Causal Inference in Machine Learning and AI

Advanced Applied Econometrics (B-KUL-D0S91A)

Paul Hünermund

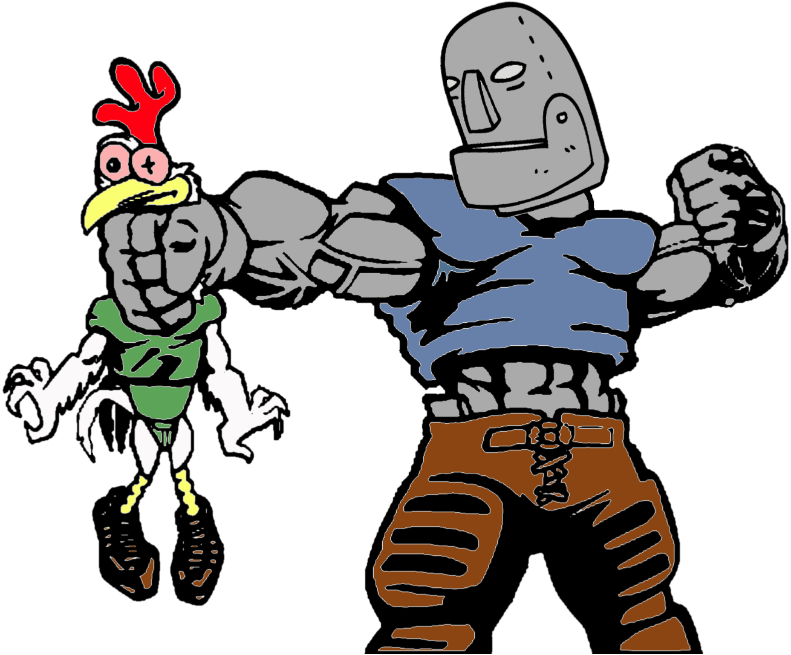
Maastricht University, School of Business and Economics, Tongersestraat 53, 6211LM Maastricht, The Netherlands

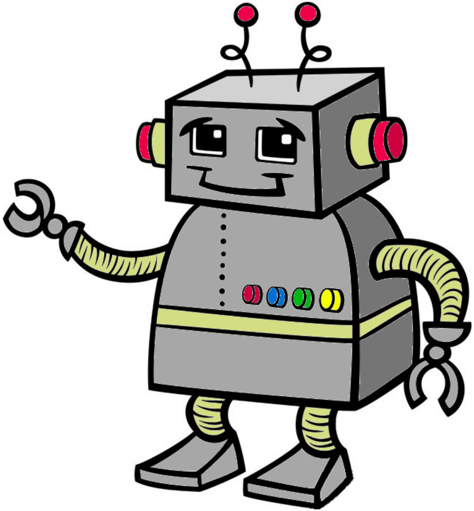
May 7, 2020

Introduction

- Causal inference is arguably the most important goal in econometrics
  - Inform policy-makers, legislators, and managers about the likely impact of their actions by uncovering quantitative relationships in statistical data (Frisch, 1933)
- Since the end of the 1980s, an extensive literature on causal inference was developed in the computer science and artificial intelligence field (Pearl and Mackenzie, 2018)
  - Builds on the graph-theoretic approach to causality developed by Pearl (1995)
  - Interest emerged from older AI techniques such as Markov random fields and Bayesian nets (Pearl, 1988)
  - Shares several mutual intellectual roots with econometrics (Strotz and Wold, 1960)
- Aim of this talk:
  - Review the (newer) advances in the causal AI literature
  - Show how management scholars can benefit from adopting these techniques
  - Foster mutual knowledge exchange between the two communities

How can we prevent a future robot from trying to make the rooster crow at 3am in order to make the sun come up?





“Beyond Curve Fitting” in Machine Learning and AI

> *“To Build Truly Intelligent Machines, Teach Them Cause and Effect”*
> — Judea Pearl, ACM Turing Award winner

- The notion of causality is a fundamental concept in human thinking
- Current ML / AI techniques remain purely prediction-based (Agrawal et al., 2018)
- In other words: machine learning is very sophisticated, high-dimensional curve fitting
- But nothing in the theoretical basis of ML allows to capture the asymmetry inherent to causal relationships
- If we want machines to be able to interact meaningfully with us, they should be equipped with a notion of cause and effect

Motivating Example: How to Estimate the Gender Pay Gap?

- The New York Times reported in March 2019:
  - *"When Google conducted a study recently to determine whether the company was underpaying women and members of minority groups, it found, to the surprise of just about everyone, that men were paid less money than women for doing similar work."*

    `https://www.nytimes.com/2019/03/04/technology/google-gender-pay-gap.html`
- The study led Google to increase the pay of its male employees to fight this blatant discrimination of men
- What's going on here? Wasn't Google just recently accused of discriminating against women, not men?
  - *"Department of Labor claims that Google systematically underpays its female employees"*

    `https://www.theverge.com/2017/4/8/15229688/department-of-labor-google-gender-pay-gap`

Simpson's Paradox

- Suppose we collected data on wages payed to 100 women and 100 men in company X. We observe the following distribution of average monthly salaries for women and men in management and non-management positions (case numbers in parentheses). And our goal is to estimate the magnitude of the gender pay gap in company X. How should we tackle this problem?

	Female	Male
Non-management	\$3163.30 (87)	\$3015.18 (59)
Management	\$5592.44 (13)	\$5319.82 (41)

Simpson's Paradox (II)

- On average, women earn less in this example

$$\left(\frac{87}{100}\cdot\$3163.30+\frac{13}{100}\cdot\$5592.44\right)-\left(\frac{59}{100}\cdot\$3015.18+\frac{41}{100}\cdot\$5319.82\right)$$
$$\approx-\$481$$

- But in each subcategory women actually have higher salaries?
  - Non-management: $\$3163.30-\$3015.18=\$148.12$
  - Management: $\$5592.44-\$5319.82=\$272.62$
- Conditioning on job position gives adjusted gender pay gap

$$\frac{87+59}{200}\cdot\$148.12+\frac{13+41}{200}\cdot\$272.62\approx\$181.74$$

- Which estimate gives us a more accurate picture of the gender pay gap?

