Econometric Approach to Causality

James J. Heckman and Rodrigo Pinto

Econ 312, Spring 2023
This draft, March 14, 2023 7:41pm

THE UNIVERSITY OF CHICAGO

- 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.

Theme: Causal Analysis Poses a Challenge to Conventional Statistics

- A recurrent question is how to interpret regression coefficients.
- In econometrics, we model:

$$\textbf{Outcomes: } Y = f(X, V, W) \tag{1}$$

  - X: observed
  - V: the latent variable connecting equations (factor)
  - W: errors uncorrelated or independent across equations

$$\textbf{Measurements: } M = \mu(V, \varepsilon) \tag{2}$$

- W, V, ε: mutually independent
- Is X a causal variable?
- Is V a causal variable?
- **What is a causal variable?**

Causality Based on a *Thought* Experiment

- Consider relationships mapping inputs X to outcomes Y using a stable map g:

$$g : X \to Y \quad \text{over the domain of } X \quad \big(Dom(X)\big). \tag{3}$$

- A map is **stable** if changing its arguments over the domain of X preserves the map.
- $Y = g(X)$, where g may be a multi-valued correspondence.

- Elementary version of (3) is:

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

- α and β don't change when X or a component of it is changed.
- **invariance** or **autonomy** of relationships (Frisch, 1938).

To Define Causality

- More than stability of maps is required.
- **Directionality** is central.

- The range of Y is a set of ***potential outcomes*** associated with X over its domain. g may be a function or a correspondence.
- Comparisons of Y for different values of X—all other factors the same—are defined as ***causal effects***.

Regression: Conditional Expectation or Thought Experiment?

Regression: Two Interpretations
(Source of Great Confusion See in Statistics: Pratt & Schlaifer, 1984)

$$Y = X\beta + U \tag{5}$$

Conditional Expectation

- Major source of confusion: (5) defined by statisticians as **conditional expectation** for $Y = X\beta + U$.

$$E(Y \mid X) = X\beta \text{ if } E(U \mid X) = 0.$$

- $E(Y \mid X) = X\beta + E(U \mid X)$ if $U \not\perp\!\!\!\perp X$.

Thought Experiments

- Interpret $Y = X\beta + U$: hypothetically vary X and U: $(U) \to Y$ via $Y = X\beta + U$.
- Not a statistical operation.
- Lies outside standard statistics.

Review: All Causes Model

- Econometricians since Haavelmo investigate the “all causes” model.
- Outcomes are generated by a deterministic mapping of observed and unobserved inputs to outputs:

$$Y = g(\underset{\text{observed}}{\underset{\downarrow}{X}}, \underset{\text{unobserved}}{\underset{\downarrow}{U}})$$

- Mapping g: $(X, U) \rightarrow Y$

THE UNIVERSITY OF CHICAGO

Example: Haavelmo, 1943, All Causes Framework

- Early work used linear models.

$$\underset{\substack{\uparrow \\ \text{outcome}}}{Y} = \underset{\substack{\uparrow \\ \text{observed} \\ \text{by} \\ \text{analyst}}}{\overset{\substack{\text{cause} \\ \downarrow}}{X}} \beta + \underset{\substack{\uparrow \\ \text{unobserved by} \\ \text{analyst}}}{\overset{\substack{\text{cause} \\ \downarrow}}{U}} \tag{6}$$

- $E(U \mid X)$ not necessarily zero.
- Distinguishing feature of the econometric approach is explicit modeling of unobservables that drive outcomes and influence inputs (X).

Fixing vs. Conditioning (Haavelmo, 1943)

- $E(Y \mid X = x)$ conditioning on $X = x$.
- $E(Y \mid X = x) = x\beta + E(U \mid X = x)$ under linearity.
- Fixing X at level $X = x$.
- X is *hypothetically manipulated* to take value x means fixing X at different levels is a **hypothetical manipulation that does not change the** U.
- $E(Y \mid X = x, U = u)$
- A **mental construct** since U not observed.
- In this thought experiment, analyst (not nature) hypothetically sets variables (X, U) to (x, u).
- Without a proper measure on Y, X, U, this is *not* a well defined statistical object.
- Our approach makes it well defined, but that is not a trivial task.

Traditional Approach That Links to Structural Models: Decomposing Unobserved Confounders

- Marschak and Andrews (1944) decompose the unobservable:

$$U = \phi \underset{\substack{\uparrow \\ \text{Source of} \\ \text{Confounding} \\ \text{("Factors in SEM")}}}{V} + \mathcal{E} \tag{7}$$

- $V \not\!\perp\!\!\!\perp X$ so $U \not\!\perp\!\!\!\perp X$
- $\mathcal{E} \perp\!\!\!\perp (V, X)$.
- $E(Y \mid X) = X\beta + \phi E(V \mid X)$.
- All estimators for causal models control for the effects of V (implicitly or explicitly).
- Factor models: measurements $M = \mu(V, \varepsilon)$ can be used to control for V.

- Consider four different possible causal models—all thought experiments: ϵ_X, ϵ_U, ϵ_V, ϵ_X all mutually independent.

Causal Model 1

$X = f_X(\epsilon_X)$
$U = f_U(\epsilon_U)$
$Y = X\beta + U$

X does not cause U, nor does U cause X.

Causal Model 2

$X = f_X(\epsilon_X, \epsilon_V)$
$U = f_U(\epsilon_U, \epsilon_V)$
$Y = X\beta + U$

X does not cause U, nor does U cause X: β is still the causal effect of X on Y. X and U are not statistically independent.

