Mean Comparison with Linear Regression

RMDA II — Spring 2026

Sebastian Tello-Trillo

University of Virginia

Today's Roadmap

Part I: Foundations

1. Mean comparison as OLS
2. Why OLS? The flexibility advantage
3. The unbiasedness condition

Part II: Tools

4. Covariates: reducing bias
5. Bad controls: what not to include
6. Signing the bias: OVB formula

Key Insight

The central question: How do we get *closer* to the causal effect when we cannot run an experiment?

From Mean Comparison to OLS

Where We Left Off: Mean Comparison

Recall: we decomposed the observed difference in means:

$$\underbrace{\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]}_{\text{What we observe}} = \underbrace{\mathbb{E}[Y_i^1 - Y_i^0 \mid D_i = 1]}_{\text{Causal Effect (ATT)}} + \underbrace{\mathbb{E}[Y_i^0 \mid D_i = 1] - \mathbb{E}[Y_i^0 \mid D_i = 0]}_{\text{Selection Bias}}$$

What we want:

The causal effect of treatment

What gets in the way:

Selection into treatment is non-random

A Running Example: Job Training Programs

Policy Application

A city offers a voluntary job training program. We want to know: **does the program increase participants' earnings?**

Training group

Average earnings: $84,937

No training group

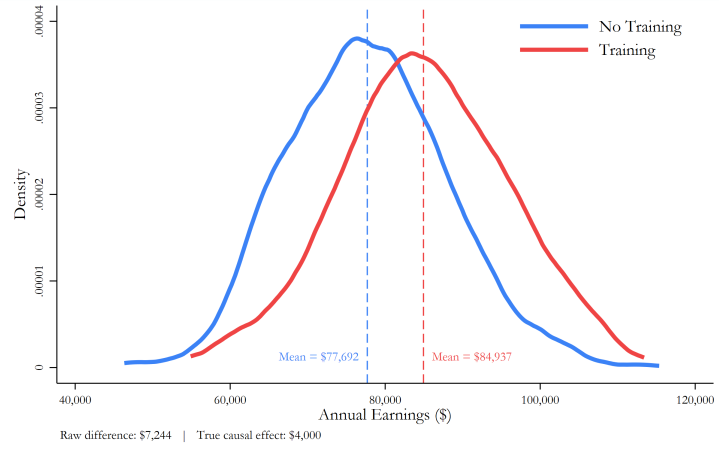
Average earnings: $77,692

Raw difference: **$7,244**

Think About It

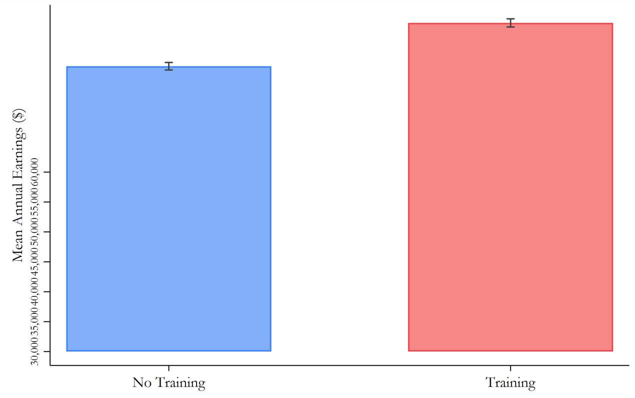
Is $7,244 the causal effect of training? What might be different about people who *choose* to enroll?

Visualizing the Raw Difference



The distributions overlap, but means differ. Is this gap causal?

Mean Comparison: The Starting Point



The bar chart gives us one number: the difference in means. But we need a more flexible tool.

Mean Comparison *is* OLS

Running this regression with a **binary** treatment variable:

$$\text{Earnings}_i = \beta_0 + \beta_1 \cdot \text{Training}_i + \varepsilon_i$$

What OLS gives us:

- $\hat{\beta}_0$ = mean of control group
- $\hat{\beta}_1$ = difference in means

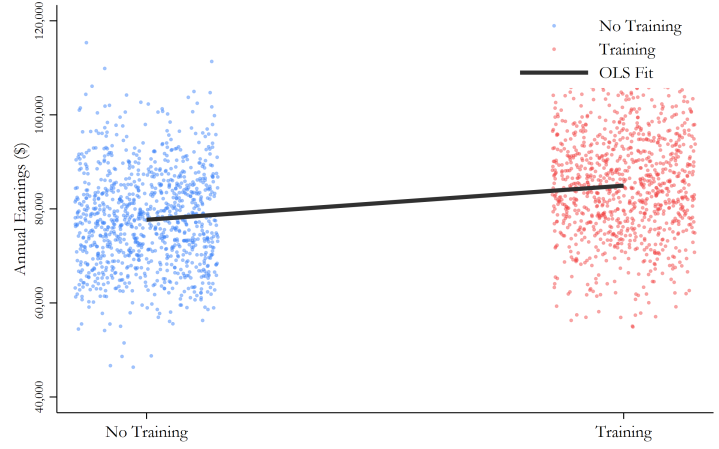
Algebraically:

- $\hat{\beta}_0 = \bar{Y}_{D=0}$
- $\hat{\beta}_1 = \bar{Y}_{D=1} - \bar{Y}_{D=0}$

Key Insight

With a binary treatment and no other variables, OLS *exactly reproduces* the difference in means. They are the same thing.

