Econ 329 - Statistical Properties of the OLS estimator

Sanjaya DeSilva

September 22, 2008
1 Overview

Recall that the true regression model is

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]  

(1)

Applying the OLS method to a sample of data, we estimate the sample regression function

\[ Y_i = b_0 + b_1 X_i + e_i \]  

(2)

where the OLS estimators are,

\[
\begin{align*}
b_1 &= \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \\
b_0 &= \bar{Y} - b_1 \bar{X}
\end{align*}
\]

2 Unbiasedness

The OLS estimate \( b_1 \) is simply a sample estimate of the population parameter \( \beta_1 \). For every random sample we draw from the population, we will get a different \( b_1 \). What then is the relationship between the \( b_1 \) we obtain from a random sample and the underlying \( \beta_1 \) of the population?

To see this, start by rewriting the OLS estimator as follows;

\[
\begin{align*}
b_1 &= \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \\
&= \frac{1}{\sum_{i=1}^{n} x_i^2} \sum_{i=1}^{n} x_i (Y_i - \bar{Y}) \\
&= \frac{1}{\sum_{i=1}^{n} x_i^2} (\sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} x_i \bar{Y})
\end{align*}
\]
\[
\sum_{i=1}^{n} x_i = 1 
\]

For \( Y_i \), we can substitute the expression for the true regression line in order to obtain a relationship between \( b_1 \) and \( \beta_1 \).

\[
b_1 = \frac{1}{\sum_{i=1}^{n} x_i^2} \left( \sum_{i=1}^{n} x_i (\beta_0 + \beta_1 X_i + u_i) \right) 
\]

\[
= \frac{1}{\sum_{i=1}^{n} x_i^2} \left( \sum_{i=1}^{n} x_i \beta_0 + \beta_1 \sum_{i=1}^{n} x_i X_i + \sum_{i=1}^{n} x_i u_i \right) 
\]

\[
= \beta_1 + \sum_{i=1}^{n} k_i u_i 
\]

where

\[
k_i = \frac{x_i}{\sum_{i=1}^{n} x_i^2} 
\]

From this expression, we see that \( b_1 \) and \( \beta_1 \) are in fact different. However, we can demonstrate that, under certain assumptions, the average \( b_1 \) in repeated sampling would equal \( \beta_1 \). To see this, take the expectation of both sides of the above expression,

\[
E(b_1) = \beta_1 + E(\sum_{i=1}^{n} k_i u_i) 
\]

If we assume that \( X_i \), and therefore \( k_i \) is non-stochastic, we can rewrite this as

\[
E(b_1) = \beta_1 + \sum_{i=1}^{n} k_i E(u_i) 
\]
If we also assume that $E(u_i) = 0$, we get

$$E(b_1) = \beta_1$$  \hspace{1cm} (14)

When the expectation of the sample estimate equals the true parameter, we say that the estimator is unbiased. To recap, we find that if $X_i$ is non-stochastic and $E(u_i) = 0$, the OLS estimator is unbiased.

However, note that these two conditions are not necessary for unbiasedness. Suppose $k_i$ is not stochastic. Then, $b_1$ is an unbiased if,

$$E(\sum_{i=1}^{n} k_iu_i) = (\sum_{i=1}^{n} E(k_iu_i)) = 0$$  \hspace{1cm} (15)

That is, if $X$ and $u$ are uncorrelated, the OLS estimator is unbiased.

### 3 Variance of the Coefficient Estimate

The variance of the $b_1$ sampling distribution is, by definition

$$Var(b_1) = E[\{b_1 - E(b_1)\}^2]$$  \hspace{1cm} (16)

We showed in the previous section that, under certain classical assumptions, $E(b_1) = \beta_1$.

Then,

$$Var(b_1) = E[\{b_1 - \beta_1\}^2] = E[\sum_{i=1}^{n} k_i^2u_i^2]$$  \hspace{1cm} (17)

Expanding terms, we get

$$Var(b_1) = E[\sum_{i=1}^{n} k_i^2u_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} 2k_i k_j u_i u_j]$$  \hspace{1cm} (18)

$$= \sum_{i=1}^{n} k_i^2E(u_i^2) + \sum_{i=1}^{n} \sum_{j \neq i} 2k_i k_j E(u_i u_j)$$  \hspace{1cm} (19)
If we make the following two additional assumptions,

1. The variance of the error term is constant, i.e. $Var(u_i) = E[u_i^2] = \sigma^2$

2. The error terms of different observations are not correlated with each other or the covariance between all error terms is zero, i.e. $E(u_iu_j) = 0$ for all $i \neq j$

