Causality and Econometrics:

Part I

James Heckman and Rodrigo Pinto

Econ 312, Spring 2022
This draft, April 26, 2022 5:32pm

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.

- 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.

- We consider two causal approaches often advocated by statisticians and computer scientists.
- 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.
- 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.
- 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 lecture 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.

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.

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

- 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.

[3] Multiple equilibria are produced in many econometric models. See, e.g., Mas-Colell et al. (1995).

[4] Marshall (1961)

- 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 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:

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

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

- 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., 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]

[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).

- 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 1

- Consider four different possible causal models – all thought experiments:

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

- In the first causal model, T does not cause U, nor does U cause T.
- Parameter β is the causal effect of varying T on Y for a fixed value of U.
- Variables T and U are statistically independent and the parameter β can be consistently estimated by OLS.

- In the second causal model, T does not cause U, nor does U cause T.
- Parameter β is still the causal effect of T on Y. However, T and U are not statistically independent because they share a common confounding variable ϵ_V and the OLS estimator of β is biased.
- This model is sometimes called a 'common cause" model with ϵ_V being 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]

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

- 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.

- 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.

- 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]

[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.

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 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.

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

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, 1958; Hamilton, 2000; Chatfield, 2000). However, the truly ambitious problem solved by policy analysts is **P4**.

[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).

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]

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

- 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.

[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."*

- The econometric approach distinguishes three tasks of econometric causal policy analysis that are often conflated in the approximating literatures:

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

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.

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.

- 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.
- 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.
- 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, 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.

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

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.

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

- 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 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 $Y(t); t \in \{0, 1\}$.

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

- The early Generalized Roy model has been generalized and extended in many ways.[18]
- 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).

[18] For instance, Heckman and Vytlacil (2007a) investigate multiple variations of the original model, Heckman, Urzúa, and Vytlavil (Heckman et al.) 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.

- 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:

$$V = f_V(\epsilon_V), \tag{4}$$
$$Z = f_Z(\epsilon_Z), \tag{5}$$
$$T = f_T(Z, V, \epsilon_T), \tag{6}$$
$$Y = f_Y(T, V, \epsilon_Y). \tag{7}$$

- 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.
- 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.[19]

[19] See Lauritzen (1996).

- 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, $\mathbb{Ch}(K) = \{W \in \mathcal{T} \text{ such that } K \in \mathbb{M}(W)\}$.

- 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 $K \in \mathcal{T}$.

- 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 on its parents.
- 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: $X \perp\!\!\!\perp Y \mid Z \Rightarrow Y \perp\!\!\!\perp$

(B) Decomposition: $X \perp\!\!\!\perp (W, Y) \mid Z \Rightarrow X \perp\!\!\!\perp$

(C) Weak Union: $X \perp\!\!\!\perp (W, Y) \mid Z \Rightarrow X \perp\!\!\!\perp$

(D) Contraction: $X \perp\!\!\!\perp W \mid (Y, Z)$ and $X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp$

(E) Intersection: $X \perp\!\!\!\perp W \mid (Y, Z)$ and $X \perp\!\!\!\perp Y \mid (W, Z) \Rightarrow X \perp\!\!\!\perp$

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

- 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.

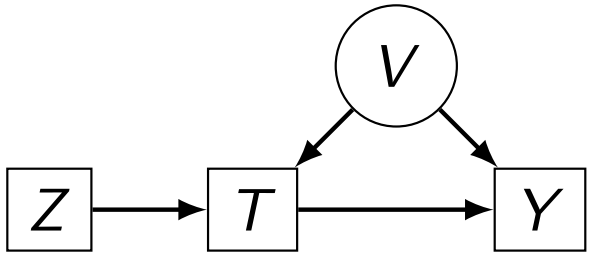
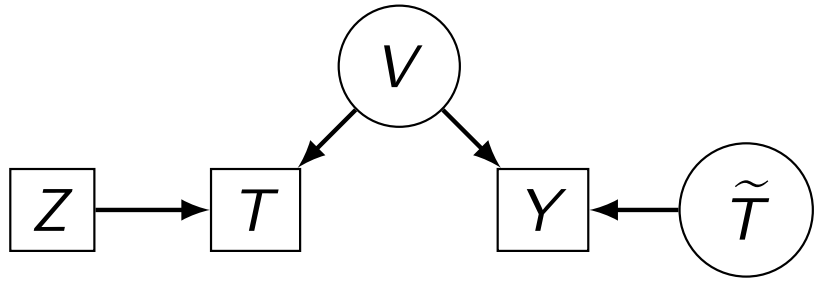
Formalizing Frisch's Insight

- 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.

