Precept 10: Matching, mediation, and dynamic treatments

Soc 504: Advanced Social Statistics

Ian Lundberg

Princeton University

April 20, 2017

Outline

1. Propensity scores
2. Coarsened exact matching
3. Mediation
4. Posters
5. Marginal structural models

Outline

1 Propensity scores

2 Coarsened exact matching

3 Mediation

4 Posters

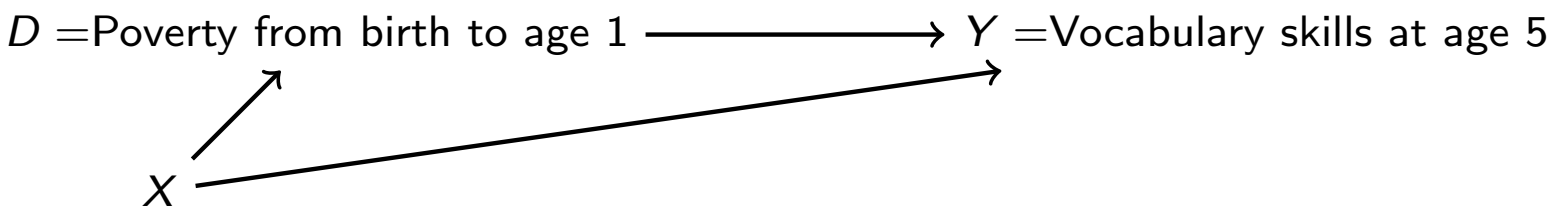
5 Marginal structural models

Running example: Poverty and child development

Question: Does exposure to poverty in childhood reduce vocabulary scores at school entry?

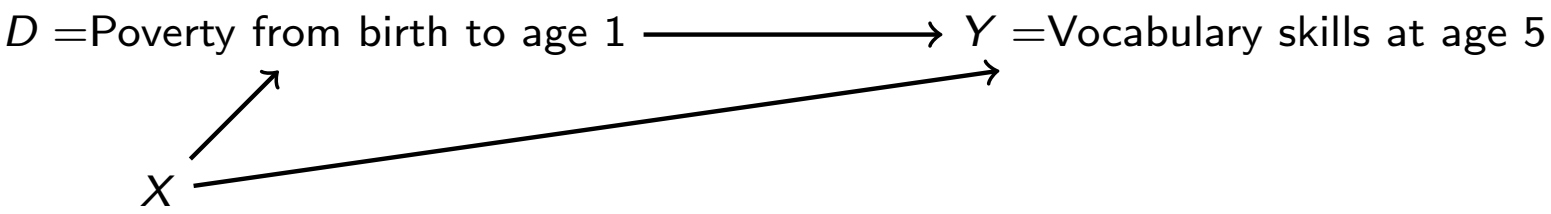
Running example: Poverty and child development

Question: Does exposure to poverty in childhood reduce vocabulary scores at school entry?



Running example: Poverty and child development

Question: Does exposure to poverty in childhood reduce vocabulary scores at school entry?



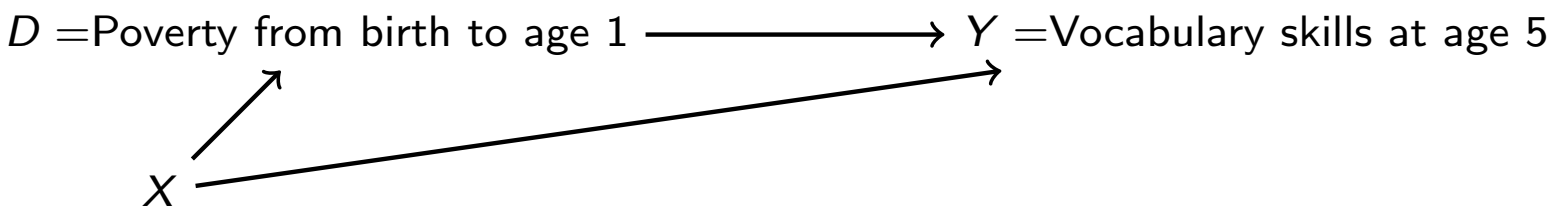
X includes mother's

- Vocabulary skills
- Race/ethnicity
- Relationship to the father at birth

and mother's and father's education.

Running example: Poverty and child development

Question: Does exposure to poverty in childhood reduce vocabulary scores at school entry?



X includes mother's

- Vocabulary skills
- Race/ethnicity
- Relationship to the father at birth

and mother's and father's education.

Identifying assumption:
$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i \mid X_i = x$

Matching: Find units comparable on X with different D

Given the identification assumption above, matching is an estimation strategy.

Matching: Find units comparable on X with different D

Given the identification assumption above, matching is an estimation strategy.

Ideally, we would find treated and control units with exactly the same X.

Matching: Find units comparable on X with different D

Given the identification assumption above, matching is an estimation strategy.

Ideally, we would find treated and control units with exactly the same X.

But X is high-dimensional. No two units match exactly on mother's vocabulary score.

Matching: Find units comparable on X with different D

Given the identification assumption above, matching is an estimation strategy.

Ideally, we would find treated and control units with exactly the same X.

But X is high-dimensional. No two units match exactly on mother's vocabulary score.

Instead, we need a way to get approximate matches.

Propensity Score as a Low-Dimensional Summary

We can summarize X with a one-dimensional summary:

$$p(x) = P[D_i = 1 \mid X_i = x]$$

The Rosenbaum and Rubin theorem states that:

$$\big(Y_i(0), Y_i(1)\big) \perp\!\!\!\perp D_i \mid p(X_i) \implies E\big(Y_i(0), Y_i(1)\big) \perp\!\!\!\perp D_i \mid X_i$$

In expectation, conditioning on $p(X_i)$ is the same in expectation as conditioning on X_i.

WARNING: Only guaranteed to help in expectation (translation: in large samples).

Steps of propensity score modeling

1. Estimate the propensity score
2. Match on or weight by the propensity score
3. Check balance and sample size
4. Repeat (1) - (3) until happy
5. Look at the outcome

Steps of propensity score modeling

1. Estimate the propensity score
2. Match on or weight by the propensity score
3. Check balance and sample size
4. Repeat (1) - (3) until happy
5. Look at the outcome

1. Estimate the propensity score

Stochastic component: $Y_i \sim \text{Bernoulli}(p_i)$
Systematic component: $\text{logit}(p_i) = X_i\beta$

1. Estimate the propensity score

Stochastic component: $Y_i \sim \text{Bernoulli}(p_i)$
Systematic component: $\text{logit}(p_i) = X_i\beta$

```
pscore.fit <- glm(pov2 ~ hv3ppvtstd_m +
                   cm1ethrace + cm1relf + cm1edu + cf1edu,
                  family = binomial(link = "logit"),
                  data = ff1)
```

1. Estimate the propensity score

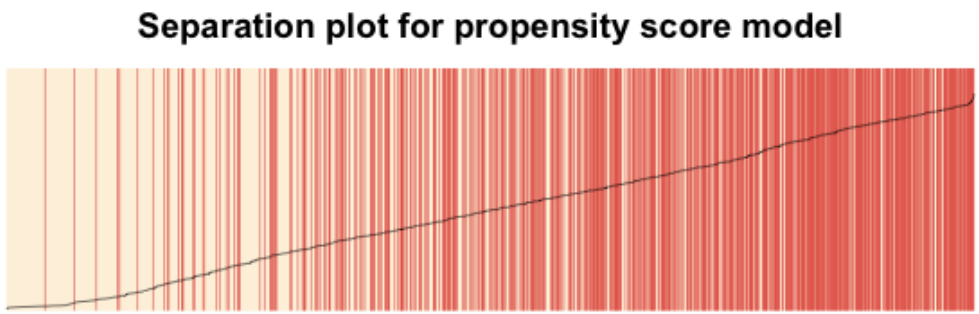
	β
(Intercept)	1.22
hv3ppvtstd_m	-0.02
cm1ethraceBlack	0.65
cm1ethraceHispanic	0.30
cm1ethraceOther	0.83
cm1relfCohabiting	0.85
cm1relfOther	1.36
cm1eduHS	-0.98
cm1eduSome college	-1.40
cm1eduCollege	-2.48
cf1eduHS	-0.24
cf1eduSome college	-0.70
cf1eduCollege	-0.77

What makes a good propensity score model?

We often check logistic regression models in terms of predictive validity.

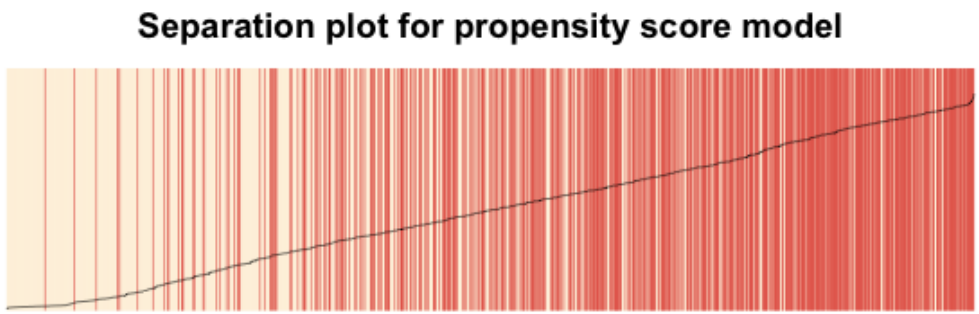
What makes a good propensity score model?

We often check logistic regression models in terms of predictive validity.



What makes a good propensity score model?