What OLS Is Doing

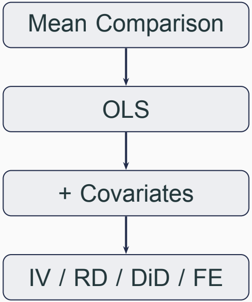


The slope of the OLS line = the difference in group means. The intercept = the control group mean.

Why Use OLS Instead of a Simple t-test?

OLS gives us the same answer, but with **more flexibility**:

1. **Add covariates** to reduce bias
2. **Control for confounders** systematically
3. **Standard errors** and inference built in
4. Foundation for **every design**: IV, RD, DiD, FE



When Can We Trust OLS?

Unbiasedness of OLS

Definition

OLS is **unbiased** if, on average across many samples:

$$\mathbb{E}[\hat{\beta}_{\text{OLS}}] = \beta_{\text{true}}$$

This means: if we could repeat our study many times with new random samples, the *average* of our estimates would equal the true parameter.

Think About It

Unbiasedness does *not* mean any single estimate is correct. It means the estimation procedure is not systematically off-target.

What is $E(\hat{\beta_{ols}})$?

$$\hat{\beta_{ols}} = \frac{Cov(D_1, Y)}{Var(D_1)}$$

From here we'll derive that:

$$E(\hat{\beta_{ols}}) = \beta_{true} + \rho(D, \epsilon)\frac{\sigma_e}{\sigma_X}$$

$$\beta_{true} + correlation(D, \epsilon)\frac{Standard\ deviation\ (\epsilon)}{Standard\ deviation\ (D)}$$

- You can find the derivation in "Advanced OLS" chapter of Real Stats

Questions

Questions

Can Standard deviation of epsilon be 0 $\sigma_\epsilon = 0$?

Can standard deviation of x be 0. $\sigma_x = 0$?

What would that imply?

The Condition for Unbiasedness

For $\hat{\beta}_1$ to be unbiased, we need:

$$\boxed{\text{Corr}(D_i, \varepsilon_i) = 0}$$

In words: the treatment variable must be **uncorrelated with the error term**.

Question

What is epsilon ϵ?

The Condition for Unbiasedness

For $\hat{\beta}_1$ to be unbiased, we need:

$$\boxed{\text{Corr}(D_i, \varepsilon_i) = 0}$$

In words: the treatment variable must be **uncorrelated with the error term**.

The error term ε_i

Everything that affects Y but is *not* in our model. These are the **unobservable characteristics**.

The condition means:

Treatment assignment is not related to things we haven't accounted for in the model.

Back to Job Training

$$\text{Earnings}_i = \beta_0 + \beta_1 \cdot \text{Training}_i + \varepsilon_i$$

What is in ε_i? Everything else that affects earnings:

- Education level
- Age and experience
- Prior earnings (proxy for ability)
- Motivation, network effects, ...

Think About It

If more educated or higher-ability people are more likely to enroll in training, then $\text{Corr}(\text{Training}, \varepsilon) > 0$ and our estimate is **biased (more positive)**.

Reducing the Bias

Goal: make Corr(D, ε) as close to zero as possible.

One strategy: pull variables *out* of ε and into the regression.

$$\text{Before:} \quad \text{Earnings}_i = \alpha_0 + \alpha_1 \cdot \text{Training}_i + \varepsilon_i$$

$$\text{After:} \quad \text{Earnings}_i = \beta_0 + \beta_1 \cdot \text{Training}_i + \beta_2 \cdot \text{Educ}_i + \beta_3 \cdot \text{Age}_i + \mu_i$$

Key Insight

By adding covariates, we remove part of what was correlated with treatment from ε. The remaining error μ_i is (hopefully) less correlated with D.

Adding Covariates

Which Covariates Should We Add?

We want variables that satisfy **two conditions**:

Condition 1	**Condition 2**
Correlated with **Y** (the outcome)	Correlated with **D** (the treatment)

If not: irrelevant to the outcome and won't reduce bias.

If not: doesn't explain selection and won't change $\hat{\beta}_1$.

Plus a secret rule (Really, condition 3) The variable must *not* be a **post-treatment outcome** (more on this shortly).

