The Consistency of OLS Estimates
ECON 30331

Bill Evans
Fall 2008

Model: (1) $y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$

We've demonstrated in class that the estimate for $\hat{\beta}_1$

$$(2) \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Recalling the properties of summations, note that numerator can be written as

$$(3) \qquad \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{n} y_i(x_i - \bar{x})$$

Note further that the true relationship between y and x is given by equation (1) above. Substituting (1) into the numerator of (2) generates

$$(4) \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i + \varepsilon_i)(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Break apart the terms in the numerator

$$(5) \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}\beta_0(x_i - \bar{x}) + \sum_{i=1}^{n}\beta_1 x_i(x_i - \bar{x}) + \sum_{i=1}^{n}\varepsilon_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

We can simplify the terms in (5) using the properties of summations:

In the first term in the numerator, note that β_0 is a constant and can be pulled outside the summation. As a result, we have the summation of a deviation from a mean, which equals zero

$$\sum_{i=1}^{n}\beta_0(x_i - \bar{x}) = \beta_0 \sum_{i-1}^{n}(x_i - \bar{x}) = \beta_0(0) = 0$$

In the second term in the numerator, β_1 is a constant and it can pulled outside the summation. Recall also that $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i(x_i - \bar{x})$ so

$$\sum_{I=1}^{n} \beta_1 x_i (x_i - \bar{x}) = \beta_1 \sum_{i=1}^{n} x_i (x_i - \bar{x}) = \beta_1 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

The first term in the numerator drops out, the second term reduces to β_1 and therefore, we can write the OLS estimate for $\hat{\beta}_1$ as

$$(6) \qquad \hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} \varepsilon_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Divide the numerator and denominator on the right hand side of (6) by (n-1)

$$(7) \qquad \hat{\beta}_1 = \beta_1 + \frac{\dfrac{\sum_{i=1}^{n} \varepsilon_i (x_i - \bar{x})}{n-1}}{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \beta_1 + \frac{\hat{\sigma}_{\varepsilon x}}{\hat{\sigma}_x^2}$$

Note that the numerator in the middle of (7) is the sample covariance of x and ε while the denominator is the sample variance of x. Re-write (7) to read. Now, what happens to the estimate $\hat{\beta}_1$ when the sample size increases.

$$(8) \qquad p\lim(\hat{\beta}_1) = \beta_1 + p\lim\left[\frac{\hat{\sigma}_{\varepsilon x}}{\hat{\sigma}_x^2}\right]$$

As $n \to \infty$, we know that $p\lim(\hat{\sigma}_{\varepsilon x}) = \sigma_{\varepsilon x}$ *and* $p\lim\left(\hat{\sigma}_x^2\right) = \sigma_x^2$, that is, $\hat{\sigma}_{\varepsilon x}$ is a consistent estimate of $\sigma_{\varepsilon x}$ and $\hat{\sigma}_x^2$ is a consistent estimate of σ_x^2. If we maintain the assumption that in large samples, $\sigma_{\varepsilon x} = 0$, then

$$(9) \qquad p\lim(\hat{\beta}_1) = \beta_1 + p\lim\left[\frac{\hat{\sigma}_{\varepsilon x}}{\hat{\sigma}_x^2}\right] = \beta_1 + \left[\frac{\sigma_{\varepsilon x}}{\sigma_x^2}\right] = \beta_1 + \left[\frac{0}{\sigma_x^2}\right] = \beta_1$$

Therefore, $\hat{\beta}_1$ is a consistent estimate.