SCORE Methods Workshop

Causal Inference: Day 1

Benjamin Grant Purzycki[a]
Joseph Bulbulia[b]

Aarhus University[a]
Victoria University[b]

2025

- Theory of causality and causal inference

- Theory of causality and causal inference
  - Counterfactuals and potential outcomes
  - EEESI workflow

Introduction

Outline

- Theory of causality and causal inference
  - Counterfactuals and potential outcomes
  - EEESI workflow
- Causal directed acyclic graphs (DAGs)

Introduction

Outline

- Theory of causality and causal inference
  - Counterfactuals and potential outcomes
  - EEESI workflow
- Causal directed acyclic graphs (DAGs)
  - Introduction to DAGs
  - DAGITTY workshop

- Theory of causality and causal inference
  - Counterfactuals and potential outcomes
  - EEESI workflow
- Causal directed acyclic graphs (DAGs)
  - Introduction to DAGs
  - DAGITTY workshop
- Simulating causal systems and estimating causal effects

Introduction

Outline

- Theory of causality and causal inference
    - Counterfactuals and potential outcomes
    - EEESI workflow
- Causal directed acyclic graphs (DAGs)
    - Introduction to DAGs
    - DAGITTY workshop
- Simulating causal systems and estimating causal effects
    - Conditional vs. marginal effects
    - Cross-sectional and longitudinal
    - G-Methods
    - Sensitivity checks

Theory of causality and causal inference

Causation is inferred

- Intervention, counterfactuals, possible worlds, potential outcomes

control	treatment
eyes are fine	bleeding eyeballs

Theory of causality and causal inference

Fundamental problem of causal inference

"...one can never observe the potential outcome under the treatment state for those observed in the control state, and one can never observe the potential outcome under the control state for those observed in the treatment state. This impossibility implies that *one can never calculate individual-level causal effects*"
[Morgan and Winship, 2015]

Group	Y^0	Y^1
Control	observable	counterfactual
Treatment	counterfactual	observable

Theory of causality and causal inference

Average Treatment Effect

$$\text{ATE} = \text{E}[Y^{X=1}] - \text{E}[Y^{X=0}]$$

Theory of causality and causal inference

Rethinking how we use statistical tools

- give up reliance of point estimates
- effectively abandon significance tests (e.g., p-values)
- estimate uncertainty
- direct vs. total effects
- "control" for counfounding and colliding effects
- reduce near-endless model specifications
- think and work harder on the front end
- expose how under-specified theories are

Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand

Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand
  - What are you trying to estimate?

Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand
  - What are you trying to estimate?
  - Intervention? Outcome?

Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand
  - What are you trying to estimate?
  - Intervention? Outcome?
  - What is the process that generates it?

Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand
  - What are you trying to estimate?
  - Intervention? Outcome?
  - What is the process that generates it?
- Estimator

Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand
  - What are you trying to estimate?
  - Intervention? Outcome?
  - What is the process that generates it?
- Estimator
  - How will you estimate the Estimand?

Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand
    - What are you trying to estimate?
    - Intervention? Outcome?
    - What is the process that generates it?
- Estimator
    - How will you estimate the Estimand?
    - statistical model

Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand
  - What are you trying to estimate?
  - Intervention? Outcome?
  - What is the process that generates it?
- Estimator
  - How will you estimate the Estimand?
  - statistical model
- Estimate

Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand
  - What are you trying to estimate?
  - Intervention? Outcome?
  - What is the process that generates it?
- Estimator
  - How will you estimate the Estimand?
  - statistical model
- Estimate
  - Apply Estimator

Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand
  - What are you trying to estimate?
  - Intervention? Outcome?
  - What is the process that generates it?
- Estimator
  - How will you estimate the Estimand?
  - statistical model
- Estimate
  - Apply Estimator
  - Interpret results in light of Estimand

