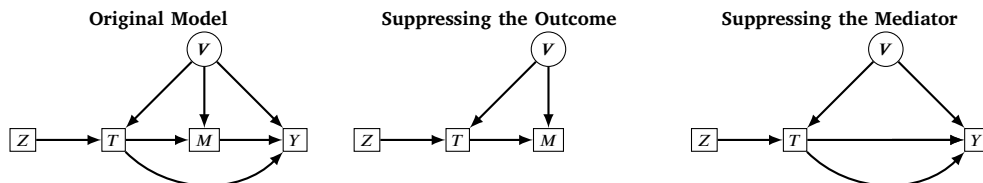
The second column of Table 17 presents the DAG generated by suppressing the final outcome. The resulting DAG depicts the IV model examined in Section 3 where the mediation M plays the role of an outcome.

The second column of Table 17 displays the DAG that results from supressing the final outcome. This DAG represents the IV model discussed in Section 3, where mediation M plays the role of an outcome. The causal effect of T on M can be identified by the methods discussed in Section 4.

The third column of Table 17 suppresses the mediator M. The resulting model is also an IV model. This means that the *total effect* of T on Y can also be identified by the methods of Section 4. Unfortunately, the IV does *not* identify the causal effect of M on Y. Consequently, mediation analysis cannot be conducted without further assumptions.

Dippel et al. (2020) address the question of whether it is possible to use an instrumental variable Z to nonparametrically identify the causal chain connecting T, M, Y while maintaining the endogeneity of the treatment T with respect to the mediator M and outcome Y. They show that the only solution to this problem is to assume the *partially confounded* mediation model of Table 18.

The partially confounded model disientangle the confounding variable V into variables V_T and V_Y, which are independent of each other ($V_T \perp\!\!\!\perp V_Y$). Here, V_T influences both T and M, whereas V_Y affects both M and Y. These unobserved variables can be understood as distinct components that are typically analyzed in variance models.

The partially confounding assumption generates an additional exogeneity condition $(M(z), Y(m,t)) \perp\!\!\!\perp Z \mid T$ while maintaining the endogeneity of the treatment T with respect to M and Y. This means that Z becomes a valid instrument for identifying the causal effect of M on Y when conditioning on the treatment variable T. If the assumption holds, the causal effect of M on T can be evaluated by the methods of Section 4. Essentially, the use of a causal model clarifies the mechanism generating the additional exogeneity condition. This clarification allows Dippel et al. (2020) to delve into the intuition, feasibility, and estimation strategies for the partially confounded mediation model. They provide a variety of examples to demonstrate situations in which the partial confounding assumption is likely to be applicable, as well as scenarios where it may not hold.

Table 18
Partially confounded model with instrumental variables.

Causal Model

$$V_T = f_{V_T}(\epsilon_{V_T})$$
$$V_Y = f_{V_Y}(\epsilon_{V_Y})$$
$$Z = f_Z(\epsilon_Z)$$
$$T = f_T(Z, V_T, \epsilon_T)$$
$$M = f_M(T, V_T, V_Y, \epsilon_M)$$
$$Y = f_Y(T, M, V_Y, \epsilon_Y)$$
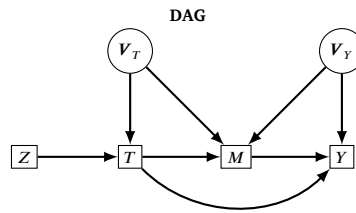
6. The do-calculus and the hypothetical model

This section compares the *do*-calculus framework (DoC) of Pearl (2009a) with the Neyman–Rubin (NR) framework of Holland (1986) and Imbens and Rubin (2015), as well as the Hypothetical Model (HM) of Heckman and Pinto (2015).

The DoC was first presented in Pearl (1995). The method employs graph theory-based algorithms to identify the probability distribution of counterfactual variables in causal models represented by DAGs.[53] In contrast with NR, DoC uses autonomous (invariant) structural equations. The method clearly describes causal relationships among model variables. Its fundamental relationships are based on thought experiments. It is not incompletely formulated in a way that leads to problematic causal interpretations, as in the NR approach.

DoC applies to any nonparametric, recursive system of structural equations. Similar to HM, DoC allows for unobserved variables. It can be applied to multiple equation causal models and a range of causal inquiries.

However, HM and the DoC differ greatly regarding counterfactual manipulations. To address the causal operation of fixing, the HM solution is based on a hypothetical model that formalizes thought experiments and places them on a sound probabilistic footing. Contrary to HM, DoC defines hypothetical models by making manipulations *within* the empirical model. It does not have a counterpart to $\widetilde{T}$, the source of hypothetical variation in HM. DoC implements the notion of setting or fixing using a new set of rules that combine graphical analysis, independence relationships and probability equalities.

For instance, the DoC uses a DAG-based criteria called *d*-separation to check for conditional independence among variables. Its definition requires some DAG terminology. 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 blocks all paths from T to Y.

6.1. The rules of DoC

As noted in Section 3.5, DoC uses the "back-door" criterion to verify matching conditions. In DoC terminology, the matching condition $Y(t) \perp\!\!\!\perp T | V$ of the Generalized Roy Model in Table 3 is expressed by the statement: "V *d*-separates Y and T in the DAG $G_{\underline{T}}$," where G be the original DAG of the Roy Model and $G_{\underline{T}}$ is a derived DAG which suppresses the arrows departing from T. The "back-door" criterion holds for confounder V of the Roy Model. This implies the matching condition in which controlling for V renders the counterfactual outcome $Y(t)$ statistically independent of treatment T.

The core machinery of the DoC consists of three DAG-based rules. Additional notation is necessary to describe these rules. Let Y, K, Z, T denote disjoint variable sets in $\mathcal{T}$. In DoC notation, $T(Z)$ denotes the variables in T that do not directly or indirectly cause Z. "Do" deletes certain links in the original graph and assumes certain conditional independence relations. This is Pearl's way to fix variables externally. DoC uses $G_{\overline{K}}$ for the derived DAG that deletes all causal arrows *arriving* at K in the original DAG G. $G_{\underline{T}}$ denotes the DAG that deletes all causal arrows emerging from T. In this notation, $G_{\overline{K},\underline{T}}$ stands for the derived DAG that suppresses all arrows *arriving* at K and emerging from T, while $G_{\overline{K,T(X)}}$ deletes all arrows arriving at K in addition to arrows arriving at $T(X)$, variables in T that are not ancestors of X.

The DoC rules combine a graphical condition with a conditional independence relation that, when satisfied, imply a probability equality. These rules are detailed in Table 19. Identifying a causal effect involves repeatedly applying these rules. We illustrate several examples of how to utilize these rules for causal analysis.

