9. Parametric g-formula: Continuous treatments

Ian Lundberg
Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

20 Sep 2022

Learning goals for today

At the end of class, you will be able to:

1. Define causal effects with continuous treatments
2. Understand the risks of extrapolation
3. Select causal estimands for continuous treatments that support the production of credible estimates

Why continuous treatments are hard

Why continuous treatments are hard

Person 1

Person 2

Person 3

Person 4

Why continuous treatments are hard

	Outcome under treatment value	
	Untreated	Treated
Person 1	∘	∘
Person 2	∘	∘
Person 3	∘	∘
Person 4	∘	∘

Why continuous treatments are hard

	Factual treatment	Outcome under treatment value	
		Untreated	Treated
Person 1	Untreated	$\circ$	$\circ$
Person 2	Treated	$\circ$	$\circ$
Person 3	Treated	$\circ$	$\circ$
Person 4	Untreated	$\circ$	$\circ$

Why continuous treatments are hard

	Factual treatment	Outcome under treatment value	
		Untreated	Treated
Person 1	Untreated	●	○
Person 2	Treated	○	●
Person 3	Treated	○	●
Person 4	Untreated	●	○

Why continuous treatments are hard

	Factual treatment	Outcome under treatment value 1	2	3	4	5	...
Person 1	3	○	○	●	○	○	○
Person 2	2	○	●	○	○	○	○
Person 3	5	○	○	○	○	●	○
Person 4	4	○	○	○	●	○	○

Why continuous treatments are hard

	Factual treatment	Outcome under treatment value
		1 2 3 4 5 . . .
Person 1	3	○○○○○○○○●○○○○○○○○○○○○
Person 2	2	○○○○●○○○○○○○○○○○○○○○○
Person 3	5	○○○○○○○○○○○○○○○○●○○○○
Person 4	4	○○○○○○○○○○○○●○○○○○○○○

Why continuous treatments are hard

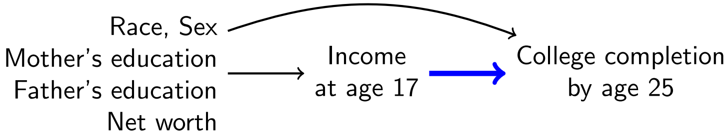
	Factual treatment	Outcome under treatment value 1 2 3 4 5 . . .
Person 1	3	○○○○○○○○●○○○○○○○○○○○○
Person 2	2	○○○○●○○○○○○○○○○○○○○○○
Person 3	5	○○○○○○○○○○○○○○○○●○○○○
Person 4	4	○○○○○○○○○○○○●○○○○○○○○

Question: Why is this hard? How would you proceed?

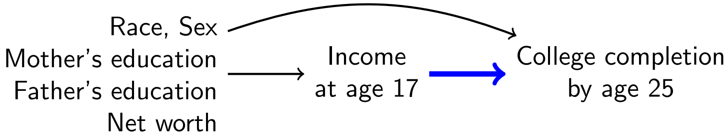
A motivating example

(current work with Jennie E. Brand)

Causal identification.

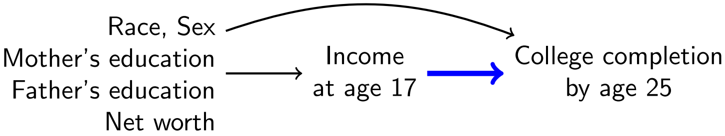


Causal identification.

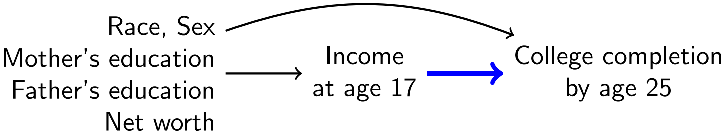


Mayer 1997
Brooks-Gunn & Duncan 1997
Duncan et al. 2010
Elwert & Pfeffer 2022

Causal identification.

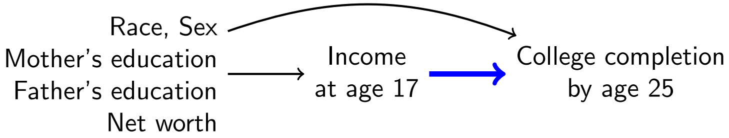


Causal identification.



Ideal estimation. What we'd like to do in a huge sample.

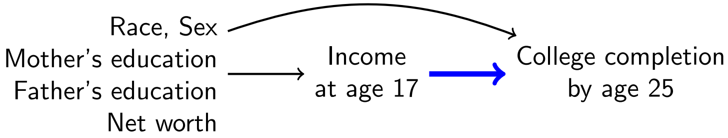
Causal identification.



Ideal estimation. What we'd like to do in a huge sample.

