Frontier Topics in Empirical Economics: Week 4
Directed Acyclic Graph

Zibin Huang [1]

[1] College of Business, Shanghai University of Finance and Economics

October 12, 2024

Introduction

- Causal inference is the central topic of applied economics
- We almost solely focus on potential outcome framework in Economics
- This framework is proposed by Donald Rubin (Imbens and Rubin, 2015; Rubin, 1974) and sometimes called "Rubin Causal Model"

Introduction

- Is this the only statistical framework dealing with causal inference issue?
  Of course NOT.
- Graphical Model is another important method (Pearl, 2009)
- This is a method highly related to computer science and AI
- Nobel Prize: AI is the future of all sciences!! LOL

Introduction

- Today we are going to learn this new framework
- How it can be applied to economic research is still a very very open question
- Imbens wrote an interesting and critical paper on it
  Imbens (2020) Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics

Introduction

- Judea Pearl is an Israeli-American computer scientist and philosopher, best known for championing the probabilistic approach to artificial intelligence and the development of Bayesian networks. In 2011, he was awarded with the Turing Award, "for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning".

Introduction

Plan for today

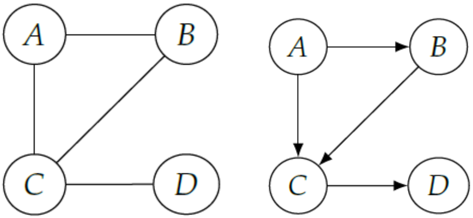
- Introduce the graphical model and the DAG framework
- Discuss the possible usage of DAG for economists: Pros and Cons
- Compare DAG and PO framework: why PO is still more popular
- An example of using DAG: Pinto (2015)
- Conclusion: How can DAG help applied economics research (open question)

DAG Approach: Introduction

- Pearl (2009): Causality, Cambridge University Press 2009
- Neal (2020): Introduction to Causal Inference Online Course
  https://www.bradyneal.com/causal-inference-course#course-textbook
- Pearl and Mackenzie (2018): The Book of Why, Allen Lane 2018

DAG Approach: Graph

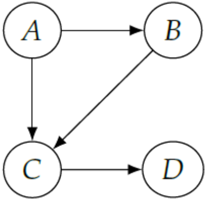
- Graph is a collection of *nodes* and *edges* that connect the nodes.
- Two nodes are called *adjacent* if they are connected by an edge.
- A directed graph's edges go out of a *parent* into a *child*.
- A *path* is any sequence of adjacent nodes, regardless of the direction of the edges. A *directed path* is a path that consists of directed edges that are all directed in the same direction.



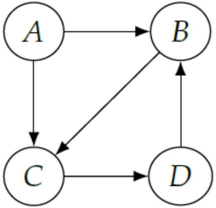
(a) Undirected Graph (b) Directed Graph

DAG Approach: Graph

- If there is a directed path that starts at node X and ends at node Y, then X is an ancestor of Y, and Y is a descendant of X.
- If there is no cycle in a directed graph, the graph is called a *directed acyclic graph* (DAG)



(c) No Cycle



(d) Cycle

DAG Approach: Bayesian Networks

- How to connect graphs to causal inference?
- The first step is to connect graphs to statistical relations: Bayesian Networks
- A Bayesian network is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a DAG

DAG Approach: Bayesian Networks

- For any PDF, a Bayesian factorization can be expressed as:

$$P(x_1, x_2, ..., x_n) = P(x_1) \prod_{i \neq 1} P(x_i | x_{i-1}, ..., x_1) \tag{1}$$

- Example: $P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_2, x_1)$
- This is like a chain
- We can simplify the model if we assume some dependency structure, e.g. $P(x_3|x_2, x_1) = P(x_3|x_2)$ if $x_1 \perp x_3 | x_2$
- When we make more assumptions, we simplify it more

DAG Approach: Bayesian Networks

- Bayesian factorization can be applied to any joint distribution of $(x_1, x_2, ..., x_n)$
- With the set of the dependency assumptions, we are giving the joint distribution a structure
- We can use a graph to represent this assumed dependency structure, system of probabilistic relations
- A one-to-one mapping between graph G and probabilistic relations P

DAG Approach: Bayesian Networks

Assumption (Minimality Assumption)

1. Given its parents in the DAG, a node X is independent of all its non-descendants (Local Markov Assumption);
2. Adjacent nodes in the DAG are dependent (Minimal independence).