In computer science, DoC is said to be "complete". This is different from the notion of completeness as defined in simultaneous equations theory discussed below in Section 7. The key DoC notion is that if a causal effect is identifiable, it can be identified by the iterative application of some sequence of the three rules (Huang and Valtorta, 2006; Shpitser and Pearl, 2006).

A major limitation of do-calculus is that it only applies to non-parametric models that can be fully characterized by a DAG. Stated otherwise, the method does not account for assumptions about the functional forms of the structural equations or covariance restrictions. This limitation hinders the application of most of the popular econometric tools used in empirical economics, such as cross-equation restrictions, separability, additivity or monotonicity assumptions. For instance, the Generalized Roy model is not

---

[53] For a recent book on the graphical approach to causality, see Peters et al. (2017). For related works on causal discovery, see Spirtes et al. (2001), Heckman and Pinto (2015), Hoyer et al. (2009), and Lopez-Paz et al. (2017).

identified by DoC because it requires assumptions such as separability. The same is true of the IV model. Separability cannot be characterized by conditional independence assumptions generated by a DAG. We now demonstrate these points.

Table 19
The three DoC rules.

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

6.2. Using Do-Calculus to investigate the Roy model

We show the limitations of the DoC for identifying the Roy model. Table 20 is instrumental in our analysis. The first column of Table 20 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 in DoC, we obtain $P(T \mid do(Z)) = P(T \mid Z)$. Summarizing:

$$G_{\underline{Z}} \Rightarrow T \perp\!\!\!\perp Z \Rightarrow \text{ by Rule 2 that } P\big(T \mid do(Z)\big) = P(T \mid Z). \tag{32}$$

Table 20
Using do-calculus to investigate the Roy Model.
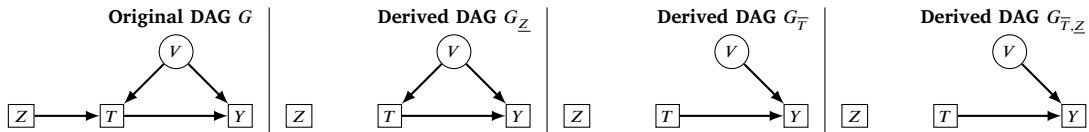
Therefore, the modified Directed Acyclic Graph (DAG), $G_{\underline{Z}}$, enables us to assert that conditioning T on $Z = z$ is equivalent to choice T when we fix Z to value z. In the NR framework, this result is obtained by the exogeneity condition $T(z) \perp\!\!\!\perp Z$, which states that the instrument Z is independent of the counterfactual choice $T(z)$ and thus $P(T \mid Z = z) = P(T(z))$ holds. Instrument Z in DAG $G_{\underline{Z}}$ is independent of both T and Y. This analysis also applies to Y. We can use (32) to obtain that $P(Y \mid do(Z)) = P(Y \mid Z)$, which means that conditioning on Z is equivalent to fixing Z. In summary, instrument Z is an external variable and the causal operation of fixing is translated to standard statistical conditioning.

The third column of Table 20 displays the DAG $G_{\overline{T}}$ which suppresses the arrow arriving at T. The LMC of Z in $G_{\overline{T}}$ implies $Z \perp\!\!\!\perp Y$. By Rule 1 of DoC, we have that $P(Y \mid do(T), Z) = P(Y \mid do(T))$. Summarizing:

$$G_{\overline{Z}} \Rightarrow Y \perp\!\!\!\perp Z \Rightarrow \text{ by Rule 1 that } P(Y \mid do(T), Z) = P\big(Y \mid do(T)\big). \tag{33}$$

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 $Y \perp\!\!\!\perp Z \mid \widetilde{T}$ of the HM framework.

The last column of Table 20 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 is $Z \perp\!\!\!\perp (T, Y, V)$ which implies that $Y \perp\!\!\!\perp Z \mid T$ holds. Using Rule 2 of the DoC we obtain:

$$G_{\overline{T}, \underline{Z}} \Rightarrow Y \perp\!\!\!\perp Z \mid T \Rightarrow \text{ by Rule 2 that } P(Y \mid do(T), do(Z)) = P(Y \mid do(T), Z). \tag{34}$$

Combining $P(Y \mid do(T), Z) = P(Y \mid do(T))$ in (33) with $P(Y \mid do(T), do(Z)) = P(Y \mid do(T), Z)$ in (34) we obtain $P(Y \mid do(T), do(Z)) = P(Y \mid do(T))$. 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 that can be performed using DoC. The method describes some key properties of the Roy model, but the application of its rules alone cannot deliver identification of treatment effects. Indeed, the assumptions necessary for securing the identification of treatment effects in the Roy model cannot be assessed by a DAG representation alone. Identifying assumptions, such as separability or monotonicity, imposes restrictions on the functional form of the choice equation which goes beyond the causal links described by a DAG.

6.3. The Front-door model

To make a more positive statement about do-calculus, it is useful to compare the identification machinery of the DoC and HM using a causal model when treatment effects *are* identified by DoC. We use the "Front-door model" of Pearl (2009a) to illustrate the differences in the approaches.

The Front-door model 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. Table 21 presents the causal representations of the model.

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 directly confounded by V since 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 21
Representations of the Front-door Model.

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

Table 22
Using do-calculus to identify the causal effect of T on Y in the Front-door model.
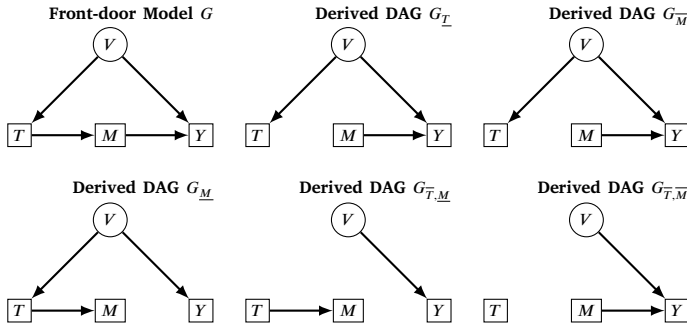
We illustrate how to use DoC to identify the distribution of the counterfactual outcome $Y(t)$. To simplify notation, suppose that all variables are discrete. The do-calculus is cumbersome. The method requires the five derived DAGs displayed in Table 22.

Identification for the counterfactual outcome is obtained by the following sequence of steps:

1. $T \perp\!\!\!\perp M$ in $G_{\underline{T}}$ holds, thus by Rule 2 we have that $P(M \mid do(T)) = P(M \mid T)$.

2. $M \perp\!\!\!\perp T$ in $G_{\overline{M}}$ holds, thus by Rule 3 we have that $P(T \mid do(M)) = P(T)$.