1) Restrict to covariate stratum $\vec{L} = \vec{\ell}$

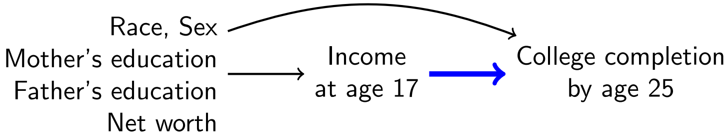
Causal identification.



Ideal estimation. What we'd like to do in a huge sample.

1) Restrict to covariate stratum $\vec{L} = \vec{\ell}$

Causal identification.

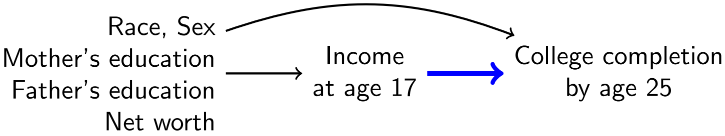


Ideal estimation. What we'd like to do in a huge sample.

1) Restrict to covariate stratum $\vec{L} = \vec{\ell}$

2) Estimate a smooth curve $\mathrm{E}(Y(a) \mid \vec{L} = \vec{\ell})$

Causal identification.

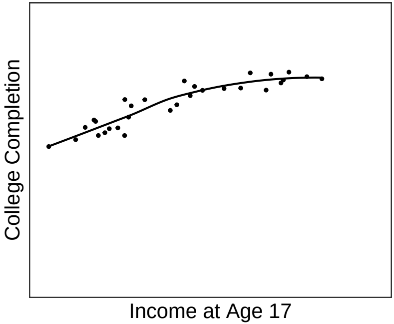


Ideal estimation.

What we'd like to do in a huge sample.

1) Restrict to covariate stratum $\vec{L} = \vec{\ell}$

2) Estimate a smooth curve $\mathrm{E}(Y(a) \mid \vec{L} = \vec{\ell})$



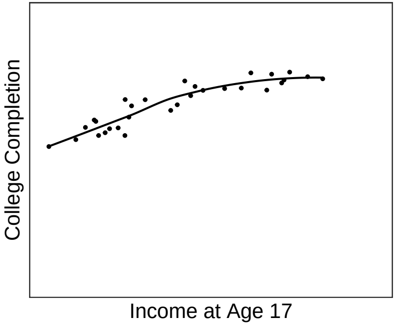
Causal identification.

Race, Sex
Mother's education
Father's education
Net worth

$\longrightarrow$ Income at age 17 $\longrightarrow$ College completion by age 25

Ideal estimation. What we'd like to do in a huge sample.

1) Restrict to covariate stratum $\vec{L} = \vec{\ell}$
2) Estimate a smooth curve $\mathrm{E}(Y(a) \mid \vec{L} = \vec{\ell})$
3) Repeat



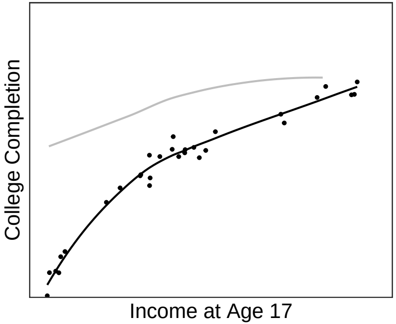
Causal identification.

Race, Sex
Mother's education
Father's education
Net worth

$\longrightarrow$ Income at age 17 $\longrightarrow$ College completion by age 25

Ideal estimation. What we'd like to do in a huge sample.

1) Restrict to covariate stratum $\vec{L} = \vec{\ell}$
2) Estimate a smooth curve $\mathrm{E}(Y(a) \mid \vec{L} = \vec{\ell})$
3) Repeat



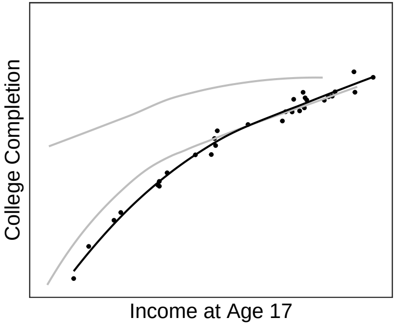
Causal identification.

Race, Sex
Mother's education
Father's education
Net worth

$\longrightarrow$ Income at age 17 $\longrightarrow$ College completion by age 25

Ideal estimation. What we'd like to do in a huge sample.

1) Restrict to covariate stratum $\vec{L} = \vec{\ell}$
2) Estimate a smooth curve $\mathrm{E}(Y(a) \mid \vec{L} = \vec{\ell})$
3) Repeat



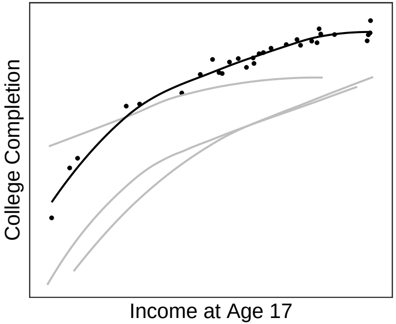
Causal identification.