Definition (Bayesian Network Factorization)

Given a probability distribution P and a DAG G satistying "Minimality Assumption", P factorizes according to G by

$$P(x_1, x_2, ..., x_n) = P(x_1) \prod_i P(x_i | pa_i)$$

where pa_i is the parents set of i.

DAG Approach: Bayesian Networks

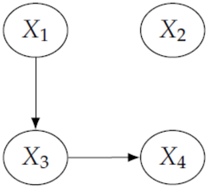
- Local Markov means that the dependence structure is "local" and "Markov"
- Minimal independence means that there is no more independence outside the network showed in the graph
- Bayesian Factorization means that: If P has a causal structure as shown in G
  - x_i only depends on parents pa_i in the graph
  - We can do Bayesian network factorization for P w.r.t. G
- We call "G represents P", "G and P are compatible", "P is Markov relative to G"

DAG Approach: Bayesian Networks

- Let's see a simple example
- Assume that we have four variables x_1, x_2, x_3, x_4
- A full decomposition is:

$$P(x_1, x_2, x_3, x_4) = P(x_1)P(x_2|x_1)P(x_3|x_2, x_1)P(x_4|x_3, x_2, x_1) \tag{2}$$

- What if we have the following DAG showing the relation among x_1, x_2, x_3, x_4?

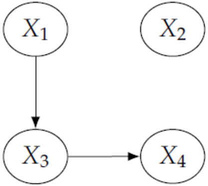


DAG Approach: Bayesian Networks

- We can then have a Bayesian Network Factorization as:

$$P(x_1, x_2, x_3, x_4) = P(x_1)P(x_2)P(x_3|x_1)P(x_4|x_3) \tag{3}$$

- Edges in the graph mean statistical dependencies



DAG Approach: Causal Graphs

- Up until now, we consider only statistical dependencies
- What about those arrows?

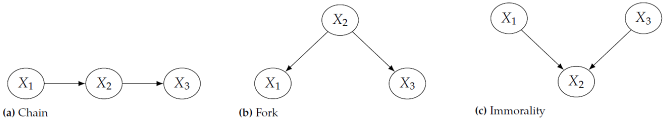
Assumption (Causal Edges Assumption)

In a directed graph, every parent is a direct cause of all its children.

- By adding causal edge assumption, we have this DAG to represent not only statistical dependencies, but causal relations
- Directed paths in DAGs correspond to causation
- A more mathematically rigorous definition is imposed on SEM

DAG Approach: Graphical Building Blocks

- Now we introduce some building blocks of the causal graph



- Flow of association is symmetric: x_1 and x_3 are associated in both chain and fork (but not immorality)
- Flow of causation is asymmetric: x_2 causes x_3 but not vice versa

DAG Approach: Graphical Building Blocks

- By conditioning on variable x_2, we can block the flow of association in chains and forks

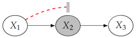


Figure 3.14: Chain with association blocked by conditioning on X_2.

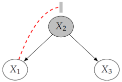


Figure 3.15: Fork with association blocked by conditioning on X_2.

- We can show that with this graph:

$$P(x_1, x_3 | x_2) = P(x_1 | x_2) P(x_3 | x_2) \tag{4}$$

DAG Approach: Graphical Building Blocks

- Things can be different in immorality
- We call X_2, the child of a immorality, as a *collider*

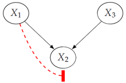


Figure 3.16: Immorality with association blocked by a collider.

- Applying Bayesian factorization:

$$\begin{aligned} P(x_1, x_3) &= \int_{x_2} P(x_1)P(x_3)P(x_2|x_1, x_3) \\ &= P(x_1)P(x_3)\int_{x_2} P(x_2|x_1, x_3) = P(x_1)P(x_3) \end{aligned} \tag{5}$$

- x_1 and x_3 are independent, without the need to conditional on x_2

DAG Approach: Graphical Building Blocks

- What's more, by conditional on x_2, you are creating dependencies!
- Controlling for post-determined variables!
- A simple example: x_1 is good-looking, x_2 is kindness, x_3 is marriage availability
- Conditional on $x_3 = 1$, you will see negative relation between x_1 and x_2!
- Well-known as bad control problem in econometrics