3. $M \perp\!\!\!\perp Y \mid T$ in $G_{\underline{M}}$ holds, thus by Rule 2 we have that $P(Y \mid T, do(M)) = P(Y \mid T, M)$.

4. Adding these results, we obtain:

$$\therefore P(Y \mid do(M)) = \sum_t P(Y \mid T = t, do(M)) P(T = t \mid do(M))$$

by Law of Iterated Expectations (L.I.E.)

$$= \sum_t P(Y \mid T = t, M) P(T = t) \text{ by steps 1,2, and 3.}$$

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

6. $Y \perp\!\!\!\perp T \mid M$ in $G_{\overline{T}, \overline{M}}$ holds, thus by Rule 3, $P(Y \mid do(T), do(M)) = P(Y \mid do(M))$.

7. Collecting these results, we have that $P(Y \mid Z, do(T)) = P(Y \mid do(Z), do(T)) = P(Y \mid do(M))$.

8. Finally, we can use previous results to obtain the following equation:

$$\therefore P(Y \mid do(T) = t) = \sum_m P(Y \mid M = m, do(T) = t) P(M = m \mid do(T) = t) \text{ by L.I.E.}$$

$$= \sum_m P(Y \mid do(M) = m, do(T) = t) P(M = m \mid do(T) = t) \text{ by step 5.}$$

$$= \sum_m P(Y \mid do(M) = m) P(M = m \mid do(T) = t) \text{ by step 7.}$$

$$= \sum_m \left(\sum_{T=t'} P(Y \mid T=t', M=m) P(T=t') \right) P(M=m \mid T=t) \text{ by step 4.}$$

6.4. The Front door model in the hypothetical model framework

We now investigate the same front-door model using the hypothetical framework. Table 23 displays the empirical and the hypothetical Front-door models as DAGs. The bottom panel of the table presents the LMC for both models:

Table 23
The empirical and hypothetical Front-door models.

LMC	LMC
$V \perp\!\!\!\perp M$	$V \perp\!\!\!\perp (M, \widetilde{T})$
$T \perp\!\!\!\perp M \mid V$	$T \perp\!\!\!\perp (M, Y, \widetilde{T}) \mid V$
$M \perp\!\!\!\perp V \mid T$	$M \perp\!\!\!\perp (T, V) \mid \widetilde{T}$
$Y \perp\!\!\!\perp T \mid (V, M)$	$Y \perp\!\!\!\perp (T, \widetilde{T}) \mid (V, M)$
	$\widetilde{T} \perp\!\!\!\perp (T, V)$

Recall that the counterfactual outcome in the hypothetical framework is denoted by the outcome Y conditioned on the hypothetical variable $\widetilde{T}$. Identification consists of expressing the counterfactual outcome distribution $P_h(Y \mid \widetilde{T} = t)$, which is defined in the hypothetical model, in terms of the observed distribution $P_e(T, M, Y)$, defined in the empirical model. The connection between the probabilities of the hypothetical and empirical models is governed by the rules (16)–(17). The first rule states that, if $Y \perp\!\!\!\perp \widetilde{T} \mid (T, W)$ holds for any variables $Y, \widetilde{T}, T, W$ in the hypothetical model, then we can equate $P_h(Y \mid \widetilde{T} = t, T = t', W) = P_e(Y \mid T = t', W)$. On the other hand, $Y \perp\!\!\!\perp T \mid (\widetilde{T}, W)$ implies that $P_h\left(Y \mid \widetilde{T} = t, T = t', W\right) = P_e(Y \mid T = t, W)$. Thus, we seek to find independence relationships of the hypothetical model that contain T and $\widetilde{T}$. The useful relations are $Y \perp\!\!\!\perp \widetilde{T} \mid (M, T)$ and $M \perp\!\!\!\perp T \mid \widetilde{T}$.[54] It is also the case $T \perp\!\!\!\perp \widetilde{T}$ holds as $\widetilde{T}$ is externally specified (exogenous) and does not cause T. We can then apply rules (16)–(17) to generate the following probability equalities:

$$Y \perp\!\!\!\perp \widetilde{T} \mid (T, M) \Rightarrow \quad P_h\left(Y \mid \widetilde{T}, T = t', M\right) = P_e\left(Y \mid T = t', M\right). \tag{35}$$

$$M \perp\!\!\!\perp T \mid \widetilde{T} \quad \Rightarrow \quad P_h\left(M \mid \widetilde{T} = t, T\right) = P_e\left(M \mid T = t\right). \tag{36}$$

$$T \perp\!\!\!\perp \widetilde{T} \mid T \quad \Rightarrow \quad P_h\left(T = t' \mid \widetilde{T}\right) = P_e\left(T = t'\right). \tag{37}$$

The causal effect of T on Y of the Front-door model is identified through the following logic:

$$P_h\left(Y \mid \widetilde{T} = t\right) = \sum_{t', m} P_h\left(Y \mid m, T = t', \widetilde{T} = t\right) P_h\left(m \mid T = t', \widetilde{T} = t\right) P_h\left(T = t' \mid \widetilde{T} = t\right). \tag{38}$$

$$= \sum_{t', m} P_e\left(Y \mid m, T = t'\right) P_e\left(m \mid T = t\right) P_e\left(T = t'\right). \tag{39}$$
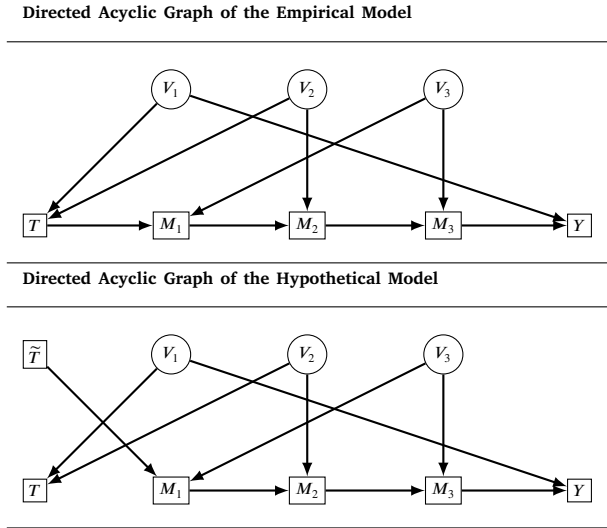
Eq. (38) is a sum of probabilities defined in the hypothetical model by application of the law of iterated expectation over T and M. Eq. (39) replaces each of the hypothetical model probabilities with empirical model probabilities listed in Eqs. (35)–(37).