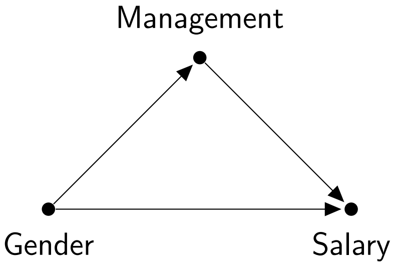
Simpson’s Paradox (III)

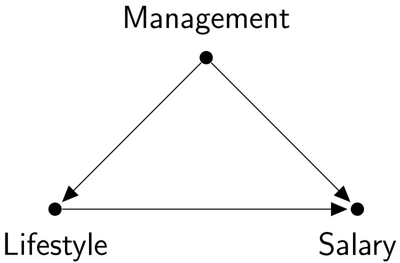
- The phenomenon that a statistical association, which holds in a population, can be reversed in every subpopulation is named after the British statistician Edward Simpson
- Simpson’s paradox well-known, for example, in epidemiology and labor economics
- Here, the unadjusted gender pay (−$481) gap gives the right answer
- But what about this example?

	Healthy Lifestyle	Unhealthy Lifestyle
Non-management	$3163.30 (87)	$3015.18 (59)
Management	$5592.44 (13)	$5319.82 (41)

Simpson's Paradox (IV)

- Here we would correctly infer that people with a healthy lifestyle earn more on average ($181.74). What is the difference between the two examples?





Simpson's Paradox (V)

- Statistics alone doesn't help us to answer this question
- Note that the joint distribution of salaries is the same in both cases
- Both problems are thus identical from a statistical point of view
- Instead, we need to make causal assumptions in order to come to a conclusion here
  - Gender affects both a person's salary level and job position
  - Whereas, life style affects salaries, but is itself affected by a person's job position
- After the course you will know how to incorporate this kind of causal knowledge in your analysis in order to solve all sorts of practical problems of causal inference

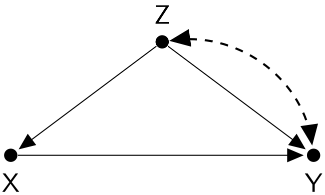
Structural Causal Models

$$
\begin{aligned}
z &\leftarrow f_Z(u_z) \\
x &\leftarrow f_X(z, u_x) \\
y &\leftarrow f_Y(x, z, u_Y)
\end{aligned}
$$

- The f_i's denote the causal mechanisms in the model
  - Are not restricted to be linear as in traditional SEM
- The u_i's refer to background factors that are determined outside of the model
- Assignment operator ($\leftarrow$) captures asymmetry of causal relationships
  - $x \leftarrow a \cdot z \;\neq\; z \leftarrow x/a$
- Similar to definition of "structure" according to Cowles foundation

Directed Acyclic Graphs

$$z \leftarrow f_Z(u_z)$$
$$x \leftarrow f_X(z, u_x)$$
$$y \leftarrow f_Y(x, z, u_Y)$$



- In a fully specified SCM, every counterfactual quantity is computable
- In most social science contexts it's hard to know the causal mechanisms f_i and distribution of background factors $P(U)$
- Therefore, restrict attention to qualitative causal information of the model, which can be encoded by a graph G
    - Nodes V: variables in the model
    - Directed edges E: causal relationships in the model

Directed Acyclic Graphs

- No functional form or distributional assumptions means that framework remains fully nonparametric
  - Particularly helpful in fields where theory is purely qualitative and no shape restrictions on can be derived (Matzkin, 2007)
- $Z \dashleftarrow\dashrightarrow Y$ is a shortcut notation for unobserved common causes $Z \leftarrow U \rightarrow Y$
- Acyclicity
  - Directed cycles such as $A \rightarrow B \rightarrow C \rightarrow A$ are excluded
  - This means there are no feedback loops
  - Otherwise A could be a cause of itself
  - Gives rise to what economists call a *recursive* model (Wold, 1954)
  - Extensions of the SCM framework to cyclic graphs exist (Spirtes et al., 2000; Pearl, 2009)

Structural Econometrics vs. PO Recursive vs. Interdependent Systems

D-Separation

- DAGs are such a useful tool because they are able to efficiently encode conditional independence relationships:

$$\begin{array}{llcl}
\underline{\text{Chain:}} & A \to B \to C & \Rightarrow & A \not\perp\!\!\!\perp C \quad \text{and} \quad A \perp\!\!\!\perp C | B \\
\underline{\text{Fork:}} & A \leftarrow B \to C & \Rightarrow & A \not\perp\!\!\!\perp C \quad \text{and} \quad A \perp\!\!\!\perp C | B \\
\underline{\text{Collider:}} & A \to B \leftarrow C & \Rightarrow & A \perp\!\!\!\perp C \quad \text{and} \quad A \not\perp\!\!\!\perp C | B
\end{array}$$

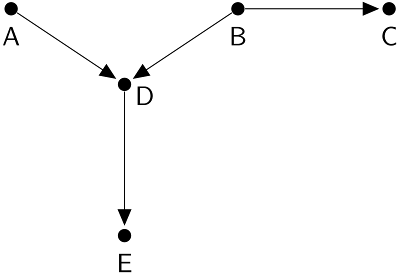
- The same holds for longer paths in the graph
  - Conditioning on a variable along a chain or fork blocks (*“d-separates”*) the path
  - Conditioning on a collider opens the path

