Econometrics I - Basic Multiple Regression

Ryan T. Godwin

University of Manitoba

Least Squares Estimator

Our first task is to estimate the parameters of our model,

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad ; \quad \boldsymbol{\epsilon} \sim N\left[\mathbf{0}, \sigma^2 I_n\right]$$

Note that there are $(k+1)$ parameters, including σ^2.

- There are many possible procedures for estimating parameters.
- Choice should be based mostly on the "sampling properties" of the resulting estimator (to be considered later). Other considerations include computational convenience, and the ease of interpretation of the estimated model.
- To begin with, we consider one possible estimation strategy – Least Squares.

For the i^{th} data-point, we have:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i,$$

and the expected *outcome* given the explanatory variables (and assumptions) is:

$$E\left(y_i|\boldsymbol{x}_i'\right) = \boldsymbol{x}_i'\boldsymbol{\beta}.$$

We'll estimate $E\left(y_i|\boldsymbol{x}_i'\right)$ by

$$\hat{y}_i = \boldsymbol{x}_i'\boldsymbol{b}.$$

In the population, the true (unobserved) disturbance is

$$\epsilon_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}.$$

When we use $\boldsymbol{b}$ to estimate $\boldsymbol{\beta}$, there will be some "estimation error", and the value,

$$e_i = y_i - \boldsymbol{x}_i'\boldsymbol{b}$$

will be called the i^{th} "residual". So,

$$y_i = (\boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i) = (\boldsymbol{x}_i'\boldsymbol{b} + e_i) = (\hat{y}_i + e_i)$$

Question: Which terms are unobserved (from the population) and which are observed (determined by the sample)?

The Least Squares criterion

We will “choose $\boldsymbol{b}$ so as to minimize the sum of squared residuals”.
Questions:

- Why *squared* residuals?
- Why not *absolute values* of residuals?
- Why not use a “minimum distance” criterion?

Minimizing the sum of squared residuals: an optimization problem

We will solve a minimization problem using the least-squares criterion in order to derive the "Least Squares" estimator. The problem that we are trying to solve can be stated as:

$$\begin{aligned} \text{Min}_{(\boldsymbol{b})} \sum_{i=1}^{n} e_i^2 &\Leftrightarrow \text{Min}_{(\boldsymbol{b})} \left(\boldsymbol{e}'\boldsymbol{e} \right) \\ &\Leftrightarrow \text{Min}_{(\boldsymbol{b})} \left[(\boldsymbol{y} - X\boldsymbol{b})'(\boldsymbol{y} - X\boldsymbol{b}) \right] \end{aligned}$$

Now, let:

$$S = (\boldsymbol{y} - X\boldsymbol{b})'(\boldsymbol{y} - X\boldsymbol{b}) = \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{b}'X'\boldsymbol{y} - \boldsymbol{y}'X\boldsymbol{b} + \boldsymbol{b}'X'X\boldsymbol{b}$$

Note that:

$$\boldsymbol{b}'X'\boldsymbol{y} = \boldsymbol{y}'X\boldsymbol{b}$$
$$(1 \times k)(k \times n)(n \times 1) = (1 \times 1)$$

So,

$$S = \boldsymbol{y}'\boldsymbol{y} - 2\left(\boldsymbol{y}'X\right)\boldsymbol{b} + \boldsymbol{b}'\left(X'X\right)\boldsymbol{b}$$

Two rules involving the differentiation of vectors and matrices that we need are:

1. $\partial\left(\boldsymbol{a}'\boldsymbol{x}\right)/\partial\boldsymbol{x} = \boldsymbol{a}$
2. $\partial\left(\boldsymbol{x}'A\boldsymbol{x}\right)/\partial\boldsymbol{x} = 2A\boldsymbol{x}$; if A is symmetric

Applying these two results:

$$\partial S/\partial \boldsymbol{b} = \mathbf{0} - 2\left(\boldsymbol{y}'X\right)' + 2\left(X'X\right)\boldsymbol{b} = 2\left[X'X\boldsymbol{b} - X'\boldsymbol{y}\right]$$

Set this to zero (for a *turning point*):

$$X'X\boldsymbol{b} = X'\boldsymbol{y}$$
$$(k \times n)(n \times k)(k \times 1) = (k \times n)(n \times 1)$$

This gives us k equations in k unknowns, sometimes called the "normal equations". Finally, provided that $\left(X'X\right)^{-1}$ exists:

$$\boldsymbol{b} = \left(X'X\right)^{-1} X'\boldsymbol{y} \tag{1}$$

$X'X$ is $(k \times k)$, and $\text{rank}(X'X) = \text{rank}(X) = k$ (by assumption). So, $(X'X)^{-1}$ exists. Need the "full rank" assumption for the Least Squares estimator, $\boldsymbol{b}$, to *exist*.

Check - have we *minimized* S?

$$\left(\frac{\partial^2 S}{\partial \boldsymbol{b} \partial \boldsymbol{b}'}\right) = \frac{\partial}{\partial \boldsymbol{b}'}\left[2X'X\boldsymbol{b} - 2X'\boldsymbol{y}\right] = 2\left(X'X\right) \quad ; \quad \text{a } (k \times k) \text{ matrix}$$

Note that $X'X$ is at least positive *semi-definite*:

$$\boldsymbol{\eta}'\left(X'X\right)\boldsymbol{\eta} = (X\boldsymbol{\eta})'(X\boldsymbol{\eta}) = (\boldsymbol{u}'\boldsymbol{u}) = \sum_{i=1}^{n} u_i^2 \geq 0$$

If $X'X$ has full rank, it will be *p.d.*, not n.d. Our full rank assumption has two implications:

1. The Least Squares estimator, $\boldsymbol{b}$, *exists*.
2. Our optimization problem leads to the *minimization* of S, not its maximization!

Least Squares estimator in scalar form

For a population model with an intercept and a single regressor, you may have seen the following formulas used in undergraduate textbooks:

$$
\begin{aligned}
b_1 &= \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} = \frac{s_{x,y}}{s_x^2} \\
b_0 &= \overline{y} - b_1\overline{x},
\end{aligned} \tag{2}
$$

where $s_{x,y}$ is the sample covariance between x and y, and s_x^2 is the sample variance of x.

Question: Why do population models typically include an intercept and how is the intercept included in matrix form?

Method of Moments

We can instead derive the least squared estimator using the *Method of Moments* (MM).

MM relies on the principle that the *sample mean* is a good way of estimating a *population mean.*

Take the simple population model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{3}$$

and take assumptions A.3 and A.5:

$$\mathrm{E}[\epsilon_i] = 0 \quad ; \quad \mathrm{E}[x_i \epsilon_i] = 0 \tag{4}$$

A.3 and A.5 imply two *moment conditions*. Replacing expectations with sample averages, the MM estimator is identical to OLS.