Omitted Variable Bias

When Relevant Variables Are Left Out

Jake Anderson

March 3, 2026

The OLS Assumptions Under Random Sampling

Assumption	Description
RS1	The model is linear: $y_i = \beta_1 + \beta_2 x_i + e_i$
RS2	The data (y_i, x_i) are i.i.d.
RS3	**Exogeneity:** $E(e_i \mid x_i) = 0$
RS4	Homoskedasticity: $\text{Var}(e_i \mid x_i) = \sigma^2$
RS5	Rank condition: x_i takes at least two values
RS6	Normality: $e_i \sim N(0, \sigma^2)$

The OLS Assumptions Under Random Sampling

Assumption	Description
RS1	The model is linear: $y_i = \beta_1 + \beta_2 x_i + e_i$
RS2	The data (y_i, x_i) are i.i.d.
RS3	**Exogeneity:** $E(e_i \mid x_i) = 0$
RS4	Homoskedasticity: $\text{Var}(e_i \mid x_i) = \sigma^2$
RS5	Rank condition: x_i takes at least two values
RS6	Normality: $e_i \sim N(0, \sigma^2)$

We have relaxed RS4 (heteroskedasticity), RS6 (non-normality), and RS2 (time series).

The OLS Assumptions Under Random Sampling

Assumption	**Description**
RS1	The model is linear: $y_i = \beta_1 + \beta_2 x_i + e_i$
RS2	The data (y_i, x_i) are i.i.d.
RS3	**Exogeneity:** $E(e_i \mid x_i) = 0$
RS4	Homoskedasticity: $\text{Var}(e_i \mid x_i) = \sigma^2$
RS5	Rank condition: x_i takes at least two values
RS6	Normality: $e_i \sim N(0, \sigma^2)$

We have relaxed RS4 (heteroskedasticity), RS6 (non-normality), and RS2 (time series).

$\implies$ Now we tackle **RS3**: what happens when $E(e_i \mid x_i) \neq 0$?

The OLS Assumptions Under Random Sampling

Assumption	Description
RS1	The model is linear: $y_i = \beta_1 + \beta_2 x_i + e_i$
RS2	The data (y_i, x_i) are i.i.d.
RS3	**Exogeneity:** $E(e_i \mid x_i) = 0$
RS4	Homoskedasticity: $\text{Var}(e_i \mid x_i) = \sigma^2$
RS5	Rank condition: x_i takes at least two values
RS6	Normality: $e_i \sim N(0, \sigma^2)$

We have relaxed RS4 (heteroskedasticity), RS6 (non-normality), and RS2 (time series).

$\Longrightarrow$ Now we tackle **RS3**: what happens when $E(e_i \mid x_i) \neq 0$?

Three sources of endogeneity: (1) omitted variable bias, (2) measurement error, (3) simultaneity.

Motivation: Test Scores

What predicts test scores? Two plausible factors:

- X_1 = prior grade (foundation from last term)
- X_2 = study time in current course

Motivation: Test Scores

What predicts test scores? Two plausible factors:

- X_1 = prior grade (foundation from last term)
- X_2 = study time in current course

The true model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Motivation: Test Scores

What predicts test scores? Two plausible factors:

- X_1 = prior grade (foundation from last term)
- X_2 = study time in current course

The true model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Three researchers run different regressions:

1. "The past is irrelevant!" $\implies Y = \beta_0 + \beta_1 X_1 + e_1$
2. "Effort is all that counts!" $\implies Y = \beta_0 + \beta_2 X_2 + e_2$
3. "Both matter!" $\implies Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

Motivation: Test Scores

What predicts test scores? Two plausible factors:

- X_1 = prior grade (foundation from last term)
- X_2 = study time in current course

The true model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Three researchers run different regressions:

1. "The past is irrelevant!" $\implies Y = \beta_0 + \beta_1 X_1 + e_1$
2. "Effort is all that counts!" $\implies Y = \beta_0 + \beta_2 X_2 + e_2$
3. "Both matter!" $\implies Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

Question: Will $\hat{\beta}_1$ be the same in Model 1 and Model 3?

A Counterexample: The Room Game

Imagine a game: you walk down a hallway; each room you enter gives 1 point; rooms are independent; each player flips a coin to decide whether to enter each room.

A Counterexample: The Room Game

Imagine a game: you walk down a hallway; each room you enter gives 1 point; rooms are independent; each player flips a coin to decide whether to enter each room.

Model: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where $\beta_1 = \beta_2 = 1$.

A Counterexample: The Room Game

Imagine a game: you walk down a hallway; each room you enter gives 1 point; rooms are independent; each player flips a coin to decide whether to enter each room.

Model: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where $\beta_1 = \beta_2 = 1$.

