Econometrics I - Basic properties of least squares

Ryan T. Godwin

University of Manitoba

Algebraic properties

First, note that the LS residuals are "orthogonal" to the regressors:

$$\underset{(k \times 1)}{X'X\boldsymbol{b} - X'\boldsymbol{y}} = \boldsymbol{0} \qquad \text{"normal equations"}$$

So,

$$-X'(\boldsymbol{y} - X\boldsymbol{b}) = -X'\boldsymbol{e} = \boldsymbol{0}$$

or,

$$\boxed{X'\boldsymbol{e} = \boldsymbol{0}} \tag{1}$$

If the model includes an intercept, then one regressor (first column of X) is a unit vector. In this case we get three further results.

Result 1: The LS residuals sum to zero

$$X'\boldsymbol{e} = \begin{pmatrix} 1 & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{nk} \end{pmatrix}' \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ x_{1k} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_i e_i \\ ? \\ ? \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

From the first element:

$$\boxed{\sum_{i=1}^{n} e_i = 0} \tag{2}$$

Result 2: The fitted regression passes through the sample mean

$$X'\boldsymbol{y} = X'X\boldsymbol{b}$$

or,

$$\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ x_{1k} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ x_{1k} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} 1 & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}.$$

So,

$$\begin{pmatrix} \sum_i y_i \\ ? \\ ? \end{pmatrix} = \begin{pmatrix} n & \sum_i x_{i2} & \dots \\ ? & \dots & ? \\ ? & \dots & ? \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}.$$

From the first row of the vector equation:

$$\sum_i y_i = nb_1 + b_2 \sum_i x_{i2} + \cdots + b_k \sum_i x_{ik}$$

or

$$\bar{y} = b_1 + b_2\overline{x}_2 + \cdots + b_k\overline{x}_k$$

Result 3: The sample mean of the fitted y-values equals the sample mean of actual y-values

the unobservable components of y_i can be replaced by estimates and residuals:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i = \boldsymbol{x}_i'\boldsymbol{b} + e_i = \widehat{y}_i + e_i.$$

So,

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \widehat{y}_i + \frac{1}{n}\sum_{i=1}^{n} e_i,$$

or,

$$\bar{y} = \overline{\hat{y}} + 0 = \overline{\hat{y}}$$

Partitioned and partial regression

Suppose the regressor matrix can be partitioned into 2 blocks:

$$\underset{(n \times 1)}{y} = \underset{(n \times k_1)\,(k_1 \times 1)}{X_1 \boldsymbol{\beta}_1} + \underset{(n \times k_2)\,(k_2 \times 1)}{X_2 \boldsymbol{\beta}_2} + \underset{(n \times 1)}{\boldsymbol{\epsilon}}$$

When multiplying or transposing matrices with a partition, the partitions behave as if they were elements in a matrix.

The model is now:

$$y = [X_1 : X_2] \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \boldsymbol{\epsilon} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

The LS estimator is still:

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

We can rewrite the LS estimator as:

$$\begin{aligned}
\boldsymbol{b} &= \left\{[X_1 : X_2]' [X_1 : X_2]\right\}^{-1} [X_1 : X_2]' \boldsymbol{y} \\
&= \left\{ \begin{bmatrix} X_1' \\ \cdots \\ X_2' \end{bmatrix} \begin{bmatrix} X_1 & : & X_2 \end{bmatrix} \right\}^{-1} \begin{bmatrix} X_1' \\ \cdots \\ X_2' \end{bmatrix} \boldsymbol{y}
\end{aligned}$$

and

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{pmatrix} X_1'\boldsymbol{y} \\ X_2'\boldsymbol{y} \end{pmatrix}.$$

The “normal equations” underlying this are

$$(X'X)\,\boldsymbol{b} = X'\boldsymbol{y}$$

or:

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix} = \begin{pmatrix} X_1'\boldsymbol{y} \\ {X_2}'\boldsymbol{y} \end{pmatrix}$$

Solve the normal equations independently for $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$!