How do we find these variables?

- It comes from our priors about how the world works
- In fancy terms, theory (Insert here your institutional knowledge or information from other fields)
- Ask yourself the question: What would make observations select in or out of treatment?

Let's go over an example

Adding Covariates: What Happens to $\hat{\beta}_1$?

	(1)	(2)	(3)	(4)	(5)
Job Training Program	7244.3***	5041.7***	4756.4***	4465.4***	4232.2***
	(472.8)	(408.3)	(296.6)	(289.2)	(274.8)
Years of Education		2556.5***	2505.1***	2510.8***	1969.4***
		(88.9)	(63.7)	(61.9)	(67.5)
Age			843.8***	844.1***	536.4***
			(20.2)	(19.6)	(27.6)
Female				3128.4***	3063.0***
				(282.8)	(268.5)
Prior Earnings ($)					0.4***
					(0.0)
Constant	77692.4***	44088.8***	13918.2***	12462.8***	11559.8***
	(320.5)	(1182.3)	(1132.4)	(1098.7)	(1055.4)
R^2	0.105	0.369	0.663	0.683	0.714
Observations	2,000	2,000	2,000	2,000	2,000

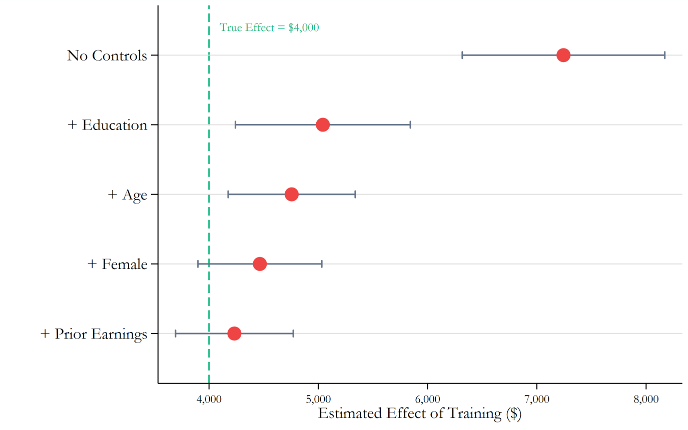
Robust standard errors in parentheses.

True causal effect = $4,000.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As we add controls, the training coefficient moves toward the true effect ($4,000).

Visualizing Bias Reduction



With no controls, we overestimate the effect. Adding covariates moves us toward the truth.

How Does Adding a Covariate Change $\hat{\beta}_1$?

With one variable: $\hat{\beta}_1 = \dfrac{\text{Cov}(X_1, Y)}{\text{Var}(X_1)}$

With multiple variables, this becomes:

$$\hat{\beta}_1 = \frac{\text{Cov}(\tilde{X}_1,\ Y)}{\text{Var}(\tilde{X}_1)}$$

where $\tilde{X}_1$ is the **residual** from regressing X_1 on all other covariates.

Key Insight

OLS isolates the "unique part" of each variable. Adding X_2 strips out variation in X_1 explained by X_2. The coefficient $\hat{\beta}_1$ is based only on the *remaining* variation.

Where Does $\tilde{X}_1$ Come From?

Start with the auxiliary regression of X_1 on all other covariates:

$$X_1 = \alpha_0 + \alpha_1 X_2 + \cdots + \alpha_k X_k + \nu$$

The residual from this regression is:

$$\tilde{X}_1 = X_1 - \hat{X}_1 = X_1 - (\hat{\alpha}_0 + \hat{\alpha}_1 X_2 + \cdots + \hat{\alpha}_k X_k) = \hat{\nu}$$

What $\tilde{X}_1$ captures:

The part of X_1 that cannot be predicted by the other covariates—its "unique" variation.

Why it matters:

If X_1 and X_2 are highly correlated, $\tilde{X}_1$ is small $\Rightarrow \hat{\beta}_1$ will change a lot when you add X_2.

Practical Implication

Think About It

If I run $Y = \beta_0 + \beta_1 X_1$ and get $\hat{\beta}_1 = 0.35$,
then run $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ and get $\hat{\beta}_1 = 0.03$,

what does this tell me?

Answer: X_1 and X_2 are highly correlated. Most of what X_1 appeared to "explain" was actually driven by its correlation with X_2.

The flip side:
If $\hat{\beta}_1$ barely changes when adding X_2, then X_1 and X_2 are not very correlated—adding X_2 did not address the source of bias related to X_1.

Bad Controls

Not All Covariates Are Good: Bad Controls

Warning

A **bad control** is a variable that is itself *affected by the treatment*—it is a **post-treatment outcome**.

Classic example: Effect of schooling on wages.