We often check logistic regression models in terms of predictive validity.



But in this case, we really care about whether the resulting propensity scores yield good covariate balance.

Steps of propensity score modeling

1. Estimate the propensity score
2. Match on or weight by the propensity score
3. Check balance and sample size
4. Repeat (1) - (3) until happy
5. Look at the outcome

2. Match on or weight by the propensity score

We will use the MatchIt package (http://gking.harvard.edu/matchit)
The code below conducts

- Nearest-neighbor
- 1-1
- propensity score matching
- for the ATT

```
match1 <- matchit(pov2 ~ hv3ppvtstd_m +
                    cm1ethrace + cm1relf + cm1edu + cf1edu,
                  data = ff1,
                  method = "nearest",
                  distance = "logit",
                  discard = "none")
```

Steps of propensity score modeling

1. Estimate the propensity score
2. Match on or weight by the propensity score
3. Check balance and sample size
4. Repeat (1) - (3) until happy
5. Look at the outcome

3. Check balance and sample size

MatchIt's `summary()` function reports lots of things you want for evaluating your matches.

```
Sample sizes:
          Control Treated
All           826     581
Matched       581     581
Unmatched     245       0
Discarded       0       0
```

We have discarded 245 control units who were deemed incomparable.
We have not discarded any treated units.

Thus, our estimates will represent the average treatment effect on the treated.

3. Check balance and sample size

We see that the children born into poor households have mothers who

- have lower vocabular scores (85 vs. 93)
- are disproportionately black (69% vs. 47%)

These differences are reduced in the matched sample.

```
Summary of balance for all data:
                  Means Treated Means Control SD Control
distance                 0.5702        0.3023     0.2308
hv3ppvtstd_m            85.3219       93.4964    12.3876
cm1ethraceWhite          0.0947        0.3123     0.4637
cm1ethraceBlack          0.6936        0.4734     0.4996
[many rows omitted]

Summary of balance for matched data:
                  Means Treated Means Control SD Control
distance                 0.5702        0.4082     0.1933
hv3ppvtstd_m            85.3219       89.9415    10.6911
cm1ethraceWhite          0.0947        0.1480     0.3554
cm1ethraceBlack          0.6936        0.6076     0.4887
[many rows omitted]
```

Can we do better?

Iterate. Back to revise step 2!

2. Match on or weight by the propensity score

We might improve balance by requiring that matched observations be within a caliper of 0.05 from each other on the propensity score.

2. Match on or weight by the propensity score

We might improve balance by requiring that matched observations be within a caliper of 0.05 from each other on the propensity score.

This will

- discard some treated units,
- change the estimand
- harm external validity
- But it may help internal validity.

2. Match on or weight by the propensity score

The code below conducts

- Nearest-neighbor
- 1-1
- propensity score matching
- for the FSATT
- With a caliper of 0.05 (rejecting matches with greater than 0.05 difference in p)

```
set.seed(08544)
match2 <- matchit(pov2 ~ hv3ppvtstd_m +
                    cm1ethrace + cm1relf + cm1edu + cf1edu,
                  data = ff1,
                  method = "nearest",
                  distance = "logit",
                  discard = "none",
                  caliper = .05)
```

3. Check balance and sample size

Our caliper restriction improved balance!

No caliper:

```
Summary of balance for matched data:
                  Means Treated Means Control SD Control
distance                 0.5702        0.4082     0.1933
hv3ppvtstd_m            85.3219       89.9415    10.6911
cm1ethraceWhite          0.0947        0.1480     0.3554
cm1ethraceBlack          0.6936        0.6076     0.4887
[many rows omitted]
```

With caliper:

```
Summary of balance for matched data:
                  Means Treated Means Control SD Control
distance                 0.4914        0.4851     0.1874
hv3ppvtstd_m            87.1134       87.8119    10.6153
cm1ethraceWhite          0.1134        0.1108     0.3143
cm1ethraceBlack          0.6933        0.6598     0.4744
```

3. Check balance and sample size

But the caliper also reduced the sample size and dropped treated units.
Now we have the Feasible Sample Average Treatment Effect on the Treated (FSATT)

```
Sample sizes:
          Control Treated
All           826     581
Matched       388     388
Unmatched     438     193
Discarded       0       0
```

But maybe some of those 438 controls could be useful? We're throwing away a lot of data.

Can we do better?

Iterate. Back to revise step 2!

2. Match on or weight by the propensity score

We might improve efficiency (get lower variance estimates) if we matched 2 controls to each treated unit whenever possible.

2. Match on or weight by the propensity score

We might improve efficiency (get lower variance estimates) if we matched 2 controls to each treated unit whenever possible.

The code below conducts

- Nearest-neighbor
- 2-1
- propensity score matching
- for the FSATT
- With a caliper of 0.05 (rejecting matches with greater than 0.05 difference in p)

```
set.seed(08544)
match3 <- matchit(pov2 ~ hv3ppvtstd_m +
                    cm1ethrace + cm1relf + cm1edu + cf1edu,
                  data = ff1,
                  method = "nearest",
                  distance = "logit",
                  discard = "none",
                  caliper = .05,
                  ratio = 2)
```

3. Check balance and sample size

Allowing 2-1 matching harmed balance.

1-1:

```
Summary of balance for matched data:
                  Means Treated Means Control SD Control
distance                 0.4914        0.4851     0.1874
hv3ppvtstd_m            87.1134       87.8119    10.6153
cm1ethraceWhite          0.1134        0.1108     0.3143
cm1ethraceBlack          0.6933        0.6598     0.4744
```

2-1:

```
Summary of balance for matched data:
                  Means Treated Means Control SD Control
distance                 0.4914        0.4843     0.1879
hv3ppvtstd_m            87.1134       87.7062    10.7413
cm1ethraceWhite          0.1134        0.1263     0.3325
cm1ethraceBlack          0.6933        0.6521     0.4768
[many rows omitted]
```

Note: These are weighted averages now, since some controls get weight of 0.5.

3. Check balance and sample size

But 2-1 matching improved sample size.
Ratio = 1

```
Sample sizes:
          Control Treated
All           826     581
Matched       388     388
Unmatched     438     193
Discarded       0       0
```

Ratio = 2

```
Sample sizes:
          Control Treated
All           826     581
Matched       508     388
Unmatched     318     193
Discarded       0       0
```

Other things you can do

To improve balance:

- Tighter caliper
- Require exact matches on some covariates
- Match with replacement

To improve efficiency:

- Full matching (weight every observation, see the `optmatch` package, can be called from `MatchIt`)
- Radius matching (match all observations within the caliper)

Steps of propensity score modeling

1. Estimate the propensity score
2. Match on or weight by the propensity score
3. Check balance and sample size
4. Repeat (1) - (3) until happy
5. Look at the outcome

5. Look at the outcome

WARNING: Only do this once you're happy with the quality of the matches. No peeking!

	Unmatched	N	Matched	N.Matched
Nonpoor	98.08	826	93.64	508
Poor	90.29	581	91.06	388

Table: Vocabulary score at age 5, by childhood poverty[1]

[1] Standard errors for these quantites are harder to get given the weights. If you use these methods, read up on how to calculate SEs appropriately!

Outline

1 Propensity scores

2 Coarsened exact matching

3 Mediation

4 Posters

5 Marginal structural models

Coarsened exact matching

We ideally want exact matches.

Propensity score matching gets approximate matches on a continuous distance.

CEM gets exact matches on a coarsened version of the variables.

Coarsened exact matching: Steps

1. Determine the amount of imbalance you are willing to accept.
2. Run CEM to see the resulting sample size. Tighter controls on imbalance
   - improve internal validity, but
   - reduce the sample size and
   - change the estimand
3. Iterate 1-2 until you are happy.
4. Then look at the outcome.

Using CEM

Load the cem package (http://gking.harvard.edu/cem) Most of these slides are motivated by the vignette: `vignette("cem")`

```
match4 <- cem(treatment = "pov2",
              data = ff1.restricted,
              drop = "hv4ppvtstd",
              eval.imbalance = T)
```

CEM output

CEM removes treated and control units in cells where both groups are not represented.
WARNING: This changes the estimand to the FSATT.

	GFALSE	GTRUE
All	826	581
Matched	443	497
Unmatched	383	84

CEM also returns several imbalance measures

```
Multivariate Imbalance Measure: L1=0.380
Percentage of local common support: LCS=52.0%

Univariate Imbalance Measures:

               statistic   type           L1 min 25% 50% 75% max
hv3ppvtstd_m -0.004646195 (diff) 5.551115e-17   0   0   0   1   0
cm1ethrace   13.172287018 (Chi2) 0.000000e+00  NA  NA  NA  NA  NA
cm1relf      29.497322844 (Chi2) 0.000000e+00  NA  NA  NA  NA  NA
cm1edu       87.510613938 (Chi2) 0.000000e+00  NA  NA  NA  NA  NA
cf1edu       43.549330476 (Chi2) 4.163336e-17  NA  NA  NA  NA  NA
```

CEM: Choosing cutpoints

CEM will coarsen continuous variables (like mother's vocabulary score) algorithmically by default.

```
> match4$breaks
$hv3ppvtstd_m
 [1]  40.00000  50.90909  61.81818  72.72727  83.63636  94.54545
 [7] 105.45455 116.36364 127.27273 138.18182 149.09091 160.00000
```

It's better to choose breaks that are theoretically meaningful!

```
match5 <- cem(treatment = "pov2",
              data = ff1.restricted,
              drop = "hv4ppvtstd",
              cutpoints = list(hv3ppvtstd_m = c(
                70, 80, 90, 100, 110, 120, 130
              )),
              eval.imbalance = T)
```

CEM: Difference in means estimator

We can use a simple difference in weighted means:

$$\widehat{FSATT} = \frac{\sum_{i:D_i=1} w_i Y_i}{\sum_{i:D_i=1} w_i} - \frac{\sum_{i:D_i=0} w_i Y_i}{\sum_{i:D_i=0} w_i}$$

```
data.frame(ff1.restricted,
           w = match5$w) %>%
  group_by(pov2) %>%
  summarize(estimate = weighted.mean(hv4ppvtstd,
                                     w = w)) %>%
  spread(key = pov2, value = estimate) %>%
  mutate(tau = `TRUE` - `FALSE`)
```

This estimator indicates that poverty reduces vocabulary scores by about 1.3 points.

