# Causal Data Science with Directed Acyclic Graphs (DAGs)

## Section 2: Structural Causal Models, Interventions, and Graphs

Dr. Paul Hünermund

Online Course at Udemy.com

# Course Outline

# Directed Acyclic Graphs

- Graph theory provides a useful mathematical language to think about causality
- A graph consists of a set of *vertices* (or nodes) $V$ and a set of *edges* (or links) $E$. Vertices represent variables in the model and edges the connections between them
- Edges can either be *undirected* or *directed* (denoted by arrowheads indicating the direction)

# Directed Acyclic Graphs (II)

- Causal relationships are generally seen as asymmetric
  - If "A causes B" is true, then "B causes A" must be false
  - Therefore we'll work with *directed graphs* most of the times
- We'll sometimes use terminology of kinship
  - X is parent of Y
  - Y is child of X
  - X is ancestor of Z
  - Z is descendant of X
- A *path* is a sequence of edges connecting two vertices
  - $A \leftarrow B \leftarrow C \rightarrow D$ is a path from $A$ to $D$
  - A path can go either along or against arrowheads
  - A path along the arrows is called *directed*: $A \rightarrow B \rightarrow C$

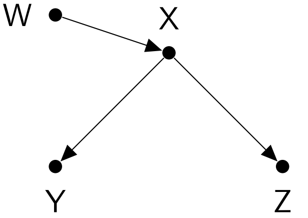
# Directed Acyclic Graphs (III)

- A directed path from a node to itself is called “directed cycle” or “feedback loop”
  - $A \rightarrow B \rightarrow C \rightarrow A$
  - A graph containing such feedback loops is called *cyclic*
  - A graph with no feedback loops is called *acyclic*
- In what follows we’ll work with *directed acyclic graphs* (DAG)
  - We exclude variables exerting a causal influence on themselves
  - DAGs are akin to what econometricians call a recursive model
  - *“We can argue that it is only fully recursive models that can be given a causal or structural interpretation”* (Maddala, 1986, p. 111)

Recursive vs. Interdependent Systems

# Structural Causal Models

- A DAG represents an underlying *structural causal models*

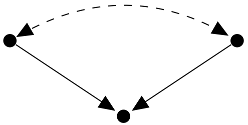



$$W = f_1(\varepsilon_1)$$
$$X = f_2(W, \varepsilon_2)$$
$$Y = f_3(X, \varepsilon_3)$$
$$Z = f_4(X, \varepsilon_4)$$

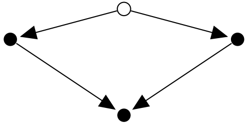
- $\varepsilon_i$'s represent unobserved disturbances (error terms)
- The $f_i$'s can be arbitrary functions. The framework is thus fully *non-parametric*
- Connection to structural equation modeling (SEM, which is mostly restricted to linear equations such as $Y = \alpha_3 + \beta_3 X + \varepsilon_3$ though)

# Structural Causal Models (II)

- Structural equations differ from algebraic equations because they change meaning under solution-preserving operations. “$\leftarrow$” instead of “$=$” would therefore be more adequate
  - “”' $X = aW$” is equivalent to “$W = \frac{1}{a}X$”
  - But “$X \leftarrow W$” and “$W \leftarrow X$” have very different meaning
- A model in which the $\varepsilon_i$'s are jointly independent is called *Markovian* (Pearl, 2009, p. 30)
- Sometimes we denote dependencies due to omitted common causes by dashed bidirected edges (the resulting model is called *semi-Markovian*).



is short for



with hollow circles denoting unobserved variables

# Conditional (In-)dependence in Graphs

- Graphs are useful because they encode conditional independence relationships irrespective of the specific functional relationships between variables of the structural causal model
- We can thus infer conditional independence relationships simply from the topology of the graph (no need for tedious algebra)
- Two variables that are directly connected by an edge in the graph are always stochastically dependent
- Chains: $X \rightarrow Z \rightarrow Y$
  - In a chain, $X$ is independent $Y$, conditional on $Z$ ($X \perp\!\!\!\perp Y|Z$)
  - This only holds if errors are independent
- Forks: $X \leftarrow Z \rightarrow Y$
  - In a fork, $X$ and $Y$ are dependent, but become indpendent, conditional on $Z$
  - $X \not\!\perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y|Z$

