Causality Part II: Further Remarks

James J. Heckman and Rodrigo Pinto

Econ 312, Spring 2021

Problem: Causal Concepts are not Well-defined in Traditional Statistics

Causal Inference	**Statistical Models**
Directional	Lacks directionality
Counterfactual	Correlational
Fixing	Conditioning

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

Some Solutions in the Literature

1. Cowles Foundation model (Haavelmo, 1943; 1944) updated by Heckman & Pinto Hypothetical Model (Theoretical Econometrics, 2014, "Causal Analysis After Haavelmo").
2. Pearl's do-calculus (series of books, 2009).
3. Neyman-Rubin model (e.g., Imbens and Rubin, 2015).

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).
   - Choice of input (X) not modeled.
   - No explicit link of inputs and outputs.

Table 1: Comparison of the Aspects of Evaluating Social Policies that are Covered by the Neyman-Rubin Approach and the Structural Approach (Treatment Effect Example)

	Neyman-Rubin Framework
Counterfactuals for objective outcomes $Y(x)$	Yes
Agent valuations of subjective outcomes	No (choice-mechanism implicit)
Models for the causes of potential outcomes	No
Ex ante versus *ex post* counterfactuals	No
Treatment assignment rules that recognize voluntary nature of participation	No
Social interactions, general equilibrium effects and contagion	No (due to "SUTVA")
Internal validity (problem **P1**)	Yes
External validity (problem **P2**)	No
Forecasting effects of new policies (problem **P3**)	No
Distributional treatment effects	No[a]
Analyze relationship between outcomes and choice equations	No (implicit)

[a] An exception is the special case of common ranks of individuals across counterfactual states: "rank invariance." See the discussion in Abbring and Heckman (2007).

Table 1: Comparison of the Aspects of Evaluating Social Policies that are Covered by the Neyman-Rubin Approach and the Structural Approach (Treatment Effect Example), Cont'd

	Structural Framework
Counterfactuals for objective outcomes $Y(x)$	Yes
Agent valuations of subjective outcomes	Yes (explicit)
Models for the causes of potential outcomes	Yes
Ex ante versus *ex post* counterfactuals	Yes
Treatment assignment rules that recognize voluntary nature of participation	Yes
Social interactions, general equilibrium effects and contagion	Yes (modeled)
Internal validity (problem **P1**)	Yes
External validity (problem **P2**)	Yes
Forecasting effects of new policies (problem **P3**)	Yes
Distributional treatment effects	Yes (for the general case)
Analyze relationship between outcomes and choice equations	Yes (explicit)

[a]An exception is the special case of common ranks of individuals across counterfactual states: "rank invariance." See the discussion in Abbring and Heckman (2007).

Linking Counterfactual Worlds to Data

How to Connect 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 statistical 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:

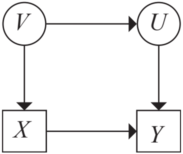
Expresses causal parameters defined in the hypothetical model using observed probabilities of the empirical model that governs the data generating process.

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.
  4. **Replaces** the input variables we seek to fix by the hypothetical variable, which conceptually can be fixed.

Empirical Model: Data Generating Process

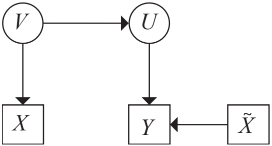
Model	DAG	LMC
$V = f_V(\omega_V)$ $U = f_U(V, \omega_U)$ $X = f_X(V, \omega_X)$ $Y = f_Y(X, U, \omega_Y)$		$Y \perp\!\!\!\perp V \mid (U, X)$ $U \perp\!\!\!\perp X \mid V$

- Can add an augmented equation $X = f'_X(Z, V, \omega_X)$.
- 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).

Associated Hypothetical Model (with Hypothetical Variable $\tilde{X}$)

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

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: What is Identification?

The capacity 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 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

Counterfactuals obtained by simple conditioning

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

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** $\mathbf{P}_{\mathrm{H}}(Y|\widetilde{X})$ 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 exists a variable V not caused by $\tilde{X}$, such that, $X \perp\!\!\!\perp Y | V, \tilde{X}$, then $E_{\mathrm{H}}(Y | V, \tilde{X} = x)$ under the hypothetical model is equal to $E_{\mathrm{H}}(Y | V, X = x)$ under empirical model.