CEM: Estimating a model

To estimate a model with CEM, you need to use the weights that come out of the process.

```
des <- svydesign(
  ids = ~0,
  weights = match5$w,
  data = ff1.restricted
)
adjusted <- svyglm(hv4ppvtstd ~ pov2 + hv3ppvtstd_m +
                     cm1ethrace + cm1relf + cm1edu + cf1edu,
                   design = des)
```

CEM results

	Unmatched	Matched
(Intercept)	73.31	70.42
pov2TRUE	-1.69	-1.10
hv3ppvtstd_m	0.26	0.28
cm1ethraceBlack	-6.23	-8.52
cm1ethraceHispanic	-5.98	-6.65
cm1ethraceOther	-2.80	-7.59
cm1relfCohabiting	0.21	0.62
cm1relfOther	0.97	2.84
cm1eduHS	0.04	0.04
cm1eduSome college	4.22	3.63
cm1eduCollege	6.69	14.17
cf1eduHS	0.90	3.66
cf1eduSome college	2.89	3.88
cf1eduCollege	4.09	23.45
N	1407.00	925.00

(For comparison, the difference in means estimator was -1.3)

WARNING about sample sizes

We might be tempted to make small bins, but this will severely restrict sample size.
With bins of width 10 for mother's vocabulary score:

	GFALSE	GTRUE
All	826	581
Matched	443	497
Unmatched	383	84

With bins of width 1 for mother's vocabulary score:

	GFALSE	GTRUE
All	826	581
Matched	170	197
Unmatched	656	384

The sample over which the FSATT is estimated is quite different!

General thoughts on CEM

- Reduces model dependence
- Best if followed by a parametric model
- Great for highlighting the challenge of multivariate balance
- You should also consider other methods. New ones are always being invented.

Outline

1 Propensity scores

2 Coarsened exact matching

3 Mediation

4 Posters

5 Marginal structural models

Takeaways about mediation

Takeaways about mediation

Brandon wants you to know

- Mediators and moderators are different

Takeaways about mediation

Brandon wants you to know

- Mediators and moderators are different
- "Direct" effects are ambiguous: there are different estimands

Takeaways about mediation

Brandon wants you to know

- Mediators and moderators are different
- "Direct" effects are ambiguous: there are different estimands
- Sequential ignorability is a hard assumption

Takeaways about mediation

Brandon wants you to know

- Mediators and moderators are different
- "Direct" effects are ambiguous: there are different estimands
- Sequential ignorability is a hard assumption

Ian wants to add

- The word "causal" does not have to be followed by the word "mechanisms."

Takeaways about mediation

Brandon wants you to know

- Mediators and moderators are different
- "Direct" effects are ambiguous: there are different estimands
- Sequential ignorability is a hard assumption

Ian wants to add

- The word "causal" does not have to be followed by the word "mechanisms."
- People often add mediation as a finishing touch on their papers as though it is easy.

Takeaways about mediation

Brandon wants you to know

- Mediators and moderators are different
- "Direct" effects are ambiguous: there are different estimands
- Sequential ignorability is a hard assumption

Ian wants to add

- The word "causal" does not have to be followed by the word "mechanisms."
- People often add mediation as a finishing touch on their papers as though it is easy. It is not.

Takeaways about mediation

Brandon wants you to know

- Mediators and moderators are different
- "Direct" effects are ambiguous: there are different estimands
- Sequential ignorability is a hard assumption

Ian wants to add

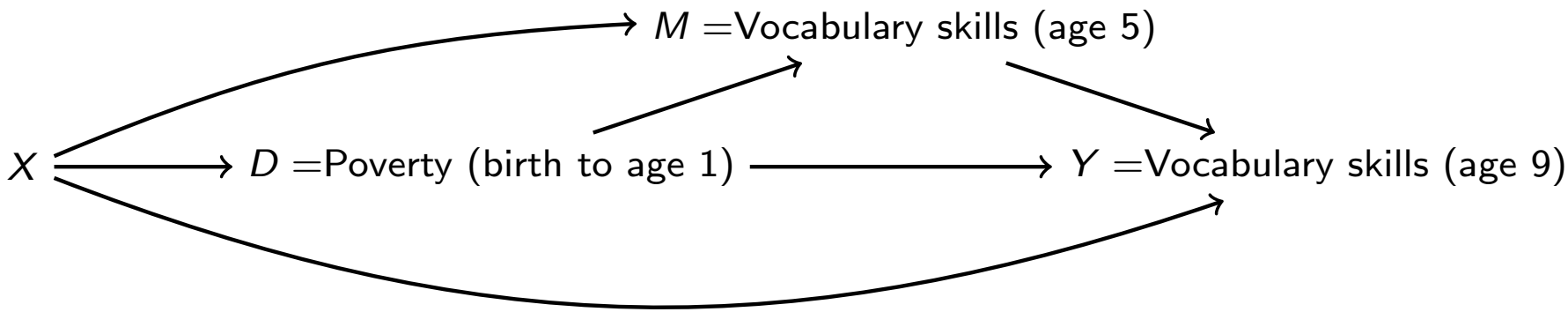
- The word "causal" does not have to be followed by the word "mechanisms."
- People often add mediation as a finishing touch on their papers as though it is easy. It is not.
- Things get really hard with multiple mediators.

Mediation

Question: Is the effect of childhood poverty on vocabulary scores at age 9 mediated by vocabulary scores at age 5?

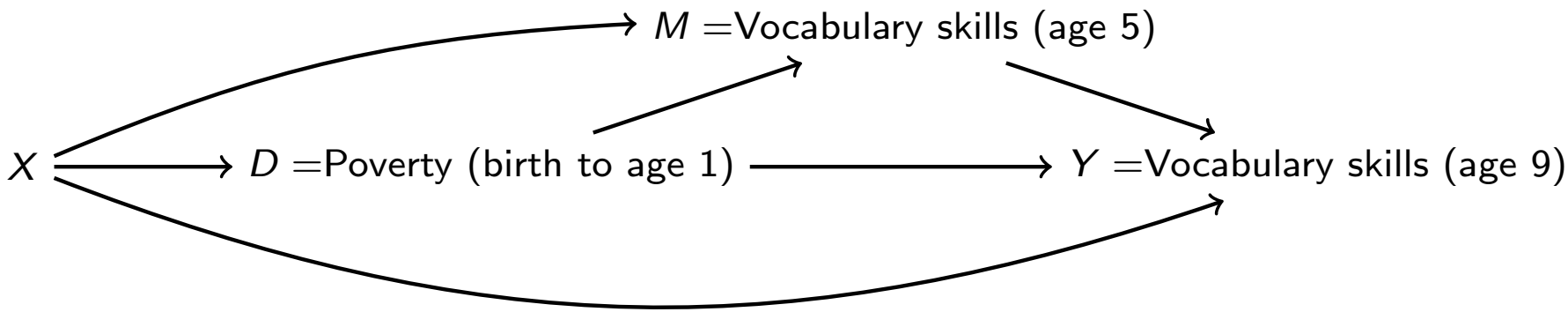
Mediation

Question: Is the effect of childhood poverty on vocabulary scores at age 9 mediated by vocabulary scores at age 5?



Mediation

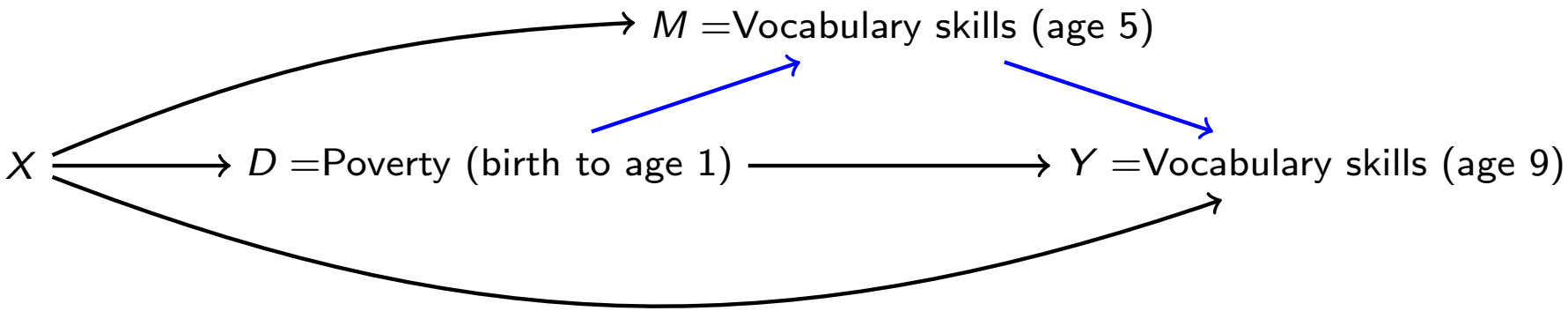
Question: Is the effect of childhood poverty on vocabulary scores at age 9 mediated by vocabulary scores at age 5?