# Colliders

- Colliders: $X \rightarrow Z \leftarrow Y$
  - If $Z$ is a collider, $X$ and $Y$ are unconditionally independent, but become dependent conditional on $Z$
  - $X \perp\!\!\!\perp Y$ and $X \not\perp\!\!\!\perp Y|Z$
- Example: "Megan Fox voted worst - but sexiest - actress of 2009" (CNN)
  - What if talent for acting and physical attractiveness are uncorrelated but both contribute equally to your chances of landing a job in Hollywood?
  - Then, $Talent \rightarrow Job \leftarrow Looks$
  - So, in a group of actors you will find many either good looking or talented people, but not likely both
  - I.e., in a sample of actors, talent and looks are negatively correlated, although both are independently distributed in the population

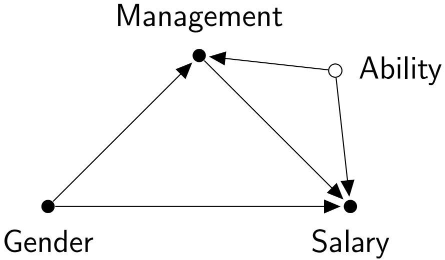
# Colliders (II)

- Example R code

```
library(tidyverse)

talent <- rnorm(1000)
looks <- rnorm(1000)
df <- data.frame(talent, looks)
df %>% summarize(correlation = cor(talent, looks))

x <- talent + looks
job <- 1*(x > quantile(x, c(.75)))
df <- cbind(df, job)

df %>% filter(job == 1) %>% summarize(correlation = cor(
    talent, looks))
```

# Colliders (III)

- Conditioning on a collider is closely related to the problem of sample selection, which we'll see later on
- It was also the problem in our gender pay gap example in the beginning




- Conditioning on *Management* creates dependency between gender and unobserved ability

# $D$-separation

- So far we only looked at graphs containing three variables. Can we somehow generalize these criteria?

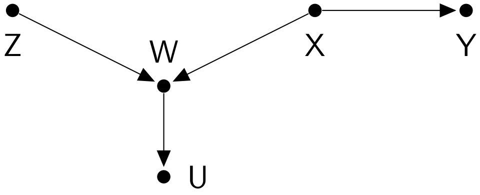
**Definition: $d$-separation (Pearl et al., 2016, p. 46)**

A path $p$ is blocked by a set of nodes $Z$ if and only if

1. $p$ contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node $B$ is in $Z$ (i.e., $B$ is conditioned on), or
2. $p$ contains a collider $A \rightarrow B \leftarrow C$ such that the collision node $B$ is not in $Z$, and no descendant of $B$ is in $Z$

- Two nodes $X$ and $Y$ are $d$-separated if every path between them is blocked
- If two nodes are $d$-separated, the variables they represent are independent

# Example: *d*-separation




- $Z$ and $Y$ are *d*-separated conditional on $\emptyset$ (empty set)
- $Z$ and $Y$ are *d*-connected conditional on $\{W\}$
- $Z$ and $Y$ are *d*-connected conditional on $\{U\}$
- $Z$ and $Y$ are *d*-separated conditional on $\{W, X\}$
- $U$ and $Y$ are *d*-separated conditional on $\{X\}$, conditional on $\{W\}$, or conditional on $\{W, X\}$

# A Working Definition of Causal Inference

*Causal inference is concerned with a very specific kind of prediction problem: predicting the results of an action, manipulation, or intervention.*

– based on Woodward (2003): "Making Things Happen"

