Omitted Variable Bias (OVB)

Prof. Mike Aguilar

https:/www.linkedin.com/in/mike-aguilar-econ

Outline

1. Case Study
2. Identifying the Bias
   - General Setting
   - Case Study
3. Measuring the Bias
   - General Setting
   - Case Study
4. OVB Sensitivity Analysis
5. "Solving" OVB

OVB - Summary

- Omitted Variable Bias (OVB) is perhaps the most pernicious problem we will confront in causal estimation
- Omitting a relevant regressor generates a biased treatment effect estimate
- We cannot see nor estimate the size of the bias, however we can...
- ...intimate the sign of the bias via domain knowledge, and...
- ...put relevant bounds on the bias

Housekeeping

- Throughout the slide deck you will see "Q", which indicates a question to you, the reader
- You will also see "A", which indicates the associated answer
- It is generally most efficient to learn this material through active participation. Whenever you encounter a "Q", be sure to try and develop your answer before turning the page to the provided "A" answer
- I recommend reading through these slides before engaging in associated coding exercises
- This slide deck is part of a broader curriculum in causal estimation and inference
- Suggested pre-requisite: foundations of regression, DAG

Outline

1. Case Study
2. Identifying the Bias
   - General Setting
   - Case Study
3. Measuring the Bias
   - General Setting
   - Case Study
4. OVB Sensitivity Analysis
5. "Solving" OVB

What is the impact of education on wages?

A local politician wants to promote high school education. If he can show that having a diploma increases wages by $\$3,000$ (or more), he can pass new supportive legislation.



- We have several hundred observations from this politician's district.
- The dataset contains a binary variable (educ) indicating high school completion
- We also have wage data

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

A:

- Outcome:

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

A:

- Outcome: wages

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

A:

- Outcome: wages
- Treatment:

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

A:

- Outcome: wages
- Treatment: complete high school (educ)

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

A:

- Outcome: wages
- Treatment: complete high school (educ)
- Treated:

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

A:

- Outcome: wages
- Treatment: complete high school (educ)
- Treated: those who completed high school

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

A:

- Outcome: wages
- Treatment: complete high school (educ)
- Treated: those who completed high school
- Control:

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

A:

- Outcome: wages
- Treatment: complete high school (educ)
- Treated: those who completed high school
- Control: those who did not complete high school

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

A:

- Outcome: wages
- Treatment: complete high school (educ)
- Treated: those who completed high school
- Control: those who did not complete high school
- Treatment Effect:

The Setup

Q: Outcome? Treatment? Treated? Control? Treatment Effect?

A:

- Outcome: wages
- Treatment: complete high school (educ)
- Treated: those who completed high school
- Control: those who did not complete high school
- Treatment Effect: Impact of High School education on wages

What's missing?

Q: What might be missing?

What's missing?

Q: What might be missing?

A:

- Regardless of education, some people have greater natural ability

What's missing?

Q: What might be missing?

A:

- Regardless of education, some people have greater natural ability
- If we don't have data on natural ability, we can't control for this direction of confounding

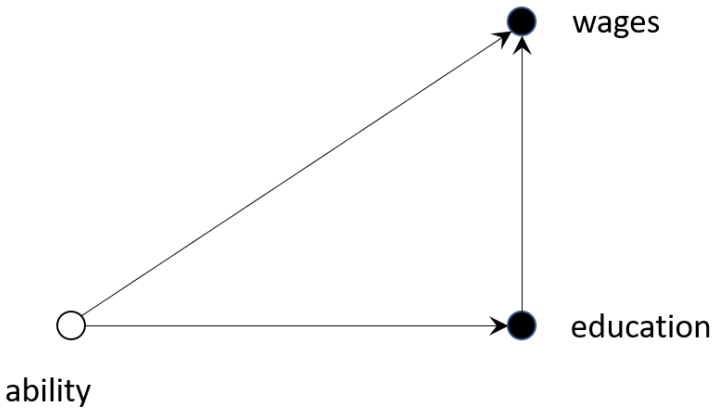
Case Study DAG

Q: Can you draw a DAG for our case study?

Case Study DAG

Q: Can you draw a DAG for our case study?

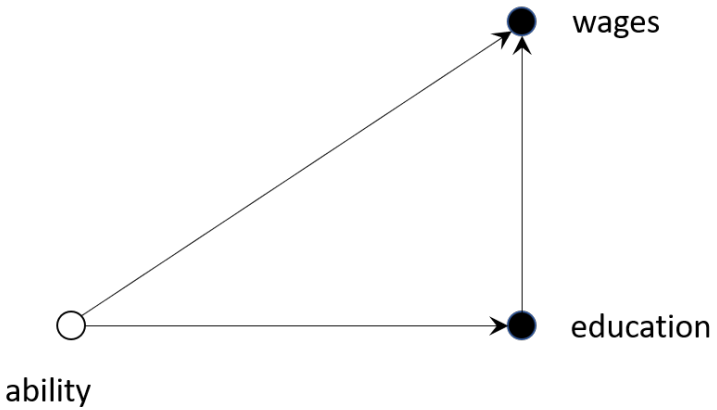
A: Here is one possibility:



Case Study DAG

Q: Can you draw a DAG for our case study?

A: Here is one possibility:

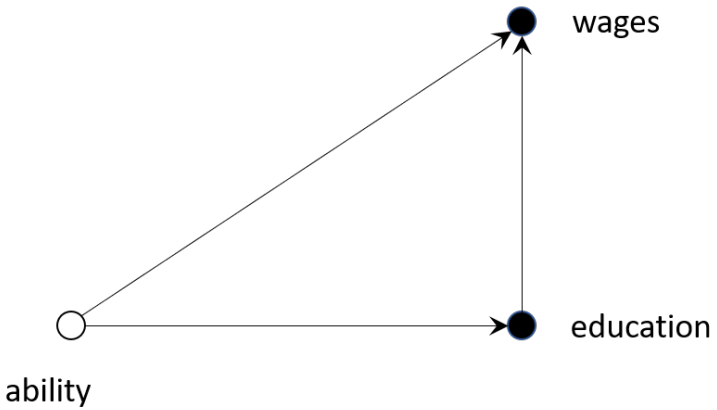


- There are likely MANY other possible DAGs, but will focus here
- "ability" is a

Case Study DAG

Q: Can you draw a DAG for our case study?

A: Here is one possibility:

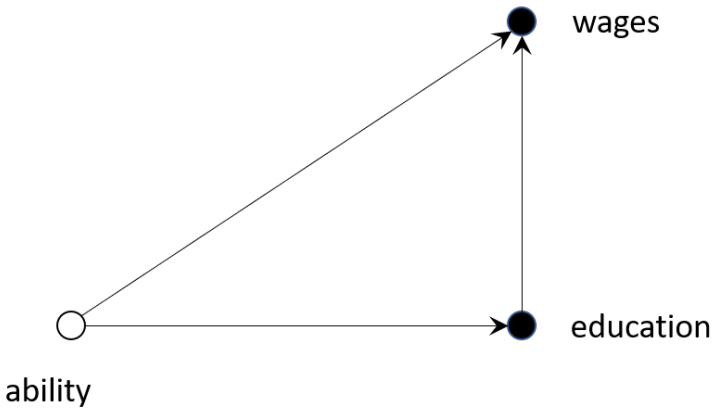


- There are likely MANY other possible DAGs, but will focus here
- "ability" is a confounder

Case Study DAG

Q: Can you draw a DAG for our case study?

A: Here is one possibility:

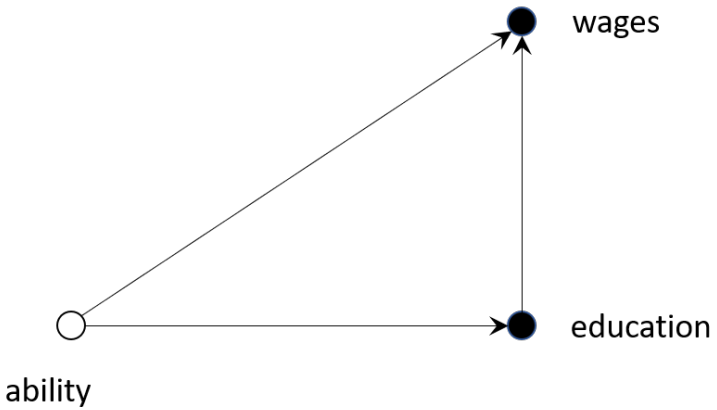


- There are likely MANY other possible DAGs, but will focus here
- "ability" is a confounder
- Notice that ability is unobserved (as indicated by the unfilled dot)

Case Study DAG

Q: Can you draw a DAG for our case study?

A: Here is one possibility:



- There are likely MANY other possible DAGs, but will focus here
- "ability" is a confounder
- Notice that ability is unobserved (as indicated by the unfilled dot)
- These greater natural abilities may impact their wages and may impact their ability to secure an education

Source of Omission

Population Regression Model: $wage = \alpha + \beta educ + \gamma abil + e$

Estimated Regression Model: $wage = \alpha + \beta educ + e$

Source of Omission

Population Regression Model: $wage = \alpha + \beta educ + \gamma abil + e$

Estimated Regression Model: $wage = \alpha + \beta educ + e$

Q: If *abil* is part of the population model (i.e. it is relevant) why would we omit it from the estimated model?