Colliders – R example

```r
# Create two independent normally distributed variables
x <- rnorm(1000)
y <- rnorm(1000)

# Construct z as being equal to one if x + y > 0, and zero
   otherwise
z <- 1*(x + y > 0)

# By design, there is no correlation between x and y
cor(x, y)

# But if we condition on z==1, we find a negative correlation
cor(x[z==1], y[z==1])
```

Testable Implications

- D-separation provides testable implications of a model



Testable implications:

$A \perp\!\!\!\perp B$ $\quad$ $A \perp\!\!\!\perp C$

$A \perp\!\!\!\perp E | D$ $\quad$ $B \perp\!\!\!\perp E | D$

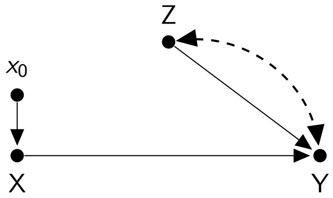
$C \perp\!\!\!\perp D | B$ $\quad$ $C \perp\!\!\!\perp E | D$

$C \perp\!\!\!\perp E | B$

- If one of these conditional independence relations do not hold in the data, the model can be rejected
- *"Causal discovery"*: try to learn compatible model from conditional independence relations found in the data

Interventions in Structural Causal Models

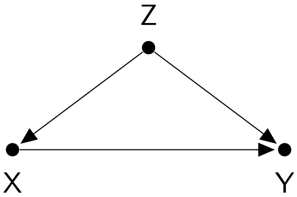
$$z \leftarrow f_Z(u_z)$$
$$x \leftarrow x_0$$
$$y \leftarrow f_Y(x, z, u_Y)$$



- Causal inference $\triangleq$ predict the effects of interventions (policy initiatives, social programs, management initiatives, etc.)
- Interventions in SCMs amount to *"wiping out"* of causal mechanisms, an idea that originally came from econometrics (Strotz and Wold, 1960)
    - Delete naturally occurring causal mechanism $f_X(\cdot)$ from model and set X to constant value x_0
    - This operation is denoted by *do-operator*: $do(X = x_0)$
- Query of interest: post-interventional distribution $P(Y = y|do(X = x))$

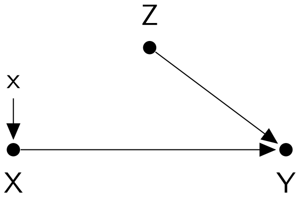
Pre- versus Post-intervention Distribution

Pre-intervention



$$
\begin{aligned}
Z &= f_z(u_z) \\
X &= f_x(Z, u_x) \\
Y &= f_y(X, Z, u_y)
\end{aligned}
$$

Post-intervention



$$
\begin{aligned}
Z &= f_z(u_z) \\
X &= x \\
Y &= f_y(X, Z, u_y)
\end{aligned}
$$

- The intervention changes the data-generating process; thus, $P(Y|X)$ (pre-intervention) is generally not equal to $P(Y|do(X))$ (post-intervention)

“Correlation doesn’t imply causation” – R example

```
# Create background factors for nodes
e_x <- rnorm(10000)
e_y <- rnorm(10000)
e_z <- rnorm(10000)

# Create nodes for the DAG: y <- x, y <- z, x <- z
z <- 1*(e_z > 0)
x <- 1*(z + e_x > 0.5)
y <- 1*(x + z + e_y > 2)
y_dox <- 1*(1 + z + e_y > 2)

# We see that P(y|do(x=1)) is not equal to P(y|x=1)
mean(y_dox)
mean(y[x==1])
```

Do-Calculus

Theorem: Rules of *do* Calculus (Pearl, 2009, p. 85)

Let G be the directed acyclic graph associated with a [structural] causal model [...], and let $P(\cdot)$ stand for the probability distribution induced by that model. For any disjoint subset of variables X, Y, Z, and W, we have the following rules.

Rule 1 (Insertion/deletion of observations):

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}.$$

Rule 2 (Action/observation exchange): Illustration

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}.$$

Rule 3 (Insertion/deletion of actions):

