1 Unbiasedness and Normality of OLS under normality

We assume the error terms are normally distributed.

In this case we use the concepts of unbiasedness and normality of the estimator, which are the comparable concept of consistency and asymptotic normality. Remember that an estimator is *consistent* if it converges to the true value as the number of observations get large.

$$\mathbf{b}_T \to \mathbf{b} \;\; \text{as } T \to \infty$$

An estimator is *unbiased* if it's expectation is equal to its true value

$$E\left[\widehat{\mathbf{b}}\right] = \mathbf{b}.$$

Similarly, it is asymptotically normal if

$$\sqrt{T}\widehat{\mathbf{b}}_T \overset{D}{\to} \mathcal{N}(b, \mathbf{\Omega}) \;\; \text{as } T \to \infty$$

where $\mathbf{\Omega}$ is the asymptotic covariance matrix,

and normal if

$$\widehat{\mathbf{b}} \sim \mathcal{N}(b, \mathbf{\Omega}),$$

where $\mathbf{\Omega}$ is the covariance matrix.

It is a simple matter to establish unbiasedness and normality of the OLS estimator under the following:

Exercise 1.

Make the following assumptions.

i The model is known to be $\mathbf{y} = \mathbf{X}\mathbf{b}_0 + \mathbf{e}$, $|\mathbf{b}_0| < \infty$.

ii $\mathbf{X}$ is a nonstochastic and finite $n \times k$ matrix.

iii $\mathbf{X}'\mathbf{X}$ is nonsingular for all $n \geq k$.

iv $E[\mathbf{e}] = 0$.

v $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_0^2\mathbf{I}),\ \sigma_0^2 < \infty$

Under these assumptions.

1. (Existence) Given (i)–(iii), $\widehat{\mathbf{b}}^{ols}$ exist for all $n \geq k$ and is unique.
2. (Unbiasedness) Given (i)–(iv), $E[\widehat{\mathbf{b}}^{ols}] = \mathbf{b}_0$
3. (Normality) Given (i)–(v) $\widehat{\mathbf{b}}^{ols} \sim \mathcal{N}(\mathbf{b}_0, \sigma_0^2(\mathbf{X}'\mathbf{X})^{-1})$.

Prove these three results.

Solution to Exercise 1.

Given the OLS estimator

$$\widehat{\mathbf{b}}^{ols} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

1. *Existence*
   - (iii) implies $\mathbf{X}'\mathbf{X}$ is finite and of full rank. Therefore $(\mathbf{X}'\mathbf{X})^{-1}$ exists.
   - (ii) implies $\mathbf{X}$ is finite
   - this and linearity of the matrix operators shows existence and uniqueness.
2. *Unbiasedness*

$$\begin{aligned} E[\widehat{\mathbf{b}}^{ols}] &= E[\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}\right] \\ &= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\mathbf{b} + \mathbf{e}\right)\right] \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{e}] \\ &= \mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} \\ &= \mathbf{b} \end{aligned}$$

   This establishes unbiasedness

3. *Distribution*

Under the model assumptions, we calculated:

$$\widehat{\mathbf{b}}^{ols} = \mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}$$

- Normality

  Under the maintained assumptions, $\mathbf{b}$ and $\mathbf{X}$ can be viewed as constants. The distribution is thus a linear combination of normally distributed variables ($\mathbf{e}$), which by a well-known result is itself normal.

- Covariance matrix:

  The covariance matrix of the OLS estimator is found as

$$\begin{aligned}
& E\left[(\widehat{\mathbf{b}}^{ols} - E[\widehat{\mathbf{b}}^{ols}])(\widehat{\mathbf{b}}^{ols} - E[\widehat{\mathbf{b}}^{ols}])'\right] \\
&= E\left[(\widehat{\mathbf{b}}^{ols} - \mathbf{b})(\widehat{\mathbf{b}}^{ols} - \mathbf{b})'\right] \\
&= E\left[\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{b}\right\}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{b}\right\}'\right] \\
&= E\left[\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\mathbf{b} + \mathbf{e}) - \mathbf{b}\right\}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\mathbf{b} + \mathbf{e}) - \mathbf{b}\right\}'\right] \\
&= E\left[\left\{(\mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{e}) - \mathbf{b}\right\}\left\{(\mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{e}) - \mathbf{b}\right\}'\right] \\
&= E\left[\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\right\}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\right\}'\right] \\
&= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E\left[\mathbf{e}\mathbf{e}'\right]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

We have now proved all but the last point of the classical OLS theorem. The remaining part is the efficiency part, which is often called the Gauss-Markov theorem.