Source of Omission

Population Regression Model: $wage = \alpha + \beta educ + \gamma abil + e$

Estimated Regression Model: $wage = \alpha + \beta educ + e$

Q: If *abil* is part of the population model (i.e. it is relevant) why would we omit it from the estimated model?
A:

1. You don't know it is relevant (model mis-specification)

Source of Omission

$$\text{Population Regression Model: } wage = \alpha + \beta educ + \gamma abil + e$$

$$\text{Estimated Regression Model: } wage = \alpha + \beta educ + e$$

Q: If *abil* is part of the population model (i.e. it is relevant) why would we omit it from the estimated model?
A:

1. You don't know it is relevant (model mis-specification)
2. You can't obtain the variable because:

Source of Omission

Population Regression Model: $wage = \alpha + \beta educ + \gamma abil + e$

Estimated Regression Model: $wage = \alpha + \beta educ + e$

Q: If *abil* is part of the population model (i.e. it is relevant) why would we omit it from the estimated model?
A:

1. You don't know it is relevant (model mis-specification)
2. You can't obtain the variable because:
   - Unobserved

Source of Omission

Population Regression Model: $wage = \alpha + \beta educ + \gamma abil + e$

Estimated Regression Model: $wage = \alpha + \beta educ + e$

Q: If *abil* is part of the population model (i.e. it is relevant) why would we omit it from the estimated model?
A:

1. You don't know it is relevant (model mis-specification)
2. You can't obtain the variable because:
   - Unobserved
   - Not available

The Problem

Population Regression Model: $wage = \alpha + \beta educ + \gamma abil + e$

Estimated Regression Model: $wage = \alpha + \beta educ + v$

The Problem

$$\text{Population Regression Model: } wage = \alpha + \beta educ + \gamma abil + e$$

$$\text{Estimated Regression Model: } wage = \alpha + \beta educ + v$$

Q: What is v in our case?

The Problem

$$\text{Population Regression Model: } wage = \alpha + \beta educ + \gamma abil + e$$

$$\text{Estimated Regression Model: } wage = \alpha + \beta educ + v$$

Q: What is v in our case?
A:

- v is the error in the Estimated Regression

The Problem

$$\text{Population Regression Model: } wage = \alpha + \beta educ + \gamma abil + e$$

$$\text{Estimated Regression Model: } wage = \alpha + \beta educ + v$$

Q: What is v in our case?
A:

- v is the error in the Estimated Regression
- Errors contain the portions of the outcome variable that are not captured by the regressors

The Problem

$$\text{Population Regression Model: } wage = \alpha + \beta educ + \gamma abil + e$$

$$\text{Estimated Regression Model: } wage = \alpha + \beta educ + v$$

Q: What is v in our case?
A:

- v is the error in the Estimated Regression
- Errors contain the portions of the outcome variable that are not captured by the regressors
- In our case, the error $v = \gamma abil + e$; i.e.

The Problem

$$\text{Population Regression Model: } wage = \alpha + \beta educ + \gamma abil + e$$

$$\text{Estimated Regression Model: } wage = \alpha + \beta educ + v$$

Q: What is v in our case?
A:

- v is the error in the Estimated Regression
- Errors contain the portions of the outcome variable that are not captured by the regressors
- In our case, the error $v = \gamma abil + e$; i.e. ability's impact upon wages as well as anything else not specified in the Population Regression Model

The Problem

$$\text{Population Regression Model: } wage = \alpha + \beta educ + \gamma abil + e$$

$$\text{Estimated Regression Model: } wage = \alpha + \beta educ + v$$

Q: What is v in our case?
A:

- v is the error in the Estimated Regression
- Errors contain the portions of the outcome variable that are not captured by the regressors
- In our case, the error $v = \gamma abil + e$; i.e. ability's impact upon wages as well as anything else not specified in the Population Regression Model
- If *abil* is related to *educ*, it's omission could bias our treatment effect estimate $\hat{\beta}$

Outline

1. Case Study
2. Identifying the Bias
   - General Setting
   - Case Study
3. Measuring the Bias
   - General Setting
   - Case Study
4. OVB Sensitivity Analysis
5. "Solving" OVB

Outline

1. Case Study
2. Identifying the Bias
   - General Setting
   - Case Study
3. Measuring the Bias
   - General Setting
   - Case Study
4. OVB Sensitivity Analysis
5. "Solving" OVB

Identifying the Bias

The consequence of omitting a relevant regressor is to introduce a "bias" to our OLS estimator.

Identifying the Bias

The consequence of omitting a relevant regressor is to introduce a "bias" to our OLS estimator.
Reminder: An estimator is biased if $E[\hat{\beta}] \neq \beta^0$, where β^0 is the population value.

When we omit a relevant regressor, OLS produces biased estimators.

Reminder on OLS Assumption

$$Y = \alpha + \beta X + e$$

Reminder on OLS Assumption

$$Y = \alpha + \beta X + e$$

Assumption	**What if this assumption doesn't hold?**
$E[e_i] = 0$	If I don't expect my errors to be zero on average, I should include that information in the model
$E[e_i X_i] = 0$	(Exogeneity) If my regressors are related to my errors, then my parameter estimates are "biased"

To understand the impact of this bias, we need to think about the interpretation of the coefficient estimators from our model through the lens of these assumptions.

Interpreting the Estimators

$Y = \alpha + \beta_1 X_1 + e$

Q: How form and interpret β_1?

Interpreting the Estimators

$Y = \alpha + \beta_1 X_1 + e$

Q: How form and interpret β_1?

$$\beta_1 = \frac{\partial Y}{\partial X_1} \text{ Sensitivity of Y to changes in } X_1$$

Interpreting the Estimators

$Y = \alpha + \beta_1 X_1 + e$

Q: How form and interpret β_1?

$$\beta_1 = \frac{\partial Y}{\partial X_1} \text{ Sensitivity of Y to changes in } X_1$$

$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$

Interpreting the Estimators

$Y = \alpha + \beta_1 X_1 + e$

Q: How form and interpret β_1?

$$\beta_1 = \frac{\partial Y}{\partial X_1} \text{ Sensitivity of Y to changes in } X_1$$

$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$

Q: How form and interpret β_1 here?

Interpreting the Estimators

$Y = \alpha + \beta_1 X_1 + e$

Q: How form and interpret β_1?

$$\beta_1 = \frac{\partial Y}{\partial X_1} \text{ Sensitivity of Y to changes in } X_1$$

$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$

Q: How form and interpret β_1 here?

A:

$$\beta_1 = \frac{\partial Y}{\partial X_1}|_{X_2} \text{ Sensitivity of Y to changes in } X_1 \ldots$$

$$\ldots \text{controlling for changes in } X_2$$

Interpreting the Estimators

$Y = \alpha + \beta_1 X_1 + e$

Q: How form and interpret β_1?

$$\beta_1 = \frac{\partial Y}{\partial X_1} \text{ Sensitivity of Y to changes in } X_1$$

$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$

Q: How form and interpret β_1 here?

A:

$$\beta_1 = \frac{\partial Y}{\partial X_1}|_{X_2} \text{ Sensitivity of Y to changes in } X_1 \ldots$$

$$\ldots \text{controlling for changes in } X_2$$

i.e. the sensitivity of Y to changes in the part of X_1 that is unrelated to X_2

Interpreting the Estimators

$Y = \alpha + \beta_1 X_1 + e$

Q: How form and interpret β_1?

$$\beta_1 = \frac{\partial Y}{\partial X_1} \text{ Sensitivity of Y to changes in } X_1$$

$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$

Q: How form and interpret β_1 here?

A:

$$\beta_1 = \frac{\partial Y}{\partial X_1}|_{X_2} \text{ Sensitivity of Y to changes in } X_1 \ldots$$

$$\ldots \text{controlling for changes in } X_2$$

i.e. the sensitivity of Y to changes in the part of X_1 that is unrelated to X_2

Let's introduce an alternative interpretation of β_1 that will be helpful as we explore OVB and possible remedies.

Deeper Meaning of β_1

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$

Estimate an auxiliary regression

$$X_1 = \pi_0 + \pi_1 X_2 + u$$

Deeper Meaning of β_1

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$

Estimate an auxiliary regression

$$X_1 = \pi_0 + \pi_1 X_2 + u$$

Note: The term "auxiliary regression" is generic. It can apply to any regression we run that complements the model of interest. Later in the notes you will see $X_2 = \pi_0 + \pi_1 X_1 + u$, which is also an "auxiliary regression".

Deeper Meaning of β_1

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$

Estimate an auxiliary regression

$$X_1 = \pi_0 + \pi_1 X_2 + u$$

Note: The term "auxiliary regression" is generic. It can apply to any regression we run that complements the model of interest. Later in the notes you will see $X_2 = \pi_0 + \pi_1 X_1 + u$, which is also an "auxiliary regression".
Q: How we do interpret u?

Deeper Meaning of β_1

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$

Estimate an auxiliary regression

$$X_1 = \pi_0 + \pi_1 X_2 + u$$

Note: The term "auxiliary regression" is generic. It can apply to any regression we run that complements the model of interest. Later in the notes you will see $X_2 = \pi_0 + \pi_1 X_1 + u$, which is also an "auxiliary regression".

Q: How we do interpret u?

A: The portion of X_1 that is unrelated to X_2

Deeper Meaning of β_1

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$

Estimate an auxiliary regression

$$X_1 = \pi_0 + \pi_1 X_2 + u$$

Note: The term "auxiliary regression" is generic. It can apply to any regression we run that complements the model of interest. Later in the notes you will see $X_2 = \pi_0 + \pi_1 X_1 + u$, which is also an "auxiliary regression".

Q: How we do interpret u?

A: The portion of X_1 that is unrelated to X_2

Q: What is the relation between u and X_2?

Deeper Meaning of β_1

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$

Estimate an auxiliary regression

$$X_1 = \pi_0 + \pi_1 X_2 + u$$

Note: The term "auxiliary regression" is generic. It can apply to any regression we run that complements the model of interest. Later in the notes you will see $X_2 = \pi_0 + \pi_1 X_1 + u$, which is also an "auxiliary regression".

