Causal Analysis After Haavelmo

James Heckman and Rodrigo Pinto

University of Oslo
Haavelmo Lecture
December 13, 2013
Oslo is the cradle of rigorous causal inference.
Two Giants

Ragnar Frisch

Trygve Haavelmo
Haavelmo’s Research Program

(a) Specify an economic model.
(b) Specify a probability model to take the economic model to data.
(c) Use Neyman-Pearson theory to test economic hypotheses.
(d) Use statistical theory to produce estimates with good properties and forecasts with interpretable stochastic variability.
Distinguish Three Policy Evaluation Questions

**P1** Evaluating the Impact of Historical Interventions on Outcomes Including Their Impact in Terms of Welfare. ("Internal Validity")

**P2** Forecasting the Impacts (Constructing Counterfactual States) of Interventions Implemented in one Environment in Other Environments, Including Their Impacts In Terms of Welfare. ("External Validity")

**P3** Forecasting the Impacts of Interventions (Constructing Counterfactual States Associated with Interventions) Never Historically Experienced to Various Environments, Including Their Impacts in Terms of Welfare.
Table 1: Three Distinct Tasks Arising in the Analysis of Causal Models

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Defining the Set of Hypotheticals or Counterfactuals</td>
<td>A Scientific Theory</td>
</tr>
<tr>
<td>2</td>
<td>Identifying Causal Parameters from Hypothetical Population Data</td>
<td>Mathematical Analysis of Point or Set Identification</td>
</tr>
<tr>
<td>3</td>
<td>Identifying Parameters from Real Data</td>
<td>Estimation and Testing Theory</td>
</tr>
</tbody>
</table>
Haavemo’s Contributions to Causality:

Two seminal papers (1943, 1944):

1. **Formalized** Yule’s credo: *Correlation is not causation.*
2. **Laid** the foundations for *counterfactual* policy analysis.
3. **Distinguished** *fixing* (causal operation) from *conditioning* (statistical operation).
4. **Distinguished** *definition* of causal parameters from their *identification* from data.
5. **Formalized** Marshall’s notion of *ceteris paribus* (1890).

**Most Important**

Causal effects are determined by the impact of *hypothetical* manipulations of an input on an output.
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, published 2011; emphasis added)
A simple example of a structural causal relationship:
(Haavelmo, 1944, *Econometrica*)

\[ Y = X_b \beta_b + X_p \beta_p + U \] (\(*)\)

\(U\): A variable unobserved by the analyst

\(X_b\): background variables, not thought to be of policy interest.

\(X_p\): policy variables (sought to be manipulated by interventions)

\(\ast\) is an “all causes” model

External manipulations define causal parameters:
Variations in \((X_p)\) that hold \(U\) and \(X_b\) fixed
If the coefficients \((\beta_b, \beta_p)\) are invariant to shifts in \((X_p)\), then \((\ast)\) is structural for \(X_p\).
Notice that OLS produces the relationship:

\[ E^*(Y \mid X_b, X_p) = X_b\beta_b + X_p\beta_p + E^*(U \mid X_b, X_p) \]

where \( E^* \) is a linear projection.

- OLS does not in general estimate a structural relationship.

- If \( E(U \mid X_b, X_p) = 0 \), OLS gives a structural estimator for \((\beta_b, \beta_p)\).
• If

\[ E^*(U \mid X_b, X_p) = E^*(U \mid X_b) \]

and the coefficients \( \beta_b \) and \( \beta_p \) are invariant to manipulations in \( X_p \) then OLS is structural for \( \beta_p \).

• But not necessarily for \( \beta_b \).
Haavelmo’s Insights:

1. What are Causal Effects?
   - *Not* empirical descriptions of actual worlds
   - *But* descriptions of hypothetical worlds.

2. How are they obtained?
   - *Through* Models—idealized thought experiments.
   - *By* varying—hypothetically—the inputs causing outcomes.

3. But what are models?
   - Frameworks defining causal relationships among variables.
   - Based on scientific knowledge.
Goal

Incorporate Haavelmo’s insights into one widely used causal framework that:

- Is consistent with his ideas
- Benefits from modern mathematical/statistical language.
Revisiting Haavemo’s Ideas on Causality

- **Insight:** express causality through a *hypothetical model* assigning independent variation to inputs determining outcomes.
- **Data:** generated by an empirical model that shares some features with the hypothetical model.
- **Identification:** relies on evaluating causal parameters defined in the *hypothetical model* using data generated by the *empirical model*.
- **New Tools:** use the modern language of Directed Acyclic Graphs (DAGs).
- **Comparison:** how a causal framework inspired by Haavelmo’s ideas relates to other approaches (Pearl, 2009).
Plan of the Talk

1. Revisit Haavelmo’s concept of causality using language of DAGs.
   - Causal Model
   - Independence Relationships based on Bayesian Networks
   - “Fixing” Operator
   - Fixing: standard statistical tools do not apply

2. Present New Causal framework inspired by Haavelmo’s concept of hypothetical variation of inputs.
   - Hypothetical Model for Examining Causality
   - Benefits of a Hypothetical Model
Plan of the Talk

3. **Compare** Haavelmo’s approach with Pearl’s *Do-calculus*.

   - Do-calculus is widely used in branches of computer science.
   - Haavelmo’s approach: thinks outside the box.
   - Do-calculus: requires complex *ad hoc* tools to work within the box.

4. **Limitations** of DAGs for econometric identification:

   - Cannot identify the standard IV Model or the Generalized Roy Model.
   - Can identify the “front door” model.
   - Rely solely on conditional independence assumptions.

5. **Review:**

   - Do-calculus of Pearl (2009) and a Framework Motivated by Haavelmo’s Analysis.
Plan of the Talk

6 Simultaneous Equations:
   • Hypothetical Models and Simultaneous Equations.

7 Conclusions on Causality:
   • Examine Haavelmo’s fundamental contributions.
   • Causal framework inspired by Haavelmo’s ideas.
   • Beyond DAGs.

8 Haavelmo’s Research Program: A Retrospective.
1. Haavelmo’s Approach to Causality
Causal Model: defined by 4 components:

1. **Random Variables** that are observed and/or unobserved by the analyst: \( T = \{ Y, U, X, V \} \). Let \( T \) denote an element of \( T, T \in T \).

2. **Error Terms** that are mutually independent: \( \epsilon_Y, \epsilon_U, \epsilon_X, \epsilon_V \).

3. **Structural Equations** that are autonomous: \( f_Y, f_U, f_X, f_V \).
   - By **Autonomy** we mean deterministic functions that are “invariant” to changes in their arguments (Frisch, 1938).

4. **Causal Relationships** that map the inputs causing each variable:
   \[
   Y = f_Y(X, U, \epsilon_Y); \quad X = f_X(V, \epsilon_X); \quad U = f_U(V, \epsilon_U); \quad V = f_V(\epsilon_V).\]

Econometric approach explicitly models **unobservables** that drive outcomes and produce selection problems. Distribution of unobservables is often the object of study.
**Directed Acyclic Graph (DAG) Representation**

Model: \( Y = f_Y(X, U, \epsilon_Y); X = f_X(V, \epsilon_X); U = f_U(V, \epsilon_U); V = f_V(\epsilon_V). \)

**Notation:**

- **Children:** Variables directly caused by other variables:
  
  \( \text{Ex: } Ch(V) = \{U, X\}, \quad Ch(X) = Ch(U) = \{Y\}. \)

- **Descendants:** Variables that directly or indirectly cause other variables:
  
  \( \text{Ex: } D(V) = \{U, X, Y\}, \quad D(X) = D(U) = \{Y\}. \)

- **Parents:** Variables that directly cause other variables:
  
  \( \text{Ex: } Pa(Y) = \{X, U\}, \quad Pa(X) = Pa(U) = \{V\}. \)
Properties of this Causal Framework

- **Recursive Property**: No variable is a descendant of itself.

Why is it useful?

Autonomy + Independent Errors
+ Recursive Property
⇒ Bayesian Network Tools Apply

- **Bayesian Network**: Translates causal links into independence relations using Statistical/Graphical Tools.