Causal Model 3

$X = f_X(\epsilon_X, U)$
$U = f_U(\epsilon_U)$
$Y = X\beta + U$

Second and third models are statistically identical in the sense that X and U are not statistically independent and the OLS estimator is biased. Model imposes a restriction on the variation in U.

Causal Model 4

$X = f_X(\epsilon_X)$
$U = f_U(\epsilon_U, X)$
$Y = X\beta + U$

Fourth model, X causes U and the OLS estimator of the parameter β does not, in genera identify the causal effect of X on because X also affects U.

- OLS estimator of β captures both direct and indirect effects of X on Y.
- Let $Y(X) = X\beta + U$: counterfactual outcome.
- Using the standard regression model as a starting point in econometrics confuses the logic of causal thought processes.

Frisch: "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

- **Separate models from estimands**
- **Structural equation models do just that**

Econometric Approach to Causality

- Developed to address questions that arise in addressing policy problems.
- Develops explicit models where economic agents choose inputs that cause outcomes.
- Investigates the mechanisms governing the choice of inputs.
- "Treatments" are inputs.
- Identification/estimation/interpretation of causal effects depends on the careful accounting of the unobserved variables that cause both input choice and outcomes.
- Structural models do that.
- Caricatures in the literature that choices of inputs by the agents analyzed involve highly stylized rational choice models or perfect information false.

Econometric Approach to Causality

Four Distinct Policy Questions

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? "Internal Validity"*

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

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. "External Validity"*

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.*

Table 1: **Three Distinct Tasks Arising in the Analysis of Causal Models**

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

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;

Table 2: Problems Addressed by Econometrics, Cont'd

(f) Accommodate multiple approaches to identification beyond randomization instrumental variables and matching, approaches 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).

Defining Causal Models

- **Causal Model:** Defined by **four** components:
  1. **Random Variables**: $\mathcal{T} = \{Y, U, X, V\}$.
  2. **Error Terms** Mutually independent: $\omega_Y, \omega_U, \omega_X, \omega_V$.
  3. **Structural Equations**: Deterministic causal relationships
  4. They are autonomous : f_Y, f_U, f_X, f_V.
- **Autonomy**: Deterministic functions invariant to changes in their arguments (Frisch, 1938).
- Different ways of arriving at the values of these arguments don't affect outputs of outcome equations.
- Also known as **"structural"** (Hurwicz, 1962).
- Keep in mind the multiple meanings of "structural" in the various literatures.

Structural Relationships

$$
\begin{aligned}
Y &= f_Y(X, U, \omega_Y), && Y \text{ observed} \\
X &= f_X(V, \omega_X), && X \text{ observed} \\
U &= f_U(V, \omega_U), && U \text{ unobserved} \\
V &= f_V(\omega_V), && V \text{ unobserved}
\end{aligned}
$$

- $Y(x) = f_Y(x, U, \omega_Y)$ is a **potential outcome** where X is fixed at x.
- All potential outcomes are outputs of structural equations.
- Can augment equation for X, for example:

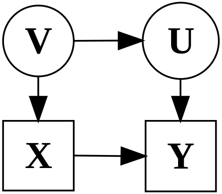
$$X = f_X(Z, V, \omega_X)$$

  where $Z \perp\!\!\!\perp (U, V, \omega_Y, \omega_U, \omega_X, \omega_V)$; Z shifts X.
- Z only affects Y through its impact on X.

Some Questions

- What statistical relationships are generated by this (or any) causal model?
- Is there an equivalence between statistical relationships and causal relationships?

Useful Tool:
Local Markov Condition (LMC):
A variable is independent of its non-descendants conditional on its parents



- For example: $Y \perp\!\!\!\perp \underbrace{V}_{\text{non-descendants}} \mid (\underbrace{X, U}_{\text{parents}})$

Link to Local Markov Condition & Graphoid Axioms

Challenges Posed to Conventional Statistics by Causal Analysis

- **Fixing:** Causal operation sets X-inputs of structural equations to x.

Data Generating Process	**Model Under Fixing** X **Fixed at x**
$V = f_V(\omega_V)$ $U = f_U(V, \omega_U)$ $X = f_X(V, \omega_X)$ $Y = f_Y(X, U, \omega_Y)$	$V = f_V(\omega_V)$ $U = f_U(V, \omega_U)$ $\boldsymbol{X = x}$ $Y = f_Y(\boldsymbol{x}, U, \omega_Y)$

Restating Preceding Analysis in This Notation: Fixing $\neq$ Conditioning

Conditioning: *Statistical* exercise that considers the dependence structure of the data generating process.

$$Y \text{ Conditioned on } X = x: E(Y \mid X = x) = E(f_Y(X, U, \omega_Y) \mid X = x)$$
$$\text{Linear Case: } E(Y \mid X = x) = x\beta + E(U \mid X = x)$$

Fixing: *causal* exercise that *hypothetically* assigns values to inputs of the autonomous equation we analyze.

