I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15081

Causality and Econometrics

James Heckman
Rodrigo Pinto

FEBRUARY 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15081

Causality and Econometrics

James Heckman
University of Chicago, Center for Economics of Human Development and IZA

Rodrigo Pinto
University of California, Los Angeles

FEBRUARY 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Causality and Econometrics*

This paper examines the econometric causal model for policy analysis developed by the seminal ideas of Ragnar Frisch and Trygve Haavelmo. We compare the econometric causal model with two popular causal frameworks: Neyman-Holland causal model and the do-calculus. The Neyman-Holland causal model is based on the language of potential outcomes and was largely developed by statisticians. 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. We make the case that economists who uncritically use these approximating frameworks often discard the substantial benefits of the econometric causal model to the detriment of more informative economic policy analyses. We illustrate the versatility and capabilities of the econometric framework using causal models that are frequently studied by economists.

JEL Classification: C10, C18

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

Corresponding author:
James J. Heckman
Department of Economics
and the Center for the Economics of Human Development
The University of Chicago
1126 E. 59th Street
Chicago IL 60637
USA

E-mail: jjh@uchicago.edu

* This research was supported in part by NIH grant NICHD R37HD065072. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Institutes of Health.

1 Introduction

Good policy analysis is causal analysis. It analyzes the factors that produce outcomes and the role of policies in doing so. It quantifies policy impacts. It elucidates the mechanisms producing outcomes in order to understand how they operate, how they might be improved and which, if any, alternative mechanisms might be used to generate outcomes. It uses all available information to give good policy advice.

It systematically explores possible counterfactual worlds. It is grounded in thought experiments – what might happen if determinants of outcomes are changed. In this regard, good policy analysis is good science. Credible hypothetical worlds are developed, analyzed, tested in 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 recently come to grips with the causal questions that have long being investigated by economists such as Ragnar Frisch and Trygve Haavelmo. As a result, private languages and procedures designed to approximate econometric models have been developed without any deep understanding of the corpus of econometric theory, and sometimes 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 to the definition or identification of counterfactuals when, in fact, they are at best handmaidens.

The current state of affairs would be of little concern if applied economists continued to draw on and extend the standard econometric model of policy evaluation. Sadly, this is not the case. Many econometricians and applied economists now emulate what they read

in statistics or computer science journals. They have forgotten or never learned their own field's foundational work to the detriment of rigorous causal policy analysis.

This paper discusses econometric policy analysis and recently developed approximations to it. Our goal is to improve the theory and practice of economic policy analysis by acquainting economists with their own rich econometric legacy and placing the recent approximations in the context of the econometric model.

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 for addressing policy problems. 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 causality within a model. The concept is simple, but requires thought processes outside statistics that are, nonetheless, quite intuitive. We discuss four distinct classes of policy problems that are addressed in econometric analyses. Some of them are either ignored or only partly addressed in the approximating literatures. We demonstrate the conceptual clarity of the econometric approach and contrast it with that of rival approaches.

In particular, we consider two causal approaches often advocated by statisticians and computer scientists. The first is the Neyman-Holland model (1923; 1958; 1974; 1986; 1996), "NR" henceforward. It uses some notions developed in rigorous econometrics but goes only part way toward implementing the full set of tools in the econometric approach to policy evaluation. It has important limitations for posing or analyzing routine policy problems outside a narrow "treatment-control" paradigm. 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–recursive models) and statistical conditional independence relationships. We demonstrate its limited capacity to address many important economic policy questions or to utilize many standard econometric estimation and identification tools.

Each of the approximating approaches has value for limited classes of problems. However, they have severe limitations when applied to the large array of problem economists routinely confront. The danger is that sole reliance on these tools eliminates serious consideration of important policy questions. The NR approach does not readily incorporate unobservables and restrictions on empirical relationships produced by economic theory that 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. They are all standard problems addressed in structural econometrics.

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

Each approximating approach has important conceptual and operational limitations compared to the econometric approach. We display the versatility and adaptability of the econometric approach and the limitations of the approximations.

This paper is organized as following. Section 2 discusses the notion of causality and 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 how the Neyman-Rubin causal model approximates the econometric model. Section 6 investigates how the do-calculus of Pearl (2009b) approximates the econometric model. Section 7 examines non-recursive models. Section 8 summarizes the paper.

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:

$$g : X \to Y \quad \text{over the domain of } X \quad \big(Dom(X)\big). \tag{1}$$

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:

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

In this example, stability means that α and β don't change when X or a component of it is changed. This is what is meant by **invariance** or **autonomy** of relationships (Frisch, 1938). It is a cornerstone of causal analysis.[1] However, more than stability of maps is required. Directionality is central. Inverting a map (when possible) may produce a stable relationship, but it is, in general, not causal. Standard examples of (1) and (2) in economics are production functions or demand equations.

The range of Y is a set of ***potential outcomes*** associated with X over its domain. g may be a function or a correspondence.[3] Potential outcomes associated with different values of X are *counterfactuals* associated with X. The key idea in causality is the notion captured in Alfred Marshall's phrase, "*ceteris paribus*" –all other else is equal.[4] 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

[1][2] The do-calculus explicitly uses autonomous structural relationships (Pearl, 2009b).
[3] Multiple equilibria are produced in many econometric models. See, e.g., Mas-Colell et al. (1995).
[4] Marshall (1961)

Generalized Roy model (1951) is an early example of a model of two potential outcomes associated with the income the same person would earn in different jobs.

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 casuality in the recent approximating literatures. For example, in an early version of the Neyman-Rubin model, Holland (1986) insists that causal effects are only defined for experimental manipulations of X. 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.[5] 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 an unobserved (by the analyst) variable:

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

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 the *statistical* relationship between Y and T (see e.g.,

[5] See Haavelmo (1943) for an early discussion of this distinction.

Holland, 1997; Pratt and Schlaifer, 1984). Doing so uses standard statistical tools to establish an empirical relationship. 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 also 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. Notice too that this example introduces considerations about the properties of random variables that are unnecessary for defining causality.

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.[6] Economists (and other scientists) use hypothetical models (thought experiments) to analyze phenomena and explore possible relationships. These and other possible relationships are not *defined* by statistical operations, although they are *estimated* using statistical methods.

To clarify these ideas, it is helpful to introduce ϵ_V, ϵ_T, ϵ_U which are unobserved (by the analyst) and mutually statistically independent random variables. They are external to the model (exogeneous) 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

[6] 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).

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 a common cause of T and U. The third causal model differs from the second model because U causes T. Nevertheless, the causal effect of T on Y remains β. The second and third models are statistically identical in the sense that T and U are not statistically independent and the OLS estimator is biased. The third model imposes a restriction on the variation in U. In the fourth model, T causes U and the OLS estimator of the parameter β does not, in general, identify the causal effect of T on Y because T also affects U. The OLS estimator of β captures both direct and indirect effects of T on Y. Let $Y(\epsilon) = t\beta + U$ be the counterfactual outcome Y when T is external set to value t.[7]

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 fixing 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 $Y(t)$ 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 do not belong to 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.

[7] $Y(t) \perp\!\!\!\perp T|U$ holds for the third model but not for the second model.

Frisch, the founding father of modern econometric causal policy analysis, clearly understood that causality is an exercise of 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*

2.3 The Econometric Approach to Causality

The econometric approach to causality develops explicit hypothetical models where inputs that cause outcomes. A common context is the study of policy evaluations in which 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 the mechanisms governing the choice of inputs is central to study the causal effect of treatment on the outcome. Identification/estimation/interpretation of empirical counterparts to the hypothetical counterfactuals require careful accounting for unobserved (by the analyst) variables (U) that cause both input choice and outcomes. Structural econometric models do just that.[8]

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.[9]

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*

[8]Caricatures sometimes made in the approximating 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.

[9]See Heckman (2008a).

large. The simplest forms of this problem are typically addressed in the approximation literatures: does a program in place "work" in terms of policy impacts?

The approximating 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 example, it seeks to know the sign and magnitude of β. However, most 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.[10] 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** 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.

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 outputs vary among environments. It replaces crude meta-analysis of treatment effects with principled explanations of mechanisms and their impacts and extrapolations of different answers to **P1**.[11] A common structural model is a useful vehicle for summarizing evidence from multiple studies.[12] Forecasting in new environments is a traditional problem in econometrics (see, e.g., Theil,

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

[11] 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 acknowledgement of a large body of econometric thought (see, e.g., Bareinboim and Pearl, 2016).

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

1958; Hamilton, 2000; Chatfield, 2000). However, the truly ambitious problem solved 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.[13]

The original impetus for the econometric approach was to conduct policy analysis for the post-World War II era using models fit on pre-World War II, Depression-era data. 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 where the policies were not in place.[14] The famous "critique" of Lucas (1976) updates Marschak's analysis to stochastic environments. McFadden (1974) is a Nobel Prize winning example of how a leading economist 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 that are often conflated in the approximating literatures:

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

[14] 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 judgements and insights beyond straight applications of fitted statistical models.

Table 1: **Three Distinct Tasks in Causal Policy 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 population	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. The models for counterfactuals do not require any statistical analysis. Identification is a separate issue required to recover β from large samples where statistical variation is not an issue. Estimation considers how to recover it in practice. Trygve Haavelmo, a student of Frisch, developed an empirically operational econometric framework for causal policy analysis that distinguished these three tasks (1943; 1944). We now state the econometric model formally using the modern notation of graph theory.

3 Econometric Causal Models

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

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 frameworks;

(c) Analyze data using a priori information from theory and/or previous studies going beyond crude statistical meta-analyses;

(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 economic theory;

(g) Exploit covariance restrictions across unobservables within and across equations to identify causal parameters;

(h) Make forecasts in new environments;

(i) Synthesize evidence across studies using common conceptual frameworks;

(j) Make forecasts of new policies never previously implemented; and

(k) Analyze the interactions across agents within markets and also within social settings (general equilibrium and peer effects).

The approximating approaches address subsets of these problems using limited toolkits.

The approximating approaches were developed to address specialized classes of problems – usually those in problem class P1. They may be very effective for analyzing the effects of causes using a limited set of tools. These studies typically focus on identifying average treatment effects or treatment on the treated. They embody Marschak's Maxim (Heckman, 2008a) that, for certain narrowly focused problems, specialized versions of the econometric approach may be highly effective. One need not necessarily implement more general models that address a wider set of questions to address specific problems. However, they are by

design, of limited value in addressing those wider problems. We now exposit the econometric causal model in depth.

3.1 Econometric Causal Framework

Heckman and Pinto (2015) develop a causal framework that formalizes Frisch's insight that causality is in the mind and places Havelmo'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 *hypothetical model* that formalizes the thought experiments of manipulating inputs that defining causality. The empirical model describes the data generating process, which differs from the hypothetical model which is an abstract model that characterizes Frish's notion of causality. They place the definition and operationalization of causality in a probabilistically consistent approach that does not require special rules or procedures invented to characterize causality used in portions of the approximating literature. Some notation is useful in describing the framework.

A causal model $\mathbb{M}$ is described as a system of structural equations like (1) that characterizes 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 $\mathcal{E} = \{\epsilon_K; K \in \mathcal{T}\}$ which contains an error term ϵ_K for each $K \in \mathcal{T}$. Error term ϵ_K shares the same dimension as K. This term is defined even if there are additional unobserved variables. They are technical assumptions designed to avoid degenerate random variables.

The 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 equation models where sets of variables are jointly determined.

Error terms are externally-specified (or exogenous). This means that error terms are not caused by any variable in $\mathcal{T}$. 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 $T = f_T(\epsilon_T)$. We impose, without loss of generality, that error terms are mutually statistically independent.[15] All variables are defined on a common probability space $(\mathcal{I}, \mathcal{F}, P)$.

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. We use $\mathcal{T}_h$, $\mathcal{E}_h$, $\mathbb{M}_h$, P_h, E_h for their counterparts in the hypothetical model.

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 of policy evaluation.[16] 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 X and an externally specified variable Z, which may be a policy variables that influences 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 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 $Y = T \cdot Y(1) + (1 - T) \cdot Y(0)$.[17] The typical solution

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

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

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

is to reformulate the problem 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. More generally, we seek to identify the probability distribution of the counterfactual outcomes $Y(t); t \in \{0, 1\}$.

The early Generalized Roy model has been generalized and extended in many ways.[18] The model is systematically ignored in the approximating literatures, despite its intellectual priority and relevance.[19] The Generalized Roy model allows the agent's decision to depend on unobserved variables V that account for subjective evaluation of the benefits of each choice (so it affects I) and to allow for multiple choices (see Heckman and Pinto, 2018; Heckman and Vytlacil, 2007a,b).