- **Statistical/Graphical Tools**:
  1. Local Markov Condition (**LMC**): A variable is independent of its non-descendants conditioned on its parents;
  2. Graphoid Axioms (**GA**): A set of independence and conditional independence relationships. Dawid (1979)
Local Markov Condition (LMC) (Kiiveri, 1984; Lauritzen, 1996)

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

\[
\text{LMC : } T \perp \perp \mathcal{T} \setminus (D(T) \cup T) | \text{Pa}(T) \ \forall \ T \in \mathcal{T}.
\]

Graphoid Axioms (GA) (Dawid, 1979)

Symmetry: \( X \perp \perp Y|Z \Rightarrow Y \perp \perp X|Z. \)

Decomposition: \( X \perp \perp (W, Y)|Z \Rightarrow X \perp \perp Y|Z. \)

Weak Union: \( X \perp \perp (W, Y)|Z \Rightarrow X \perp \perp Y|(W, Z). \)

Contraction: \( X \perp \perp W|(Y, Z) \) and \( X \perp \perp Y|Z \)

\[ \Rightarrow X \perp \perp (W, Y)|Z. \]

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

\[ \Rightarrow X \perp \perp (W, Y)|Z. \]

Redundancy: \( X \perp \perp Y|X. \)
• **Application** of these tools generates:

\[ Y \perp \perp V|(U, X), \ U \perp \perp X|V. \]
Analysis of Counterfactuals: The Fixing Operator

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

<table>
<thead>
<tr>
<th>Standard Model</th>
<th>Model under Fixing</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V = f_V(\epsilon_V)$</td>
<td>$V = f_V(\epsilon_V)$</td>
</tr>
<tr>
<td>$U = f_U(V, \epsilon_U)$</td>
<td>$U = f_U(V, \epsilon_U)$</td>
</tr>
<tr>
<td>$X = f_X(V, \epsilon_X)$</td>
<td>$X = x$</td>
</tr>
<tr>
<td>$Y = f_Y(X, U, \epsilon_Y)$</td>
<td>$Y = f_Y(x, U, \epsilon_Y)$</td>
</tr>
</tbody>
</table>

- **Importance**: Establishes a framework for counterfactuals.

- **Counterfactual**: $Y(x)$ represents outcome $Y$ when $X$ is fixed at $x$.

- **Linear Case**: $Y = X\beta + U + \epsilon_Y$ and $Y(x) = x\beta + U + \epsilon_Y$.
Joint Distributions

1 Model Representation under Fixing:

\[ Y = f_Y(x, U, \epsilon_Y); \quad X = x; \quad U = f_U(V, \epsilon_U); \quad V = f_V(\epsilon_V). \]

2 Standard Joint Distribution Factorization:

\[ Pr(Y, V, U|X = x) = Pr(Y|U, V, X = x)Pr(U|V, X = x)\cdot Pr(V|X = x). \]

\[ = Pr(Y|U, V, X = x)Pr(U|V)Pr(V|X = x) \]

because \( U \perp X|V \) by LMC.

\[ = Pr(Y|U, X = x)Pr(U|V)Pr(V|X = x) \]

because \( Y \perp V|X, U \) by LMC.

3 Factorization under Fixing \( X \) at \( x \):

\[ Pr(Y, V, U|X \text{ fixed at } x) = Pr(Y|U, X = x)Pr(U|V)Pr(V). \]
• **Conditioning** $X$ at $x$ affects the distribution of $V$.
• **Fixing** $X$ at $x$ does **not** affect the distribution of $V$. 
The definition of causal model permits the following operations:

1. Through **iterated substitution** we can represent all variables as functions of error terms.
2. This representation **clarifies** the concept of fixing.
### Representing the Models Through Their Error Terms

<table>
<thead>
<tr>
<th>Standard Model</th>
<th>Model under Fixing</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V = f_V(\epsilon_V)$</td>
<td>$V = f_V(\epsilon_V)$</td>
</tr>
<tr>
<td>$U = f_U(f_V(\epsilon_V), \epsilon_U)$</td>
<td>$U = f_U(f_V(\epsilon_V), \epsilon_U)$</td>
</tr>
<tr>
<td>$X = f_X(f_V(\epsilon_V), \epsilon_X)$</td>
<td>$X = x$</td>
</tr>
</tbody>
</table>

### Outcome Equation

**Standard Model:** $Y = f_Y(f_X(f_V(\epsilon_V), \epsilon_X), f_U(f_V(\epsilon_V), \epsilon_U), \epsilon_Y)$.

**Model under Fixing:** $Y = f_Y(x, f_U(f_V(\epsilon_V), \epsilon_U), \epsilon_Y)$. 

James Heckman and Rodrigo Pinto

Causal Analysis
Understanding the Fixing Operator

1. Cumulative error distribution function: $Q_\epsilon$.

2. **Conditioning**: $(Y = f_Y(f_X(f_U(\epsilon_U), \epsilon_X), f_U(\epsilon_U), \epsilon_Y))$

   \[
   \therefore E(Y|X=x) = \int_A f_Y(f_X(f_V(\epsilon_V), \epsilon_X), f_U(f_V(\epsilon_V), \epsilon_U), \epsilon_Y) \frac{dQ_\epsilon(\epsilon)}{\int_A dQ_\epsilon}
   \]

   \[A = \{\epsilon ; f_X(f_V(\epsilon_V), \epsilon_X) = x\}\]

   Imposes restrictions on values of error terms.

3. **Fixing**: $(Y = f_Y(x, \epsilon_X), f_U(\epsilon_U), \epsilon_Y))$

   \[
   \therefore E(Y(x)) = \int f_Y(x, \epsilon_X), f_U(f_V(\epsilon_V), \epsilon_U), \epsilon_Y) dQ_\epsilon(\epsilon)
   \]

   Imposes no restriction on values assumed by the error terms.
Fixing ≠ Conditioning

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

\[ Y \text{ conditioned on } X : \quad Y|X = x \]

Linear Case: \[ E(Y|X = x) = x\beta + E(U|X = x); \]
\[ E(\epsilon_Y|X = x) = 0. \]

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

\[ Y \text{ when } X \text{ is fixed at } x \Rightarrow Y(x) = f_Y(x, U, \epsilon_Y) \]

Linear Case: \[ E(Y(x)) = x\beta + E(U); \]
\[ E(\epsilon_Y) = 0. \]
Average Causal Effects: $X$ is fixed at $x, x'$:

$$ATE = E(Y(x)) - E(Y(x'))$$
Fixing: A Causal (not statistical) Operation

- **Problem:** Fixing is a Causal Operation defined **Outside** of standard statistics.

- **Not Readily Comprehended by Statisticians:** Its justification/representation does not follow from standard statistical arguments.

- **Consequence:** Frequent source of **confusion** in statistical discussions.

- **Question:** How can we make statistics converse with causality?

- **Solution:** *Think outside the box*
Thinking Outside the Box

- This was Haavelmo’s great insight.
  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** the motto “Causality is in the Mind.”
  6. **Clarifies** the notion of identification.

**Identification:**
Expresses causal parameters defined in the hypothetical model using probabilities of the empirical model that govern the data-generating process.
2. Empirical and Hypothetical Causal Models
The Hypothetical Model

Formalizing Haavelmo’s Insight

**Empirical Model:** Governs the data-generating process.

**Hypothetical Model:** Abstract model used to examine causality.

The hypothetical model has the following properties:

1. **Same** set of structural equations as the empirical model.
2. **Appends** a hypothetical variable 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.
The Hypothetical Model

- **Hypothetical Variable:** \( \tilde{X} \) replaces the \( X \)-inputs of structural equations.

- **Characteristic:** \( \tilde{X} \) is an external variable; i.e., no parents.

- **Usage:** Hypothetical variable \( \tilde{X} \) enables analysts to examine fixing using standard tools of probability.

- **Notation:**
  1. **Empirical Model:** \((T_E, Pa_E, D_E, Ch_E, Pr_E, E_E)\) denote the variable set, parents, descendants, children, probability and expectation of the empirical model.
  2. **Hypothetical Model:** \((T_H, Pa_H, D_H, Ch_H, Pr_H, E_H)\) denote the variable set, parents, descendants, children, probability and expectation of the hypothetical model.
The hypothetical model is not a speculative departure from the empirical data-generating process but an expanded version of it.
Example of the Hypothetical Model for Fixing $X$

The Associated Hypothetical Model

$$Y = f_Y(\tilde{X}, U, \epsilon_Y); \ X = f_X(V, \epsilon_X); \ U = f_U(V, \epsilon_U); \ V = f_V(\epsilon_V).$$

<table>
<thead>
<tr>
<th>Empirical Model</th>
<th>Hypothetical Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Empirical Model Diagram" /></td>
<td><img src="image2" alt="Hypothetical Model Diagram" /></td>
</tr>
</tbody>
</table>

LMC

| $Y \perp\!\!\!\!\perp V|(U, X)$ | $Y \perp\!\!\!\!\perp (X, V)|(U, \tilde{X})$ |
| $U \perp\!\!\!\!\perp X|V$ | $U \perp\!\!\!\!\perp (X, \tilde{X})|V$ |
| $\tilde{X} \perp\!\!\!\!\perp (U, V, X)$ | $X \perp\!\!\!\!\perp (U, Y, \tilde{X})|V$ |
Benefits of a Hypothetical Model

- **Formalizes** Haavelmo’s insight of Hypothetical variation;

- **Statistical Analysis:** Bayesian Network Tools apply (Local Markov Condition; Graphoid Axioms);

- **Clarifies** the definition of causal parameters;
  1. Causal parameters are defined under the hypothetical model;
  2. Observed data is generated through the empirical model;

- **Distinguishes** definition from 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.
Benefits of a Hypothetical Model

1. **Versatility:** Targets causal links, not variables.