The identification of the counterfactual outcomes in the Front-door Model stems from the three independence relationships in (35)–(37). These independence relationships illustrate two properties that are at the core of the identification result. The first property is that the independence relationships alternate between T and $\widetilde{T}$ in the positions of conditioning variable and independent variable on the right-hand side. We term this property as *alternate conditionals*. The second property is that the sequence of conditioning variables on the right-hand side of (35)–(37) form a sequence $Y \rightarrow M \rightarrow T$ that starts at the targeted outcome Y and arrives at the treatment using the variable M to bridge these variables. We term this use of M as the *bridging* property.

Identification is secured whenever the properties of alternating conditionals and the bridging properties hold. We illustrate these ideas for the complex mediation model in Table 24. The model has three observed mediating variables M_1, M_2, M_3 (instead of M) and three unobserved, confounding variables V_1, V_2, V_3 (instead of V).

---

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

Table 24
Using 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:

$$Y \perp\!\!\!\perp \widetilde{T} \mid (T, M_3, M_2, M_1) \quad \Rightarrow P_h\big(Y \mid \widetilde{T}, T = t', M_3, M_2, M_1\big) = P_e\big(Y \mid T = t', M_3, M_2, M_1\big). \tag{40}$$

$$M_3 \perp\!\!\!\perp T \mid (\widetilde{T}, M_2, M_1) \quad \Rightarrow P_h\big(M_3 \mid \widetilde{T} = t, T, M_2, M_1\big) = P_e\big(M_3 \mid T = t, M_2, M_1\big). \tag{41}$$

$$M_2 \perp\!\!\!\perp \widetilde{T} \mid (T, M_1) \quad \Rightarrow P_h\big(M_2 \mid \widetilde{T}, T = t', M_1)\big) = P_e\big(M_2 \mid T = t', M_1)\big). \tag{42}$$

$$M_1 \perp\!\!\!\perp T \mid \widetilde{T} \quad \Rightarrow P_h\big(M_1 \mid \widetilde{T} = t, T\big) = P_e\big(M_1 \mid T = t\big). \tag{43}$$

$$T \perp\!\!\!\perp \widetilde{T} \mid T \quad \Rightarrow P_h\big(T = t' \mid \widetilde{T}\big) = P_e\big(T = t'\big). \tag{44}$$

The set of independence relationships (40)–(44) exhibits the alternate conditionals property. 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 provides a bridging sequence: $Y \rightarrowtail M_3 \rightarrowtail M_2 \rightarrowtail M_1 \rightarrowtail T$. The law of iterated expectations and independence relationships (40)–(44) enable us to express the counterfactual probability $P_h(Y \mid \widetilde{T})$ as:

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

Connection rules (16)–(17) enable us to translate hypothetical probabilities into the empirical probabilities as listed in (40)–(44). The resulting identification equation is presented below. It displays the alternative pattern of values t and t' in the same fashion as the identification equation of the Front-door model:

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

6.5. Comparing DoC and HM frameworks

Both DoC and HM employ structural equations and describe causal models with both observed and unobserved variables. They clearly separate the task of defining counterfactuals and identifying them. Both frameworks enable analysts to disentangle the tasks of causal analysis in Table 1. Both frameworks employ economic theory to define causal models (Task 1) and the structural equations that underlie the approach.

There are, however, some distinct practices in DoC and HM. When DoC fixes a treatment variable, it *eliminates* the equation for T in constructing the joint distribution of variables. All of the DoC analyses are done within the empirical model so generated.

HM does *not* eliminate the equation for the treatment variable. Instead, it *adds* a hypothetical variable. The presence of both treatment and hypothetical variables in the HM framework facilitates the study of the causal effects. HM can be used to readily analyze both external manipulation and conditioning, such as the treatment on the treated, whereas this is outside the scope of DoC. It facilitates the examination of causal inference for direct and indirect effects in which the hypothetical variable replaces some but not all the treatment inputs of the structural equations. DoC invents new rules to undertake those tasks for each combination of conditioning variables.

The identification of causal effects (Task 2) requires connecting the hypothetical model with the empirical model. HM uses two statistical implications to connect the probability distributions of the hypothetical and empirical models. HM analyses remain within the realm of standard statistical theory and do not require the invocation of non-probabilistic DAG-based rules.

The DoC machinery consists of three DAG-based rules. It constructs a series of possible DAGs. Each of them constitutes a causal model that modifies the empirical model. Each modification of the empirical model corresponds to introducing a new set of conditional independence relationships. The search for the combinations of DAGs and conditional independence relationships that are required to identify counterfactuals grows exponentially. An algorithm has been developed to perform this task.[55] Calculations with HM are simpler than those based on DoC. They rely on a single modification of the original DAG, as encoded in the hypothetical model, instead of a growing list of DAGs to implement the three guiding rules of DoC.

DoC relies critically on DAGs, conditional independence relationships, and a special set of rules. The HM machinery remains within the statistical realm to make statistics converse with causality. In doing so, the method is capable of accommodating assumptions that explore the rich variety of functional form restrictions, distributional assumptions, and cross-equation and cross-variable relationships that lie outside the scope of DoC.

7. Simultaneous causality

The Generalized Roy model is usually expressed as a recursive model.[56] However, simultaneous causality is a property of many economic models. Examples of such models include those for social interactions, general equilibrium, Walrasian market clearing, or simultaneous play in models of industrial organization, which are staples of economic theory (see, e.g., Mas-Colell et al., 1995; Tamer, 2003). Such models are ignored in discussions of causality in the NR literature. The NR approach invokes the Stable Unit Treatment Value Assumption (SUTVA), which excludes the possibility of interactions among agents.[57] Such interactions are usually termed "confounders" and are treated as a problem rather than a source of information about economic and social behavior.

It is instructive to consider these models because they challenge the approaches used in the statistical literature but are easily analyzed by rigorous econometric causal models. The pioneering econometric models studied by the Cowles Commission featured simultaneity.[58] Haavelmo's (1943) paper explicitly analyzed causality in a simultaneous system. Many of the core ideas in simultaneous equations models are ignored or remain unknown to the followers of the statistical approaches, which rely on recursive formulations, which are considered to be essential features of causal models.

Simultaneous causality is an essential feature of many structural equation econometric models.[59] The LISREL model of Jöreskog (1973) allows for simultaneity, measurement error and latent variables proxied by measurements as discussed in Section 4.