- 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 $\widetilde{T}$ 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 $\widetilde{T}$ which replaces the T-input of variables directly caused by T.

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

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

- Notationally, $\mathcal{T}_h = \mathcal{T}_e \cup \{\widetilde{T}\}$ such that $\mathbb{M}_h(\widetilde{T}) = \varnothing$ and for each $K \in \mathcal{T}$ we have that $\mathbb{M}_h(K) = \{\widetilde{T}\} \cup \{\mathbb{M}_e(K) \setminus \{T\}\}$ if $K \in \mathbb{Ch}_e(T)$ and $\mathbb{M}_h(K) = \mathbb{M}_e(K)$ otherwise.
- Table 4 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.

- The hypothetical variable $\widetilde{T}$ is external.
- It has no parents.
- According to (8), the hypothetical variable $\widetilde{T}$ is independent of all its non-descendants, and, in particular, $\widetilde{T} \perp\!\!\!\perp T$ 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 $\widetilde{T} \to Y$ link.

- *Counterfactuals* are generated by hypothetical (external) manipulations of treatments.
- These are produced in the hypothetical model by *conditioning* on the hypothetical variable $\widetilde{T}$.
- For instance, the distribution of the counterfactual outcome Y when the treatment is externally set to a value $t \in \text{supp}(T)$ is $P_h(Y \mid \widetilde{T} = t)$ and the counterfactual outcome mean is given by $E_h(Y \mid \widetilde{T} = t)$.
- 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 $ATE = E_h(Y \mid \widetilde{T} = 1) - E_h(Y \mid \widetilde{T} = 0)$.
- 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
$TOT = E_h(Y \mid \widetilde{T} = 1, T = 1) - E_h(Y \mid \widetilde{T} = 0, T = 1)$.

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 $Y(t) = f_Y(t, V, \epsilon_Y)$.

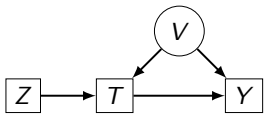
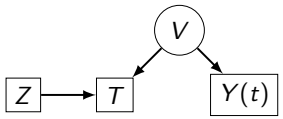
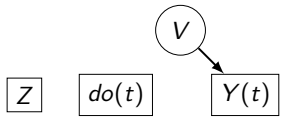
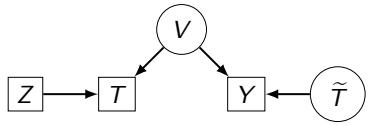
- There is no direct empirical counterpart to this concept without further analysis.
- Fixing does not eliminate the structural equation for treatment variable T.
- It only modifies the outcome equation by replacing the random variable T by a fixed treatment value $t \in \text{supp}(T)$.
- Thus the variable T is still present in the causal model when fixing is applied.

- The do-operator of Pearl (2009, 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.

- The following table 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.

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

- 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:

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

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

- 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 (2009) 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 $\widetilde{T}$ and any variable Y that is a descendant of $\widetilde{T}$ in the hypothetical model, we have that:

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

where $\left(Y \mid \widetilde{T}=t, K\right)_{\mathbb{M}_h}$ denotes the variable Y conditional on K and on the event $\widetilde{T}=t$ in the hypothetical model, $\left(Y(t) \mid K\right)_{\mathbb{M}_{e^*}}$ and $\left(Y(t) \mid K\right)_{\mathbb{M}_{e^\dagger}}$ denote the counterfactual outcome under fixing and doing respectively.

Identification of Counterfactual Outcomes

- 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:[20]

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

- 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 $\widetilde{T}$.
- 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 $\widetilde{T}$.