2. **Simplicity:** Does not require definition of any statistical operation outside the realm of standard statistics.

3. **Completeness:** Automatically generates Pearl’s do-calculus when it applies (Pinto 2013).

**Most Important**

Fixing in the empirical model is translated to statistical conditioning in the hypothetical model:

\[
E_E(Y(x)) = E_H(Y|\tilde{X} = x)
\]

Causal Operation Empirical Model  Statistical Operation Hypothetical Model
Identification

- *Hypothetical Model* allows analysts to define and examine causal parameters.

- *Empirical Model* generates observed/unobserved data.

- **Identification**: Express causal parameters of the hypothetical model through observed probabilities in the empirical model.

- **Requires**: connecting probability distributions of *Hypothetical* and *Empirical* Models.
How to Connect the Empirical and Hypothetical Models?

1. By **sharing** the same error terms and structural equations, conditional probabilities of some variables of the hypothetical model can be written in terms of the probabilities of the empirical model.

2. Conditional **independence properties** of the variables in the hypothetical model facilitate connecting the hypothetical and empirical models.
Connecting the Hypothetical and Empirical Models

**Same** error terms and structural equations generate the following relationships:

1. Distribution of **non-children** of $\tilde{X}$ (i.e., $T \in \mathcal{T}_E \setminus Ch_H(\tilde{X})$) are the same in hypothetical and empirical models:

   $$Pr_H(T | Pa_H(T)) = Pr_E(T | Pa_E(T)), T \in (\mathcal{T}_E \setminus Ch_H(\tilde{X}))$$

2. Distribution of **children** of $\tilde{X}$ (i.e. $T \in Ch_H(\tilde{X})$) are the same in hypothetical and empirical models whenever $X$ and $\tilde{X}$ are conditioned on the same value $x$:

   $$Pr_H(T | Pa_H(T) \setminus \{\tilde{X}\}, \tilde{X} = x) = Pr_E(T | Pa_E(T) \setminus \{X\}, X = x)$$
Connecting the Empirical and Hypothetical Models

Thus

1. Distribution of non-descendants of $\tilde{X}$ are the same in the hypothetical and empirical models.

2. Distributions of variables conditional on $X$ and $\tilde{X}$ at the same value of $x$ in the empirical model and in the hypothetical model are the same.

3. Distribution of an outcome $Y \in \mathcal{T}_E$ when $X$ is fixed at $x$ is the same as the distribution of $Y$ conditional on $\tilde{X} = x$ in $Y \in \mathcal{T}_H$. 
Connecting the Hypothetical and Empirical Models

1 **L-1:** Let $W, Z$ be any disjoint set of variables in $\mathcal{T}_E \setminus D_H(\tilde{X})$ then:

$$Pr_H(W|Z) = Pr_H(W|Z, \tilde{X}) = Pr_E(W|Z) \forall \{W, Z\} \subset \mathcal{T}_E \setminus D_H(\tilde{X})$$

2 **T-1:** Let $W, Z$ be any disjoint set of variables in $\mathcal{T}_E$ then:

$$Pr_H(W|Z, X = x, \tilde{X} = x) = Pr_E(W|Z, X = x) \forall \{W, Z\} \subset \mathcal{T}_E$$

3 **C-1:** Let $\tilde{X}$ be uniformly distributed in the support of $X$ and let $W, Z$ be any disjoint set of variables in $\mathcal{T}_E$. Then

$$Pr_H(W|Z, X = \tilde{X}) = Pr_E(W|Z) \forall \{W, Z\} \subset \mathcal{T}_E$$

4 **Matching:** Let $Z, W$ be any disjoint set of variables in $\mathcal{T}_E$ such that, in the hypothetical model, $X \perp\!\!\!\!\!\!\perp W|(Z, \tilde{X})$, then

$$Pr_H(W|Z, \tilde{X} = x) = Pr_E(W|Z, X = x).$$
Matching: A Solution to the Problem of Causal Inference that Connects the Empirical and Hypothetical Models

Matching

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

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

$$E_E(Y(x)) = \int E_H(Y | V = v, \tilde{X} = x) dQ_V(v)$$

In Hypothetical Model

$$= \int E_E(Y | V = v, X = x) dQ_V(v)$$

In Empirical Model

But if $V$ is not observed, the model is not identified without further assumptions.
Some Remarks on Our Causal Framework

- Statistical relationships among variables emerge as **consequences** of applying LMC and GA.
- Causal effects **are associated** by the causal links replaced by hypothetical variables.
- Our framework allows for the same hypothetical variable to operate through **distinct causal effects** (such as **mediation**).
Models for Mediation Analysis
(Keep $\epsilon, U, V$ Implicit)

1. Empirical Model

2. Total Effect of $X$ on $Y$

3. Indirect Effect of $X$ on $Y$

4. Direct Effect of $X$ on $Y$

James Heckman and Rodrigo Pinto
Causal Analysis
Notation for the Do-calculus

More notation is needed to define these rules:

DAG Notation

Let $X$, $Y$, $Z$ be arbitrary disjoint sets of variables (nodes) in a causal graph $G$.

- $G_{\overline{X}}$: DAG that modifies $G$ by deleting the arrows pointing to $X$. $\overline{X} \iff X \leftarrow$
- $G_{X}$: DAG that modifies $G$ by deleting arrows emerging from $X$. $\overline{X} \iff X \rightarrow$
- $G_{\overline{X}, Z}$: DAG that modifies $G$ by deleting arrows pointing to $X$ and emerging from $Z$. 
Examples of DAG Notation

$G$

$G_X$

$G_{X}$

$G_{X, U}$
Do-calculus Rules

- Assumes the Local Markov Condition and independence of $\epsilon$.

Let $G$ be a DAG and let $X$, $Y$, $Z$, $W$ be any disjoint sets of variables. The do-calculus rules are:

- **Rule 1:** *Insertion/deletion of observations (better: “variables”):*
  
  $Y \perp\!\!\!\!\perp Z|(X, W)$ under $G_{\overline{X}}$
  
  $\Rightarrow Pr(Y|do(X), Z, W) = Pr(Y|do(X), W)$.

- **Rule 2:** *Action/observation exchange:*
  
  $Y \perp\!\!\!\!\perp Z|(X, W)$ under $G_{\overline{X}, Z}$
  
  $\Rightarrow Pr(Y|do(X), do(Z), W) = Pr(Y|do(X), Z, W)$.

- **Rule 3:** *Insertion/deletion of actions:*
  
  $Y \perp\!\!\!\!\perp Z|(X, W)$ under $G_{\overline{X}, \overline{Z(W)}}$
  
  $\Rightarrow Pr(Y|do(X), do(Z), W) = Pr(Y|do(X), W)$,

  where $Z(W)$ is the set of $Z$-nodes that are not ancestors of any $W$-node in $G_{\overline{X}}$. 
Graphs for Applying the Do-Calculus

\[ G \]

\[ G_{\bar{X}} \]

\[ G_{\bar{X}, Z} \]

\[ G_{\bar{X}, Z(W)} \]
Let $G$ be a DAG; then, for any disjoint sets of variables $X, Y, Z, W$:

**Rule 1: Insertion/deletion of observations**

If $Y \perp \perp Z|(X, W)$ under $G_X$ then

$Pr(Y|do(X), Z, W) = Pr(Y|do(X), W)$

**Simple Example**

Statistical Relation

Graphical Criterion

Equivalent Probability Expression
4. Comparing the do-Calculus and Haavelmo’s Approach
Application of the do-calculus to a DAG entails several distinct steps:

1. First, the DAG of interest is specified.
2. Second, graphs generated by the “bar operations,” e.g., $G_{\bar{X}}$ or $G_{\overline{X}}$, etc., are created.
3. Third, the Local Markov Condition is applied to generate the implied conditional independence relationships among the variables.
4. Fourth, the Rules 1–3 are applied to define the implied causal parameters, if any.
5. Step 5, examine if there are empirical counterparts to the components of the hypothetical model so that those components are identified.
• It only examines the consequences of conditional independence. See Figure 1.
Figure 1: Illustrating the Do-Calculus

Step 1

Original DAG

Step 2

Possible subgraphs

G_X

Step 3

LMC

G_Y

Conditional Independence Relationships

Step 4

Apply

Rule 1

Rule 2

Rule 3

Step 5

Implied Causal Parameters (if any)

Step 6

Identification:
Find Empirical Counterparts (if any)
Cumbersome Approach to Defining and Identifying Causal Parameters