The structural systems typically consist of two parts: (a) an autonomous structural system expressed in terms of latent variables (Bollen, 2002) and (b) a measurement system. The measurement system proxies the latent variables using observed measurements. The first part of the structural system consists of structure for person i:

$$\boldsymbol{B}\boldsymbol{\eta}_i = \boldsymbol{\alpha}_\eta + \boldsymbol{\Gamma}\boldsymbol{\chi}_i + \boldsymbol{\omega}_i \tag{45}$$

where $\boldsymbol{\omega}_i$, $\boldsymbol{\eta}_i$, $\boldsymbol{\chi}_i$ are vectors of latent variables. The measurement system consists of vectors of measurements:

$$\text{Measurement:} \begin{cases} \boldsymbol{y}_i = \boldsymbol{\alpha}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \ (\text{measurement for } \boldsymbol{\eta}_i) \\ \boldsymbol{x}_i = \boldsymbol{\alpha}_x + \boldsymbol{Y}_\chi + \boldsymbol{\xi}_i \ (\text{measurement for } \boldsymbol{\chi}_i) \end{cases}$$

where ε_i, and ξ_i are vectors of latent variables. These models have been extended to time series and panel data settings (see e.g. Goldberger and Duncan, 1973; Bollen, 1989; Hansen and Sargent, 1982).

Bollen and Pearl (2013) exposit this system of equations as a causal model with simultaneity and show how various measurement systems use factor models and other approaches to proxy the latent variables which may be the variables measured with error or omitted variables, like ability in an earnings equation, or technical efficiency in a production function. They dispel many misguided criticisms of the structural approach lodged by advocates of the NR approach. These systems are equipped to use cross-equation

---

[55] See Pearl (2009a).
[56] See, however, Heckman (1978), Brock and Durlauf (2007).
[57] See, for instance, Imbens and Rubin (2015), Holland (1986).
[58] See, e.g., Koopmans et al. (1950).
[59] See Goldberger (1972), Haavelmo (1944, 1943), Koopmans et al. (1950), Goldberger and Duncan (1973).

restrictions and covariance restrictions to secure the identification of causal parameters. Hansen and Sargent (1982) is an example of this approach applied to time series models.

This literature is rich and we lack the space to exposit it thoroughly. We note that as previously discussed linear equation versions of these models provide a framework for proxying V. It is also an approach for studying mediation where analysts can study how interventions on χ_i percolate through equation system (45). Schennach (2020) summarizes a large literature on nonparametric factor and proxy models.

Instead of a general exposition of these systems, we refer the reader to Bollen and Pearl (2013) and consider a simple two-equation simultaneous equations model of the sort exposited by Haavelmo (1943). We consider a system of two autonomous (structurally-invariant) causal equations:

$$Y_1 = g_{Y_1}(Y_2, X_1, U_1, \epsilon_1) \tag{46}$$

$$Y_2 = g_{Y_2}(Y_1, X_2, U_2, \epsilon_2) \quad U_1 \not\!\perp U_2. \tag{47}$$

We use this system to demonstrate how causality can be analyzed in simultaneous systems. Again, ϵ_1 and ϵ_2 are mutually independent error terms that are also independent of U_1, U_2, X_1, and X_2. To conserve on notation, we keep ϵ_1 and ϵ_2 implicit in what follows.

This system of equations represents two maps: $g_{Y_1}: (Y_2, X_1, U_1) \to Y_1; g_{Y_2}: (Y_1, X_1, U_2) \to Y_2$. Y_1 and Y_2 could be actions of a pair of interacting agents.[60] To simplify the discussion, we assume that both equations are twice continuously differentiable. This is a convenience and not a necessity. The model of Eqs. (46)–(47) is treated in a special way in the DoC approach. Models with multiple simultaneous equations are standard in the literature (see, e.g., Koopmans et al., 1950; Theil, 1958, 1971; Fisher, 1966; Goldberger and Duncan, 1973; Bollen, 1989).

As previously noted, Eqs. (46) and (47) are assumed to be structural, i.e., invariant under manipulations of their arguments, so they are stable, autonomous maps. Policies consist of manipulations of their arguments. Autonomy is one part of the *SUTVA* assumption in the NR model.[61]

In the classical model of market clearing equilibrium, Y_1 is price; Y_2 is quantity and X_1, X_2, U_1, and U_2 are causal determinants. Eqs. (46) and (47) are generated by thought experiments varying the arguments and tracing out the outcomes. Thus, Y_1 in (46) could be the market price that is consistent with hypothetical values Y_2, X_1, U_1. (47) is the analogous relationship for market quantity Y_2. The addition of unobserved (by the economist) variables U_1 and U_2 is made in anticipation of empirical applications. In the peer effects literature, Y_1 and Y_2 are behaviors of two interacting agents (e.g., smoking or drug use).

In terms of our previous notation, the variable set for the empirical model is $\mathcal{T}_e = \{Y_1, Y_2, X_1, X_2, U_1, U_2\}$. $\mathbb{M}_e(Y_1) = \{Y_2, X_1, U_1\}$ and $\mathbb{M}_e(Y_2) = \{Y_1, X_2, U_2\}$. The empirical and hypothetical models are displayed as DAGs in Table 25 given by[62]:

Table 25
Empirical and hypothetical causal models.
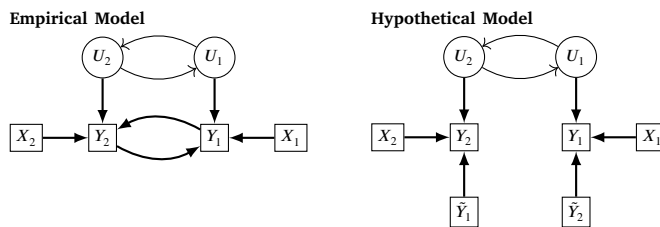
The LMC condition does not apply, so that the Bayesian net approach is not useful in this context. "Fixing" and the hypothetical model approach readily extend to a system of simultaneous equations for Y_1 and Y_2, whereas the fundamentally recursive methods based on DAGs require special treatment.

7.1. Completeness

"Completeness" assumes the existence of at least a local solution for Y_1 and Y_2 in terms of (X_1, X_2, U_1, U_2):

$$Y_1 = \phi_1(X_1, X_2, U_1, U_2) \tag{48}$$

$$Y_2 = \phi_2(X_1, X_2, U_1, U_2). \tag{49}$$

These are *reduced form* equations (see, e.g., Koopmans et al., 1950; Matzkin, 2008, 2013). They inherit the autonomy properties of the structural equations. Completeness is a property that guarantees the conceptual possibility of simultaneity, which is not necessarily guaranteed. If it fails, the existence of consistent solutions to (46) and (47) is not guaranteed. Nonetheless, autonomous correspondences may still exist, and they can be used to make set-valued causal inferences.[63]

---

[60] In the literature on peer effects, simultaneous equation problems are relabeled as "reflection problems". See Moffitt (2001), Manski (1993).

[61] The other part is the absence of simultaneity or general equilibrium effects. See Heckman (2008a).

[62] U_2 and U_1 are reciprocally related.

[63] See, e.g., Heckman (1978), Tamer (2003), Mas-Colell et al. (1995), Quandt (1988).

