Causal Frameworks for Complex Causal Models

Rodrigo Pinto and James J. Heckman

Econ 312, Spring 2021
University of Chicago
Lecture 2 : Causal Calculus

About the Lecture

- **Scope** Introduce Causal Frameworks that enhance the Causal Inference
- **Goal** Tools that enable the analysis of arbitrarily complex causal models.
- **Benefit** Enable to explore nonparametric identification of models with multiple variables
- **Beyond** Matching, IV, and Mediation Models
- **Insights** Expand the way you can model an empirical inquiry

Related Literature

1. Pearl (1995)
   Causal Diagrams for Empirical Research
2. Pearl (2012)
   The Do-Calculus Revisited
3. Jaber, Zhang, Bareinboin (2018)
   Causal Identification under Markov Equivalence
4. Heckman and Pinto (2020)
   Causal Calculus for the Hypothetical Model Framework
5. Chalak and White (2011)
   Extended class of instrumental variables for the estimation of causal effects
6. Richardson Evans and Robins (2017)
   Nested Markov Properties for Acyclic Directed Mixed Graphs

Softwares

R Package

- Tikka and Karvanen (2019)
  Identifying Causal Effects with the R Package `causaleffect`

Online Resource

- Online Software `DAGitty`
  `www.http://dagitty.net/`
- Johannes Textor (2020)
  Drawing and Analyzing Causal DAGs with DAGitty

Basic Concepts Review

Defining Causal Models

Causal Model: defined by a **4** components:

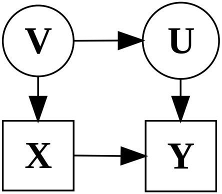
1. **Random Variables** that are observed and/or unobserved by the analyst: $\mathcal{T} = \{Y, U, X, V\}$.
2. **Error Terms** that are mutually independent: $\epsilon_Y, \epsilon_U, \epsilon_X, \epsilon_V$.
3. **Structural Equations** that are autonomous : f_Y, f_U, f_X, f_V.
4. **Causal Relationships** that map the inputs causing each variable: $Y = f_Y(X, U, \epsilon_Y); X = f_X(V, \epsilon_X); U = f_U(V, \epsilon_U); V = f_V(\epsilon_V)$.

Econometric approach explicitly models **unobservables** that are often the main object of study.

Structural Relationships / Autonomous Functions

$Y = f_Y(X, U, \epsilon_Y),$	Y observed
$X = f_X(V, \epsilon_X),$	X observed
$U = f_U(V, \epsilon_U),$	U unobserved
$V = f_V(\epsilon_V),$	V unobserved

Directed Acyclic Graph (DAG) representation

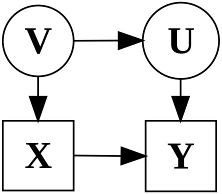


Properties of this Causal Framework

- **Recursive Property :** No variable is descendant of itself.
- Autonomy + Indep. Errors + Recursivity ⇒ **Bayesian Network**
  1. **Local Markov Condition** (LMC): a variable is independent of its non-descendants conditioned on its parents.
  2. **Graphoid Axions** (GA): new independence relations based on the LMC relations
- **Benefit** of Bayesian Network Tools translates causal links into independence relations

A Useful Tool: Local Markov Condition (LMC):
(Kiiveri, 1984, Lauritzen, 1996)

LMC: A variable is independent of its non-descendants conditional on its parents



- For example: $Y \perp\!\!\!\perp \underbrace{V}_{\text{non-descendants}} | \underbrace{(X,U)}_{\text{parents}}$
- A fully non-parametric causal model can be equivalently described by its LMCs.

Additional Tool: Graphoid Axioms (GA)
(Dawid, 1979)

Primary GA rules:

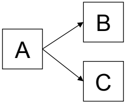
Weak Union: $X \perp\!\!\!\perp (W,Y)|Z \Rightarrow X \perp\!\!\!\perp Y|(W,Z)$.
Contraction: $X \perp\!\!\!\perp W|(Y,Z)$ and $X \perp\!\!\!\perp Y|Z \Rightarrow X \perp\!\!\!\perp (W,Y)|Z$.
Intersection: $X \perp\!\!\!\perp W|(Y,Z)$ and $X \perp\!\!\!\perp Y|(W,Z) \Rightarrow X \perp\!\!\!\perp (W,Y)|Z$

Remaining GA rules:

Symmetry: $X \perp\!\!\!\perp Y|Z \Rightarrow Y \perp\!\!\!\perp X|Z$.
Decomposition: $X \perp\!\!\!\perp (W,Y)|Z \Rightarrow X \perp\!\!\!\perp Y|Z$.
Redundancy: $X \perp\!\!\!\perp Y|X$.

Building Blocks of Causal Relations

Common Cause

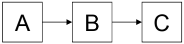


Example:
Village size (A)
causes babies (B)
and storks (C)

CI:
B and C
conditionally
independent
given A

$$B \perp\!\!\!\perp C | A$$

Chain

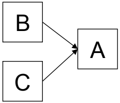


Example:
Smoking (A)
causes tar (B)
causes cancer (C)

CI:
A and C
conditionally
independent
given B

$$A \perp\!\!\!\perp C | B$$

Collider



Example:
Firing squad
(B & C)
shoot prisoner (A)

CI:
B and C
conditionally
dependent
given A

$$B \perp\!\!\!\perp C$$

but $B \not\!\perp\!\!\!\perp C | A$

More general approaches?

Is there a more general approach to investigate independencies? Yes!

1. A linear-algebraic tool for conditional independence inference Inference
   Tanaka, Studeny, Takemura and Sei (2015)
2. Efficient Algorithms for Conditional Independence Inference
   Bouckaert, Hemmecke, Lindner, Studeny (2010)
3. Probabilistic Conditional Independence Structures
   Studeny (2005)
4. See Lauritzen (1996) for the general theory of Bayesian Networks.

Analysis of Counterfactuals – The Fixing Operator

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

Standard Model	Model under Fixing
$V = f_V(\epsilon_V)$	$V = f_V(\epsilon_V)$
$U = f_U(V, \epsilon_U)$	$U = f_U(V, \epsilon_U)$
$X = f_X(V, \epsilon_X)$	$\boldsymbol{X = x}$
$Y = f_Y(X, U, \epsilon_Y)$	$Y = f_Y(\boldsymbol{x}, U, \epsilon_Y)$

- **Importance:** Establishes the framework for counterfactuals.
- **Counterfactual:** $Y(x)$ represents outcome Y when X is fixed at x.
- **Linear Case:** $Y = X\beta + U + \epsilon_Y$ and $Y(x) = x\beta + U + \epsilon_Y$;

Questions

Causal Calculus

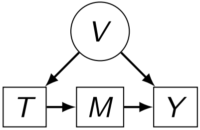
What can you gain from additional structure?