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\overline{Z(W)}}},$$

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

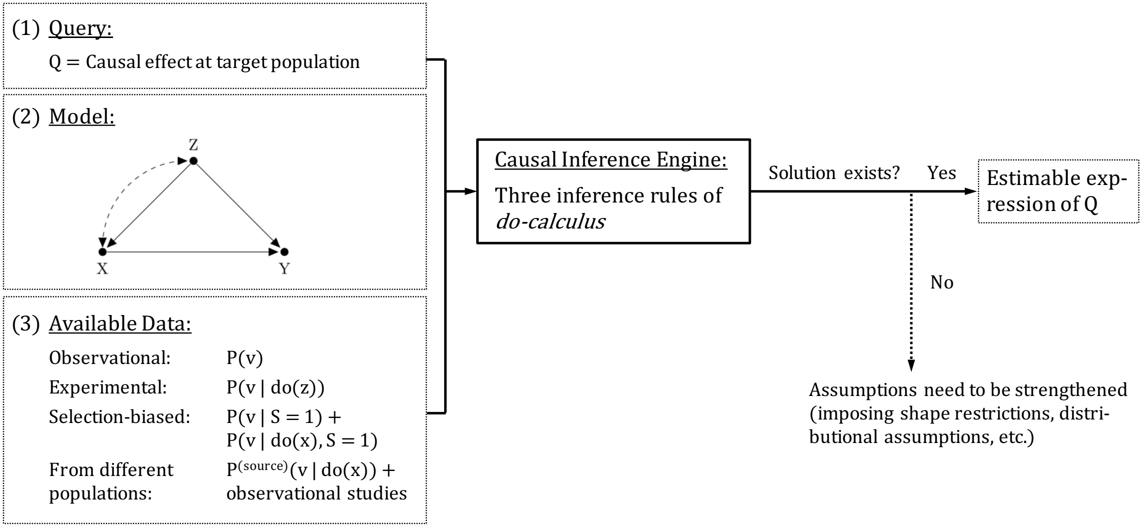
Do-Calculus

- *Do-calculus* is a powerful symbolic machinery that provides a set of inference rules by which sentences involving do-interventions can be transformed into other sentences (Pearl, 2009; Pearl et al., 2016)
  - Apply the rules of *do*-calculus repeatedly until a do-expression is translated into an equivalent expression that can be estimated from the data
- It can be shown that do-calculus is *complete* for many identification tasks
  - I.e., if a causal effect is identifiable there exists a sequence of steps applying the rules of *do*-calculus that transforms the causal effect formula into an expression that includes only observable quantities (Shpitser and Pearl, 2006; Huang and Valtorta, 2006)
  - Put differently, if *do*-calculus fails, the causal effect is guaranteed to be unidentifiable
- Furthermore, it can be shown that many do-calculus tasks can be fully automated (Bareinboim and Pearl, 2016)

Applications of Do-Calculus

- The three inference rules of do-calculus can be used to solve many recurrent problems in applied econometric work
  1. Dealing with confouding bias
  2. Using surrogate experiments to deal with confounding (generalized IV)
  3. Recover from selection bias
  4. Extrapolate causal knowledge across heterogeneous settings ("external validity")
- This also illustrates that all of these problems are essentially causal questions

The Data Fusion Process



(1) Query:

Q = Causal effect at target population

(2) Model:

(3) Available Data:

Observational:	$P(v)$
Experimental:	$P(v \mid do(z))$
Selection-biased:	$P(v \mid S = 1)$ + $P(v \mid do(x), S = 1)$
From different populations:	$P^{(source)}(v \mid do(x))$ + observational studies

Causal Inference Engine:

Three inference rules of *do-calculus*

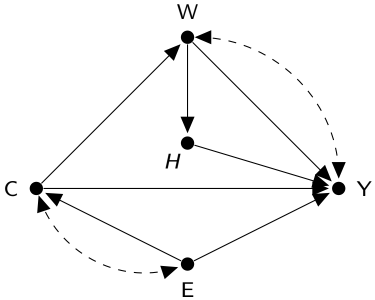
Estimable expression of Q

Assumptions need to be strengthened (imposing shape restrictions, distributional assumptions, etc.)

Confounding Bias

Task: Use the rules of do-calculus to transform $Q = P(y|do(x))$ into an expression that only contains standard probability objects

- We also call this task **"identification"** (Pearl, 2009; Matzkin, 2007)
- Take the stylized example of the college wage premium
  - C: college degree
  - Y: earnings
  - W: occupation
  - H: work-related health
  - E: other socio-economic factors



Confounding Bias

- We could find $P(y|do(c))$ by applying the rules of do-calculus Do-Calculus Derivation
- An easier solution can be found, however, by recognizing that there are only two *backdoor paths* that create a spurious association between C and Y
  1. $C \leftarrow E \rightarrow Y$
  2. $C \dashleftarrow\!\dashrightarrow E \rightarrow Y$
- We can close both of these backdoor paths by adjusting for E
- The causal effect can then be identified by the adjustment formula

$$P(Y = y|do(C = c)) = \sum_{e} P(Y = y|C = c, E = e)P(E = e)$$

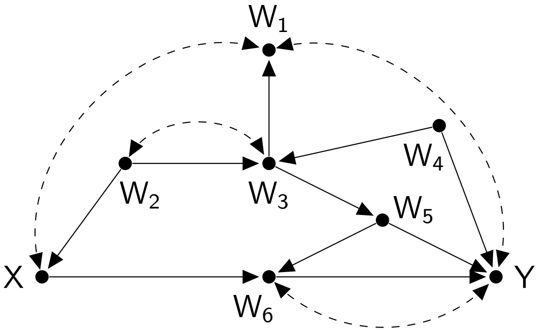
Backdoor Adjustment

Definition: The Backdoor Criterion (Pearl et al., 2016, p. 61)

Given an ordered pair of of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the backdoor criterion relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.

- Intuition: block all spurious paths between X and Y while leaving direct paths unperturbed and creating no new spurious paths
- Note how adjusting for occupation W would open up the path $C \rightarrow W \dashleftarrow\dashrightarrow Y$ (**"collider bias"**)
- Finding suitable adjustment sets Z can be easily automated (Textor et al., 2011)

Backdoor Adjustment



- Minimum sufficient adjustment sets:

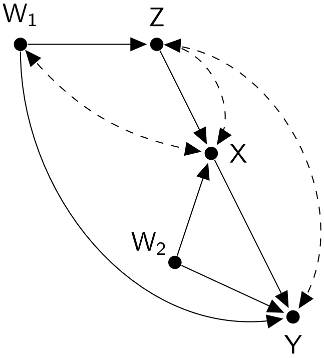
$$Z = \{\{W_2\}, \{W_2, W_3\}, \{W_2, W_4\}, \{W_3, W_4\}, \{W_2, W_3, W_4\}, \{W_2, W_5\},$$
$$\{W_2, W_3, W_5\}, \{W_4, W_5\}, \{W_2, W_4, W_5\}, \{W_3, W_4, W_5\}, \{W_2, W_3, W_4, W_5\}\}$$

Backdoor Adjustment – R example

```
# Define DGP as in previous R example

# Adjustment formula: P(y|do(x=1)) = P(y|x=1, z=1)*P(z=1) + P(y|x
    =1, z=0)*P(z=0)
mean(y[x==1 & z==1]) * mean(z==1) + mean(y[x==1 & z==0]) * mean(z
    ==0)

# Estimation via inverse probability weighting
df <- df %>% group_by(z) %>% mutate(weight = mean(x))
weight <- df$weight
y_weighted <- y / weight
```