Q: How we do interpret u?

A: The portion of X_1 that is unrelated to X_2

Q: What is the relation between u and X_2?

A: The model assumes they are orthogonal.

Deeper Meaning of β_1

Estimate β_1 in two steps

Deeper Meaning of β_1

Estimate β_1 in two steps

$$(1) X_1 = \pi_0 + \pi_1 X_2 + u$$
$$(2) Y = \alpha + \beta \hat{u} + e$$

Deeper Meaning of β_1

Estimate β_1 in two steps

$$(1) X_1 = \pi_0 + \pi_1 X_2 + u$$
$$(2) Y = \alpha + \beta \hat{u} + e$$

The resulting $\hat{\beta} =$

Deeper Meaning of β_1

Estimate β_1 in two steps

$$\begin{aligned}(1) X_1 &= \pi_0 + \pi_1 X_2 + u \\ (2) Y &= \alpha + \beta \hat{u} + e\end{aligned}$$

The resulting $\hat{\beta} = \frac{\widehat{Cov}(Y, \hat{u})}{\widehat{Var}(\hat{u})}$.

Deeper Meaning of β_1

Estimate β_1 in two steps

$$(1) X_1 = \pi_0 + \pi_1 X_2 + u$$
$$(2) Y = \alpha + \beta \hat{u} + e$$

The resulting $\hat{\beta} = \frac{\widehat{Cov}(Y, \hat{u})}{\widehat{Var}(\hat{u})}$.

Q: How do you interpret this estimator?

Deeper Meaning of β_1

Estimate β_1 in two steps

$$(1) X_1 = \pi_0 + \pi_1 X_2 + u$$
$$(2) Y = \alpha + \beta \hat{u} + e$$

The resulting $\hat{\beta} = \frac{\widehat{Cov}(Y, \hat{u})}{\widehat{Var}(\hat{u})}$.

Q: How do you interpret this estimator?

A: The sensitivity of Y to changes in u.

Deeper Meaning of β_1

Estimate β_1 in two steps

$$(1) X_1 = \pi_0 + \pi_1 X_2 + u$$
$$(2) Y = \alpha + \beta \hat{u} + e$$

The resulting $\hat{\beta} = \frac{\widehat{Cov}(Y, \hat{u})}{\widehat{Var}(\hat{u})}$.

Q: How do you interpret this estimator?

A: The sensitivity of Y to changes in u. i.e. the sensitivity of Y to changes in the part of X_1 that is unrelated to X_2.

Deeper Meaning of β_1

Estimate β_1 in two steps

$$(1) X_1 = \pi_0 + \pi_1 X_2 + u$$
$$(2) Y = \alpha + \beta \hat{u} + e$$

The resulting $\hat{\beta} = \frac{\widehat{Cov}(Y, \hat{u})}{\widehat{Var}(\hat{u})}$.

Q: How do you interpret this estimator?

A: The sensitivity of Y to changes in u. i.e. the sensitivity of Y to changes in the part of X_1 that is unrelated to X_2.

Recall: This is the exact same as the way we interpreted $\beta_1 = \frac{\partial Y}{\partial X_1}|_{X_2}$

Deeper Meaning of β_1

Estimate β_1 in two steps

$$(1) X_1 = \pi_0 + \pi_1 X_2 + u$$
$$(2) Y = \alpha + \beta \hat{u} + e$$

The resulting $\hat{\beta} = \frac{\widehat{Cov}(Y, \hat{u})}{\widehat{Var}(\hat{u})}$.

Q: How do you interpret this estimator?

A: The sensitivity of Y to changes in u. i.e. the sensitivity of Y to changes in the part of X_1 that is unrelated to X_2.

Recall: This is the exact same as the way we interpreted $\beta_1 = \frac{\partial Y}{\partial X_1}|_{X_2}$
To find the β_1 we don't need to model all the drivers of Y. We only need the "unique" parts of X_1.

Outline

1. Case Study
2. Identifying the Bias
   - General Setting
   - Case Study
3. Measuring the Bias
   - General Setting
   - Case Study
4. OVB Sensitivity Analysis
5. "Solving" OVB

Violating Exogeneity

Back to our wage example.
True population model is

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + e$$

Violating Exogeneity

Back to our wage example.
True population model is

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + e$$

However, since ability is not observed, we estimate

$$wage = \beta_0 + \beta_1 educ + v$$

with $v = \beta_2 abil + e$.

Violating Exogeneity

Back to our wage example.
True population model is

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + e$$

However, since ability is not observed, we estimate

$$wage = \beta_0 + \beta_1 educ + v$$

with $v = \beta_2 abil + e$.

Since education is related to ability (e.g. people with higher ability tend to have higher education level), the regressor educ is related to v.

Violating Exogeneity

Back to our wage example.
True population model is

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + e$$

However, since ability is not observed, we estimate

$$wage = \beta_0 + \beta_1 educ + v$$

with $v = \beta_2 abil + e$.

Since education is related to ability (e.g. people with higher ability tend to have higher education level), the regressor educ is related to v.

So the marginal effect of educ on wage is NOT β_1. There is a bias due to omitting ability.

Outline

1. Case Study
2. Identifying the Bias
   - General Setting
   - Case Study
3. Measuring the Bias
   - General Setting
   - Case Study
4. OVB Sensitivity Analysis
5. "Solving" OVB

Outline

1. Case Study
2. Identifying the Bias
   - General Setting
   - Case Study
3. Measuring the Bias
   - General Setting
   - Case Study
4. OVB Sensitivity Analysis
5. "Solving" OVB

Measuring the OVB

CAUTION:

- We CANNOT measure the magnitude of the bias in practice.

Measuring the OVB

CAUTION:

- We CANNOT measure the magnitude of the bias in practice.
- We can only get a sense of the direction (i.e. sign) of the bias.

Measuring the OVB

CAUTION:

- We CANNOT measure the magnitude of the bias in practice.
- We can only get a sense of the direction (i.e. sign) of the bias.
- The following investigation is meant as a pedagogical tool to develop a method for assessing the sign of the bias. It CANNOT be reproduced in practice.

Setting up the Problem

Let's use the notation from the Mastering Metrics text.
Long:

Setting up the Problem

Let's use the notation from the Mastering Metrics text.
Long: $Y = \alpha^{l} + \beta^{l} X_1 + \gamma^{l} X_2 + e^{l}$

Setting up the Problem

Let's use the notation from the Mastering Metrics text.
Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short:

Setting up the Problem

Let's use the notation from the Mastering Metrics text.
Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$

Setting up the Problem

Let's use the notation from the Mastering Metrics text.
Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
We want β^l, but we can only attain β^s. We will show that β^s is actually β^l plus a bias.

Setting up the Problem

Let's use the notation from the Mastering Metrics text.
Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
We want β^l, but we can only attain β^s. We will show that β^s is actually β^l plus a bias.

$$\hat{\beta}^s = \frac{\widehat{Cov}(Y, X_1)}{\widehat{Var}(X_1)}$$

Setting up the Problem

Let's use the notation from the Mastering Metrics text.
Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
We want β^l, but we can only attain β^s. We will show that β^s is actually β^l plus a bias.

$$\begin{aligned}\hat{\beta}^s &= \frac{\widehat{Cov}(Y, X_1)}{\widehat{Var}(X_1)} \\ &= \frac{\widehat{Cov}(\hat{\alpha}^l + \hat{\beta}^l X_1 + \hat{\gamma} X_2 + \hat{e}^l, X_1)}{\widehat{Var}(X_1)}\end{aligned}$$

Setting up the Problem

Let's use the notation from the Mastering Metrics text.
Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
We want β^l, but we can only attain β^s. We will show that β^s is actually β^l plus a bias.

$$\begin{aligned}
\hat{\beta}^s &= \frac{\widehat{Cov}(Y, X_1)}{\widehat{Var}(X_1)} \\
&= \frac{\widehat{Cov}(\hat{\alpha}^l + \hat{\beta}^l X_1 + \hat{\gamma} X_2 + \hat{e}^l, X_1)}{\widehat{Var}(X_1)} \\
&= \frac{\widehat{Cov}(\hat{\alpha}^l, X_1) + \widehat{Cov}(\hat{\beta}^l X_1, X_1) + \widehat{Cov}(\hat{\gamma}^l X_2, X_1) + \widehat{Cov}(\hat{e}^l, X_1)}{\widehat{Var}(X_1)}
\end{aligned}$$

Setting up the Problem

Let's use the notation from the Mastering Metrics text.
Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
We want β^l, but we can only attain β^s. We will show that β^s is actually β^l plus a bias.

$$
\begin{aligned}
\hat{\beta}^s &= \frac{\widehat{Cov}(Y, X_1)}{\widehat{Var}(X_1)} \\
&= \frac{\widehat{Cov}(\hat{\alpha}^l + \hat{\beta}^l X_1 + \hat{\gamma} X_2 + \hat{e}^l, X_1)}{\widehat{Var}(X_1)} \\
&= \frac{\widehat{Cov}(\hat{\alpha}^l, X_1) + \widehat{Cov}(\hat{\beta}^l X_1, X_1) + \widehat{Cov}(\hat{\gamma}^l X_2, X_1) + \widehat{Cov}(\hat{e}^l, X_1)}{\widehat{Var}(X_1)} \\
&= \frac{\hat{\beta}^l \widehat{Var}(X_1) + \hat{\gamma}^l \hat{Cov}(X_2, X_1)}{\widehat{Var}(X_1)} =
\end{aligned}
$$