A General Method to Examine Complex Models

Merging Statical Theory with Causal Analysis

How can we use the SMC to identify the Front-door Model?

$$V = f_V(\epsilon_V)$$
$$T = f_X(V, \epsilon_T)$$
$$M = f_M(T, \epsilon_M)$$
$$Y = f_Y(M, V, \epsilon_Y)$$



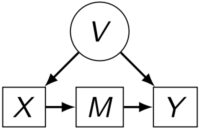
Two Counterfactuals:

$$M(t) = f_M(t, \epsilon_M) \Rightarrow M(t) \perp\!\!\!\perp T$$
$$Y(m) = f_Y(m, V, \epsilon_Y) \text{ but } M \perp\!\!\!\perp V|T \Rightarrow Y(m) \perp\!\!\!\perp M|T$$

Thus the following equalities hold:

- $P(M(t)) = P(M|T = t)$
- $E(Y(m)|T = t) = E(Y|M = m, T = t)$

Identifying the Counterfactual Mean $E(Y(t))$



Outcome $Y = f_Y(M, V, \epsilon_y)$ generates the following counterfactual:

$$\therefore Y(t) = f_Y(M(t), V, \epsilon_y) \Rightarrow E(Y(t)) = \int E(Y(m)) dF_{M(t)}(m)$$

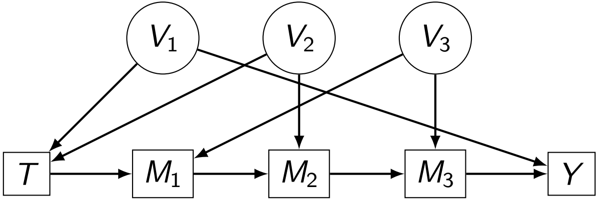
But $P(M(t)) = P(M|T = t)$ and

$$E(Y(m)|T = t') = E(Y|M = m, T = t')$$

$$\Rightarrow E(Y(m)) = \int E(Y|M = m, T = t') dF_T(t')$$

$$\Rightarrow \mathbf{E}(Y(t)) = \int_m \underbrace{\left(\int_{t'} \mathbf{E}(Y|M = m, T = t') dF_T(t') \right)}_{E(Y(m))} \underbrace{dF_{M|T=t}(m)}_{dF_{M(t)}(m)}$$

What about this model?



- X is endogenous, $Y(x) \not\perp\!\!\!\perp X$, indeed, ALL variables are endogenous
- No instruments
- Yet, causal effects are identified:

$$\mathbf{E}(Y(t)) = \int_{t'} \int_{m_1} \int_{m_2} \int_{m_3} \mathbf{E}(Y|m_3, m_2, m_1, \mathbf{T} = \mathbf{t}')$$

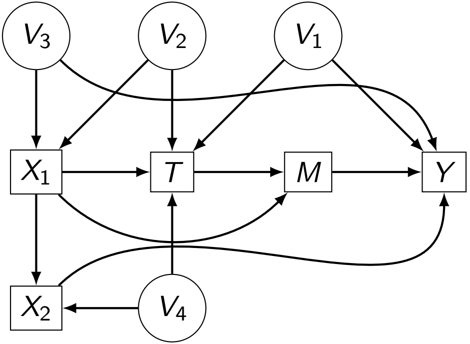
$$dF_{M_3|m_2,m_1,\mathbf{T}=\mathbf{t}}(m_3)$$

$$dF_{M_2|m_1,\mathbf{T}=\mathbf{t}'}(m_2)$$

$$dF_{M_1|\mathbf{T}=\mathbf{t}}(m_1)$$

$$dF_T(\mathbf{t'})$$

And what about this model?



$$\mathbf{E}(Y(t)) = \int_{t'} \int_{m} \int_{x_1} \mathbf{E}(Y|m, x_1, \mathbf{T} = \mathbf{t}') dF_{M|x_1, \mathbf{T}=\mathbf{t}}(m) dF_{X_1|\mathbf{T}=\mathbf{t}'}(x_1) dF_T(\mathbf{t}')$$

What about a General Framework for Causal Calculus?

- The goal of a framework for causal calculus is to deliver a standard methodology that applies to any DAG.
- A set general of rules that can be used to assess counterfactual outcomes whenever those are identified.
- A methodology/algorithm that be coded, so the researcher dos not need to investigate case by case.
- Such framework is useful to investigate which properties of DAGs are necessary/sufficient to render identification of causal parameters.
- Most important, a framework that facilitates to investigate the identification of causal effect is more complex DAGs.

Moreover, Fixing is not Well-defined in Statistics

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 considers the dependence structure of the data generating process.

- Fixing has direction while conditioning does not.
- **Question:** How can we make statistics converse with causality?
- **Answer:** The hypothetical model

Questions

The Hypothetical Model Framework

The Causal Calculus using The Hypothetical Framework

Merging Statical Theory and Causal Analysis

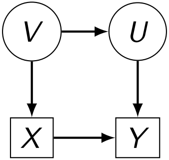
- The mismatch between statistical theory and causal inference motivated the study of the Hypothetical Model Framework
- The framework merges statical theory and causal analysis without the necessity of defining new tools of analysis

Properties of the Hypothetical Model

- **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.
- **Simplicity:** the method does not rely on additional tools of analysis beyond standard statistical theory
- **Identification:** relies on evaluating causal parameters defined in the *hypothetical model* using data generated by the *empirical model*.

Example of Data Generating Model (DAG) Representation

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



- The Local Markov Condition (LMC) generates two independence conditions:
- $Y \perp\!\!\!\perp V|(U, X)$ and $U \perp\!\!\!\perp X|V$

Defining The Hypothetical Model

The hypothetical model stems from 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.

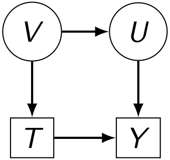
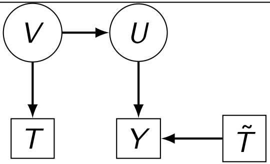
Usage:

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

Example of the Hypothetical Model for fixing T

The Associated Hypothetical Model

$$Y = f_Y(\tilde{T}, U, \epsilon_Y);\ T = f_T(V, \epsilon_T);\ U = f_U(V, \epsilon_U);\ V = f_V(\epsilon_V).$$

Empirical Model	Hypothetical Model
	

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

Why the hypothetical variable is useful?

Properties the Hypothetical Model:

1. **Hypothetical Variable:** $\tilde{T}$ replaces the T-inputs of structural equations.
2. **Characteristic:** $\tilde{T}$ is an **external variable**, i.e., no parents.
3. **Thus:** Hypothetical variable has independent variation.
4. **Usage:** hypothetical variable $\tilde{T}$ enables analysts to examine fixing using standard tools of probability (conditioning).

Main Benefit

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

$$\underbrace{E_{\mathrm{E}}(Y(t))}_{\text{Causal Operation Empirical Model}} = \underbrace{E_{\mathrm{H}}(Y|\tilde{T}=t)}_{\text{Statistical Operation Hypothetical Model}}$$

- Causality is defined **within** Statistics/Probability
- **No additional Tools Required.**

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.

Connecting Hypothetical and Empirical Models: Two Useful Conditions

Only two conditions **suffice** to investigate the identification of causal parameters!