$$Y \text{ when } X \text{ is fixed at } x \Rightarrow Y(x) = f_Y(x, U, \omega_Y)$$

$$\text{Linear Case: } E(Y(x)) = x\beta + E(U); E(\omega_Y) = 0.$$

Average Causal Effects for Fixed *x*

X is *fixed* at x, x' :

$$ATE = \mathrm{E}(Y(x)) - \mathrm{E}(Y(x'))$$

Another Way to Understand the Challenge: Consider Joint Distributions

1. **Model Representation under Fixing:**

$$Y = f_Y(x, U, \omega_Y); X = x; U = f_U(V, \omega_U); V = f_V(\omega_V).$$

2. **Standard Joint Distribution Factorization:**

$$\begin{aligned} P(Y, V, U \mid X = x) &= P(Y \mid U, V, X = x) P(U \mid V, X = x) P(V \mid X = x) \\ &= P(Y \mid U, V, X = x) P(U \mid V) \mathbf{P}(\mathbf{V} \mid \mathbf{X} = \mathbf{x}) \\ &\text{because } U \perp\!\!\!\perp X \mid V \text{ by LMC.} \end{aligned}$$

3. **Factorization under Fixing X at x:**

$$P(Y, V, U \mid X \text{ fixed at } x) = P(Y \mid U, V, X \text{ fixed at } x) P(U \mid V) \mathbf{P}(\mathbf{V}).$$

- **Conditioning** on X: Includes relationship of X with the distribution of V.
- **Fixing**: X does **not** affect the distribution of V.

Fixing Cannot Be Defined by Standard Probability Theory

- Fixing is a **causal operator**, not a statistical operator.
- Fixing does not affect the distribution of parent (predecessor) variables.
- Conditioning is a statistical operator.
- It embraces the distribution of **all** variables.
- Fixing has a causal direction.
- Conditioning has no direction.
- $\therefore$ researchers accustomed to reasoning in terms of conditional probability have a hard time understanding fixing.

Problem: Causal Concepts Are Not Well-Defined in Traditional Statistics

Causal Inference	Statistical Models
Directional	Lack directionality
Counterfactual	Correlational
Fixing	Conditioning

1. **Fixing:** *Causal* operation that assigns values to the inputs of structural equations associated to the variable we fix.
2. **Conditioning:** *Statistical* exercise that encompasses the dependence structure of the data generating process.
3. How to make statistics converse with causality?

Causal Frameworks

1. Hypothetical model (Heckman & Pinto, 2015)
   - Framework fully integrated into standard probability theory.
2. Do-Calculus (Pearl, 2009)
   - Defines new rules outside of standard probability and statistics.
3. Neyman-Rubin model
   - Does not use structural equations (no mechanisms generating outcomes).
   - Choice of input (X) not modeled.
   - No explicit link of inputs and outputs.
   - "Effects of causes" (Holland, 1986)

Will Focus on (1) Today

Linking Counterfactual Worlds to Data

Connecting Statistics with Causality
Econometric Approach

1. **New Model:** Define a **Hypothetical Model** with desired independent variation of inputs.
2. **Usage:** Hypothetical model allows us to examine causality.
3. **Characteristic:** Usual statisticral tools apply.
4. **Benefit:** Fixing translates to statistical conditioning.
5. **Formalizes** Frisch motto "*Causality is in the Mind*".
6. **Clarifies** the notion of identification.

Identification:

Expresses causal parameters defined in the hypothetical model using observed probabilities of the empirical model that governs the data generating process.

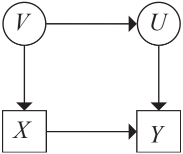
Defining the Hypothetical Model

Empirical Model: Governs the data generating process.
Hypothetical Model: Abstract model used to examine causality.

- The hypothetical model uses:
  1. **Same** set of structural equations as the empirical model.
  2. **Appends hypothetical variables that we *fix*.**
  3. **Hypothetical variable** not caused by any other variable in the system.
  4. **Replaces** the input variables we seek to fix by the hypothetical variable, which conceptually can be fixed.

The hypothetical model framework does not require any tool outside of standard probability theory, provided we endow the space of hypotheticals with a probability measure.

Empirical Model: Data Generating Process

Model	DAG	LMC
$V = f_V(\omega_V)$ $U = f_U(V, \omega_U)$ $X = f_X(V, \omega_X)$ $Y = f_Y(X, U, \omega_Y)$		$Y \perp\!\!\!\perp V \mid (U, X)$ $U \perp\!\!\!\perp X \mid V$

- Can add an augmented equation $X = f'_X(Z, V, \omega_X)$.
- Z plays the role of an instrumental variable.
- Models choices of inputs.

Define a Hypothetical Variable $\tilde{X}$

- $\tilde{X}$ replaces X as input of outcome Y.
- $Y = f_Y(\tilde{X}, U, \omega_Y)$ instead of $Y = f_Y(X, U, \omega_Y)$.
- Generates new Local Markov Conditions (LMC).

- $\tilde{X}$ is a hypothetical version of X.

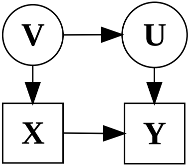
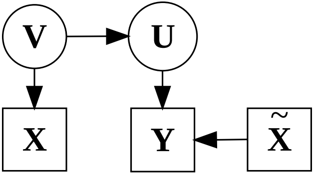
$$Y \perp\!\!\!\perp (X, V) \mid U, \tilde{X}$$
$$U \perp\!\!\!\perp (X, \tilde{X}) \mid V$$
$$\tilde{X} \perp\!\!\!\perp (U, V, X)$$
$$X \perp\!\!\!\perp (U, Y, \tilde{X}) \mid V$$