The causal response of Y_2 on Y_1 when Y_2 is fixed at y_2 is generated by

$$Y_1(y_2) = g_{Y_1}(y_2, X, U_1).$$

Symmetrically, the causal response of Y_1 on Y_2 when Y_1 is fixed at y_1 is generated by:

$$Y_2(y_1) = g_{Y_2}(y_1, X, U_2).$$

The relationships (46) and (47) can be defined even if they might not be identified or estimated. The *completeness assumption* says that there are values of X_1, X_2, U_1, U_2 that generate values of Y_1, Y_2 consistent with (46) and (47). These involve hypothetical variations. For certain models, no such sets of variables may exist, and the models are termed incomplete.

7.2. Can we hypothetically vary Y_2 and Y_1?

If Y_2 and Y_1 are simultaneously determined, the notion of varying Y_2 to change Y_1 may seem impossible. Pearl (2009b) maintains his focus on recursive models and addresses this problem in a very special way by assuming structural invariance and "shutting one equation down", assuming the rest of the system remains unchanged. Thus, for example, Eq. (47) is suspended, but (46) is maintained. This is consistent with the logic of do-calculus, which eliminates relationships from systems, assuming invariance of the remaining system. He sets Y_2 to a constant that can be manipulated in (46). This thought experiment converts a simultaneous system into a recursive system with all other equations assumed to hold as before. After Y_2 is fixed, the do-calculus can be applied.

This approach is cumbersome and strains credibility in many economic contexts (e.g.,in a two person world, person 1 influences 2, but not vise versa), but it is logically possible.[64] It is unnecessary if exclusions in (46) and (47) are used. To show this, we define exclusion of X_2 in (46) as $\frac{\partial g_{Y_1}}{\partial X_2} = 0$ for all values of Y_2, X_1, X_2, U_1.[65] Exclusion of X_1 in (47) is defined as $\frac{\partial g_{Y_2}}{\partial X_1} = 0$ for all values of Y_1, X_1, X_2, U_2. Implicit is the assumption that components of X_1 and X_2 can be varied. Under completeness and exclusion X_2 from (47), by the chain rule, the causal effect of Y_2 on Y_1 is

$$\frac{\partial g_{Y_1}}{\partial Y_2} = \frac{\partial Y_1}{\partial X_2} \Big/ \frac{\partial Y_2}{\partial X_2} = \frac{\partial \phi_1}{\partial X_2} \Big/ \frac{\partial \phi_2}{\partial X_2}.$$

We may define and identify the causal effects for Y_1 on Y_2 in an analogous fashion. Variations in X_1 and X_2 that respect completeness define the causal parameters when the components of X_1 and X_2 can be independently varied.[66] No implausible "shutting down" of any equation in a system while assuming autonomy (structural invariance) of the remaining system is required.

This logic is now standard and is the basis for an estimation technique, "indirect least squares" (see Theil, 1958 and Tinbergen, 1930, 1939). It demonstrates the flexibility of the econometric approach for defining and identifying causal parameters outside the narrow world of DAGs. Fisher (1966) gives a range of approaches for identifying systems like (46) and (47) and more general versions using restrictions within and across equations for observables and unobservables.

7.3. Econometric mediation analysis

We have already discussed mediation analyses in recursive models. These notions extend to models with simultaneity. Under completeness, reduced form (48) characterizes the **net effect** of a policy change X_1:

$$\frac{\partial Y_1}{\partial X_1} = \frac{\partial \phi_1(X_1, X_2, U_1, U_2)}{\partial X_1}. \tag{50}$$

Following Klein and Goldberger (1955) and Wright (1921, 1934), we can conduct "mediation analyses" that address problem **P-2** and trace the impact of an externally manipulated X_1 on Y_1, both through its direct effect on (46) and its indirect effect through Y_2:

$$\frac{\partial Y_1}{\partial X_1} = \underbrace{\underbrace{\left(\frac{\partial g_{Y_1}}{\partial Y_2}\right)}_{\substack{\text{From} \\ \text{Structure}}} \underbrace{\left(\frac{\partial Y_2}{\partial X_1}\right)}_{\substack{\text{From Reduced} \\ \text{Form}}}}_{\substack{\text{Indirect effect} \\ \text{through } Y_2}} + \underbrace{\underbrace{\frac{\partial g_{Y_1}}{\partial X_1}}_{\text{From Structure}}}_{\text{Direct effect}} = \frac{\partial \phi_1(X_1, X_2, U_1, U_2)}{\partial X_1}.$$

This approach can be readily applied to recursive systems and general multiple equation systems. Reliance on linear equations, while traditional in the literature, is not necessary and nonparametric approaches are available.[67]

Mediation analysis is a staple of econometric policy evaluation to examine all channels of influence of variables (see, e.g., Theil, 1958). All of the tools used to analyze simultaneous equations are available to estimate these models (See e.g., Amemiya, 1985; Fisher, 1966; Matzkin, 2007). Klein and Goldberger (1955) is a classic example of dynamic mediation analysis in a Keynesian model of the time series of consumption and investment in the U.S. economy.

---

[64] In a market for a good, shutting down the supply equation would likely alter the properties of demand curves as agents would face a different market structure altering their expectations. Construction of a theory consistent counterfactual world would entail such considerations.

[65] Or more generally, X_2 is not an argument of g_{Y_1}.

[66] The completeness condition is part of the hypothetical model thought experiment. In some contexts it may be ruled out as not credible.

[67] See Matzkin (2008, 2015, 2013) for nonparametric analyses of such systems.

8. Conclusion

This paper presents the basic framework of the econometric model for causal analysis. We discuss the definition of causal parameters and approaches to their identification within it. We consider two recent approaches to causality that are used in the non-economic literature on causal inference and their relationship with the econometric approach.

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

The two competing statistical approaches are: (a) the Neyman–Rubin (NR) approach rooted in the statistics of experiments, and (b) the *do-calculus* (DoC) that originated in computer science. Both address some of the same problems tackled by the econometric approach. Each has important, but different, limitations. Neither has the flexibility or clarity of the econometric approach.

All start from the basic intuitive definition of a causal effect as a *ceteris paribus* consequence of a change in inputs on outcomes, where the change can be a policy. However, the rules for constructing and identifying counterfactuals are very different in these approaches.

The do-calculus (DoC) invokes a special set of rules for identifying causal parameters that lie outside of probability theory and that use a limited class of identifying assumptions for behavioral equations. It relies heavily on recursive directed-acyclic-graphs and assumptions about conditional independence relationships. Its rigid rules preclude the use of many traditional techniques of identification and estimation.