In broader terms, does school readiness explain the link between family income and academic achievement?

Or do poor kids continue to face direct disadvantages from poverty during elementary school years?

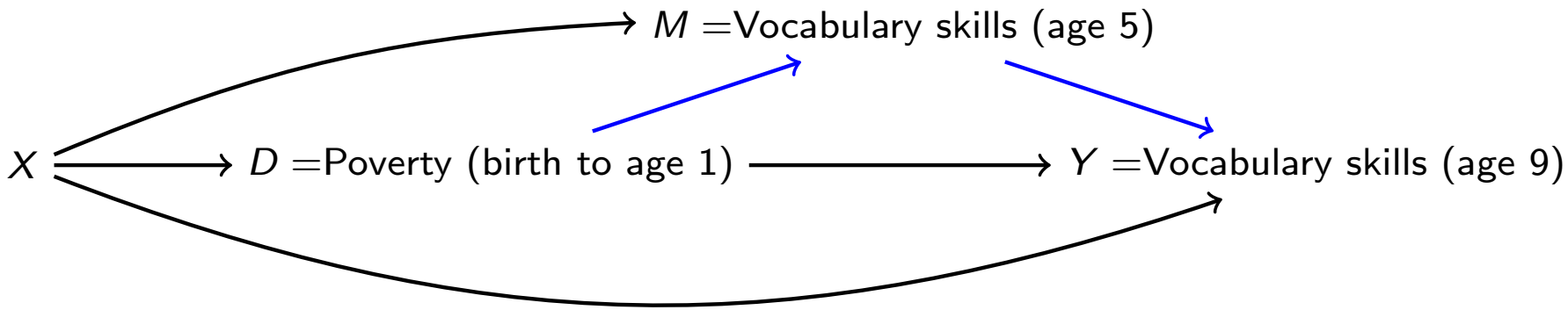
Natural indirect effect



Natural indirect effect (NIE):

$$\delta_i(d) = Y_i(d, M_i(1)) - Y_i(d, M_i(0))$$

Natural indirect effect



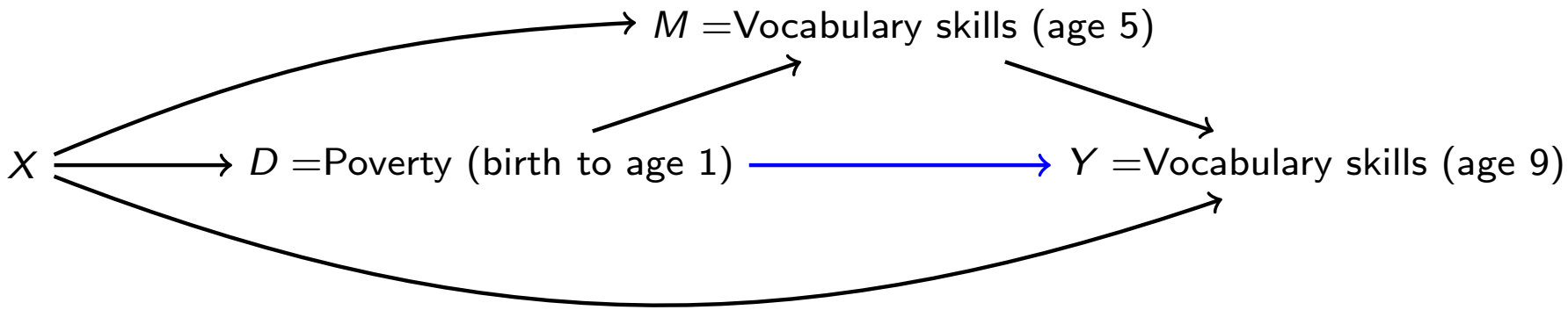
Natural indirect effect (NIE):

$$\delta_i(d) = Y_i(d, M_i(1)) - Y_i(d, M_i(0))$$

In words: (vocabulary skills at age 9 if you had the school readiness you would have if born into poverty) -

(vocabulary skills at age 9 if you had the school readiness you would have if not born into poverty)

Natural direct effect



Natural direct effect (NDE):

$$\eta_i(d) = Y_i(1, M_i(d)) - Y_i(0, M_i(d))$$

In words: (vocabulary skills at age 9 if you were born in poverty, but school readiness was held at the observed value) -

(vocabulary skills at age 9 if you were not born into poverty, but school readiness was held at the observed value)

Quantities that cannot be observed

In causal inference, there are always values we don't get to see.

Quantities that cannot be observed

In causal inference, there are always values we don't get to see.

If a unit is treated, then $Y_i(1)$ is observed but $Y_i(0)$ is unobserved.

Quantities that cannot be observed

In causal inference, there are always values we don't get to see.

If a unit is treated, then $Y_i(1)$ is observed but $Y_i(0)$ is unobserved.

This is the fundamental problem of causal inference.

Quantities that cannot be observed

In causal inference, there are always values we don't get to see.

If a unit is treated, then $Y_i(1)$ is observed but $Y_i(0)$ is unobserved.

This is the fundamental problem of causal inference.

But, we could have observed $Y_i(0)$ if unit i had been given control.

Quantities that cannot be observed

In causal inference, there are always values we don't get to see.

If a unit is treated, then $Y_i(1)$ is observed but $Y_i(0)$ is unobserved.

This is the fundamental problem of causal inference.

But, we could have observed $Y_i(0)$ if unit i had been given control.

Mediation requires identification of quantities that can never be observed.

Quantities that cannot be observed

Suppose $d = 1$: we want to decompose effects among those who are treated.

The NDE is:

$$\eta_i(1) = \underbrace{Y_i(1, M_i(1))} - \underbrace{Y_i(0, M_i(1))}$$

The NIE is:

$$\delta_i(1) = \underbrace{Y_i(1, M_i(1))} - \underbrace{Y_i(1, M_i(0))}$$

Quantities that cannot be observed

Suppose $d = 1$: we want to decompose effects among those who are treated.

The NDE is:

$$\eta_i(1) = \underbrace{Y_i(1, M_i(1))}_{\text{Observed!}} - \underbrace{Y_i(0, M_i(1))}$$

The NIE is:

$$\delta_i(1) = \underbrace{Y_i(1, M_i(1))} - \underbrace{Y_i(1, M_i(0))}$$

Quantities that cannot be observed

Suppose $d = 1$: we want to decompose effects among those who are treated.

The NDE is:

$$\eta_i(1) = \underbrace{Y_i(1, M_i(1))}_{\text{Observed!}} - \underbrace{Y_i(0, M_i(1))}_{\text{Never observed}}$$

The NIE is:

$$\delta_i(1) = \underbrace{Y_i(1, M_i(1))} - \underbrace{Y_i(1, M_i(0))}$$

Quantities that cannot be observed

Suppose $d = 1$: we want to decompose effects among those who are treated.

The NDE is:

$$\eta_i(1) = \underbrace{Y_i(1, M_i(1))}_{\text{Observed!}} - \underbrace{Y_i(0, M_i(1))}_{\text{Never observed}}$$

The NIE is:

$$\delta_i(1) = \underbrace{Y_i(1, M_i(1))}_{\text{Observed!}} - \underbrace{Y_i(1, M_i(0))}_{}$$

Quantities that cannot be observed

Suppose $d = 1$: we want to decompose effects among those who are treated.

The NDE is:

$$\eta_i(1) = \underbrace{Y_i(1, M_i(1))}_{\text{Observed!}} - \underbrace{Y_i(0, M_i(1))}_{\text{Never observed}}$$

The NIE is:

$$\delta_i(1) = \underbrace{Y_i(1, M_i(1))}_{\text{Observed!}} - \underbrace{Y_i(1, M_i(0))}_{\text{Never observed}}$$

Sequential Ignorability, Part 1

The treatment is independent of the potential outcomes and potential mediators, conditional on a set of covariates:

$$\{Y_i(d', m), M_i(d)\} \perp\!\!\!\perp D_i | X_i = x$$

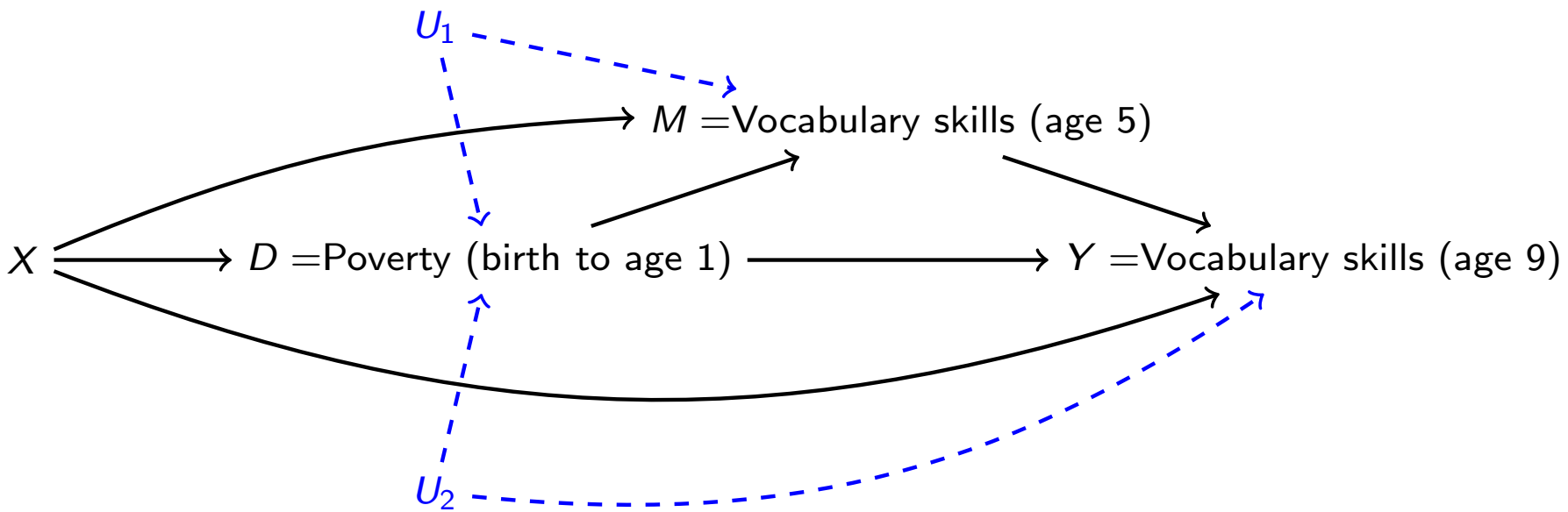
This holds in a (well-executed) experiment.