- Application of these steps can, in principle, generate a variety of causal parameters which may or may not be identified.
- Many applications of the do-calculus do not clearly separate identification of causal parameters from their definition.
- That separation is a hallmark of the Haavelmo approach.
- For high-dimensional graphs, there can be a high-dimensional set of subgraphs and implied causal parameters.
- To make the discussion specific, it is fruitful to apply the three rules to a DAG for the IV model depicted in Table 2.
- All variables are assumed to be scalar.
Table 2: An IV Model

<table>
<thead>
<tr>
<th>Empirical Model</th>
<th>Modified Empirical Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G )</td>
<td>( G_X )</td>
</tr>
<tr>
<td>( Z )</td>
<td>( Z )</td>
</tr>
<tr>
<td>( X )</td>
<td>( X )</td>
</tr>
<tr>
<td>( U )</td>
<td>( U )</td>
</tr>
<tr>
<td>( Y )</td>
<td>( Y )</td>
</tr>
</tbody>
</table>
Applying Rule 1

- For $G_X$, confirm that $Y \perp \!\!\!\!\!\perp Z|X$.
- LMC applied to $Z \Rightarrow Z \perp \!\!\!\!\!\perp (U, X, Y)$.
- Thus $Y \perp \!\!\!\!\!\perp Z|X$ from symmetry and weak union properties of graphoid axioms.
- By Rule 1,

$$Pr(Y|do(X), Z) = Pr(Y|do(X)).$$

Thus,

$$Y \perp \!\!\!\!\!\perp Z|do(X).$$
• This defines the causal parameter associated with $G_X$ but by itself does not identify it.

• Pearl (2009, chapters 3 and 5) acknowledges that this IV model is not identified with his rules.
Using the Haavelmo approach, we create a hypothetical model shown in Figure 2.

Figure 2: Haavelmo’s IV Model

where $\tilde{X} \perp \perp (Y, U, X, Z)$. 
• Without application of any of Pearl’s rules, it is immediate, applying LMC to $Z$,

\[ Y \perp\!
\perp Z|\tilde{X} \]

and the causal parameter $Pr(Y|\tilde{X})$ is well defined.

• Identified using standard IV methods widely used in econometrics (see, e.g., Matzkin, 2013).
Applying Rule 2

Illustrate the application of Rule 2. Consider the graph $G_{\overline{X},Z}$ depicted in Figure 3.

Figure 3: Graph for $G_{\overline{X},Z}$
In this case, the graphs for $G_{X,Z}$ and for $G_{X}$ coincide.

By Rule 2, since $Y \perp \perp Z|X$,

$$Pr(Y|do(X), do(Z)) = Pr(Y|do(X), Z).$$

By Rule 1, it follows that

$$Pr(Y|do(X), Z) = Pr(Y|do(X)).$$
Applying Rule 2 again

It is also instructive to consider the application of Rule 2 to $G_{U,X}$, shown in Figure 4.

Figure 4: Graph for $G_{U,X}$
From $Y \perp X \mid U$ and Rule 2,

$$Pr(Y \mid do(U), do(X)) = Pr(Y \mid do(U), X).$$

(1)
Next consider $G_U$:

**Figure 5**: Graph for $G_U$

![Graph for $G_U$](image)
From $Y \perp \!\!\!\!\!\!\perp U|X$ (from applying Rule 1 when $U$ takes the place of $Z$ and $X$ takes the place of $W$),

$$\Pr(Y|do(U), X) = \Pr(Y|U, X).$$  \hfill (2)

Putting together equations (1) and (2), it follows

$$\Pr(Y|do(U), do(X)) = \Pr(Y|U, X).$$

Justifies matching on $U$ and $X$ to identify the causal parameter $\Pr(Y|do(U), do(X))$. 
Haavelmo’s Approach

- Compare this analysis to one based on Haavelmo’s insights.
- It follows immediately from the graph in Figure 2 that $Y \perp \perp X|U, \tilde{X}$.
- Thus,

$$Pr(Y|U, \tilde{X}, X) = Pr(Y|U, X),$$

and matching identifies the causal parameter in a straightforward way.
Rule 3

Consider the analysis of $G_{X, Z(W)}$, where we add $W$ to illustrate the interpretation of $Z(W)$. See Figure 6.

Figure 6: Graph for $G_{X, Z(W)}$
• There is no link from $W$ to $Z$ in this graph because it is based on $Z(W)$.

• Applying Rule 3, because $Y \perp \perp Z|(X, W)$,

$$Pr(Y|do(X), do(Z), W) = Pr(Y|do(X), W).$$
Haavelmo’s Approach

- The definition of the causal parameter $Pr(Y|\tilde{X}, W)$ is an immediate consequence of $\tilde{X} \perp\!\!\!\!\!\perp X, U, W$. 
Another Simple Do-Calculus Exercise

1. LMC applied to $X$ under $G_X$ generates

$$X \perp \perp (U, Y) \mid V.$$ 

2. Now if $X \perp \perp (U, Y) \mid V$ holds under $G_X$, then, by Rule 2

$$Pr(Y \mid do(X), V) = Pr(Y \mid X, V). \tag{3}$$

$$\therefore E(Y \mid do(X) = x) = \int E(Y \mid V = v, do(X) = x)dQ_V(v)$$

Using do(X); i.e., fixing $X$

$$= \int E(Y \mid V = v, X = x)dQ_V(v)$$

Replace “do” with Standard Statistical Conditioning by Equation (3)
The “Front-Door Model”
Table 3: “Front-Door” Empirical and Hypothetical Models

1. Pearl’s “Front-Door” Empirical Model

- $\mathcal{T} = \{U, X, M, Y\}$
- $\epsilon = \{\epsilon_U, \epsilon_X, \epsilon_M, \epsilon_Y\}$
- $Y = f_Y(M, U, \epsilon_Y)$
- $X = f_X(U, \epsilon_X)$
- $M = f_M(X, \epsilon_M)$
- $U = f_U(\epsilon_U)$

2. Our Version of the “Front-Door” Hypothetical Model

- $\mathcal{T} = \{U, X, M, Y, \tilde{X}\}$
- $\epsilon = \{\epsilon_U, \epsilon_X, \epsilon_M, \epsilon_Y\}$
- $Y = f_Y(M, U, \epsilon_Y)$
- $X = f_X(U, \epsilon_X)$
- $M = f_M(\tilde{X}, \epsilon_M)$
- $U = f_U(\epsilon_U)$
Table 3 (continued): “Front-Door” Empirical and Hypothetical Models

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Empirical Model</th>
<th>Hypothetical Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pa(U) = ∅</td>
<td>Pa(U) = Pa(\tilde{X}) = ∅</td>
<td>Pa(U) = ∅</td>
</tr>
<tr>
<td>Pa(X) = {U}</td>
<td>Pa(X) = ∅</td>
<td>Pa(X) = {U}</td>
</tr>
<tr>
<td>Pa(M) = {X}</td>
<td>Pa(M) = {\tilde{X}}</td>
<td>Pa(M) = {\tilde{X}}</td>
</tr>
<tr>
<td>Pa(Y) = {M, U}</td>
<td>Pa(Y) = {M, U}</td>
<td>Pa(Y) = {M, U}</td>
</tr>
<tr>
<td>Y \perp\perp X</td>
<td>(M, U)</td>
<td>Y \perp\perp (\tilde{X}, X)</td>
</tr>
<tr>
<td>M \perp\perp U</td>
<td>X</td>
<td>M \perp\perp (U, X)</td>
</tr>
<tr>
<td></td>
<td>X \perp\perp (M, \tilde{X}, Y)</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>\tilde{X} \perp\perp (X, U)</td>
<td></td>
</tr>
<tr>
<td>\Pr_E(Y, M, X, U) =</td>
<td>\Pr_H(Y, M, X, U, \tilde{X}) =</td>
<td></td>
</tr>
<tr>
<td>\Pr_E(Y</td>
<td>M, U)\Pr_E(X</td>
<td>U)\Pr_E(M</td>
</tr>
<tr>
<td>\Pr_E(Y, M, U</td>
<td>do(X) = x) =</td>
<td>\Pr_H(Y, M, U, X</td>
</tr>
<tr>
<td>\Pr_E(Y</td>
<td>M, U)\Pr_E(M</td>
<td>X = x)\Pr_E(U)</td>
</tr>
</tbody>
</table>
• The do-calculus identifies $Pr(Y|do(X))$ through four steps.
• Steps 1, 2, and 3 identify $Pr(M|do(X))$, $Pr(Y|do(M))$, and $Pr(Y|M, do(X))$, respectively.
• Step 4 uses the first three steps to identify $Pr(Y|do(X))$. 
Application of the Rules
Step 1

- Invoking LMC for variable $M$ of DAG $G_X$, (DAG 1 of Table 4) generates $X \perp\!\!\!\!\!\!\!\perp M$.
- Thus, by Rule 2 of the do-calculus, we obtain $Pr(M|do(X)) = Pr(M|X)$.  