The Generalized Roy model consists of four variables $\mathcal{T}_e = \{Z, V, T, Y\}$. Z is an external policy vector that causes the treatment T, which in turn causes an outcome Y. Z plays the role of an instrumental variable. It causes Y only through its effects on T. V is an external set of confounding variables that jointly cause T and Y. Variables Z, T, Y are observed by the analyst; V is not. V is a source of selection bias in treatment choice, which makes evaluation of the causal effect of T on Y more difficult. The observed relationship between T and Y may be due to the common effect of V on both T, Y instead of the causal effect of T on Y. For now, we suppress the X variables for the sake of notational simplicity. We reintroduce such variables when relevant to our discussion.

The Roy model can be represented by the mapping $\mathbb{M}(Z) = \mathbb{M}(V) = \varnothing$, $\mathbb{M}(T) = \{V, Z\}$, $\mathbb{M}(Y) = \{V, T\}$, which imply the following structural equations:

$$V = f_V(\epsilon_V), \tag{4}$$

$$Z = f_Z(\epsilon_Z), \tag{5}$$

$$T = f_T(Z, V, \epsilon_T), \tag{6}$$

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

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

$$Y = f_Y(T, V, \epsilon_Y). \tag{7}$$

The independence of error terms ϵ_V, ϵ_Z, ϵ_T, ϵ_Y implies that $Z \perp\!\!\!\perp V$ and $Y \perp\!\!\!\perp Z \mid (T, V)$ hold where "$\perp\!\!\!\perp$" denotes independence. This model is recursive. We consider fully simultaneous models in a later section. The theory of Bayesian Networks offers useful tools for investigating the statistical properties of recursive causal models.[20]

We now describe some basic concepts used in that literature that underly the do-calculus and link Pearl's approach and the theory of Bayesian meta-analysis (Spiegelhalter et al., 1993) to the structural economics literature. $\mathbb{M}(K)$ are called *parents* of a variable $K \in \mathcal{T}$. Parents of K's parents are $\mathbb{M}^2(K) = \cup_{W \in \mathbb{M}(K)} \mathbb{M}(W)$. *Ancestors* of K include all higher order parental variables that lead to K, $\mathbb{A}(K) = \cup_{n=1}^{N} \mathbb{M}^n(K)$ for N such that $\mathbb{M}^N(K)$ contains only external variables.

The variables directly caused by K are called *children* of K, $\mathbb{Ch}(K) = \{W \in \mathcal{T}$ such that $K \in \mathbb{M}(W)\}$. The second order of children of K are $\mathbb{Ch}^2(K) = \cup_{W \in \mathbb{Ch}(K)} \mathbb{Ch}(W)$. *Descendants* of K include all the higher order children traced to K, $\mathbb{D}(K) = \cup_{n=1}^{N} \mathbb{Ch}^n(K)$ for N such that $\mathbb{Ch}^{N+1}(K) \subset \cup_{n=1}^{N} \mathbb{Ch}^n(K)$.

In this notation, the Generalized Roy model is a recursive (acyclic) model in which no variable is a descendant of itself, namely $K \notin \mathbb{D}(K)$ for each $K \in \mathcal{T}$. As we show below, causality does not require recursivity.

A useful property of recursive models is the *Local Markov Condition* (Kiiveri et al., 1984; Pearl, 1988). It states that a variable K is independent of its non-descendants conditional

[20] See Lauritzen (1996).

on its parents:[21]

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

For example, the outcome Y in the Generlized Roy model (4)–(7) has no descendants and its parents are $\mathbb{M}_e(Y) = \{V, T\}$. The LMC for Y is thus $Y \perp\!\!\!\perp Z \mid (T, V)$. Z has no parents and its descendants are T, Y. Thus, its LMC is $Z \perp\!\!\!\perp V$. In the literature outside economics, these recursive features are viewed by some as essential to the definition of causality when, as we show, they are not.

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 exogenously assigns values to the treatment variable. 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.

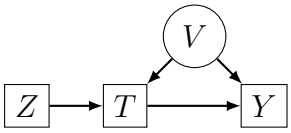
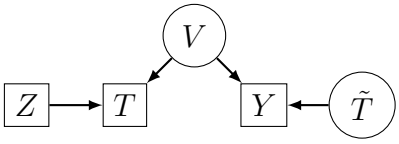
Frisch's thought experiment is conceptually simple. However, it is a causal operation outside the scope of statistical theory. In statistics, random variables are fully characterized by their joint distributions. This information by itself is insufficient for causal analysis as it lacks directionality – 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. In particular, assigning values to T differs from conditioning on T because conditioning changes the distribution of Z, V, whereas fixing T does not.

[21] Additional independence relationships may be generated by the Graphoid Axioms of Dawid (1979). These consist of five 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 $X \perp\!\!\!\perp Y \mid (W, Z) \Rightarrow X \perp\!\!\!\perp (W, Y) \mid Z.$

Frisch's thought experiment can be formalized and cast into a rigorous probability framework by a hypothetical model that adds an externally-specified hypothetical variable $\tilde{T}$ which causes the children of T (instead of T itself). The hypothetical model $\mathbb{M}_h$ has the same equations and the same distributions of error terms of the empirical model $\mathbb{M}_e$. It differs from the empirical model by appending a hypothetical variable $\tilde{T}$ which replaces the T-input of variables directly caused by T. Notationally, we have that $\mathcal{T}_h = \mathcal{T}_e \cup \{\tilde{T}\}$ such that $\mathbb{M}_h(\tilde{T}) = \varnothing$ and for each $K \in \mathcal{T}$ we have that $\mathbb{M}_h(K) = \{\tilde{T}\} \cup \{\mathbb{M}_e(K) \setminus \{T\}\}$ if $K \in \mathbb{Ch}_e(T)$ and $\mathbb{M}_h(K) = \mathbb{M}_e(K)$ otherwise. Table 3 represents the empirical Generalized Roy model and its hypothetical counterpart as DAGs (Directed acyclic graphs). Causal relationships are described by directed arrows, circles denote unobserved (by the analyst) variables, and squares denote observed variables. Below each DAG, we present the LMC for each variable of each model.

Table 3: Generalized Roy Model: Empirical and Hypothetical Causal Models

	Empirical Model	**Hypothetical Model**
		
	LMC	**LMC**
V :	$V \perp\!\!\!\perp Z$	$V \perp\!\!\!\perp (Z, \tilde{T})$
Z :	$Z \perp\!\!\!\perp V$	$Z \perp\!\!\!\perp (V, Y, \tilde{T})$
T :	$T \perp\!\!\!\perp \varnothing \mid (Z, V)$	$T \perp\!\!\!\perp (\tilde{T}, Y) \mid (Z, V)$
Y :	$Y \perp\!\!\!\perp Z \mid (T, V)$	$Y \perp\!\!\!\perp (Z, T) \mid (\tilde{T}, V)$
$\tilde{T}$:	(not defined for the model)	$\tilde{T} \perp\!\!\!\perp (T, V, Z)$

The hypothetical variable $\tilde{T}$ is external. Therefore it has no parents. According to (8), the hypothetical variable $\tilde{T}$ is independent of all its non-descendants, and, in particular, $\tilde{T} \perp\!\!\!\perp T$ always holds. The hypothetical model is defined by a thought experiment, whereas the empirical model is the data-generated process. The hypothetical model breaks the direct $T \to Y$ link and replaces it with a $\tilde{T} \to Y$ link.

Counterfactuals are generated by hypothetical (external) manipulations of treatments. These are produced in the hypothetical model by *conditioning* on the hypothetical variable $\tilde{T}$. For instance, the distribution of the counterfactual outcome Y when the treatment is externally set to a value $t \in \text{supp}(T)$ is $P_h(Y \mid \tilde{T} = t)$ and the counterfactual outcome mean is given by $E_h(Y \mid \tilde{T} = t)$. These are in contrast to the empirical counterparts $P_e(Y \mid T = t)$ and $E_e(Y \mid T = t)$.

Treatment effects are often (but not inevitably) defined at the population level by expected values of counterfactual differences. To fix ideas, suppose that T is a binary variable that indicates college graduation and Y denotes adulthood income. The average treatment effect of college on income is given by $ATE = E_h(Y \mid \tilde{T} = 1) - E_h(Y \mid \tilde{T} = 0)$. Treatment-on-the-treated (TOT) is the average causal effect of college on income by those who choose to go to college ($T = 1$), which is given by $TOT = E_h(Y \mid \tilde{T} = 1, T = 1) - E_h(Y \mid \tilde{T} = 0, T = 1)$.

The hypothetical model describes an external manipulation that entails several causal parameters. In the example, of the Generalized Roy model, we focus on an external variation of the treatment variable T that causes a single variable, the outcome Y. The model is suitable for investigating counterfactual distributions and forming ATE, TT, and several other causal parameters. The hypothetical variable can also be used in empirical models where the treatment directly causes multiple variables. The hypothetical model can be used to investigate all causal links of a treatment variable or a subset of these links conditioned on various populations.

Alternative Counterfactual Approaches

Counterfactual analysis modifies the original empirical model to characterise the causal operation of external manipulation. Such modification are often a source of confusion as they do not follow from any standard statistical tool. This is why causal analysis can be so challenging for people trained exclusively in statistics. Using the hypothetical model is just one of several approaches that supplement statistical theory in an effort to assess causality.

We describe two additional approaches that can be used to define counterfactuals: the fixing operator and the do-operator of Pearl (2012). Both *fix* and *do* operators formalize the notion of counterfactuals by suppressing some aspect of the original empirical model.

The fix operator is commonly used in economics (Heckman and Pinto, 2015). It is implicit in Haavelmo's pioneering paper (1943). It defines counterfactuals by *deleting the causal link* between treatment T and its children. In the empirical model of equations (4)–(7), the counterfactual outcome $Y(t)$ is obtained by *fixing* the T-argument of the outcome equation (7) to a value $t \in \text{supp}(T)$, so that $Y(t) = f_Y(t, V, \epsilon_Y)$. There is no direct empirical counterpart to this concept without further analysis. Fixing does not eliminate the structural equation for treatment variable T. It only modifies the outcome equation by replacing the random variable T by a fixed treatment value $t \in \text{supp}(T)$. Thus the variable T is still present in the causal model when fixing is applied.

The do-operator of Pearl (2009b, 2012) resembles fixing in the sense that it replaces all the T-inputs of the structural equations for all the variables directly caused by T. The do-operator differs from fixing by ***deleting*** ("shutting down") the structural equation for treatment variable T (Pearl, 2012).

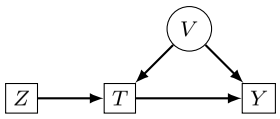
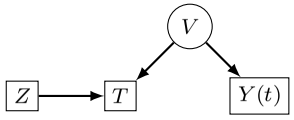
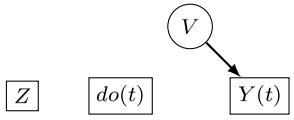
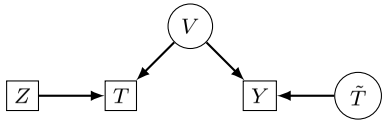
Neither *fix* nor the *do* operator are well-defined in statistics. They differ from statistical conditioning because conditioning on $T = t$ would, in general, change the distribution of all model variables (i.e. V, Y and Z) in the empirical model while *fixing* or *doing* T to a value t does not change the distribution of its ancestors V, Z.

Table 4 compares the different approaches for generating counterfactuals for the Generalized Roy model. The first column presents the original empirical model. The second and third columns present the models generated by the *fix* and the *do* operators respectively. The last column presents the hypothetical model.