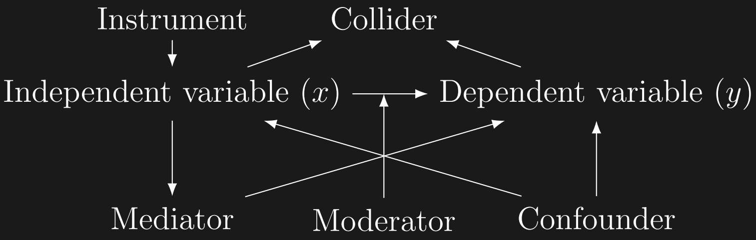
Theory of causality and causal inference

Causal inference is EEESI [Bendixen and Purzycki, 2025]

- Estimand
  - What are you trying to estimate?
  - Intervention? Outcome?
  - What is the process that generates it?
- Estimator
  - How will you estimate the Estimand?
  - statistical model
- Estimate
  - Apply Estimator
  - Interpret results in light of Estimand
- SImulate

Directed Acyclic Graphs

Variable types



Directed Acyclic Graphs

Biases in estimates

Confounder

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ X \arrow[rr] & & Y \end{tikzcd}$$

"fork"

Mediator

$$\begin{tikzcd} & Z \arrow[dr] & \\ X \arrow[ur] \arrow[rr] & & Y \end{tikzcd}$$

"pipe"

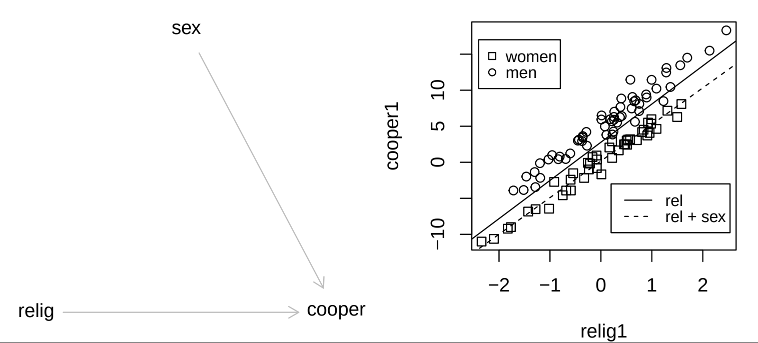
Collider

$$\begin{tikzcd} & Z & \\ X \arrow[ur] \arrow[rr] & & Y \arrow[ul] \end{tikzcd}$$

"collider"

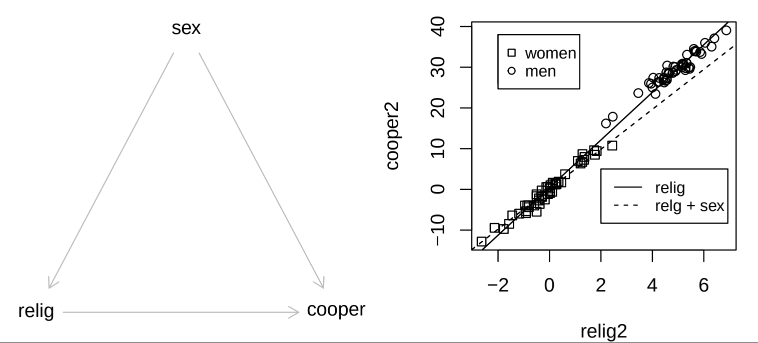
Directed Acyclic Graphs

A note on "controls" (prediction)



Directed Acyclic Graphs

A note on “controls” (causal inference)



Directed Acyclic Graphs

Controls: The Good, The Bad, and the Inconsequential [Cinelli et al., 2024]

$n = 100$
$\beta = 0.5$
$X, Y, Z \sim \text{Normal}(0, 1)$
error terms $\sim \text{Normal}(0, 1)$
iter $= 1,000$

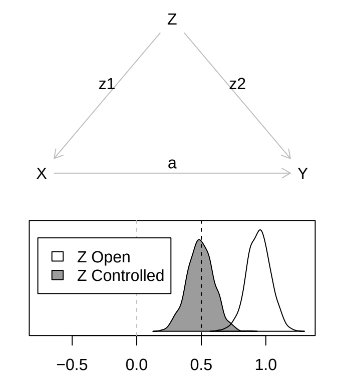
```
Z <- rnorm(n, 0, 1)
X <- beta * Z + err1
Y <- beta * X + beta * Z + err2
lm(Y ~ X)
lm(Y ~ X + Z)
```