1.1 Gauss Markov theorem

When we do estimation, we naturally want to have the estimates that uses the information in the data as good as possible.

To formalize this we need a concept of efficiency. One of the most common is to use the *mean squared error* as a measure of efficiency, which in this case will be to look at

$$E\left[(\widehat{\mathbf{b}} - \mathbf{b})(\widehat{\mathbf{b}} - \mathbf{b})'\right]$$

In univariate cases the comparison of two estimators is simple, but for multivariate cases we need to know how to compare the mean squared errors of the estimators we are comparing.

Let

$$\mathbf{V}(\widehat{\mathbf{b}}) = E\left[(\widehat{\mathbf{b}} - \mathbf{b})(\widehat{\mathbf{b}} - \mathbf{b})'\right]$$

be the covariance matrix of $\mathbf{b}$.

If we compare two estimators $\widehat{\mathbf{b}}$ and $\tilde{\mathbf{b}}$, we say that $\widehat{\mathbf{b}}$ is more efficient than $\tilde{\mathbf{b}}$ if

$$\mathbf{V}(\tilde{\mathbf{b}}) - \mathbf{V}(\widehat{\mathbf{b}})$$

is a positive semidefinite matrix.

The best known case of an efficiency result is the Gauss-Markov theorem, and I do not think I can avoid going over this theorem.

Exercise 2.

Consider the following result, also called the Gauss-Markov Theorem.

Assume

i The model is known to be $\mathbf{y} = \mathbf{X}\mathbf{b}_0 + \mathbf{e}$, $|\mathbf{b}_0| < \infty$.

ii $\mathbf{X}$ is a nonstochastic and finite $n \times k$ matrix.

iii $\mathbf{X}'\mathbf{X}$ is nonsingular for all $n \geq k$.

iv $E[\mathbf{e}] = 0$.

v $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I}),\ \sigma_0^2 < \infty$

Under these assumptions show that the OLS estimator

$$\mathbf{b}^{ols} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

is the Best Linear Unbiased Estimator (BLUE) in the sense that the variance–covariance matrix of any other unbiased estimator exceeds that of $\mathbf{b}^{ols}$ by a positive semidefinite matrix, regardless of the value of $\mathbf{b}_0$

Solution to Exercise 2.

Let $\tilde{\mathbf{b}}$ be some other linear unbiased estimator. If we can show

$$\mathbf{V}(\tilde{\mathbf{b}}) - \mathbf{V}(\widehat{\mathbf{b}}^{ols})$$

is positive semidefinite, we have proved the theorem.

Step 1. Since $\tilde{\mathbf{b}}$ is linear, we can express it as

$$\tilde{\mathbf{b}} = \mathbf{A}\mathbf{y},$$

where $\mathbf{A}$ is *some* matrix.

Let

$$\mathbf{C} = \mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Then we can write

$$\tilde{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \mathbf{C}\mathbf{y}$$

Note that $(\mathbf{X}'\mathbf{X})^{-1}\ \mathbf{X}'\ \mathbf{y}$ is the OLS estimate.

Replace $\mathbf{y}$ with its value under true (known) model $\mathbf{y} = \mathbf{X}\mathbf{b}_0 + \mathbf{e}$:

$$\begin{aligned}
\tilde{\mathbf{b}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{b}_0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} + \mathbf{C}\mathbf{X}\mathbf{b}_0 + \mathbf{C}\mathbf{e} \\
&= \mathbf{b}_0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} + \mathbf{C}\mathbf{X}\mathbf{b}_0 + \mathbf{C}\mathbf{e}
\end{aligned}$$