Now run the short regression: $Y = \beta_0 + \beta_1 x_1 + e$

A Counterexample: The Room Game

Imagine a game: you walk down a hallway; each room you enter gives 1 point; rooms are independent; each player flips a coin to decide whether to enter each room.

Model: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where $\beta_1 = \beta_2 = 1$.

Now run the short regression: $Y = \beta_0 + \beta_1 x_1 + e$

$\implies \hat{\beta}_1 \approx 1$ (no bias!) Because $x_1 \perp x_2$ (room choices are independent).

A Counterexample: The Room Game

Imagine a game: you walk down a hallway; each room you enter gives 1 point; rooms are independent; each player flips a coin to decide whether to enter each room.

Model: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where $\beta_1 = \beta_2 = 1$.

Now run the short regression: $Y = \beta_0 + \beta_1 x_1 + e$

$\implies \hat{\beta}_1 \approx 1$ (no bias!) Because $x_1 \perp x_2$ (room choices are independent).

$\implies$ OVB requires both: (1) omitted variable affects Y, and (2) omitted variable is correlated with included X.

The OVB Formula

Short (omits X_2): $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$

The OVB Formula

Short (omits X_2): $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$
Long (includes X_2): $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

The OVB Formula

Short (omits X_2): $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$
Long (includes X_2): $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$
Auxiliary (how are X_1 and X_2 related?): $X_{2i} = \delta_0 + \delta_1 X_{1i} + v_i$

The OVB Formula

Short (omits X_2): $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$
Long (includes X_2): $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$
Auxiliary (how are X_1 and X_2 related?): $X_{2i} = \delta_0 + \delta_1 X_{1i} + v_i$

The OVB formula:

$$\hat{\beta}_1^{\text{short}} = \hat{\beta}_1^{\text{long}} + \underbrace{\hat{\beta}_2}_{\substack{\text{effect of } X_2 \\ \text{on } Y}} \times \underbrace{\hat{\delta}_1}_{\substack{\text{relationship} \\ \text{of } X_2 \text{ to } X_1}}$$

The OVB Formula

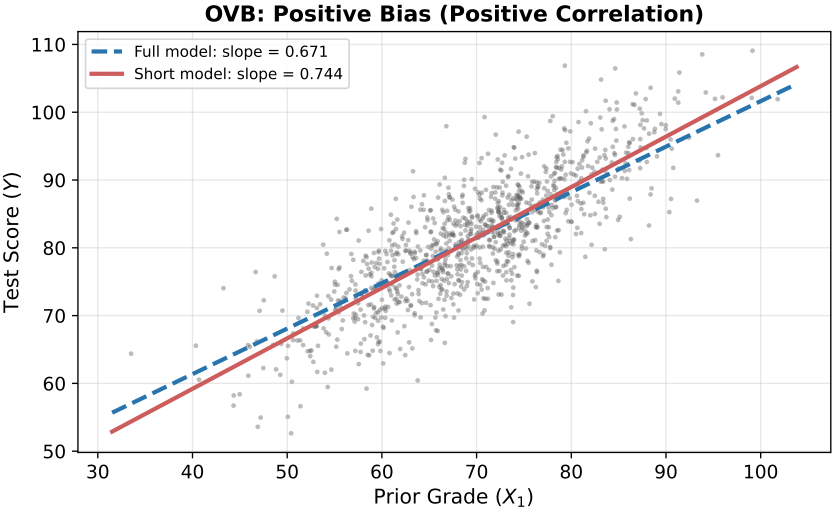
Short (omits X_2): $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$
Long (includes X_2): $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$
Auxiliary (how are X_1 and X_2 related?): $X_{2i} = \delta_0 + \delta_1 X_{1i} + v_i$

The OVB formula:

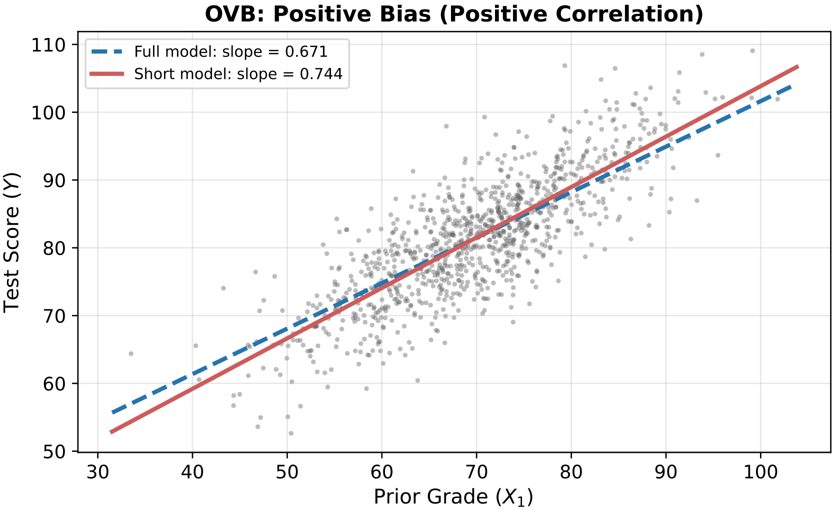
$$\hat{\beta}_1^{\text{short}} = \hat{\beta}_1^{\text{long}} + \underbrace{\hat{\beta}_2}_{\substack{\text{effect of } X_2 \\ \text{on } Y}} \times \underbrace{\hat{\delta}_1}_{\substack{\text{relationship} \\ \text{of } X_2 \text{ to } X_1}}$$