Directed Acyclic Graphs

Controls: The Good, The Bad, and the Inconsequential [Cinelli et al., 2024]

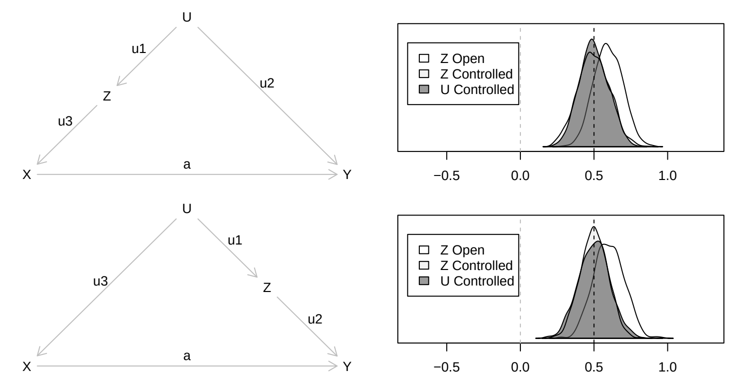
$n = 100$
$\beta = 0.5$
$X, Y, Z \sim \text{Normal}(0, 1)$
error terms $\sim \text{Normal}(0, 1)$
iter $= 1,000$

```
Z <- rnorm(n, 0, 1)
X <- beta * Z + err1
Y <- beta * X + beta * Z + err2
lm(Y ~ X)
lm(Y ~ X + Z)
```



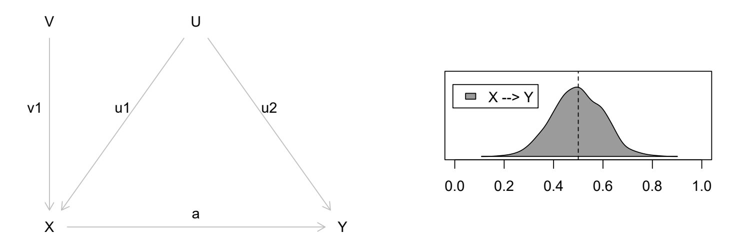
Directed Acyclic Graphs

Controls: The Good (Confounder)



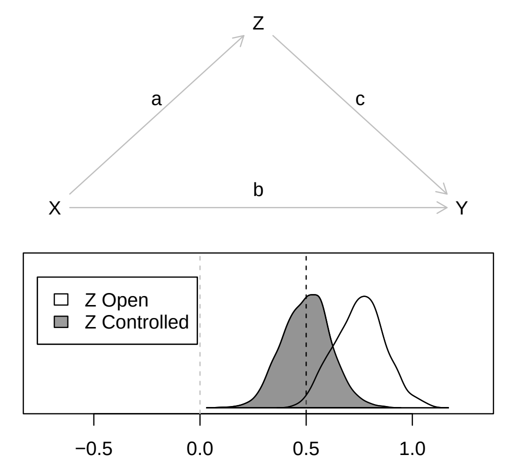
Directed Acyclic Graphs

Controls: The Conditionally Good (Instrumental)



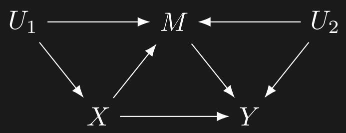
Directed Acyclic Graphs

Controls: The Conditionally Good (Mediator)