The expression for the variance of $b_1$ reduces to the following elegant form,

$$Var(b_1) = \frac{n}{\sum_{i=1}^{n} x_i^2} \sigma^2$$

(20)

Note that the variance of the slope coefficient depends on two things. Variance of the slope coefficient increases as

1. The variance of the error term increases

2. The sum of squared variation in the independent variable decreases, i.e. the X variable is clustered around the mean.

3.1 Estimate of The Variance of the Error Term

Even though the above expression is elegant, it is impossible to compute the variance of the slope estimate because we don’t know the variance of the underlying error term. We get around this problem by estimating the variance of the error term, i.e. $\sigma^2$ using the residuals obtained from OLS. It can be shown that, under certain classical assumptions,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n - 2}$$

(21)
is an unbiased estimator of $\sigma^2$, i.e.

$$E[\hat{\sigma}^2] = E\left[ \frac{\sum_{i=1}^{n} e_i^2}{n - 2} \right] = \sigma^2$$  \hspace{1cm} (22)

For the formal proof, see Gujarati Appendix. Note that this proof also depends crucially on the classical assumptions.

Note that the denominator of this unbiased estimator is the SSR. The estimator itself is often called the Mean Square Residual. The square root of the estimator is called the standard error of the regression (SER) and is typically used as an estimate of the standard deviation of the error term.

## 4 The Efficiency of the OLS estimator

Under the classical assumptions, the OLS estimator $b_1$ can be written as a linear function of $Y$;

$$b_1 = \sum k_i Y_i$$  \hspace{1cm} (23)

where

$$k_i = \frac{x_i}{\sum \bar{x}_i^2}$$  \hspace{1cm} (24)

Our goal now is to show that this OLS estimator has a lower variance than any other linear estimator, i.e. the OLS estimator is efficient or best. To do so, consider any other linear unbiased estimator,

$$b_1^* = \sum w_i Y_i$$  \hspace{1cm} (25)

where $w_i$ is some other function of the two variables.
The expected value of this estimator is,

\[ E(b_1^*) = \sum w_i E(Y_i) = \beta_0 \sum w_i + \beta_1 \sum w_i X_i \]  

(26)

Because \( b_1^* \) is unbiased,

\[ E(b_1^*) = \beta_1 \]  

(27)

For this to be the case, it follows that,

\[ \sum w_i = 0 \]  

(28)

\[ \sum w_i X_i = 1 \]  

(29)

It follows from these two identities that,

\[ \sum w_i x_i = \sum w_i (X_i - \bar{X}) = \sum w_i X_i - \bar{X} \sum w_i = 1 \]  

(30)

The variance of \( b_1^* \) is

\[ Var(b_1^*) = Var(\sum w_i Y_i) = \sum w_i^2 Var(Y_i) = \sigma^2 \sum w_i^2 \]  

(31)

If we rewrite the variance as,

\[ Var(b_1^*) = \sigma^2 \sum (w_i - k_i + k_i)^2 \]  

(32)

and expand this expression

\[ Var(b_1^*) = \sigma^2 \left( \sum (w_i - k_i)^2 + \sum k_i^2 + 2 \sum k_i (w_i - k_i) \right) \]  

(33)

Note that,

\[ \sum k_i w_i = \frac{\sum x_i w_i}{\sum x_i^2} = \frac{1}{\sum x_i^2} \]  

(34)

under the unbiasedness assumption made earlier.
In addition,

\[ \sum k_i^2 = \frac{\sum x_i^2}{\sum x_i^2} = 1 \]  \hspace{1cm} (35)

Therefore, the variance of \( b_1^* \) simplifies to,

\[ \text{Var}(b_1^*) = \sigma^2 \left( \sum (w_i - k_i)^2 + \sum k_i^2 \right) \]  \hspace{1cm} (36)

This expression is minimized when

\[ w_i = k_i \]  \hspace{1cm} (37)

and the minimum variance is,

\[ \text{Var}(b_1^*) = \sigma^2 \sum k_i^2 \]  \hspace{1cm} (38)

This completes the proof that, under the classical assumptions, the OLS estimator is has the least variance among all linear unbiased estimators.

### 4.1 Consistency

We established that the OLS estimator is unbiased and efficient under classical assumptions. We can also show easily that the OLS estimator is consistent under the same assumption.