James Heckman and Rodrigo Pinto  
Causal Analysis
Table 4: Do-calculus and the Front-Door Model

<table>
<thead>
<tr>
<th>1. Modified Front-Door Model $G_X = G_M$</th>
<th>2. Modified Front-Door Model $G_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Diagram" /></td>
<td><img src="image2.png" alt="Diagram" /></td>
</tr>
<tr>
<td>$(Y, M) \perp \perp X \mid U$</td>
<td>$(X, M) \perp \perp Y \mid U$</td>
</tr>
<tr>
<td>$(X, U) \perp \perp M$</td>
<td>$(Y, U) \perp \perp M \mid X$</td>
</tr>
<tr>
<td><strong>3. Modified Front-Door Model $G_{X,M}$</strong></td>
<td><strong>4. Modified Front-Door Model $G_{X,M}$</strong></td>
</tr>
<tr>
<td><img src="image3.png" alt="Diagram" /></td>
<td><img src="image4.png" alt="Diagram" /></td>
</tr>
<tr>
<td>$(X, M) \perp \perp (Y, U)$</td>
<td>$(Y, M, U) \perp \perp X$</td>
</tr>
<tr>
<td></td>
<td>$U \perp \perp M$</td>
</tr>
</tbody>
</table>
Step 2

- Invoking LMC for variable $M$ of DAG $G_M$, (DAG 1 of Table 4) generates $X \perp\!\!\!\!\!\!\!\perp M$.
- Thus, by Rule 3 of the do-calculus, $Pr(X|do(M)) = Pr(X)$.
- In addition, applying LMC for variable $M$ of DAG $G_M$, (DAG 2 of Table 4) generates $M \perp\!\!\!\!\!\!\!\!\perp Y|X$.
- Thus, by Rule 2 of the do-calculus, $Pr(Y|X, do(M)) = Pr(Y|X, M)$. 
Therefore, \( Pr(Y|do(M)) = \sum_{x' \in \text{supp}(X)} Pr(Y|X = x', do(M)) Pr(X = x'|do(M)) \)

\[ = \sum_{x' \in \text{supp}(X)} Pr(Y|X = x', M) Pr(X = x'), \]

where "supp" means support.
Step 3

• Invoking LMC for variable $M$ of DAG $G_{X,M}$, (DAG 3) generates $Y \perp\!\!\!\!\!\!\!\!\!\!\perp M|X$.

• Thus, by Rule 2 of the do-calculus,
  $Pr(Y|M, do(X)) = Pr(Y|do(M), do(X))$.

• In addition, applying LMC for variable $X$ of DAG $G_{X,M}$, (DAG 4 of Table 4) generates $(Y, M, U) \perp\!\!\!\!\!\!\!\!\!\!\perp X$.

• By weak union and decomposition, we obtain $Y \perp\!\!\!\!\!\!\!\!\!\!\perp X|M$.

• Thus, by Rule 3 of the do-calculus, we obtain that
  $Pr(Y|do(X), do(M)) = Pr(Y|do(M))$.

• Thus,
  $Pr(Y|M, do(X)) = Pr(Y|do(M), do(X)) = Pr(Y|do(M))$. 
Step 4

• Collect the results from the three previous steps to identify $Pr(Y|do(X))$ from observed data:

\[
Pr(Y|do(X) = x) = \sum_{m \in \text{supp}(M)} Pr(Y|do(M) = m, do(X) = x) Pr(M = m|do(X) = x)
\]

• In this fashion, we use the do-calculus to identify the desired causal parameter.
Lemma 1

*In the Front-Door hypothetical model,*

(1) \( Y \perp\!\!\!\!\!\!\!\!\perp \tilde{X} \mid M \),

(2) \( X \perp\!\!\!\!\!\!\!\!\perp M \),

and

(3) \( Y \perp\!\!\!\!\!\!\!\!\perp \tilde{X} \mid (M, X) \).
Proof.

- By LMC for $X$, we obtain $(Y, M, \tilde{X}) \perp\!
\perp X\mid U$.

- By LMC for $Y$ we obtain $Y \perp\!
\perp (X, \tilde{X})\mid(M, U)$.

- By Contraction applied to $(Y, M, \tilde{X}) \perp\!
\perp X\mid U$ and $Y \perp\!
\perp (X, \tilde{X})\mid(M, U)$ we obtain $(Y, X) \perp\!
\perp \tilde{X}\mid(M, U)$.

- By LMC for $U$ we obtain $(M, \tilde{X}) \perp\!
\perp U$.

- By Contraction applied to $(M, \tilde{X}) \perp\!
\perp U$ and $(Y, M, \tilde{X}) \perp\!
\perp X\mid U$ we obtain $(X, U) \perp\!
\perp (M, \tilde{X})$.

- The second relationship in the Lemma is obtained by Decomposition.

- In addition, by Contraction on $(Y, X) \perp\!
\perp \tilde{X}\mid(M, U)$ and $(M, \tilde{X}) \perp\!
\perp U$ we obtain $(Y, X, U) \perp\!
\perp \tilde{X}\mid M$.

- The two remaining conditional independence relationships of the Lemma are obtained by Weak Union and Decomposition.
Applying these results,

\[ Pr_H(Y|\tilde{X} = x) \]

\[ = \sum_{m \in \text{supp}(M)} Pr_H(Y|M = m, \tilde{X} = x) Pr_H(M = m|\tilde{X} = x) \]

\[ = \sum_{m \in \text{supp}(M)} Pr_H(Y|M = m) Pr_H(M = m|\tilde{X} = x) \]

\[ = \sum_{m \in \text{supp}(M)} \left( \sum_{x' \in \text{supp}(X)} Pr_H(Y|X = x', M = m) Pr_H(X = x'|M = m) \right) Pr_H(M = m|\tilde{X} = x) \]

\[ = \sum_{m \in \text{supp}(M)} \left( \sum_{x' \in \text{supp}(X)} Pr_H(Y|X = x', \tilde{X} = x', M = m) Pr_H(X = x'|M = m) \right) Pr_H(M = m|\tilde{X} = x) \]

\[ = \sum_{m \in \text{supp}(M)} \left( \sum_{x' \in \text{supp}(X)} \frac{Pr_E(Y|M, X = x') Pr_E(X = x')}{\text{by T-1}} \right) \frac{Pr_E(M = m|X = x)}{\text{by Matching}}. \]
• The second equality comes from relationship (1) \( Y \perp \perp \tilde{X}|M \) of Lemma 1.

• The fourth equality comes from relationship (2) \( X \perp \perp M \) of Lemma 1.

• The fifth equality comes from relationship (3) \( Y \perp \perp \tilde{X}|(M, X) \) of Lemma 1.

• The last equality links the distributions of the hypothetical model with the ones of the empirical model.

• The first term uses T-1 to equate
\[
Pr_H(Y|X = x', \tilde{X} = x', M = m) = Pr_E(Y|M, X = x').
\]

• The second term uses the fact that \( X \) is not a child of \( \tilde{X} \); thus, by L-1,
\[
Pr_H(X = x') = Pr_E(X = x').
\]

• Finally, the last term uses Matching applied to \( M \). Namely, LMC for \( M \) generates \( M \perp X|\tilde{X} \) in the hypothetical model. Then, by Matching,
\[
Pr_H(M|\tilde{X} = x) = Pr_E(M|X = x).
\]
Summary
The Do-calculus

- **Stays within the box of the empirical model.**
- **Attempt:** Counterfactual manipulations using the empirical model.
- **Goal:** Expressions obtained from a hypothetical model.
- **Tools:** Uses causal/graphical/statistical rules outside statistics.
- **Fixing:** Uses $do(X) = x$ for fixing $X$ at $x$ in the DAG for all $X$-inputs (does not allow to target causal links separately).
- **Flexibility:** Does not easily define complex treatments, such as treatment on the treated; i.e.,

$$E_E(Y|X = 1, \tilde{X} = 1) - E_E(Y|X = 1, \tilde{X} = 0).$$

**Claim:** Identification using the hypothetical model is transparent and does not require additional causal rules, only standard statistical tools.
The Do-calculus: Summary

1. **Haavelmo approach**: Thinks outside the *causal* box.
2. **Do-calculus**: Requires complex new statistical rules.
Characteristics of Pearl’s Do-Calculus

1. **Information:** DAG only provides information on the causal relationships among variables. LMC and GA give conditional independence relationships.

2. **Not Suited** for examining implications for identification from assumptions from functional forms, even non-parametric functional form restrictions.