The first panel of the table displays the structural equations for each model. The empirical model is our benchmark. The *fix* and *do* operators can be understood as sub-models that

Table 4: Generalized Roy Model: Approaches to Generating Counterfactuals

e: empirical model; e^*: model when treatment fixed; $e^\dagger$: model when T is "done"-*do* (T); h: hypothetical model

	Empirical Models			**Hypothetical Model**
	Original Model (e)	Fixing T at t (e^*)	$do(t)$ $(e^\dagger)$	Hypothetical Var. $\tilde{T}$ (h)
	Structural Equations			
V : Z : T : Y : $\tilde{T}$:	$V = f_V(\epsilon_V)$ $Z = f_Z(\epsilon_Z)$ $T = f_T(Z, V, \epsilon_T)$ $Y = f_Y(T, V, \epsilon_Y)$	$V = f_V(\epsilon_V)$ $Z = f_Z(\epsilon_Z)$ $T = f_T(Z, V, \epsilon_T)$ $Y(t) = f_Y(t, V, \epsilon_Y)$	$V = f_V(\epsilon_V)$ $Z = f_Z(\epsilon_Z)$ $do(T = t)$ $Y(t) = f_Y(t, V, \epsilon_Y)$	$V = f_V(\epsilon_V)$ $Z = f_Z(\epsilon_Z)$ $T = f_T(Z, V, \epsilon_T)$ $Y = f_Y(\tilde{T}, V, \epsilon_Y)$ $\tilde{T} = f_{\tilde{T}}(\epsilon_{\tilde{T}})$
	Directed Acyclic Graphs (DAGs)			
				
	Local Markov Conditions			
V : Z : T : Y : $\tilde{T}$:	$V \perp\!\!\!\perp Z$ $Z \perp\!\!\!\perp V$ $T \perp\!\!\!\perp \varnothing \mid (Z, V)$ $Y \perp\!\!\!\perp Z \mid (T, V)$ (not defined for the model)	$V \perp\!\!\!\perp Z$ $Z \perp\!\!\!\perp (V, Y(t))$ $T \perp\!\!\!\perp Y(t) \mid (Z, V)$ $Y(t) \perp\!\!\!\perp (Z, T) \mid V$ (not defined for the model)	$V \perp\!\!\!\perp Z$ $Z \perp\!\!\!\perp (V, Y(t))$ (not defined for the model) $Y(t) \perp\!\!\!\perp Z \mid V$ (not defined for the model)	$V \perp\!\!\!\perp (Z, \tilde{T})$ $Z \perp\!\!\!\perp (V, Y, \tilde{T})$ $T \perp\!\!\!\perp (\tilde{T}, Y) \mid (Z, V)$ $Y \perp\!\!\!\perp (Z, T) \mid (\tilde{T}, V)$ $\tilde{T} \perp\!\!\!\perp (T, V, Z)$
	Factorial Decomposition of the Joint Probability Distributions			
	$P_e(Y, T, V, Z) =$ $P_e(Y \mid T, V) P_e(T \mid Z, V) P_e(V) P_e(Z)$	$P_{e^*}(Y(t), T, V, Z) =$ $P_{e^*}(Y(t) \mid V) P_{e^*}(T \mid V, Z) P_{e^*}(V) P_{e^*}(Z)$	$P_{e^\dagger}(Y(t), V, Z) =$ $P_{e^\dagger}(Y(t) \mid V) P_{e^\dagger}(V) P_{e^\dagger}(Z)$	$P_h(Z, V, T, \tilde{T}, Y) =$ $P_h(Y \mid \tilde{T}, V) P_h(T \mid Z, V) P_h(V) P_h(Z) P_h(\tilde{T})$

remove some elements of the empirical model. The *fix* operator replaces the T-input of the outcome equation by a value $t \in \text{supp}(T)$. It has the same number of variables as the empirical model, but (counterfactually) evaluates them for a fixed value of T. The *do*-operator suppresses ("shuts down") the treatment equations altogether. It *eliminates* the treatment variable. The hypothetical model *adds* the hypothetical variable $\tilde{T}$, which replaces the T-input of the outcome equation.

The second panel displays the DAGs for each model. The first column displays the DAG for the empirical model. The DAG for the *fix* operator (second column) removes the arrow that arises from T into Y. Otherwise stated, the fix operator breaks the causal link of the treatment variable but maintains all of the variables of the empirical model. The *do* operator *excludes* the variable T. Its DAG suppresses all arrows arriving into or out of T. This is why the commonplace concept of "treatment on the treated" is so challenging for the *do-calculus* and requires special manipulations.[22] The DAG of the hypothetical model is similar to the DAG for the fix operator. It also breaks the causal link arising from T by replacing the treatment T by the hypothetical variable $\tilde{T}$.

The third panel presents the LMC of each of the model variables. The independence conditions depend on the variables in each counterfactual model. The outcome LMC of fixing model generates the following independence relationship:

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

This is sometimes called a *matching condition*. It states that the counterfactual outcome $Y(t)$ is independent of the treatment variable T conditional on the confounding variable V. The corresponding matching condition for the hypothetical model is:

$$Y \perp\!\!\!\perp T \mid (\tilde{T}, V). \tag{10}$$

Matching conditions (9) and (10) are equivalent. They play primary roles in devising methods to identify treatment effects.

[22] See Shpitser and Pearl (2009).

The *do* operator eliminates the treatment T from the set of model variables. It does not generates a matching condition like that in (9) or (10). Instead, Pearl (2009b) develops a DAG criteria to check for analogs to matching conditions in the empirical model. In the language of the do-calculus, matching conditions (9)–(10) are described by the private language "V d-separates Y and T." The elimination of the treatment T from the analysis does not permit researchers to investigate parameters such as the TOT because the treatment effect is conditioned on the values of the treatment. Shpitser and Pearl (2009) solve this problem by supplementing the counterfactual model with additional special structure.

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_{e^*} for the model generated by the fix operator, $P_{e^\dagger}$ for the *do* operator and P_h for the hypothetical model. The factorizations differ according to the number of variables in each counterfactual model.

All models share the same distributions of error terms. Consequently, the joint distribution of the ancestors of T, that is (V, Z), is the same across all models. The distribution of the counterfactual outcome $Y(t)$ depends only on V and ϵ_Y. Therefore, the distribution of the counterfactual outcomes is the same regardless of whether we use the *fix* or the *do* operator.

One benefit of the hypothetical model is that is enables analysts to use probability to converse with causality without introducing new (and unnecessary) concepts. It translates the probabilistically ill-defined causal operations of *fixing* or *doing* into standard statistical conditioning. Formally, for any set K of non-descendant variables of $\tilde{T}$ and any variable Y that is a descendant of $\tilde{T}$ in the hypothetical model, we have that:

$$\begin{aligned} \Big(Y \mid \tilde{T} = t, K\Big)_{\mathbb{M}_h} &\stackrel{d}{=} \Big(Y(t) \mid K\Big)_{\mathbb{M}_{e^*}} \text{ and} \\ \Big(Y \mid \tilde{T} = t, \big\{K \setminus \{T\}\big\}\Big)_{\mathbb{M}_h} &\stackrel{d}{=} \Big(Y(t) \mid \big\{K \setminus \{T\}\big\}\Big)_{\mathbb{M}_{e^\dagger}} \end{aligned} \tag{11}$$

where $\big(Y \mid \tilde{T} = t, K\big)_{\mathbb{M}_h}$ denotes the variable Y conditional on K and on the event $\tilde{T} = t$ in the hypothetical model, $\big(Y(t) \mid K\big)_{\mathbb{M}_{e^*}}$ and $\big(Y(t) \mid K\big)_{\mathbb{M}_{e^\dagger}}$ denote the counterfactual

outcome under fixing and doing respectively. In particular, we have that $\big(Y \mid \tilde{T} = t\big)_{\mathbb{M}_h} \stackrel{d}{=} \big(Y(t)\big)_{\mathbb{M}_{e^*}} \stackrel{d}{=} \big(Y(t)\big)_{\mathbb{M}_{e^\dagger}}$.

Even though all the models share many common concepts, they differ greatly regarding the machinery used to *identify* causal effects.

Identification of Counterfactual Outcomes

We now move to Task 2 in Table 1. Counterfactuals are said to be identified if they can be expressed in terms of the probability distributions of the observed data generated by the empirical model. Thus identification requires the analyst to connect the probability distribution of the hypothetical model with the probability distributions of the empirical model. A connection between empirical and hypothetical models is made if we can justify the following criteria: for any disjoint set of variables Y, W in $\mathcal{T}$ and any subsets $\mathcal{A}, \mathcal{A}' \subset \text{supp}(T)$ we have that:[23]

$$Y \perp\!\!\!\perp \widetilde{T} \mid (T, W) \Rightarrow P_h\Big(Y \mid \widetilde{T} \in \mathcal{A}, T \in \mathcal{A}', W\Big) = P_h(Y \mid T \in \mathcal{A}', W) = P_e(Y \mid T \in \mathcal{A}', W). \quad (12)$$

$$Y \perp\!\!\!\perp T \mid (\widetilde{T}, W) \Rightarrow P_h\Big(Y \mid \widetilde{T} \in \mathcal{A}, T \in \mathcal{A}', W\Big) = P_h(Y \mid \widetilde{T} \in \mathcal{A}, W) = P_e(Y \mid T \in \mathcal{A}, W). \quad (13)$$

Equations (12)-(13) state that we can move from the hypothetical model to the empirical model whenever the independence relationships (12): $Y \perp\!\!\!\perp \widetilde{T} \mid (T, W)$ or (13): $Y \perp\!\!\!\perp T \mid (\widetilde{T}, W)$ apply. The relationships are symmetric in the roles played by T and $\tilde{T}$. While $Y \perp\!\!\!\perp \widetilde{T} \mid (T, W)$ is an independence relationship between some variable Y and $\widetilde{T}$ conditioned on T, the independence $Y \perp\!\!\!\perp \widetilde{T} \mid (T, W)$ is an independence relationship between Y and T conditioned on $\widetilde{T}$.

Equations (12)-(13) are useful for describing the intuitive properties of the hypothetical model. Since the hypothetical variable $\widetilde{T}$ is externally specified and independent of all its non-descendants, which include the treatment T, $K \perp\!\!\!\perp \widetilde{T} \mid T$ holds for any variable K not caused by $\tilde{T}$. According to (13), we have that for $P_h(K \mid T \in \mathcal{A}') = P_e(K \mid T \in \mathcal{A}')$ and

[23] See Heckman and Pinto (2015) for a proof.

for $\mathcal{A}' = \text{supp}(T)$ we have that $P_h(K) = P_e(K)$. In other words, hypothetical variation of treatment does not change the distribution of its non-descendants.

Consider the hypothetical Roy model of Table 3. The LMC of Y generates the independence relationship $Y \perp\!\!\!\perp T \mid (\tilde{T}, V)$. Variable V is a matching variable. Conditioning on it generates the useful relation:

$$P_h(Y \mid \widetilde{T} = t, V) = P_{e^*}(Y(t) \mid V) = P_e(Y \mid T = t, V). \tag{14}$$

The first equality is justified by (11). It relates conditioning in the hypothetical model to fixing in the empirical model. The second equality is justfied by (13). If $Y \perp\!\!\!\perp T \mid (\tilde{T}, V)$ holds, we can access the counterfactual outcome by conditioning on V. Otherwise stated, if the confounding variable V were observed and we could condition on it, we would be able to evaluate the counterfactual outcome. Moreover, V is not a descendant of $\tilde{T}$, which implies that $P_h(V) = P_e(V)$. Thus if V were observed, the probability distribution of the counterfactual $P_h(Y \mid \tilde{T} = t)$ would be obtained by integrating $P_e(Y \mid T = t, V = v)$ over the values v in the support of V.

The econometric literature provides an unusually rich menu of strategies to eliminate the confounding effects of V not available in the approximating literature. We discuss some of this menu in the next section.

4 Identification of Counterfactuals in the Generalized Roy Model

The Generalized Roy model is a laboratory for exploring the large toolkit of the econometric approach to identifying counterfactuals compared to what is possible in the approximating paradigms. We describe several of these approaches here.

Equation (14) 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. They control for V by shifting T without affecting the distribution of V. However, the IV model described by equations (4)–(7) with Z as an instrument does not identify interesting counterfactuals without additional assumptions.

The literature on policy evaluation in structural settings provides a large array of additional tools that facilitate identification of the causal effect of T on Y. For example, the simplest identifying assumption is linearity. If the treatment and the outcome functions are linear, so $T = \alpha_0 + \alpha_1 V + \epsilon_T$, and $Y = \beta_0 + \beta_1 T + \beta_2 V + \epsilon_Y$, where α_0, α_1, β_0, β_1, β_2 are scalar parameters, the causal effect of T on Y is given by β_1. It is identified by the covariance ratio $cov(Y, Z)/cov(T, Z)$ and can be estimated by the Two-Stage Least Squares (2SLS) Regression. This tool has been available to economists since the 1950s.[24]

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 so that treating β_1 as a constant does not capture the essential heterogeneity of treatment effects across agents. The analogue to 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 heterogenous treatment effects and then consider alternative approaches.

Instrumental Variables

Heckman and Vytlacil (1999, 2005) address this problem assuming a separable choice equation. Their approach enables analysts to control for V and, in turn, identify counterfactual outcomes. Their local Instrumental Variable (LIV) Model considers a binary treatment $T \in \{0, 1\}$. Their *separability assumption* arises from 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

[24] See Amemiya (1985); Hansen (2021); Theil (1953, 1958, 1971). Theil (1953) invented this method.

the choice equation as:

$$T = \mathbf{1}\big[P(Z) \geq U\big]; \quad P(Z) = P_e\big(T = 1 \mid Z\big), \tag{15}$$

where $P(Z) = P_e(T = 1 \mid Z)$ is 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 computation with respect to the empirical model. Variable U has a uniform distribution if $\phi(V)$ is absolutely continuous; that is, $U \sim unif([0,1])$. The structural approach uses unobservables. The Neyman-Rubin approach does not. The *do-calculus* uses them, but in a limited way, and rules out separability that is used to obtain (15). This approach to unobservables precludes the use of methods that are fruitful in the econometric approach.