Sequential Ignorability, Part 1

The treatment is independent of the potential outcomes and potential mediators, conditional on a set of covariates:

$$\{Y_i(d', m), M_i(d)\} \perp\!\!\!\perp D_i | X_i = x$$

This holds in a (well-executed) experiment.
It says that there is no U_1 or U_2 below.



Sequential Ignorability, Part 2

The mediator is ignorable with respect to the outcome, conditional on the treatment:

$$Y_i(d', m) \perp\!\!\!\perp M_i(d) | D_i = d, X_i = x$$

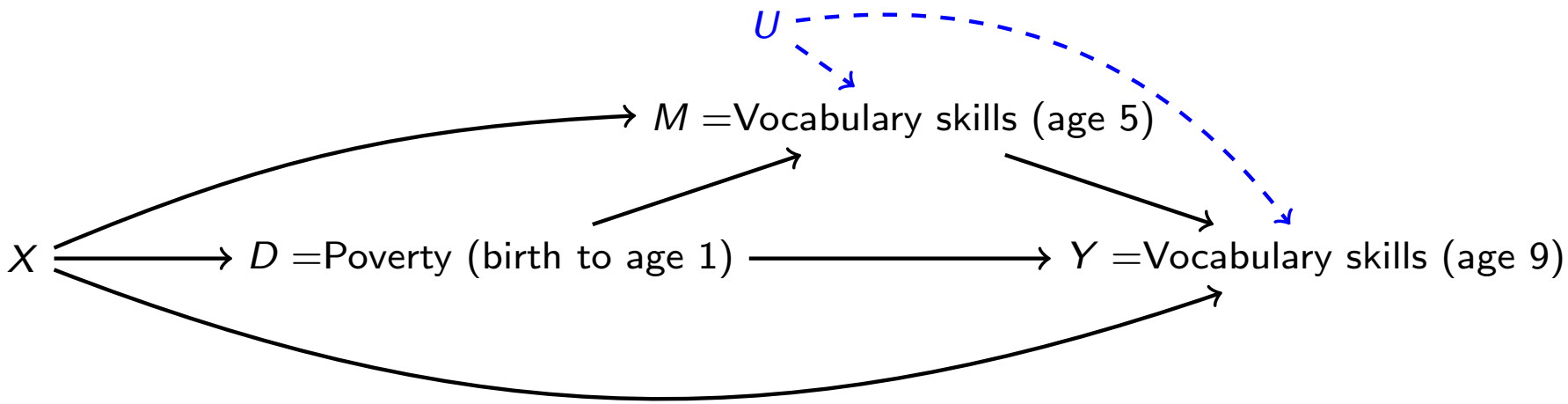
This must hold for all values of d, d'.

Sequential Ignorability, Part 2

The mediator is ignorable with respect to the outcome, conditional on the treatment:

$$Y_i(d', m) \perp\!\!\!\perp M_i(d) | D_i = d, X_i = x$$

This must hold for all values of d, d'.
It says that there is no U below.

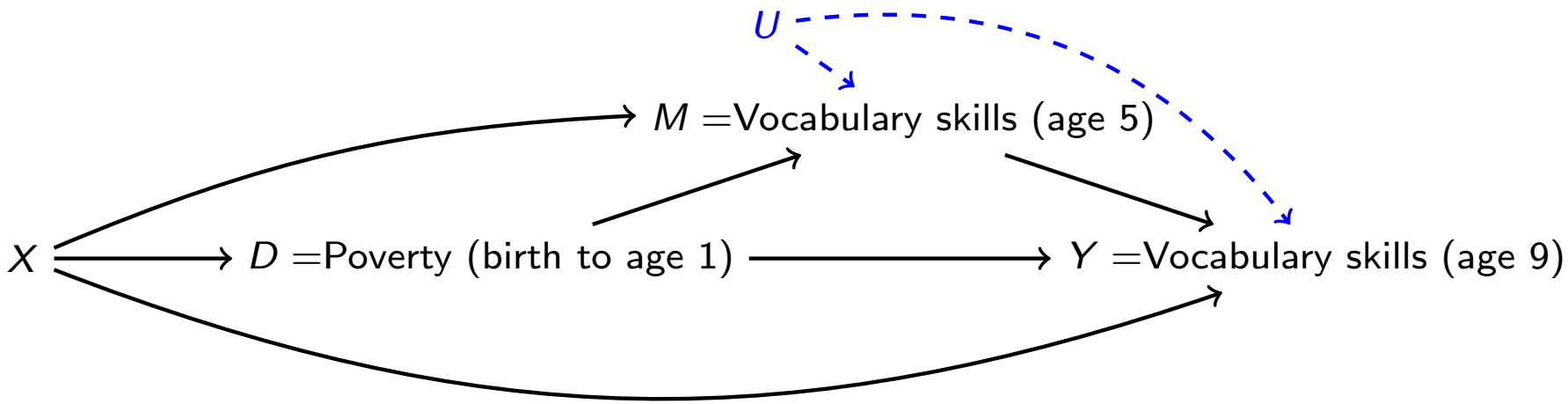


Sequential Ignorability, Part 2

The mediator is ignorable with respect to the outcome, conditional on the treatment:

$$Y_i(d', m) \perp\!\!\!\perp M_i(d) | D_i = d, X_i = x$$

This must hold for all values of d, d'.
It says that there is no U below.



Ian finds this assumption rarely believable.

Estimation via `mediation` package

Specify a model for the mediator:

```
mediator.model <- lm(hv4ppvtstd ~ cm2povco + hv3ppvtstd_m +
                       cm1ethrace + cm1relf + cm1edu + cf1edu,
                     data = med)
```

Specify a model for the outcome:

```
outcome.model <- lm(hv5_ppvtss ~ hv4ppvtstd + cm2povco + hv3ppvtstd_m +
                      cm1ethrace + cm1relf + cm1edu + cf1edu,
                    data = med)
```

Call `mediate`

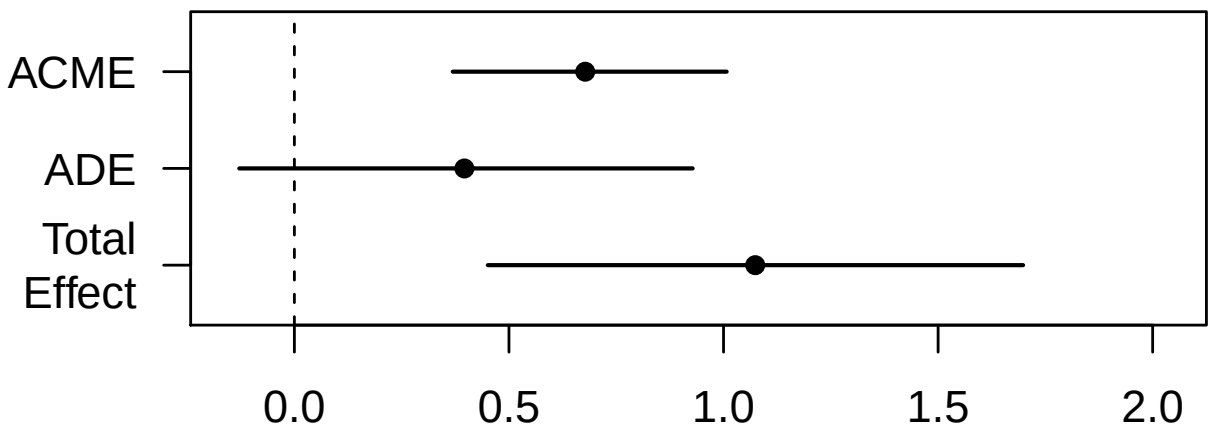
```
mediation <- mediate(
  model.m = mediator.model,
  model.y = outcome.model,
  treat = "cm2povco",
  mediator = "hv4ppvtstd"
)
```

Results

```
> summary(mediation)

Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

               Estimate 95% CI Lower 95% CI Upper p-value
ACME              0.678        0.369        1.007    0.00
ADE               0.396       -0.128        0.928    0.14
Total Effect      1.074        0.451        1.698    0.00
Prop. Mediated    0.625        0.385        1.255    0.00

Sample Size Used: 1407

Simulations: 1000
```