For any disjoint set of variables Y, W, we have that:

Rule 1: $Y \perp\!\!\!\perp \widetilde{T} | (T, W) \Rightarrow$

$$\mathbf{P}_H(Y|\widetilde{T}, T = t', W) = \mathbf{P}_H(Y|T = t', W) = \mathbf{P}_E(Y|T = t', W)$$

Rule 2: $Y \perp\!\!\!\perp T | (\widetilde{T}, W) \Rightarrow$

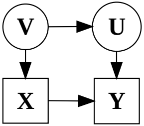
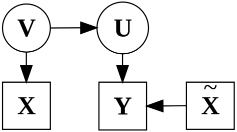
$$\mathbf{P}_H(Y|\widetilde{T} = t, T, W) = \mathbf{P}_H(Y|\widetilde{T} = t, W) = \mathbf{P}_E(Y|T = t, W)$$

- If $Y \perp\!\!\!\perp \widetilde{T} | (T, W)$ or $Y \perp\!\!\!\perp T | (\widetilde{T}, W)$ occurs
- We can connect hypothetical and empirical models!

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** $\mathbf{P}_{\mathrm{H}}(Y|\widetilde{T})$ in terms of other variables.
4. **Connect** this expression to the empirical model using $Y \perp\!\!\!\perp \widetilde{T}|(T, W)$ or $Y \perp\!\!\!\perp T|(\widetilde{T}, W)$

Example of the Hypothetical Model for Fixing X

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

1. $E_E(Y(x) \mid V) = E_H(Y \mid \widetilde{X} = x, V)$ by the main property of the HM
2. $X \perp\!\!\!\perp (U, Y, \tilde{X}) \mid V \Rightarrow X \perp\!\!\!\perp Y \mid (\tilde{X}, V)$ holds by LMC
3. $E_H(Y \mid \widetilde{X} = x, V) = E_E(Y \mid X = x, V)$ by rule 2

Rule 2 is a Matching Property

If there exist V such that, $T \perp\!\!\!\perp Y|V, \tilde{T}$, then $E_{\mathrm{H}}(Y|V, \tilde{T} = t)$ in hypothetical model is equal to $E_{\mathrm{E}}(Y(t)|T = t)$ in empirical model.

- Main Property of the Hypothetical Model implies that counterfactual outcome $E_{\mathrm{E}}(Y(x))$ can be expressed as

$$E_{\mathrm{E}}(Y(t)) = \int E_{\mathrm{H}}(Y|V = v, \tilde{T} = t) dF_V(v)$$

- LMC for the hypothetical model generates $Y \perp\!\!\!\perp T|(V, \tilde{T})$.
- By Rule 2, $E_{\mathrm{H}}(Y|V = v, \tilde{T} = t) = E_{\mathrm{E}}(Y|V = v, T = t)$
- Thus, the counterfactual outcome $E_{\mathrm{E}}(Y(t))$ can be obtained by:

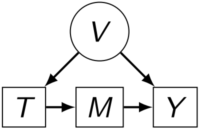
$$E_{\mathrm{E}}(Y(t)) = \underbrace{\int E_{\mathrm{E}}(Y|V = v, T = t) dF_V(v)}_{\text{In Empirical Model by Rule 2}}$$

Causal Model 1: Revisiting the Front-door Model

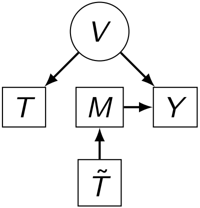
Empirical Front-door Model	**Hypothetical Front-door Model**
Observed Variables	Observed Variables
$T = f_T(V, \epsilon_T)$ $M = f_M(T, \epsilon_M)$ $Y = f_Y(V, M, \epsilon_Y)$	$T = f_T(V, \epsilon_T)$ $M = f_M(\widetilde{T}, \epsilon_M)$ $Y = f_Y(V, M, \epsilon_Y)$ $Y = f_Y(V, M, \epsilon_Y)$
Exogenous Variables	Exogenous Variables
V	$V, \widetilde{T}$
Unobserved Variables	Unobserved Variables
$V = f_V(\epsilon_V)$	$V = f_V(\epsilon_V)$

Independence Relations for Front-Door Model

Empirical Model



Hypothetical Model



$V \perp\!\!\!\perp -\mid-$	$V \perp\!\!\!\perp (M, \tilde{T})$
$T \perp\!\!\!\perp -\mid V$	$T \perp\!\!\!\perp (M, Y, \tilde{T})\mid V$
$M \perp\!\!\!\perp V\mid T$	$M \perp\!\!\!\perp (T, V)\mid \tilde{T}$
$Y \perp\!\!\!\perp T\mid(V, M)$	$Y \perp\!\!\!\perp (T, \tilde{T})\mid(V, M)$
	$\tilde{T} \perp\!\!\!\perp (T, V)$

Useful independence relations in the Front-Door hypothetical model:

1. $Y \perp\!\!\!\perp \tilde{T}\mid(M, T)$ (due to $Y \perp\!\!\!\perp \tilde{T}\mid M$ & $(\tilde{T}, M) \perp\!\!\!\perp (T, V)$)
2. $M \perp\!\!\!\perp T\mid\tilde{T}$
3. $\tilde{T} \perp\!\!\!\perp T$

General Identification Criteria

1. Given a Causal Model represented by a DAG,
2. The counterfactual outcome $Y(t)$ is identified if
3. There exists a set of observable variable K that bridges
4. The conditional independence $Y \perp\!\!\!\perp \widetilde{T} | (T, K)$ into $T \perp\!\!\!\perp \widetilde{T}$.
5. Moreover, the identification formula for $Y(t)$ can be expressed as an alternate pattern.

Conditions for Causal Model 1 (Front-door)

Variable M bridges the independence $Y \perp\!\!\!\perp \widetilde{T} | (T, M)$ to $\widetilde{T} \perp\!\!\!\perp T$:

$$
\begin{array}{rlllcr}
& & & & & \text{Connection} \\
Y & \perp\!\!\!\perp & \widetilde{T} & | (T, M) & \Rightarrow & \mathbf{P}_H(Y|\widetilde{T}, T = t', M) = \mathbf{P}_E(Y|T = t', M) \\
M & \perp\!\!\!\perp & T & | \widetilde{T} & \Rightarrow & \mathbf{P}_H(M|\widetilde{T} = t, T) = \mathbf{P}_E(M|T = t) \\
T & \perp\!\!\!\perp & \widetilde{T} & | T & \Rightarrow & \mathbf{P}_H(T = t'|\widetilde{T}) = \mathbf{P}_E(T = t')
\end{array}
$$

The identification Formula that follows the alternate pattern:

$$
\begin{aligned}
& P_H(Y|\widetilde{T} = t) = \\
& = \sum_{t',m} P_H(Y|m, T = t', \widetilde{T} = t) P_H(m|T = t', \widetilde{T} = t) P_H(T = t'|\widetilde{T} = t) \\
& = \sum_{t',m} P_E(Y|m, T = t') P_E(m|T = t) P_E(T = t'),
\end{aligned}
$$

Identifying Equations (Front-door Model)

