Bent E. Sørensen

ECONOMETRICS I, Spring 2017

Bias of the OLS estimator when the regressor is measured with error.

Consider a regression model of form

$$y_i \;=\; \alpha \;+\; \beta x_i \;+ u_i \;.$$

Under the standard OLS assumptions (x_i fixed, $Eu_i = 0$, $Eu_i u_j = 0$ when $i \neq j$ and constant variance of the u_is) the efficient OLS-estimator of β (based on N observations) is

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \;.$$

(Note: you can assume the variables are demeaned if you want simpler notation.)
Now, because

$$y_i - \bar{y} = \alpha + \beta x_i + u_i - (\alpha + \beta\bar{x} + \bar{u}) = \beta(x_i - \bar{x}) + (u_i - \bar{u}) \;,$$

we have

$$\hat{\beta} \;=\; \frac{\sum(x_i - \bar{x})(\beta(x_i - \bar{x}) + (u_i - \bar{u}))}{\sum(x_i - \bar{x})^2} \;,$$

or

$$\hat{\beta} - \beta \;=\; \frac{\sum(x_i - \bar{x})(u_i - \bar{u})}{\sum(x_i - \bar{x})^2} = \frac{\frac{1}{N}\sum(x_i - \bar{x})(u_i - \bar{u})}{\frac{1}{N}\sum(x_i - \bar{x})^2}.$$

For $N \to \infty$, we have $\frac{1}{N}\sum(x_i - \bar{x})(u_i - \bar{u}) \to 0$ and $\frac{1}{N}\sum(x_i - \bar{x})^2 \to \mathrm{var}(x)$, so the right hand side converges to zero; i.e., the OLS estimator is *consistent* ($\hat{\beta} \to \beta$).
If x_i is measured with error, this consistency result does not hold. Assume

$$x_i^* = x_i + e_i \;,$$

where e_i is a "classical measurement error" where $Ee_i = 0$, $Ee_ie_j = 0; i \neq j$ and $Ee_iu_j = 0; \forall i, j$. Now, if you regress y on x^* using the OLS formula, $\hat{\beta}$ will be biased towards zero; i.e. $E|\hat{\beta}| < E|\beta|$.

This is easy to demonstrate: We have

$$\hat{\beta} \;=\; \frac{\sum(x_i^* - \bar{x^*})(\beta(x_i - \bar{x}) + (u_i - \bar{u}))}{\sum(x_i^* - \bar{x^*})^2}$$

$$= \frac{\beta \frac{1}{N}\sum(x_i - \bar{x})(x_i - \bar{x}) + \beta \frac{1}{N}\sum(e_i - \bar{e})(x_i - \bar{x}) + \frac{1}{N}\sum(x_i^* - \bar{x^*})(u_i - \bar{u})}{\frac{1}{N}\sum(x_i - \bar{x} + e_i - \bar{e})^2} \ ,$$

where the second and third terms in the numerator converges to 0 by the law of large numbers. We then have

$$\hat{\beta} \ \approx \ \beta \frac{\frac{1}{N}\sum(x_i - \bar{x})^2}{\frac{1}{N}\sum(x_i - \bar{x})^2 + \frac{1}{N}\sum(e_i - \bar{e})^2 + \frac{1}{N}\sum((x_i - \bar{x})(e_i - \bar{e}))} \to \beta \frac{var(x)}{var(x) + var(e)} \ .$$

This demonstrates that $\hat{\beta}$ converges to the true β times a term smaller than 1.

And extended version of this note, that explains the relevance for permanent income theory, can be found off my macro II class-page.