DAG Approach: Graphical Building Blocks

- Homework: Prove that by conditional on x_2, we have x_1 and x_3 to be dependent. That is, $P(x_1, x_3|x_2) \neq P(x_1|x_2) \cdot P(x_3|x_2)$

DAG Approach: Blocked Path and d-separation

Definition (Blocked Path)

A path between X and Y is blocked by a conditioning set Z if either of the following is true:
1. Along the path, there is a chain $\rightarrow W \rightarrow$ or a fork $\leftarrow W \rightarrow$ where $W \in Z$;
2. There is a collider W that both itself and its descendants are not conditioned on in Z;

- Association flows along unblocked paths, does NOT flow along blocked paths!

Definition (d-separation)

Two sets of nodes X and Y are d-separated by a set of nodes Z if all of the paths between nodes in X and nodes in Y are blocked by Z

- d-separation means conditional independence!!
- All association flows between X and Y are blocked by Z

DAG Approach: Blocked Path and d-separation

- Theorem 1.2.4, 1.2.5 in Pearl (2009), Theorem 3.1 in Neal (2020)

Theorem (d-separation and statistical independence)

If X and Y are d-separated in a DAG G conditional on Z, then X and Y are independent conditioned on Z in every distribution compatible with G:

$$X \perp_G Y | Z \Rightarrow X \perp_P Y | Z, \forall P \text{ compatible with } G$$

Conversely, if X and Y are independent conditional on Z in all P compatible with G, then X and Y are d-separated in G conditional on Z:

$$\forall P \text{ compatible with } G, X \perp_P Y | Z \Rightarrow X \perp_G Y | Z$$

- This theorem is a bridge, telling you how to express statistical independence in a graph!!

DAG Approach: Blocked Path and d-separation

- Associations flow along unblocked paths
- Causations flow along directed unblocked paths
- Identification: how to net causation out of associations?
- By ensuring that there is no non-causal association between X and Y!
- If X and Y are d-separated in the augmented graph where we remove outgoing edges from X
- In another word, all non-causal paths are blocked

DAG Approach: do-operator

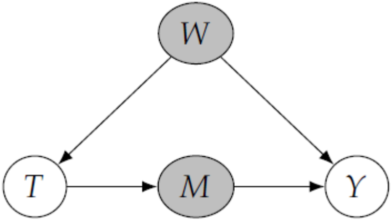
- We define operator "$do(T = t)$" as an intervention to give the whole population treatment t
- We denote it in terms of potential outcomes as:

$$P(y|do(t)) = P(Y = y|do(T = t)) = P(Y(t) = y) \tag{6}$$

- $P(y|do(t))$ means the distribution of the potential outcome $Y(t)$
- Identification of a causal model: If we can reduce an expression Q with *do* to one without *do*, then Q is identifiable.
- Just like we can reduce an expression with potential outcomes to an expression without them in PO framework

DAG Approach: Backdoor Adjustment

- Non-directed unblocked paths from T to Y are "backdoor paths"
- If some variable set W blocks all backdoor paths from T to Y and does not contain any descendants of T, we say W satisfies "the backdoor criterion"



DAG Approach: Backdoor Adjustment

- Backdoor Adjustment Theorem

Theorem (Backdoor Adjustment)

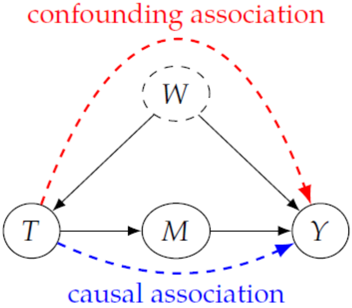
If W satisfies the backdoor criterion, we can identify the causal effect of T on Y by:

$$P(y|do(t)) = \int_w P(y|t,w)P(w)$$

- W is what we usually call "control variables"
- The backdoor criterion is similar to the "selection on observables" assumption

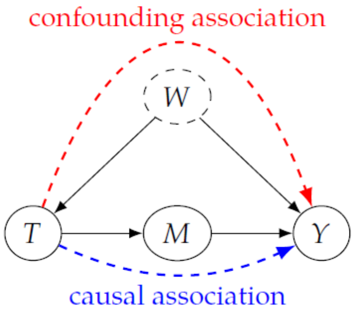
DAG Approach: Frontdoor Adjustment

- Another very interesting identification method in DAG is frontdoor adjustment
- This is totally new to economists
- Assume that we have the following DAG



DAG Approach: Frontdoor Adjustment

- If W is unobserved, we can identify effect of T on Y in three steps
  - 1. Identify effect of T on M
  - 2. Identify effect of M on Y (control for T)
  - 3. Combine step 1 and 2



DAG Approach: Frontdoor Adjustment

Definition (Frontdoor Criterion)

A set of variables M satisfies the frontdoor criterion relative to T and Y if:
1. M completely mediates the causal effect of T on Y;
2. There is no unblocked backdoor path from T to M;
3. All backdoor paths from M to Y are blocked by T.