3. **Identification:** If the information in a DAG is sufficient to identify causal effects, then:

   There exists a **sequence** of application of the rules of the Do-Calculus that **generates** a formula for causal effects based on observational quantities (Huang and Valtorta 2006, Shpitser and Pearl 2006). “Completeness” of the DAG.
Characteristics of Pearl’s Do-Calculus

4 \textbf{Limitation:} Does not allow for additional information outside the DAG framework.

i \textbf{Only} applies to the information content of a DAG.

ii \textbf{IV} is not identified through Do-calculus

iii \textbf{Why?} Requires assumptions outside DAGs: linearity, monotonicity, separability.

\underline{Link to Appendix I}
The “do-calculus” cannot identify the Generalized Roy Model.

Link to Appendix II
# 5. Summarizing Do-calculus of Pearl (2009) and Haavelmo's Framework

## Common Features of Haavelmo and Do-Calculus:

<table>
<thead>
<tr>
<th>Autonomy</th>
<th>(Frisch, 1938)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errors Terms:</td>
<td>$\epsilon$ mutually independent</td>
</tr>
<tr>
<td>Statistical Tools:</td>
<td>LMC and Graphoid Axioms apply</td>
</tr>
<tr>
<td>Counterfactuals:</td>
<td>Fixing or Do-operator is a causal, not statistical, operation</td>
</tr>
</tbody>
</table>

## Distinctive Features of Haavelmo and Do-Calculus:

<table>
<thead>
<tr>
<th><strong>Haavelmo</strong></th>
<th><strong>Do-calculus</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Approach:</td>
<td>Thinks outside the box of the empirical model by constructing a new hypothetical model motivated by, but distinct from, the empirical model where fixing can be analyzed using standard tools of probability</td>
</tr>
<tr>
<td></td>
<td>Stays inside the box of the empirical model; creates complex nonstandard rules to introduce fixing into a probabilistic framework</td>
</tr>
<tr>
<td>Introduces:</td>
<td>Constructs a hypothetical model</td>
</tr>
<tr>
<td></td>
<td>Graphical rules</td>
</tr>
<tr>
<td>Identification:</td>
<td>Connects $Pr_H$ and $Pr_E$</td>
</tr>
<tr>
<td></td>
<td>Iteration of do-calculus rules</td>
</tr>
<tr>
<td>Versatility:</td>
<td>Basic statistical principles apply</td>
</tr>
<tr>
<td></td>
<td>Creates new rules of statistics</td>
</tr>
</tbody>
</table>
6. Hypothetical Models and Simultaneous Equations
The Simultaneous Equations Model

A system of two equations:

\[ Y_1 = g_{Y_1}(Y_2, X_1, U_1) \]  \hspace{1cm} (7a)
\[ Y_2 = g_{Y_2}(Y_1, X_2, U_2). \]  \hspace{1cm} (7b)

- **Variables:** \( T_E = \{ Y_1, Y_2, X_1, X_2, U_1, U_2 \} \).
- **Assumptions:** \( U_1 \perp \perp U_2 \) and \((U_1, U_2) \perp \perp (X_1, X_2)\).
- **LMC** condition breaks down.
- **Matzkin (2008)** relaxes these assumptions and identifies causal effects for \( U_1 \not\perp \!\!\!\!\perp U_2 \) and \((U_1, U_2) \not\perp \!\!\!\!\perp (X_1, X_2)\).

**Most Important**

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).
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

$$Y_1(y_2) = g_{Y_1}(y_2, X_1, U_1).$$

**By Symmetry:**

$$Y_2(y_1) = g_{Y_2}(y_1, X_2, U_2).$$

**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 (7a)-(7b).
- $(\tilde{Y}_1, \tilde{Y}_2) \perp \perp (X_1, X_2, U_1, U_2)$; and $\tilde{Y}_1 \perp \perp \tilde{Y}_2$.
- $\mathcal{T}_H = \{ \tilde{Y}_1, \tilde{Y}_2, Y_1, Y_2, X_1, X_2, U_1, U_2 \}$.
- 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
  \[ Pr_H(Y_1|\tilde{Y}_2 = y_2). \]

- **Average causal effect** of $Y_2$ on $Y_1$ when $Y_2$ is fixed at $y_2$ and $y'_2$ values:
  \[ E_H(Y_1|\tilde{Y}_2 = y_2) - E_H(Y_1|\tilde{Y}_2 = y'_2) \]

- **Notation**: $E_H$ denotes expectation over the probability measure $Pr_H$ of the hypothetical model.
Empirical Model for Simultaneous Equations
Some Hypothetical Models for $Y_2$ and $Y_1$, Respectively
Causal Analysis
Completeness Assumption in Simultaneous Equations

• **Common Assumption**: completeness—the existence of at least a local solution for $Y_1$ and $Y_2$ in terms of $(X_1, X_2, U_1, U_2)$:

$$Y_1 = \phi_1(X_1, X_2, U_1, U_2) \quad (8a)$$

$$Y_2 = \phi_2(X_1, X_2, U_1, U_2). \quad (8b)$$

• **Reduced form** equations (see, e.g., Matzkin, 2008, 2013).

• **Inherit** the autonomy properties of the structural equations.
If $X_1$ and $X_2$ are disjoint, use ILS to identify:

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

From reduced forms:

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

Thus:

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

But cannot be identified by the rules of the do-calculus.

Source: Heckman (2008)
7. Pearl’s Lament
“In almost every one of his recent articles James Heckman stresses the importance of counterfactuals as a necessary component of economic analysis and the hallmark of econometric achievement in the past century. For example, the first paragraph of the HV article reads: ‘they [policy comparisons] require that the economist construct counterfactuals. Counterfactuals are required to forecast the effects of policies that have been tried in one environment but are proposed to be applied in new environments and to forecast the effects of new policies.’ Likewise, in his Sociological Methodology article (2005), Heckman states: ‘Economists since the time of Haavelmo (1943, 1944) have recognized the need for precise models to construct counterfactuals… The econometric framework is explicit about how counterfactuals are generated and how interventions are assigned…’ ”
“And yet, despite the proclaimed centrality of counterfactuals in econometric analysis, a curious reader will be hard pressed to identify even one econometric article or textbook in the past 40 years in which counterfactuals or causal effects are formally defined. Needed is a procedure for computing the counterfactual $Y(x,u)$ in a well-posed, fully specified economic model, with $X$ and $Y$ two arbitrary variables in the model. By rejecting Haavelmo’s definition of $Y(x,u)$, based on surgery, Heckman commits econometrics to another decade of division and ambiguity, with two antagonistic camps working in almost total isolation.”
“Economists working within the potential-outcome framework of the Neyman-Rubin model take counterfactuals as primitive, unobservable variables, totally detached from the knowledge encoded in structural equation models (e.g., Angrist 2004; Imbens 2004). Even those applying propensity score techniques, whose validity rests entirely on the causal assumption of ‘ignorability,’ or unconfoundedness, rarely know how to confirm or invalidate that assumption using structural knowledge....”
“Economists working within the structural equation framework (e.g., Kennedy 2003; Mittelhammer et al. 2000; Intriligator et al. 1996) are busy estimating parameters while treating counterfactuals as metaphysical ghosts that should not concern ordinary mortals. They trust leaders such as Heckman to define precisely what the policy implications are of the structural parameters they labor to estimate, and to relate them to what their colleagues in the potential-outcome camp are doing.... Economists will do well resurrecting the ideas of Haavelmo (1943), Marschak (1950), and Strotz and Wold (1960) and reinvigorating them with the logic of graphs and counterfactuals presented in this book.”
8. Structural Models vs. the Neyman-Rubin Model

Link to Appendix III
Table 5: Comparison of the Neyman-Rubin Approach and the Structural Approach

<table>
<thead>
<tr>
<th></th>
<th>Neyman-Rubin Framework</th>
<th>Structural Framework (originating with Haavelmo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counterfactuals for objective outcomes ($Y_0, Y_1$)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Agent valuations of subjective outcomes</td>
<td>No (choice-mechanism implicit)</td>
<td>Yes</td>
</tr>
<tr>
<td>Models for the causes of potential outcomes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Ex ante versus ex post</em> counterfactuals</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment assignment rules that recognize voluntary nature of participation</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
**Table 6: Comparison of the Neyman-Rubin Approach and the Structural Approach**

<table>
<thead>
<tr>
<th>Evaluation of returns at the margin of various policies</th>
<th>Neyman-Rubin Framework</th>
<th>Structural Framework (originating with Haavelmo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social interactions, general equilibrium effects and contagion</th>
<th>Neyman-Rubin Framework</th>
<th>Structural Framework (originating with Haavelmo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No (assumed away)</td>
<td>Yes (modeled)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Internal validity (problem P1)</th>
<th>Neyman-Rubin Framework</th>
<th>Structural Framework (originating with Haavelmo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>External validity (problem P2)</th>
<th>Neyman-Rubin Framework</th>
<th>Structural Framework (originating with Haavelmo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Forecasting effects of new policies (problem P3)</th>
<th>Neyman-Rubin Framework</th>
<th>Structural Framework (originating with Haavelmo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distributional treatment effects</th>
<th>Neyman-Rubin Framework</th>
<th>Structural Framework (originating with Haavelmo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No(^a)</td>
<td>Yes (for the general case)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analyze relationship between outcomes and choice equations</th>
<th>Neyman-Rubin Framework</th>
<th>Structural Framework (originating with Haavelmo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No (implicit)</td>
<td>Yes (explicit)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) An exception is the special case of common ranks of individuals across counterfactual states: “rank invariance.” See the discussion in Abbring and Heckman (2007).
9. Marschak’s Maxim and Link of Treatment Effect Literature to the Structural Literature
Treatment Effects vs. Economically Interpretable Effects

(Frisch, 1933)
10. Haavelmo’s Contributions to Empirical Economics
11. Conclusion
Examined Haavelmo’s fundamental contributions

- **Distinction** between causation and correlation (first formal analysis).