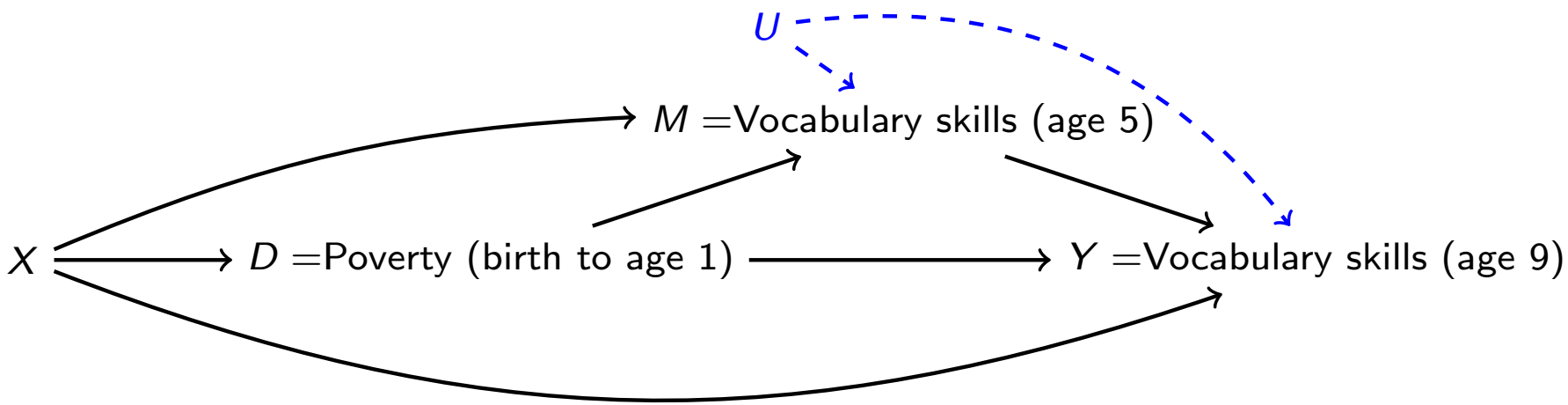
Plotting results

```
plot(mediation)
```



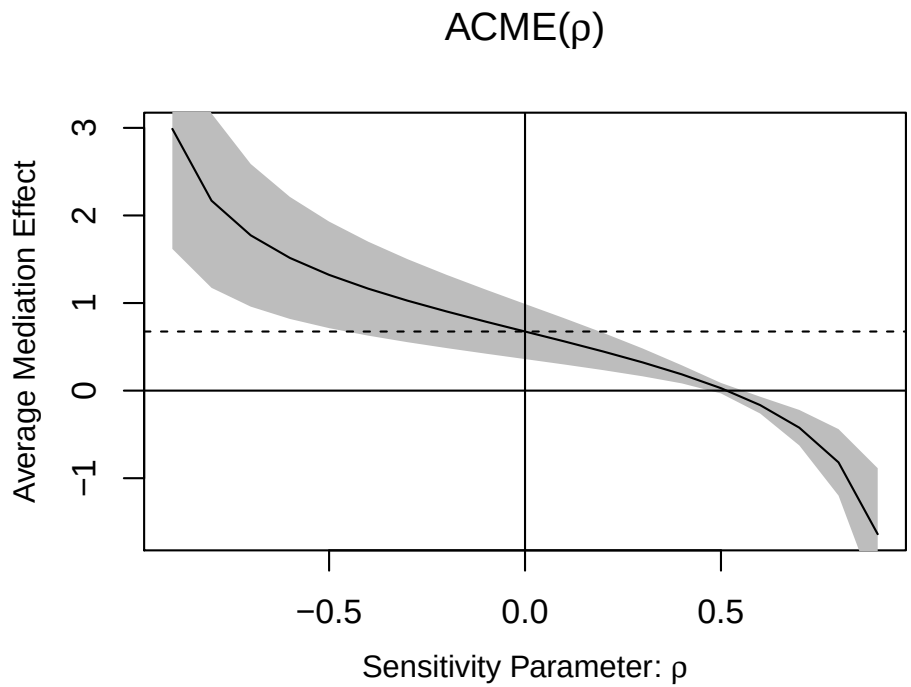
Sensitivity analysis

Violations of sequential ignorability, part 2, will create correlation in the error terms of the mediator and the outcome models.
We can parameterize this to explore the sensitivity of our results to violoations of this assumption.



Sensitivity analysis

```
sensitivity <- medsens(mediation)
plot(sensitivity)
```



This plots the sensitivity to correlation in the error terms of the mediator and outcome models that would arise from violation of sequential ignorability, part 2. If you use this in your own work, you should read the papers [link] and also think carefully about what degree of violation is plausible!

Outline

1 Propensity scores

2 Coarsened exact matching

3 Mediation

4 Posters

5 Marginal structural models

Posters

- Plan ahead - they take time to prepare!
- Make use of figures
- Minimize amount of text
- Tell a story!
- We are putting examples on Blackboard.
- We will discuss mine now!

Outline

1 Propensity scores

2 Coarsened exact matching

3 Mediation

4 Posters

5 Marginal structural models

Question

What is the cumulative effect of poverty on vocabulary skills?

Question

What is the cumulative effect of poverty on vocabulary skills?

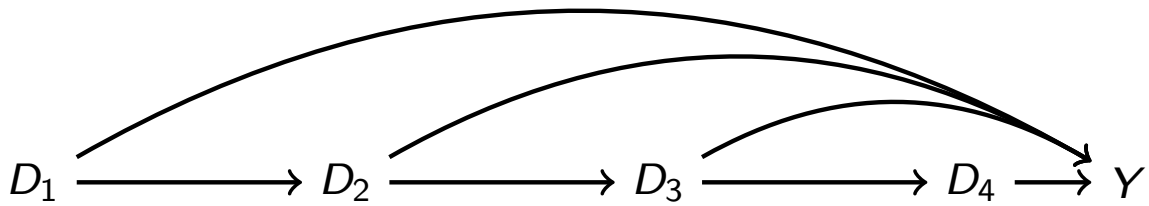
L = Counfounders (family structure)
D = Poverty
Y = Vocabulary skills at child age 9
Subscripts = time

$$D_1 \longrightarrow D_2 \longrightarrow D_3 \longrightarrow D_4 \longrightarrow Y$$

Question

What is the cumulative effect of poverty on vocabulary skills?

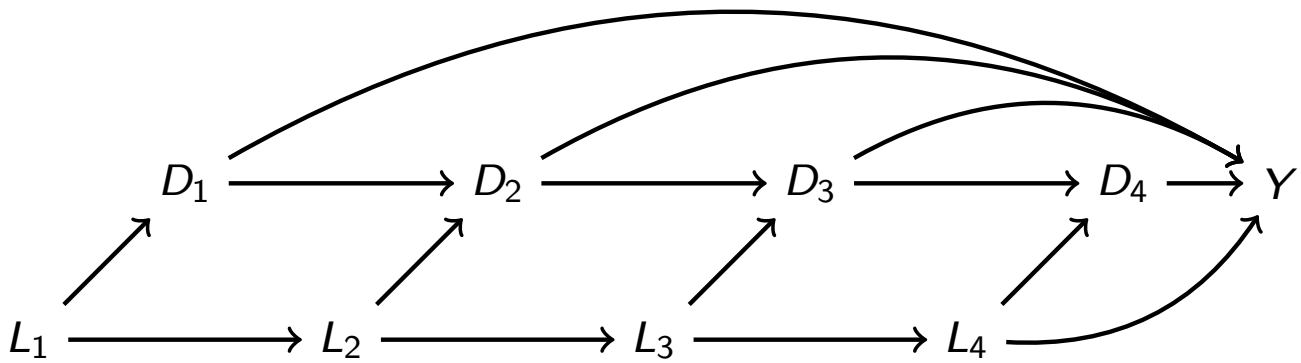
L = Counfounders (family structure)
D = Poverty
Y = Vocabulary skills at child age 9
Subscripts = time



Question

What is the cumulative effect of poverty on vocabulary skills?

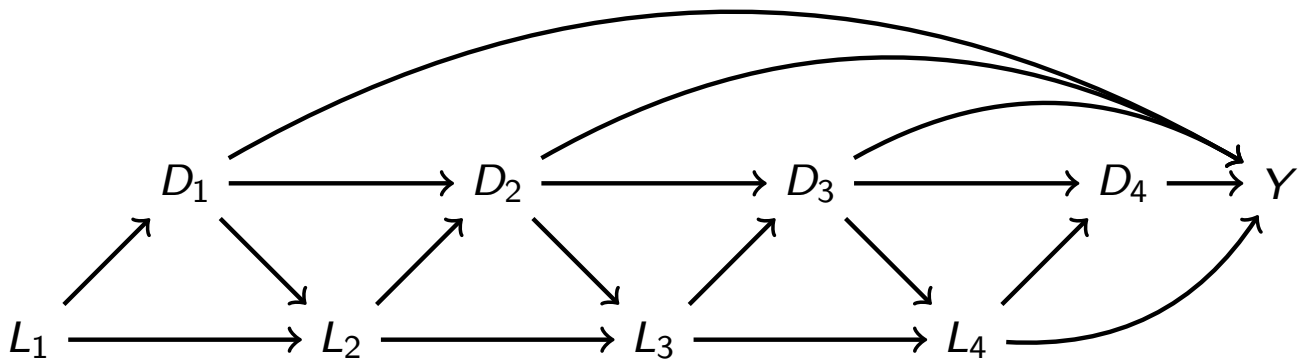
L = Counfounders (family structure)
D = Poverty
Y = Vocabulary skills at child age 9
Subscripts = time



Question

What is the cumulative effect of poverty on vocabulary skills?

