ECONOMETRICS NOTES

Unit 0 – LECTURE 2

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

The regression coefficients are special types of random variable. We will demonstrate this using the simple regression model in which *Y* depends on *X*. The two equations show the true model and the fitted regression.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

We will investigate the properties of the ordinary least squares (OLS) estimator of the slope coefficient, shown above.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

***Y* has two components: a nonrandom component that depends on *X* and the parameters, and the random component *u*. Since b_2 depends on *Y*, it indirectly depends on *u*.**

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

If the values of *u* in the sample had been different, we would have had different values of *Y*, and hence a different value for b_2. We can in theory decompose b_2 into its nonrandom and random components.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum (X_i - \bar{X})([\beta_1 + \beta_2 X_i + u_i] - [\beta_1 + \beta_2 \bar{X} + \bar{u}])}{\sum (X_i - \bar{X})^2}$$

The first step is to substitute for *Y* and its sample mean from the true model.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum (X_i - \bar{X})([\beta_1 + \beta_2 X_i + u_i] - [\beta_1 + \beta_2 \bar{X} + \bar{u}])}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum (X_i - \bar{X})(\beta_2 (X_i - \bar{X}) + u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

The β_1 terms in the second factor cancel. We rearrange the remaining terms.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(\beta_2 (X_i - \bar{X}) + u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum \beta_2 (X_i - \bar{X})^2 + (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

We expand the numerator (multiply through out the brackets)

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(\beta_2 (X_i - \bar{X}) + u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum \beta_2 (X_i - \bar{X})^2 + (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$= \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

Hence we decompose b_2 into the true value β_2 and an error term that depends on the values of *X* and *u*.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(\beta_2 (X_i - \bar{X}) + u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum \beta_2 (X_i - \bar{X})^2 + (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$= \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

The error term depends on the value of the disturbance term in every observation in the sample, and thus it is a special type of random variable.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(\beta_2 (X_i - \bar{X}) + u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum \beta_2 (X_i - \bar{X})^2 + (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$= \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

The error term is responsible for the variations of b_2 around its fixed component β_2. If we wish, we can express the decomposition more tidily.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

This is the decomposition so far.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X})u_i - \sum (X_i - \bar{X})\bar{u}$$

The next step is to make a small simplification of the numerator of the error term. First, we expand it as shown.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X})u_i - \sum (X_i - \bar{X})\bar{u}$$

$$= \sum (X_i - \bar{X})u_i - \bar{u}\sum (X_i - \bar{X})$$

The mean value of *u* is a common factor of the second summation, so it can be taken outside.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\begin{aligned}\sum (X_i - \bar{X})(u_i - \bar{u}) &= \sum (X_i - \bar{X})u_i - \sum (X_i - \bar{X})\bar{u} \\ &= \sum (X_i - \bar{X})u_i - \bar{u}\sum (X_i - \bar{X}) \\ &= \sum (X_i - \bar{X})u_i\end{aligned}$$

$$\sum (X_i - \bar{X}) = \left(\sum X_i\right) - n\bar{X} = n\bar{X} - n\bar{X} = 0$$

$$\bar{X} = \frac{\sum X_i}{n}$$

The second term then vanishes because the sum of the deviations of *X* around its sample mean is automatically zero. Why is this the case?

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X})u_i$$

$$\Delta = \sum (X_j - \bar{X})^2$$

$$b_2 = \beta_2 + \frac{\sum (X_i - \bar{X})u_i}{\Delta}$$

Thus we can rewrite the decomposition as shown. For convenience, the denominator of the error term has been denoted Δ.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X})u_i$$

$$\Delta = \sum (X_j - \bar{X})^2$$

$$b_2 = \beta_2 + \frac{\sum (X_i - \bar{X})u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum (X_i - \bar{X})u_i$$

A further small rearrangement of the expression for the error term.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X})u_i$$

$$\Delta = \sum (X_j - \bar{X})^2$$

$$b_2 = \beta_2 + \frac{\sum (X_i - \bar{X})u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum (X_i - \bar{X})u_i$$

$$= \beta_2 + \sum \left(\frac{1}{\Delta}\right)(X_i - \bar{X})u_i$$

Another re-arrangement.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X})u_i$$

$$\Delta = \sum (X_j - \bar{X})^2$$

$$b_2 = \beta_2 + \frac{\sum (X_i - \bar{X})u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum (X_i - \bar{X})u_i$$

$$= \beta_2 + \sum \left(\frac{1}{\Delta}\right)(X_i - \bar{X})u_i = \beta_2 + \sum \left(\frac{X_i - \bar{X}}{\Delta}\right)u_i$$