Directed Acyclic Graphs

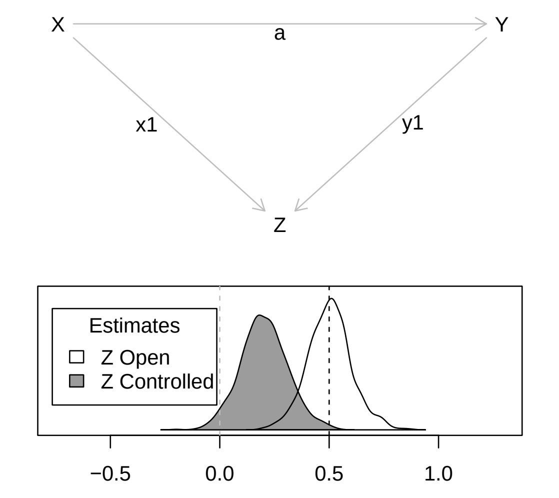
Note on Mediation



- Direct vs. total effects
    - Direct: M and U_2
    - Total: U_1 but *not* M

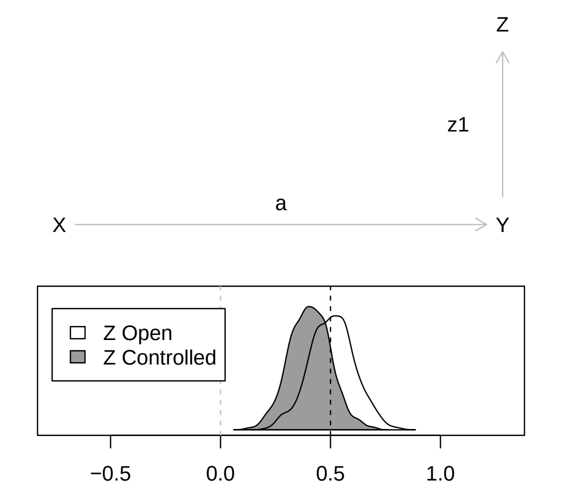
Directed Acyclic Graphs

Controls: The Bad (Collider)



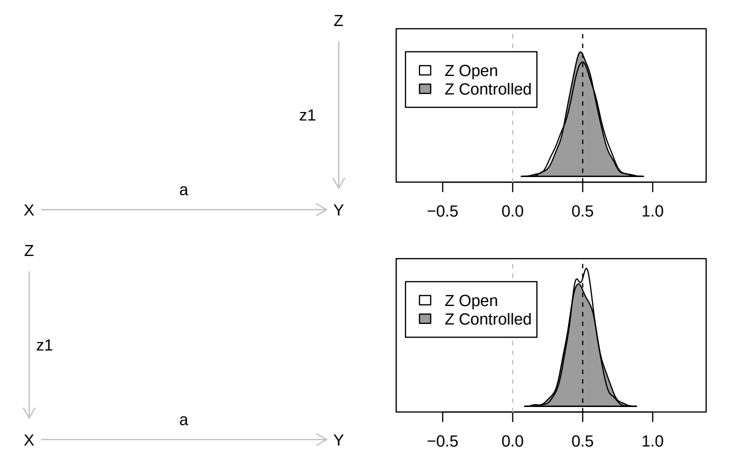
Directed Acyclic Graphs

Controls: The Bad (Post-Treatment)



Directed Acyclic Graphs

Controls: The Inconsequential



Modeling Causal Systems

What to include in your DAG

- theoretically relevant variables

Modeling Causal Systems

What to include in your DAG

- theoretically relevant variables
- precedent relevant variables

Modeling Causal Systems

What to include in your DAG

- theoretically relevant variables
- precedent relevant variables
- meta-analytic variables

Modeling Causal Systems

What to include in your DAG

- theoretically relevant variables
- precedent relevant variables
- meta-analytic variables
- sources of bias (e.g., sampling, confounders, etc.)

Modeling Causal Systems

What to include in your DAG

- theoretically relevant variables
- precedent relevant variables
- meta-analytic variables
- sources of bias (e.g., sampling, confounders, etc.)
- common sense/gumption variables

Modeling Causal Systems

What to include in your DAG

- theoretically relevant variables
- precedent relevant variables
- meta-analytic variables
- sources of bias (e.g., sampling, confounders, etc.)
- common sense/gumption variables
- de-biasing or instrumental variables