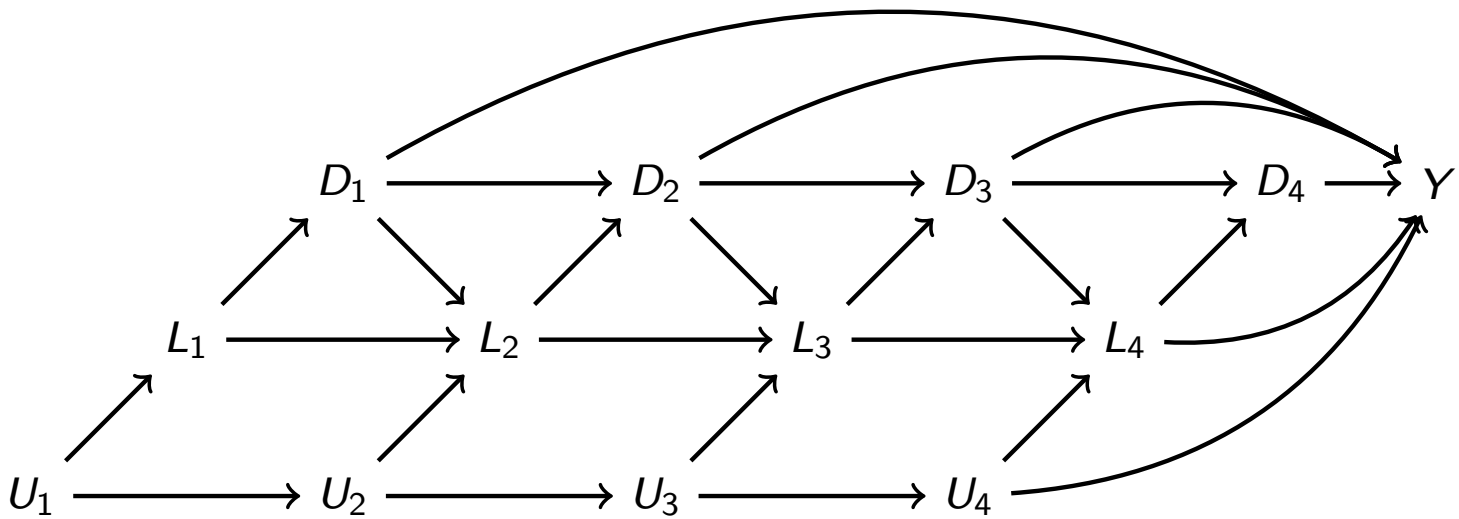
L = Counfounders (family structure)
D = Poverty
Y = Vocabulary skills at child age 9
Subscripts = time



Question

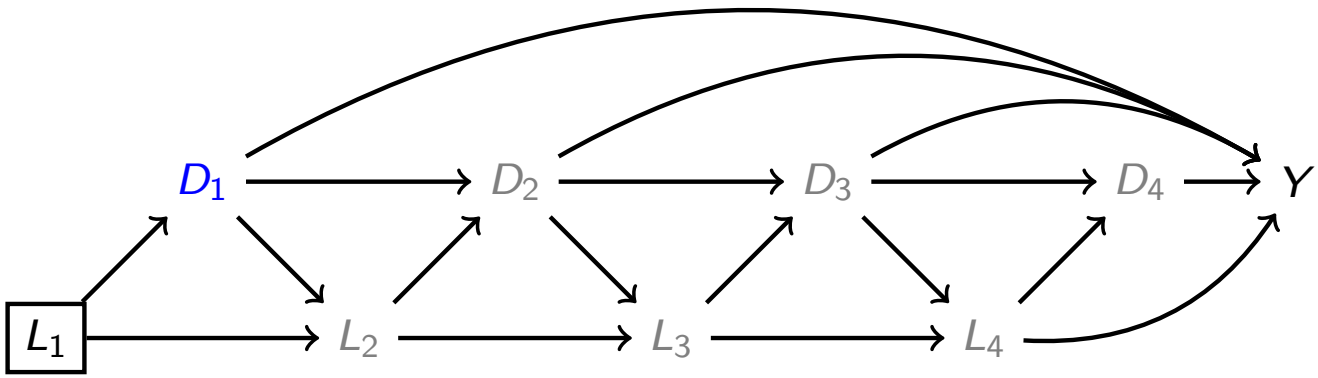
What is the cumulative effect of poverty on vocabulary skills?

L = Counfounders (family structure)
D = Poverty
Y = Vocabulary skills at child age 9
Subscripts = time



$D_1 \rightarrow Y$

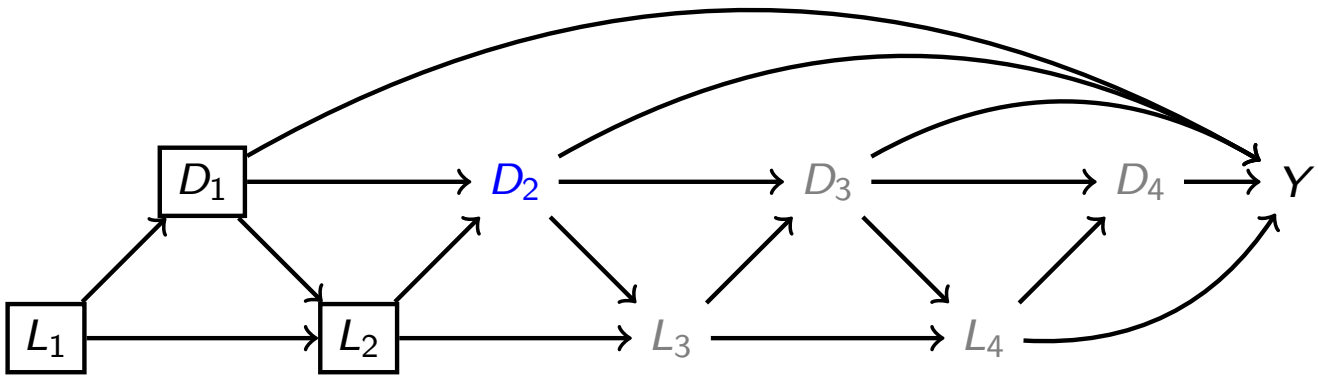
What is the cumulative effect of poverty on vocabulary skills?



$$p_1(L_1) = P(D_1 = 1 \mid L_1)$$

$D_2 \rightarrow Y$

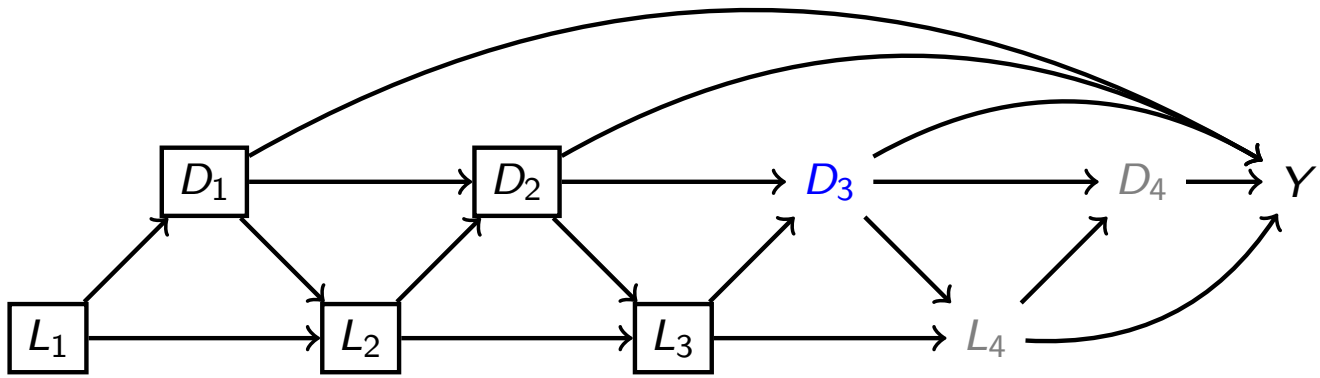
What is the cumulative effect of poverty on vocabulary skills?



$$p_2(L_1, L_2, D_1) = P(D_2 = 1 \mid L_1, L_2, D_1)$$

$D_3 \rightarrow Y$

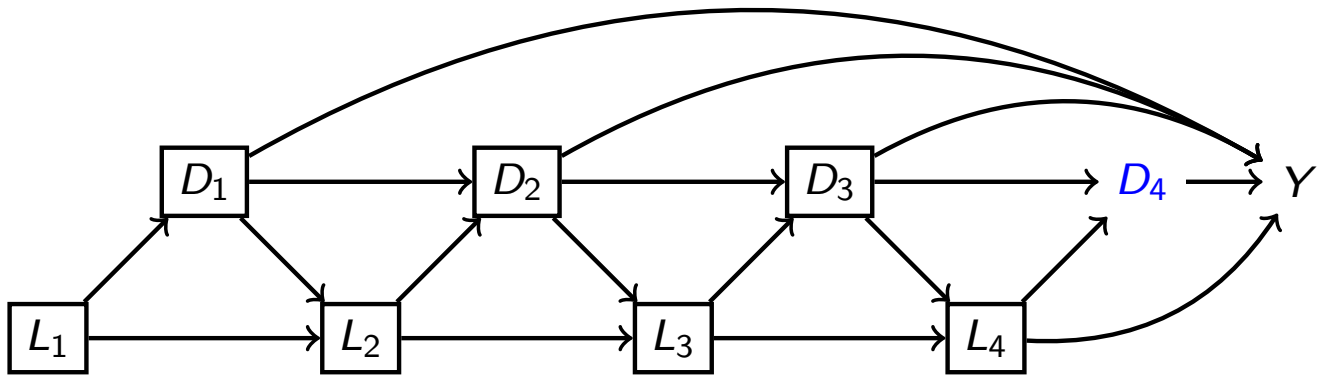
What is the cumulative effect of poverty on vocabulary skills?



$$p_3(L_1, L_2, L_3, D_1, D_2) = P(D_2 = 1 \mid L_1, L_2, L_3, D_1, D_2)$$

$D_4 \rightarrow Y$

What is the cumulative effect of poverty on vocabulary skills?



$$p_4(L_1, L_2, L_3, L_4, D_1, D_2, D_3) = P(D_2 = 1 \mid L_1, L_2, L_3, L_4, D_1, D_2, D_3)$$

Weighting treatment regimes

In ordinary causal inference, we can weight by the inverse probability of treatment.