Setting up the Problem

Let's use the notation from the Mastering Metrics text.
Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
We want β^l, but we can only attain β^s. We will show that β^s is actually β^l plus a bias.

$$\begin{aligned}
\hat{\beta}^s &= \frac{\widehat{Cov}(Y, X_1)}{\widehat{Var}(X_1)} \\
&= \frac{\widehat{Cov}(\hat{\alpha}^l + \hat{\beta}^l X_1 + \hat{\gamma} X_2 + \hat{e}^l, X_1)}{\widehat{Var}(X_1)} \\
&= \frac{\widehat{Cov}(\hat{\alpha}^l, X_1) + \widehat{Cov}(\hat{\beta}^l X_1, X_1) + \widehat{Cov}(\hat{\gamma}^l X_2, X_1) + \widehat{Cov}(\hat{e}^l, X_1)}{\widehat{Var}(X_1)} \\
&= \frac{\hat{\beta}^l \widehat{Var}(X_1) + \hat{\gamma}^l \hat{Cov}(X_2, X_1)}{\widehat{Var}(X_1)} = \hat{\beta}^l + \hat{\gamma}^l \frac{\widehat{Cov}(X_2, X_1)}{\widehat{Var}(X_1)}
\end{aligned}$$

Measuring the OVB

Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$

$$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \frac{\widehat{Cov}(X_2, X_1)}{\widehat{Var}(X_1)}$$

Measuring the OVB

Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$

$$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \frac{\widehat{Cov}(X_2, X_1)}{\widehat{Var}(X_1)}$$

Let's consolidate for ease of interpretation

Measuring the OVB

Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$

$$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \frac{\widehat{Cov}(X_2, X_1)}{\widehat{Var}(X_1)}$$

Let's consolidate for ease of interpretation

$$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \hat{\pi}_1$$

Measuring the OVB

Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$

$$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \frac{\widehat{Cov}(X_2, X_1)}{\widehat{Var}(X_1)}$$

Let's consolidate for ease of interpretation

$$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \hat{\pi}_1$$

where $X_2 = \pi_0 + \pi_1 X_1 + u$, $\hat{\pi}_1 = \frac{\widehat{Cov}(X_2, X_1)}{\widehat{Var}(X_1)}$
Notice that this is another "auxiliary regression", distinct from one seen earlier, used here for a different purpose.

Direction of the OVB

If domain experts can tell us the signs of γ' and π_1, we can guess the direction of the OVB

Direction of the OVB

If domain experts can tell us the signs of γ^l and π_1, we can guess the direction of the OVB

Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \hat{\pi}_1$

	$\pi_1 > 0$	$\pi_1 < 0$
$\gamma^l > 0$		
$\gamma^l < 0$		

Table: Summary of Bias in β^s

Direction of the OVB

If domain experts can tell us the signs of γ^l and π_1, we can guess the direction of the OVB

Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \hat{\pi}_1$

	$\pi_1 > 0$	$\pi_1 < 0$
$\gamma^l > 0$	Positive Bias	
$\gamma^l < 0$		

Table: Summary of Bias in β^s

Direction of the OVB

If domain experts can tell us the signs of γ^l and π_1, we can guess the direction of the OVB

Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \hat{\pi}_1$

	$\pi_1 > 0$	$\pi_1 < 0$
$\gamma^l > 0$	Positive Bias	Negative Bias
$\gamma^l < 0$		

Table: Summary of Bias in β^s

Direction of the OVB

If domain experts can tell us the signs of γ^l and π_1, we can guess the direction of the OVB

Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \hat{\pi}_1$

	$\pi_1 > 0$	$\pi_1 < 0$
$\gamma^l > 0$	Positive Bias	Negative Bias
$\gamma^l < 0$	Negative Bias	

Table: Summary of Bias in β^s

Direction of the OVB

If domain experts can tell us the signs of γ^l and π_1, we can guess the direction of the OVB

Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \hat{\pi}_1$

	$\pi_1 > 0$	$\pi_1 < 0$
$\gamma^l > 0$	Positive Bias	Negative Bias
$\gamma^l < 0$	Negative Bias	Positive Bias

Table: Summary of Bias in β^s

Direction of the OVB

If domain experts can tell us the signs of γ^l and π_1, we can guess the direction of the OVB

Long: $Y = \alpha^l + \beta^l X_1 + \gamma^l X_2 + e^l$
Short: $Y = \alpha^s + \beta^s X_1 + e^s$
$\hat{\beta}^s = \hat{\beta}^l + \hat{\gamma}^l \hat{\pi}_1$

	$\pi_1 > 0$	$\pi_1 < 0$
$\gamma^l > 0$	Positive Bias	Negative Bias
$\gamma^l < 0$	Negative Bias	Positive Bias

Table: Summary of Bias in β^s

Note that if the bias is positive then we can say that β^s is OVER-estimating the true marginal impact of X_1 upon Y. (Analogously, a negative bias UNDER-estimates).

Outline

1. Case Study
2. Identifying the Bias
   - General Setting
   - Case Study
3. Measuring the Bias
   - General Setting
   - Case Study
4. OVB Sensitivity Analysis
5. "Solving" OVB

Identifying the Long and Short Regressions

Q: Can you write the "long" and "short" regressions in our education and wage case study?

Identifying the Long and Short Regressions

Q: Can you write the "long" and "short" regressions in our education and wage case study?

A:

$$\text{(Long) } \textit{wage} =$$
$$\text{(Short) } \textit{wage} =$$

Identifying the Long and Short Regressions

Q: Can you write the "long" and "short" regressions in our education and wage case study?

A:

$$\text{(Long) } wage = \alpha^l + \beta^l educ + \gamma^l abil + e^l$$

$$\text{(Short) } wage =$$

Identifying the Long and Short Regressions

Q: Can you write the "long" and "short" regressions in our education and wage case study?

A:

$$\text{(Long) } wage = \alpha^l + \beta^l educ + \gamma^l abil + e^l$$
$$\text{(Short) } wage = \alpha^s + \beta^s educ + e^s$$

Identifying the Auxiliary Regression

Q: Can you write the auxiliary reg. that will produce π_1 in our case study?

Identifying the Auxiliary Regression

Q: Can you write the auxiliary reg. that will produce π_1 in our case study?
A:

$$abil = \pi_0 + \pi_1 educ + u$$

Identifying the Auxiliary Regression

Q: Can you write the auxiliary reg. that will produce π_1 in our case study?
A:

$$abil = \pi_0 + \pi_1 educ + u$$

Q: What is the meaning of π_1 in our case study where *educ* is a binary (e.g. graduated high school or not), and define graduating high school as our "treatment"? What does a large π_1 suggest about our efforts to estimate a treatment effect?

Identifying the Auxiliary Regression

Q: Can you write the auxiliary reg. that will produce π_1 in our case study?
A:

$$abil = \pi_0 + \pi_1 educ + u$$

Q: What is the meaning of π_1 in our case study where *educ* is a binary (e.g. graduated high school or not), and define graduating high school as our "treatment"? What does a large π_1 suggest about our efforts to estimate a treatment effect?
A:

$$\pi_1 = E[abil|educ = 1] - E[abil|educ = 0]$$

Identifying the Auxiliary Regression

Q: Can you write the auxiliary reg. that will produce π_1 in our case study?
A:

$$abil = \pi_0 + \pi_1 educ + u$$

Q: What is the meaning of π_1 in our case study where *educ* is a binary (e.g. graduated high school or not), and define graduating high school as our "treatment"? What does a large π_1 suggest about our efforts to estimate a treatment effect?
A:

$$\pi_1 = E[abil|educ = 1] - E[abil|educ = 0]$$

In this setting, π_1 captures the imbalance of the covariate *abil* across our treated and control samples. A large π_1 indicates large covariate imbalance, making it more important to control for this confounder.

Identifying the Bias

$$\beta^s = \beta^l + \gamma^l \pi_1$$

Identifying the Bias

$$\beta^s = \beta^l + \gamma^l \pi_1$$

Q: What do you suspect to be the sign of the bias, and its meaning in our case study?

Identifying the Bias

$$\beta^s = \beta^l + \gamma^l \pi_1$$

Q: What do you suspect to be the sign of the bias, and its meaning in our case study?

A:

- $\gamma^l > 0$:

Identifying the Bias

$$\beta^s = \beta^l + \gamma^l \pi_1$$

Q: What do you suspect to be the sign of the bias, and its meaning in our case study?
A:

- $\gamma^l > 0$: higher natural ability generates higher wages

Identifying the Bias

$$\beta^s = \beta^l + \gamma^l \pi_1$$

Q: What do you suspect to be the sign of the bias, and its meaning in our case study?
A:

- $\gamma^l > 0$: higher natural ability generates higher wages
- $\pi_1 > 0$: $E[abil|educ = 1] - E[abil|educ = 0] = \pi_1$; i.e.

Identifying the Bias

$$\beta^s = \beta^l + \gamma^l \pi_1$$

Q: What do you suspect to be the sign of the bias, and its meaning in our case study?
A:

- $\gamma^l > 0$: higher natural ability generates higher wages
- $\pi_1 > 0$: $E[abil|educ = 1] - E[abil|educ = 0] = \pi_1$; i.e. I suspect that those who graduated high school likely have higher natural abilities than those who didn't

Identifying the Bias

$$\beta^s = \beta^l + \gamma^l \pi_1$$

Q: What do you suspect to be the sign of the bias, and its meaning in our case study?
A:

- $\gamma^l > 0$: higher natural ability generates higher wages
- $\pi_1 > 0$: $E[abil|educ = 1] - E[abil|educ = 0] = \pi_1$; i.e. I suspect that those who graduated high school likely have higher natural abilities than those who didn't
- Bias > 0: Implying that β^s is likely overstating the true treatment effect of education on wages

Outline

1. Case Study
2. Identifying the Bias
   - General Setting
   - Case Study
3. Measuring the Bias
   - General Setting
   - Case Study
4. OVB Sensitivity Analysis
5. "Solving" OVB

Framework for OVB Sensitivity Analysis

- We cannot estimate the size of OVB
- But we can conduct sensitivity analysis
- The following is adapted upon Cinelli and Hazlett (2020)

How much bias is relevant?