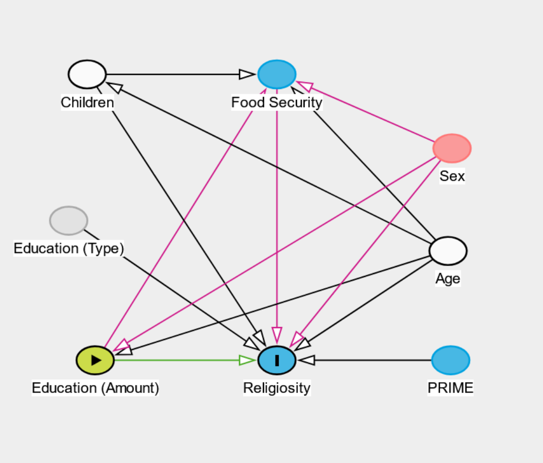
Modeling Causal Systems

What to include in your DAG

- theoretically relevant variables
- precedent relevant variables
- meta-analytic variables
- sources of bias (e.g., sampling, confounders, etc.)
- common sense/gumption variables
- de-biasing or instrumental variables
- many are defined by the structure of your model!

Simulating Causal Systems

DAGITTY workshop

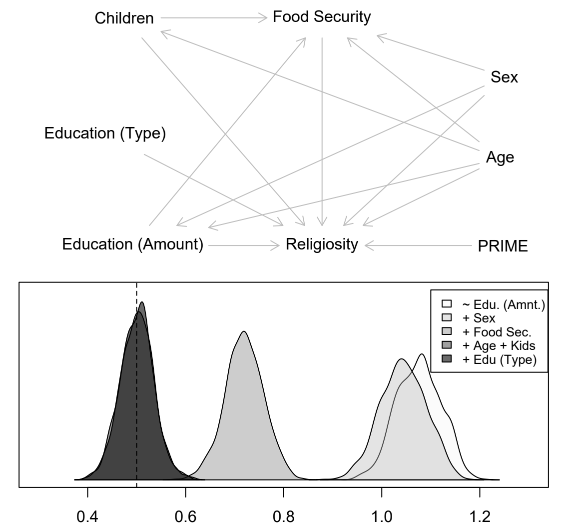


Simulating Causal Systems

R workshop

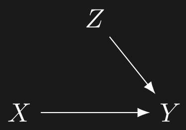
```r
fd <- function(n, beta) {
  e_rel <- rnorm(n, 0, 1) # noise
  e_edu <- rnorm(n, 0, 1)
  e_mat <- rnorm(n, 0, 1)
  e_qua <- rnorm(n, 0, 1)
  e_sex <- rnorm(n, 0, 1)
  e_age <- rnorm(n, 0, 1)
  e_kid <- rnorm(n, 0, 1)
  SEX <- rbinom(n, 1, .5) # sex
  XIA <- rbinom(n, 1, .5) # edu type
  PRI <- rbinom(n, 1, .5) # prime
  AGE <- rnorm(n, 0, 1)
  EDU <- beta * SEX + beta * AGE + e_edu
  MAT <- beta * SEX * beta * EDU + beta * AGE + e_mat
  KID <- beta * MAT + beta * AGE + e_kid # children
  REL <- beta * EDU + beta * XIA + beta * MAT + beta * SEX + beta * AGE +
    beta * PRI + beta * KID + e_rel # religiosity
  df <- data.frame(SEX, XIA, PRI, AGE, EDU, MAT, REL)
  open0 <- coef(lm(REL ~ EDU, dat = df))[2]
  open1 <- coef(lm(REL ~ EDU + SEX, dat = df))[2]
  open2 <- coef(lm(REL ~ EDU + SEX + MAT, dat = df))[2]
  open3 <- coef(lm(REL ~ EDU + SEX + MAT + AGE + KID, dat = df))[2]
  closed <- coef(lm(REL ~ EDU + SEX + MAT + AGE + KID + XIA, dat = df))[2]
  return(c(open0, open1, open2, open3, closed))
}
```

Simulating Causal Systems
R workshop



Directed Acyclic Graphs

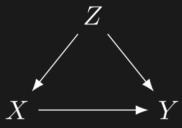
Assumptions of experimental data



- Estimand obtained when:
    - random assignment of X
    - no unmeasured confounders
    - no reverse causation
    - no measurement error
    - $\text{ATE} = \text{E}[Y^{X=1}] - \text{E}[Y^{X=0}]$