One more re-arrangement.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X})u_i$$

$$\Delta = \sum (X_j - \bar{X})^2$$

$$b_2 = \beta_2 + \frac{\sum (X_i - \bar{X})u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum (X_i - \bar{X})u_i$$

$$= \beta_2 + \sum \left(\frac{1}{\Delta}\right)(X_i - \bar{X})u_i = \beta_2 + \sum \left(\frac{X_i - \bar{X}}{\Delta}\right)u_i$$

$$= \beta_2 + \sum a_i u_i$$

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum (X_j - \bar{X})^2}$$

Thus we have shown that b_2 is equal to the true value and plus a weighted linear combination of the values of the disturbance term in the sample, where the weights are functions of the values of *X* in the observations in the sample.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X})u_i$$

$$\Delta = \sum (X_j - \bar{X})^2$$

$$b_2 = \beta_2 + \frac{\sum (X_i - \bar{X})u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum (X_i - \bar{X})u_i$$

$$= \beta_2 + \sum \left(\frac{1}{\Delta}\right)(X_i - \bar{X})u_i = \beta_2 + \sum \left(\frac{X_i - \bar{X}}{\Delta}\right)u_i$$

$$= \beta_2 + \sum a_i u_i$$

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum (X_j - \bar{X})^2}$$

As you can see, every value of the disturbance term in the sample affects the sample value of b_2.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X})u_i$$

$$\Delta = \sum (X_j - \bar{X})^2$$

$$b_2 = \beta_2 + \frac{\sum (X_i - \bar{X})u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum (X_i - \bar{X})u_i$$

$$= \beta_2 + \sum \left(\frac{1}{\Delta}\right)(X_i - \bar{X})u_i = \beta_2 + \sum \left(\frac{X_i - \bar{X}}{\Delta}\right)u_i$$

$$= \beta_2 + \sum a_i u_i$$

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum (X_j - \bar{X})^2}$$

Before moving on, it may be helpful to clarify a mathematical technicality. In the summation in the denominator of the expression for a_i, the subscript has been changed to j. Why?

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2} = \frac{X_i - \bar{X}}{\left(X_1 - \bar{X}\right)^2 + \ldots + \left(X_n - \bar{X}\right)^2}$$

$$\Delta = \sum\left(X_j - \bar{X}\right)^2$$

$$b_2 = \beta_2 + \frac{\sum\left(X_i - \bar{X}\right)u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum\left(X_i - \bar{X}\right)u_i$$

$$= \beta_2 + \sum\left(\frac{1}{\Delta}\right)\left(X_i - \bar{X}\right)u_i = \beta_2 + \sum\left(\frac{X_i - \bar{X}}{\Delta}\right)u_i$$

$$= \beta_2 + \sum a_i u_i$$

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum\left(X_j - \bar{X}\right)^2}$$

The denominator is the sum, from 1 to *n*, of the squared deviations of *X* from its sample mean. This is made explicit in the version of the expression in the box at the top of the slide.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2} = \frac{X_i - \bar{X}}{\left(X_1 - \bar{X}\right)^2 + \ldots + \left(X_n - \bar{X}\right)^2}$$

$$\Delta = \sum\left(X_j - \bar{X}\right)^2$$

$$b_2 = \beta_2 + \frac{\sum\left(X_i - \bar{X}\right)u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum\left(X_i - \bar{X}\right)u_i$$

$$= \beta_2 + \sum\left(\frac{1}{\Delta}\right)\left(X_i - \bar{X}\right)u_i = \beta_2 + \sum\left(\frac{X_i - \bar{X}}{\Delta}\right)u_i$$

$$= \beta_2 + \sum a_i u_i$$

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum\left(X_j - \bar{X}\right)^2}$$

