Some algebraic results for OLS estimation in the bivariate regression model

Suppose we have n observations on a random variable X

$$X_1, X_2, \ldots, X_n$$

and n observations on a random variable Y

$$Y_1, Y_2, \ldots, Y_n$$

The sample means are

$$\boxed{\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t} \quad \text{and} \quad \boxed{\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t}$$

When discussing OLS estimation in the bivariate regression model, we need to manipulate expressions like

$$\sum_{t=1}^{n} (X_t - \bar{X})(Y_t - \bar{Y})$$

and

$$\sum_{t=1}^{n} (X_t - \bar{X})^2$$

You must make sure you understand how to derive the following results:

Result 1:

$$\begin{aligned}\sum_{t=1}^{n} (X_t - \bar{X})(Y_t - \bar{Y}) &= \sum_{t=1}^{n} (X_t - \bar{X}) Y_t \\ &= \sum_{t=1}^{n} X_t Y_t - n\bar{X}\bar{Y}\end{aligned}$$

Result 2:

$$\begin{aligned}\sum_{t=1}^{n} (X_t - \bar{X})^2 &= \sum_{t=1}^{n} (X_t - \bar{X}) X_t \\ &= \sum_{t=1}^{n} X_t^2 - n\bar{X}^2\end{aligned}$$

Derivation of Result 1:

$$\sum_{t=1}^{n} (X_t - \overline{X})(Y_t - \overline{Y})$$

$$= \sum_{t=1}^{n} \left\{ (X_t - \overline{X}) Y_t - (X_t - \overline{X}) \overline{Y} \right\}$$

$$= \sum_{t=1}^{n} (X_t - \overline{X}) Y_t - \sum_{t=1}^{n} (X_t - \overline{X}) \overline{Y}$$

$$= \sum_{t=1}^{n} (X_t - \overline{X}) Y_t - \overline{Y} \left\{ \sum_{t=1}^{n} X_t - n\overline{X} \right\}$$

$$= \sum_{t=1}^{n} (X_t - \overline{X}) Y_t - \overline{Y} \left\{ \cancel{\sum_{t=1}^{n} X_t - \sum_{t=1}^{n} X_t} \right\}$$

$$= \sum_{t=1}^{n} (X_t - \overline{X}) Y_t$$ (FIRST PART OF RESULT 1)

$$= \sum_{t=1}^{n} \left\{ X_t Y_t - \overline{X} Y_t \right\}$$

$$= \sum_{t=1}^{n} X_t Y_t - \overline{X} \sum_{t=1}^{n} Y_t$$

$$= \sum_{t=1}^{n} X_t Y_t - n \overline{X}\,\overline{Y}$$ (SECOND PART OF RESULT 1)

Derivation of Result 2:

$$\sum_{t=1}^{n} (X_t - \overline{X})^2$$

$$= \sum_{t=1}^{n} (X_t - \overline{X})(X_t - \overline{X})$$

$$= \sum_{t=1}^{n} \left\{ (X_t - \overline{X}) X_t - (X_t - \overline{X}) \overline{X} \right\}$$

$$= \sum_{t=1}^{n} (X_t - \overline{X}) X_t - \underbrace{\overline{X} \sum_{t=1}^{n} (X_t - \overline{X})}_{\text{THIS TERM EQUALS ZERO (SEE ABOVE)}}$$

$$= \sum_{t=1}^{n} (X_t - \overline{X}) X_t$$ (FIRST PART OF RESULT 2)

$$= \sum_{t=1}^{n} X_t^2 - \overline{X} \sum_{t=1}^{n} X_t$$

$$= \sum_{t=1}^{n} X_t^2 - n \overline{X}^2$$ (SECOND PART OF RESULT 2)

In Lecture 14, we will show that for a model

$$Y_t = \alpha + \beta X_t \qquad t = 1, \ldots, n$$

the Ordinary Least Squares (OLS) estimate of β is

$$\boxed{\hat{\beta} = \frac{\sum_{t=1}^{n} (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^{n} (X_t - \bar{X})^2}}$$

Using the above results, this can be written equivalently as

$$\boxed{\hat{\beta} = \frac{\sum_{t=1}^{n} X_t Y_t - n\bar{X}\bar{Y}}{\sum_{t=1}^{n} X_t^2 - n\bar{X}^2}}$$

or

$$\boxed{\hat{\beta} = \frac{\sum_{t=1}^{n} (X_t - \bar{X}) Y_t}{\sum_{t=1}^{n} (X_t - \bar{X})^2}}$$