Directed Acyclic Graphs

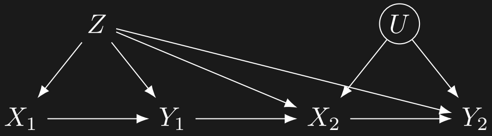
Assumptions of cross-sectional data



- Estimand obtained when:
  - random assignment of X
  - no unmeasured confounders
  - no reverse causation
  - no measurement error
  - $\text{ATE} = \text{E}[Y^{X=1}] - \text{E}[Y^{X=0}]$
- Sensitivity analysis

Directed Acyclic Graphs

Assumptions of longitudinal data



Estimand obtained when:

- random assignment of X
- no unmeasured confounders
- no reverse causation
- no measurement error
- $\text{ATE} = \text{E}[Y^{X=1}] - \text{E}[Y^{X=0}]$

Marginal vs. Conditional Effects

G-Methods

g-computation

1. Predict values where $X = 0$
2. Predict values where $X = 1$
3. Calculate ATE
4. Bootstrap/Go Bayes

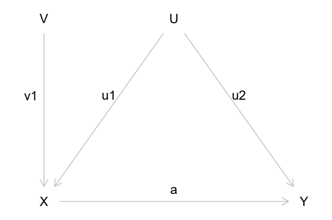
```r
set.seed(777)
n <- 1e4
bZ <- 2
bX <- 2
Z <- rnorm(n, 0, 0.5)
X <- rbinom(n, 1, plogis(0.5 + Z * bZ))
Y <- rnorm(n, 10 + X * bX + Z * bZ)
d <- data.frame(Y, X, Z)

model <- lm(Y ~ X + Z, data = d)
d$EX1 <- predict(model,
                 newdata = transform(d, X = 1))
d$EX0 <- predict(model,
                 newdata = transform(d, X = 0))
mean(d$EX1) - mean(d$EX0)

# bootstrap
iter <- 100

bootstrap.fun <- function(data, indices) {
  d <- data[indices, ] # resample
  model <- lm(Y ~ X + Z, data = d) # fit model
  d$EX1 <- predict(model, newdata = transform(d, X = 1))
  d$EX0 <- predict(model, newdata = transform(d, X = 0))
  ate <- with(d, mean(EX1) - mean(EX0)) # ate
  return(ate)
}

bootstrap.res <- boot(data = d,
                      statistic = bootstrap.fun,
                      R = iter)

(bootstrap.sum <- boot.ci(bootstrap.res, type = c("norm")))
```

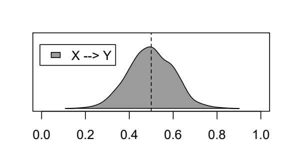
G-Methods

Instrumental variables revisited

Assumptions

- V is independent of U
- V is a cause of X
- V does not cause Y other than through X





$$\frac{\mathrm{E}[Y|V=1]-\mathrm{E}[Y|V=0]}{\mathrm{E}[X|V=1]-\mathrm{E}[Y|V=0]}$$

G-Methods

Inverse probability weighting (IPW)

1. Estimate probability of $X = 1$ as function of confound
2. Invert probs. and stabilize weights
3. Fit outcome model with weights
4. Bootstrap/Go Bayes

Treated: $\frac{\Pr(X_i)}{\Pr(X_i|Z_i)}$

Untreated: $\frac{1-\Pr(X_i)}{1-\Pr(X_i|Z_i)}$

G-Methods

Doubly robust estimation

-

Acknowledgments

- Theiss Bendixen
- Joseph Bulbulia
- Daniel Major-Smith
- Richard McElreath
- Julia Rohrer
- Brooke, Don, and Mike

SCORE Methods Workshop

Causal Inference: Day 2

Joseph Bulbulia[a]
Benjamin Grant Purzycki[b]

Victoria University[a]
Aarhus University[b]

2025

Bibliography I

Bendixen, T. and Purzycki, B. G. (2025).
The Data Analyst's Guide to Cause and Effect: An Introduction to Applied Causal Inference.
Sage, London.

Cinelli, C., Forney, A., and Pearl, J. (2024).
A crash course in good and bad controls.
Sociological Methods & Research, 53(3):1071–1104.

Morgan, S. and Winship, C. (2015).
Counterfactuals and causal inference.
Cambridge University Press, Cambridge.