Written this way, the meaning of the denominator is clear, but the form is clumsy. Obviously, we should use Σ–notation to compress it.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2} = \frac{X_i - \bar{X}}{\left(X_1 - \bar{X}\right)^2 + \ldots + \left(X_n - \bar{X}\right)^2}$$

$$\Delta = \sum\left(X_j - \bar{X}\right)^2$$

$$b_2 = \beta_2 + \frac{\sum\left(X_i - \bar{X}\right)u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum\left(X_i - \bar{X}\right)u_i$$

$$= \beta_2 + \sum\left(\frac{1}{\Delta}\right)\left(X_i - \bar{X}\right)u_i = \beta_2 + \sum\left(\frac{X_i - \bar{X}}{\Delta}\right)u_i$$

$$= \beta_2 + \sum a_i u_i$$

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum\left(X_j - \bar{X}\right)^2}$$

For the Σ-notation, we need to choose an index symbol that changes as we go from the first squared deviation to the last. We can use anything we like, EXCEPT *i*, because we are already using *i* for a completely different purpose in the numerator.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2} = \frac{X_i - \bar{X}}{\left(X_1 - \bar{X}\right)^2 + \ldots + \left(X_n - \bar{X}\right)^2}$$

$$\Delta = \sum\left(X_j - \bar{X}\right)^2$$

$$b_2 = \beta_2 + \frac{\sum\left(X_i - \bar{X}\right)u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum\left(X_i - \bar{X}\right)u_i$$

$$= \beta_2 + \sum\left(\frac{1}{\Delta}\right)\left(X_i - \bar{X}\right)u_i = \beta_2 + \sum\left(\frac{X_i - \bar{X}}{\Delta}\right)u_i$$

$$= \beta_2 + \sum a_i u_i$$

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum\left(X_j - \bar{X}\right)^2}$$

We have used *j* here, but this was quite arbitrary. We could have used anything for the summation index (except *i*), as long as the meaning is clear. We could have used a smiley ☻ instead (please don't).

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X})u_i$$

$$\Delta = \sum (X_j - \bar{X})^2$$

$$b_2 = \beta_2 + \frac{\sum (X_i - \bar{X})u_i}{\Delta} = \beta_2 + \frac{1}{\Delta}\sum (X_i - \bar{X})u_i$$

$$= \beta_2 + \sum \left(\frac{1}{\Delta}\right)(X_i - \bar{X})u_i = \beta_2 + \sum \left(\frac{X_i - \bar{X}}{\Delta}\right)u_i$$

$$= \beta_2 + \sum a_i u_i$$

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum (X_j - \bar{X})^2}$$

The error term depends on the value of the disturbance term in every observation in the sample, and thus it is a special type of random variable.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$\sum (X_i - \bar{X})(u_i - \bar{u}) = \sum (X_i - \bar{X}) u_i$$

$$\Delta = \sum (X_j - \bar{X})^2$$

$$b_2 = \beta_2 + \frac{\sum (X_i - \bar{X}) u_i}{\Delta} = \beta_2 + \frac{1}{\Delta} \sum (X_i - \bar{X}) u_i$$

$$= \beta_2 + \sum \left(\frac{1}{\Delta} \right) (X_i - \bar{X}) u_i = \beta_2 + \sum \left(\frac{X_i - \bar{X}}{\Delta} \right) u_i$$

$$= \beta_2 + \sum a_i u_i$$

$$a_i = \frac{X_i - \bar{X}}{\Delta} = \frac{X_i - \bar{X}}{\sum (X_j - \bar{X})^2}$$

We will show that the error term has expected value zero, and hence that the ordinary least squares (OLS) estimator of the slope coefficient in a simple regression model is unbiased.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \sum a_t u_i$$

$$a_i = \frac{X_i - \bar{X}}{\sum_{j=1}^{n} (X_j - \bar{X})^2}$$

$$E(b_2) = E(\beta_2) + E\left(\sum a_i u_i\right)$$

The expected value of b_2 is equal to the expected value of β_2 and the expected value of the weighted sum of the values of the disturbance term.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \sum a_t u_i$$

$$E\left(\sum a_i u_i\right) = E\left(a_1 u_1 + \ldots + a_n u_n\right) = E\left(a_1 u_1\right) + \ldots + E\left(a_n u_n\right) = \sum E\left(a_i u_i\right)$$