Theorem (Frontdoor Adjustment)

If T, M, Y satisfy the frontdoor criterion, then we have

$$P(y|do(t)) = \sum_m P(m|t) \sum_{t'} P(y|m, t')P(t')$$

- We can identify the original treatment effect if we have a complete mediator

DAG Approach: Non-parametric Identification

- But backdoor and frontdoor criteria are just sufficient conditions for causal identification
- They are not necessary
- Can we find a set of necessary conditions?
- If there is such a set, we can decide whether a causal effect is identifiable or not in any causal system
- Here it comes: do-calculus

DAG Approach: Non-parametric Identification

- Denote $G_{\overline{X}}$ as take graph G and then remove all incoming edges to X
- Denote $G_{\underline{X}}$ as take graph G and then remove all outgoing edges to X

Theorem (Rules of do-calculus)

(1) Rule 1: $P(y|do(t), z, w) = P(y|do(t), w), \text{if } Y \perp_{G_{\overline{T}}} Z|T, W$
(2) Rule 2: $P(y|do(t), do(z), w) = P(y|do(t), z, w), \text{if } Y \perp_{G_{\overline{T}\underline{Z}}} Z|T, W$
(3) Rule 3: $P(y|do(t), do(z), w) = P(y|do(t), w), \text{if } Y \perp_{G_{\overline{TZ(W)}}} Z|T, W$

DAG Approach: Non-parametric Identification

Theorem (Identification of Causal Effect)

A causal effect Q is identifiable in a model characterized by a graph G if there exists a finite sequence of transformations, each conforming to one of the inference rules 1, 2, or 3, that reduce Q into a standard ("do"-free) probability expression involving observed quantities.

- do-calculus is complete. You can use these three rules to identify all identifiable causal estimands.
- Caution: we consider only non-parametric identification here!

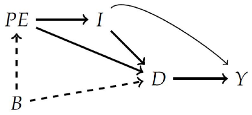
DAG Approach: Non-parametric Identification

What are the intuitions of the three rules?

- Rule 1 (deletion of var): $P(y|do(t), z, w) = P(y|do(t), w), \text{if } Y \perp_{G_{\overline{T}}} Z|T, W$
  - Erase $do(t)$, this is just an extension of d-separation under the Markov assumption
  - $P(y|z, w) = P(y|w), \text{if } Y \perp_{G} Z|W$
- Rule 2 (do-var exchange): $P(y|do(t), do(z), w) = P(y|do(t), z, w), \text{if } Y \perp_{G_{\overline{T}\underline{Z}}} Z|W$
  - Erase $do(t)$, this is just an extension of the backdoor adjustment
  - $P(y|do(z), w) = P(y|z, w), \text{if } Y \perp_{G_{\underline{Z}}} Z|T, W$
  - W can block all non-causal links between Z and Y
- Rule 3 (deletion of action): $P(y|do(t), do(z), w) = P(y|do(t), w), \text{if } Y \perp_{G_{\overline{TZ(W)}}} Z|T, W$

DAG Approach: An Example

- An example: College (D) return on wages (Y)
- Which variable do we need to control for?



- PE: parental education
- I: family income
- B: unobserved background factors, such as genetics, family environment, mental ability, etc.

DAG in Economics

- In general, Imbens believes that " *These frameworks are complementary, with different strengths that make them particularly appropriate for different questions.*"
- Two major advantages of DAG framework:
  - DAG illustrates causal assumptions in an explicit and clear way
    Especially if you are interested in mediation/surrogates.
  - Machinery developed in DAG (do-calculus) allows researchers to investigate causal queries in a systematic way
    Especially for complex models with large number of variables.

DAG in Economics

- What are pros and cons of using DAG in Economics?