- Insert $\tilde{X}$ into structural model
- Invoke Autonomy

Example of Heckman-Pinto Approach
Example of the Hypothetical Model for Fixing X

The Associated Hypothetical Model

$$Y = f_Y(\tilde{X}, U, \omega_Y); X = f_X(V, \omega_X); U = f_U(V, \omega_U); V = f_V(\omega_V).$$

Empirical Model	Hypothetical Model
	
Local Markov Condition	Local Markov Condition
$Y \perp\!\!\!\perp V \mid (U, X)$	$Y \perp\!\!\!\perp (X, V) \mid (U, \tilde{X})$
$U \perp\!\!\!\perp X \mid V$	$U \perp\!\!\!\perp (X, \tilde{X}) \mid V$
	$\tilde{X} \perp\!\!\!\perp (U, V, X)$
	$X \perp\!\!\!\perp (U, Y, \tilde{X}) \mid V$

The Hypothetical Model and the Data Generating Process

The hypothetical model is not a speculative departure from the empirical data-generating process but an **expanded** version of it.

- Expands the number of random variables in the model.
- Allows for thought experiments.
- Allows us to manipulate $\tilde{X}$ while conditioning on X.
- Adding additional hypothetical variables.

Benefits of a Hypothetical Model

- **Formalizes** Haavelmo's insight of Hypothetical variation;
- **Statistical Analysis:** Bayesian Network Tools apply (Local Markov Condition; Graphoid Axioms, etc.);
- **Clarifies** the definition of causal parameters;
  1. Causal parameters are defined by the hypothetical model;
  2. Observed data is generated through empirical model;
- **Distinguish** definition of causal parameters from their identification;
  1. Identification requires us to **connect** the hypothetical and empirical models.
  2. Allows us to evaluate causal parameters defined in the Hypothetical model using data generated by the Empirical Model.

Identification

- **Hypothetical Model** allows analysts to define and examine causal parameters.
- **Empirical Model** generates observed/unobserved data;

Clarity

The ability to express causal parameters of the hypothetical model through observed probabilities in the empirical model.

Tools: What does Identification require?

Probability laws that connect *Hypothetical* and *Empirical* Models.

THE UNIVERSITY OF CHICAGO

The Hypothetical Model vs. Empirical Model

- Distributions of variables in hypothetical/empirical models **differ**.
  - $\mathbf{P}_E$ for the probabilities of the empirical model
  - $\mathbf{P}_H$ for the probabilities of the hypothetical model

Model Identified When

$$\mathbf{P}_E(Y \mid X = x) = \mathbf{P}_H(Y \mid \widetilde{X} = x).$$

Causal parameters are defined as conditional probabilities in the hypothetical model $\mathbf{P}_H$ and are said to be identified if those can be expressed in terms of the distribution of observed data generated by the empirical model $\mathbf{P}_E$.

How To Use This Causal Framework?
Rules of Engagement

1. **Define** the empirical and associated hypothetical model.
2. **Hypothetical Model:** Generate statistical relationships (LMC, GA).
3. **Express** $\mathbf{P}_{\mathrm{H}}(Y \mid \widetilde{X})$ in terms of other variables.
4. **Connect** this expression to the empirical model.

Controlling for V Is the Key

(1) Matching
(2) IV (regression discontinuity; RCT)
(3) Factor models
  - Extract V from measures (e.g., Bartlett scores)
  - Joint factor models (LISREL CFA)

Example: Matching
Connecting Empirical and Hypothetical Models

Matching Property

If there exist a variable V not caused by $\tilde{X}$, such that $X \perp\!\!\!\perp Y | V, \tilde{X}$, then $E_{\text{H}}(Y|V, \tilde{X} = x)$ under the hypothetical model is equal to $E_{\text{H}}(Y \mid V, X = x)$ under empirical model.

- **Obs:** LMC for the hypothetical model generates $X \perp\!\!\!\perp Y | V, \tilde{X}$.
- Thus, by matching, treatment effects $E_{\text{H}}(Y(x))$ can be obtained by:

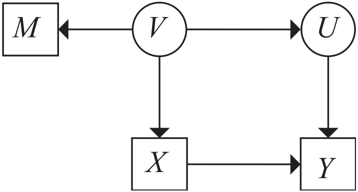
$$E_{\text{H}}(Y(x)) = \underbrace{\int E_{\text{H}}(Y \mid V = v, \tilde{X} = x) dF_V(v)}_{\text{In Hypothetical Model}}$$

$$= \underbrace{\int E_{\text{E}}(Y \mid V = v, X = x) dF_V(v)}_{\text{In Empirical Model}}$$

Controlling for V

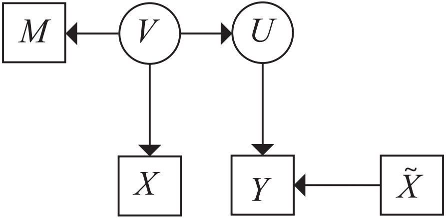
- But if V is unobserved, then the model is unidentified without further assumptions.
- A variety of methods exist for unknown or mismeasured V.

Latent Variable Model
Empirical Model



$$V = f_V(\omega_V)$$
$$M = f_M(V, \omega_M)$$
$$U = f_U(V, \omega_U)$$
$$X = f_X(V, \omega_X)$$
$$Y = f_Y(X, U, \omega_Y)$$

Latent Variable Model
Hypothetical Model



- The underlying idea is:

$$Y \perp\!\!\!\perp X \mid (U, \tilde{X}) \text{ by LMC, and } U \perp\!\!\!\perp (X, \tilde{X}) \mid V \text{ by LMC}$$

$$Y \perp\!\!\!\perp X \mid (U, \tilde{X}) \text{ and } U \perp\!\!\!\perp (X, \tilde{X}) \mid V \Rightarrow Y \perp\!\!\!\perp X \mid (V, \tilde{X})$$

by Graphoid Axioms.