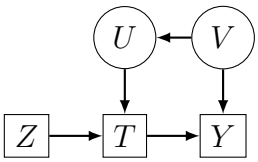
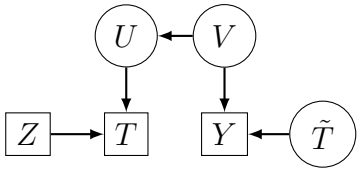
The hypothetical and empirical models for the Generalized Roy model that include the unobserved variable U are displayed in Table 5. The LMC of T in the hypothetical Roy model of Table 5 implies that $Y \perp\!\!\!\perp T \mid (Z, \tilde{T}, U)$. The LMC of Z implies $Y \perp\!\!\!\perp Z \mid (U, \tilde{T})$. These two independence relationships imply, by contraction property D, that $Y \perp\!\!\!\perp T \mid (\tilde{T}, U)$. Following the same analysis of V as (14), $Y \perp\!\!\!\perp T \mid (\tilde{T}, U)$ implies that:

$$P_h(Y \mid \widetilde{T} = t, U) = P_{e^*}(Y(t) \mid U) = P_e(Y \mid T = t, U). \tag{16}$$

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 $Y \perp\!\!\!\perp T \mid (\tilde{T}, V) \Rightarrow Y \perp\!\!\!\perp T \mid (\tilde{T}, U)$.[25]

[25]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**
$V:$	$V \perp\!\!\!\perp Z$	$V \perp\!\!\!\perp (Z, \tilde{T})$
$Z:$	$Z \perp\!\!\!\perp (U, V)$	$Z \perp\!\!\!\perp (V, U, Y, \tilde{T})$
$U:$	$U \perp\!\!\!\perp Z \mid V$	$U \perp\!\!\!\perp (Y, Z, \tilde{T}) \mid V$
$T:$	$T \perp\!\!\!\perp V \mid (Z, U)$	$T \perp\!\!\!\perp (\tilde{T}, V, Y) \mid (Z, U)$
$Y:$	$Y \perp\!\!\!\perp (Z, U) \mid (T, V)$	$Y \perp\!\!\!\perp (Z, U, T) \mid (\tilde{T}, V)$
$\tilde{T}:$	(not defined for the model)	$\tilde{T} \perp\!\!\!\perp (T, V, U, Z)$

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 $E_{e^*}(g(Y(t)) \mid U = u)$ 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 $E_h(g(Y \mid \tilde{T} = t, U = u)$ 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:

$$\begin{aligned} &E_h\big(Y \mid \tilde{T} = 1, U = u\big) - E_h\big(Y \mid \tilde{T} = 0, U = u\big) \\ &\equiv E_{e^*}\big(Y(1) - Y(0) \mid U = u\big) = \frac{\partial E_e(Y \mid P(Z))}{\partial P(Z)}\bigg|_{P(Z)=u}. \end{aligned} \tag{17}$$

where e^* refers to the distribution generated by fixing 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 $ATE \equiv E_h(Y \mid \tilde{T} = 1) - E_h(Y \mid \tilde{T} = 0) = \int_0^1 \big(E_h(Y \mid T = 1, U = u) - E_h(Y \mid T = 0, U = u)\big) du$.

Stratification

A recurrent theme in this section is that identification of counterfactual outcomes hinges on controlling for the confounding variable V. The solution of the LIV model invokes separability assumption (15) which generates a balancing score U for V. According to (18), 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 $u \in (0, 1)$.

Equation (17) assumes that the sample propensity score has enough variation around the value $u \in (0, 1)$. Consequently, the equation is not directly applicable to discrete instruments. One approach to overcome this limitation is to use the discrete counterpart of equation (17). 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:

$$\begin{aligned} \frac{E_e(Y \mid Z = z') - E_e(Y \mid Z = z)}{P_e(T = 1 \mid Z = z') - P_e(T = 1 \mid Z = z)} &= \frac{\int_u^{u'} E_{e^*}\Big(Y(1) - Y(0) \mid U = u\Big) du}{u' - u} \\ &= E_{e^*}(Y(1) - Y(0) \mid u \leq U \leq u'). \end{aligned} \tag{18}$$

Equation (18) 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.[26]

A consequence of (18) 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 ($P(z_1) = 1$):

$$E_e(Y \mid Z = z_1) - E_e(Y \mid Z = z_0) = E_{e^*}\Big(Y(1) - Y(0) \mid 0 \leq U \leq 1\Big)$$

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

$$= E_h(Y \mid \tilde{T} = 1) - E_h(Y \mid \tilde{T} = 0) = ATE.$$

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

Another approach to controlling for V exploits the discrete nature of the IV 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 $P(z_1) < \cdots < P(z_N)$.[27] 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* $\boldsymbol{S} = [T(z_1), \ldots, T(z_N)]'$ is the random vector of potential choices across all Z-values.

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

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

Similar to U, the response vector $\boldsymbol{S}$ is a balancing score for V. The independence relationship $Y \perp\!\!\!\perp T \mid (\tilde{T}, \boldsymbol{S})$ holds, which implies that $P_h(Y \mid \widetilde{T} = t, \boldsymbol{S}) = P_e(Y \mid T = t, \boldsymbol{S})$. Heckman and Pinto (2018) show that the response vector $\boldsymbol{S}$ 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 $\boldsymbol{S}$ is just a list of responses to treatments for a person of a given V.

The values of $\boldsymbol{S}$ are called response-types or strata.[28] The separability assumption eliminates some of 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 deniers. They invoke a monotonicity condition that is equivalent to the separability assumption.

[27] The increasing ordering of propensity scores is assumed without loss of generality.
[28] The concept was developed by Robins (1986) and Frangakis and Rubin (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 identification results.

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*.[29] 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 $Y \perp\!\!\!\perp T \mid (\tilde{T}, X)$. According to (13), we have that $P_h(Y \mid \tilde{T} = t, X) = P_{e^*}(Y(t) \mid X) = P_e(Y \mid T = t, X)$ which means that the counterfactual outcome is identified by conditioning on X. Matching variables X are assumed not to be a descendant of the hypothetical variable $\tilde{T}$, thus $P_h(X) = P_e(X)$ and the probability distribution of the counterfactual outcome is given by $P_{e^*}(Y(t)) = \int (P_e(Y \mid T = t, X = x) dF_{e,X}(x)$.

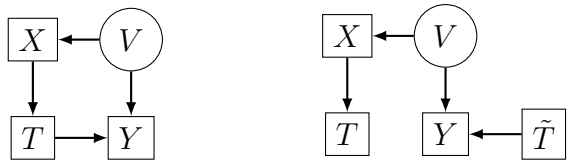
The average causal effect of a binary treatment $T \in \{0, 1\}$ is evaluated by the weighted average of mean difference between the treated and not-treated participants that *match* on X, namely, $ATE = \int \Big(E_e(Y \mid T = 1, X = x) - E_e(Y \mid T = 0, X = x)\Big) dF_{e,X}(x)$.[30]

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

[30] Heckman et al. (1998) incorporated 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.

Table 6: Matching Model: Empirical and Hypothetical Causal Models

Empirical Model **Hypothetical Model**



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 via 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 used in the randomization protocol. In observational studies, a 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.

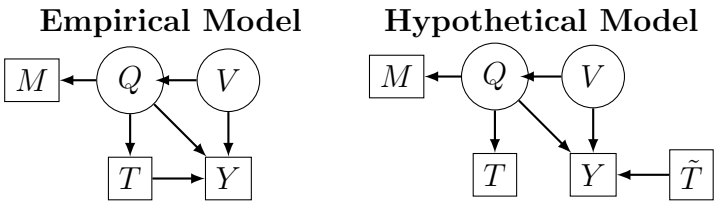
Matching on Proxied Unobservables

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 hypothetical model, $Y \perp\!\!\!\perp T \mid (\tilde{T}, Q)$, 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 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 $M = f_M(Q, \epsilon_M)$. The identification strategy is to explore 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) apply 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); 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



Control Function

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) 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 $K = f_K(T, V, \epsilon_K)$ represents unobserved skills caused by the treatment T and the unobserved confounding variable V. In addition, let the outcome equation be *additive* in K, that is to say that the outcome Y can be written as $Y = f_Y(T, \epsilon_Y) + \psi(K)$, The model is displayed as a DAG in Table 8. The LMC of Y in the hypothetical model implies that $Y \perp\!\!\!\perp T \mid (\tilde{T}, K)$. This means that K is a matching variable. The control function approach seeks to control for variable V by estimating the function $\psi(K)$ 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 $\psi(K)$ 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

$F_V(V) \sim unif([0,1])$. Note that the expected value of the outcome conditional on $T=1$ gives the *conditional* counterfactual mean:

$$E_e(Y \mid Z,\, T=1) = E_{e^*} h\big(Y(1) \mid Z,\, T=1\big) = E_h(Y \mid \tilde{T}=1,\, Z, T=1),$$

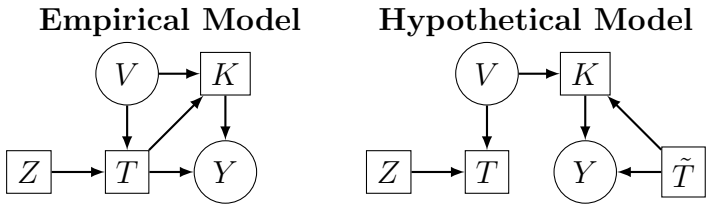
where the first term is observed, the second term uses fixing and the last one uses the hypothetical model. Under separability and outcome additivity, we can express $E_h(Y(1) \mid \tilde{T}=1, Z, T=1)$ as:

$$\begin{aligned}
E_h\big(Y \mid \tilde{T}=1, Z=z, T=1\big) &= E_h\big(f_Y(\tilde{T}, \epsilon_Y) \mid \tilde{T}=1\big) + E_h\big(\psi(K) \mid \tilde{T}=1, Z=z, T=1\big), \\
&= E_h\big(f_Y(1, \epsilon_Y)\big) + E_h\big(\psi\big(f_K(1, V, \epsilon_K)\big) \mid Z=z, T=1\big), \\
&\quad \Big(\text{setting } E_h(f_Y(1,\epsilon_Y)) = \alpha_1\Big) \\
&= \alpha_1 + E_h\Big(\psi\big(f_K(1, V, \epsilon_K)\big) \mid P(z) > F_V(V)\Big), \\
&= \alpha_1 + E_e\Big(\psi\big(f_K(1, V, \epsilon_K)\big) \mid P(z) > F_V(V)\Big), \\
\therefore E_h\big(Y \mid \tilde{T}=1, Z, T=1\big) &= \alpha_1 + \underbrace{f_1(P(Z))}_{\text{control function}}, \text{ where } f_1\big(P(Z)\big) = E_h\big(\psi\big(f_K(1, V, \epsilon_K)\big) \mid Z, T=1\big)
\end{aligned}$$

where the first equality uses the additivity assumption, the second uses the fact the $\tilde{T}$ is an external variable, the third uses the separability assumption, the fourth switches the hypothetical model into the empirical model as V, ϵ_K, Z are non-descendants of $\tilde{T}$. The last equation gives the expectation $E_h(Y \mid \tilde{T}=1, Z, T=1)$ as a function of the propensity score $P(Z)$. Control function $f_1(P(Z))$ can be estimated from observed data and the expected value of the counterfactual outcome can be evaluated as

$$E_h(Y(1)) = \int_0^1 \alpha_1 + f_1(p) dF_{P(Z)}(p).$$

Table 8: Control Function: Empirical and Hypothetical Causal Models



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* (Durlauf et al., 2020; 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

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

	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	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	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 and in long panels can identify V
Mixing Distributions[g]	$V \perp\!\!\!\perp X$	No (intervals of V)	Yes (Mixtures)

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

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).[31] 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. Causal models are characterized by statistical independence relationships among counterfactual counterparts of observed variables, never precisely defined.

The NR approach lacks the clarity of interpretation offered by causal models described by structural equations. It is very often difficult to map the independence relationships of a NR model into the actual causal relationships produced by economic theory. In particular, NR makes it difficult to assess the credibility of assumptions that ensure the identification of causal effects.