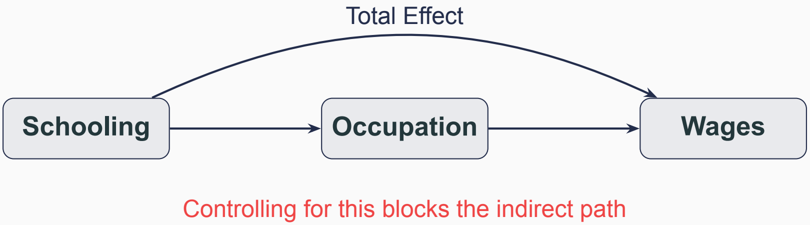
$$\text{Wages}_i = \beta_0 + \beta_1 \cdot \text{Schooling}_i + \varepsilon_i \qquad (\beta_{\text{true}} > 0)$$

Now add occupation:

$$\text{Wages}_i = \beta_0 + \beta_1 \cdot \text{Schooling}_i + \beta_2 \cdot \text{Occupation}_i + \varepsilon_i$$

Problem: Schooling $\rightarrow$ Occupation $\rightarrow$ Wages. By controlling for occupation, we block part of the *causal channel* through which schooling affects wages.

Why Bad Controls Are Dangerous



Key Insight

Tip: If a variable could be an *outcome* in the causal chain from treatment to Y, do not control for it. Ask: "Could the treatment have caused this variable to change?"

Bad Controls in Action

	Good Controls	With Bad Control
Job Training Program	4232.2***	4031.8***
	(274.8)	(319.6)
Years of Education	1969.4***	1921.5***
	(67.5)	(78.6)
Age	536.4***	538.0***
	(27.6)	(27.7)
Female	3063.0***	3062.8***
	(268.5)	(268.2)
Prior Earnings ($)	0.4***	0.4***
	(0.0)	(0.0)
Occupation Tier FE	No	Yes
R^2	0.714	0.715
Observations	2,000	2,000

Robust standard errors in parentheses.

Occupation tier is a post-treatment variable (bad control).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column (1): good controls. Column (2): adds occupation tier (post-treatment). Training effect shrinks.

Real-World Example: Gender Pay Gap at Google

In 2017, Google claimed there was *no* gender pay gap after controlling for job title, location, and performance.

Their regression:

$$\text{Wage}_i = \alpha_0 + \beta_1 \text{Female}_i + \sum_d \beta_d \text{Job}_d + \ldots$$

Former Employees Are Suing Google Over Alleged Gender Discrimination

After the *New York Times* detailed the employee spreadsheets on Friday, Google spokesperson Gina Scigliano told Gizmodo that its own data shows, when you take “location, tenure, job role, level and performance” into account, that “women are paid 99.7% of what men are paid at Google.” Scigliano described the *Times* story as “extremely flawed.”

Warning

Job title is a **post-treatment outcome**. If gender affects which jobs people get, controlling for job title blocks that causal channel.

It’s a Textbook Example of What Not to Do

Beware Bad Control

Perhaps more controls are the answer. Why not control for occupation for example? Many data sets that report earnings also classify workers’ jobs, such as whether they’re employed as a manager or laborer. Surely occupations are strong predictors of both schooling and earnings, possibly capturing traits that distinguish Mick and Johan from more average Joes. By the logic of OVB, therefore, we should control for occupation, a matter easily accomplished by including dummy variables to indicate the types of jobs held.

In spite of the fact that occupation is strongly correlated with both schooling and wages, occupation dummies are *bad controls* in a wage equation. The fact that Master Joshway works today as a professor and not a busboy (as he once did) is in part a reward for his extravagant schooling. It’s a mistake to eliminate this benefit from our calculation by comparing only professors or busboys when attempting to quantify the economic value of schooling. Even in a world where all professors earn a uniform one million dollars a year (may it soon come to pass) and all busboys earn a uniform $10,000, an experiment that randomly assigns schooling would show that schooling raises wages. The channel by which wages are increased in this notional experiment is the shift from lowly busboyhood to elevated professorness.

Including occupation dummies in a wage equation is the canonical example of a bad control.

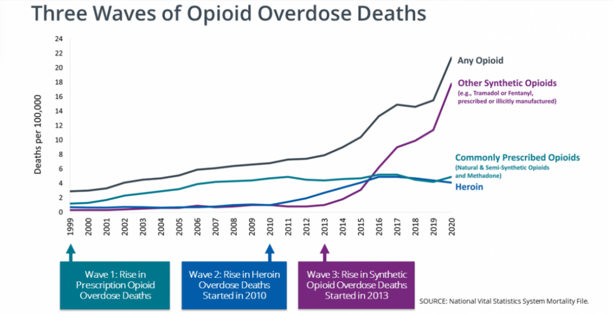
See: *Mastering ’Metrics*, p. 215 | Sally Hudson’s tweet

Real-World Application: Deaths of Despair

Application: Are Opioid Deaths "Deaths of Despair"?