$$E\left(b_2\right) = E\left(\beta_2\right) + E\left(\sum a_i u_i\right)$$

$$= \beta_2 + \sum E\left(a_i u_i\right)$$

β_2 is fixed so it is unaffected by taking expectations. The first expectation rule (supplementary notes) states that the expectation of a sum of several quantities is equal to the sum of their expectations.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \sum a_t u_i$$

$$a_i = \frac{X_i - \bar{X}}{\sum_{j=1}^{n} (X_j - \bar{X})^2}$$

$$E(b_2) = E(\beta_2) + E\left(\sum a_i u_i\right)$$

$$= \beta_2 + \sum E(a_i u_i) = \beta_2 + \sum a_i E(u_i)$$

Now for each *i*, $E(a_i u_i) = a_i E(u_i)$. This is a really important step and we can make it only with Model A.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \sum a_t u_i$$

$$a_i = \frac{X_i - \bar{X}}{\sum_{j=1}^{n} (X_j - \bar{X})^2}$$

$$E(b_2) = E(\beta_2) + E\left(\sum a_i u_i\right)$$

$$= \beta_2 + \sum E(a_i u_i) = \beta_2 + \sum a_i E(u_i)$$

Under Model A, we are assuming that the values of *X* in the observations are nonstochastic. It follows that each a_i is nonstochastic, since it is just a combination of the values of *X*.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \sum a_t u_i$$

$$a_i = \frac{X_i - \bar{X}}{\sum_{j=1}^{n} (X_j - \bar{X})^2}$$

$$E(b_2) = E(\beta_2) + E\left(\sum a_i u_i\right)$$

$$= \beta_2 + \sum E(a_i u_i) = \beta_2 + \sum a_i E(u_i)$$

Thus it can be treated as a constant, allowing us to take it out of the expectation using the second expected value rule (Supplementary notes).

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \sum a_t u_i$$

$$a_i = \frac{X_i - \bar{X}}{\sum_{j=1}^{n} (X_j - \bar{X})^2}$$

$$E(b_2) = E(\beta_2) + E\left(\sum a_i u_i\right)$$

$$= \beta_2 + \sum E(a_i u_i) = \beta_2 + \sum a_i E(u_i) = \beta_2$$

Under CLR Assumptions, $E(u_i) = 0$ for all i, and so the estimator is unbiased. The proof of the unbiasedness of the estimator of the intercept will be left as an exercise.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

It is important to realize that the OLS estimators of the parameters are not the only unbiased estimators. We will give an example of another.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

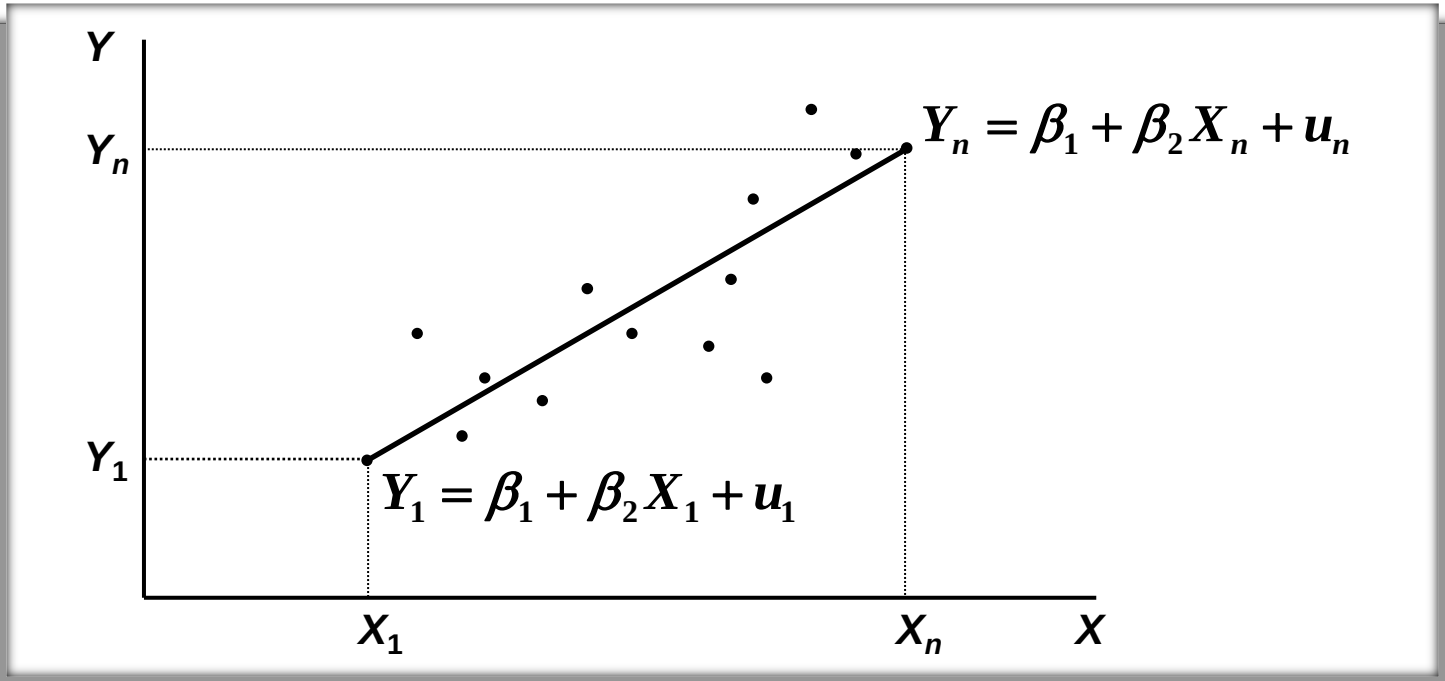
True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{Y_n - Y_1}{X_n - X_1}$$