Race, Sex
Mother's education
Father's education
Net worth

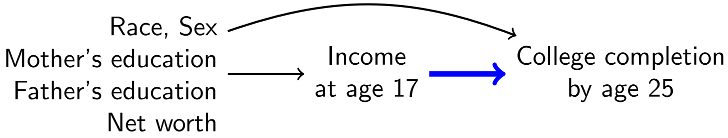
$\longrightarrow$ Income at age 17 $\longrightarrow$ College completion by age 25

Ideal estimation. What we'd like to do in a huge sample.

1) Restrict to covariate stratum $\vec{L} = \vec{\ell}$
2) Estimate a smooth curve $\mathrm{E}(Y(a) \mid \vec{L} = \vec{\ell})$
3) Repeat



Causal identification.

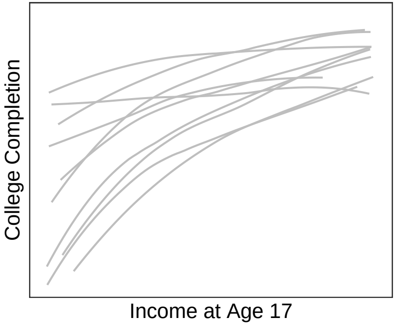


Ideal estimation. What we'd like to do in a huge sample.

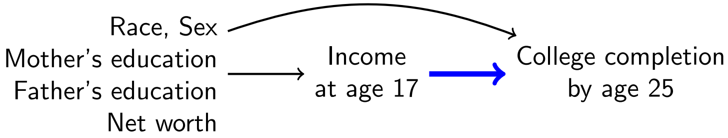
1) Restrict to covariate stratum $\vec{L} = \vec{\ell}$

2) Estimate a smooth curve $\mathrm{E}(Y(a) \mid \vec{L} = \vec{\ell})$

3) Repeat



Causal identification.

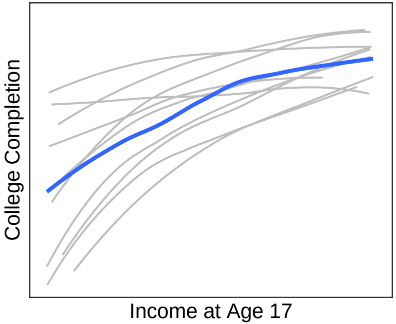


Ideal estimation. What we'd like to do in a huge sample.

1) Restrict to covariate stratum $\vec{L} = \vec{\ell}$

2) Estimate a smooth curve $\mathrm{E}(Y(a) \mid \vec{L} = \vec{\ell})$

3) Repeat



Dose-response curve

The previous slide introduced the **dose-response curve**

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

- Take everyone in the population

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

- Take everyone in the population
- Give them treatment value a

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

- Take everyone in the population
- Give them treatment value a
- What is the average outcome $\text{E}(Y^a)$?

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

- Take everyone in the population
- Give them treatment value a
- What is the average outcome $\mathrm{E}(Y^a)$?
- Draw the curve for all values of a

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

- Take everyone in the population
- Give them treatment value a
- What is the average outcome $\text{E}(Y^a)$?
- Draw the curve for all values of a

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

- Take everyone in the population
- Give them treatment value a
- What is the average outcome $\text{E}(Y^a)$?
- Draw the curve for all values of a

How one estimates:

- Define the adjustment set $\vec{L}$

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

- Take everyone in the population
- Give them treatment value a
- What is the average outcome $\text{E}(Y^a)$?
- Draw the curve for all values of a

How one estimates:

- Define the adjustment set $\vec{L}$
- Estimate $\text{E}(\text{E}(Y \mid A = a, \vec{L}))$

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

- Take everyone in the population
- Give them treatment value a
- What is the average outcome $\text{E}(Y^a)$?
- Draw the curve for all values of a

How one estimates:

- Define the adjustment set $\vec{L}$
- Estimate $\text{E}(\text{E}(Y \mid A = a, \vec{L}))$
  - Model Y given A and $\vec{L}$

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

- Take everyone in the population
- Give them treatment value a
- What is the average outcome $\text{E}(Y^a)$?
- Draw the curve for all values of a

How one estimates:

- Define the adjustment set $\vec{L}$
- Estimate $\text{E}(\text{E}(Y \mid A = a, \vec{L}))$
  - Model Y given A and $\vec{L}$
  - Set A to the value a

Dose-response curve

The previous slide introduced the **dose-response curve**

Target trial:

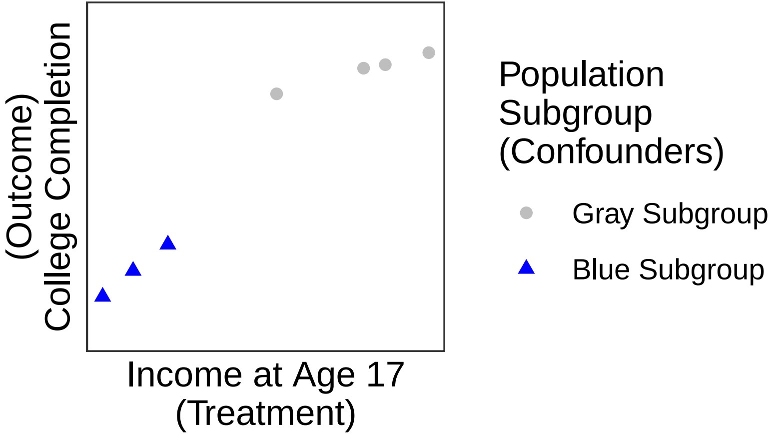
- Take everyone in the population
- Give them treatment value a
- What is the average outcome $\mathrm{E}(Y^a)$?
- Draw the curve for all values of a

How one estimates:

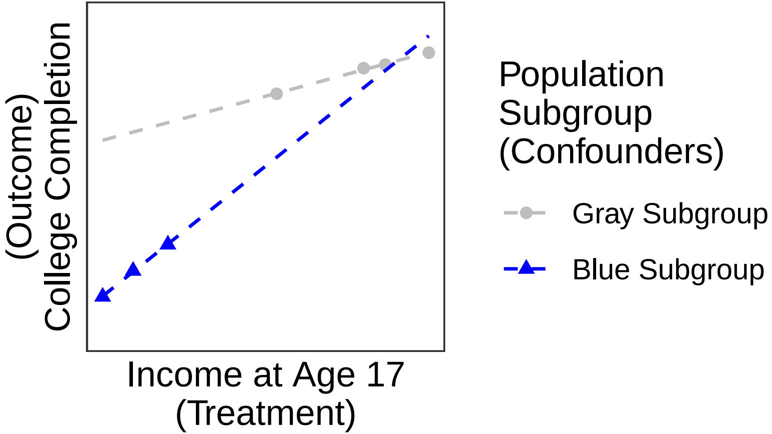
- Define the adjustment set $\vec{L}$
- Estimate $\mathrm{E}(\mathrm{E}(Y \mid A = a, \vec{L}))$
    - Model Y given A and $\vec{L}$
    - Set A to the value a
    - Make a prediction for everyone. Report the average

Warning: In strongly confounded settings, the dose-response curve is a dangerous goal. Here is why.

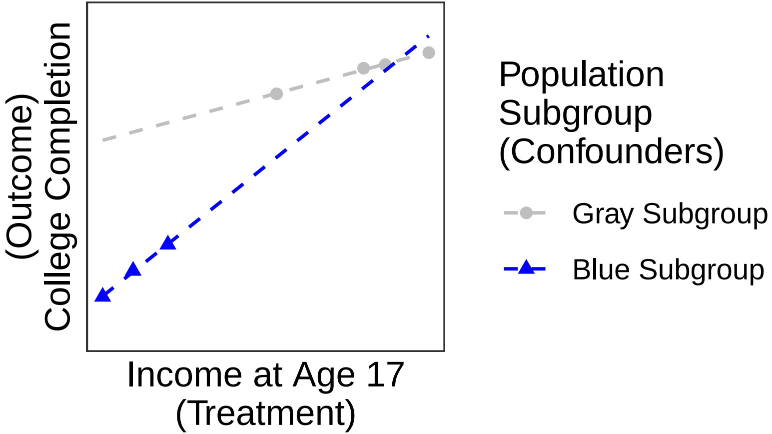
Sparse data can lead to **extrapolation**



Sparse data can lead to **extrapolation**



Sparse data can lead to **extrapolation**

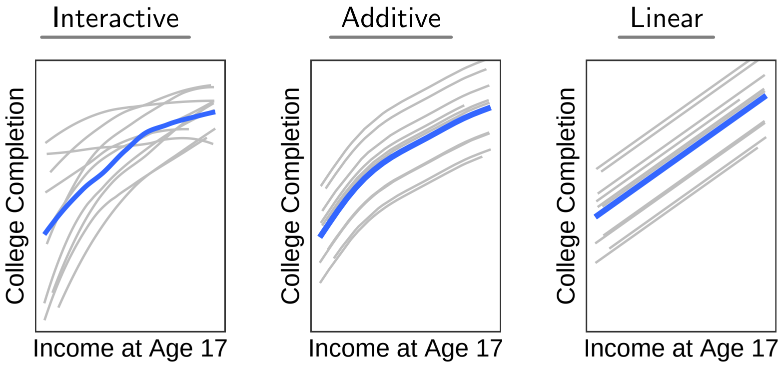


Question: What should one do?

