Bias and Efficiency of Least Squares Regression Coefficients

In addition to the overall fit of the model, we now need to ask how accurate each individual OLS coefficient estimate is.

To do this need to make some assumptions about the behaviour of the (true) residual term that underlies our view of the world (Gauss-Markov assumptions)

1. \( E(u_i) = 0\)
   - the expected (average or mean) value of the residual is zero
   - sometimes positive, sometimes negative, but there is never any systematic behaviour in this random variable so that on average its value is zero

2. \( \text{Var}(u / X) = \sigma^2 = \text{constant for all } X \text{ values in the data set} \) (homoskedasticity)
   - think of a value of the X variable and look at the different values of the residual at this value of X. The distribution of these residual values around this point should be no different than the distribution of residual values at any other value of X
   
   This is a useful assumption since it implies that no particular value of x carries any more information about the behaviour of Y than any other

3. \( \text{Cov}(u_i, u_j) = 0 \) for all \( i \neq j \)
   (no autocorrelation)
   - there is no systematic association between values of the residuals so knowledge of the value of one residual imparts no information about the value of any other residual.

4. \( \text{Cov}(X, u_i) = 0 \)
   - there is zero covariance (association) between the residual and any value of X – ie X and the residual are independent
   
   This means that we can assess the individual contributions of X and u to explaining Y

   (Note that this assumption is automatically satisfied if X is non-stochastic ie non-random so can be treated like a constant, measured with certainty)

Given these 4 assumptions we can proceed to establish the properties of OLS estimates
The 1st desirable feature of any estimate of any coefficient is that it should, on average, be as accurate an estimate of the true coefficient as possible. Accuracy in this context is given by the “bias”

This means that we would like the expected, or average, value of the estimator to equal the true (unknown) value for the population of interest

\[ E(\hat{\beta}) = \beta \]

ie if continually re-sampled and re-estimated the same model and plotted the distribution of estimates then would expect the mean value of these estimates to equal the true value (which would only be obtained if sampled everyone in the relevant population)

Given the OLS estimate of the slope

\[ \hat{b}_1 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \]

Using the rules of covariance

\[ \text{Cov}(X,Y) = \text{Cov}(X, b_0 + b_1 X + u) = \text{Cov}(X, b_0) + \text{Cov}(X, b_1 X) + \text{Cov}(X, u) \]

since \( b_0 \) is a constant it has no variance no matter the value of \( X \), so \( \text{Cov}(X, b_0) = 0 \)

since \( b_1 \) is a constant can take it outside the bracket

so \( \text{Cov}(X, b_1 X) = b_1 \text{Cov}(X, X) = b_1 \text{Var}(X) \)

Hence

\[ \text{Cov}(X,Y) = b_1 \text{Var}(X) + \text{Cov}(X,u) \]

Sub. this into OLS slope formula

\[ \hat{b}_1 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = \frac{b_1 \text{Var}(X) + \text{Cov}(X,u)}{\text{Var}(X)} \]

\[ \hat{b}_1 = \frac{b_1 \text{Var}(X)}{\text{Var}(X)} + \frac{\text{Cov}(X,u)}{\text{Var}(X)} \]

which since \( \text{Cov}(X,u) \) is assumed =0, implies that

\[ \hat{b}_1 = \frac{b_1 \text{Var}(X)}{\text{Var}(X)} = b_1 \]

Now need expected values to establish the extent of any bias. It follows that
1) \( E(\hat{b}_1) = b_1 \)

so that, on average, the OLS estimate of the slope will be equal to the true (unknown) value

ie OLS estimates are unbiased

don’t need to sample entire population since OLS on a sub-sample will give an unbiased estimate of the truth

(Can show unbiased property also holds for ols estimate of constant – see problem set 2)

**Precision of OLS Estimates**

If OLS done a 100 different (random) samples would not expect to get same result every time – but the average of those estimates would equal the true value

Measure *efficiency* of any estimate by its dispersion –

- based on the variance (or more usually its square root – the standard error)

Given 2 (unbiased) estimates will prefer the one whose range of estimates are more concentrated around the true value.

Can show (see Gujarati Chap. 3 for proof) that

Variance of the OLS estimates of the intercept and the slope are

\[
\text{Var}(\hat{\beta}_0) = \frac{\sigma_u^2}{N} \left[ 1 + \frac{\bar{X}^2}{\text{Var}(X)} \right]
\]

and

\[
\text{Var}(\hat{\beta}_1) = \frac{\sigma_u^2}{N\text{Var}(X)}
\]

(where \( \sigma_u^2 = \text{Var}(u) = \text{variance of true (not estimated) residuals} \)

This formula makes intuitive sense since

i) the variance of the OLS estimate of the slope is proportional to the variance of the residuals, \( \sigma_u^2 \)

- the more there is random unexplained behaviour in the population, the less precise the estimates
ii) the larger the sample size, N, the lower (the more efficient) the variance of the OLS estimate
   - more information means estimates likely to be more precise

iii) the larger the variance in the X variable the more precise (efficient) the OLS estimates
   - the more variation in X the more likely it is to capture any variation in the Y variable

In practice never know variation of true residuals. Can show, however, that an unbiased estimate of $\sigma^2_u$ is given by

$$s^2 = \frac{N}{N-k} \text{Var}(\hat{u}) = \frac{RSS}{N-k} = \sum_{i=1}^{N} \hat{u}_i^2$$

Sub. this into (1) and (2) gives the formula for the precision of the OLS estimates.

At same time usual to take square root to give standard errors of the estimates (standard deviation refers to the known variance, standard error refers to the estimated variance)

(3) $$s.e.(\hat{\beta}_0) = \sqrt{\frac{s^2}{N} \left(1 + \frac{\bar{X}^2}{\text{Var}(X)}\right)}$$

and

(4) $$s.e.(\hat{\beta}_1) = \sqrt{\frac{s^2}{N \text{Var}(X)}}$$

(*** learn this ***)

**Gauss-Markov Theorem**

Given Gauss-Markov assumptions 1-4 (see above) hold, then can prove a very important result:
that OLS estimates will have the smallest variance of all (linear) unbiased estimators

- there may be other ways of obtaining unbiased estimates, but OLS estimates will have the smallest standard errors