Another drawback is that the NR framework lacks fundamental tools of econometric causal analysis. It does not explicitly model unobserved variables in structural models. This feature substantially limits the use of the tools exposited in Section 4. 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,[32] and cross-equation restrictions on both observable and unobservable model components, 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, IV and its many surrogates and differences-in-differences (see Imbens and Rubin, 2015). This section illustrates drawbacks of NR in analyzing core policy questions.

[31] (Rosen, 1986) explains that SUTVA - Stable Unit Treatment Value Assumption - is a mixture of two two distinct concepts regarding function autonomy and no interaction among agents.

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

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 (4)–(7) using two counterfactuals: $T_i(z)$ is the potential treatment when the instrument Z is set to value $z \in \text{supp}(Z)$; and $Y_i(t, z)$ is the potential outcome of agent i when Z is set to value $z \in \text{supp}(Z)$ and choice T is set to $t \in \text{supp}(T)$. It does not explicitly characterize the choice equation. It prides itself on being nonparametric, although some proponents claim that assuming linearity is an assumption, even when models are fundamentally nonlinear.[33]

The NR framework characterises 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 $z, z' \in \text{supp}(Z)$ and for 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 $Z \perp\!\!\!\perp (Y(t), T(z))$.

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. The 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 (19) of Imbens and Angrist (1994). It states that a change in an instrument induces agents to change their treatment

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

choice towards the same direction. Notationally, for any $z, z' \in \text{supp}(Z)$, we have that:

$$T_i(z) \geq T_i(z') \ \forall i \in \mathcal{I} \quad \text{or} \quad T_i(z) \leq T_i(z') \ \forall i \in \mathcal{I} \tag{19}$$

Vytlacil (2002) shows that the monotonicity condition (19) 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):

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

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.

The power of analysis 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.

Table 10: Some Causal Parameters as Weighted Average the MTE

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

Source: Heckman and Vytlacil (2005).

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 equations (4)–(7) and its monotonicity criteria is equivalent to a separability condition. Its apparent simplicity is due to its lack of explicit statement of its assumptions.

The Matching Model in the NR

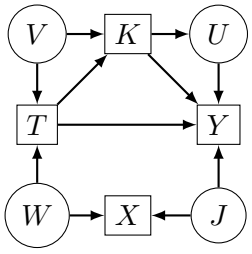
A common identification approach in NR is a *matching* assumption on observed variabes. It states that the treatment choice T is independent of counterfactual outcomes $Y(t)$ when conditioning on observed pre-treatment variables X, that is, $Y(t) \perp\!\!\!\perp T \mid X$.[34] Intuitively, the assumption 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 as-

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

sumption 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 infer that increasing the number of matching variables may only decrease the potential bias generated by unobserved confounders. This statement is known to be false.[35] However it is rather difficult to investigate the truth of this claim using the NR framework. The causal model of Table 11 clarifies this point.

Table 11: Hypothetical Matching Model

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

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

The causal model of Table 11 exemplifies the difficulty of performing causal investigation 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 easily

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

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

Mediation Models under NR: An example

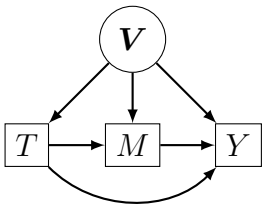
Mediation models originate in the path analysis and simultaneous equations literatures.[36] They trace the impacts of interventions on outcomes through their multiple channels of operation. Identifying the causal models generated by NR assumptions is often a daunting task and the economic content of these assumptions is often far from clear. We examine several mediation models to illustrate this fact 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. $\boldsymbol{V}$ 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 $M(t) = f_M(t, \boldsymbol{V}, \epsilon_M)$. The counterfactual outcome when the treatment is fixed at t and the mediator is fixed at $m \in \{0, 1\}$ is $Y(t, m) = f_Y(t, m, \boldsymbol{V}, \epsilon_Y)$. The counterfactual outcome when we fix only T at t is $Y(t) = f_Y(t, M(t), \boldsymbol{V}, \epsilon_Y)$.

Table 12: Mediation Model with Confounding Variable

Causal Model

$$\boldsymbol{V} = f_V(\epsilon_V)$$
$$T = f_T(\boldsymbol{V}, \epsilon_T)$$
$$M = f_M(T, \boldsymbol{V}, \epsilon_M)$$
$$Y = f_Y(T, M, \boldsymbol{V}, \epsilon_Y)$$

DAG

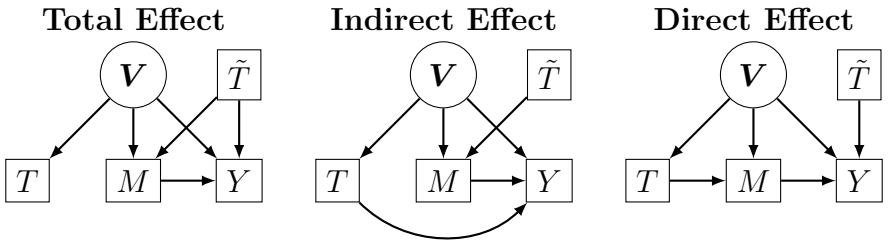


The goal of mediation models is to decompose the total effect of T on Y into an indirect effect that includes the effect of T on M and M on Y and a direct effect not mediated by M. To facilitate the discussion, let T and M denote binary variables taking values in $\{0, 1\}$. The

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

average (total) effect of T on Y is $E_{e*}(Y(1) - Y(0))$. We can also define the average direct effect of T on Y as $E_{e*}(Y(1, M) - Y(0, M)) = \sum_{m=0}^{1} E_{e*}(Y(1, m) - Y(0, m))P_e(M = m)$ and the average indirect effect as $E_{e*}(Y(T, 0) - Y(T, 1)) = \sum_{t=0}^{1} E_{e*}(Y(t, 1) - Y(t, 0))P_e(T = t)$.[37] Table 13 displays there hypothetical models suitable for examining the total, direct and indirect effects. The first DAG corresponds to the total effect. The hypothetical variable $\tilde{T}$ 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 $\tilde{T}$ replaces only the T-input of outcome equation.

Table 13: Hypothetical Models for the Mediation Model: Total, Direct and Indirect Effects



The confounding variable $\boldsymbol{V}$ prevents the identification of the counterfactual means $E_{e^*}(M(t))$ and $E_{e^*}(Y(t, m))$. A solution to this identification problem using NR is the Sequential Ignorability (SI):[38]

$$\big(Y(t', m), M(t)\big) \perp\!\!\!\perp T, \tag{22}$$

$$Y(t', m) \perp\!\!\!\perp M(t) \mid T, \tag{23}$$

[37] Alternatively, we can then define the direct effect and indirect effects for a given t by (20) and (21) respectively.

$$DE(t) = E_{e*}\big(Y(1, M(t)) - Y(0, M(t))\big) = \int E_{e*}\big(Y(1, m) - Y(0, m)\big) dF_{M(t)}(m) \tag{20}$$

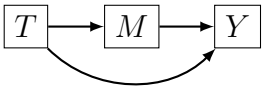
$$IE(t) = E_{e*}\big(Y(t, M(0)) - Y(0, M(1))\big) = \int E_{e*}(Y(t, m)) dF_{M(1)}(m) - \int E_{e*}(Y(t, m)) dF_{M(0)}(m). \tag{21}$$

[38] See Imai et al. (2011, 2010) for the properties of these assumptions.

for any $t, t' \in \text{supp}(T)$ and $m \in \text{supp}(M)$. SI (22)–(23) enables analysts to identify counterfactual means by statistical conditioning $E_e(M(t)) = E_{e^*}(M \mid T = t)$ and $E_{e^*}(Y(t,m)) = E_e(Y \mid T = t, M = m)$.

SI assumptions (22)–(23) can be understood as an application of the matching condition to mediation models. Assumption (22) 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 (23) 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 (Imai et al., 2010). SI assumptions (22)–(23) 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

Causal Model	**DAG**
$T = f_T(\epsilon_T)$ $M = f_M(T, \epsilon_M)$ $Y = f_Y(T, M, \epsilon_Y)$	

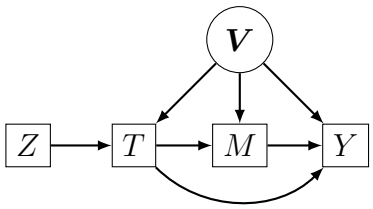
SI assumptions (22)–(23) are rather strong. They can be weakened if instrumental variables are available as depicted in Table 15. We use the model to exemplify a case in which NR assumptions are logically possible but generate a causal model that is difficult to justify using any plausible argument. The structural model enables the analyst to interpret the statistical assumptions using behavioral theory.

Table 15: Mediation Model with Instrumental Variables

Causal Model

$\boldsymbol{V} = f_V(\epsilon_V)$
$Z = f_Z(\epsilon_Z)$
$T = f_T(Z, \boldsymbol{V}, \epsilon_T)$
$M = f_M(T, \boldsymbol{V}, \epsilon_M)$
$Y = f_Y(T, M, \boldsymbol{V}, \epsilon_Y)$

DAG



The mediation model with IV has four counterfactuals, $T(z)$, $M(t)$, $Y(t)$, $Y(t,m)$ previously defined. In 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 $\boldsymbol{V}$.[39]. Suppressing Y generates an IV model where M plays the role of the outcome.

To dig more deeply, investigate the case of a binary instrument $Z \in \{0,1\}$. The response vector $\boldsymbol{S}_i = [T_i(0), T_i(1)]'$ 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 $\boldsymbol{S}$, the treatment choice T depends only on the instrument Z. The exogeneity condition states Z is independent of the counterfactual outcome $Y(t)$. Thus

$$T \perp\!\!\!\perp Y(t) \mid \boldsymbol{S}. \tag{24}$$

$\boldsymbol{S}$ is a balancing score for $\boldsymbol{V}$.

Yamamoto (2014) uses the language of NR to identify mediation effects using instrumental variables. His solution merges SI (22)-(23) with the matching property of the response vector $\boldsymbol{S}$ in (24). He advocates an assumption that he terms the *local average causal mediation effects (LACME) assumption*:

$$(Y(t,m), M(t')) \perp\!\!\!\perp T \mid (\boldsymbol{S} = [0,1]'), \tag{25}$$

$$Y(t,m) \perp\!\!\!\perp M(t') \mid (T, \boldsymbol{S} = [0,1]'). \tag{26}$$

[39]Note that if we were to suppress M from the DAG of Table 15, we would obtain the empirical model of Table 3

LACME (25)–(26) adds the the response vector $\boldsymbol{S}$ as an additional conditioning variable to the SI independence relationships in (22)-(23). Assumption (25) is a simple extension of the matching property of $\boldsymbol{S}$ from the IV model of Table 14 to the mediator model of Table 15. Under monotonicity (19), the LACME assumption identifies the direct and indirect mediation effects for compliers.

It is easy to interpret LACME in terms of NR assumptions: assumptions(25)–(26) are a weaker version of SI (22)-(23) 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 places the unobserved response vector $\boldsymbol{S}$ into the mediation model of Table 12. The response vector $\boldsymbol{S}$ plays the role of a balancing score for $\boldsymbol{V}$ only for choice T.[40] The addition of the response vector does not result in any loss of generality. The second DAG displays the mediation model under LACME. According to assumption (26), the response vector $\boldsymbol{S}$ plays the role of a balancing score for T and M. In addition, LACME prevents $\boldsymbol{V}$ from jointly causing M, Y and implies that $\boldsymbol{S}$ directly causes M, Y. It is hard to translate LACME into credible causal relationships.

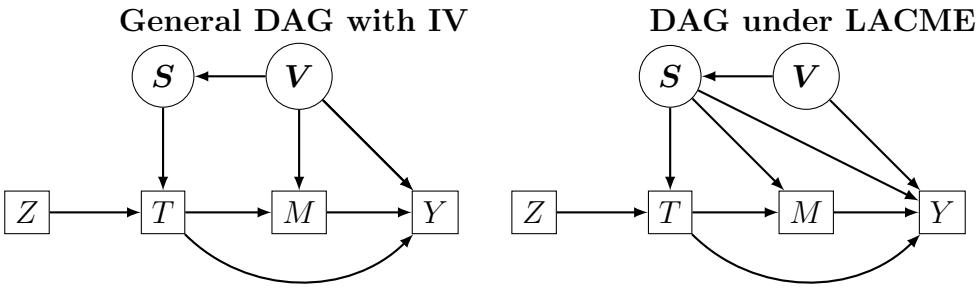
$\boldsymbol{S} = [T(0), T(1)]'$ is expressed as a function of the confounding variable $\boldsymbol{V}$ because $T(z)$ is a function of $\boldsymbol{V}$. Note that the choice T is expressed as a function of $\boldsymbol{S}$ and Z because $T = \big[\mathbf{1}[Z=0], \mathbf{1}[Z=1]\big]\boldsymbol{S}$. The response vector $\boldsymbol{S} = [T(0), T(1)]'$ is expressed as a function of the confounding variable $\boldsymbol{V}$ because $T(z)$ is a function of $\boldsymbol{V}$. The resulting DAG does not include more information than the original model of Table 12 because $\boldsymbol{S}$ is unobserved.