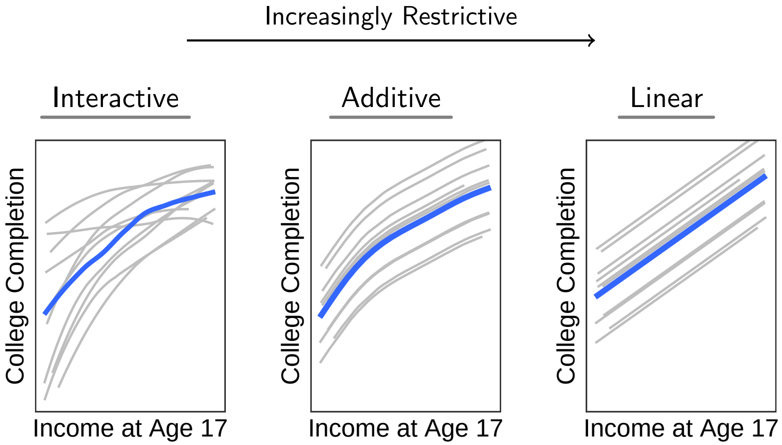
Sparse data can lead to **extrapolation**: Three solutions

1. Extrapolate transparently: Additive model
2. Reduce extrapolation: Small shift
3. Avoid extrapolation: Incremental causal effects

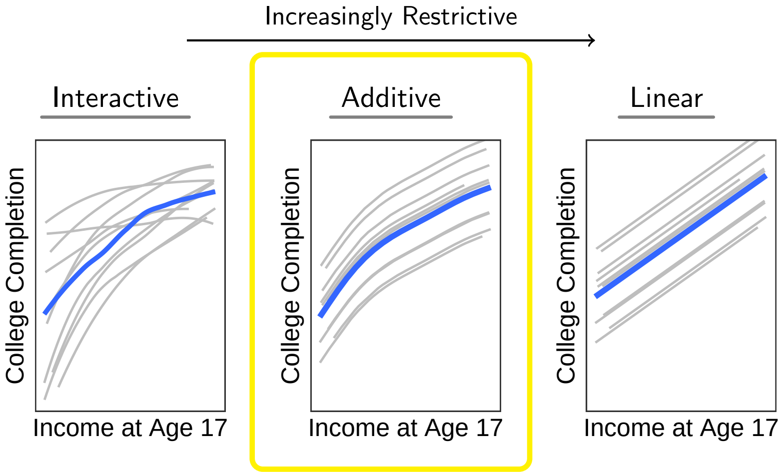
1) Extrapolate transparently: Additive model



1) Extrapolate transparently: Additive model



1) Extrapolate transparently: Additive model



2) Reduce extrapolation: Small shift

- For each person, do not consider all treatment values a

2) Reduce extrapolation: Small shift

- For each person, do not consider all treatment values a
- Consider a shift from A to

2) Reduce extrapolation: Small shift

- For each person, do not consider all treatment values a
- Consider a shift from A to
  - $A + \delta$ (additive shift)

2) Reduce extrapolation: Small shift

- For each person, do not consider all treatment values a
- Consider a shift from A to
  - $A + \delta$ (additive shift)
  - kA (multiplicative shift)

2) Reduce extrapolation: Small shift

- For each person, do not consider all treatment values a
- Consider a shift from A to
  - $A + \delta$ (additive shift)
  - kA (multiplicative shift)

2) Reduce extrapolation: Small shift

- For each person, do not consider all treatment values a
- Consider a shift from A to
  - $A + \delta$ (additive shift)
  - kA (multiplicative shift)

Example: The effect of adding 1: $\mathrm{E}(Y^{A+1} - Y^{A})$

2) Reduce extrapolation: Small shift

- For each person, do not consider all treatment values a
- Consider a shift from A to
  - $A + \delta$ (additive shift)
  - kA (multiplicative shift)

Example: The effect of adding 1: $\mathrm{E}(Y^{A+1} - Y^{A})$

Advantages:

2) Reduce extrapolation: Small shift

- For each person, do not consider all treatment values a
- Consider a shift from A to
  - $A + \delta$ (additive shift)
  - kA (multiplicative shift)

Example: The effect of adding 1: $E(Y^{A+1} - Y^A)$

Advantages:

- Only two potential outcomes for each person

2) Reduce extrapolation: Small shift

- For each person, do not consider all treatment values a
- Consider a shift from A to
  - $A + \delta$ (additive shift)
  - kA (multiplicative shift)

Example: The effect of adding 1: $\mathrm{E}(Y^{A+1} - Y^{A})$

Advantages:

- Only two potential outcomes for each person
- One of them is observed

2) Reduce extrapolation: Small shift

- For each person, do not consider all treatment values a
- Consider a shift from A to
  - $A + \delta$ (additive shift)
  - kA (multiplicative shift)

Example: The effect of adding 1: $\mathrm{E}(Y^{A+1} - Y^{A})$

Advantages:

- Only two potential outcomes for each person
- One of them is observed
- The other is not far away

3) Avoid extrapolation: Incremental causal effects[1]

[1] Rothenhäusler, D., & Yu, B. (2019). Incremental causal effects. arXiv preprint arXiv:1907.13258.