Link to Local Markov Condition & Graphoid Axioms

- Now we can use M to control for V under additional assumption $\Rightarrow Y \perp\!\!\!\perp X \mid (\rho(M), \tilde{X})$, where $\rho(M) = V$.
- X "purged" of V $[X_{-V}]$: $X_{-V} = \tilde{X}$

Ways to Control for V:

- Directly measure it and condition on it (Replacement function)
- Eliminate it by some operation (e.g. Fixed effect model; difference out V)
- In

$$Y = X\beta + U$$
$$U = \psi V + \varepsilon$$

  instrument X (shift X holding U fixed; vary X indepdently of V)
- RCT is an instrument for X

Linear Equation Examples: Some Ways to Eliminate V from Heckman & Robb (1985)

(1) Replacement functions:

$$M = Z\gamma + V$$

$$(M, Z) \text{ observed, } V \perp\!\!\!\perp Z$$

$$M - Z\gamma = V$$

Substitute for V: $Y = X\beta + \phi V + \varepsilon$

Assume $\sum_{X, M-Z\gamma} = \sum_{X,V}$ is full rank.

(2) Substitute directly (use estimated V or just condition on M, Z), need M, Z to estimate or control for V.

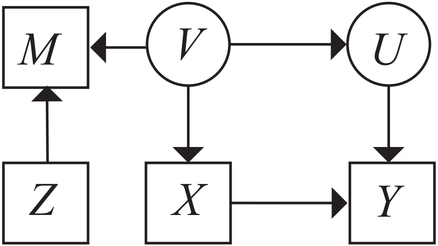
(3) **Control Function**: Condition on a function of M and Z.

(4) $E(Y \mid X, V) = X\beta + \phi E(U \mid V)$: model the U, V dependence (selection corrections).

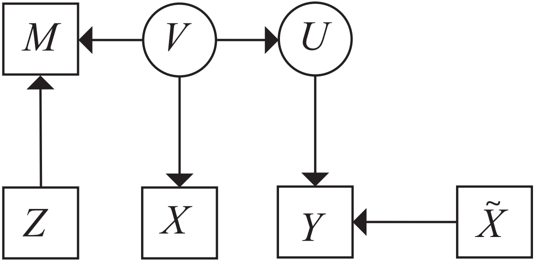
(5) Factor model: $M = \lambda Z + \ V + \varepsilon$

1. Bartlett method
2. Fixed effect

Figure 1: Factor Model:



The Hypothetical Model



Well Known Methods to Control for V

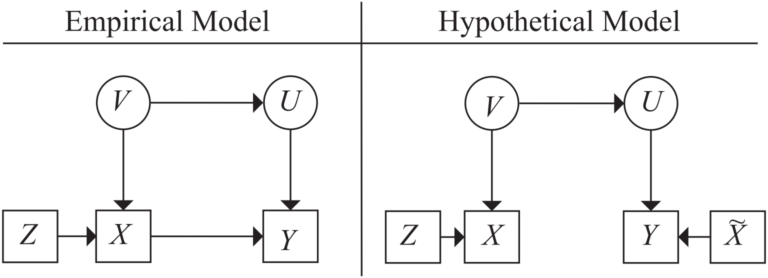
(5) Randomized Control Trials (RCTs)

- Controls for V by randomly assigning values to X, which implies $X \perp\!\!\!\perp V$.

(6) Instrumental variables (IV)

- Explores an exogenous random variable Z that causes X, but does not directly cause any other variable of the system.

The IV Model



LMC Empirical Model	LMC Hypothetical Model
$Y \perp\!\!\!\perp V, Z \mid (X, U)$	$Y \perp\!\!\!\perp (V, X, Z) \mid (U, \tilde{X})$
$Z \perp\!\!\!\perp (V, U)$	$Z \perp\!\!\!\perp (V, U, Y, \tilde{X})$
$U \perp\!\!\!\perp (Z, X) \mid V$	$U \perp\!\!\!\perp (Z, X, \tilde{X}) \mid V$
	$\tilde{X} \perp\!\!\!\perp (U, V, X, Z)$

Source: Heckman & Pinto (2013). IV outside the range of Do-calculus.

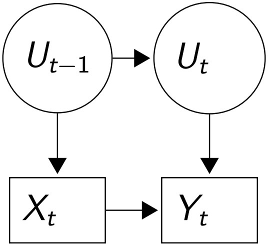
(7) Time series/panel methods (replacement functions)

$$
\begin{aligned}
Y_t &= X_t\beta + U_t, \quad \text{where } U_t \not\perp\!\!\!\perp X_t \\
U_t &= \rho U_{t-1} + \varepsilon_t, \quad \text{so } U_{t-1} \text{ plays role of } V_t \\
\varepsilon_t &\perp\!\!\!\perp (U_{t-1}, X_{t-1}, \dots), \quad \text{but } U_{t-1} \not\perp\!\!\!\perp X_t \\
Y_t &= X_t\beta + \rho U_{t-1} + \varepsilon_t. \\
\text{But } Y_{t-1} - X_{t-1}\beta &= U_{t-1} \quad \text{(replacement function)} \\
Y_t &= \rho Y_{t-1} + X_t\beta - \rho X_{t-1}\beta + \varepsilon_t
\end{aligned}
$$