The second DAG displays the mediation model under LACME. From assumption (26), the response vector $\boldsymbol{S}$ plays the role of a matching variable for the causal effect of M on Y. It plays the role of a balancing score for $\boldsymbol{V}$ for T, M, and Y. The assumption prevents $\boldsymbol{V}$

[40] This property is based on the discreteness of the instrument.

from jointly causing M, Y and implies that $\boldsymbol{S}$ directly causes M, Y. It is hard to produce interpretable models that justify $\boldsymbol{S}$ as a cause of M or Y. LACME is an unmotivated but statistically useful assumption.

Table 16: Mediation Model including $\boldsymbol{S}$ and the Mediation Model under LACME Assumption



Using Structural Equations to Identify the Mediation Model with IV

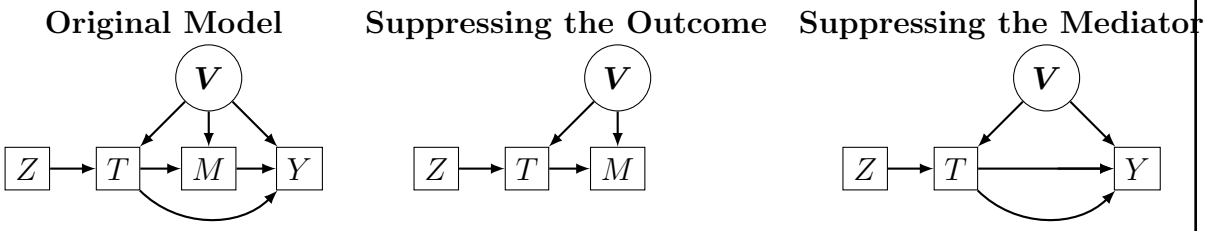
Dippel, Gold, Heblich, and Pinto (2020) study the identification of causal effects for the mediation model with an instrumental variable. Their analysis illustrates the gain in clarity and scrutiny when a causal model is expressed by structural equations instead of NR statistical independence relationships.

A typical empirical setting of an IV model consist of one instrument and various outcomes. A mediation model with an instrument arises when 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.

The second column of Table 17 presents the DAG generated by suppressing the final outcome. The resulting DAG is an IV model like that examined in Section 3. 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.

Table 17: Dissecting the Mediation Model



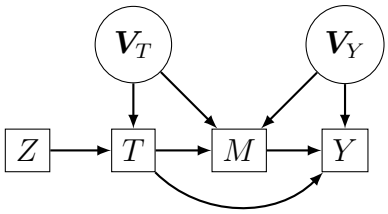
Dippel, Gold, Heblich, and Pinto (2020) address the question of whether it is possible to use an instrumental variable Z to nonparametrically identify the causal chain connecting T, M, Y while maintaining the endogeneity of the treatment T with respect to the mediator M and outcome Y. They show that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

Table 18: Partially Confounded Model with Instrumental Variables

Causal Model

$$\begin{aligned}
\boldsymbol{V}_T &= f_{V_T}(\epsilon_{V_T}) \\
\boldsymbol{V}_Y &= f_{V_Y}(\epsilon_{V_Y}) \\
Z &= f_Z(\epsilon_Z) \\
T &= f_T(Z, \boldsymbol{V}_T, \epsilon_T) \\
M &= f_M(T, \boldsymbol{V}_T, \boldsymbol{V}_Y, \epsilon_M) \\
Y &= f_Y(T, M, \boldsymbol{V}_Y, \epsilon_Y)
\end{aligned}$$

DAG



The partially confounded assumption is that $\boldsymbol{V}_T \perp\!\!\!\perp \boldsymbol{V}_Y$. The assumption generates an additional exogeneity condition $(M(z), Y(m,t)) \perp\!\!\!\perp Z \mid (T = t)$ while maintaining the endogeneity of the treatment T with respect to M and Y. This means that Z is 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. Dippel, Gold, Heblich, and Pinto (2020) discuss the intuition, plausibility, and estimation of the partially confounded mediation model. They illustrate a

range of examples where the partially confounding assumption may hold and where it does not.

6 The Do-Calculus and the Hypothetical Model

This section compares the *do*-calculus (DoC) of Pearl (2009b) with the Neyman-Rubin (NR) framework of Holland (1986); Imbens and Rubin (2015) and the Hypothetical Model (HM) approach 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.[41] In contrast with NR, DoC is based on autonomous structural equations. The method clearly describes the causal relationships between model variables and does not encounter the problematic causal interpretations of the NR approach.

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

The HM and the DoC differ greatly regarding counterfactual manipulations. To address the causal operation of fixing, the HM solution uses a hypothetical model that formalizes the notion of thought experiments and places it on a sound probabilistic footing. Contrary to HM, DoC defines hypothetical models by making manipulations *within* the empirical model. The method implements the notion of setting or fixing using a set of rules that combine graphical analysis, independence relationships and probability equalities.

Some notation is required to explain the method. Let G denote a DAG that represents the original causal model. Let Y, K, X, T denote disjoint variable sets in $\mathcal{T}$. In DoC notation, $T(X)$ denotes the variables in T that do not directly or indirectly cause X. The DoC uses

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

$G_{\overline{K}}$ for the derived DAG that deletes all causal arrows arriving at K in the original DAG G. $G_{\underline{T}}$ denotes the DAG that deletes all causal arrows emerging from T. In this notation, $G_{\overline{K},\underline{T}}$ stands for the derived DAG that suppresses all arrows arriving at K and emerging from T, while $G_{\overline{K,T(X)}}$ deletes all arrows arriving at K in addition to arrows arriving at $T(X)$, namely, arriving at variables in T that are not ancestors of X.

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

The Three DoC Rules

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

The process of checking if a causal effect is identified requires reiterative use of these rules. We present several examples of how to use the DoC method below.

In computer science, the DoC is said to be "complete." This is different from the notion of completeness as defined in simultaneous equations theory discussed in Section 7. The 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).

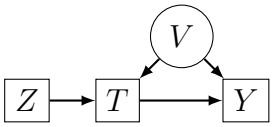
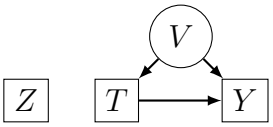
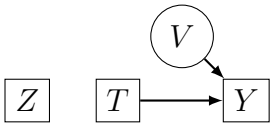
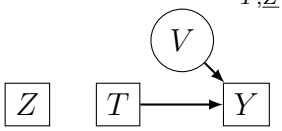
A major limitation of do-calculus is that it only applies to non-parametric models that can be fully characterized by a DAG. Otherwise stated, the method does not account for assumptions about the functional forms of the structural equations or cross covariance restrictions. This limitation hinders the 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 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. By the rules of do-calculus, the IV model and the Roy model are not identified. We now demonstrate these points.

Using Do-Calculus to Investigate the Roy Model

We show the limitations of the DoC for identifying the Roy model.

Table 19: Using Do-Calculus to Investigate the Roy Model

Original DAG G	**Derived DAG** $G_{\underline{Z}}$	**Derived DAG** $G_{\overline{T}}$	**Derived DAG** $G_{\overline{T},\underline{Z}}$
			

The first column of Table 19 presents the DAG of the original Roy model, which is denoted by G. The second column displays the DAG $G_{\underline{Z}}$ which suppresses the arrow arising from Z. The LMC of Z on DAG $G_{\underline{Z}}$ is $Z \perp\!\!\!\perp (Y,T)$. From Rule 2 of DoC, we obtain $P(T \mid do(Z)) = P(T \mid Z)$. Summarizing:

$$G_{\underline{Z}} \Rightarrow T \perp\!\!\!\perp Z, \Rightarrow \text{ by Rule 2 } P\big(T \mid do(Z)\big) = P(T \mid Z). \tag{27}$$

This says that Z is statistically independent of T when we fix Z. In the NR framework, this is the exogeneity condition $T(z) \perp\!\!\!\perp Z$, namely, that the instrument Z is independent of the counterfactual choice $T(z)$. Instrument Z in DAG $G_{\underline{Z}}$ is independent of both T and Y. Thus we can replace T by Y in (27) to obtain $P(Y \mid do(Z)) = P(Y \mid Z)$. This means that conditioning on Z is equivalent to fixing Z. Indeed the instrument Z is an external variable and the causal operation of fixing is translated to standard statistical conditioning.

The third column of Table 19 displays the DAG $G_{\overline{T}}$ which suppresses the arrow arriving at T. LMC of Z on $G_{\overline{T}}$ implies $Z \perp\!\!\!\perp Y$. By Rule 1 of DoC, we have that $P(Y \mid do(T), Z) = P(Y \mid do(T))$. Summarizing:

$$G_{\overline{Z}} \Rightarrow Y \perp\!\!\!\perp Z, \Rightarrow \text{ by Rule 1 } P(Y \mid do(T), Z) = P\big(Y \mid do(T)\big). \tag{28}$$

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

The last column of Table 19 displays the DAG $G_{\overline{T},\underline{Z}}$ which suppresses the arrow arriving at T and arising from Z. Note that the DAGs $G_{\overline{T},\underline{Z}}$ and $G_{\overline{T}}$ are the same. The LMC of Z for $G_{\overline{T}}$ implies $Z \perp\!\!\!\perp Y$. By Rule 1 of DoC, we have that $P(Y \mid do(T), Z) = P(Y \mid do(T))$. In summary:

$$G_{\overline{Z}} \Rightarrow Y \perp\!\!\!\perp Z, \Rightarrow \text{ by Rule 1 } P(Y \mid do(T), Z) = P\big(Y \mid do(T)\big). \tag{29}$$

This means that Z is statistically independent of Y when we fix T. This statement is the exogeneity condition $Y(t) \perp\!\!\!\perp Z$ or the independence relationship $Y \perp\!\!\!\perp Z \mid \tilde{T}$ of the HM framework. The LMC of Z is $Z \perp\!\!\!\perp (T, Y, V)$ which implies that $Z \perp\!\!\!\perp T$ holds. Using Rule 2 of the DoC we obtain:

$$G_{\overline{T},\underline{Z}} \Rightarrow Y \perp\!\!\!\perp Z \mid T, \text{ so Rule 2 } P(Y \mid do(T), do(Z)) = P(Y \mid do(T), Z). \tag{30}$$

Combining $P(Y \mid do(T), Z) = P(Y \mid do(T))$ in (29) with $P(Y \mid do(T), do(Z)) = P(Y \mid do(T), Z)$ in (30) 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 analysis that can be performed using DoC. DoC describes some key properties of the Roy model, but application of its rules alone cannot deliver identification of treatment effects. Unfortunately, the type of assumptions that would secure the identification of treatment effects in the Roy model are ruled out by DoC.

The Front-door Model

To make a more positive statement, 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 (2009b) to illustrate the differences in the approaches.

The Front-Door model (31)–(34) consists of three observed variables T, M, Y and an unobserved confounding variable V. Treatment T causes a mediator M which in turn causes outcome Y. Confounding variable V causes T, Y but not M.[42]

$$V = f_V(\epsilon_V) \tag{31}$$

$$T = f_T(V, \epsilon_T) \tag{32}$$

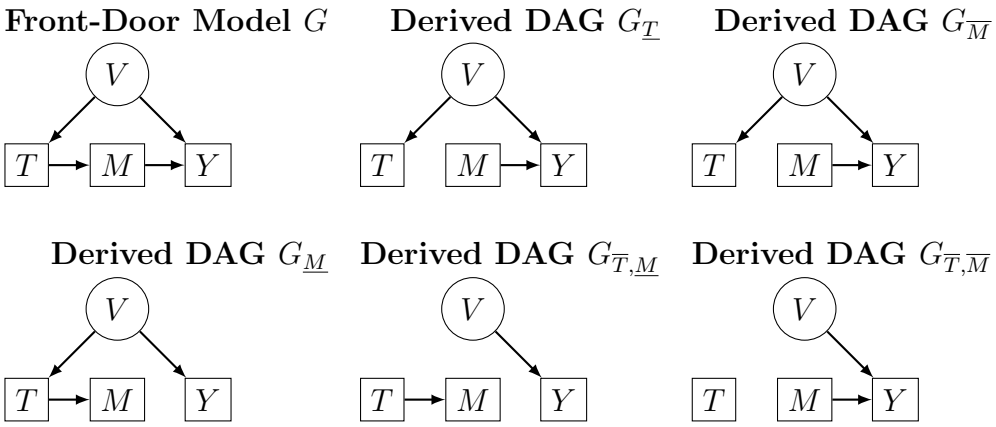
$$M = f_T(M, \epsilon_M) \tag{33}$$

$$Y = f_Y(M, V, \epsilon_Y) \tag{34}$$

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

[42] As before, the error terms $\epsilon_V, \epsilon_T, \epsilon_M, \epsilon_Y$ in the front-door model (31)–(34) are mutually statistically independent.