Recall the policymaker can pass the legislation if he can say the impact of the high school program is at least $\$3,000$ (denote as 3)



How much bias is relevant?

Recall the policymaker can pass the legislation if he can say the impact of the high school program is at least $\$3,000$ (denote as 3)



- Given your model
  $wages = \alpha + \beta educ + v$

How much bias is relevant?

Recall the policymaker can pass the legislation if he can say the impact of the high school program is at least $\$3,000$ (denote as 3)



- Given your model
  $wages = \alpha + \beta educ + v$
- You estimate $\hat{\beta} = 4$

How much bias is relevant?

Recall the policymaker can pass the legislation if he can say the impact of the high school program is at least $\$3,000$ (denote as 3)



- Given your model
  $wages = \alpha + \beta educ + v$
- You estimate $\hat{\beta} = 4$
- You know this is likely

How much bias is relevant?

Recall the policymaker can pass the legislation if he can say the impact of the high school program is at least $\$3,000$ (denote as 3)



- Given your model
  $wages = \alpha + \beta educ + v$
- You estimate $\hat{\beta} = 4$
- You know this is likely overstated given $\pi_1 \gamma > 0$

How much bias is relevant?

Recall the policymaker can pass the legislation if he can say the impact of the high school program is at least $\$3,000$ (denote as 3)



- Given your model
  $wages = \alpha + \beta educ + v$
- You estimate $\hat{\beta} = 4$
- You know this is likely overstated given $\pi_1 \gamma > 0$
- Q: How much bias is "relevant" to the policy maker?

How much bias is relevant?

Recall the policymaker can pass the legislation if he can say the impact of the high school program is at least $\$3,000$ (denote as 3)



- Given your model
  $wages = \alpha + \beta educ + v$
- You estimate $\hat{\beta} = 4$
- You know this is likely overstated given $\pi_1 \gamma > 0$
- Q: How much bias is "relevant" to the policy maker?
- A: If the bias is 1 or more, the policymaker can't pass the legislation

Bounding the Relevant Bias

We know

$$\beta^s = \beta^l + \gamma\pi_1$$

Bounding the Relevant Bias

We know

$$\beta^s = \beta^l + \gamma\pi_1$$

Rewrite slightly

$$\beta^l = \beta^s - \gamma\pi_1$$

Bounding the Relevant Bias

We know

$$\beta^s = \beta^l + \gamma\pi_1$$

Rewrite slightly

$$\beta^l = \beta^s - \gamma\pi_1$$

The client can take action if the treatment effect ≥ 3

$$\beta^l = \beta^s - \pi_1\gamma^l \geq 3$$

Bounding the Relevant Bias

We know

$$\beta^s = \beta^l + \gamma\pi_1$$

Rewrite slightly

$$\beta^l = \beta^s - \gamma\pi_1$$

The client can take action if the treatment effect ≥ 3

$$\beta^l = \beta^s - \pi_1\gamma^l \geq 3$$

Given our estimate of 4, we can put a bound on the relevant bias

Bounding the Relevant Bias

We know

$$\beta^s = \beta^l + \gamma\pi_1$$

Rewrite slightly

$$\beta^l = \beta^s - \gamma\pi_1$$

The client can take action if the treatment effect ≥ 3

$$\beta^l = \beta^s - \pi_1\gamma^l \geq 3$$

Given our estimate of 4, we can put a bound on the relevant bias

$$\pi_1\gamma^l \leq 1$$

Bounding the Relevant Bias

We know

$$\beta^s = \beta^l + \gamma\pi_1$$

Rewrite slightly

$$\beta^l = \beta^s - \gamma\pi_1$$

The client can take action if the treatment effect ≥ 3

$$\beta^l = \beta^s - \pi_1\gamma^l \geq 3$$

Given our estimate of 4, we can put a bound on the relevant bias

$$\pi_1\gamma^l \leq 1$$

If the bias is less than 1, the client can take action.

Generating the Relevant Bias

Here is one way the relevant bias of 1 could manifest

Generating the Relevant Bias

Here is one way the relevant bias of 1 could manifest

Q: What is the meaning of $\pi_1 = 2$ and $\gamma = .5$ ($\pi_1\gamma = 2 \times .5 = 1$)?

Generating the Relevant Bias

Here is one way the relevant bias of 1 could manifest

Q: What is the meaning of $\pi_1 = 2$ and $\gamma = .5$ ($\pi_1\gamma = 2 \times .5 = 1$)?
A:

- $\pi_1 = 2$:

Generating the Relevant Bias

Here is one way the relevant bias of 1 could manifest

Q: What is the meaning of $\pi_1 = 2$ and $\gamma = .5$ ($\pi_1\gamma = 2 \times .5 = 1$)?
A:

- $\pi_1 = 2$: Ability of the high school graduates is two units higher on average than that of the non high school graduates.

Generating the Relevant Bias

Here is one way the relevant bias of 1 could manifest

Q: What is the meaning of $\pi_1 = 2$ and $\gamma = .5$ ($\pi_1\gamma = 2 \times .5 = 1$)?
A:

- $\pi_1 = 2$: Ability of the high school graduates is two units higher on average than that of the non high school graduates.
- $\gamma = .5$:

Generating the Relevant Bias

Here is one way the relevant bias of 1 could manifest

Q: What is the meaning of $\pi_1 = 2$ and $\gamma = .5$ ($\pi_1\gamma = 2 \times .5 = 1$)?
A:

- $\pi_1 = 2$: Ability of the high school graduates is two units higher on average than that of the non high school graduates.
- $\gamma = .5$: A one unit increase in ability would increase wages by \$500 (recall we are measuring in thousands of dollars)

Generating the Relevant Bias

Here is one way the relevant bias of 1 could manifest

Q: What is the meaning of $\pi_1 = 2$ and $\gamma = .5$ ($\pi_1\gamma = 2 \times .5 = 1$)?
A:

- $\pi_1 = 2$: Ability of the high school graduates is two units higher on average than that of the non high school graduates.
- $\gamma = .5$: A one unit increase in ability would increase wages by \$500 (recall we are measuring in thousands of dollars)

Anything greater than these values would invalidate the policymaker's proposed legislation.

Generating the Relevant Bias

Here is one way the relevant bias of 1 could manifest

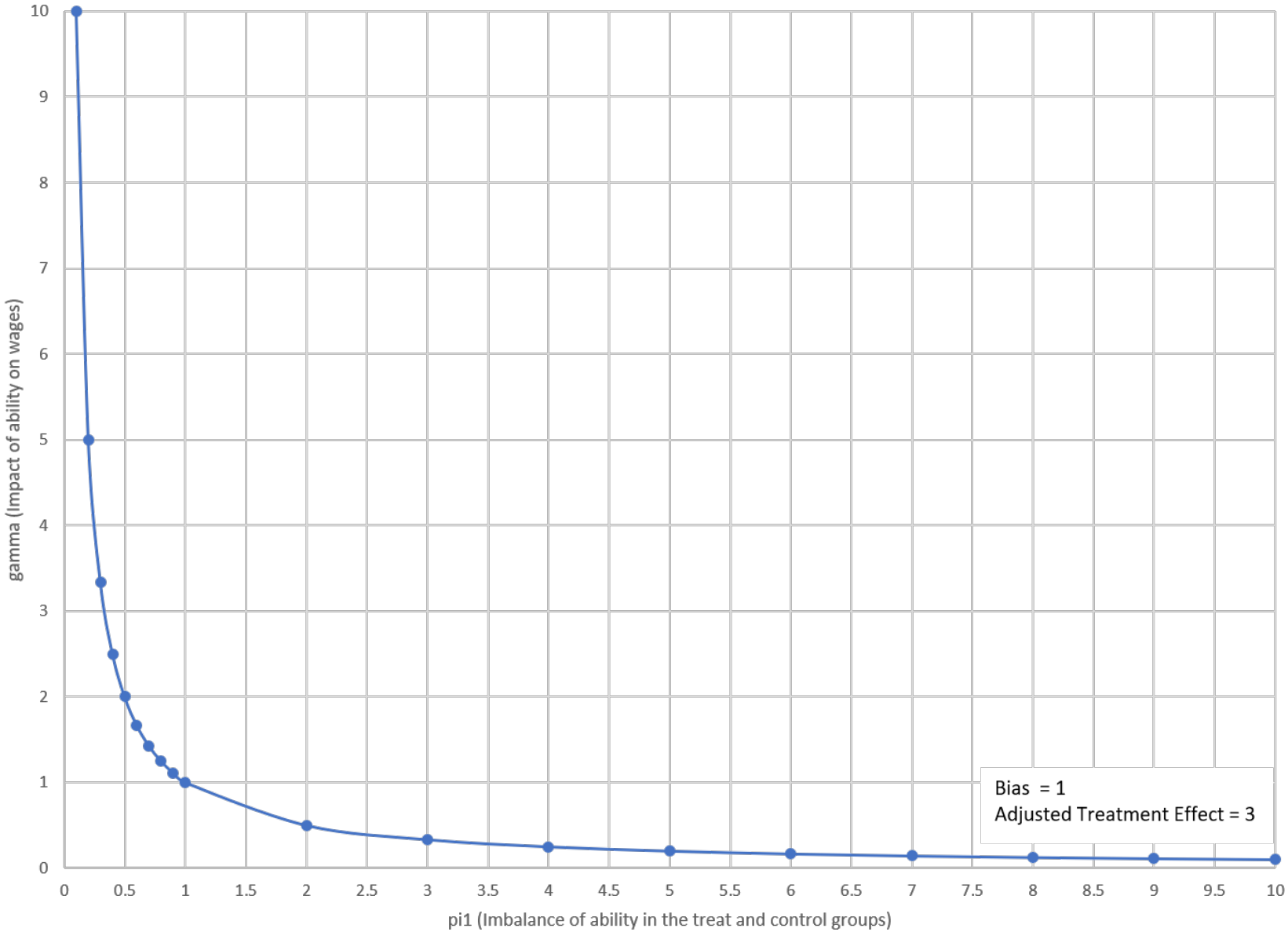
Q: What is the meaning of $\pi_1 = 2$ and $\gamma = .5$ ($\pi_1\gamma = 2 \times .5 = 1$)?
A:

- $\pi_1 = 2$: Ability of the high school graduates is two units higher on average than that of the non high school graduates.
- $\gamma = .5$: A one unit increase in ability would increase wages by \$500 (recall we are measuring in thousands of dollars)

Anything greater than these values would invalidate the policymaker's proposed legislation.We don't know these values. We can probe subject matter experts and explore other complimentary analyses to make an educated guess.

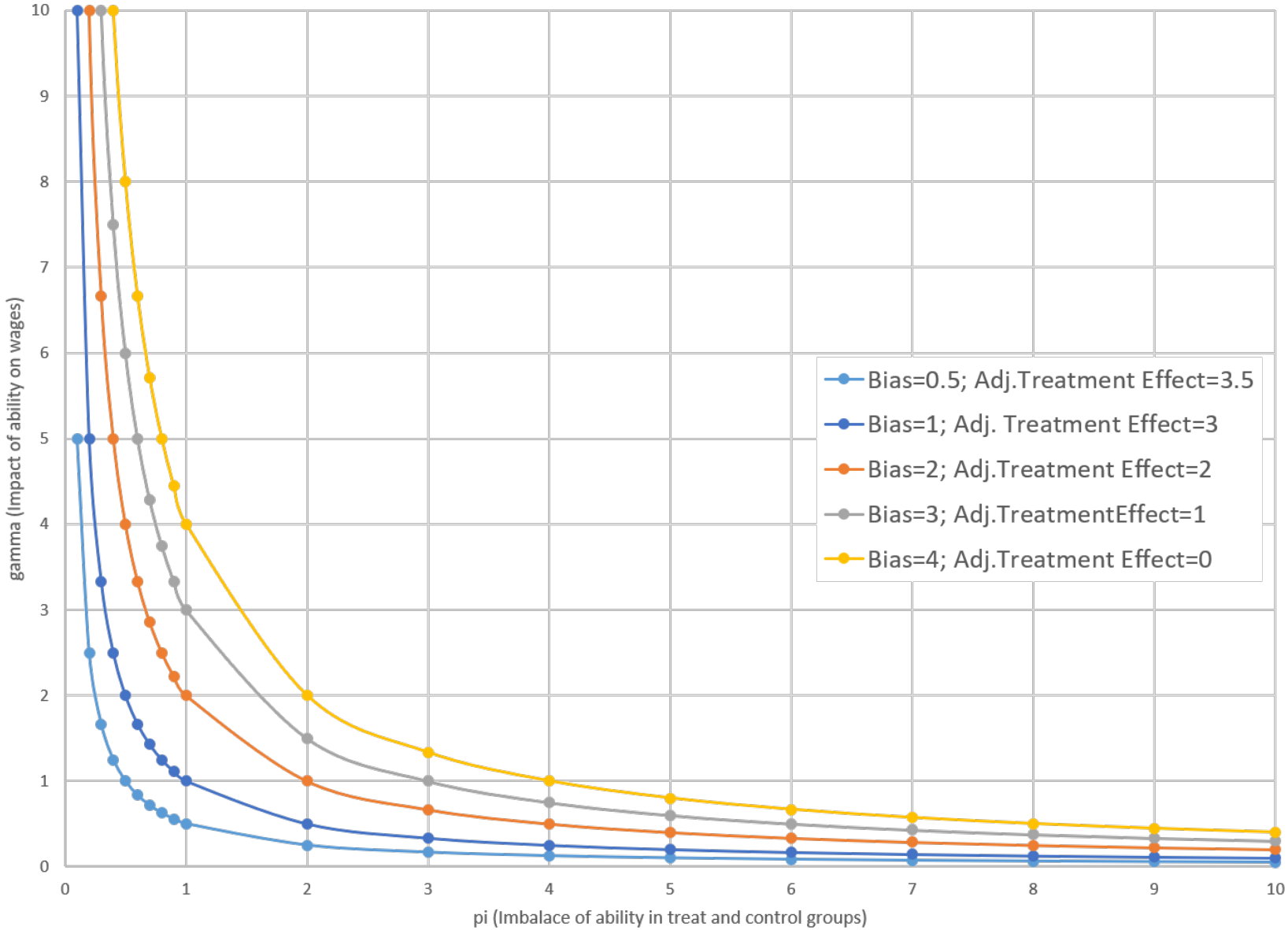
Contour for Bias=1

Myriad ways in which a bias of 1 could be generated



Contours for a range of Bias

We can generate a sequence of these surface plots for various scenarios

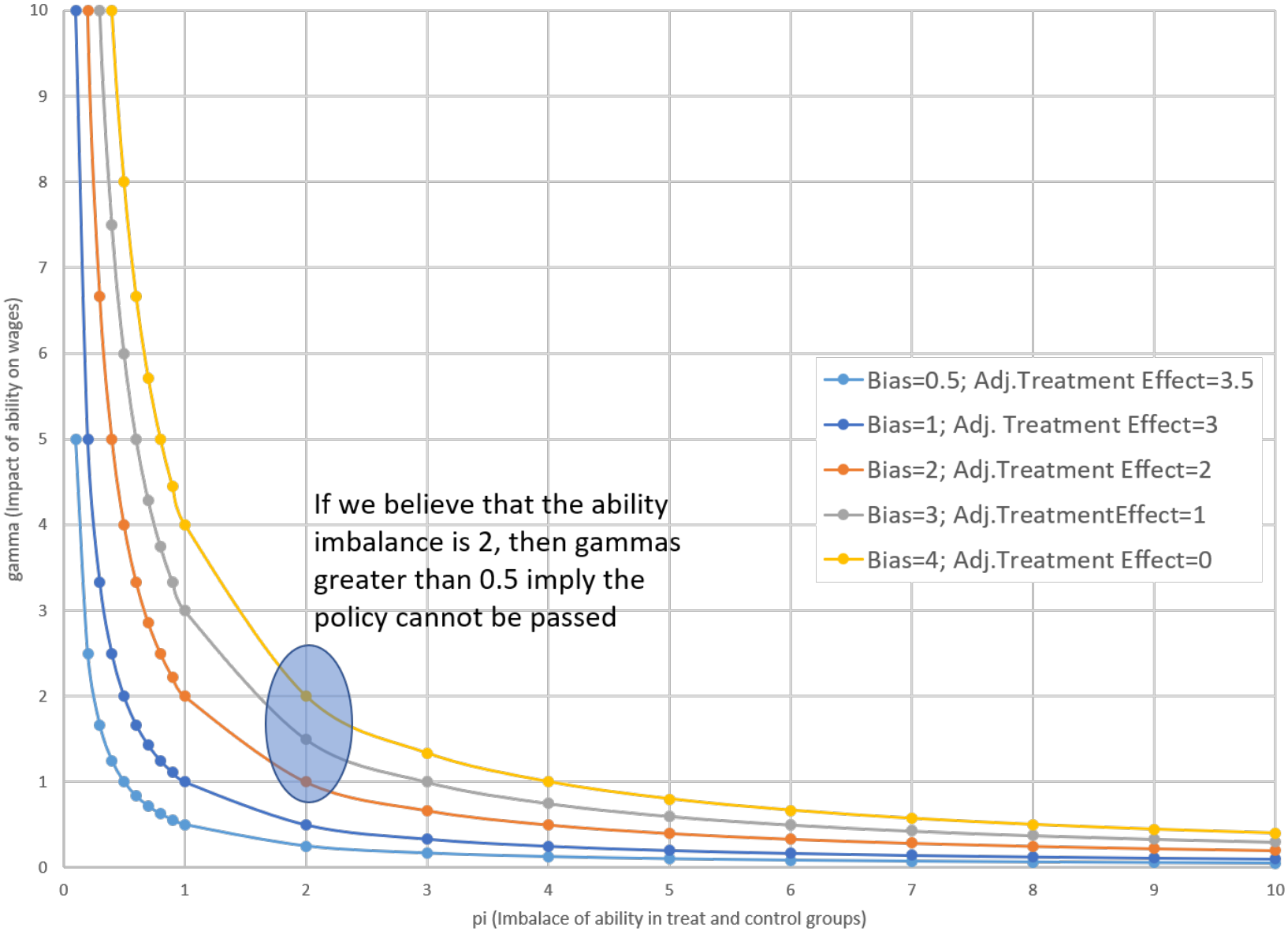


Putting a range on the source of relevant bias

Q: If we believe $\pi_1 = 2$, what range of γ is acceptable to the stakeholder?

Putting a range on the source of relevant bias

Q: If we believe $\pi_1 = 2$, what range of γ is acceptable to the stakeholder?