Since $\tilde{\mathbf{b}}$ is unbiased, $E[\tilde{\mathbf{b}}] = \mathbf{b}_0$, which implies

$$\begin{aligned}
\mathbf{b}_0 &= E[\tilde{\mathbf{b}}] \\
&= E[\mathbf{b}_0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} + \mathbf{C}\mathbf{X}\mathbf{b}_0 + \mathbf{C}\mathbf{e}] \\
&= \mathbf{b}_0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{e}] + \mathbf{C}\mathbf{X}\mathbf{b}_0 + \mathbf{C}E[\mathbf{e}] \\
&= \mathbf{b}_0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} + \mathbf{C}\mathbf{X}\mathbf{b}_0 + \mathbf{C}\mathbf{0} \\
&= \mathbf{b}_0 + \mathbf{C}\mathbf{X}\mathbf{b}_0
\end{aligned}$$

which implies

$$\mathbf{C}\mathbf{X}\mathbf{b}_0 = \mathbf{0}$$

Since this have to hold for all $\mathbf{b}_0$, we must have

$$\mathbf{C}\mathbf{X} = \mathbf{0}$$

Step 2.

$$\begin{aligned}
&E[(\tilde{\mathbf{b}} - \mathbf{b}_0)(\tilde{\mathbf{b}} - \mathbf{b}_0)'] \\
&= E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \mathbf{C}\mathbf{y} - b_0)((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \mathbf{C}\mathbf{y} - b_0)'] \\
&= E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\mathbf{b}_0 + \mathbf{e}) + \mathbf{C}(\mathbf{X}\mathbf{b}_0 + \mathbf{e}) - b_0)((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\mathbf{b}_0 + \mathbf{e}) + \mathbf{C}(\mathbf{X}\mathbf{b}_0 + \mathbf{e}) - b_0)'] \\
&= E[(\mathbf{b}_0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}) + \mathbf{C}\mathbf{X}\mathbf{b}_0 + \mathbf{C}\mathbf{e} - \mathbf{b}_0)(\mathbf{b}_0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} + \mathbf{C}\mathbf{X}\mathbf{b}_0 + \mathbf{C}\mathbf{e} - \mathbf{b}_0)'] \\
&= E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} + \mathbf{C}\mathbf{X}\mathbf{b}_0 + \mathbf{C}\mathbf{e})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} + \mathbf{C}\mathbf{X}\mathbf{b}_0 + \mathbf{C}\mathbf{e}))' \\
&= E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} + \mathbf{C}\mathbf{e})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} + \mathbf{C}\mathbf{e})'] \\
&= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\mathbf{e}'\mathbf{C}'] + E[\mathbf{C}\mathbf{e}\mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] + E[\mathbf{C}\mathbf{e}\mathbf{e}'\mathbf{C}'] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{C}' + \mathbf{C}\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{C}\sigma^2\mathbf{I}\mathbf{C}' \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\mathbf{C}' + \sigma^2\mathbf{C}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{C}\mathbf{C}' \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{C}\mathbf{C}' \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{C}\mathbf{C}'
\end{aligned}$$

In several steps we have used that $\mathbf{CX} = \mathbf{0}$ to cancel terms.

Note that $\mathbf{CC}'$ is positive semidefinite by observing that for any vector ω

$$\omega'\mathbf{CC}'\omega = (\omega'\mathbf{C})'(\mathbf{C}'\omega)$$

Since $\omega\mathbf{C}$ is a vector, this is a sum of squared terms, and

$$\omega'\mathbf{CC}'\omega \geq 0 \ \forall \ \omega$$

Step 3:

$$\begin{aligned} \mathbf{V}(\tilde{\mathbf{b}}) - \mathbf{V}(\widehat{\mathbf{b}}^{ols}) &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{CC}' - \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2\mathbf{CC}', \end{aligned}$$

which we have just shown to be positive definite.

This concludes the proof that $\widehat{\mathbf{b}}^{ols}$ is BLUE.

Readings/Sources (Theil, 1971, Ch 3) (Davidson and MacKinnon, 1993, 3.2, 5.5)

References

Russel Davidson and James G MacKinnon. *Estimation and Interference in Econometrics.* Oxford University Press, 1993.

Henri Theil. *Principles of econometrics.* Wiley, 1971.