Table 20: 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 $P_h(Y(t))$. For sake of notational simplicity, suppose that all variables are discrete. The do-calculus is cumbersome. The method requires the five derived DAGs displayed in Table 20. The identification formula of the counterfactual outcome is obtained by the following sequence of steps:

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

$$\therefore P_e(Y \mid do(M)) = \sum_t P_{e^\dagger}(Y \mid T = t, do(M)) P_{e^\dagger}(T = t \mid do(M))$$

$$\text{by Law of Iterated Expectations (L.I.E.)}$$

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

5. $Y \perp\!\!\!\perp M \mid T$ in $G_{\overline{T},\underline{M}}$ holds, thus by Rule 2, $P_{e^\dagger}(Y \mid M, do(T)) = P_{e^\dagger}(Y \mid do(M), do(T))$
6. $Y \perp\!\!\!\perp T \mid M$ in $G_{\overline{T},\overline{M}}$ holds, thus by Rule 3, $P_{e^\dagger}(Y \mid do(T), do(M)) = P_{e^\dagger}(Y \mid do(M))$

7. Collecting these results, we have that $P_{e^\dagger}(Y \mid Z, do(T)) = P_{e^\dagger}(Y \mid do(Z), do(T)) = P_{e^\dagger}(Y \mid do(M))$.

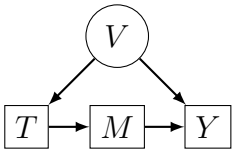
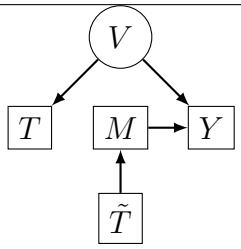
8. Finally, we can use previous results to obtain the following equation:

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

The Front Door Model in the Hypothetical Model Framework

We now investigate the same front-door model using the hypothetical framework. Table 22 displays the hypothetical model associated with the Front-door model (31)–(34) as a DAG. The bottom panel of Table 22 presents the LMC for both models.

Table 21: The Empirical and Hypothetical Front-door Models1

Empirical Model	**Hypothetical Model**
	
LMC	**LMC**
$V \perp\!\!\!\perp -\mid -$	$V \perp\!\!\!\perp (M, \tilde{T})$
$T \perp\!\!\!\perp -\mid V$	$T \perp\!\!\!\perp (M, Y, \tilde{T}) \mid V$
$M \perp\!\!\!\perp V \mid T$	$M \perp\!\!\!\perp (T, V) \mid \tilde{T}$
$Y \perp\!\!\!\perp T \mid (V, M)$	$Y \perp\!\!\!\perp (T, \tilde{T}) \mid (V, M)$
	$\tilde{T} \perp\!\!\!\perp (T, V)$

We seek to identify the counterfactual outcome $P_h(Y \mid \tilde{T} = t)$, i.e., to express $P_h(Y \mid \tilde{T} = t)$ in terms of the observed distribution $P_e(T, M, Y)$. Identification requires us to connect the probability distributions of the hypothetical and the empirical models. To do so we seek independence relationships that contain T and $\tilde{T}$, that is, so that $Y \perp\!\!\!\perp \tilde{T} \mid (M, T)$ and $M \perp\!\!\!\perp T \mid \tilde{T}$ hold.[43] It is also the case $T \perp\!\!\!\perp \tilde{T}$ holds as $\tilde{T}$ is externally specified (exogenous) and does not cause T. We can then apply rules (12)–(13) to generate the following probability equalities:

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

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

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

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

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

$$= \sum_{t', m} P_e\big(Y \mid m, T = t'\big) P_e\big(m \mid T = t\big) P_e\big(T = t'\big) \tag{39}$$

Equation (38) is a sum of probabilities defined in the hypothetical model by to application of the law of iterated expectation over T and M. Equation (39) replaces each of the hypothetical model probabilities with empirical model probabilities using rules (12)-(13).

Understanding the Identification Criteria

The identification of the counterfactual outcomes in the Front-door Model stems from the three independence relationships in (35)–(37). These independence relationships comply with two general properties that facilitate the identification of the counterfactual outcome. We clarify the underlying properties that secure identification.

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

The first property is called *alternate conditionals*. It refers to the fact that the first relationship (35) is an independence relationship regarding T conditional on $\widetilde{T}$. The second relationship (36) is an independence relationship of $\widetilde{T}$ conditional on T. The last relationship (37) cycles back. It is an independence relationship regarding T conditional on $\widetilde{T}$. This property enables us to translate the probabilities of the hypothetical model into the probabilities of the empirical model via the connection rules (12)–(13).

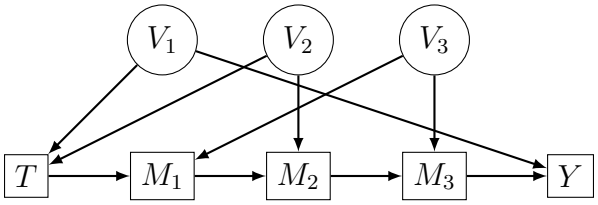
The property of *alternate conditionals* describes an alternating feature to the identification equation (39). The first term of (39) is conditioned on $T = t'$ which refers to the first conditional T in (35). The identification equation (39) sums t' over the support of T. The second term of (39) is conditioned on the treatment value $T = t$. which refers to the second conditional T in (36). The value t remains fixed in the summation as it is the value used to define the counterfactual $(Y \mid \tilde{T} = t)$. The last term in (39) alternates. It is conditioned on $T = t'$ which refers to the last conditional T in (37) and t' varies in the summation.

The second property of the set of independence relationships is called *bridging* and it refers to the variables other than $(T, \widetilde{T})$. The first independence relationship (35) starts with the outcome Y and conditions on the variable M. The second relationship (36) starts with M and conditions on no other variable besides T or $\widetilde{T}$). We say that variable M bridges the path between Y and $(T, \widetilde{T})$, that is, $Y \rightarrowtail M \rightarrowtail (T, \widetilde{T})$. In general terms, *bridging* refers to a sequence of nested sets $\mathcal{T}_1 \subset \cdots \subset \mathcal{T}_K$ of observed variables in $\mathcal{T}$ such that the property of *alternate conditionals* $Y \perp\!\!\!\perp \tilde{T} \mid (T, \mathcal{T}_K)$, $(\mathcal{T}_K \setminus \mathcal{T}_{K-1}) \perp\!\!\!\perp T \mid (\tilde{T}, \mathcal{T}_{K-1}), \ldots$, until $\mathcal{T}_1 \perp\!\!\!\perp T \mid (\tilde{T})$, or $\mathcal{T}_1 \perp\!\!\!\perp \tilde{T} \mid T$ holds. Identification is secured whenever a set of conditional independence relationships among observe variables in the hypothetical model exhibits the alternate conditionals and the bridging properties.

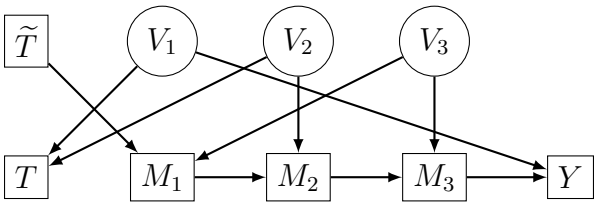
We illustrate these ideas for the complex mediation model of Table 22. The model has three observed mediating variables M_1, M_2, M_3 (instead of M) and three unobserved, confounding variables V_1, V_2, V_3 (instead of V).

Table 22: Using the HM to Identify Counterfactuals

Directed Acyclic Graph of the Empirical Model



Directed Acyclic Graph of the Hypothetical Model



The following conditional independence relationships hold for the hypothetical model:

$$Y \perp\!\!\!\perp \widetilde{T} \mid (T, M_3, M_2, M_1) \tag{40}$$

$$M_3 \perp\!\!\!\perp T \mid (\widetilde{T}, M_2, M_1) \tag{41}$$

$$M_2 \perp\!\!\!\perp \widetilde{T} \mid (T, M_1) \tag{42}$$

$$M_1 \perp\!\!\!\perp T \mid \widetilde{T} \tag{43}$$

$$T \perp\!\!\!\perp \widetilde{T} \mid T \tag{44}$$

The set of independence relationships (40)–(44) is a set of *alternate conditionals*. The first relationship is conditioned on T, the second on $\widetilde{T}$, followed by T and so on.

The bridging property also holds. The right-hand variable of each independence relationship gives the bridging sequence: $Y \rightarrowtail M_3 \rightarrowtail M_2 \rightarrowtail M_1 \rightarrowtail T$. We can define the nested sets $\mathcal{T}_1 = \{M_1\}$, $\mathcal{T}_2 = \{M_1, M_2\}$, $\mathcal{T}_3 = \{M_1, M_2, M_3\}$, to rewritten (40)–(44) as:

$$Y \perp\!\!\!\perp \widetilde{T} \mid (T, \mathcal{T}_3) \tag{45}$$

$$\mathcal{T}_3 \setminus \mathcal{T}_2 \perp\!\!\!\perp T \mid (\widetilde{T}, \mathcal{T}_2, M_1) \tag{46}$$

$$\mathcal{T}_2 \setminus \mathcal{T}_1 \perp\!\!\!\perp \widetilde{T} \mid (T, \mathcal{T}_1) \tag{47}$$

$$\mathcal{T}_1 \perp\!\!\!\perp T \mid \widetilde{T} \tag{48}$$

$$T \perp\!\!\!\perp \widetilde{T} \mid T \tag{49}$$

The law of iterated expectations and independence relationships (40)–(44) enable us to express the counterfactual probability $P_h(Y \mid \widetilde{T})$ as:

$$\textbf{Hypothetical Model} \quad P_h(Y \mid \widetilde{T} = t) = \sum_{t', m_3, m_2, m_1} A_h \cdot B_h \cdot C_h \cdot D_h \cdot E_h,$$

where:

$$A_h = P_h(Y \mid m_3, m_2, m_1, T = t', \widetilde{T} = t)$$

$$B_h = P_h(M_3 = m_3 \mid m_2, m_1, T = t', \widetilde{T} = t)$$

$$C_h = P_h(M_2 = m_2 \mid m_1, T = t', \widetilde{T} = t)$$

$$D_h = P_h(M_1 = m_1 \mid T = t', \widetilde{T} = t)$$

$$E_h = P_h(T = t' \mid \widetilde{T} = t)$$

The connection rules (12)–(13) enable us to translate hypothetical probabilities into empirical probabilities. The identification equation displays the alternative pattern of values t and t' in the same fashion as the identification equation of the Front-door model:

$$\textbf{Empirical Model} \quad P_e(Y(t)) = \sum_{t', m_3, m_2, m_1} A_e \cdot B_e \cdot C_e \cdot D_e \cdot E_e,$$

where:

$$A_e = P_e(Y \mid m_3, m_2, m_1, T = t')$$

$$B_e = P_e(M_3 = m_3 \mid m_2, m_1, T = t)$$

$$C_e = P_e(M_2 = m_2 \mid m_1, T = t')$$

$$D_e = P_e(M_1 = m_1 \mid T = t)$$

$$E_e = P_e(T = t')$$

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 scientific knowledge 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 variable from the joint distribution of variables. All the DoC analysis is 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. They readily analyze both external manipulation and conditioning, such as the treatment on the treated, whereas this is outside the scope of DoC. It facilitates 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 needs to invent 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 employs two statistical implications to connect the probability distributions of the hypothetical and empirical models. HM implications remain within the realm of standard statistical theory and do not require 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 are required to identify counterfactuals grows exponentially. An algorithm has been developed to perform this task.[44] Calculations with HM are simpler than those based on DoC. They rely on a single modification of the original DAG, as encoded in

[44]See Pearl (2009b).

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 to accommodate assumptions that explore functional form restrictions or distributional assumptions outside the scope of DoC.