Categorical Variables:

$$E_H(Y|\widetilde{T} = t) =$$
$$= \sum_{t',m} E_E(Y|m, T = t')P_E(m|T = t)P_E(T = t'),$$

Continuous Case:

$$\mathbf{E}(Y(t)) = \int_{t'} \int_m E(Y|m, T = t')dF_{M|T=t}(m)dF_T(t')$$

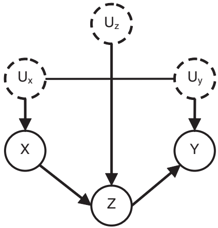
Previous Equation:

$$\mathbf{E}(Y(t)) = \int_m \underbrace{\left(\int_{t'} \mathbf{E}(Y|M = m, T = t')dF_T(t')\right)}_{E(Y(m))} \underbrace{dF_{M|T=t}(m)}_{dF_{M(t)}(m)}$$

What if We Assume Linearity? (Chalack and White, 2013)
The Front Door Model

$$
\begin{aligned}
&(1)\quad X \overset{c}{=} \alpha_x U_x \\
&(2)\quad Z \overset{c}{=} \gamma_z X + \alpha_z U_z \\
&(3)\quad Y \overset{c}{=} Z'\delta_o + U_y'\alpha_o,
\end{aligned}
$$

where $U_z \perp (U_x, U_y)$ and $U_x \not\perp U_y$.



GRAPH 10 (G_{10}) Conditional instruments

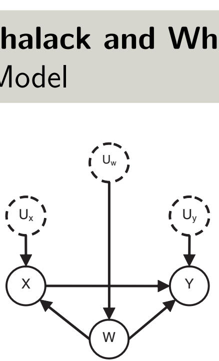
$$\hat{\beta}_n^{XI|C} \equiv \{(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z})\} \times \{[\mathbf{Z}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}]^{-1}[\mathbf{Z}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}]\}.$$

$$E(Y(t)) = \sum_m \mathbf{P}(M = m|X = x)\left(\sum_{x'} E(Y|m, x')\,\mathbf{P}(X = x')\right)$$

What if We Assume Linearity? (Chalack and White, 2013)
The Matching Model

$$\begin{aligned}
&(1)\quad W \overset{c}{=} \alpha_w U_w \\
&(2)\quad X \overset{c}{=} \gamma_x W + \alpha_x U_x \\
&(3)\quad Y \overset{c}{=} X'\beta_o + W'\gamma_o + U_y'\alpha_o,
\end{aligned}$$

where U_w, U_x, and U_y are jointly independent.



GRAPH 7*b* (G_{7b}) Conditioning instruments

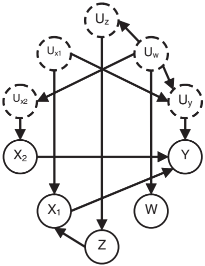
$$\hat{\beta}_n^{XC|I} \equiv \{\mathbf{X}'(\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{X}\}^{-1}\{\mathbf{X}'(\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{Y}\}.$$

- This is the Frisch–Waugh–Lovell (double errors) theorem

What if We Assume Linearity?
Conditional Instruments

(1) $U_z \overset{c}{=} \delta_z U_w$
(2) $U_{x_2} \overset{c}{=} \delta_{x_2} U_w$
(3) $U_y \overset{c}{=} \delta_{y_1} U_w + \delta_{y_2} U_{x_1}$
(4) $W \overset{c}{=} \alpha_w U_w$
(5) $Z \overset{c}{=} \alpha_z U_z$
(6) $X_1 \overset{c}{=} \gamma_{x_1} Z + \alpha_{x_1} U_{x_1}$
(7) $X_2 \overset{c}{=} \alpha_{x_2} U_{x_2}$
(8) $Y \overset{c}{=} X_1'\beta_1 + X_2'\beta_2 + U_y'\alpha_o,$

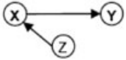
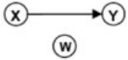
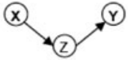
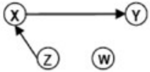
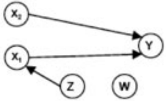
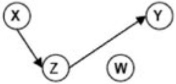
where $U_{x_1} \perp U_w$.



GRAPH 14 (G_{14}) Conditionally exogenous instruments and causes, given conditioning instruments (XCI|I)

$$\hat{\beta}_n^{XCI|I} \equiv [\tilde{\mathbf{Z}}'(\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{X}]^{-1}[\tilde{\mathbf{Z}}'(\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{Y}].$$

- $(Z, X_2) \perp\!\!\!\perp U_y | W$ holds under linearity
- Conditional Instruments: $\tilde{Z} = [Z, X_2]$

Causal structure among observables	(Conditional) exogeneity	Estimator
	XC: $X \perp U_y$	$\hat{\beta}_n^{XC} \equiv (X'X)^{-1}(X'Y)$
	XI: $Z \perp U_y$	$\hat{\beta}_n^{XI} \equiv (Z'X)^{-1}(Z'Y)$
	XC|I: $X \perp U_y \mid W$	$\hat{\beta}_n^{XC\|I} \equiv [X'(I - W(W'W)^{-1}W')X]^{-1}$ $[X'(I - W(W'W)^{-1}W')Y]$
	XI|C: $Z \perp U_y \mid X$ and $X \perp U_z$	$\hat{\beta}_n^{XI\|C} \equiv (X'X)^{-1}(X'Z)$ $\times [Z'(I - X(X'X)^{-1}X')Z]^{-1}$ $\times [Z'(I - X(X'X)^{-1}X')Y]$
	XI|I: $Z \perp U_y \mid W$	$\hat{\beta}_n^{XI\|I} \equiv [Z'(I - W(W'W)^{-1}W')X]^{-1}$ $\times [Z'(I - W(W'W)^{-1}W')Y]$
	XCI|I: $(Z, X_2) \perp U_y \mid W$	$\hat{\beta}_n^{XCI\|I} \equiv [\tilde{Z}'(I - W(W'W)^{-1}W')X]^{-1}$ $\times [\tilde{Z}'(I - W(W'W)^{-1}W')Y]$ where $X \equiv [X_1', X_2']'$ and $\tilde{Z} \equiv [Z', X_2']'$
	XI|CI: $Z \perp U_y \mid (X, W)$ and $X \perp U_z$	$\hat{\beta}_n^{XI\|CI} \equiv [(X'X)]^{-1}(X'Z)$ $\times [Z'(I - \tilde{W}(\tilde{W}'\tilde{W})^{-1}\tilde{W}')Z]^{-1}$ $\times [Z'(I - \tilde{W}(\tilde{W}'\tilde{W})^{-1}\tilde{W}')Y]$ where $\tilde{W} \equiv [X', W']'$