- **Distinguished** definition of causal parameters (though process of creating hypothetical models) from their identification from data.

- **Explained** that causal effects of inputs on outputs are defined under abstract models that assign independent variation to inputs.

- **Clarified** concepts that are still muddled in some quarters of statistics.

- **Formalizes** Frisch’s notion that causality is in the mind.
Causal Framework Inspired by Haavelmo’s Ideas

- **Contribution:** Causal framework inspired by Haavelmo.
- **Introduce** hypothetical models for examining causal effects.
- **Assigns** independent variation to inputs determining outcomes.
- **Enables** us to discuss causal concepts such as fixing using an intuitive approach.
- **Fixing** is easily translated to statistical conditioning.
- **Eliminates** the need for additional extra-statistical graphical/statistical rules to achieve identification (in contrast with the do-calculus).
- **Identification** relies on evaluating causal parameters defined in the *hypothetical model* using data generated by the *empirical model*.
- **Achieved** by applying standard statistical tools to fundamentally recursive Bayesian Networks.
Beyond DAGs

- We discuss the limitations of methods of identification that rely on the fundamentally recursive approach of Directed Acyclic Graphs.

- Haavelmo’s framework can be extended to the fundamentally non-recursive framework of the simultaneous equations model without violating autonomy.

- Simultaneous equations are fundamentally non-recursive and fall outside of the framework of Bayesian causal nets and DAGs.

- Haavelmo’s approach also covers simultaneous causality whereas other frameworks cannot, except through ad hoc rules such as “shutting down” equations.

- Haavelmo’s framework allows for a variety of econometric methods can be used to secure identification of this class of models (see, e.g., Matzkin, 2012, 2013.)

- Failed to produce an algorithm for learning from the data—remaining an open issue.
• Showed benefits of the structural approach compared to Neyman-Rubin model
• Limitations of Haavelmo’s Neyman-Pearson approach for learning from data
12. Appendix I
IV: Formalized by Reiersol (1945)
Failure of Do-Calculus
Does Not Generate Standard IV Results

The simplest instrumental variable model consists of four variables:

1. A confounding variable $U$ that is external and unobserved,
2. An external instrumental variable $Z$,
3. An observed variable $X$ caused by $U$ and $Z$,
4. An outcome $Y$ caused by $U$ and $X$. 
4.1 Do-Calculus Non-identification of the IV Model

- **Limitation:** IV model is not identified by literature that relies exclusively on DAGs.

- **Why?** IV identification relies on assumptions outside the scope of the DAG literature.

- **LMC** generates the conditional independence relationships: $Y \perp\!\!\!\!\!\!\!\!\perp Z | (U, X)$ and $U \perp\!\!\!\!\!\!\!\!\perp Z$.

- **TSLS:** $X \not\perp\!\!\!\!\!\!\!\!\perp Z$ holds; thus, the IV model satisfies the necessary criteria to apply the method of Two Stage Least Squares (TSLS).

- **Assumption Outside of DAGs:** TSLS identifies the IV model under linearity.
The Do-Calculus applied to the IV Model generates:

1. \( \Pr(Y|\text{do}(X), \text{do}(Z)) = \Pr(Y|\text{do}(X), Z) = \Pr(Y|\text{do}(X)), \)
2. \( \Pr(Y|\text{do}(Z)) = \Pr(Y|Z) \)

Only establishes the exogeneity of the instrumental variable \( Z \). Insufficient to identify \( \Pr(Y|\text{do}(X)) \).
Return to main text.
13. Appendix II
1. Causal Effects and the Do-calculus
Identification of Treatment Effects of a DAG

Pearl’s Do-Calculus:

1. **Purpose:** Identify casual effects from non-experimental data.

2. **Application:** Bayesian network structure, i.e., Directed Acyclic Graph (DAG) that represents causal relationships.

3. **Tools:** Three inference rules that translate graphical relations of a DAG into causal independence conditional relations (Pearl 1995, and Pearl 2000).

4. **Identification method:** Iteration of do-calculus rules to generate a function that describes treatment effects statistics as a function of the observed variables only (Tian and Pearl 2002, Tian and Pearl 2003).
Characteristics of Pearl’s Do-Calculus

**Completeness**

If some causal effect of a DAG is **identifiable**, then there exists a **sequence** of application of the **Do-Calculus rules** that can generate a formula that translates causal effects into an equation that only relies on observational quantities (Huang and Valtorta 2006, Shpitser and Pearl 2006).

**Limitation**

**Only** works for DAGs. **Does not** allow for additional information outside the DAG framework that could generate identification of causal distributions. **Only** applies to the information content of a DAG.
2. The Do-Calculus (Pearl, 1995)
Identifying Causal Effects for Markovian Models

Tools needed:

- **Rules**: A set of graphical/statistical rules that convert expressions of causal inference into probability equations.

- **Complete algorithm**: an algorithm based on these rules that, for any causal effect in question, we can generate an expression involving observed conditional probabilities or report that the causal effect is not identifiable.
Some Notation

DAG Notation
Let $X$, $Y$, $Z$ be arbitrary disjoint sets of variables (nodes) in a causal graph $G$.

- $G_{X}$: DAG that modifies $G$ by deleting the arrows pointing to $X$.
- $G_{X}^{-}$: DAG that modifies $G$ by deleting arrows emerging from $X$.
- $G_{X}^{-,Z}$: DAG that modifies $G$ by deleting arrows pointing to $X$ and emerging from $Z$.

Probability Notation
Probability of $Y$ when $X$ is fixed at $x$ and $Z$ is observed.

$$Pr(Y|do(X) = x, Z) = \frac{Pr(Y, Z|do(X) = x)}{Pr(Z|do(X) = x)}$$
This figure represents causal relations between four variables. Arrows represent direct causal relations. Circles represent unobserved variables. Squares represent observed variables.

The “If” Question

DAG defines causal relations among variables. It answers the question if a variable causes or is caused by any other variable.
Example of DAG Notation

\[ G_X = G_Z \]

\[ G_Z \]

\[ G_{X,Z} \]

\[ G_{X,Z} \]
Three Rules of Do-Calculus (Pearl, 2000)

Let \( G \) be a DAG then for any disjoint sets of variables \( X, Y, Z, W \):

- **Rule 1:** Insertion/deletion of observations
  
  \[
  Y \perp \perp Z|(X, W) \text{ under } G_{\overline{X}} \Rightarrow Pr(Y|do(X), Z, W) = Pr(Y|do(X), W)
  \]

- **Rule 2:** Action/observation exchange
  
  \[
  Y \perp \perp Z|(X, W) \text{ under } G_{\overline{X},\overline{Z}} \Rightarrow Pr(Y|do(X), do(Z), W) = Pr(Y|do(X), Z, W)
  \]

- **Rule 3:** Insertion/deletion of actions
  
  \[
  Y \perp \perp Z|(X, W) \text{ under } G_{\overline{X},\overline{Z(W)}} \Rightarrow Pr(Y|do(X), do(Z), W) = Pr(Y|do(X), W)
  \]

where \( Z(W) \) is the set of \( Z \)-nodes that are not ancestors of any \( W \)-node in \( G_{\overline{X}} \).
Understanding Rules of Do-Calculus

Let $G$ be a DAG then for any disjoint sets of variables $X, Y, Z, W$:

**Rule 1:** Insertion/deletion of observations

If $Y \perp \! \! \! \perp Z \mid (X, W)$ under $G_X$ then

$$Pr(Y \mid do(X), Z, W) = Pr(Y \mid do(X), W)$$

Equivalent Probability Expression
3. Do-Calculus Exercise
Using the Do-Calculus: Task 1 – Compute $Pr(Z|do(X))$

$X \perp Z$ in $G_X$, by Rule 2, $Pr(Z|do(X)) = Pr(Z|X)$.
Using the Do-Calculus: Task 2 – Compute $Pr(Y|do(Z))$