Identification by Surrogate Experiments

- In many settings, simple covariate adjustment is not feasible
- At the same time, conducting an RCT in X might not be feasible
- What if we are able to experimentally manipulate a third variable Z?
- Surrogate experiments are ubiquitous in economics
  - E.g., "encouragement designs" in development economics (Duflo et al., 2008)



Identification by Surrogate Experiments

Task: (z-identification) Use the rules of do-calculus to transform $Q = P(y|do(x))$ into an expression that only contains standard probability objects and $do(Z)$

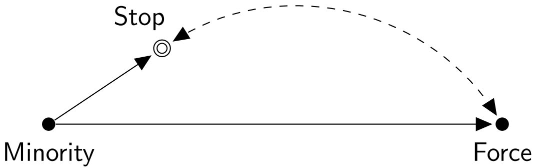
- Do-calculus provides a complete solution for the z-identification problem (Bareinboim and Pearl, 2012a)
- In the previous graph the solution is given by

$$P(y|do(x)) = \sum_{w_1, w_2} P(y|do(z), x, w_1, w_2)P(w_1)P(w_2)$$

- $\mathcal{Z}$-identification implies that the post-intervention distribution is *nonparametrically* identified
  - No shape restrictions (e.g., monotonicity) or distributional assumptions required
  - Standard IV only identifies a LATE (Imbens and Angrist, 1994) zID vs. IV

Selection Bias

- Non-random, selection-biased data is a frequent problem in economics
- For example, Knox et al. (2019) criticize papers that try to measure the degree of racial-bias in policing with the help of administrative records
  - Problem: An individual only appears in the data, if it was stopped by the police
  - If there is a racial bais in policing, stopping can be the result of minority status
  - There are unobserved confounders, such as officers' suspicion, between the selection variable and outcome



Selection Bias

Task: Use the rules of do-calculus to transform $Q = P(y|do(x))$ into an expression that only contains probabilities conditional on $S = 1$

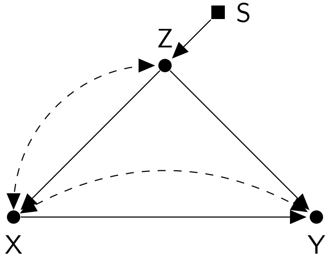
- There is a principled solution for dealing with selection bias in DAGs based on do-calculus (Bareinboim and Pearl, 2012b; Bareinboim et al., 2014; Bareinboim and Tian, 2015) Example Derivation
- Compared to standard approaches in econometrics, these results do not rely on
  - functional-form assumptions about the selection propensity score $P(S|PA)$ (Heckman, 1979)
  - or, ignorability of the selection mechanism (Angrist, 1997)
- In addition, there is the possibility to combine biased and unbiased data in order to increase identifying power (Bareinboim et al., 2014; Correa et al., 2017)
  - In many applied settings, unbiased records of covariates are available from secondary data sources (e.g., census data)

Transportability

- Causal knowledge is usually acquired in different contexts than it is supposed to be used (e.g., in a laboratory experiment)
- If domains differ structurally in important ways, how can we be sure that causal knowledge remains valid across contexts?
- This problem is known under the rubric of *"transportability"* in the causal AI field
- Economists more frequently use the term *"external validity"*
- Example: Banerjee et al. (2007) study the effect of a randomized remedial education program for third and fourth graders in two Indian cities: Mumbai and Vadodara
    - They find similar effects on math skills, but effect positive impact on language proficiency is much smaller in Mumbai compared to Vadodara

Transportability

- Banerjee et al. (2007) explain this result by baseline reading skills that were higher in Mumbai, because families are wealthier there and schools are better equipped
- What do we do if we do not have a second experiment to validate our results?
- We can incorporate knowledge about structural differences across domains by a selection node (■) in a causal diagram
  - Captures the notion that domains differ either in the distribution of background factors $P(U_i)$ or causal mechanisms f_i in the underlying structural causal model



Transportability

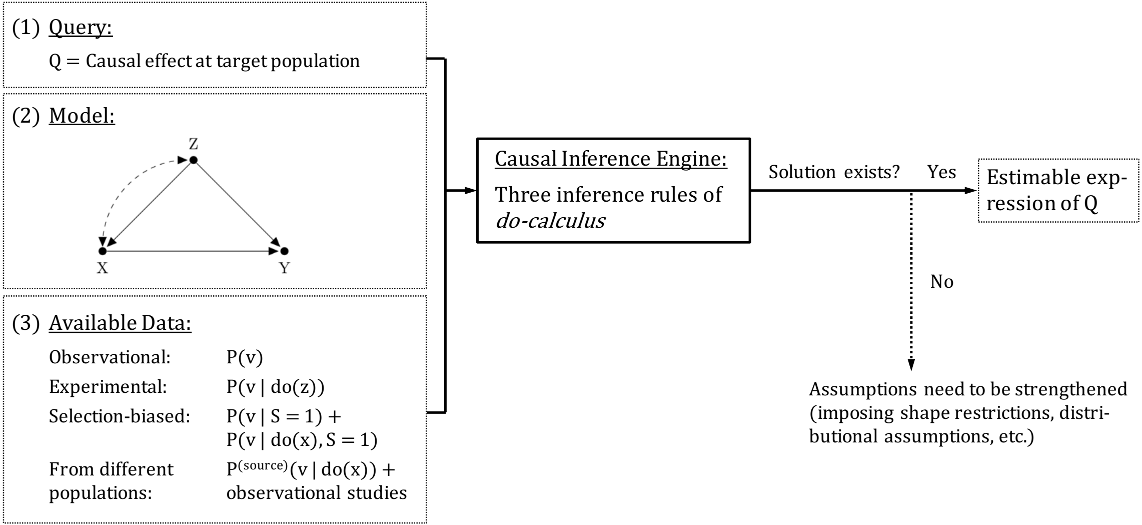
Task: Use the rules of do-calculus to express causal query $Q = P^*(y|do(x))$ in target domain with the help of causal knowledge in a source domain (Pearl and Bareinboim, 2011)