- Can identify β, ρ under no-collinearity assumptions

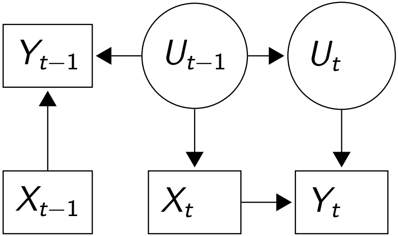
THE UNIVERSITY OF CHICAGO

Time Series Unit Model at Time t



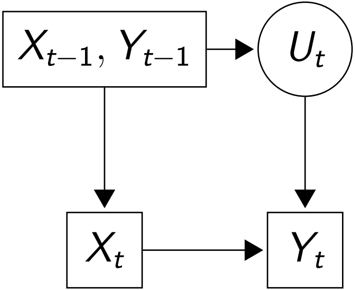
- U_{t-1} plays the role of V_t
- $Y_t = \beta X_t + U_t$
- $U_t = \rho U_{t-1} + \epsilon_t$

Time Series Model with Additional Lag



- $Y_{t-1} = \beta X_{t-1} + U_{t-1}$

Time Series Model with Replacement Function



- $U_{t-1} = \beta X_{t-1} - Y_{t-1}$
- $U_t = \rho Y_{(t-1)} - \rho\beta X_{(t-1)} + \epsilon_t$
- $Y_t = \beta X_t + U_t \;=\; \beta X_t + \rho(Y_{t-1} - \beta X_{t-1}) + \epsilon_t$

THE UNIVERSITY OF CHICAGO

Hypothetical Models and Simultaneous Equations

The Simultaneous Equation Model (Haavelmo, 1944)

A system of two equations:

$$Y_1 = g_{Y_1}(Y_2, X_1, U_1) \tag{8}$$
$$Y_2 = g_{Y_2}(Y_1, X_2, U_2). \tag{9}$$

- **Variables:** $\mathcal{T}_{\mathrm{E}} = \{Y_1, Y_2, X_1, X_2, U_1, U_2\}$.
- **Assumptions:** $U_1 \perp\!\!\!\perp U_2$ and $(U_1, U_2) \perp\!\!\!\perp (X_1, X_2)$. (made only to simplify the argument)
- **LMC** condition breaks down.
- **Matzkin (2008)** relaxes these assumptions and identifies causal effects for $U_1 \not\perp\!\!\!\perp U_2$ and $(U_1, U_2) \not\perp\!\!\!\perp (X_1, X_2)$.

Fixing readily extends to a system of simultaneous equations for Y_1 and Y_2, whereas the fundamentally recursive methods based on DAGs do not (Pearl, 2009).

Completeness Assumption

- **Common Assumption:** completeness—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):

$$Y_1 = \phi_1(X_1, X_2, U_1, U_2) \tag{11}$$
$$Y_2 = \phi_2(X_1, X_2, U_1, U_2). \tag{12}$$

- **Reduced form** equations (see, e.g., Matzkin, 2008, 2013).
- **Inherit** the autonomy properties of the structural equations.

Characteristics of the Simultaneous Equation Model

- **Autonomy:** the causal effect of Y_2 on Y_1 when Y_2 is fixed at y_2 is given by

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

- **Symmetrically:**

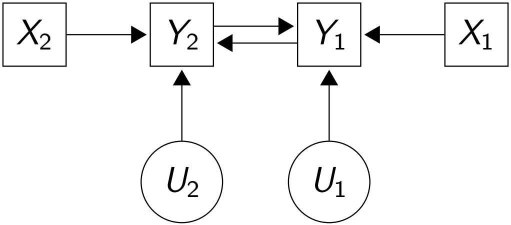
$$Y_2(y_1) = g_{Y_2}(y_1, X, U_2).$$

- **Define** hypothetical random variables $\tilde{Y}_1, \tilde{Y}_2$ such that:
  - $\tilde{Y}_1, \tilde{Y}_2$ replaces the Y_1, Y_2 inputs on Equations (8) and (9).
  - $(\tilde{Y}_1, \tilde{Y}_2) \perp\!\!\!\perp (X_1, X_2, U_1, U_2)$; and $\tilde{Y}_1 \perp\!\!\!\perp \tilde{Y}_2$.
  - $\mathcal{T}_{\text{H}} = \{\tilde{Y}_1, \tilde{Y}_2, Y_1, Y_2, X_1, X_2, U_1, U_2\}$.
  - Assume a common support for (Y_1, Y_2) and $(\tilde{Y}_1, \tilde{Y}_2)$.

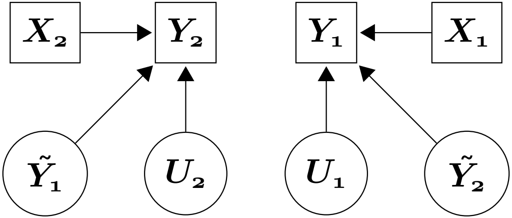
Counterfactuals of the Simultaneous Equation Model

- **Distribution** of Y_1 when Y_2 is fixed at y_2 is given by
$$\mathbf{P}_{\mathrm{H}}(Y_1 \mid \tilde{Y}_2 = y_2).$$
- **Average causal effect** of Y_2 on Y_1 when Y_2 is fixed at y_2 and y_2' values:
$$E_{\mathrm{H}}(Y_1 \mid \tilde{Y}_2 = y_2) - E_{\mathrm{H}}(Y_1 \mid \tilde{Y}_2 = y_2')$$
- **Notation:** E_{H} denotes expectation over the probability measure $\mathbf{P}_{\mathrm{H}}$ of the hypothetical model.