$$X_1'X_1\boldsymbol{b}_1 + X_1'X_2\boldsymbol{b}_2 = X_1'\boldsymbol{y} \tag{3}$$

$${X_2}'X_1\boldsymbol{b}_1 + {X_2}'X_2\boldsymbol{b}_2 = {X_2}'\boldsymbol{y} \tag{4}$$

From 3:

$$(X_1' X_1)\, \boldsymbol{b}_1 = X_1' \boldsymbol{y} - X_1' X_2 \boldsymbol{b}_2,$$

or:

$$\begin{aligned} \boldsymbol{b_1} &= (X_1' X_1)^{-1} X_1' \boldsymbol{y} - (X_1' X_1)^{-1} X_1' X_2 \boldsymbol{b}_2 \\ &= ({X_1}' X_1)^{-1} [X_1' \boldsymbol{y} - X_1' X_2 \boldsymbol{b}_2] \end{aligned} \tag{5}$$

If $X_1' X_2 = 0$, then $\boldsymbol{b_1} = (X_1' X_1)^{-1} X_1' \boldsymbol{y}$.

Why do the "partial" and "full" regression estimators coincide in this case?

Now, substitute 5 into 4:

$$\left(X_2'X_1\right)\left[\left(X_1'X_1\right)^{-1}X_1'\boldsymbol{y}-\left(X_1'X_1\right)^{-1}X_1'X_2\boldsymbol{b}_2\right]+\left(X_2'X_2\right)\boldsymbol{b}_2=X_2'\boldsymbol{y},$$

or

$$\left[\left(X_2'X_2\right)-\left(X_2'X_1\right)\left(X_1'X_1\right)^{-1}\left(X_1'X_2\right)\right]\boldsymbol{b}_2=X_2'\boldsymbol{y}-\left(X_2'X_1\right)\left(X_1'X_1\right)^{-1}X_1'\boldsymbol{y}$$

and so:

$$\boldsymbol{b}_2=\left[\left(X_2'X_2\right)-\left(X_2'X_1\right)\left(X_1'X_1\right)^{-1}\left(X_1'X_2\right)\right]^{-1}\left[X_2'\left(I-X_1\left(X_1'X_1\right)^{-1}X_1'\right)\boldsymbol{y}\right]$$

Define:

$$M_1=\left(I-X_1\left(X_1'X_1\right)^{-1}X_1'\right)$$

Then, we can write:

$$\boldsymbol{b}_2 = \left(X_2{}' M_1 X_2\right)^{-1} X_2{}' M_1 \boldsymbol{y}$$

If we repeat the whole exercise, with X_1 and X_2 interchanged, we get:

$$\boldsymbol{b}_1 = \left(X_1{}' M_2 X_1\right)^{-1} X_1{}' M_2 \boldsymbol{y}$$

where:

$$M_2 = \left(I - X_2 \left(X_2' X_2\right)^{-1} X_2'\right)$$

M_1 and M_2 are idempotent matrices. For example, $M_1 M_1 = M_1 M_1' = M_1 = M_1' M_1$.

Due to the idempotent property, we can write:

$$\boldsymbol{b}_1 = \left(X_1^{*\prime} X_1^*\right)^{-1} X_1^{*\prime} \boldsymbol{y}_1^* \quad ; \quad \boldsymbol{b}_2 = \left(X_2^{*\prime} X_2^*\right)^{-1} X_2^{*\prime} \boldsymbol{y}_2^*$$

where:

$$X_1^* = M_2 X_1 \quad ; \quad X_2^* = M_1 X_2 \quad ; \quad y_1^* = M_2 y \quad ; \quad y_2^* = M_1 y$$

The Frisch-Waugh-Lovell theorem states that $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ from the above two equations correspond to the LS estimator $\boldsymbol{b}$ in the full, un-partitioned, regression model. Why are these results useful?

The M matrix is sometimes referred to as a "residual-maker" matrix. That is, $M_Q \boldsymbol{q}$ produces the LS residuals from a regression of the vector $\boldsymbol{q}$ on the matrix Q.

Let $M_1 = \left(I - X_1 \left(X_1' X_1 \right)^{-1} X_1' \right)$. Prove that $M_1 \boldsymbol{y}$ is equal to the LS residuals of a regression of $\boldsymbol{y}$ on X_1.