DAG in Economics: Clarity

Pro 1: Clarity

- Unconfoundedness

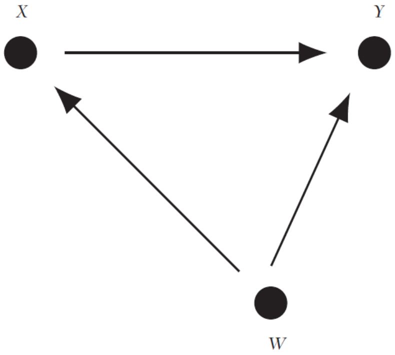


Figure 2. Unconfoundedness

DAG in Economics: Clarity

- IV strategy

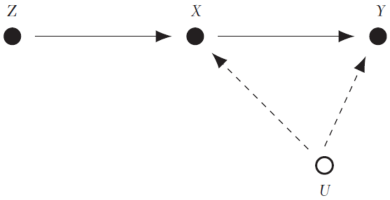


Figure 3. Instrumental Variables

DAG in Economics: Complicated Model

Pro 2: Tool to analyze complicated causal model

- An example of a complicated model

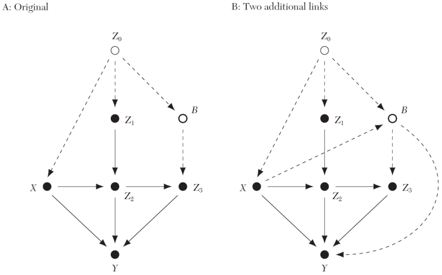


Figure 4. Two Examples of Complex DAGs

DAG in Economics: Complicated Model

- Structural Equation Modeling
- Given a DAG, we write down a linear equation system

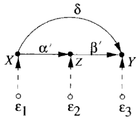


Figure 5.13 Diagram representing model M' of (5.12)–(5.14).

$$x = \varepsilon_1, \tag{5.12}$$

$$z = \alpha' x + \varepsilon_2, \tag{5.13}$$

$$y = \beta' z + \delta x + \varepsilon_3. \tag{5.14}$$

DAG in Economics: Complicated Model

- Imbens' concern: do we really need such huge model and SEM in econ?
- He argues that economists don't like SEM without economic meaning
- Structural modeling in econ uses economic theory more deeply than DAGs can capture
- DAGs cannot easily show shape restrictions (monotonicity of variables etc)

DAG in Economics: Complicated Model

- Structural in econ is different from Structural in some other fields
- We want to regularize data by theory, and care about primitive parameters
- DAGs can deal with SEM, but not structural models in econ
- Personally, I agree with this: how can you illustrate a dynamic discrete choice model using DAGs?

DAG in Economics: Frontdoor Criterion

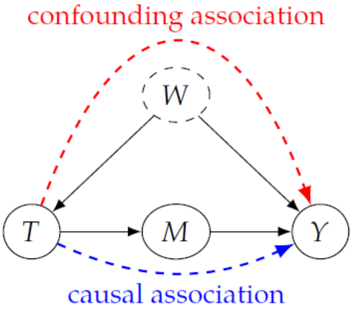
Pro 3: Frontdoor Criterion

- Frontdoor adjustment can be an interesting identification strategy for economists
- It relies on the existence of a complete mediator
- How to apply this method to economics is still an open question
- Too hard to find such a DAG in real life (a complete mediator)

DAG in Economics: Frontdoor Criterion

Complete mediator is rare to find

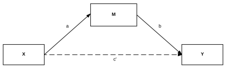
- What if T affects Y in other ways?
- What if some unobserved U affects both Y and M?
- What if W can also affect M?



DAG in Economics: Mediation

Pro 4: Systematic analysis of mediation effect

- DAG may shed lights on identifying mediation effect
- The question remains: we need to impose strong causal structure assumption
- Still much better than "mediation effect test" (I really hate it...)
- Mediation effect test forces you to admit a very simple causal structure just to implement an on-the-shelf test
- This is a typical behavior of regression monkey
- DAG allows you to "have a causal structure" based on your economic context



DAG in Economics

Con 1: DAG needs ex ante causal structure

- DAG develops machinery for identification given two inputs
  - Knowledge of joint distribution of all observed variables
  - Structure of the causal model
- Little is said about why we have this model structure (before model) and inference (after model)
- These can be unfriendly to economists