Empirical Model for Simultaneous Equations



Some Hypothetical Models for Y_2 and Y_1, Respectively



Pearl (2009): Solves simultaneity problems by "shutting one equation down" assuming remaining parameters are the same as before.

- In social systems, this thought experiment is often not credible.
- This is strictly not required.

Definition and Identification: Nonlinear Case

- In a general nonlinear model,

$$
\begin{aligned}
Y_1 &= g_{Y_1}(Y_2, X_1, X_2, U_1) \\
Y_2 &= g_{Y_2}(Y_1, X_1, X_2, U_2),
\end{aligned}
$$

- Exclusion is defined as $\dfrac{\partial g_{Y_1}}{\partial X_2} = 0$ for all (Y_2, X_1, X_2, U_1) and $\dfrac{\partial g_{Y_2}}{\partial X_1} = 0$ for all (Y_1, X_1, X_2, U_2).

- Assuming the existence of local solutions, we can solve these equations to obtain:

$$\begin{aligned} Y_1 &= \varphi_1(X_1, X_2, U_1, U_2) \\ Y_2 &= \varphi_2(X_1, X_2, U_1, U_2) \end{aligned}$$

- By the chain rule we can write:

$$\frac{\partial g_{Y_1}}{\partial Y_2} = \frac{\partial Y_1}{\partial X_1} \bigg/ \frac{\partial Y_2}{\partial X_1} = \frac{\partial \varphi_1}{\partial X_1} \bigg/ \frac{\partial \varphi_2}{\partial X_1}.$$

- We may define and identify causal effects for Y_1 on Y_2 using partials with respect to X_2 in an analogous fashion.

If X_1 and X_2 are disjoint (made only to simplify exposition):

$$\frac{\partial Y_1}{\partial X_2} = \frac{\partial g_{Y_1}(Y_2, X_1, U_1)}{\partial Y_2} \quad \frac{\partial Y_2}{\partial X_2}$$

$$\begin{aligned}\frac{\partial Y_1}{\partial X_2} &= \frac{\partial g_{Y_1}(Y_2, X_1, U_1)}{\partial X_2} \\ &= \frac{\partial g_{Y_1}(\cdot)}{\partial Y_2(\cdot)} \quad \frac{\partial Y_2(\cdot)}{\partial X_2}\end{aligned}$$

$$\frac{\frac{\partial Y_1}{\partial X_2}}{\frac{\partial Y_2}{\partial X_2}} = \frac{\frac{\partial \phi_1(\cdot)}{\partial X_2}}{\frac{\partial \phi_2(\cdot)}{\partial X_2}} = \frac{\partial g_{Y_1}(\cdot)}{\partial Y_2}$$

Cannot be identified by the rules of the do-calculus. No "wiping out" needed as in Pearl (2009).

Econometric Mediation Analysis

- Build on Wright(1921, 1934), Klein and Goldberger (1955), and Theil (1958).
- Reduced form estimates the **net effect** of a policy change X_1,

$$\frac{\partial Y_1}{\partial X_1} = \frac{\partial \phi_1(X_1, X_2, U_1, U_2)}{\partial X_1}. \tag{14}$$

- Using this analysis, the system can trivially be used to conduct mediation analyses.

$$\frac{\partial Y_1}{\partial X_1} = \underbrace{\left(\frac{\partial g_{Y_1}}{\partial Y_2}\right)}_{\substack{\text{Identified}\\ \text{through}\\ \text{exclusion}\\ \text{in structure}}} \underbrace{\left(\frac{\partial Y_2}{\partial X_1}\right)}_{\substack{\text{Identified}\\ \text{from reduced}\\ \text{form}}} + \underbrace{\frac{\partial g_{Y_1}}{\partial X_1}}_{\substack{\text{Identified}\\ \text{from}\\ \text{structure}}} = \frac{\partial \phi_1(X_1, X_2, U_1, U_2)}{\partial X_1}$$

Linear Example as in Haavelmo (1944)

- Linear model in terms of parameters (Γ, B), observables (Y, X) and unobservables U:

$$\Gamma Y + BX = U \qquad E(U) = 0 \tag{15}$$

- Y is a vector of internal and interdependent variables.
- X is external and exogenous $(E(U \mid X) = 0)$.
- Γ is a full rank matrix.

Some Properties

- Linear-in-the-parameters "all causes" model for vector Y.
- Causes are X and U.
- The "structure" is (Γ, B), Σ_U, where Σ_U is the variance-covariance matrix of U.
- In the Cowles Commission analysis it is assumed that Γ, B, Σ_U are **invariant** to classes of changes in X and modifications of the distribution of U.
- Autonomy (Frisch, 1938).
- Later defined as part of the "SUTVA" (1986) assumption.
- However, the model obviously involves interaction among agents, something ruled out by "SUTVA."

Two Agent Economic Model

- Consider a two-agent model of social interactions.
- Y_1 is the outcome for agent 1; Y_2 is the outcome for agent 2.

$$Y_1 = \alpha_1 + \gamma_{12} Y_2 + \beta_{11} X_1 + \beta_{12} X_2 + U_1 \tag{16}$$

$$Y_2 = \alpha_2 + \gamma_{21} Y_1 + \beta_{21} X_1 + \beta_{22} X_2 + U_2. \tag{17}$$

- Social interactions model ("reflection problem") is a version of the standard simultaneous equations problem with enhanced error structure.