Weaknesses

Q: What are some weaknesses of this approach?

Weaknesses

Q: What are some weaknesses of this approach?

A:

- The stakeholders/subject matter experts may not have a sense of the potential size of π_1 nor γ

Weaknesses

Q: What are some weaknesses of this approach?

A:

- The stakeholders/subject matter experts may not have a sense of the potential size of π_1 nor γ
- Omitted variables often don't have well defined units of measurement. In our case, the units of ability are not well defined.

Weaknesses

Q: What are some weaknesses of this approach?

A:

- The stakeholders/subject matter experts may not have a sense of the potential size of π_1 nor γ
- Omitted variables often don't have well defined units of measurement. In our case, the units of ability are not well defined.
- Not as intuitive if continuous treatment
- Can't accommodate multiple omitted confounders that interact

Weaknesses

Q: What are some weaknesses of this approach?

A:

- The stakeholders/subject matter experts may not have a sense of the potential size of π_1 nor γ
- Omitted variables often don't have well defined units of measurement. In our case, the units of ability are not well defined.
- Not as intuitive if continuous treatment
- Can't accommodate multiple omitted confounders that interact

Recasting with partial R^2

Cinelli and Hazlett (2020) attempt to overcome (some of) these weaknesses by recasting the sensitivity analysis away from coefficient values towards partial R^2, by creating RV and RV_α. See their paper for derivations. Here we will focus on the applications.

- RV stands for Residual Variation, which is a transformed partial R^2
- We can think of residual variation as the amount of variation in the outcome (and treatment) variable that is not explained by the observed confounders

Reminder on Partial R^2

$$R^2_{Y,X_2|X_1} = \frac{SSE(X_1) - SSE(X_1, X_2)}{SSE(X_1)}$$

where SSE is the sum of squared errors $\sum(\hat{Y}_i - Y_i)^2$

Reminder on Partial R^2

$$R^2_{Y,X_2|X_1} = \frac{SSE(X_1) - SSE(X_1, X_2)}{SSE(X_1)}$$

where SSE is the sum of squared errors $\sum(\hat{Y}_i - Y_i)^2$ In our case study we might consider

$$R^2_{wage,abil|educ} = \frac{SSE(educ) - SSE(educ, abil)}{SSE(educ)}$$

Reminder on Partial R^2

$$R^2_{Y,X_2|X_1} = \frac{SSE(X_1) - SSE(X_1, X_2)}{SSE(X_1)}$$

where SSE is the sum of squared errors $\sum(\hat{Y}_i - Y_i)^2$ In our case study we might consider

$$R^2_{wage,abil|educ} = \frac{SSE(educ) - SSE(educ, abil)}{SSE(educ)}$$

i.e. the part of the variation in *wage* explained by *abil* that is not explained by *educ*

Applying RV to our case

You estimated $\hat{\beta}^s = 4$. Assume accompanying $RV = 20\%$ and $RV_{.05} = 12\%$

Applying RV to our case

You estimated $\hat{\beta}^s = 4$. Assume accompanying $RV = 20\%$ and $RV_{.05} = 12\%$

- $RV = 20\%$ implies that all of the unobserved confounders (including ability) would need to explain 20% or more of the remaining variability (aka Residual Variation; RV) in both wages and educ in order to bring the treatment effect of $\hat{\beta}^s = 4$ down to zero.

Applying RV to our case

You estimated $\hat{\beta}^s = 4$. Assume accompanying $RV = 20\%$ and $RV_{.05} = 12\%$

- $RV = 20\%$ implies that all of the unobserved confounders (including ability) would need to explain 20% or more of the remaining variability (aka Residual Variation; RV) in both wages and educ in order to bring the treatment effect of $\hat{\beta}^s = 4$ down to zero.
  - Q: What if $RV = 95\%$? Does that give you more / less confidence in your estimates?

Applying RV to our case

You estimated $\hat{\beta}^s = 4$. Assume accompanying $RV = 20\%$ and $RV_{.05} = 12\%$

- $RV = 20\%$ implies that all of the unobserved confounders (including ability) would need to explain 20% or more of the remaining variability (aka Residual Variation; RV) in both wages and educ in order to bring the treatment effect of $\hat{\beta}^s = 4$ down to zero.
  - Q: What if $RV = 95\%$? Does that give you more / less confidence in your estimates?
  - A: More. If $RV = 95\%$ that implies it would require what's missing in our model to explain a lot of what we're observing.

Applying RV to our case

You estimated $\hat{\beta}^s = 4$. Assume accompanying $RV = 20\%$ and $RV_{.05} = 12\%$

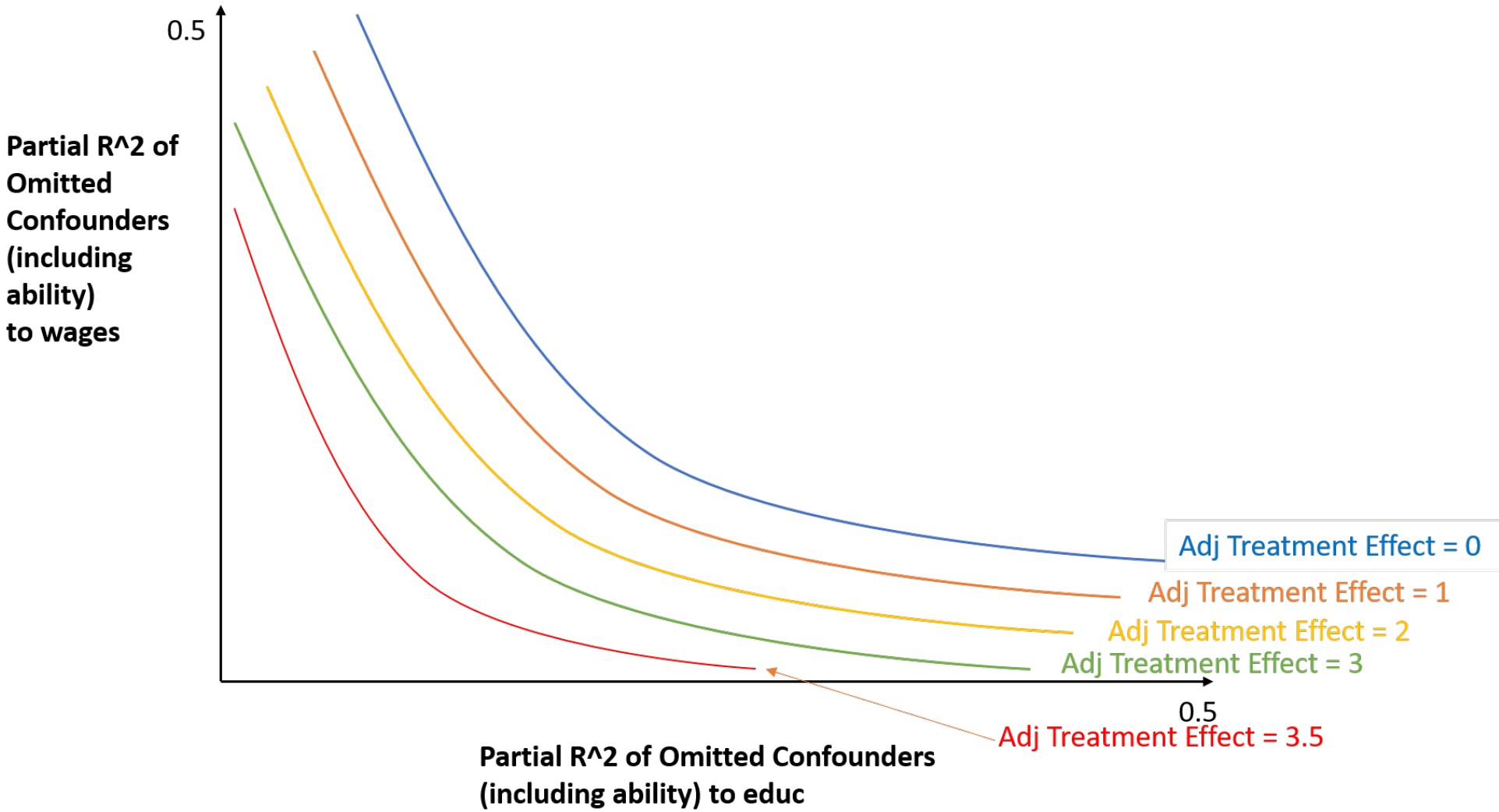
- $RV = 20\%$ implies that all of the unobserved confounders (including ability) would need to explain 20% or more of the remaining variability (aka Residual Variation; RV) in both wages and educ in order to bring the treatment effect of $\hat{\beta}^s = 4$ down to zero.
  - Q: What if $RV = 95\%$? Does that give you more / less confidence in your estimates?
  - A: More. If $RV = 95\%$ that implies it would require what's missing in our model to explain a lot of what we're observing. Higher RV show more robustness to OVB.