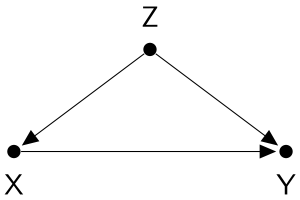
# Interventions

- Interventions in structural causal models are defined by the *do-operator*
- Notation: $P(Y|do(X=x))$ stands for "the probability distribution of $Y$ if we fix (set) $X$ to the specific value $x$"
- Many (if not most) questions we try to answer with data involve some form of intervention, manipulation, or action
    - $P(\text{Firm Growth} \mid do(\text{Public R\&D Grant}))$
    - $P(\text{Wages} \mid do(\text{Schooling}))$
    - $P(\text{Firm Performance} \mid do(\text{M\&A}))$
    - $P(\text{Click-through Rate} \mid do(\text{Advertising}))$
    - $P(\text{Air Quality} \mid do(\text{Banning Diesel Cars}))$

# Interventions (II)

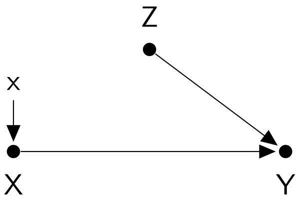
- In graphical models, intervening on a variable $X$ is similar to a kind of surgery in which we remove all edges into that variable

Pre-intervention




$$Z = f_z(\varepsilon_z)$$
$$X = f_x(Z, \varepsilon_x)$$
$$Y = f_y(X, Z, \varepsilon_y)$$

Post-intervention




$$Z = f_z(\varepsilon_z)$$
$$X = x$$
$$Y = f_y(X, Z, \varepsilon_y)$$

# Interventions (III)

## Definition: Causal Effect (Pearl, 2009, p. 70)

Given two disjoint sets of variables, $X$ and $Y$, the causal effect of $X$ on $Y$, denoted either as $P(y|\hat{x})$ or $P(y|do(x))$ is a function from $X$ to the space of probability distributions on $Y$. For each realization $x$ of $X$, $P(y|\hat{x})$ gives the probability of $Y = y$ induced by deleting from the [structural causal] model [...] all equations corresponding to variables in $X$ and substituting $X = x$ in the remaining equations.

- Sometimes, causal effects of interventions are defined as

$$E(Y = y|do(X = x'')) - E(Y = y|do(X = x'))$$

which can always be computed from the general function $P(y|do(x))$

# Interventions (IV)

- Carrying out the intervention ourselves, in a randomized control trial, is not always feasible (too expensive, impractical, or unethical)
- How can we then identify the effect of interventions purely from observational data?
  - We want to know $P(y|do(x))$ but all we have is data $P(x, y, z)$
  - And we know that $P(y|do(x)) \neq P(y|x)$ ("correlation is not causation")
  - No fancy machine learning algorithm will ever solve this problem
- We need to find a way to transform $P(y|do(x))$ into an expression that only contains, observed, "do-free" quantities (topic of section 4)
- What if we only have data that is measured with selection bias or that stems from a different population (topic of sections 5 and 6)?

# References I

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

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

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

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

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

Maddala, G. S. (1986). *Limited-Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs.

# References II

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

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

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

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.

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.

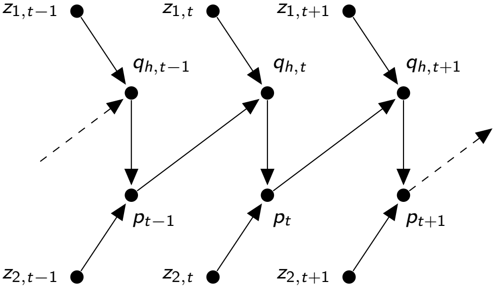
Woodward, J. (2003). *Making Things Happen*. Oxford Studies in Philosophy of Science. Oxford University Press.

# 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

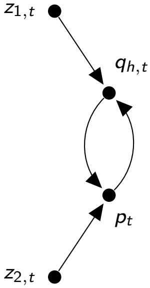
# Recursive Versus Interdependent Systems

- Strotz and Wold (1960) discuss the example of the cobweb model:




$p_{t-1} = p_t$

$\Rightarrow$




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

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

# 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 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)