$Z \perp\!\!\!\!\!\!\!\!\!\!\!\!\perp X$ in $G_{\overline{Z}}$, by Rule 3, $Pr(X|do(Z)) = Pr(X)$

$Z \perp\!\!\!\!\!\!\!\!\!\!\!\!\perp Y|X$ in $G_{\overline{Z}}$, by Rule 2, $Pr(Y|X, do(Z)) = Pr(Y|X, Z)$

$\therefore Pr(Y|do(Z)) = \sum\limits_X Pr(Y|X, do(Z)) Pr(X|do(Z))$

$= \sum\limits_X Pr(Y|X, Z) Pr(X)$
Using the Do-Calculus: Task 3 – Compute $Pr(Y|Z, do(X))$

$Y \perp \perp Z|X$ in $G_{X,Z}$, by Rule 2,

$$Pr(Y|Z, do(X)) = Pr(Y|do(Z), do(X))$$

$Y \perp \perp X|Z$ in $G_{X,Z}$, by Rule 3,

$$Pr(Y|do(X), do(Z)) = Pr(Y|do(Z))$$

$\therefore Pr(Y|Z, do(X)) = Pr(Y|do(Z), do(X)) = Pr(Y|do(Z))$
Using the Do-Calculus: Task 4 – Compute $Pr(Y|do(X))$

\[
\therefore Pr(Y|do(X)) = \sum_Z Pr(Y|Z, do(X)) Pr(Z|do(X))
\]
\[
= \sum_Z \left( Pr(Y|do(Z), do(X)) Pr(Z|do(X)) \right) \quad \text{Task 3}
\]
\[
= \sum_Z \left( Pr(Y|do(Z)) Pr(Z|do(X)) \right) \quad \text{Task 3}
\]
\[
= \sum_Z \left( \sum_{X'} Pr(Y|X', Z) Pr(X') \right) Pr(Z|X) \quad \text{Task 1}
\]
4. Do-Calculus and The Roy Model
Generalized Roy Model

The Generalized Roy Model stems from six variables:

1. **V**: Unobserved confounding variable $V$ not caused by any variable;
2. **X**: observed pre-treatment variables $X$ caused by $V$;
3. **Z**: instrumental variable $Z$ caused by $X$;
4. **D**: treatment choice $D$ that caused by $Z$, $V$ and $X$;
5. **U**: unobserved variable $U$ caused by $T$, $V$ and $X$;
6. **Y**: outcome of interest $Y$ caused by $T$, $U$ and $X$. 
This figure represents causal relations of the Generalized Roy Model. Arrows represent direct causal relations. Circles represent unobserved variables. Squares represent observed variables.
Key Aspects of the Generalized Roy Model

1. \( D \) is caused by \( Z, V \);
2. \( U \) mediates the effects of \( V \) on \( Y \) (that is \( V \) causes \( U \));
3. \( D \) and \( U \) cause \( Y \) and
4. \( Z \) (instrument) not caused by \( V, U \) and does not directly cause \( Y, U \).

We are left to examine the cases whether:

1. \( V \) causes \( X \) (or vice-versa),
2. \( X \) causes \( Z \) (or vice-versa),
3. \( X \) causes \( D \),
4. \( X \) causes \( U \),
5. \( D \) causes \( U \), and
6. \( X \) causes \( Y \).

The combinations of all these causal relations generate 144 possible models (Pinto, 2013).
Key Aspects of the Generalized Roy Model (Pinto, 2013)

Dashed lines denote causal relations that may not exist or, if they exist, the causal direction can go either way. Dashed arrows denote causal relations that may not exist, but, if they exist, the causal direction must comply the arrow direction.
Marginalizing the Generalized Roy Model

- We examine the identification of causal effects of the Generalized Roy Model using a simplified model w.l.o.g.
- Suppress variables $X$ and $U$.
- This simplification is usually called marginalization in the DAG literature (Koster (2002), Lauritzen (1996), Wermuth (2011)).
Marginalizing the Generalized Roy Model

\[ G = G_Z \]

This figure represents causal relations of the Marginalized Roy Model. Arrows represent direct causal relations. Circles represent unobserved variables. Squares represent observed variables.

**Note:** \( Z \) is exogenous, thus conditioning on \( Z \) is equivalent to fixing \( Z \).
Examining the Marginalized Roy Model – 1/4

- $Y \independent Z$ in $G_X$, by **Rule 1**
  \[ Pr(Y|do(X), Z) = Pr(Y|do(X)) \]

- $Y \independent Z$, in $G_{X,Z}$, by **Rule 3**
  \[ Pr(Y|do(X), Z) = Pr(Y|do(X)) \]

- $Y \independent Z|X$ in $G_{X,Z}$, by **Rule 2**
  \[ Pr(Y|do(X), do(Z)) = Pr(Y|do(X), Z) \]

\[ G_X = G_{X,Z} = G_{X,Z} \]

\[ \text{U} \rightarrow \text{X} \rightarrow \text{Y} \]

- $Z$

James Heckman and Rodrigo Pinto  
Causal Analysis  

THE UNIVERSITY OF CHICAGO
• Under $G_X$, $Y \not\perp\!\!\!\!\perp X$, thus **Rule 2** does not apply.
• Under $G_{X,Z}$, $Y \not\perp\!\!\!\!\perp X|Z$, thus **Rule 2** does not apply.
Examining the Marginalized Roy Model – 3/4

- $G_Z \Rightarrow Y \perp Z$, thus by **Rule 2** $Pr(Y|do(Z)) = Pr(Y|Z)$.
Examining the Marginalized Roy Model – 4 of 4 Modifications

- Under $G_{X,Z}$, $Y \not\perp (X, Z)$, thus Rule 2 does not apply.

$G_{X,Z}$

$U$

$Z$

$X$

$Y$
5. Conclusion

The Do-Calculus applied to the Marginalized Roy Model generates:

1. \( \Pr(Y|do(X), do(Z)) = \Pr(Y|do(X), Z) = \Pr(Y|do(X)) \),
2. \( \Pr(Y|do(Z)) = \Pr(Y|Z) \)

These relations only corroborate the exogeneity of the instrumental variable \( Z \) and are not sufficient to identify \( \Pr(Y|do(X)) \).

Identification of the Roy Model

To identify the Roy Model, we make assumption on how \( Z \) impacts \( X \), i.e. monotonicity/separability. These assumptions cannot be represented in a DAG. These assumptions are associated with properties of how \( Z \) causes \( X \) and not only if \( Z \) causes \( X \).

Return to main text.
14. Appendix III
Neyman and Rubin postulate counterfactuals \( \{ Y(s, \omega) \}_{s \in S} \), counterfactuals for person \( \omega \). The fine structure is missing describing the determinants of \( \omega \) without modeling the factors determining the \( Y(s, \omega) \) as is done in the econometric approach.
• Rubin and Neyman offer no model of the choice of which outcome is selected.
The “Rubin model” assumes

(R-1) \( \{ Y(s, \omega) \}_{s \in S} \), a set of counterfactuals defined for \textit{ex post} outcomes. It does not analyze agent valuations of outcomes nor does it explicitly specify treatment selection rules, except for contrasting randomization with nonrandomization;

(R-2) Invariance of counterfactuals for objective outcomes to the mechanism of assignment within a policy regime;

(R-3) No social interactions or general equilibrium effects for objective outcomes;

(R-4) No simultaneity in causal effects, i.e., outcomes cannot cause each other reciprocally.
Two further implicit assumptions in the application of the model are that internal validity problem P1 is the only evaluation problem of interest and that mean causal effects are the only objects of interest.
Its signature features are:

1. Development of an explicit framework for outcomes $Y(s, \omega)$, $s \in S$, measurements and the choice of outcomes where the role of unobservables ("missing variables") in creating selection problems and justifying estimators is made explicit.

2. The analysis of subjective evaluations of outcomes $R(s, \omega)$, $s \in S$, and the use of choice and compliance data to infer them.

3. The analysis of *ex ante* and *ex post* realizations and evaluations of treatments. This analysis enables analysts to model and identify regret and anticipation by agents. Points 2 and 3 introduce agent decision making into the treatment effect literature.
Development of models for identifying entire distributions of treatment effects (ex ante and ex post) rather than just the traditional mean parameters focused on by many statisticians. These distributions enable analysts to determine the proportion of people who benefit from treatment, a causal parameter not considered in the statistical literature on treatment effects (see Abbring and Heckman (2007) for a survey of methods used to identify distributions of treatment effects).

6 Models for simultaneous causality.

7 Definitions of parameters made without appeals to hypothetical experimental manipulations.

8 Clarification of the need for invariance of parameters with respect to different classes of manipulations to answer different policy questions.

Return to main text.