DAG in Economics: DAG and Traditional Methods in Economics

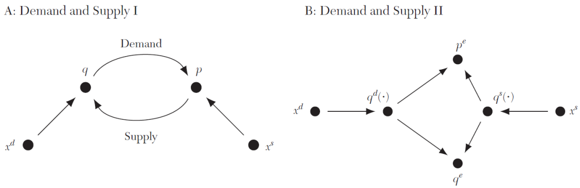
Con 2: DAG does not fit into IV very well

- DAG is unable to clearly express some parts of the IV method
  - Shape restrictions like monotonicity assumption is not naturally captured in DAG
  - LATE theorem is not easily derived in a DAG approach
- PO can naturally express IV
- Or in general, the inability to fit IV shows two weaknesses of DAG
  - DAG is not convenient in expressing economic-related structural assumptions
  - PO can deal with heterogeneity issue (LATE) better
- DAG does not add much insight in RDD

DAG in Economics: Simultaneity

Con 3: DAG cannot capture some equilibrium concept in Economics

- DAG by definition is not cyclical
- They naturally cannot capture equilibrium behavior
- They cannot express simultaneity issue
- Here is an attempt from Imbens, though not so successful



DAG in Economics

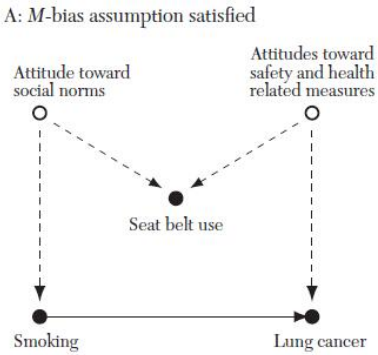
- Why have we not seen too much usage of DAG in applied econ?
- 1. PO framework has several features fitting applied econ better
  - Some common assumptions (monotonicity) are easily captured in PO but not DAG
  - PO connects easily to traditional econ approaches
  - Applied Econ focuses on models with relatively few variables
  - PO accounts for heterogeneity better
  - PO connects closely to the implementation of the method and its inference
- 2. No substantive empirical examples are provided
  - We do not see concrete examples of implementing DAG in econ questions
  - Most of examples from Pearl are "toy models"

Personal View: Clarity in Illustration

- But personally I think DAG does much better in showing "bad control" problem
- Simpson Paradox is illustrated very clearly in DAG
- In PO, we usually select control variables pretty arbitrarily
- Angrist only gives a rule-of-thumb: only control variables happening before treatment

Personal View: Clarity in Illustration

- Although this is useful, it is actually wrong: M-bias



Personal View: Clarity in Illustration

- We can see here the essence is that we should not control for colliders
- Colliders do not necessarily happen after treatment
- M-bias is the case when collider is a pre-determined variable

Personal View: Clarity in Illustration

- DAG gives us a powerful tool to select controls, given our assumptions of causal structure
- It forces us to firmly and explicitly consider our causal structure and show them in a transparent way

An Application in Economics: Pinto (2015)

- Pinto (2015) Selection Bias in a Controlled Experiment: The Case of Moving to Opportunity
- This is the only applied ECON paper I've ever read using DAG and Bayesian Networks
- Sadly, in his latest version, Pinto deletes all DAG stuffs...
- There are more than DAG in this paper $\rightarrow$ Choice model and IV
- Pinto shows an interesting method to use WARP to achieve the identification
- We will discuss it later

Final Conclusion

- DAG approach fully deserves the attention of all economists
- It has advantages in clearly illustrating causal structures, guiding the selection of controls, and dealing with models with large number of variables
- However, it still has many weaknesses compared with PO in applying to economics
- Especially, it lacks of concrete examples in applying this method in economics
- It is still an open question to all economists! Chances here!

References

Imbens, Guido W. 2020. "Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics." *Journal of Economic Literature* 58 (4):1129–1179.

Imbens, Guido W and Donald B Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Neal, Brady. 2020. "Introduction to Causal Inference." *Course Lecture Notes (draft)* .

Pearl, Judea. 2009. *Causality*. Cambridge university press.

Pearl, Judea and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. Basic books.

Pinto, Rodrigo. 2015. "Selection Bias in a Controlled Experiment: The Case of Moving to Opportunity." *Unpublished Ph. D. Thesis, University of Chicago, Department of Economics* .

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5):688.