Policy Application

The "deaths of despair" hypothesis: rising mortality from drugs, alcohol, and suicide is driven by **declining economic conditions** in affected communities.



Deaths of Despair?

The proposed idea is that these deaths are driven by changes in economic conditions.

Think About It

How would one measure changes in economic conditions?

What would be the regression that you would run?

The Research Design: Kitchen Sink OLS

Regression:

$$\Delta M_k = \alpha_0 + \beta \, \Delta \vec{E}_k + \varepsilon_k$$

Outcome:

ΔM_k: Change in mortality across counties

Explanatory variables: $\Delta \vec{E}_k$:

- Changes in poverty rate
- Changes in median income
- Changes in home prices
- Changes in unemployment
- Changes in import exposure

Think About It

What is the main assumption for interpreting $\hat{\beta}$ as causal? What might be in ε_k?

Results: Economic Conditions and Drug Mortality

The coefficients of interest, $\ddot{\beta}$, show estimated economic effects on mortality growth. A requirement for unbiased estimates is

The author progressively adds covariates to see how the economic conditions coefficient changes.

Results: After Adding Controls

that the supplementary covariates must adequately control for influences on mortality trends that are spuriously correlated with $\Delta \boldsymbol{E}$ such that $\text{cov}(\Delta E_k, \varepsilon_k) = 0$. However, the estimates will be attenuated if these additional controls include variables that are caused by changes in economic conditions. This is not a problem for the predetermined regressors, $\boldsymbol{X}_0$, but could be an issue for $\Delta \boldsymbol{X}$. For example, individuals, particularly young ones, often migrate to areas with better economic conditions (Greenwood, 1997), implying that changes in age-specific population shares (or in other population characteristics) may be affected by economic performance. Similarly, localities hit by international trade shocks experience reductions in the supply of marriageable men and increases in the fraction of children born to unwed mothers

Economic variables explain very little of the variation in drug mortality changes, once supply-side factors are controlled for.

Results: Table 1

Table 1
Estimated Effect of Economic Conditions on Changes in Various Death Rates, 1999–2015.

Economic Proxy	All Drugs	Opioid Analgesics	Illicit Opioids	DSA
Measures Included Separately				
Δ in Poverty Rate	2.205***	0.798***	1.334***	2.320***
	(0.560)	(0.242)	(0.446)	(0.752)
Δ in Median Household Income	2.068***	0.679***	1.136**	2.515***
	(0.546)	(0.254)	(0.496)	(0.773)
Δ in Median Home Price	2.289***	0.908**	1.158*	2.840***
	(0.649)	(0.354)	(0.627)	(0.680)
Δ in Unemployment Rate	1.370***	0.295**	1.069***	1.144
	(0.464)	(0.131)	(0.253)	(0.765)
Δ in Import Exposure	0.572	0.398**	0.168	0.570
	(0.414)	(0.182)	(0.328)	(0.511)
Measures Included Together				
Δin Poverty Rate	1.102**	0.519**	0.782*	0.793
	(0.515)	(0.259)	(0.403)	(0.599)
Δ in Median Household Income	0.206	−0.097	−0.043	0.751
	(0.671)	(0.329)	(0.543)	(0.951)
Δ in Median Home Price	1.465*	0.710*	0.626	1.959**
	(0.805)	(0.409)	(0.668)	(0.883)
Δ in Unemployment Rate	0.307	−0.143	0.536	−0.146
	(0.452)	(0.213)	(0.365)	(0.669)
Δ in Import Exposure	0.212	0.269	−0.065	0.214
	(0.392)	(0.177)	(0.343)	(0.470)
P-Value	<0.001	<0.001	<0.001	<0.001
R^2	0.082	0.059	0.051	0.060
Multiple Proxy Estimate	2.949***	1.164***	1.710***	3.256
	(0.798)	(0.239)	(0.428)	(2.395)
% of Total Δ Explained	32.5%	27.6%	25.6%	26.3%
Dep. Var. Mean [SD]	10.37	3.58	6.27	15.39
	[9.06]	[4.22]	[6.67]	[12.38]

Results: Table 2

Table 2
Estimated Effect of Economic Conditions on 1999–2015 Change in Total Drug Death Rate, with Various Sets of Controls.

Economic Conditions Proxy	(a)	(b)	(c)
Δ in Poverty Rate	1.102**		
	(0.515)		
Δ in Median Household Income	0.206		
	(0.671)		
Δ in Median Home Price	1.465*		
	(0.805)		
Δ in Unemployment Rate	0.307		
	(0.452)		
Δ in Import Exposure	0.212		
	(0.392)		
P-Value	<0.001		
R^2	0.082		
Multiple Proxy Estimate	2.949***		
	(0.798)		
% of Total Δ Explained	32.5%		
Additional Controls	None		

Results: Table 2

Table 2
Estimated Effect of Economic Conditions on 1999–2015 Change in Total Drug Death Rate, with Various Sets of Controls.