3) Avoid extrapolation: Incremental causal effects[1]

Take the limit of an extremely small additive shift:

$$\frac{\mathsf{E}\left(Y^{A+\delta} - Y^{A}\right)}{\delta}$$

[1] Rothenhäusler, D., & Yu, B. (2019). Incremental causal effects. arXiv preprint arXiv:1907.13258.

3) Avoid extrapolation: Incremental causal effects[1]

Take the limit of an extremely small additive shift:

$$\frac{\mathsf{E}\left(Y^{A+\delta} - Y^{A}\right)}{\delta}$$

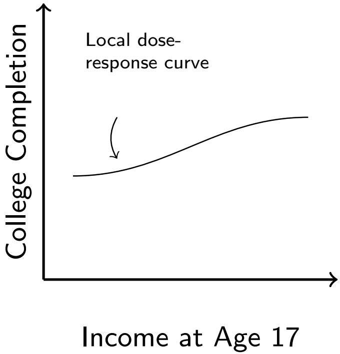
This is the **local slope**
of the **local dose-response curve**,
averaged over the population
(see next slide).

[1] Rothenhäusler, D., & Yu, B. (2019). Incremental causal effects. arXiv preprint arXiv:1907.13258.

3) Avoid extrapolation: Incremental causal effects

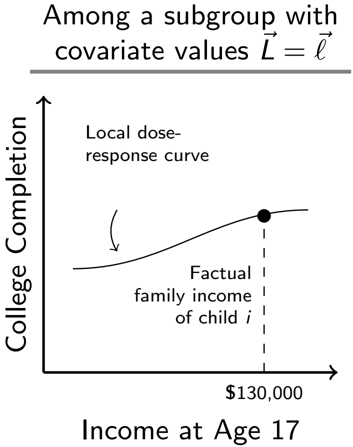
3) Avoid extrapolation: Incremental causal effects

Among a subgroup with covariate values $\vec{L} = \vec{\ell}$



Rothenhausler & Yu 2019

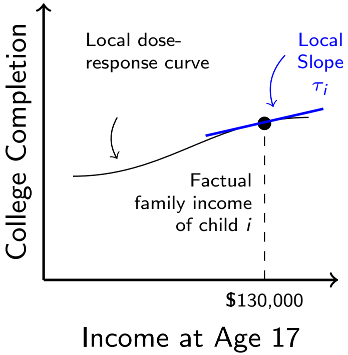
3) Avoid extrapolation: Incremental causal effects



Rothenhausler & Yu 2019

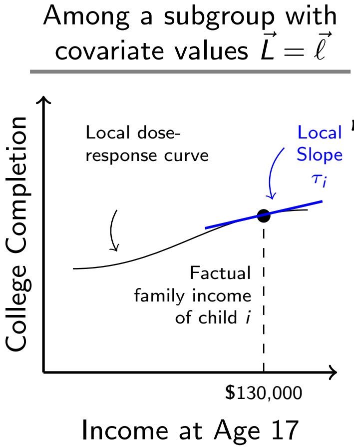
3) Avoid extrapolation: Incremental causal effects

Among a subgroup with covariate values $\vec{L} = \vec{\ell}$



Rothenhausler & Yu 2019

3) Avoid extrapolation: Incremental causal effects



Report the
population
average
of these

Rothenhausler & Yu 2019

Example from a current working paper

[2]This estimation idea comes from: Friedberg, R., Tibshirani, J., Athey, S., & Wager, S. (2020). Local linear forests. Journal of Computational and Graphical Statistics, 30(2), 503-517.

Example from a current working paper

Locally linear estimation within coarsened subgroups[2]

[2] This estimation idea comes from: Friedberg, R., Tibshirani, J., Athey, S., & Wager, S. (2020). Local linear forests. Journal of Computational and Graphical Statistics, 30(2), 503-517.

Example from a current working paper

Locally linear estimation within coarsened subgroups[2]

1. Learn a random forest
   - Predict college completion
   - Predictors include covariates $\vec{L}$ and treatment A

[2]This estimation idea comes from: Friedberg, R., Tibshirani, J., Athey, S., & Wager, S. (2020). Local linear forests. Journal of Computational and Graphical Statistics, 30(2), 503-517.