$\Longrightarrow$ The bias is $\hat{\beta}_2 \times \hat{\delta}_1$. If either is zero, there is no bias.

Visualizing OVB: Positive Bias

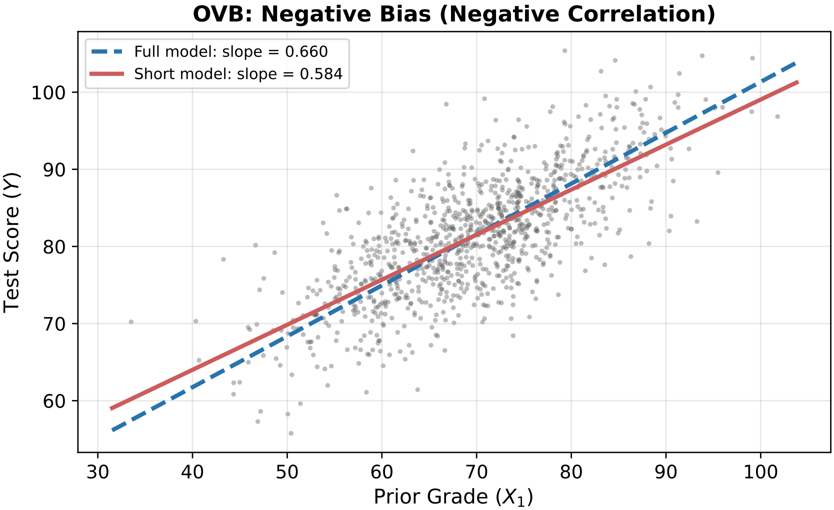


Visualizing OVB: Positive Bias

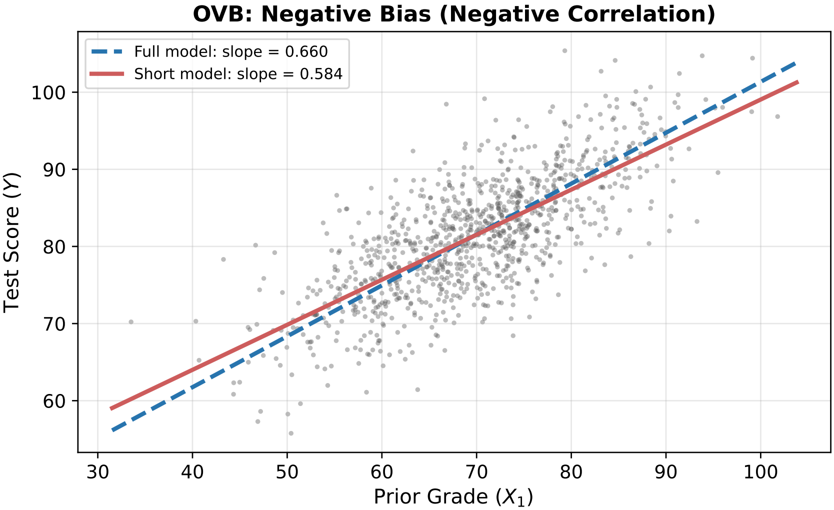


When X_1 and X_2 are **positively** correlated and $\beta_2 > 0$: the short model overestimates β_1 because it

Visualizing OVB: Negative Bias



Visualizing OVB: Negative Bias



When X_1 and X_2 are **negatively** correlated and $\beta_2 > 0$: the short model underestimates β_1 because the

Direction of Bias: The Sign Table

Direction of Omitted Variable Bias

β_{omitted} \ $\text{Corr}(X_{\text{included}}, X_{\text{omitted}})$	**Positive (+)**	**Negative (−)**
Positive (+)	**Positive bias (overestimate)**	**Negative bias (underestimate)**
Negative (−)	**Negative bias (underestimate)**	**Positive bias (overestimate)**

$$\hat{\beta}_1^{\text{short}} = \hat{\beta}_1^{\text{long}} + \hat{\beta}_2 \times \hat{\delta}_1$$