7 Simultaneous Causality

The Generalized Roy model is usually expressed as a recursive model.[45] However, simultaneous causality is a property of many economic models. Examples of such models include social interactions, general equilibrium, Walrasian market clearing, or simultaneous play in Nash models of industrial organization are staples of economic theory (see, e.g., Mas-Colell et al., 1995). These type of models are ignored in most discussions of causality in the NR literature. The NR approach commonly invokes the Stable Unit Treatment Value Assumption (SUTVA), which excludes the possibility of interaction between agents.[46]

It is instructive to consider these models because they challenge the approximating approaches in the literature, but are easily analyzed in econometric causal policy analysis. The pioneering econometric models featured simultaneity. Many of the core ideas are ignored or remain unknown to the followers of the approximating approaches, which rely on recursive formulations, and are considered as essential features of causal models. In fact, these are at best only convenient assumptions for analyzing causal models, used as special by economists for generations.[47]

[45] See, however, Brock and Durlauf (2007); Heckman (1978).
[46] See, for instance, Imbens and Rubin (2015).
[47] See Strotz and Wold (1960).

Simultaneous causality is an essential feature of structural equation models.[48] 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 system expressed in terms of latent variables (Bollen, 2002) and (b) a measurement system. The measurement system proxies the latent variables. The first part of the structural system consists of structure for person i:

$$\boldsymbol{\eta_i} = \boldsymbol{\alpha_\eta} + \boldsymbol{\beta\eta_i} + \boldsymbol{\Gamma\chi_i} + \boldsymbol{\omega_i} \tag{50}$$

where $\boldsymbol{\eta_i}$, $\boldsymbol{\varepsilon_i}$, $\boldsymbol{\chi_i}$ are vectors of latent variables. The measurement system consists of vectors of measurements:

$$\text{Measurement:} \begin{cases} \boldsymbol{y_i} = \boldsymbol{\alpha_y} + \boldsymbol{\Lambda_y\eta_i} + \boldsymbol{\varepsilon_i} \ \text{ (measurement for } \eta_i) \\ \boldsymbol{\chi_i} = \boldsymbol{\alpha_x} + \boldsymbol{\lambda_x U} = \boldsymbol{\xi_i} \ \text{ (measurement for } \chi_i) \end{cases}$$

These models have been extended to time series and panel data settings (see e.g. Bollen, 1989; Goldberger and Duncan, 1973).

In a valuable paper, Bollen and Pearl (2013) exposit this system of equations as a causal model with simultaneity and show how various measurent 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 restrictions and covariance restrictions to secure identification of causal parameters.

This literature is rich and we lack the space to exposit it thoroughly. We note that these systems illustrate–in linear equation models–an approach for proxying V as previously discussed. It is also an approach for studying mediation where analysts can study how

[48] See Goldberger (1972) and Goldberger and Duncan (1973).

interventions on $\boldsymbol{\chi_i}$ percolate through equation system (43). Schennach (2020) summarizes a large literature on nonparametric factors and proxy models.

Instead of a general exposition of these systems, we consider a simple simultaneous equations model due to Haavelmo (1944). We consider a system of two autonomous causal (structural) equations:

$$Y_1 = g_{Y_1}(Y_2, X_1, U_1, \epsilon_1) \tag{51}$$

$$Y_2 = g_{Y_2}(Y_1, X_2, U_2, \epsilon_2) \quad U_1 \not\perp\!\!\!\perp U_2. \tag{52}$$

We use this system to demonstrate how causality can be analyzed in simultaneous systems.

This system of equations gives two maps: $g_{Y_1} : (Y_2, X_1, U_1) \to Y_2; g_{Y_2} : (Y_1, X_1, U_2) \to Y_2$. Y_1 and Y_2 could be actions of a pair of interacting agents.[49] To simplify the discussion, we assume that both equations are twice continuously differentiable. This is a convenience and not a necessity. The model of equations (51)–(52) are treated in a special way in the DoC approach. We focus on a two equation system to simplify the exposition. Models with multiple simultaneous equations are standard in the literature (see, e.g., Bollen, 1989; Fisher, 1966; Goldberger and Duncan, 1973; Koopmans et al., 1950; Theil, 1958, 1971).

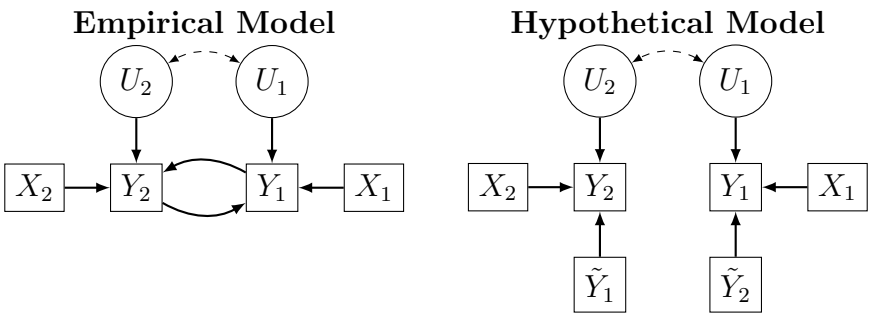
Equations (51) and (52) are assumed to be structural, i.e., invariant under manipulations of their arguments, so they are stable, autonomous maps. Policies consist of manipulations of their arguments.

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. Equations (51) and (52) are generated by thought experiments varying the arguments and tracing out the outcomes. Thus, (51) is the market price that is consistent with hypothetical values Y_2, X_1, U_1. (52) is the analogous relationship for quantity. 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).

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

In terms of our previous notation, the variable set 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 23 given by:

Table 23: Empirical and Hypothetical Causal Models



The LMC condition breaks down so the Bayesian net approach fails. "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):

$$Y_1 = \phi_1(X_1, X_2, U_1, U_2) \tag{53}$$

$$Y_2 = \phi_2(X_1, X_2, U_1, U_2). \tag{54}$$

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 (51) and (52) is not guaranteed. Nonetheless autonomous correspondences may still exist and they can be used to make set-valued causal inferences.[50]

[50] See, e.g., Heckman (1978); Quandt (1988); Tamer (2003).

The causal effect 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 effect 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 (51) and (52) 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 (51) and (52). These involve hypothetical variations. For certain models no such sets of variables may exist.

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 (2009a) preserves 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, equation (52) is suspended, but (51) 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 (51). This thought experiment converts a simultaneous system into a recursive system with all other equations assumed to hold as before.

This approach is cumbersome and strains credibility in many interlinked economic contents (e.g., person 1 influences 2, but not vice versa) but is logically possible. It is unnecessary if exclusions in (51) and (52) are used. To show this, we define exclusion of X_2 in (51) as $\frac{\partial g_{Y_1}}{\partial X_2} = 0$ for all (Y_2, X_1, X_2, U_1).[51] Exclusion of X_1 in (52) is defined as $\frac{\partial g_{Y_2}}{\partial X_1} = 0$ for all (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 (52), by the chain rule, the causal effect of Y_2 on

[51] Or more generally, X_2 is not an argument of gY_N.

Y_1 is

$$\frac{\partial g_{Y_1}}{\partial Y_2} = \frac{\partial Y_1}{\partial X_2} \Big/ \frac{\partial Y_2}{\partial X_2} = \frac{\partial \varphi_1}{\partial X_2} \Big/ \frac{\partial \varphi_2}{\partial X_2}.$$

We may define and identify 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.[52] No implausible "shutting down" of any equation in a system and assuming autonomy 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 (51) and (52) 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 forms (53) and (54) estimate the **net effect** of a policy change X_1:

$$\frac{\partial Y_1}{\partial X_1} = \frac{\partial \phi_1(X_1, X_2, U_1, U_2)}{\partial X_1}. \tag{55}$$

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 (51) and its indirect effect through Y_2:

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

[52] Assuming that the completeness condition is part of the thought experiment. In some contexts it may be ruled out as not credible.

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.[53]

Mediation 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).

8 Conclusion

This paper presents the basic framework of the econometric model for causal policy analysis. We discuss the definition of causal parameters and approaches to their identification within it. We consider two approximations to it that are current in the 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 adequately characterize the lessons of economic theory and allow for testing it, for synthesizing evidence on it from multiple sources, constructing credible policy counterfactuals, including forecasting policy impacts in new environments and forecasting the likely impacts of policies never previously implemented. The econometric approach delineates the definition of causal parameters and their identification as two separate tasks.

The two approximating approaches are: (a) the Neyman-Rubin approach rooted in the statistics of experiments, and (b) the do-calculus that originated in computer science. Both arc recent developments that attempt to address some of the same problems tackled by the econometric approach. Each has important, but different, limitations. Neither has the flexibility or clarity of the econometric approach.

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

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

The do-calculus 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. Its rigid rules preclude the use of many traditional techniques of identification and estimation.

The Neyman-Rubin 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. 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 or to analyze well-posed economic questions with it using the large toolkit of modern econometrics.

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

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

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

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

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

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

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

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

Bareinboim, E. and J. Pearl (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences 113*(27), 7345–7352.

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

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

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

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psych 53*(1), 605–634.

Bollen, K. A. and J. Pearl (2013). Eight myths about causality and structural equation models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research*, Chapter 15, pp. 301–328. Springer, Dordrecht.

Brock, W. A. and S. N. Durlauf (2007). Identification of binary choice models with social interactions. *Journal of Econometrics 140*(1), 52–75. Analysis of spatially dependent data.

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

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

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

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

Chatfield, C. (2000). *Time-Series Forecasting*. CRC Press.

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

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

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

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

Dippel, C., R. Gold, S. Heblich, and R. Pinto (2020). Mediation analysis in iv settings with a single instrument. *Unpublished Manuscript*.

Durlauf, S., L. P. Hansen, J. J. Heckman, and R. L. Matzkin (2020). *Handbook of Econometrics*. Elsevier.

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

Fisher, F. M. (1966). *The Identification Problem in Econometrics*. New York: McGraw-Hill.

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

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

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

Frisch, R. (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.

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

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

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

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

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

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. (Forthcoming, 2021). Econometrics. Princeton University Press.

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

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

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

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

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

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

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

Heckman, J. J. and E. E. Leamer (2001). *Handbook of Econometrics*, Volume 5 of *Handbooks in Economics*. Amsterdam: North Holland.

Heckman, J. J. and E. E. Leamer (2007). *Handbook of Econometrics*, Volume 6AB of *Handbooks in Economics*. Amsterdam: North Holland.

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

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

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

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

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

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

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

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

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

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

Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs, and to forecast their effects in new environments. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 71, pp. 4875–5143. Amsterdam: Elsevier B. V.

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

Holland, P. W. (1997). Some reflections on Freedmans critiques. In *Topics in the Foundation of Statistics*, pp. 50–57. Springer.

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

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

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

Imai, K., L. Keele, D. Tingley, and T. Yamamoto (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review 105*, 765–789.

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

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

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

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

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

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

Knight, F. (1921). *Risk, Uncertainty and Profit*. New York: Houghton Mifflin Company.

Koopmans, T. C., H. Rubin, and R. B. Leipnik (1950). Measuring the equation systems of dynamic economics. In T. C. Koopmans (Ed.), *Statistical Inference in Dynamic Economic Models*, Number 10 in Cowles Commission Monograph, Chapter 2, pp. 53–237. New York: John Wiley & Sons.

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

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

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

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

Manski, C. F. (1993, July). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies 60*(3), 531–542.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Olley, G. S. and A. Pakes (1996, November). 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*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

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

Pearl, J. (2009a). Causal inference in statistics: An overview. *Statistics Surveys 3*, 96–146.

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

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., D. Jazzing, and B. Schölkopf (2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge, MA: MIT Press.

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

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

Quandt, R. E. (1958, December). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association 53*(284), 873–880.

Quandt, R. E. (1988). *The Econometrics of Disequilibrium*. New York: Blackwell.

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

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

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

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

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

Rubin, D. B. (1978, January). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics 6*(1), 34–58.

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

Shpitser, I. and J. Pearl (2006, November). Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference, AAAI-06/IAAI-06*, Proceedings of the National Conference on Artificial Intelligence, pp. 1219–1226. 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference, AAAI-06/IAAI-06 ; Conference date: 16-07-2006 Through 20-07-2006.

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

Spiegelhalter, D. J., A. Dawid, S. L. Lauritzen, and R. G. Cowell (1993). Bayesian analysis in expert systems. *Statistical Science*, 219–247.

Strotz, R. H. and H. O. A. Wold (1960, April). Recursive vs. nonrecursive systems: An attempt at synthesis (part i of a triptych on causal chain systems). *Econometrica 28*(2), 417–427.

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

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

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

Theil, H. (1958). *Economic Forecasts and Policy*. Number 15 in Contributions to Economic Analysis. Amsterdam: North-Holland Publishing Company.

Theil, H. (1971). *Principles of Econometrics*. New York: Wiley.

Tinbergen, J. (1930, October). Bestimmung und deutung von angebotskurven ein beispiel. *Zeitschrift für Nationalökonomie 1*(5), 669–679.

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

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

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research 20*, 557–585.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics 5*(3), 161–215.

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