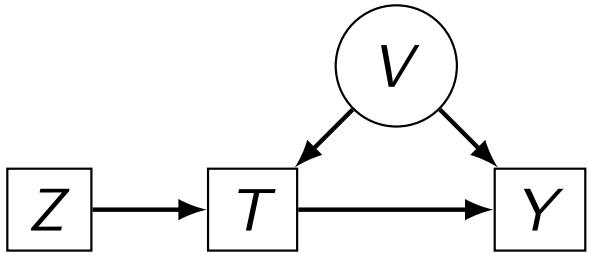
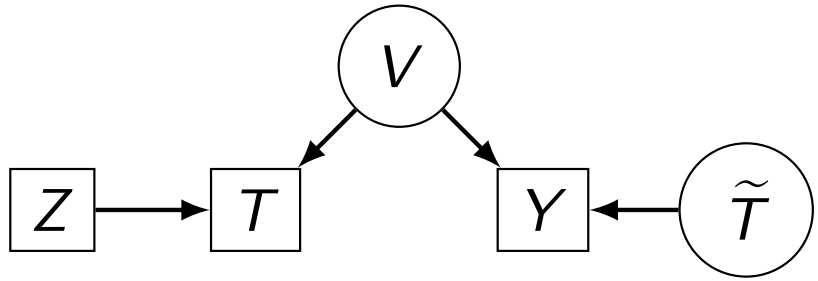
- Equations (12)-(13) are useful for describing the intuitive properties of the hypothetical model.
- Since the hypothetical variable $\widetilde{T}$ is externally specified and independent of all its non-descendants, which include the treatment T, $K \perp\!\!\!\perp \widetilde{T} \mid T$ holds for any variable K not caused by $\widetilde{T}$.
- According to (13), we have that for $P_h(K \mid T \in \mathcal{A}') = P_e(K \mid T \in \mathcal{A}')$ and 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 4.
- The LMC of Y generates the independence relationship $Y \perp\!\!\!\perp T \mid (\widetilde{T}, V)$.
- Variable V is a matching variable. Conditioning on it generates the useful relation:

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

- 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).

Table 4: Generalized Roy Model: Empirical and Hypothetical Causal Models (Repeated)

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

- If $Y \perp\!\!\!\perp T \mid (\widetilde{T}, V)$ 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 $\widetilde{T}$, which implies that $P_h(V) = P_e(V)$.
- Thus if V were observed, the probability distribution of the counterfactual $P_h(Y \mid \widetilde{T} = t)$ 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.

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.[21]

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

- 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 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, $U \sim \textit{unif}([0,1])$.
- 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.

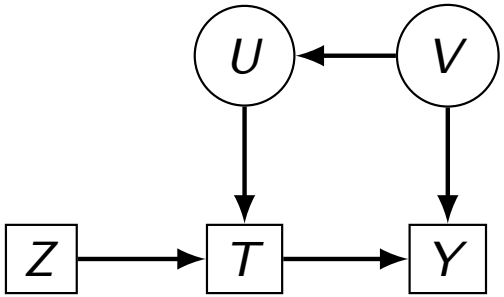
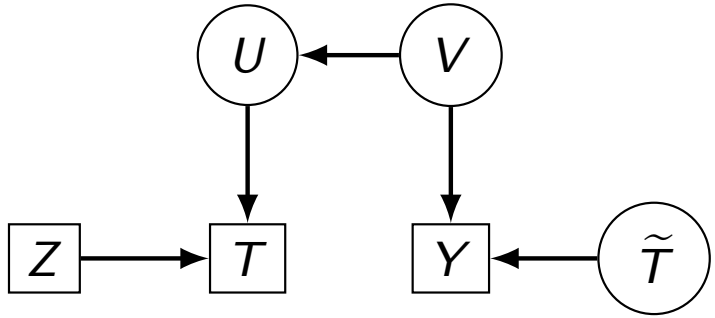
- The hypothetical and empirical models for the Generalized Roy model that include the unobserved variable U are displayed in Table 5.
- The LMC of T in the hypothetical Roy model of Table 5 implies that $Y \perp\!\!\!\perp T \mid (Z, \widetilde{T}, U)$.
- The LMC of Z implies $Y \perp\!\!\!\perp Z \mid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