Someone who had never heard of regression analysis, seeing a scatter diagram of a sample of observations, might estimate the slope by joining the first and the last observations, and dividing the increase in the height by the horizontal distance between them.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

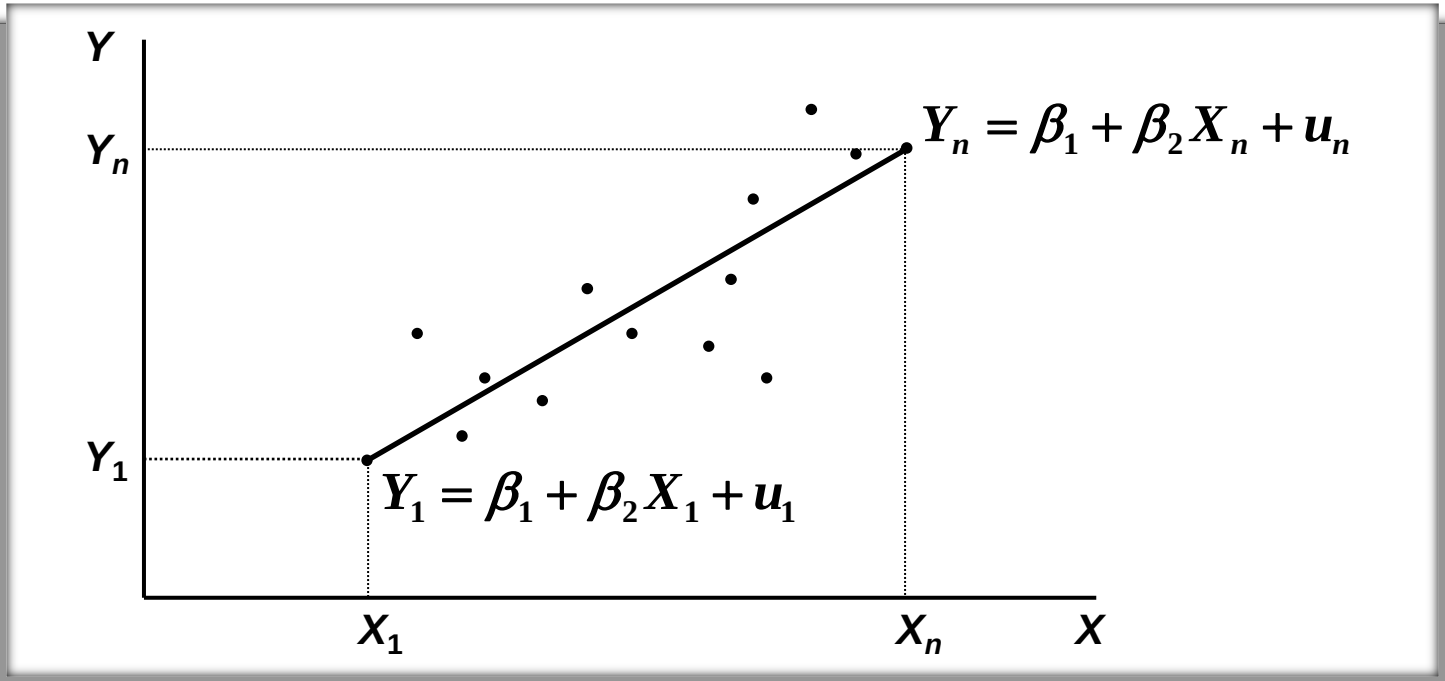
True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{Y_n - Y_1}{X_n - X_1}$$