Example from a current working paper

Locally linear estimation within coarsened subgroups[2]

1. Learn a random forest
   - Predict college completion
   - Predictors include covariates $\vec{L}$ and treatment A
2. Each leaf is a coarsened subgroup of $\{\vec{L}, A\}$

[2] This estimation idea comes from: Friedberg, R., Tibshirani, J., Athey, S., & Wager, S. (2020). Local linear forests. Journal of Computational and Graphical Statistics, 30(2), 503-517.

Example from a current working paper

Locally linear estimation within coarsened subgroups[2]

1. Learn a random forest
   - Predict college completion
   - Predictors include covariates $\vec{L}$ and treatment A
2. Each leaf is a coarsened subgroup of $\{\vec{L}, A\}$
3. Within each leaf, estimate OLS with $\{\vec{L}, A\}$ as predictors

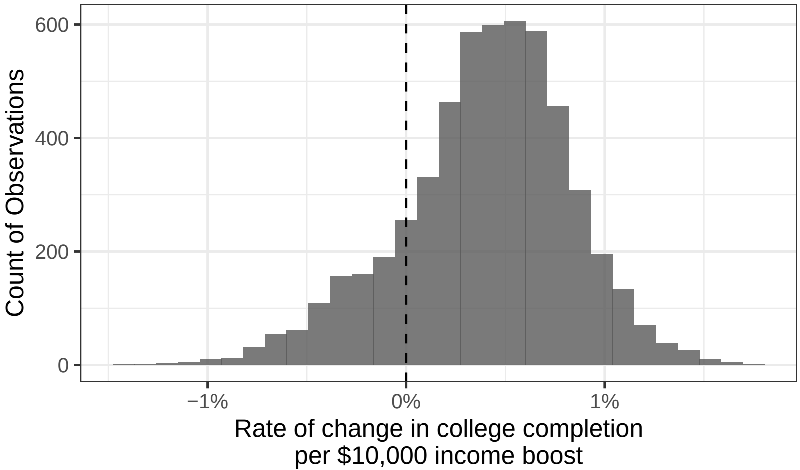
[2] This estimation idea comes from: Friedberg, R., Tibshirani, J., Athey, S., & Wager, S. (2020). Local linear forests. Journal of Computational and Graphical Statistics, 30(2), 503-517.

Example from a current working paper

Locally linear estimation within coarsened subgroups[2]

1. Learn a random forest
   - Predict college completion
   - Predictors include covariates $\vec{L}$ and treatment A
2. Each leaf is a coarsened subgroup of $\{\vec{L}, A\}$
3. Within each leaf, estimate OLS with $\{\vec{L}, A\}$ as predictors
4. Coefficient on the treatment $\approx$ incremental causal effect

[2] This estimation idea comes from: Friedberg, R., Tibshirani, J., Athey, S., & Wager, S. (2020). Local linear forests. Journal of Computational and Graphical Statistics, 30(2), 503-517.


600
400
200
0
Count of Observations
−1%
0%
1%
Rate of change in college completion
per $10,000 income boost

Describing the covariates of those with

Small Effects	Big Effects
-0.1% to 0.3% increase in P(College) per $10,000 income boost	0.6% to 0.9% increase in P(College) per $10,000 income boost

Describing the covariates of those with

	Small Effects	Big Effects
	-0.1% to 0.3% increase in P(College) per $10,000 income boost	0.6% to 0.9% increase in P(College) per $10,000 income boost
% Black		

Describing the covariates of those with

	Small Effects	Big Effects
	-0.1% to 0.3% increase in P(College) per $10,000 income boost	0.6% to 0.9% increase in P(College) per $10,000 income boost
% Black	19%	

Describing the covariates of those with

	Small Effects	Big Effects
	-0.1% to 0.3% increase in P(College) per $10,000 income boost	0.6% to 0.9% increase in P(College) per $10,000 income boost
% Black	19%	34%

Describing the covariates of those with

	Small Effects -0.1% to 0.3% increase in P(College) per $10,000 income boost	Big Effects 0.6% to 0.9% increase in P(College) per $10,000 income boost
% Black	19%	34%
% Hispanic	18%	

Describing the covariates of those with

	Small Effects -0.1% to 0.3% increase in P(College) per $10,000 income boost	Big Effects 0.6% to 0.9% increase in P(College) per $10,000 income boost
% Black	19%	34%
% Hispanic	18%	25%

Describing the covariates of those with

	Small Effects -0.1% to 0.3% increase in P(College) per $10,000 income boost	Big Effects 0.6% to 0.9% increase in P(College) per $10,000 income boost
% Black	19%	34%
% Hispanic	18%	25%
% Father Absent	21%	