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

- 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
  $Y \perp\!\!\!\perp T \mid (\widetilde{T}, V) \Rightarrow Y \perp\!\!\!\perp T \mid (\widetilde{T}, U)$.[22]

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

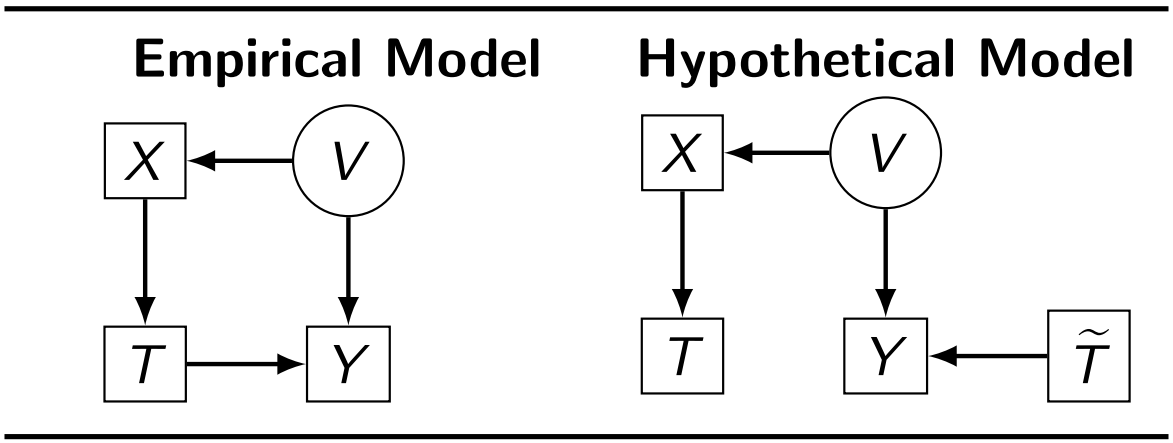
The Matching Assumption

- 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*.[23]
- Another (structural) way to state this is that X spans the space of V.

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

- Table 6 presents the empirical and the hypothetical models that justify the matching assumption.
- The LMC of T in the hypothetical model implies that $Y \perp\!\!\!\perp T \mid (\widetilde{T}, X)$.
- According to (13), we have that $P_h(Y \mid \widetilde{T} = t, X) = P_{e^*}(Y(t) \mid X) = P_e(Y \mid T = t, X)$ 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 $\widetilde{T}$, thus $P_h(X) = P_e(X)$ and the probability distribution of the counterfactual outcome is given by $P_{e^*}(Y(t)) = \int (P_e(Y \mid T = t, X = x) dF_{e,X}(x)$.

Table 6: Matching Model: Empirical and Hypothetical Causal Models



- 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,

$$ATE = \int \Big(E_e(Y \mid T = 1, X = x) - E_e(Y \mid T = 0, X = x) \Big) dF_{e,X}(x).^{24}$$

[24] 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.

- 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.