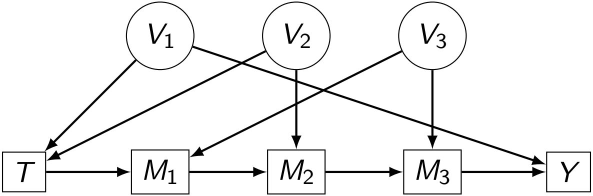
Questions

Example: Causal Model 2

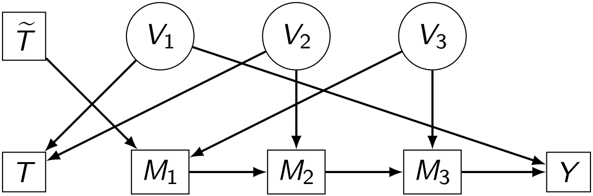
Empirical Model Observed Variables	**Hypothetical Model** Observed Variables
$T = f_T(V_1, V_2, \epsilon_T)$	$T = f_T(V_1, V_2, \epsilon_T)$
$M_1 = f_{M_1}(V_3, T, \epsilon_{M_1})$	$M_1 = f_{M_1}(V_3, T, \epsilon_{M_1})$
$M_2 = f_{M_2}(V_2, M_1, \epsilon_{M_2})$	$M_2 = f_{M_2}(V_2, M_1, \epsilon_{M_2})$
$M_3 = f_{M_3}(V_3, M_2, \epsilon_{M_3})$	$M_3 = f_{M_3}(V_3, M_2, \epsilon_{M_3})$
$Y = f_Y(V_1, M_3, \epsilon_Y)$	$Y = f_Y(V_1, M_3, \epsilon_Y)$
Exogenous Variables	Exogenous Variables
V_1, V_2, V_3	$V_1, V_2, V_3, \widetilde{T}$

DAG of Causal Model 2

Directed Acyclic Graph of the Empirical Model



Directed Acyclic Graph of the Hypothetical Model



Causal Model 2 - Connecting Hypothetical and Empirical

Applying LMC and GA to the hypothetical model generates the following indep. relations:

$$
\begin{aligned}
Y &\perp\!\!\!\perp \widetilde{T} \mid (T, M_3, M_2, M_1) \\
M_3 &\perp\!\!\!\perp T \mid (\widetilde{T}, M_2, M_1) \\
M_2 &\perp\!\!\!\perp \widetilde{T} \mid (T, M_1) \\
M_1 &\perp\!\!\!\perp T \mid \widetilde{T} \\
\widetilde{T} &\perp\!\!\!\perp T \text{ always hold}
\end{aligned}
$$

Observe that:

- The sequence of observed variables $M_1 \to M_2 \to M_3$ forms a **bridge**
- from $Y \perp\!\!\!\perp \widetilde{T} \mid (T, M_3, M_2, M_1)$ (initial relation)
- to $\widetilde{T} \perp\!\!\!\perp T$ (final relation)

Causal Model 2 - Connecting Hypothetical and Empirical

Using the two probability rules, we can achieve identification:

Hypothetical Model

$$P_H(Y|\widetilde{T} = t) = \sum_{t', m_3, m_2, m_1} P_H(Y|m_3, m_2, m_1, T = t', \widetilde{T} = t) \, P_H(M_3 = m_3|m_2, m_1, T = t', \widetilde{T} = t) \, P_H(M_2 = m_2|m_1, T = t', \widetilde{T} = t) \, P_H(M_1 = m_1|T = t', \widetilde{T} = t) \, P_H(T = t'|\widetilde{T} = t)$$

Empirical Model

(alternate pattern)

$$P_E(Y(t)) = \sum_{t', m_3, m_2, m_1} P_E(Y|m_3, m_2, m_1, T = t') \, P_E(M_3 = m_3|m_2, m_1, T = t) \, P_E(M_2 = m_2|m_1, T = t') \, P_E(M_1 = m_1|T = t) \, P_E(T = t')$$

Small Detour: On the Do-Calculus

- Creates a special set of rules that combine:
    1. Graphical conditions
    2. Conditional independence statements
    3. Probability equalities as postulates

In contrast, 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

Major Achievement: The do-calculus is Complete!

Limitation of the Do-Calculus: IV model is not Identified

- The necessary assumptions the identify the IV model are monotonicity/separability conditions
- These are functional form assumptions
- They refer to properties of the structural functions
- Beyond the DAG information
  (Causal direction among variables remains the same)
- **The do-calculus cannot identify the IV model**
- The algorithm simply returns that the IV model is not identified

Causal Model 2 - Comparison Hypothetical vs Do-Calculus

Equation from do-calculus is different, but equivalent:

Hypothetical Model (alternate pattern):

$$P_E(Y(t)) = \sum_{m_3, m_2, m_1, t'} P_E(Y|m_3, m_2, m_1, t')P_E(m_3|m_2, m_1, t)$$
$$P_E(m_2|m_1, t')P_E(m_1|t)P_E(t')$$

Do-calculus:

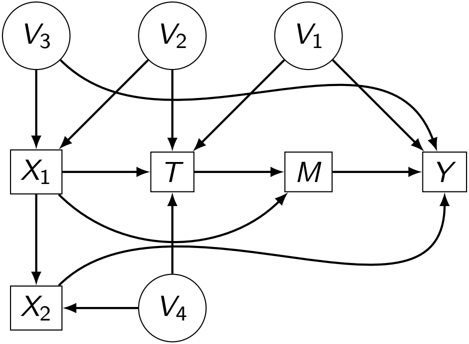
$$P_E(Y(t)) = \sum_{m_1, m_2, m_3} P_E(m_1|t)P_E(m_3|t, m_1, m_2)\cdot$$
$$\left(\sum_{t'} P_E(t')P_E(m_2|t', m_1)\right)\cdot$$
$$\left(\sum_{t', m_2'} P_E(t')P_E(m_2'|t', m_1)P_E(Y|t', m_1, m_2', m_3)\right)$$

Causal Model 3

Empirical Model Observed Variables	Hypothetical Model Observed Variables
$X_1 = f_{X_1}(V_2, V_3, \epsilon_{X_1})$ $X_2 = f_{X_2}(V_4, X_1, \epsilon_{X_2})$ $T = f_T(V_1, V_2, V_4, X_1, \epsilon_T)$ $M = f_M(X_1, T, \epsilon_M)$ $Y = f_Y(V_1, V_3, X_2, M, \epsilon_Y)$	$X_1 = f_{X_1}(V_2, V_3, \epsilon_{X_1})$ $X_2 = f_{X_2}(V_4, X_1, \epsilon_{X_2})$ $T = f_T(V_1, V_2, V_4, X_1, \epsilon_T)$ $M = f_M(X_1, \widetilde{T}, \epsilon_M)$ $Y = f_Y(V_1, V_3, X_2, M, \epsilon_Y)$
Exogenous Variables	Exogenous Variables
V_1, V_2, V_3, V_4	$V_1, V_2, V_3, V_4, \widetilde{T}$
Unobserved Variables $V_1 = f_{V_1}(\epsilon_{V_1}),\ V_2 = f_{V_2}(\epsilon_{V_2}),$ $V_3 = f_{V_3}(\epsilon_{V_3}),\ V_4 = f_{V_4}(\epsilon_{V_4})$	Unobserved Variables $V_1 = f_{V_1}(\epsilon_{V_1}),\ V_2 = f_{V_2}(\epsilon_{V_2})$ $V_3 = f_{V_3}(\epsilon_{V_3}),\ V_4 = f_{V_4}(\epsilon_{V_4})$

