Measurement Error

Attenuation Bias and the Reliability Ratio

Jake Anderson

March 3, 2026

Motivation: Study Time and Office Hours

Suppose test scores depend on **true study time** (x_i^*):

$$y_i = \beta_1 + \beta_2 x_i^* + v_i$$

But we cannot observe x_i^* directly. Instead, we use a proxy: **office hours attendance** (x_i).

The proxy measures true study time with error:

$$x_i = x_i^* + u_i$$

where u_i is measurement error with $E(u_i) = 0$ and $\text{Var}(u_i) = \sigma_u^2$.

Some students study a lot but never come to office hours. Others show up frequently but don't study much otherwise.

$\implies$ Office hours attendance is a **noisy** version of the true variable we care about.

Why Measurement Error Causes Endogeneity

Substitute $x_i^* = x_i - u_i$ into the true model:

$$y_i = \beta_1 + \beta_2(x_i - u_i) + v_i = \beta_1 + \beta_2 x_i + \underbrace{(v_i - \beta_2 u_i)}_{e_i}$$

Now check whether the regressor x_i is correlated with the composite error e_i:

$$\begin{aligned}
\text{Cov}(x_i, e_i) &= \text{Cov}(x_i^* + u_i,\ v_i - \beta_2 u_i) \\
&= \underbrace{\text{Cov}(x_i^*, v_i)}_{=0} - \beta_2 \underbrace{\text{Cov}(x_i^*, u_i)}_{=0} + \underbrace{\text{Cov}(u_i, v_i)}_{=0} - \beta_2 \underbrace{\text{Cov}(u_i, u_i)}_{=\sigma_u^2} \\
&= -\beta_2 \sigma_u^2 \neq 0
\end{aligned}$$

$\Longrightarrow$ If $\beta_2 > 0$, there is a **negative** correlation between x_i and e_i. OLS underestimates β_2.

Attenuation Bias Formula

As $N \to \infty$, the OLS estimator converges to:

$$b_2 \xrightarrow{p} \beta_2 \cdot \underbrace{\frac{\sigma^2_{x^*}}{\sigma^2_{x^*} + \sigma^2_u}}_{\lambda}$$

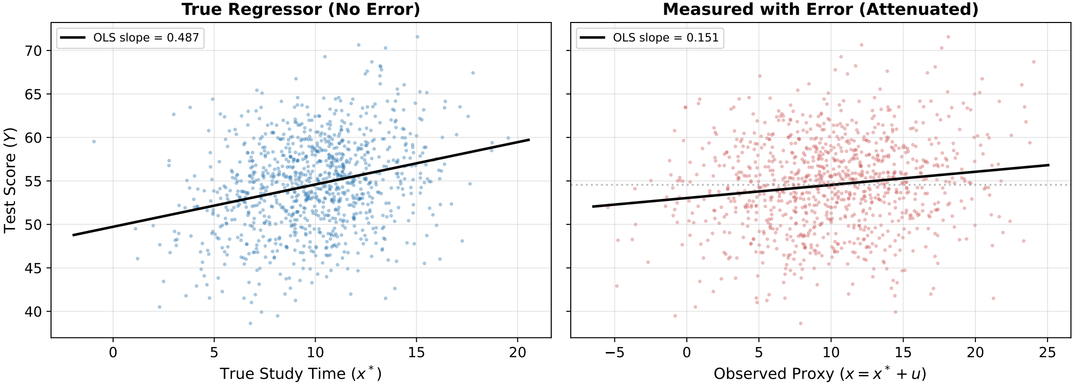
where λ is the **reliability ratio**, always between 0 and 1.