An estimator is consistent if its variances reaches zero as the sample size increases. In order to see this, start with the expression for the variance,

\[ \text{Var}(b_1) = \frac{\sigma^2}{\sum x_i^2} \]  \hspace{1cm} (39)

Divide both the denominator and numerator by \( n \).

\[ \text{Var}(b_1) = \frac{\sigma^2/n}{\sum x_i^2/n} \]  \hspace{1cm} (40)
As \( n \to \infty \), the numerator approaches zero whereas the denominator remains positive. Therefore,

\[
\lim_{n \to \infty} \text{Var}(b_1) = 0 \quad (41)
\]

5 Gauss-Markov Theorem and Classical Assumptions

To recap, we have demonstrated that the OLS estimator,

\[
b_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum k_i Y_i \quad (42)
\]

has the following properties;

1. Unbiased, i.e.

\[
E(b_1) = \beta_1 \quad (43)
\]

2.

\[
\text{Var}(b_1) = \frac{\sigma^2}{\sum x_i^2} = \sigma^2 \sum k_i^2 \quad (44)
\]

3. Best or efficient, i.e. has lower variance than any other linear unbiased estimator, i.e.

\[
\text{Var}(b_1) < \text{Var}(b_1^*) \quad (45)
\]

where \( b_1^* = \sum w_i Y_i \) and \( w_i \) is any other function of \( x_i \).

4. Consistent, i.e.

\[
\lim_{n \to \infty} \text{Var}(b_1) = 0 \quad (46)
\]
if the following classical assumptions are satisfied,

1. The underlying regression model is linear in parameters, has an additive error and is correctly specified, i.e.

\[ Y_i = \beta_0 + \beta_1 f(X_i) + u_i \quad (47) \]

2. The X variable is non-stochastic, i.e. fixed in repeated sampling.

3. The expected value of the error term is zero, i.e.

\[ E(u_i) = 0 \quad (48) \]

Note that the intercept term, \( \beta_0 \) ensures that this condition is met. Consider

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \quad (49) \]

\[ E(u_i) = k \quad (50) \]

This is equivalent to a model where

\[ Y_i = \beta_0^* + \beta_1 X_i + u_i^* \quad (51) \]

\[ \beta_0^* = \beta_0 + 3 \quad (52) \]

\[ E(u_i) = 0 \quad (53) \]

Note also that the first three conditions are sufficient for OLS to be unbiased.

4. The explanatory variable, \( X_i \) is uncorrelated with the error term \( u_i \), i.e.

\[ Cor(X, e) = E[x_i u_i] = 0 \quad (54) \]
Note that this assumption is necessary for OLS to be unbiased. Even if $x_i$ is non-stochastic, we can obtain unbiased coefficients if $x_i$ is uncorrelated with the error term. Such correlation occurs typically if $X_i$ is endogenous, i.e. determined by other variables. If both $X_i$ and $Y_i$ are determined by the same unobserved variables, this assumption is violated. If $X_i$ and $Y_i$ are determined by each other, i.e. simultaneous equations, this assumption is also violated. For example, if

$$Y = \beta_0 + \beta_1 X_i + u_i \quad (55)$$
$$X_i = \delta_0 + \delta_1 Y_i + \epsilon_i \quad (56)$$

$\text{Cor}(X_i, u_i) \neq 0$ if $\delta_1 \neq 0$ and/or $\text{Cor}(u_i, \epsilon_i) \neq 0$

5. The error term is homoskedastic, i.e. the conditional variance is a constant.

$$\text{Var}(u_i|X_i) = E(u_i^2|X_i) = \sigma^2 \quad (57)$$

6. The error term is serially uncorrelated, i.e. the error term of one observation is not correlated with the error term of any other observation.

$$\text{Cov}(u_i, u_j|X_i, X_j) = E(u_i u_j|X_i, X_j) = 0 \forall i \neq j \quad (58)$$

The assumptions of serially uncorrelated and homoskedastic errors allow us to obtain an unbiased estimator for the variance of the error term, and a simple OLS formula for the variance of the coefficient estimate. In addition, we need these two assumptions to demonstrate that OLS is efficient. In fact, we will see later that other GLS methods are efficient when these assumptions are violated.
There are a few other assumptions that are necessary to obtain OLS coefficients and standard errors;

1. At least one degree of freedom, i.e. the number of observations must exceed the number of parameters \((n > k + 1)\) where \(k\) is the number of \(X\) variables. In the simple regression with one \(X\) variable, this means there should be at least three observations.

2. No \(X\) variable should be a deterministic linear function of other \(X\) variables, i.e. no multicollinearity. This condition applies only to multiple regressions where there are more than one \(X\) variable, and is discussed later.

3. There should be some variation in the \(X\) variable. If the \(X\) variable does not vary, it is impossible to estimate the slope of a regression line.