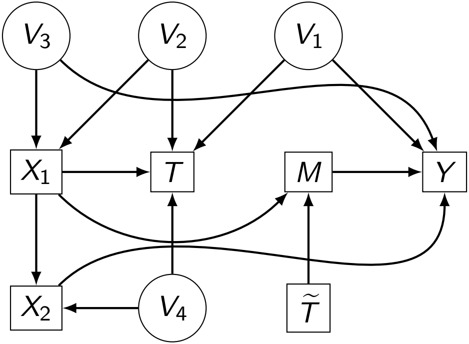
DAG of Empirical Model 3

Directed Acyclic Graph of the Empirical Model



DAG of Hypothetical Model 3

Directed Acyclic Graph of the Hypothetical Model



$$Y \perp\!\!\!\perp \widetilde{T} | (\widetilde{T}, X_1, M) \quad (1)$$

$$M \perp\!\!\!\perp T | (\widetilde{T}, X_1) \quad (2)$$

$$X_1 \perp\!\!\!\perp \widetilde{T} | \widetilde{T} \quad (3)$$

Causal Model 3 - Connecting Hypothetical and Empirical

LMC and GA give you the following conditions:

$$
\begin{aligned}
Y &\perp\!\!\!\perp \widetilde{T} \mid (T, M, X_1) \\
M &\perp\!\!\!\perp T \mid (\widetilde{T}, X_1) \\
X_1 &\perp\!\!\!\perp \widetilde{T} \mid T \\
\widetilde{T} &\perp\!\!\!\perp T \text{ always hold}
\end{aligned}
$$

- The sequence of observed variables $M \to X_1$ forms a **bridge**
- from $Y \perp\!\!\!\perp \widetilde{T} | (T, X_1, M)$ (initial relation)
- to $\widetilde{T} \perp\!\!\!\perp T$ (final relation)

Causal Model 3 - Connecting Hypothetical and Empirical

Using the two probability rules, we can achieve identification:

Hypothetical Model

$$P_H(Y|\widetilde{T} = t) = \sum_{t',m,x_1} P_H(Y|m, x_1, T = t', \widetilde{T} = t) \, P_H(M = m|x_1, T = t', \widetilde{T} = t) \, P_H(X_1 = x_1|T = t', \widetilde{T} = t) \, P_H(T = t'|\widetilde{T} = t)$$

Empirical Model

(alternate pattern)

$$P_E(Y(t)) = \sum_{t',m,x_1} P_E(Y|m, x_1, T = t') \, P_E(M = m|x_1, T = t) \, P_E(X_1 = x_1|T = t') \, P_E(T = t')$$

Causal Model 3 - Do-calculus Identifying Equation

Equation from do-calculus is different, but equivalent:

Using Hypothetical Model (alternate pattern):

$$P_E(Y(t)) =$$

$$\sum_{m, x_1, t'} P_E(Y|m, x_1, T = t')P_E(m|x_1, T = t)P_E(x_1|T = t')P_E(T = t')$$

Using Do-calculus:

$$P_E(Y(t)) = \sum_{x_1, x_2, m} P_E(m|x_1, T = t)P_E(x_2|x_1)P_E(x_1)$$

$$\cdot \frac{\left(\sum_{t'} P_E(Y|x_1, T = t', x_2, m)P_E(x_2|x_1, T = t')P_E(T = t'|x_1)P_E(x_1)\right)}{\left(P_E(x_2|x_1)P_E(x_1)\right)}$$

Questions/Break

The Do-calculus

The most substantial contribution to the theory of causality is the last decades.

- Do-calculus (Pearl, 1995) three causal inference rules.
- Software
  1. Free Software DAGitty (only for indepence conditions, matching variables and finding Instruments)
  2. R-package `causaleffect`

The Do-calculus

- **Goal:** Counterfactual manipulations using the empirical model.
- **No** 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., $E_{\mathrm{E}}(Y|X=1,\tilde{X}=1) - E_{\mathrm{E}}(Y|X=1,\tilde{X}=0)$.

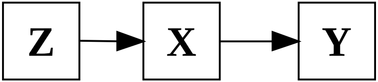
Difference: Identification using the hypothetical model does not require additional **causal rules**, only standard statistical tools.

Definition the Do-operator (which is Fixing)

The **Do-operator** is based on the **Truncated Factorization** of the probability factor of the fixed variable is deleted:
Let $X \subset V$: Then
$\Pr(V(x) = v) = \Pr(V_1 = v_1, \ldots, V_{m+n} = v_{m+n}, |do(X) = x)$ and:

$$\Pr(V(x) = v) = \begin{cases} \prod_{V_i \in V \setminus X} P(V_i = v_i | pa(V_i)) & \text{if } v \text{ is consistent with } x; \\ 0 & \text{if } v \text{ is inconsistent with } x. \end{cases}$$

Example of the Do-operator



- **Variables:** Y, X, Z
- **Factorization:**

$$\begin{aligned} \Pr(Y, X, Z) &= \Pr(Y|Z, X) \Pr(X|Z) \Pr(Z) \\ &= \Pr(Y|X) \Pr(X|Z) \Pr(Z) \end{aligned}$$

- **Do-operator:** $\Pr(Z, Y|do(X) = x) = \Pr(Y|X = x) \Pr(Z)$
- **Conditional operator:**

$$\begin{aligned} \Pr(Y, Z|X = x) &= \Pr(Y|Z, X = x) \Pr(Z|X = x) \\ &= \Pr(Y|X = x) \Pr(Z|X = x) \end{aligned}$$

Do-operator targets variables, not causal links.

Comparison: Hypothetical Model and Do-Operator

Fixing within Standard Probability Theory

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

$$\underbrace{E_{\mathrm{E}}(Y(x))}_{\text{Causal Operation Empirical Model}} = \underbrace{E_{\mathrm{H}}(Y|\tilde{X}=x)}_{\text{Statistical Operation Hypothetical Model}}$$

do-Operator and Statistical Conditioning

Let $\tilde{X}$ be the hypothetical variable in G_{H} associated with variable X in the empirical model G_{E}, such that $Ch_{\mathrm{H}}(\tilde{X}) = Ch_{\mathrm{E}}(X)$, then:

$$\mathbf{P}_{\mathrm{H}}(\mathcal{T}_{\mathrm{E}} \setminus \{X\}|\tilde{X}=x) = \mathbf{P}_{\mathrm{E}}(\mathcal{T}_{\mathrm{E}} \setminus \{X\}|do(X)=x).$$

Defining the Do-calculus

What is the do-calculus?

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

1. **Goal:** Identify causal effects from non-experimental data.
2. **Application:** Bayesian network structure, i.e., Directed Acyclic Graph (DAG) that represents causal relationships.
3. **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

1. **Information:** DAG only provides information on the causal relation among variables.
2. **Not Suited** for examining assumptions on functional forms.
3. **Identification:** If this information is sufficient to identify causal effects, then:
4. **Completeness**:
   1. There exists a **sequence** of application of the Do-Calculus that
   2. **generates** a formula for causal effects based on observational quantities (Huang and Valtorta 2006, Shpitser and Pearl 2006)
5. **Limitation:** Does not allow for additional information outside the DAG framework.
   1. **Only** applies to the information content of a DAG.
   2. **IV** is not identified through Do-calculus
   3. **Why?** requires assumptions outside DAG: linearity, monotonicity, separability.