- Bareinboim and Pearl (2013a) develop a complete nonparametric solution for this task based on the selection diagram Example
- Moreover, there is the possibility to combine causal knowledge from several different source domains (Bareinboim and Pearl, 2013b)
  - Meta-analyses are becoming increasingly popular in economics (Card et al., 2010; Dehejia et al., 2015)
  - However, by simply averaging out results, they completely disregard potential domain heterogeneity
- Possibility to combine transportability with idea z-identification to what is called "mz-transportability" (Bareinboim and Pearl, 2014)

Algrithmatization of Causal Inference

- There exist algorithmic solutions for all the inference tasks just discussed
  - Dealing with confounding bias (Tian and Pearl, 2002; Shpitser and Pearl, 2006)
  - $\mathcal{Z}$-Identification (Bareinboim and Pearl, 2012a)
  - Selection bias (Bareinboim and Tian, 2015)
  - Transportability (Bareinboim and Pearl, 2013a, 2014)
- Input:
  1. A causal query Q
  2. The model in form of a diagram
  3. The type of data available
- Output: an estimable expression of Q
  - Most algorithms inherit *completeness* property from do-calculus
- Analyst can fully concentrate on the modeling and the scientific content, the identification is done automatically

The Data Fusion Process



(1) Query:

$Q =$ Causal effect at target population

(2) Model:

(3) Available Data:

Observational: $P(v)$

Experimental: $P(v \mid do(z))$

Selection-biased: $P(v \mid S = 1) +$ $P(v \mid do(x), S = 1)$

From different populations: $P^{(source)}(v \mid do(x)) +$ observational studies

Causal Inference Engine:

Three inference rules of *do-calculus*

Solution exists? Yes

Estimable expression of Q

No

Assumptions need to be strengthened (imposing shape restrictions, distributional assumptions, etc.)

Conclusion

- Graphical models of causation provide a unified framework for causal inference that allow to solve many of the recurrent problems econometricians face
  - Unique perspective on many applied problems
  - DAGs and do-calculus are a powerful tool for causal inference
  - Combine the strengths of both structural econometrics (identifying assumptions are stated clearly, no *"black box"* character) and the potential outcomes framework (fully nonparametric, easy to apply)
- Possibilities to fully automatize the identification task ($\Rightarrow$ "causal AI")
- Paper: "Causal Inference and Data-Fusion in Econometrics" (Hünermund and Bareinboim, 2019)
- Teaching material available at: https://p-hunermund.com/teaching/

Thank you

Personal Website:	p-hunermund.com
Twitter:	@PHuenermund
Email:	p.hunermund@maastrichtuniversity.nl

Structural Econometrics vs. Potential Outcomes

- Econometrics is currently dominated by two competing streams
- Structural econometrics
  - In practice, relies on distributional assumptions and (parametric) shape restrictions
  - Work by, e.g., Matzkin (2007) that aims to relax parametric assumptions, but
    - still relies on (weaker) shape restrictions, and is not widely adopted in applied work
- Potential outcomes framework (Rubin, 1974; Imbens and Rubin, 2015)
  - Does impose crucial identifying assumptions (e.g., ignorability) without reference to an underlying model ("black box character")
    - A feature that has been frequently criticized by the structural camp (e.g., by Rosenzweig and Wolpin, 2000 and Heckman and Urzua, 2009)
  - In practice, causal inference in PO boils down to the four "tricks of the trade" (matching, IV, RDD, difference-in-differences)

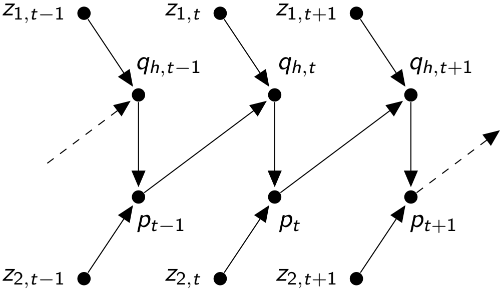
⇒ DAGs are a perfect "middle ground" between structural econometrics and PO

Recursive Versus Interdependent Systems

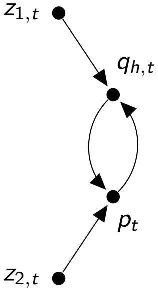
- DAGs represent recursive systems, but many standard models in economics are interdependent (Marshallian cross, game theory, etc.)
- This connects to an old debate within econometrics about the causal interpretation of recursive versus interdependent models that emerged in the aftermath of Haavelmo's celebrated 1943 paper
- One central argument (Bentzel and Hansen, 1955; Strotz and Wold, 1960):
  - Individual equations in an interdependent model do not have a causal interpretation *in the sense of a stimulus-response relationship* (Strotz and Wold, 1960, p. 417)
  - Interdependent systems with equilibrium conditions are regarded as a *shortcut* (Wold, 1960; Imbens, 2014) description of the underlying dynamic behavioral processes

Recursive Versus Interdependent Systems

- In this context, Strotz and Wold (1960) discuss the example of the cobweb model:



$$p_{t-1} = p_t$$
$$\Rightarrow$$



$$\begin{aligned} q_{h,t} &\leftarrow \gamma + \delta p_{t-1} + \nu z_{1,t} + u_{1,t}, \\ p_t &\leftarrow \alpha - \beta q_{h,t} + \varepsilon z_{2,t} + u_{2,t}. \end{aligned}$$

$$\begin{aligned} q_{h,t} &\leftarrow \gamma + \delta p_t + \nu z_{1,t} + u_{1,t} \\ p_t &\leftarrow \alpha - \beta q_{h,t} + \varepsilon z_{2,t} + u_{2,t} \end{aligned}$$