The estimator is thus (Y_n–Y_1) divided by (X_n–X_1). We will investigate whether it is biased or unbiased.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

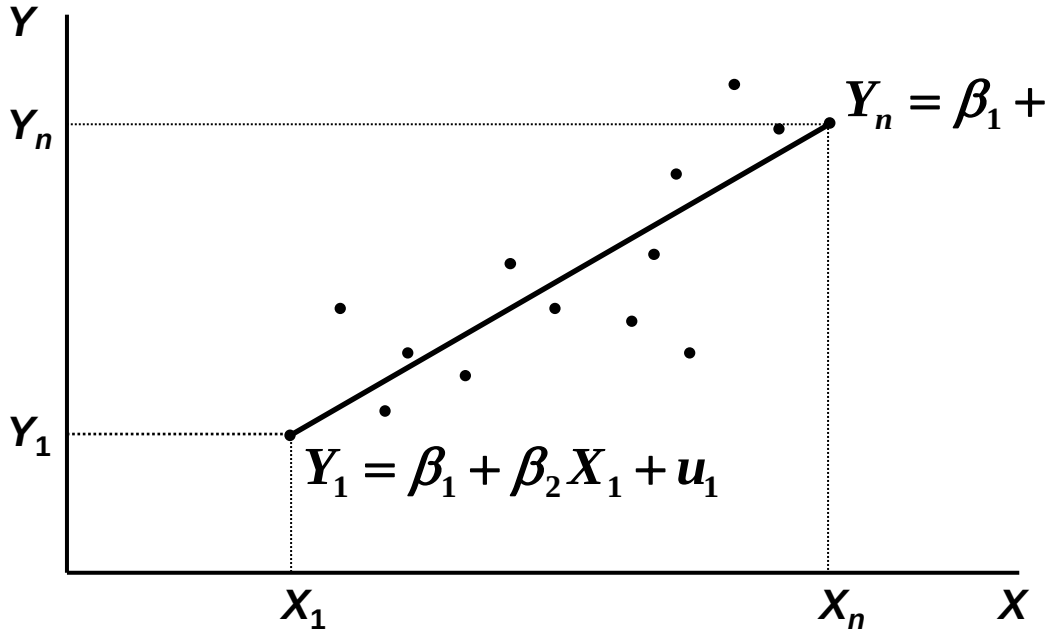
True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{Y_n - Y_1}{X_n - X_1} = \frac{(\beta_1 + \beta_2 X_n + u_n) - (\beta_1 + \beta_2 X_1 + u_1)}{X_n - X_1}$$



To do this, we start by substituting for the *Y* components in the expression.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{Y_n - Y_1}{X_n - X_1} = \frac{(\beta_1 + \beta_2 X_n + u_n) - (\beta_1 + \beta_2 X_1 + u_1)}{X_n - X_1}$$

$$= \frac{\beta_2 (X_n - X_1) + (u_n - u_1)}{X_n - X_1} = \beta_2 + \frac{u_n - u_1}{X_n - X_1}$$

The β_1 terms cancel out and the rest of the expression simplifies as shown. Thus we have decomposed this naïve estimator into two components, the true value and an error term. This decomposition is parallel to that for the OLS estimator, but the error term is different.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{Y_n - Y_1}{X_n - X_1} = \frac{(\beta_1 + \beta_2 X_n + u_n) - (\beta_1 + \beta_2 X_1 + u_1)}{X_n - X_1}$$

$$= \frac{\beta_2 (X_n - X_1) + (u_n - u_1)}{X_n - X_1} = \beta_2 + \frac{u_n - u_1}{X_n - X_1}$$