Economic Conditions Proxy	(a)	(b)	(c)
Δ in Poverty Rate	1.102**	0.638	0.736**
	(0.515)	(0.397)	(0.361)
Δ in Median Household Income	0.206	–0.604	0.171
	(0.671)	(0.434)	(0.393)
Δ in Median Home Price	1.465*	0.337	0.115
	(0.805)	(0.441)	(0.350)
Δ in Unemployment Rate	0.307	0.160	–0.185
	(0.452)	(0.257)	(0.312)
Δ in Import Exposure	0.212	–0.283	–0.302
	(0.392)	(0.237)	(0.262)
P-Value	<0.001	0.496	0.003
R^2	0.082	0.431	0.441
Multiple Proxy Estimate	2.949***	0.431	0.792*
	(0.798)	(0.488)	(0.436)
% of Total Δ Explained	32.5%	4.8%	8.7%
Additional Controls	None	$\mathbf{X}_{1999}$, $\Delta\mathbf{X}$	$\mathbf{X}_{1999}$, $\Delta\mathbf{X}^I$

Conclusions

Drivers of the fatal drug epidemic

Christopher J. Ruhm

Frank Batten School of Leadership & Public Policy, University of Virginia, 235 McCormick Road, P.O. Box 400893, Charlottesville, VA, 22904-4893, United States

A R T I C L E I N F O

Article history:
Received 22 June 2018
Received in revised form
23 December 2018
Accepted 9 January 2019
Available online 14 January 2019

Keywords:
Drug mortality
Opioid deaths
Drug deaths
Economic conditions
Macroeconomy
Drug fatalities
Opioid mortality

A B S T R A C T

This study examines the contributions of the medium-run evolution of local economies and of changes in the “drug environment’ in explaining county-level changes in drug and related mortality rates from 1999 to 2015. A primary finding is that drug mortality rates did increase more in counties experiencing relative economic decline than in those with more robust growth, but that the relationship is weak and mostly accounted for by confounding factors. In the preferred estimates, less than one-tenth of the rise in drug and opioid-involved fatality rates is explained and the contribution is even smaller, quite possibly zero, when allowing for plausible selection on unobservables. Conversely, the risk of drug deaths varies systematically over time across population subgroups in ways that are consistent with an important role for the public health environment related to the availability and cost of drugs. In particular, the relative risk and share of drug mortality increased rapidly for males and younger adults, compared to their counterparts, when the primary driver of the fatal drug epidemic transitioned from prescription to illicit opioids. These results suggest that efforts to improve local economies, while desirable for other reasons, are not likely to yield significant reductions in overdose mortality, but with greater potential for interventions directly addressing the drug environment.

Signing the Bias

We Can't Eliminate All Bias—But We Can Sign It

- We've seen how adding covariates reduces bias
- But we can *never* include everything—some bias always remains
- However, we can often determine the **direction** of the remaining bias

Key Insight

Even if you cannot pin down the exact magnitude of the bias, knowing whether your estimate is **too high** or **too low** is valuable for policy conclusions.

Signing the Bias: Example

- Imagine someone wanted to estimate the following:

$$\text{PoliceForce}_i = \alpha + \gamma_1 \text{BodyCamera}_i + \epsilon$$

- Say this person finds the following

$$\text{PoliceForce}_i = \alpha + 0.30\,\text{BodyCamera}_i + \epsilon$$

Think About It

What's one omitted variable that would be affecting the "causal" interpretation of γ_1?

Income as Omitted Variable

- *Colloquial Answer: "Police departments that have BWC tend to be in higher income neighborhoods, and those are the neighborhoods that you would also see less use of force, **and so the effect would be smaller**."*
- That argument may be true, but the direction of the bias may result in a different conclusion.

Think About It

How do we know if the argument is pushing the estimate in one direction or another?

Finding the Sign of the Bias

- In order to find the sign of the bias we need two pieces of information:
  - (1) What's the sign of the correlation between the omitted variable and the outcome? $\text{Corr}(OV, Y)$
  - (2) What's the sign of the correlation between the omitted variable and the main explanatory variable? $\text{Corr}(OV, D)$
- The sign of the bias is then $\text{Corr}(OV, Y) \times \text{Corr}(OV, D)$
- So if $\text{Corr}(OV, Y) = +$ and $\text{Corr}(OV, D) = -$, then the sign of the bias is $(-)$

Example: Body Cameras and Use of Force

A researcher estimates:

$$\text{UseOfForce}_i = \alpha + 0.30 \cdot \text{BodyCamera}_i + \varepsilon$$

Suspected omitted variable: Neighborhood income

Omitted Variable	Corr(OV, BodyCam)	Corr(OV, Force)
Neighborhood Income	$(+)$	$(-)$
Bias direction:		$(+) \times (-) = (-)$

Implication: $\gamma_1 = \beta_1 +$ (negative bias), so $\beta_1 > \gamma_1$.