Direction of Bias: The Sign Table

Direction of Omitted Variable Bias

Corr(X_{included}, X_{omitted})

β_{omitted}	**Positive (+)**	**Negative (−)**
Positive (+)	**Positive bias (overestimate)**	**Negative bias (underestimate)**
Negative (−)	**Negative bias (underestimate)**	**Positive bias (overestimate)**

$$\hat{\beta}_1^{\text{short}} = \hat{\beta}_1^{\text{long}} + \hat{\beta}_2 \times \hat{\delta}_1$$

OVB Simulation: Test Scores

True DGP: $Y = 30 + 0.65X_1 + 0.40X_2 + \varepsilon$, with $\text{Corr}(X_1, X_2) \approx 0.4$

OVB Simulation: Test Scores

True DGP: $Y = 30 + 0.65X_1 + 0.40X_2 + \varepsilon$, with $\text{Corr}(X_1, X_2) \approx 0.4$

	Model 1 (X_1 only)	**Model 2** (X_2 only)	**Model 3** (Full)
$\hat{\beta}_1$ (Prior Grade)	≈ 0.73	—	≈ 0.65
$\hat{\beta}_2$ (Study Time)	—	≈ 0.55	≈ 0.40

OVB Simulation: Test Scores

True DGP: $Y = 30 + 0.65X_1 + 0.40X_2 + \varepsilon$, with $\text{Corr}(X_1, X_2) \approx 0.4$

	Model 1 (X_1 only)	**Model 2** (X_2 only)	**Model 3** (Full)
$\hat{\beta}_1$ (Prior Grade)	≈ 0.73	—	≈ 0.65
$\hat{\beta}_2$ (Study Time)	—	≈ 0.55	≈ 0.40

- Models 1 and 2 **overestimate** their coefficients: each attributes some of the omitted variable's effect to the included one
- Model 3 recovers estimates close to the true values

OVB Simulation: Test Scores

True DGP: $Y = 30 + 0.65X_1 + 0.40X_2 + \varepsilon$, with $\text{Corr}(X_1, X_2) \approx 0.4$

	Model 1 (X_1 only)	**Model 2** (X_2 only)	**Model 3** (Full)
$\hat{\beta}_1$ (Prior Grade)	≈ 0.73	—	≈ 0.65
$\hat{\beta}_2$ (Study Time)	—	≈ 0.55	≈ 0.40

- Models 1 and 2 **overestimate** their coefficients: each attributes some of the omitted variable's effect to the included one
- Model 3 recovers estimates close to the true values

$\implies$ Both $\beta > 0$ and Corr > 0, so bias is positive in both short models.

Summary: Omitted Variable Bias

The OVB formula: $\hat{\beta}_1^{\text{short}} = \hat{\beta}_1^{\text{long}} + \hat{\beta}_2 \times \hat{\delta}_1$

Summary: Omitted Variable Bias

The OVB formula: $\hat{\beta}_1^{\text{short}} = \hat{\beta}_1^{\text{long}} + \hat{\beta}_2 \times \hat{\delta}_1$

Direction of bias depends on two signs: effect of the omitted variable on Y ($\hat{\beta}_2$) and its correlation with the included regressor ($\hat{\delta}_1$).

Summary: Omitted Variable Bias

The OVB formula: $\hat{\beta}_1^{\text{short}} = \hat{\beta}_1^{\text{long}} + \hat{\beta}_2 \times \hat{\delta}_1$

Direction of bias depends on two signs: effect of the omitted variable on Y ($\hat{\beta}_2$) and its correlation with the included regressor ($\hat{\delta}_1$).

Two conditions for OVB (both must hold):

1. The omitted variable affects Y ($\beta_2 \neq 0$)
2. The omitted variable is correlated with included X ($\delta_1 \neq 0$)

Summary: Omitted Variable Bias

The OVB formula: $\hat{\beta}_1^{\text{short}} = \hat{\beta}_1^{\text{long}} + \hat{\beta}_2 \times \hat{\delta}_1$

Direction of bias depends on two signs: effect of the omitted variable on Y ($\hat{\beta}_2$) and its correlation with the included regressor ($\hat{\delta}_1$).

Two conditions for OVB (both must hold):

1. The omitted variable affects Y ($\beta_2 \neq 0$)
2. The omitted variable is correlated with included X ($\delta_1 \neq 0$)

Solutions:

- Include the omitted variable (if observable)
- Use instrumental variables (if a valid instrument exists)
- Use fixed effects (if the omitted variable is time-invariant)

Thank you!

jakeanderson@g.ucla.edu