The Neyman–Rubin (NR) approach eschews the benefits of structural equations and many fruitful strategies for their identification. Reflecting its origins, it casts all policy problems into a "treatment-control" framework. Randomized experiments rather than thought experiments are foundational elements in this approach. In some versions, it conflates issues of definition with issues of identification. Its lack of reliance on structural equations with explicit links to theory and explicit analyses of unobservables, makes it difficult to interpret estimates obtained from it to analyze well-posed economic questions with it using the large toolkit of modern econometrics, or to synthesize studies within a common framework.

Econometrics has a rich body of theory and tools to address policy problems. Applied economists would do well by using the impressive set of conceptual tools available from econometric theory.

References

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

Abadie, A., 2005. Semiparametric difference-in-differences estimators. Rev. Econom. Stud. 72 (1), 1–19.

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

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

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

Angrist, J.D., Imbens, G.W., Rubin, D., 1996. Identification of causal effects using instrumental variables. J. Amer. Statist. Assoc. 91 (434), 444–455.

Angrist, J.D., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton, NJ.

Bareinboim, E., Pearl, J., 2016. Causal inference and the data-fusion problem. Proc. Natl. Acad. Sci. 113 (27), 7345–7352.

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? Q. J. Econ. 119 (1), 249–275.

Bjerkholt, O., Dupont, A., 2010. Ragnar Frisch's conception of econometrics. Hist. Political Econ. 42 (1), 21–73.

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

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

Bollen, K.A., 1989. Structural Equations with Latent Variables. Wiley, New York.

Bollen, K.A., 2002. Latent variables in psychology and the social sciences. Annu. Rev. Psych. 53 (1), 605–634.

Bollen, K.A., Pearl, J., 2013. Eight myths about causality and structural equation models. In: Morgan, S.L. (Ed.), Handbook of Causal Analysis for Social Research. Springer, Dordrecht, pp. 301–328.

Brock, W.A., Durlauf, S.N., 2007. Identification of binary choice models with social interactions. J. Econometrics 140 (1), 52–75.

Buchinsky, M., Pinto, R., 2021. Using economic incentives to generate monotonicity criteria of IV models. NBER Unpublished Manuscript, UCLA.

Bursztyn, L., Yang, D.Y., 2021. Misperceptions about others. Working Paper 29168, NBER Unpublished.

Cameron, S.V., Heckman, J.J., 1998. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. J. Polit. Econ. 106 (2), 262–333.

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

Chatfield, C., 2000. Time-Series Forecasting. CRC Press.

Cox, D.R., 1958. Planning of Experiments. Wiley, New York.

Cunha, F., Heckman, J.J., Navarro, S., 2005. Separating uncertainty from heterogeneity in life cycle earnings, the 2004 hicks lecture. Oxf. Econ. Pap. 57 (2), 191–261.

Cunha, F., Heckman, J.J., Schennach, S.M., 2010. 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.

Dawid, A.P., 1979. Conditional independence in statistical theory (with discussion). J. R. Stat. Soc. Ser. B (Stat. Methodol.) 41 (1), 1–31.

Dippel, C., Gold, R., Heblich, S., Pinto, R., 2020. Mediation analysis in IV settings with a single instrument. Unpublished Manuscript.

Ekeland, I., Heckman, J.J., Nesheim, L., 2004. Identification and estimation of hedonic models. J. Polit. Econ. 112 (S1), S60–S109, Paper in Honor of Sherwin Rosen: A Supplement to Volume 112.

Fisher, R.A., 1935. The Design of Experiments. Oliver and Boyd, London.

Fisher, F., 1966. The Identification Problem in Econometrics. In: Economics handbook series, McGraw-Hill.

Frangakis, C.E., Rubin, D., 2002. Principal stratification in causal inference. Biometrics 58 (1), 21–29.

Frisch, R., 1930. A dynamic approach to economic theory: Lectures by ragnar Frisch at yale university. Lectures at Yale University beginning September, 1930. Mimeographed, 246 pp. Frisch Archives, Department of Economics, University of Oslo.

Frisch, R., 1933. In: Bjerkholt, O., Dupont-Kieffer, A. (Eds.), Problems and Methods of Econometrics: The Poincaré Lectures of Ragnar Frisch, 1933. Routledge, New York, New York, published 2009.

Frisch, R., 1933. Problèmes et méthodes de l'économétrie. Eight lectures given at Institut Henri Poincaré, University of Paris, March-April 1933. Frisch Archive, 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., Verma, T., Pearl, J., 1990. Identifying independence in Bayesian networks. Networks 20 (5), 507–534.

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

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

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

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

Haavelmo, T., 1944. The probability approach in econometrics. Econometrica 12 (Supplement), iii–vi and 1–115.

Hamilton, J.D., 2000. Time Series Analysis. Princeton University Press.

Hansen, B.E., 2022. Econometrics. Princeton University Press, Princeton University Press.

Hansen, L.P., Sargent, T.J., 1982. Instrumental variables procedures for estimating linear rational expectations models. J. Monet. Econ. 9 (3), 263–296.

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

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

Heckman, J.J., 2008a. Econometric causality. Internat. Statist. Rev. 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., Humphries, J.E., Veramendi, G., 2018. Returns to education: the causal effects of education on earnings, health, and smoking. Journal of Political Economy 126 (S1), S197–S246.

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

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

Heckman, J.J., Leamer, E.E., 2001. Handbook of Economics. In: Handbooks in Economics, vol. 5, North Holland, Amsterdam.

Heckman, J.J., Leamer, E.E., 2007. Handbook of Econometrics. In: Handbooks in Economics, vol. 6AB, North Holland, Amsterdam.

Heckman, J.J., Navarro, S., 2004. Using matching, instrumental variables, and control functions to estimate economic choice models. Rev. Econ. Stat. 86 (1), 30–57.

Heckman, J.J., Pinto, R., 2015. Causal analysis after haavelmo. Econom. Theory 31 (1), 115–151.

Heckman, J., Pinto, R., 2018. Unordered monotonicity. Econometrica 86, 1–35.

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

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

Heckman, J.J., Singer, B.S., 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52 (2), 271–320.

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

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

Heckman, J.J., Vytlacil, E.J., 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects. Proc. Natl. Acad. Sci. 96 (8), 4730–4734.

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

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

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

Holland, P.W., 1986. Statistics and causal inference. J. Amer. Statist. Assoc. 81 (396), 945–960.

Holland, P.W., 1997. Some reflections on freedman's critiques. In: Topics in the Foundation of Statistics. Springer, pp. 50–57.

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

Huang, Y., Valtorta, M., 2006. Pearl's calculus of intervention is complete. In: Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence. UAI '06, AUAI Press, Arlington, Virginia, USA, pp. 217–224.

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

Imai, K., Keele, L., Tingley, D., Yamamoto, T., 2011. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. Am. Political Sci. Rev. 105, 765–789.