The true reduction in force is **larger** than what we observe.

OVB in Action: Simulated Body Camera Data

	Biased	With Income Control
Body Camera	-0.748***	-0.514***
	(0.091)	(0.095)
Neighborhood Income ($)		-0.000***
		(0.000)
Constant	3.833***	4.728***
	(0.066)	(0.153)
R^2	0.121	0.196
Observations	500	500

Robust standard errors in parentheses.

True effect of body cameras = -0.5.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Without income control, the camera effect is biased away from zero. Adding income moves toward the true effect (-0.5).

Body Cameras: What Did We Learn?

Key Insight

Recall: *"Police departments that have BWC tend to be in higher income neighborhoods, and those are the neighborhoods that you would also see less use of force, and so the effect would be smaller."*

In fact, when we account for neighborhood income, the effects should be **larger**, not smaller. The colloquial intuition about the direction was wrong—signing the bias formally shows the true effect is larger in magnitude than the naïve estimate.

Omitted Variable Bias: The Formula

Suppose the **true model** is:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_2 + \nu_i$$

But we can only estimate the **short regression** (omitting X_2):

$$Y_i = \gamma_0 + \gamma_1 D_i + \varepsilon_i$$

Using the **auxiliary regression** $X_2 = \delta_0 + \delta_1 D_i + \tau_i$, we can show:

$$\boxed{\gamma_1 = \beta_1 + \underbrace{\beta_2 \cdot \delta_1}_{\text{Bias}}}$$

β_2: effect of omitted var on Y

δ_1: relationship between omitted var and D

Where Does the Formula Come From?

Start with the true model and substitute:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 \underbrace{(\delta_0 + \delta_1 D_i + \tau_i)}_{X_2} + \nu_i$$

$$= (\beta_0 + \beta_2\delta_0) + (\beta_1 + \beta_2\delta_1) D_i + (\beta_2\tau_i + \nu_i)$$

$$= \quad \gamma_0 \quad + \quad \gamma_1 \quad D_i + \quad \varepsilon_i$$

Therefore:

$$\gamma_1 = \beta_1 + \beta_2\delta_1 = \text{True effect} + \text{Bias}$$

Bias is zero when either $\beta_2 = 0$ (omitted var doesn't affect Y) or $\delta_1 = 0$ (omitted var is uncorrelated with D).

Practice: Study Time and Test Scores

A researcher estimates:

$$
\begin{aligned}
&(1) \quad \Pr(\text{Pass}) = \beta_0 + \beta_1 \cdot \text{StudyTime} \\
&(2) \quad \Pr(\text{Pass}) = \gamma_0 + \gamma_1 \cdot \text{StudyTime} + \gamma_2 \cdot \text{HWGrade}
\end{aligned}
$$

Given:

- StudyTime is *negatively* correlated with HWGrade
- HWGrade is *positively* correlated with passing

Think About It

Is $\hat{\beta}_1$ (from regression 1) biased **more positive** or **more negative**?

Signing the Bias: A Practical Tool

The sign of the bias is:

$$\text{sign(Bias)} = \text{sign}(\beta_2) \times \text{sign}(\delta_1)$$

	$\text{Corr}(OV, D) > 0$	$\text{Corr}(OV, D) < 0$
$\text{Corr}(OV, Y) > 0$	Bias > 0 (more positive)	Bias < 0 (more negative)
$\text{Corr}(OV, Y) < 0$	Bias < 0 (more negative)	Bias > 0 (more positive)

Just two pieces of information—both based on *theory* about how the world works—let you determine the direction of bias.

Summary

What We Learned Today

1. **Mean comparison = OLS** with a binary treatment (no covariates)
2. OLS is unbiased when $\text{Corr}(D, \varepsilon) = 0$ — rarely satisfied in observational data
3. **Adding covariates** pulls confounders from ε, reducing bias
4. **Bad controls** are post-treatment variables—never include them
5. The **OVB formula**: $\gamma_1 = \beta_1 + \beta_2 \cdot \delta_1$

Policy Application

Next class: We move beyond "kitchen sink" OLS to designs that can *credibly* achieve $\text{Corr}(D, \varepsilon) = 0$: instrumental variables, regression discontinuity, and more.