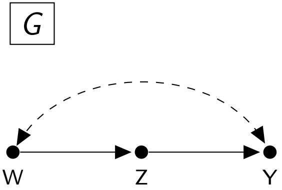
Recursive Versus Interdependent Systems

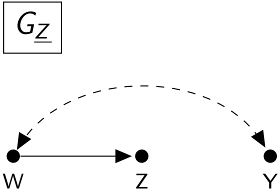
- However, equilibrium assumption $p_{t-1} = p_t$ carries no behavioral interpretation
- Individual equations in interdependent system do not represent autonomous causal relationships in the stimulus-response sense (Heckman and Pinto, 2013)
  - Endogenous variables are determined jointly by all equations in the system (Matzkin, 2013)
  - Not possible, e.g., to directly manipulate p_t to bring about a desired change in $q_{h,t}$
- Equilibrium models can of course still be useful for learning about causal parameters
- But, if individual mechanisms are supposed to be interpreted as stimulus-response relationships, cyclic patterns need to be excluded (Woodward, 2003; Cartwright, 2007)
  - For this reason, potential outcomes framework (Rubin, 1974; Imbens and Rubin, 2015) also implicitly maintains the assumption of acyclicity (Heckman and Vytlacil, 2007)

Back

Do-Calculus Rule 2

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \qquad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}$$





Assume we are interested in the query $Q = P(y|do(z), w)$.

Denote the resulting graph when all arrows emitted by Z in G are deleted by $G_{\underline{Z}}$.

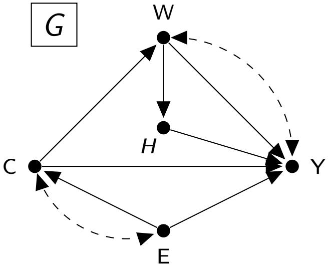
In $G_{\underline{Z}}$, W blocks the only backdoor path between Z and Y: $Z \leftarrow W \dashleftarrow\dashrightarrow Y$.

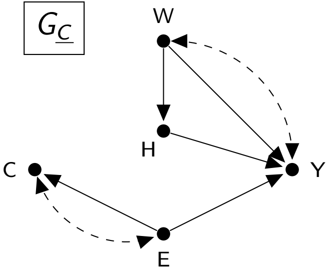
Thus, by d-separation $(Y \perp\!\!\!\perp Z|W)_{G_{\underline{Z}}}$ and therefore the second rule of do-calculus applies.

Consequently, we can get rid of the do-operator by setting $P(y|do(z), w) = P(y|z, w)$. The latter expression is estimable from observational data.

Back

Do-Calculus Example





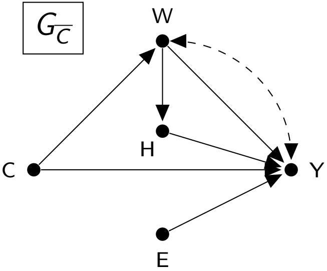
Consider the causal effect of C on Y in graph G. There are two backdoor paths in G, which can both be blocked by E. Conditioning and summing over all values of E yields

$$P(y|do(c)) = \sum_e P(y|do(c), e)P(e|do(c)).$$

By rule 2 of do-calculus

$$P(y|do(c), e) = P(y|c, e), \quad \text{since } (Y \perp\!\!\!\perp C|E)_{G_{\underline{C}}}.$$

Do-Calculus Example



By rule 3 of do-calculus

$$P(e|do(c)) = P(e), \quad \text{since } (E \perp\!\!\!\perp C)_{G_{\overline{C}}}.$$

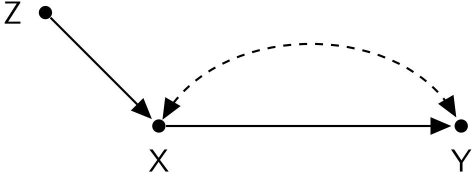
It follows that

$$P(y|do(c)) = \sum_{e} P(y|c,e)P(e).$$

The right-hand-side expression is do-free and can therefore be estimated from observational data.

Back

Z-Identification Versus Instrumental Variables

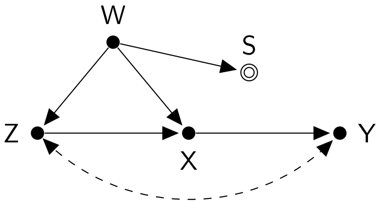


- This is the canonical IV setup with endogenous X ($X \dashleftarrow\dashrightarrow Y$), Z is both relevant ($Z \rightarrow X$) and excludable ($Z \not\rightarrow Y$)
- But effect of X on Y is *not* z-identifiable (condition (ii) in Theorem 3 of Bareinboim and Pearl (2012a) is violated)
  - IV estimator is not nonparametrically identified (Balke and Pearl, 1995)
  - We need to either introduce shape restrictions (e.g., monotonicity; Imbens and Angrist, 1994) or resort to partial identification (Manski, 1990)

Back

Selection Bias Example Derivation

Take the following DAG augmented with selection node S:



By the first rule of do-calculus, since $(S, W \perp\!\!\!\perp Y)$ in $G_{\overline{X}}$ (the resulting graph when all incoming arrows in X are deleted),

$$P(y|do(x)) = P(y|do(x), w, S = 1),$$

$$= \sum_{z} P(y|do(x), z, w, S = 1)P(z|do(x), w, S = 1),$$

where the second line on the previous slide follows from conditioning on Z.