Two extreme cases:

- If $\sigma^2_u = 0$ (no error): $\lambda = 1$ and $b_2 \to \beta_2$ (no bias)
- If $\sigma^2_u \to \infty$ (pure noise): $\lambda \to 0$ and $b_2 \to 0$

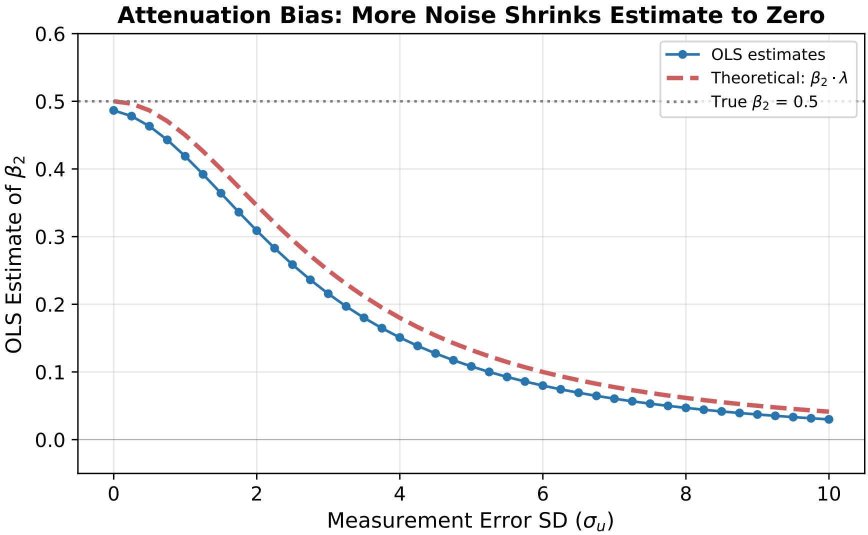
$\Longrightarrow$ Measurement error **always shrinks the coefficient toward zero**. This is called **attenuation bias**. More data does not help.

Visualizing Measurement Error

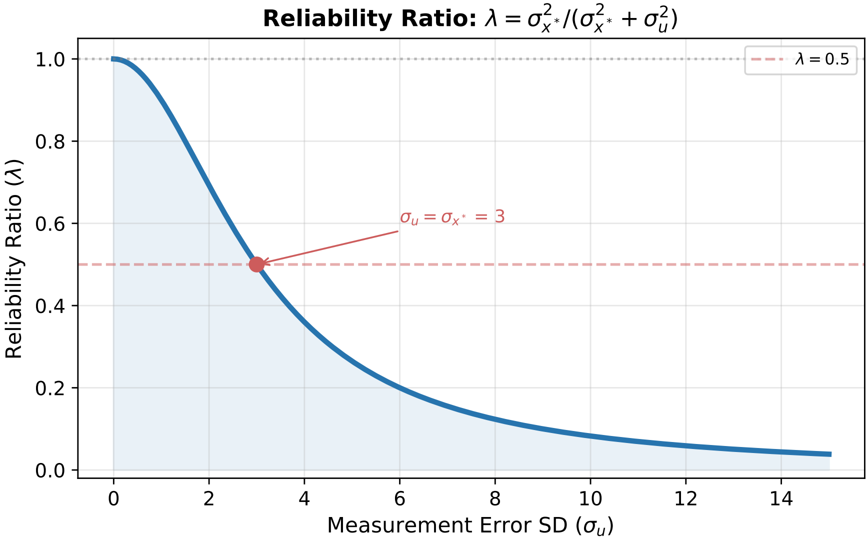


Left: using true x^*, OLS recovers the correct slope. Right: using observed x, the scatter is wider and the slope is attenuated.

Simulation: Attenuation Bias in Action



The Reliability Ratio



Measurement Error: Summary

- Measurement error in X creates endogeneity: $\text{Cov}(x_i, e_i) = -\beta_2 \sigma_u^2$
- **Attenuation bias**: the coefficient shrinks toward zero

$$b_2 \xrightarrow{p} \beta_2 \cdot \lambda, \qquad \lambda = \frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_u^2} \in (0, 1)$$

- The reliability ratio λ determines how much of the true effect survives
- More data does **not** fix it: the bias persists even as $N \to \infty$
- Solutions: find better measures of the true variable, or use instrumental variables

Thank you!

jakeanderson@g.ucla.edu