Notation for the Do-calculus

More notation is needed to define these rules:

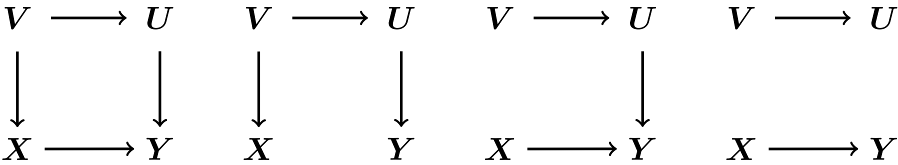
DAG Notation

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.
- $G_{\underline{X}}$: DAG that modifies G by deleting arrows emerging *from* X.
- $G_{\overline{X},\underline{Z}}$: DAG that modifies G by deleting arrows pointing *to* X and *emerging* from Z.

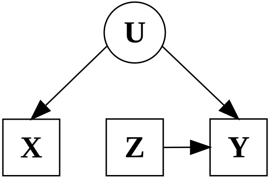
Examples of DAG Notation

G $\quad G_{\underline{X}}$ $\quad G_{\overline{X}}$ $\quad G_{\overline{X},\underline{U}}$

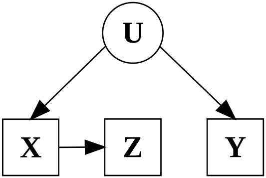


Example of DAG Notation

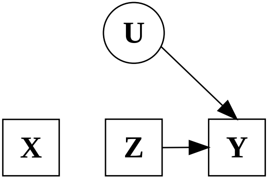
$G_{\underline{X}} = G_{\overline{Z}}$



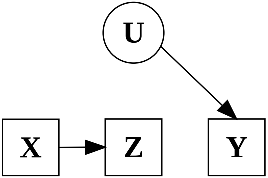
$G_{\underline{Z}}$



$G_{\overline{X},\overline{Z}}$



$G_{\overline{X},\underline{Z}}$



Do-calculus Rules

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

$$Y \perp\!\!\!\perp Z | (X, W) \text{ under } G_{\overline{X}} \Rightarrow \mathbf{P}(Y | do(X), Z, W) = \mathbf{P}(Y | do(X), W)$$

- **Rule 2:** *Action/observation exchange:*

$$Y \perp\!\!\!\perp Z | (X, W) \text{ under } G_{\overline{X}, \underline{Z}} \Rightarrow \mathbf{P}(Y | do(X), do(Z), W) = \mathbf{P}(Y | do(X), Z, W)$$

- **Rule 3:** *Insertion/deletion of actions:*

$$Y \perp\!\!\!\perp Z | (X, W) \,\&\, G_{\overline{X}, \overline{Z(W)}} \Rightarrow \mathbf{P}(Y | do(X), do(Z), W) = \mathbf{P}(Y | do(X), W)$$

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

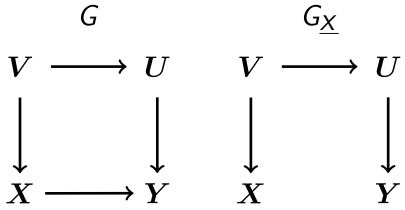
Understanding the Rules of Do-Calculus

Let G be a DAG then for any disjoint sets of variables X, Y, Z, W :
Rule 1: Insertion/deletion of observations

$$\text{If } \underbrace{Y \perp\!\!\!\perp Z|(X, W)}_{\text{Statistical Relation}} \text{ under } \underbrace{G_{\overline{X}}}_{\text{Graphic Criterion}} \text{ then}$$

$$\underbrace{\Pr(Y|do(X), Z, W) = \Pr(Y|do(X), W)}_{\text{Equivalent Probability Expression}}$$

Do-Calculus Exercise



1. LMC to X under $G_{\underline{X}}$ generates $X \perp\!\!\!\perp (U, Y)|V \Rightarrow X \perp\!\!\!\perp (U, Y)|V$.
2. Now if $X \perp\!\!\!\perp (U, Y)|V$ holds under $G_{\underline{X}}$, then, by **Rule 2**,

$$\mathbf{P}(Y|do(X), V) = \mathbf{P}(Y|X, V).$$

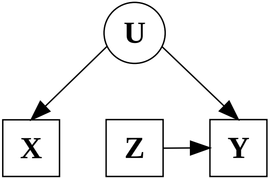
$$\therefore E(Y|do(X) = x) = \underbrace{\int E(Y|V = v, do(X) = x) dF_V(v)}_{\text{Using do(X),i.e. Fixing } X}$$

$$= \underbrace{\int E(Y|V = v, X = x) dF_V(v)}_{\text{Replace ``do'' with Standard Statistical Conditioning}}$$

Do-Calculus Exercise : The Front-door Model

Using the Do-Calculus : Task 1 – Compute $\Pr(Z|do(X))$

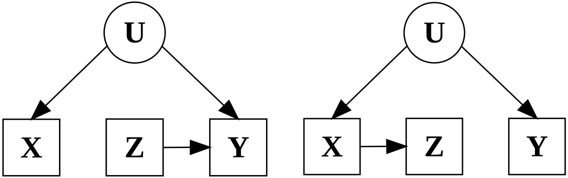
$G_{\underline{X}}$



$X \perp\!\!\!\perp Z$

- $X \perp\!\!\!\perp Z$ in $G_{\underline{X}}$, by **Rule 2**, $\Pr(Z|do(X)) = \Pr(Z|X)$.

Using the Do-Calculus : Task 2 – Compute $\Pr(Y|do(Z))$



$G_{\overline{Z}}$: $Z \perp\!\!\!\perp X$

$G_{\underline{Z}}$: $Z \perp\!\!\!\perp Y|X$

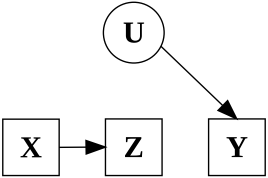
- $Z \perp\!\!\!\perp X$ in $G_{\overline{Z}}$, by **Rule 3**, $\Pr(X|do(Z)) = \Pr(X)$
- $Z \perp\!\!\!\perp Y|X$ in $G_{\underline{Z}}$, by **Rule 2**, $\Pr(Y|X, do(Z)) = \Pr(Y|X, Z)$

Adding these results, we have that:

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

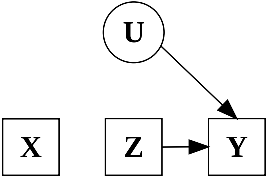
Using the Do-Calculus : Task 3 – Compute $\Pr(Y|Z, do(X))$

