Econometrics: Deriving OLS Estimators

Ryan Safner

Fall 2017

1 Deriving The OLS Estimators

The population linear regression model is:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

The errors (ϵ_i) are unobserved, but for candidate values of $\hat{\beta}_0$ and $\hat{\beta}_1$, we can obtain an estimate of the residual. Algebraically, the error is:

$$\hat{\epsilon_i} = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \tag{1}$$

Recall our goal is to find $\hat{\beta}_0$ and $\hat{\beta}_1$ that *minimizes* the sum of squared errors (SSE):

$$SSE = \sum_{i=1}^{n} \hat{\epsilon_i}^2 \tag{2}$$

So our minimization problem is:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \tag{3}$$

Using calculus, we take the partial derivatives and set it equal to 0 to find a minimum. The first order conditions are:

$$\frac{\partial SSE}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{4}$$

$$\frac{\partial SSE}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i = 0 \tag{5}$$

1.1 Finding $\hat{\beta}_0$

Working with the first FOC, equation 4, divide both sides by -2:

$$\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{6}$$

Then expand the summation across all terms and divide by n:

$$\underbrace{\frac{1}{n} \sum_{i=1}^{n} Y_i}_{\bar{Y}} - \underbrace{\frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_0}_{\hat{\beta}_0} - \underbrace{\frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_1 X_i}_{\hat{\beta}_1 \bar{X}} = 0 \tag{7}$$

Note the first term is $\bar{Y}$, the second is $\hat{\beta}_0$, the third is $\hat{\beta}_1\bar{X}$.[1] So we can rewrite as:

$$\bar{Y} - \hat{\beta}_0 - \beta_1 = 0 \tag{8}$$

Rearranging:

$$\hat{\beta}_0 = \bar{Y} - \bar{X}\beta_1 \tag{9}$$

1.2 Finding $\hat{\beta}_1$

To find $\hat{\beta}_1$, take the second FOC, equation 5 and divide by -2:

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)X_i = 0 \tag{10}$$

From equation 9, substitute in for $\hat{\beta}_0$:

$$\sum_{i=1}^{n}\Big(Y_i - [\bar{Y} - \hat{\beta}_1\bar{X}] - \hat{\beta}_1 X_i\Big)X_i = 0 \tag{11}$$

Combining similar terms:

$$\sum_{i=1}^{n}\Big([Y_i - \bar{Y}] - [X_i - \bar{X}]\hat{\beta}_1\Big)X_i = 0 \tag{12}$$

Distribute X_i and expand terms into the subtraction of two sums (and pull out $\hat{\beta}_1$ as a constant in the second sum:

$$\sum_{i=1}^{n}[Y_i - \bar{Y}]X_i - \hat{\beta}_1\sum_{i=1}^{n}[X_i - \bar{X}]X_i = 0 \tag{13}$$

Move the second term to the righthand side:

$$\sum_{i=1}^{n}[Y_i - \bar{Y}]X_i = \hat{\beta}_1\sum_{i=1}^{n}[X_i - \bar{X}]X_i \tag{14}$$

Divide to keep just $\hat{\beta}_1$ on the right:

$$\frac{\sum_{i=1}^{n}[Y_i - \bar{Y}]X_i}{\sum_{i=1}^{n}[X_i - \bar{X}]X_i} = \hat{\beta}_1 \tag{15}$$

Note that from the properties of summation operators:

$$\sum_{i=1}^{n}[Y_i - \bar{Y}]X_i = \sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})$$

and:

$$\sum_{i=1}^{n}[X_i - \bar{X}]X_i = \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X}) = \sum_{i=1}^{n}(X_i - \bar{X})^2$$

Plug in these two facts:

$$\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \hat{\beta}_1 \tag{16}$$

[1] From the rules about summation operators, we define the mean of a random variable X as $\bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i$. The mean of a constant, like β_0 or β_1 is itself.

1.3 Proof of Unbiasedness

Begin with equation:

$$\hat{\beta}_1 = \frac{\sum Y_i X_i}{\sum X_i^2} \tag{17}$$

Substitute for Y_i:

$$\hat{\beta}_1 = \frac{\sum (\beta_1 X_i + \epsilon_i) X_i}{\sum X_i^2} \tag{18}$$

Distribute X_i in the numerator:

$$\hat{\beta}_1 = \frac{\sum \beta_1 X_i^2 + \epsilon_i X_i}{\sum X_i^2} \tag{19}$$

Separate the sum into additive pieces:

$$\hat{\beta}_1 = \frac{\sum \beta_1 X_i^2}{\sum X_i^2} + \frac{\epsilon_i X_i}{\sum X_i^2} \tag{20}$$

β_1 is constant, so we can pull it out of the first sum:

$$\hat{\beta}_1 = \beta_1 \frac{\sum X_i^2}{\sum X_i^2} + \frac{\epsilon_i X_i}{\sum X_i^2} \tag{21}$$

Simplifying the first term, we are left with:

$$\hat{\beta}_1 = \beta_1 + \frac{\epsilon_i X_i}{\sum X_i^2} \tag{22}$$

Now if we take expectations of both sides:

$$E[\hat{\beta}_1] = E[\beta_1] + E\left[\frac{\epsilon_i X_i}{\sum X_i^2}\right] \tag{23}$$

β_1 is a constant, so the expectation of β_1 is itself.

$$E[\hat{\beta}_1] = \beta_1 + E\left[\frac{\epsilon_i X_i}{\sum X_i^2}\right] \tag{24}$$

Using the properties of expectations, we can pull out $\frac{1}{\sum X_i^2}$ as a constant:

$$E[\hat{\beta}_1] = \beta_1 + \frac{1}{\sum X_i^2} E\left[\sum \epsilon_i X_i\right] \tag{25}$$

Again using the properties of expectations, we can put the expectation inside the summation operator (the expectation of a sum is the sum of expectations):

$$E[\hat{\beta}_1] = \beta_1 + \frac{1}{\sum X_i^2} \sum E[\epsilon_i X_i] \tag{26}$$

Under the exogeneity condition, the correlation between X_i and ϵ_i is 0.

1.4 Variance of $\hat{\beta}_1$

2 Algebraic Properties of OLS Esetimators

The OLS residuals ($\hat{\epsilon}$) and predicted values $\hat{Y}$ are chosen by the minimization problem to satisfy:

1. The expected value (average) error is 0:

$$E(\epsilon_i) = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon_i} = 0$$

2. The covariance between X and the errors is 0:

$$\hat{\sigma}_{X,\epsilon} = 0$$

   Note the first two properties imply strict *exogeneity*. That is, this is only a valid model if X and ϵ are not correlated.

3. The expected predicted value of Y is equal to the expected value of Y:

$$\bar{\hat{Y}} = \frac{1}{n}\sum_{i=1}^{n}\hat{Y_i} = \bar{Y}$$

4. Total sum of squares is equal to the explained sum of squares plus sum of squared errors:

$$TSS = ESS + SSE$$

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y_i} - \bar{Y})^2 + \sum_{i=1}^{n}\epsilon^2$$

   Recall R^2 is $\frac{ESS}{TSS}$ or $1 - SSE$