Imai, K., Keele, L., Yamamoto, T., 2010. Identification, inference and sensitivity analysis for causal mediation effects. Statist. Sci. 25 (1), 51–71.

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

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

Jöreskog, K.G., 1973. Analysis of covariance structures. In: Multivariate Analysis–III. Elsevier, pp. 263–285.

Kiiveri, H., Speed, T.P., Carlin, J.B., 1984. Recursive causal models. J. Australian Math. Soc. (Ser. A) 30–52.

Klein, L.R., Goldberger, A.S., 1955. An Econometric Model of the United States, 1929–1952. North-Holland Publishing Company, Amsterdam.

Knight, F., 1921. Risk, Uncertainty and Profit. Houghton Mifflin Company, New York.

Koopmans, T.C., Rubin, H., Leipnik, R.B., 1950. Measuring the equation systems of dynamic economics. In: Koopmans, T.C. (Ed.), Statistical Inference in Dynamic Economic Models. In: Cowles Commission Monograph, (no. 10), John Wiley & Sons, New York, pp. 53–237.

Lauritzen, S.L., 1996. Graphical Models. Clarendon Press, Oxford, UK.

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

Lewbel, A., 2019. The identification zoo: Meanings of identification in econometrics. J. Econ. Lit. 57 (4), 835–903.

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

Lucas, Jr., R.E., 1976. Econometric policy evaluation: A critique. In: Brunner, K., Meltzer, A.H. (Eds.), The Phillips Curve and Labor Markets. In: Carnegie-Rochester Conference Series on Public Policy, vol. 1, North-Holland, Amsterdam.

Manski, C.F., 1993. Identification of endogenous social effects: The reflection problem. Rev. Econom. Stud. 60 (3), 531–542.

Marschak, J., 1953. Economic measurements for policy and prediction. In: Hood, W.C., Koopmans, T.C. (Eds.), Studies in Econometric Method. Yale University Press, New Haven, CT, pp. 1–26.

Marshall, A., 1961. Principles of Economics, ninth (Valorium) Edition London, Macmillan for the Royal Economic Society.

Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic Theory. Oxford University Press, New York.

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

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

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. Annu. Rev. Econ. 5 (1), 457–486.

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

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.

Moffitt, R.A., 2001. Policy interventions, low-level equilibria, and social interactions. In: Durlauf, S., Young, P. (Eds.), Social Dynamics. vol. 4, (45–82), MIT Press, pp. 6–17.

Mogstad, M., Torgovitsky, A., 2018. Identification and extrapolation of causal effects with instrumental variables. Annu. Rev. Econ. 2, 577–613.

Morgan, S.L., Winship, C., 2015. Counterfactuals and Causal Inference. Cambridge University Press.

Mullins, J., 2024. A structural meta-analysis of welfare reform experiments and their impacts on children. Under review, Journal of Political Economy.

Nerlove, M., 1967. Recent empirical studies of the CES and related production functions. In: The Theory and Empirical Analysis of Production. National Bureau of Economic Research, pp. 55–136.

Neyman, J., 1923. Statistical problems in agricultural experiments. J. R. Stat. Soc. II (Supplement) (2), 107–180.

Olley, G.S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64 (6), 1263–1297.

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

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

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

Pearl, J., 2009a. Causality: Models, Reasoning, and Inference, second ed. Cambridge University Press, New York.

Pearl, J., 2009b. Causal inference in statistics: An overview. Stat. Surv. 3, 96–146.

Pearl, J., 2009c. 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., Jazzing, D., Schölkopf, B., 2017. Elements of Causal Inference: Foundations and Learning Algorithms. MIT Press, Cambridge, MA.

Prakasa Rao, B.L.S., 1992. Identifiability in Stochastic Models: Characterization of Probability Distributions. In: Probability and mathematical statistics, Academic Press, Boston.

Pratt, J.W., Schlaifer, R., 1984. On the nature and discovery of structure. J. Amer. Statist. Assoc. 79 (385), 9–33.

Quandt, R.E., 1958. The estimation of the parameters of a linear regression system obeying two separate regimes. J. Amer. Statist. Assoc. 53 (284), 873–880.

Quandt, R.E., 1988. The Econometrics of Disequilibrium. Blackwell, New York.

Robins, J., 1986. A new approach to causal inference in mortality studies with a sustained exposure period: Application to control of the healthy worker survivor effect. Math. Model. 7 (9–12), 1393–1512.

Rosen, S., 1986. The theory of equalizing differences. In: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics. vol. 1, North-Holland, New York, pp. 641–692.

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

Roy, A., 1951. Some thoughts on the distribution of earnings. Oxf. Econ. Pap. 3 (2), 135–146.

Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66 (5), 688–701.

Rubin, D.B., 1978. Bayesian inference for causal effects: The role of randomization. Ann. Statist. 6 (1), 34–58.

Schennach, S.M., 2020. Mismeasured and unobserved variables. In: Durlauf, S.N., Hansen, L.P., Heckman, J.J., Matzkin, R.L. (Eds.), In: Handbook of Econometrics, vol. 7, (A), Elsevier, pp. 487–565.

Shpitser, I., Pearl, J., 2006. Identification of joint interventional distributions in recursive semi-Markovian causal models. In: Proceedings of the 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference, AAAI-06/IAAI-06. In: 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., Pearl, J., 2009. Effects of treatment on the treated: Identification and generalization. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, UAI 2009. AUAI Press, pp. 514–521.

Spirtes, P., Glymour, C., Scheines, R., 2001. Causation, prediction, and search. The MIT Press.

Tamer, E., 2003. Incomplete simultaneous discrete response model with multiple equilibria. Rev. Econom. Stud. 70 (1), 147–165.

Telser, L.G., 1964. Iterative estimation of a set of linear regression equations. J. Amer. Statist. Assoc. 59 (307), 845–862.

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

Theil, H., 1958. Economic Forecasts and Policy. In: Contributions to Economic Analysis, (no. 15), North-Holland Publishing Company, Amsterdam.

Theil, H., 1971. Principles of Econometrics. Wiley, New York.

Tinbergen, J., 1930. Bestimmung und Deutung von Angebotskurven Ein Beispiel. Zeitschrift für Nationalökonomie 1 (5), 669–679.

Tinbergen, J., 1939. Statistical Testing of Business Cycle Theories: Part II: Business Cycles in the United States of America, 1919–1932. League of Nations, Economic Intelligence Service, Geneva.

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

Wright, S., 1921. Correlation and causation. J. Agric. Res. 20, 557–585.

Wright, S., 1934. The method of path coefficients. Ann. Math. Stat. 5 (3), 161–215.

Yamamoto, T., 2014. Identification and estimation of causal mediation effects with treatment noncompliance. Unpublished Manuscript, MIT Department of Political Science.