$G_{\overline{X},\underline{Z}}$



$Y \perp\!\!\!\perp Z|X$

$G_{\overline{X},\overline{Z}}$



$Y \perp\!\!\!\perp X|Z$

- $Y \perp\!\!\!\perp Z|X$ in $G_{\overline{X},\underline{Z}}$, by **Rule 2**, $\Pr(Y|Z, do(X)) = \Pr(Y|do(Z), do(X))$
- $Y \perp\!\!\!\perp X|Z$ in $G_{\overline{X},\overline{Z}}$, by **Rule 3**, $\Pr(Y|do(X), do(Z)) = \Pr(Y|do(Z))$

Adding these results, we have that:

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

Using the Do-Calculus : Final Task – Compute $\Pr(Y|do(X))$

Using Tasks 1,2 and 3, we have that:

$$\begin{aligned}
\therefore \Pr(Y|do(X)) &= \sum_{Z} \Pr(Y|Z, do(X)) \Pr(Z|do(X)) \\
&= \sum_{Z} \underbrace{\Pr(Y|do(Z), do(X))}_{\textbf{Task 3}} \Pr(Z|do(X)) \\
&= \sum_{Z} \underbrace{\Pr(Y|do(Z))}_{\textbf{Task 3}} \Pr(Z|do(X)) \\
&= \sum_{Z} \underbrace{\left(\sum_{X'} \Pr(Y|X', Z) \Pr(X') \right)}_{\textbf{Task 2}} \underbrace{\Pr(Z|X)}_{\textbf{Task 1}}
\end{aligned}$$

Summarizing Do-calculus of Pearl (2009) and Hypothetical Model Framework

Hypothetical Model	**Do-calculus**
Features in Common	*Features in Common*
Autonomy	**Autonomy**
(Frisch, 1938)	(Frisch, 1938)
Errors Terms:	**Error Terms:**
ϵ mutually independent	ϵ mutually independent
Statistical Tools:	**Statistical Tools:**
LMC and GA apply	LMC and GA apply
Counterfactuals:	**Counterfactuals:**
Fixing is a Causal Operation	Uses "do" for Fixing
Complete Method	**Complete Method**
Solution: Haavelmo's Inspired	Solution: Graphical/Statistical rules
Where They Depart	*Where They Depart*
Introduces	**Creates**
P_H (hypothetical model)	Three Graphical/Statistical rules
Identification:	**Identification:**
Connect P_H and P_E	Reiteration of do-calculus rules
Versatility:	**Versatility:**
Standard Statistical Tools apply	Standard Statistical Tools do not apply
	Need an extra statistical/graphical theory

Research Questions

1. The Do-calculus is complete, the hypothetical model was not shown to be complete.

2. Go beyond an algorithm.
   - Every DAG can be described by a Binary Matrix
   - Generate a criteria, i.e. a formula (not an algorithm) that determines if a causal effect is identified or not
   - Only need to test if the bridge pattern holds
   - The identification formula is immediate given the pattern

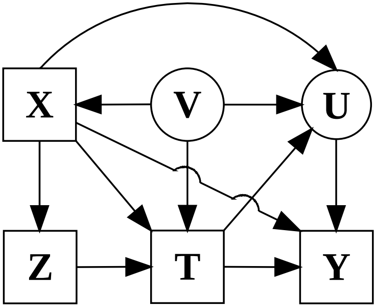
Do-Calculus Exercise : The Roy Model

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. **T:** treatment choice T 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.

Generalized Roy Model



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

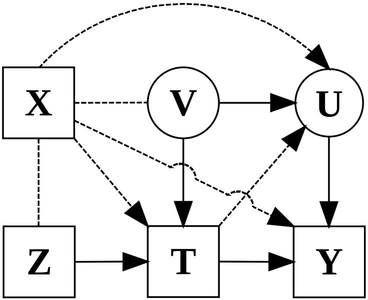
1. T is caused by Z, V;
2. U mediates the effects of V on Y (that is V causes U);
3. T and U cause Y and
4. Z (instrument) not caused by V, U and does not directly cause Y, U.

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 T,
4. X causes U,
5. T 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)



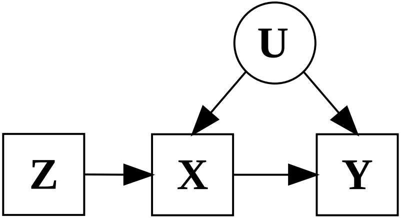
Dashed lines denote causal relations that may not exist or, if they exist, the causal direction can go either way. Dashed arrows denote causal relations that may not exist, but, if they exist, the causal direction must comply the arrow direction.

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

$$G = G_{\overline{Z}}$$



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

- $Y \perp\!\!\!\perp Z$ in $G_{\overline{X}}$, by **Rule 1**

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

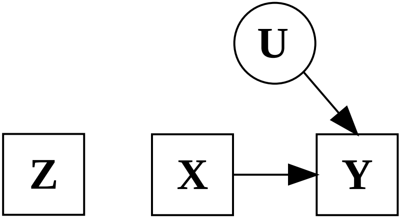
- $Y \perp\!\!\!\perp Z$, in $G_{\overline{X},\overline{Z}}$, by **Rule 3**

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

- $Y \perp\!\!\!\perp Z|X$ in $G_{\overline{X},\underline{Z}}$, by **Rule 2**

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

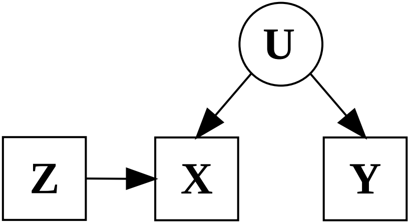
$$G_{\overline{X}} = G_{\overline{X},\overline{Z}} = G_{\overline{X},\underline{Z}}$$



Examining the Marginalized Roy Model – 2/4

- Under $G_{\overline{X}}$, $Y \not\perp\!\!\!\perp X$, thus **Rule 2** does not apply.
- Under $G_{\underline{X},\overline{Z}}$, $Y \not\perp\!\!\!\perp X|Z$, thus **Rule 2** does not apply.

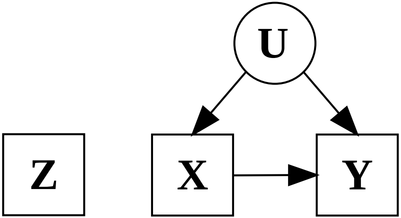
$$G_{\underline{X}} = G_{\underline{X},\overline{Z}}$$



Examining the Marginalized Roy Model – 3/4

- $G_{\underline{Z}} \Rightarrow Y \perp\!\!\!\perp Z$, thus by **Rule 2** $\Pr(Y|do(Z)) = \Pr(Y|Z)$.

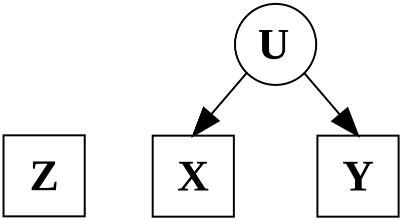
$G_{\underline{Z}}$



Examining the Marginalized Roy Model – 4 of 4 Modifications

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

$G_{\underline{X},\underline{Z}}$



Conclusion of Do-calculus and the Roy Model

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.

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

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

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