This is a propensity score adjustment like matching.

In marginal structural models, we weight by the inverse probability of the dynamic treatment regime.

$$
\begin{aligned}
w_i &= \frac{1}{P(D_1 = d_1, D_2 = d_2, D_3 = d_3, D_4 = d_4)} \\
&= \frac{1}{\prod_{t=1}^{4} p_{it}^{D_{it}} (1 - p_{it})^{1 - D_{it}}}
\end{aligned}
$$

Often we actually use a stabilized weight. Read up on this if you want to use it!

Our example

What is the cumulative effect of poverty on vocabulary skills?

1. Estimate logit for poverty at age 3 given the past

Our example

What is the cumulative effect of poverty on vocabulary skills?

1. Estimate logit for poverty at age 3 given the past
   - Also one given only baseline covariates for stabilized weights

Our example

What is the cumulative effect of poverty on vocabulary skills?

1. Estimate logit for poverty at age 3 given the past
   - Also one given only baseline covariates for stabilized weights
2. Estimate a logit for poverty at age 5 given the past

Our example

What is the cumulative effect of poverty on vocabulary skills?

1. Estimate logit for poverty at age 3 given the past
   - Also one given only baseline covariates for stabilized weights
2. Estimate a logit for poverty at age 5 given the past
   - Also one given only baseline covariates for stabilized weights

Our example

What is the cumulative effect of poverty on vocabulary skills?

1. Estimate logit for poverty at age 3 given the past
   - Also one given only baseline covariates for stabilized weights
2. Estimate a logit for poverty at age 5 given the past
   - Also one given only baseline covariates for stabilized weights
3. Calculate stablized weights

Our example

What is the cumulative effect of poverty on vocabulary skills?

1. Estimate logit for poverty at age 3 given the past
   - Also one given only baseline covariates for stabilized weights
2. Estimate a logit for poverty at age 5 given the past
   - Also one given only baseline covariates for stabilized weights
3. Calculate stablized weights

$$sw = \frac{P(D_1 \mid X)P(D_3 \mid X)P(D_5 \mid X)}{P(D_1 \mid X)P(D_3 \mid X, D_1, L_1)P(D_5 \mid X, D_1, L_1, D_3, L_3)}$$

Our example

What is the cumulative effect of poverty on vocabulary skills?

1. Estimate logit for poverty at age 3 given the past
   - Also one given only baseline covariates for stabilized weights
2. Estimate a logit for poverty at age 5 given the past
   - Also one given only baseline covariates for stabilized weights
3. Calculate stablized weights

$$sw = \frac{P(D_1 \mid X)P(D_3 \mid X)P(D_5 \mid X)}{P(D_1 \mid X)P(D_3 \mid X, D_1, L_1)P(D_5 \mid X, D_1, L_1, D_3, L_3)}$$

4. Estimate a weighted model, conditional on $\sum D_i$ and on X, but not conditional on L.

Fitting the weighted model

```
svyglm(ppvt ~ sumPov +
         cm1ethrace + cm1relf +
         cm1edu + cf1edu + pcg2,
       design = svydesign(ids = ~0,
                          weights = sw,
                          data = msm))
```

Fitting the weighted model

```
svyglm(ppvt ~ sumPov +
         cm1ethrace + cm1relf +
         cm1edu + cf1edu + pcg2,
       design = svydesign(ids = ~0,
                          weights = sw,
                          data = msm))
```

Answer: Each year spent in poverty causes a 2 point reduction in vocabulary score at age 9.

Marginal structural models: Takeaways

- Useful if you want to estimate the effect of a treatment that changes over time
- Natural extension of propensity score weighting

Marginal structural models: Takeaways

- Useful if you want to estimate the effect of a treatment that changes over time
- Natural extension of propensity score weighting
- Marginal in the sense that a unit increase in treatment at any period is associated with the same outcome effect.

Marginal structural models: Takeaways

- Useful if you want to estimate the effect of a treatment that changes over time
- Natural extension of propensity score weighting
- Marginal in the sense that a unit increase in treatment at any period is associated with the same outcome effect.
- Structural in the sense of being causally identified.

Marginal structural models: Takeaways

- Useful if you want to estimate the effect of a treatment that changes over time
- Natural extension of propensity score weighting
- Marginal in the sense that a unit increase in treatment at any period is associated with the same outcome effect.
- Structural in the sense of being causally identified.
- Weights can be unstable and sensitive to modeling decisions

Marginal structural models: Takeaways

- Useful if you want to estimate the effect of a treatment that changes over time
- Natural extension of propensity score weighting
- Marginal in the sense that a unit increase in treatment at any period is associated with the same outcome effect.
- Structural in the sense of being causally identified.
- Weights can be unstable and sensitive to modeling decisions
- MSMs are a special case of a broader framework for dynamic treatment regimes invented by Jamie Robins. Related models are known as structural nested models and G-estimation

Marginal structural models: Takeaways

- Useful if you want to estimate the effect of a treatment that changes over time
- Natural extension of propensity score weighting
- Marginal in the sense that a unit increase in treatment at any period is associated with the same outcome effect.
- Structural in the sense of being causally identified.
- Weights can be unstable and sensitive to modeling decisions
- MSMs are a special case of a broader framework for dynamic treatment regimes invented by Jamie Robins. Related models are known as structural nested models and G-estimation
- Read more before using!

Causal inference takeaways

Matching

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal explanation and dynamic treatment regimes

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal explanation and dynamic treatment regimes

- Moderators: pre-treatment variables correlated with treatment effects

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal explanation and dynamic treatment regimes

- Moderators: pre-treatment variables correlated with treatment effects
- Mediators: post-treatment variables through which treatment effects flow

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal explanation and dynamic treatment regimes

- Moderators: pre-treatment variables correlated with treatment effects
- Mediators: post-treatment variables through which treatment effects flow
- Establishing mediation is hard, requiring a strong sequential ignorability assumption

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal explanation and dynamic treatment regimes

- Moderators: pre-treatment variables correlated with treatment effects
- Mediators: post-treatment variables through which treatment effects flow
- Establishing mediation is hard, requiring a strong sequential ignorability assumption
- Controlled direct effects are easier to estimate, but less natural

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal explanation and dynamic treatment regimes

- Moderators: pre-treatment variables correlated with treatment effects
- Mediators: post-treatment variables through which treatment effects flow
- Establishing mediation is hard, requiring a strong sequential ignorability assumption
- Controlled direct effects are easier to estimate, but less natural
- Dynamic treatment regimes likewise involve strong assumptions

Posters

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal explanation and dynamic treatment regimes

- Moderators: pre-treatment variables correlated with treatment effects
- Mediators: post-treatment variables through which treatment effects flow
- Establishing mediation is hard, requiring a strong sequential ignorability assumption
- Controlled direct effects are easier to estimate, but less natural
- Dynamic treatment regimes likewise involve strong assumptions

Posters

- Use minimal text

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal explanation and dynamic treatment regimes

- Moderators: pre-treatment variables correlated with treatment effects
- Mediators: post-treatment variables through which treatment effects flow
- Establishing mediation is hard, requiring a strong sequential ignorability assumption
- Controlled direct effects are easier to estimate, but less natural
- Dynamic treatment regimes likewise involve strong assumptions

Posters

- Use minimal text
- Use lots of figures

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal explanation and dynamic treatment regimes

- Moderators: pre-treatment variables correlated with treatment effects
- Mediators: post-treatment variables through which treatment effects flow
- Establishing mediation is hard, requiring a strong sequential ignorability assumption
- Controlled direct effects are easier to estimate, but less natural
- Dynamic treatment regimes likewise involve strong assumptions

Posters

- Use minimal text
- Use lots of figures
- Look at examples

Causal inference takeaways

Matching

- Matching helps conceptually to clarify identification assumptions
- But those identification assumptions are the same as in regression
- Matching is a useful estimation strategy for reducing model dependence
- Weighting is often good, but check that you don't have a few enormous weights.

Causal explanation and dynamic treatment regimes

- Moderators: pre-treatment variables correlated with treatment effects
- Mediators: post-treatment variables through which treatment effects flow
- Establishing mediation is hard, requiring a strong sequential ignorability assumption
- Controlled direct effects are easier to estimate, but less natural
- Dynamic treatment regimes likewise involve strong assumptions

Posters

- Use minimal text
- Use lots of figures
- Look at examples
- Start early!