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

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

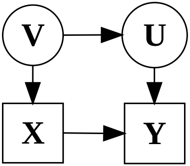
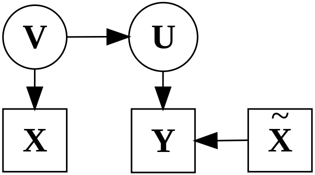
$$= \underbrace{\int E_{\mathrm{E}}(Y | V = v, X = x) dF_V(v)}_{\text{In Empirical Model}}$$

Controlling for V

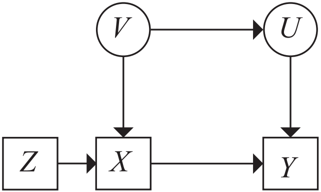
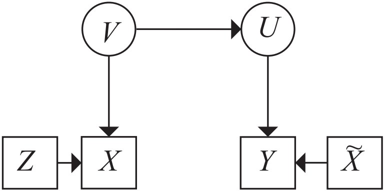
- But if V is unobserved, then the model is unidentified without further assumptions.
- A variety of methods exist for unknown or mismeasured V.

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

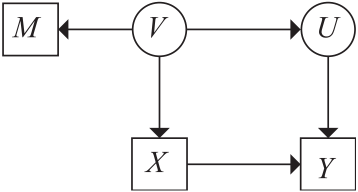
The IV Model

Empirical Model	Hypothetical Model
	

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

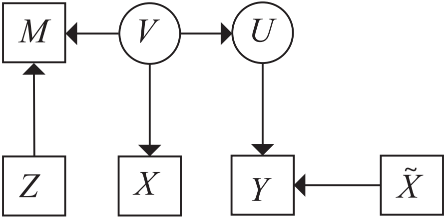
Source: Heckman & Pinto (2013). IV outside the range of Do-calculus.

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:

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

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

by Graphoid Axioms.

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

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

(1) Replacement functions:

$$M = Z\gamma + V$$
$$(M,Z) \text{ observed, } V \perp\!\!\!\perp Z$$
$$M - Z\gamma = V$$

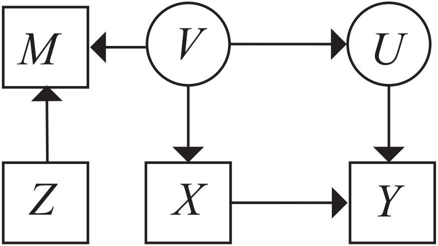
Substitute for V : $Y = X\beta + \phi V + \varepsilon$
Assume $\sum_{X,M-Z\gamma} = \sum_{X,V}$ is full rank.

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

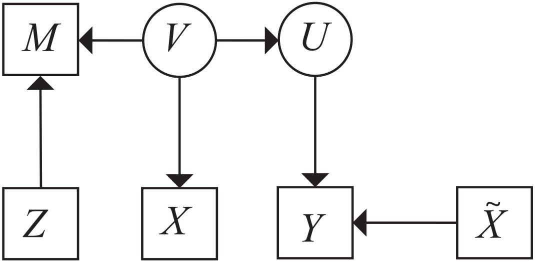
(3) Factor model: $M = \lambda Z + \Lambda V + \varepsilon$

1. Bartlett method
2. Fixed effect

Figure 1: Factor Model:



The Hypothetical Model



Well Known Methods to Control for V

(4) Randomized Control Trials (RCTs)

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

(5) Instrumental variables (IV)

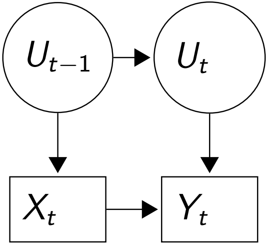
- Uses an exogenous random variable Z that causes X, but does not directly cause any other variable of the system.

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

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

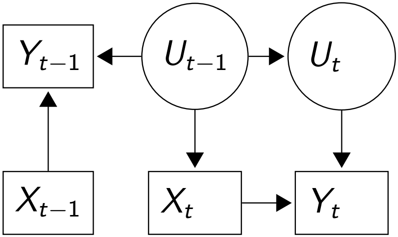
- Can identify β, ρ under no-collinearity assumptions

Time Series Unit Model for Time t



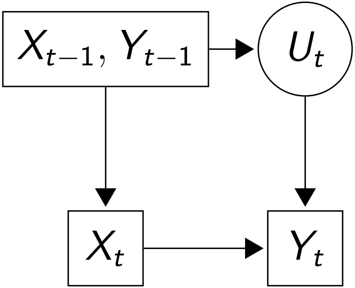
- U_{t-1} plays the role of V_t
- $Y_t = \beta X_t + U_t$
- $U_t = \rho U_{t-1} + \epsilon_t$

Time Series Model with Additional Lag



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

Time Series Model with Replacement Function



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

Hypothetical Models and Simultaneous Equations

The Simultaneous Equation Model (Haavelmo, 1944)

A system of two equations:

$$Y_1 = g_{Y_1}(Y_2, X_1, U_1) \tag{1}$$
$$Y_2 = g_{Y_2}(Y_1, X_2, U_2). \tag{2}$$

- **Variables:** $\mathcal{T}_E = \{Y_1, Y_2, X_1, X_2, U_1, U_2\}$.
- **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 $(U_1, U_2) \not\perp\!\!\!\perp (X_1, X_2)$.