Selection Bias Example Derivation

Applying rule 2, since $(Y \perp\!\!\!\perp X | W, Z)$ in $G_{\underline{X}}$ (the resulting graph when all arrows emitted by X are deleted), we can eliminate the do-operator in the first term

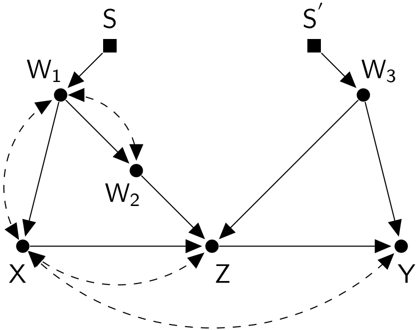
$$= \sum_{z} P(y|x, z, w, S = 1) P(z|do(x), w, S = 1).$$

Finally, because $(Z \perp\!\!\!\perp X | W)$ in $G_{\overline{X}}$, it follows from rule 3 that

$$= \sum_{z} P(y|x, z, w, S = 1) P(z|w, S = 1).$$

Back

Transportability Example



The causal effect of X on Y in the target domain π^* can be found by the algorithm developed in Bareinboim and Pearl (2013a) is given by

$$P^*(y|do(x)) = \sum_{z,w2,w3} P(y|do(x), z, w2, w3)P(z|do(x), w2, w3)P^*(w2, w3)$$

Back

References I

Agrawal, A., Gans, J., and Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press.

Angrist, J. D. (1997). Conditional independence in sample selection models. *Economics Letters*, 54:103–112.

Balke, A. and Pearl, J. (1995). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92:1171–1176.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quartely Journal of Economics*, 122(3):1235–1264.

Bareinboim, E. and Pearl, J. (2012a). Causal inference by surrogate experiments: z-identifiability. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, pages 113–120.

References II

Bareinboim, E. and Pearl, J. (2012b). Controlling selection bias in causal inference. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, pages 100–108.

Bareinboim, E. and Pearl, J. (2013a). A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference*, 1(1):107–134.

Bareinboim, E. and Pearl, J. (2013b). Meta-transportability of causal effects: A formal approach. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 31, Scottsdale, AZ.

Bareinboim, E. and Pearl, J. (2014). Transportability from multiple environments with limited experiments: Completeness results. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances of Neural Information Processing Systems*, volume 27, pages 280–288.

Bareinboim, E. and Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352.

References III

Bareinboim, E. and Tian, J. (2015). Recovering causal effects from selection bias. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Bareinboim, E., Tian, J., and Pearl, J. (2014). Recovering from selection bias in causal and statistical inference. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Bentzel, R. and Hansen, B. (1954 - 1955). On recursiveness and interdependency in economic modeks. *The Review of Economic Studies*, 22(3):153–168.

Card, D., Kluve, J., and Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *The Economic Journal*, 120:452–477.

Cartwright, N. (2007). *Hunting Causes and Using Them*. Cambridge University Press.

Correa, J. D., Tian, J., and Bareinboim, E. (2017). Generalized adjustment under confounding and selection biases. Technical Report R-29-L.

References IV

Dehejia, R., Pop-Eleches, C., and Samii, C. (2015). From local to global: External validity in a fertility natural experiment. NBER Working Paper No. 21459.

Duflo, E., Glennerster, R., and Kremer, M. (2008). Using randomization in development economics research: A toolkit. In *Handbook of Development Economics*, volume 4, chapter 61. Elsevier.

Frisch, R. (1933). Editor's note. *Econometrica*, 1(1):1–4.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11(1):1–12.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Heckman, J. J. and Pinto, R. (2013). Causal Analysis after Haavelmo. *Econometric Theory*, 31:115–151.

References V

Heckman, J. J. and Urzua, S. (2009). Comparing IV with Structural Models: What Simple IV Can And Cannot Identify. NBER Working Paper 14706.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric evaluation of social programs, part 1: Causal models, structural models and econometric policy evaluation. In *Hanbook of Econometrics*, volume 6B. Elsevier B.V.

Huang, Y. and Valtorta, M. (2006). Pearl's calculus of interventions is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI2006)*.

Imbens, G. W. (2014). Instrumental variables: An econometrician's perspective. *Statistical Science*, 29(3):323–358.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

References VI

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Knox, D., Lowe, W., and Mummolo, J. (2019). The bias is built in: How administrative records mask racially biased policing.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *American Economic Review, Papers and Proceedings*, 80:319–323.

Matzkin, R. L. (2007). Nonparametric identification. In *Handbook of Econometrics*, volume 6B.

Matzkin, R. L. (2013). Nonparametric identification in structural economic models. *Annual Review of Economics*, 5:457–486.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–709.

References VII

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

Pearl, J. and Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*.

Pearl, J., Glymour, M., and Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. John Wiley & Sons Ltd, West Sussex, United Kingdom.

Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books, New York.

Rosenzweig, M. R. and Wolpin, K. I. (2000). Natural "Natural Experiments" in Economics. *Journal of Economic Literature*, 38:827–874.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66:688–701.

References VIII

Shpitser, I. and Pearl, J. (2006). Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models. In *Twenty-First National Conference on Artificial Intelligence*.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. The MIT Press, Cambride, MA, 2nd edition.

Strotz, R. H. and Wold, H. O. A. (1960). Recursive vs. nonrecursive systems: An attempt at synthesis (part i of a triptych on causal chain systems). *Econometrica*, 28(2):417–427.

Textor, J., Hardt, J., and Knüppel, S. (2011). DAGitty: A Graphical Tool for Analyzing Diagrams. *Epidemiology*, 5(22):745.

Tian, J. and Pearl, J. (2002). A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 567–573, Menlo Park, CA. AAAI Press/The MIT Press.

References IX

Wold, H. (1954). Causality and econometrics. *Econometrica*, 22(2):162–177.

Wold, H. O. A. (1960). A generalization of causal chain models (part iii of a triptych on causal chain systems). *Econometrica*, 28(2):443–463.

Woodward, J. (2003). *Making Things Happen*. Oxford Studies in Philosophy of Science. Oxford University Press.