Describing the covariates of those with

	Small Effects -0.1% to 0.3% increase in P(College) per $10,000 income boost	Big Effects 0.6% to 0.9% increase in P(College) per $10,000 income boost
% Black	19%	34%
% Hispanic	18%	25%
% Father Absent	21%	50%

Describing the covariates of those with

	Small Effects -0.1% to 0.3% increase in P(College) per $10,000 income boost	Big Effects 0.6% to 0.9% increase in P(College) per $10,000 income boost
% Black	19%	34%
% Hispanic	18%	25%
% Father Absent	21%	50%
% Neither Parent Finished College	74%	

Describing the covariates of those with

	Small Effects -0.1% to 0.3% increase in P(College) per \$10,000 income boost	Big Effects 0.6% to 0.9% increase in P(College) per \$10,000 income boost
% Black	19%	34%
% Hispanic	18%	25%
% Father Absent	21%	50%
% Neither Parent Finished College	74%	90%

Describing the covariates of those with

	Small Effects -0.1% to 0.3% increase in P(College) per $10,000 income boost	Big Effects 0.6% to 0.9% increase in P(College) per $10,000 income boost
% Black	19%	34%
% Hispanic	18%	25%
% Father Absent	21%	50%
% Neither Parent Finished College	74%	90%
Median Income	$77,751	

Describing the covariates of those with

	Small Effects -0.1% to 0.3% increase in P(College) per $10,000 income boost	Big Effects 0.6% to 0.9% increase in P(College) per $10,000 income boost
% Black	19%	34%
% Hispanic	18%	25%
% Father Absent	21%	50%
% Neither Parent Finished College	74%	90%
Median Income	$77,751	$31,433

Continuous treatments: Concluding thoughts

There exists a common strategy:

Continuous treatments: Concluding thoughts

There exists a common strategy:

$$\begin{aligned}\text{Outcome} = \beta_0 + \beta_1 \times \text{Income} &+ \beta_2 \times (\text{Confounder 1}) \\ &+ \beta_3 \times (\text{Confounder 2}) \\ &+ \ldots + \epsilon\end{aligned}$$

Continuous treatments: Concluding thoughts

There exists a common strategy:

$$\begin{aligned}\text{Outcome} = \beta_0 + \underbrace{\beta_1}_{\text{“effect” of income}} \times \text{Income} &+ \beta_2 \times (\text{Confounder 1}) \\ &+ \beta_3 \times (\text{Confounder 2}) \\ &+ \ldots + \epsilon\end{aligned}$$

Continuous treatments: Concluding thoughts

There exists a common strategy:

$$\begin{aligned}\text{Outcome} = \beta_0 + \underbrace{\beta_1}_{\text{“effect” of income}} \times \text{Income} &+ \beta_2 \times (\text{Confounder 1}) \\ &+ \beta_3 \times (\text{Confounder 2}) \\ &+ \ldots + \epsilon\end{aligned}$$

But the effect actually

— varies across people **heterogeneity**

Continuous treatments: Concluding thoughts

There exists a common strategy:

$$\begin{aligned}\text{Outcome} = \beta_0 + \underbrace{\beta_1}_{\text{“effect” of income}} \times \text{Income} &+ \beta_2 \times (\text{Confounder 1}) \\ &+ \beta_3 \times (\text{Confounder 2}) \\ &+ \ldots + \epsilon\end{aligned}$$

But the effect actually

— varies across people **heterogeneity**

— varies across income values **nonlinearity**

Continuous treatments: Concluding thoughts

There exists a common strategy:

$$\begin{aligned}\text{Outcome} = \beta_0 + \underbrace{\beta_1}_{\text{“effect” of income}} \times \text{Income} &+ \beta_2 \times (\text{Confounder 1}) \\ &+ \beta_3 \times (\text{Confounder 2}) \\ &+ \ldots + \epsilon\end{aligned}$$

But the effect actually

— varies across people **heterogeneity**

— varies across income values **nonlinearity**

Nonlinear and heterogeneous patterns
offer opportunities for new discoveries

Learning goals for today

At the end of class, you will be able to:

1. Define causal effects with continuous treatments
2. Understand the risks of extrapolation
3. Select causal estimands for continuous treatments that support the production of credible estimates

Let me know what you are thinking

tinyurl.com/CausalQuestions

Office hours TTh 11am-12pm and at
calendly.com/ianlundberg/office-hours
Come say hi!

If time allows: Discuss

How would you define the estimand?
Pick one or invent your own example.

1. More time exercising causes lower blood pressure.
2. More time sleeping improves focus in class.
3. Reading more pages per day promotes vocabulary development.
4. Smoking more cigarettes causes higher risk of lung cancer.

Things to consider:

- What treatment values are compared?
- Over whom will you average those counterfactual outcomes?
- Are there dangers of extrapolation? How would you mitigate them?