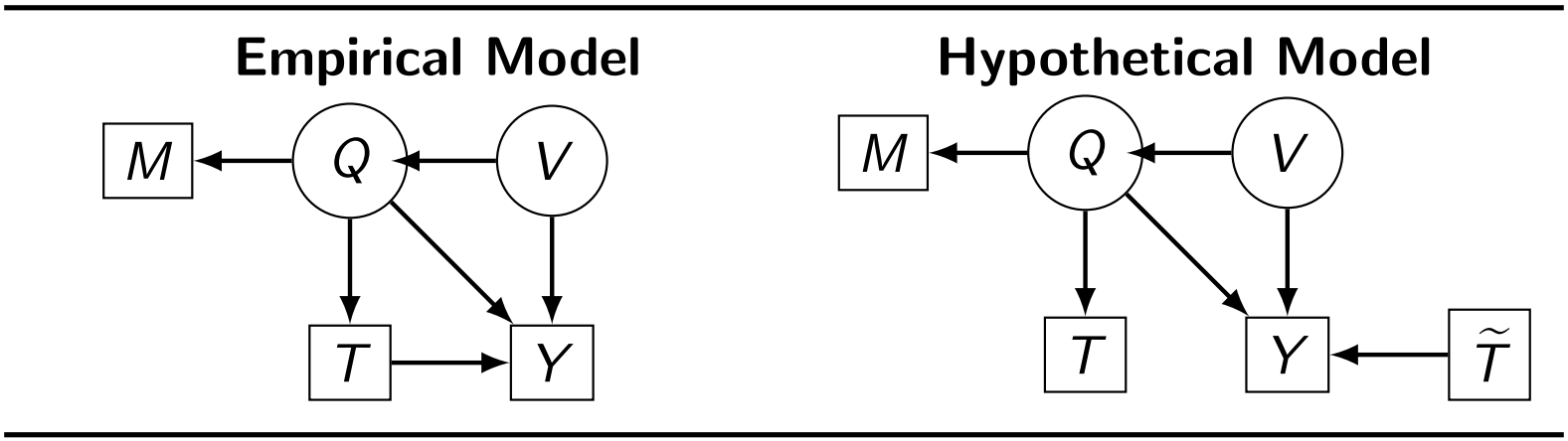
Matching on Proxied Unobservables

- 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, $Y \perp\!\!\!\perp T \mid (\widetilde{T}, Q)$, and its respective counterpart in the empirical model, $Y(t) \perp\!\!\!\perp T \mid Q$, hold. Variable Q has two additional properties: (1) it may cause outcome Y; and (2) it may be measured with error by the observed variable M.

- A common setup where Q arises is in the evaluation of college 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 $Y \perp\!\!\!\perp T \mid (\widetilde{T}, K)$.
- 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:

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

where the first term is observed, the second term uses fixing and the last one uses the hypothetical model.

- Under separability and outcome additivity, we can express $E_h(Y(1) \mid \widetilde{T} = 1, Z, T = 1)$ as:

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

where

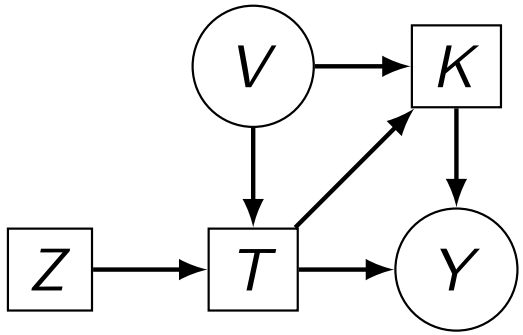
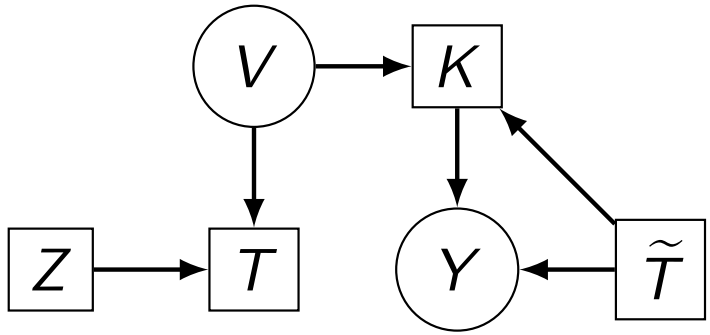
$$
f_1\big(P(Z)\big) = E_h\big(\psi\big(f_K(1, V, \epsilon_K)\big) \mid Z, T = 1\big)
$$

where the first equality uses the additivity assumption, the second uses the fact the $\widetilde{T}$ is an external variable, the third uses the separability assumption, the fourth switches the hypothetical model into the empirical model as V, ϵ_K, Z are non-descendants of $\widetilde{T}$.

- The last equation gives the expectation $E_h(Y \mid \widetilde{T} = 1, Z, T = 1)$ as a function of the propensity score $P(Z)$.
- Control function $f_1(P(Z))$ 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

Empirical Model	**Hypothetical Model**
	

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.

Panel data Analysis and Other Approaches

- 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

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

	Method Assumes	Need Instrument (Z)?	Identify Distribution of V?
Matching[a]	V, X known	No	Yes (V observed)
Control Functions[b]	V estimated, X, Z known (continuous T); Bounds on quantiles of V estimated (discrete case)	Yes	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)