Applying RV to our case

You estimated $\hat{\beta}^s = 4$. Assume accompanying $RV = 20\%$ and $RV_{.05} = 12\%$

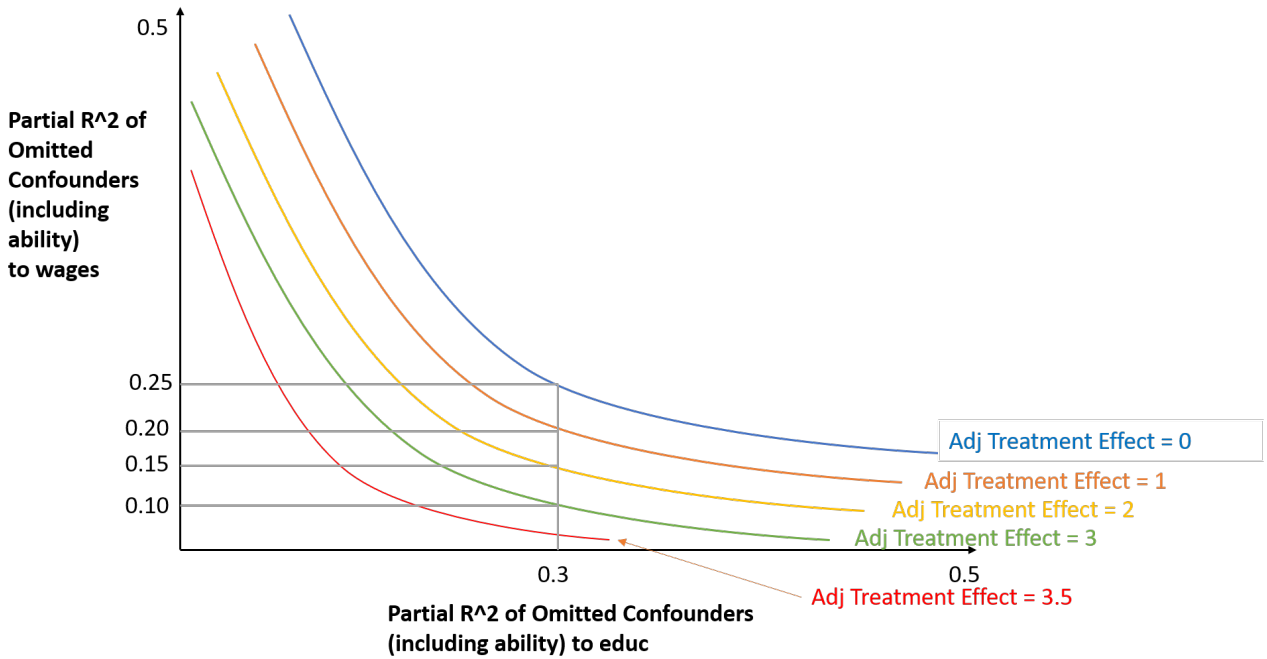
- $RV = 20\%$ implies that all of the unobserved confounders (including ability) would need to explain 20% or more of the remaining variability (aka Residual Variation; RV) in both wages and educ in order to bring the treatment effect of $\hat{\beta}^s = 4$ down to zero.
  - Q: What if $RV = 95\%$? Does that give you more / less confidence in your estimates?
  - A: More. If $RV = 95\%$ that implies it would require what's missing in our model to explain a lot of what we're observing. Higher RV show more robustness to OVB.
- $RV_{0.05} = 12\%$ implies that at 5% significance, the unobserved confounders would need to explain only 12% of the residual variation in both wages and educ in order to make our estimate of $\hat{\beta}^s = 4$ statistically insignificant.

Partial R^2 Contours



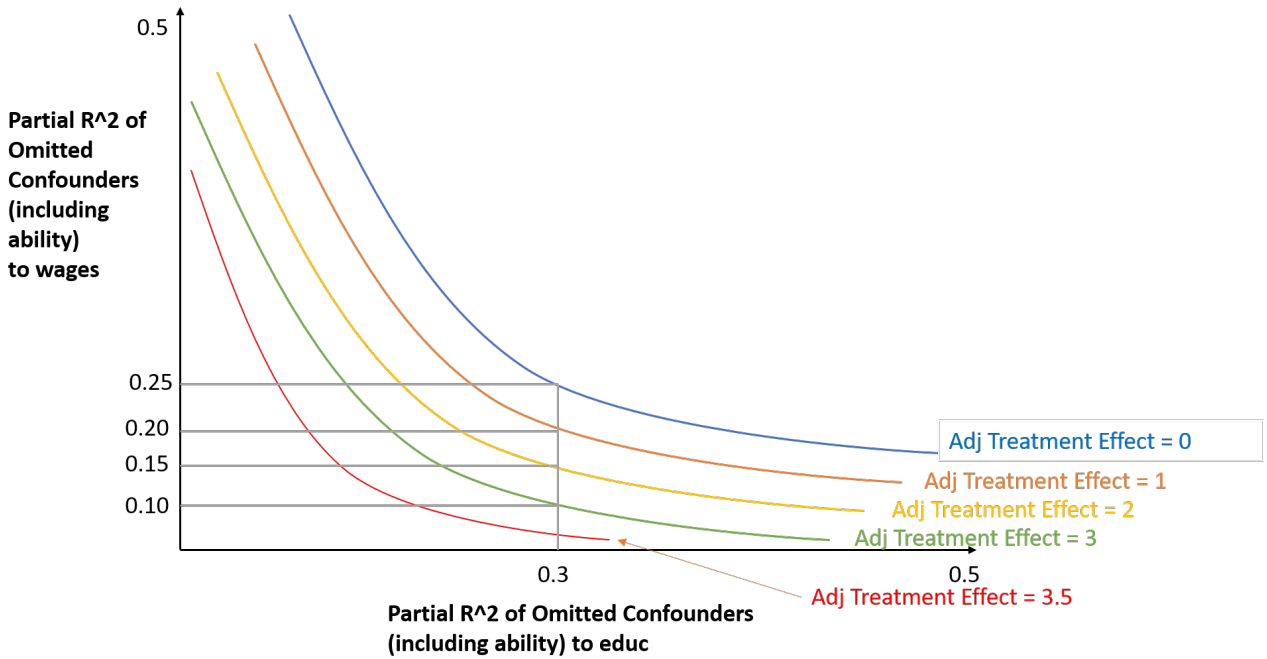
Application to our Case Study

Q: Assume that ability can explain 30% of the residual variation in education. What range of partial R^2 of ability for wages is acceptable to the policy maker?



Application to our Case Study

Q: Assume that ability can explain 30% of the residual variation in education. What range of partial R^2 of ability for wages is acceptable to the policy maker?



A: If we believe that ability (and other omitted confounders) jointly explain more than 10% of the residual variation in wages (i.e. variation not already explained by our model), then the treatment effect would be less than $\$3,000$ and the policy maker wouldn't be able to pass legislation.

Outline

1. Case Study
2. Identifying the Bias
   - General Setting
   - Case Study
3. Measuring the Bias
   - General Setting
   - Case Study
4. OVB Sensitivity Analysis
5. "Solving" OVB

Proxies as a Solution

Consider again the population regression model

$$wage = \alpha + \beta educ + \gamma abil + e$$

wherein ability is omitted.

Proxies as a Solution

Consider again the population regression model

$$wage = \alpha + \beta educ + \gamma abil + e$$

wherein ability is omitted. Perhaps we have available a proxy variable.

Proxies as a Solution

Consider again the population regression model

$$wage = \alpha + \beta educ + \gamma abil + e$$

wherein ability is omitted. Perhaps we have available a proxy variable.

$$wage = \alpha + \beta educ + \gamma IQ + e$$

Proxies as a Solution

Consider again the population regression model

$$wage = \alpha + \beta educ + \gamma abil + e$$

wherein ability is omitted. Perhaps we have available a proxy variable.

$$wage = \alpha + \beta educ + \gamma IQ + e$$

Warning: The proxy not only has to be available, but also must capture all of the information contained in the omitted regressor.

Proxies as a Solution

Consider again the population regression model

$$wage = \alpha + \beta educ + \gamma abil + e$$

wherein ability is omitted. Perhaps we have available a proxy variable.

$$wage = \alpha + \beta educ + \gamma IQ + e$$

Warning: The proxy not only has to be available, but also must capture all of the information contained in the omitted regressor.
Is IQ a suitable proxy?

Proxies as a Solution

Consider again the population regression model

$$wage = \alpha + \beta educ + \gamma abil + e$$

wherein ability is omitted. Perhaps we have available a proxy variable.

$$wage = \alpha + \beta educ + \gamma IQ + e$$

Warning: The proxy not only has to be available, but also must capture all of the information contained in the omitted regressor.
Is IQ a suitable proxy?

Perhaps something like Emotional Intelligence (EQ) is part of your innate ability to earn high wages, but is not captured by an IQ test.

OVB "Solutions"

OVB is a pernicious problem in causality. Some might say it cannot be solved, only alleviated.

OVB "Solutions"

OVB is a pernicious problem in causality. Some might say it cannot be solved, only alleviated.

- Proxies for omitted. Suitable proxies are difficult to find.

OVB "Solutions"

OVB is a pernicious problem in causality. Some might say it cannot be solved, only alleviated.

- Proxies for omitted. Suitable proxies are difficult to find.
- Instrumental Variables (IV). Suitable instruments are difficult to find.

OVB "Solutions"

OVB is a pernicious problem in causality. Some might say it cannot be solved, only alleviated.

- Proxies for omitted. Suitable proxies are difficult to find.
- Instrumental Variables (IV). Suitable instruments are difficult to find.
- Fixed Effects. Requires appropriate panel data.

OVB "Solutions"

OVB is a pernicious problem in causality. Some might say it cannot be solved, only alleviated.

- Proxies for omitted. Suitable proxies are difficult to find.
- Instrumental Variables (IV). Suitable instruments are difficult to find.
- Fixed Effects. Requires appropriate panel data.
- Natural Experiments. Not within the analyst's control to create a natural experiment

OVB - Summary

- Omitted Variable Bias (OVB) is perhaps the most pernicious problem we will confront in causal estimation
- Omitting a relevant regressor generates a biased treatment effect estimate
- We cannot see nor estimate the size of the bias, however we can...
- ...intimate the sign of the bias via domain knowledge, and...
- ...put relevant bounds on the bias