Reduced Form

- Under completeness, the reduced form outcomes of the model after social interactions are solved out can be written as:

$$Y_1 = \pi_{10} + \pi_{11} X_1 + \pi_{12} X_2 + \mathcal{E}_1, \tag{18}$$

$$Y_2 = \pi_{20} + \pi_{21} X_1 + \pi_{22} X_2 + \mathcal{E}_2. \tag{19}$$

$$\pi_{11} = \frac{\beta_{11} + \gamma_{12}\beta_{21}}{1 - \gamma_{12}\gamma_{21}}, \quad \pi_{12} = \frac{\beta_{12} + \gamma_{12}\beta_{22}}{1 - \gamma_{12}\gamma_{21}},$$

$$\pi_{21} = \frac{\gamma_{21}\beta_{11} + \beta_{21}}{1 - \gamma_{12}\gamma_{21}}, \quad \pi_{22} = \frac{\gamma_{21}\beta_{12} + \beta_{22}}{1 - \gamma_{12}\gamma_{21}}$$

$$\mathcal{E}_1 = \frac{U_1 + \gamma_{12}U_2}{1 - \gamma_{12}\gamma_{21}},$$

$$\mathcal{E}_2 = \frac{\gamma_{21}U_1 + U_2}{1 - \gamma_{12}\gamma_{21}}.$$

Example of Exclusion in Linear Model

$$\beta_{12} = 0$$

$$\pi_{12} = \frac{\gamma_{12}\beta_{22}}{1 - \gamma_{12}\gamma_{21}}$$

$$\pi_{22} = \frac{\beta_{22}}{1 - \gamma_{12}\gamma_{21}}$$

$$\frac{\pi_{12}}{\pi_{22}} = \gamma_{12} \quad \text{(causal effect of } Y_2 \text{ on } Y_1\text{)}$$

THE UNIVERSITY OF CHICAGO

Summary

- Understanding causal content of $Y = X\beta + U$.
- Answer is a major challenge to conventional statistics.
- The received literature often conflates definition, identification, and estimation.
- The econometric approach delineates these three tasks.

Table 3: **Three Distinct Tasks Arising in the Analysis of Causal Models**

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

Benefits of Hypothetical Models

- Separate issues of estimation from those of definition and identification.
- Understand mechanisms generating outcomes motivates identification and estimation strategies. (Example: latent variables.)
- Can address in a common framework problems of
    1. Internal validity
    2. External validity (autonomy)
    3. Forecasting worlds never previously experienced
- These are treated as separate issues in some literatures.

Summary of Causal Frameworks

- **Heckman/Pinto: Haavelmo** (Hypothetical Model)
  Mechanisms clarified—all three policy problems addressed.
  - Introduce hypothetical variables: output of thought experiments.
  - Endows these hypothetical models with well-defined probability measures.
  - **Add these** to empirical model space.
  - Shows how to connect the empirical with the hypothetical (identification).
  - Same framework can be used to forecast out-of-sample and combine samples and forecast impacts of new policies never previous experienced.

Some Further Reading

- Heckman, James J. "Econometric Causality." *International Statistical Review* 76, no. 1 (2008): 1-27.
- Heckman, James, and Rodrigo Pinto. "Causal Analysis After Haavelmo." *Econometric Theory* 31, no. 1 (2015): 115.
- Pinto, Rodrigo and James J. Heckman. "The Econometric Model for Causal Policy Analysis." Under review, *Journal of Econometrics* (2022).

Local Markov Condition (LMC)

(Kiiveri, 1984, Lauritzen, 1996)

- If a model is acyclical, i.e., $Y \notin D(Y) \ \forall \, Y \in \mathcal{T}$ then any variable is independent of its non-descendants, conditional on its parents:

$$\textbf{LMC}: Y \perp\!\!\!\perp \underbrace{\mathcal{T} \setminus (D(Y) \cup Y)}_{\text{set difference}} \mid Pa(Y) \quad \forall \, Y \in \mathcal{T}.$$

Graphoid Axioms (GA)

(Dawid, 1979)

Symmetry: $X \perp\!\!\!\perp Y \mid Z \Rightarrow Y \perp\!\!\!\perp X \mid Z.$

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

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

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

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

Redundancy: $X \perp\!\!\!\perp Y \mid X.$

Return to main text

Local Markov Condition (LMC)
(Kiiveri, 1984, Lauritzen, 1996)

- If a model is acyclical, i.e., $Y \notin D(Y) \ \forall \ Y \in \mathcal{T}$ then any variable is independent of its non-descendants, conditional on its parents:

$$\textbf{LMC}: Y \perp\!\!\!\perp \underbrace{\mathcal{T} \setminus (D(Y) \cup Y)}_{\text{set difference}} \mid Pa(Y) \quad \forall \ Y \in \mathcal{T}.$$

Graphoid Axioms (GA)

(Dawid, 1979)

Symmetry: $X \perp\!\!\!\perp Y \mid Z \Rightarrow Y \perp\!\!\!\perp X \mid Z.$

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

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

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

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

Redundancy: $X \perp\!\!\!\perp Y \mid X.$

THE UNIVERSITY OF CHICAGO

Return to main text