Resources

- **Textbook:** Chapter 5 and Chapter 14.4
- Case study on page 141
- Google gender pay gap analysis

TABLE 14.1 Effect of Omitting X_2 on Coefficient Estimate for X_1

Correlation of X_1 and X_2	β_2 Effect of omitted variable on Y: > 0	0	< 0
> 0	Overstate coefficient	No bias	Understate coefficient
0	No bias	No bias	No bias
< 0	Understate coefficient	No bias	Overstate coefficient

Cell entries show sign of bias for omitted variable bias problem in which a single variable (X_2) is omitted.

The true equation is Equation 14.8 and the estimated model is Equation 14.9. If $\beta_2 > 0$ and X_1 and X_2 are positively correlated, $\hat{\beta}_1^{\text{Omit}X_2}$ (the expected value of the coefficient on X_1 from a model that omits X_2) will be larger than the actual value of β_1.

Appendix: Standard Errors and Covariates

Standard Errors: The Bivariate Case

Start with the simple regression (one explanatory variable, no covariates):

$$Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i$$

The standard error of $\hat{\beta}_1$ is:

$$\text{se}(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\sum_{i=1}^{N}(X_{1i} - \bar{X}_1)^2}} = \frac{\hat{\sigma}_\varepsilon}{\sqrt{N \cdot \text{Var}(X_1)}}$$

Two things make the standard error **smaller**:

- **Less noise** in the outcome: smaller $\hat{\sigma}_\varepsilon$
- **More variation** in X_1: larger $\text{Var}(X_1)$

Standard Errors: Adding a Covariate

Now add one covariate X_2: $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu_i$

The standard error of $\hat{\beta}_1$ becomes:

$$\text{se}(\hat{\beta}_1) = \frac{\hat{\sigma}_\mu}{\sqrt{N \cdot \text{Var}(X_1)}} \cdot \frac{1}{\sqrt{1 - R^2_{X_1 \sim X_2}}}$$

where $R^2_{X_1 \sim X_2}$ is the R^2 from regressing X_1 on X_2.

Key Insight

There is a new term: $\frac{1}{\sqrt{1-R^2_{X_1 \sim X_2}}}$ — the **variance inflation factor (VIF)**. It is always ≥ 1, and grows as X_1 and X_2 become more correlated.

The Standard Error Trade-off

Adding covariates has **two competing effects** on $\text{se}(\hat{\beta}_1)$:

Good news:

Adding a relevant covariate reduces $\hat{\sigma}_\mu$ (less unexplained variation in Y), which *shrinks* the standard error.

Bad news:

If X_2 is correlated with X_1, the VIF term $\frac{1}{\sqrt{1-R^2_{X_1\sim X_2}}}$ grows, which *inflates* the standard error.

	Bivariate	With covariate X_2
Residual variance	$\hat{\sigma}^2_\varepsilon$	$\hat{\sigma}^2_\mu \leq \hat{\sigma}^2_\varepsilon$
Effective variation in X_1	$\text{Var}(X_1)$	$\text{Var}(X_1)(1 - R^2_{X_1\sim X_2})$

Net effect depends on the relative magnitudes. In practice, adding genuine confounders usually helps more than it hurts.

Example: Job Training and Standard Errors

Back to our job training data. Compare two regressions:

$$(1) \quad \text{Earnings}_i = \beta_0 + \beta_1 \cdot \text{Training}_i + \varepsilon_i$$

$$(2) \quad \text{Earnings}_i = \beta_0 + \beta_1 \cdot \text{Training}_i + \beta_2 \cdot \text{Education}_i + \mu_i$$

	No Controls	+ Education	+ Educ. & Age
Job Training Program	7244.3***	5041.7***	4756.4***
	(472.8)	(408.3)	(296.6)
Years of Education		2556.5***	2505.1***
		(88.9)	(63.7)
Age			843.8***
			(20.2)
Constant	77692.4***	44088.8***	13918.2***
	(320.5)	(1182.3)	(1132.4)
Root MSE ($\hat{\sigma}$)	10560.0	8869.4	6481.7
R^2	0.105	0.369	0.663
Observations	2,000	2,000	2,000

Robust standard errors in parentheses.

Notice how the SE on Training changes as we add covariates.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

What Happened to the Standard Error?

Residual variance fell:

Education explains part of the variation in earnings that was in ε, so $\hat{\sigma}_\mu^2 < \hat{\sigma}_\varepsilon^2$.

This pushes the SE **down**.

Effective variation in Training fell:

Training and education are positively correlated ($R^2_{T\sim E} > 0$), so the VIF > 1.

This pushes the SE **up**.

Key Insight

In this case, the reduction in residual variance dominates—the standard error *decreases* when we add education. This is the typical pattern when you add a genuine confounder.

Try it yourself! The do file `lecture4_figures.do` generates this table so you can experiment with different covariates.