THE UNIVERSITY OF CHICAGO

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):

$$Y_1 = \phi_1(X_1, X_2, U_1, U_2) \tag{4}$$
$$Y_2 = \phi_2(X_1, X_2, U_1, U_2). \tag{5}$$

- **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:**

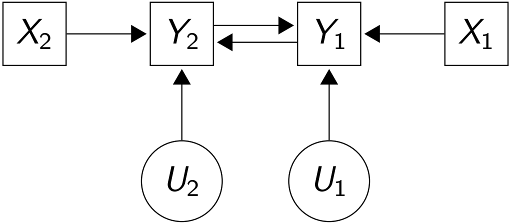
$$Y_2(y_1) = g_{Y_2}(y_1, X, U_2).$$

- **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 (1) and (2).
  - $(\tilde{Y}_1, \tilde{Y}_2) \perp\!\!\!\perp (X_1, X_2, U_1, U_2)$; 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 $(\tilde{Y}_1, \tilde{Y}_2)$.

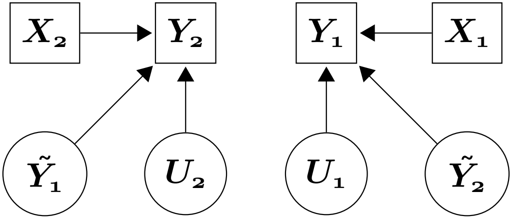
Counterfactuals of the Simultaneous Equation Model

- **Distribution** of Y_1 when Y_2 is fixed at y_2 is given by
$$\mathbf{P}_{\mathrm{H}}(Y_1|\tilde{Y}_2 = y_2).$$
- **Average causal effect** of Y_2 on Y_1 when Y_2 is fixed at y_2 and y_2' values:
$$E_{\mathrm{H}}(Y_1|\tilde{Y}_2 = y_2) - E_{\mathrm{H}}(Y_1|\tilde{Y}_2 = y_2')$$
- **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



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):

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

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

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

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,

$$\frac{\partial Y_1}{\partial X_1} = \frac{\partial \phi_1(X_1, X_2, U_1, U_2)}{\partial X_1}. \tag{7}$$

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

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

THE UNIVERSITY OF CHICAGO

Linear Example as in Haavelmo (1944)

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

$$\Gamma Y + BX = U \qquad E(U) = 0 \tag{8}$$

- 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.

$$Y_1 = \alpha_1 + \gamma_{12} Y_2 + \beta_{11} X_1 + \beta_{12} X_2 + U_1 \quad (9)$$
$$Y_2 = \alpha_2 + \gamma_{21} Y_1 + \beta_{21} X_1 + \beta_{22} X_2 + U_2. \quad (10)$$

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

THE UNIVERSITY OF CHICAGO

Reduced Form

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

$$Y_1 = \pi_{10} + \pi_{11} X_1 + \pi_{12} X_2 + \mathcal{E}_1, \tag{11}$$
$$Y_2 = \pi_{20} + \pi_{21} X_1 + \pi_{22} X_2 + \mathcal{E}_2. \tag{12}$$

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

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

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

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

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 2: **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	} Statistical Analysis
3: Estimation	Estimating parameters from real data	Estimation and testing theory	} Statistical Analysis

- Today we focused on Task 1 and a bit on Task 2.
- Much of the literature starts at Task 3.

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

- **Rubin:** "Potential outcomes"
  - No model of selection of inputs.
  - Focuses on policy problem P1.
  - Issues of support, extrapolation, external validity, forecasting outcomes never experienced are settled on ad hoc basis.
  - No models of mechanisms: "effects of causes, *not* causes of effects."
  - No model of unobservables connecting equations or models of systems of behavioral relationships.
  - This framework is a special and restricted case of structural equations.
  - "Potential outcomes" in fact are outputs of structural equations but not recognized as such by followers of this approach.

THE UNIVERSITY OF CHICAGO

Summary of Causal Frameworks

- **Pearl:** Do-calculus uses structural models
  - Defines causality by invoking ad hoc rules.
  - Rules are outside statistics and probability theory.
  - The ad hoc rules of "do calculus" operate on empirical models to generate causal models.
  - He starts and ends with the data generation process invoking special rules for the variables.
  - Does not work with hypothetical models.

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.