Error-in-Variables Problem

Suppose the true model, for $i = 1, \ldots, n$, is

$$y_i = \beta x_i + \varepsilon_i,$$

where x_i is the true regressor and ε_i is noise, independent of x_i.

Noisy regressor

However, we observe only a noisy version of x_i:

$$w_i = x_i + u_i,$$

where u_i is **measurement error** with mean zero and variance σ_u^2, independent of both x_i and ε_i.

Naive regression

If we regress y_i on w_i, the estimated slope is

$$\hat{\beta}_{\text{naive}} = \frac{\text{Cov}(y, w)}{\text{Var}(w)}.$$

However

$$\begin{aligned} \text{Cov}(y, w) &= \text{Cov}(\beta x + \varepsilon, x + u) = \beta \text{ Var}(x) = \beta \sigma_x^2, \\ \text{Var}(w) &= \text{Var}(x) + \text{Var}(u) = \sigma_x^2 + \sigma_u^2, \end{aligned}$$

such that

$$\hat{\beta}_{\text{naive}} = \beta \left(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \right) \leq 1.$$

The slope is biased toward zero (attenuation bias).