$$E(b_2) = E(\beta_2) + E\left(\frac{u_n - u_1}{X_n - X_1}\right)$$

We now take expectations to investigate unbiasedness.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{Y_n - Y_1}{X_n - X_1} = \frac{(\beta_1 + \beta_2 X_n + u_n) - (\beta_1 + \beta_2 X_1 + u_1)}{X_n - X_1}$$

$$= \frac{\beta_2 (X_n - X_1) + (u_n - u_1)}{X_n - X_1} = \beta_2 + \frac{u_n - u_1}{X_n - X_1}$$

$$E(b_2) = E(\beta_2) + E\left(\frac{u_n - u_1}{X_n - X_1}\right)$$

$$= \beta_2 + \frac{1}{X_n - X_1} E(u_n - u_1)$$

The denominator of the error term can be taken outside because the values of *X* are nonstochastic.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{Y_n - Y_1}{X_n - X_1} = \frac{(\beta_1 + \beta_2 X_n + u_n) - (\beta_1 + \beta_2 X_1 + u_1)}{X_n - X_1}$$

$$= \frac{\beta_2 (X_n - X_1) + (u_n - u_1)}{X_n - X_1} = \beta_2 + \frac{u_n - u_1}{X_n - X_1}$$

$$E(b_2) = E(\beta_2) + E\left(\frac{u_n - u_1}{X_n - X_1}\right)$$

$$= \beta_2 + \frac{1}{X_n - X_1} E(u_n - u_1) = \beta_2$$

Given CLR Assumption, the expectations of u_n and u_1 are zero. Therefore, despite being naïve, this estimator is unbiased.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{Y_n - Y_1}{X_n - X_1} = \frac{(\beta_1 + \beta_2 X_n + u_n) - (\beta_1 + \beta_2 X_1 + u_1)}{X_n - X_1}$$

$$= \frac{\beta_2 (X_n - X_1) + (u_n - u_1)}{X_n - X_1} = \beta_2 + \frac{u_n - u_1}{X_n - X_1}$$

$$E(b_2) = E(\beta_2) + E\left(\frac{u_n - u_1}{X_n - X_1}\right)$$

$$= \beta_2 + \frac{1}{X_n - X_1} E(u_n - u_1) = \beta_2$$

It is intuitively easy to see that we would not prefer the naïve estimator to OLS. Unlike OLS, which takes account of every observation, it employs only the first and the last and is wasting most of the information in the sample.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{Y_n - Y_1}{X_n - X_1} = \frac{(\beta_1 + \beta_2 X_n + u_n) - (\beta_1 + \beta_2 X_1 + u_1)}{X_n - X_1}$$

$$= \frac{\beta_2 (X_n - X_1) + (u_n - u_1)}{X_n - X_1} = \beta_2 + \frac{u_n - u_1}{X_n - X_1}$$

$$E(b_2) = E(\beta_2) + E\left(\frac{u_n - u_1}{X_n - X_1}\right)$$

$$= \beta_2 + \frac{1}{X_n - X_1} E(u_n - u_1) = \beta_2$$

The naïve estimator will be sensitive to the value of the disturbance term *u* in those two observations, whereas the OLS estimator combines all the disturbance term values and takes greater advantage of the possibility that to some extent they cancel each other out.

RANDOM COMPONENTS, UNBIASEDNESS OF THE REGRESSION COEFFICIENTS

True model

$$Y = \beta_1 + \beta_2 X + u$$

Fitted model

$$\hat{Y} = b_1 + b_2 X$$

$$b_2 = \frac{Y_n - Y_1}{X_n - X_1} = \frac{(\beta_1 + \beta_2 X_n + u_n) - (\beta_1 + \beta_2 X_1 + u_1)}{X_n - X_1}$$

$$= \frac{\beta_2 (X_n - X_1) + (u_n - u_1)}{X_n - X_1} = \beta_2 + \frac{u_n - u_1}{X_n - X_1}$$

$$E(b_2) = E(\beta_2) + E\left(\frac{u_n - u_1}{X_n - X_1}\right)$$

$$= \beta_2 + \frac{1}{X_n - X_1} E(u_n - u_1) = \beta_2$$

More rigorously, it can be shown that the population variance of the naïve estimator is greater than that of the OLS estimator, and that the naïve estimator is therefore less efficient.

END OF LECTURE