1 Introduction

We now compare the properties of the LS estimator to other potential estimators, for both small and large samples sizes.

2 Small-Sample Properties

2.1 Gauss-Markov Theorem

The Gauss-Markov Theorem states that, provided the Classical assumptions hold, the ordinary least squares (OLS) estimator $b$ is the minimum variance estimator among all linear unbiased estimators. Sometimes it is said that the OLS estimator is BLUE (Best Linear Unbiased Estimator).

Proof.

First, we need to show that $b$ is unbiased. We know

$$b = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \epsilon) = \beta + (X'X)^{-1}X'\epsilon.$$  

Taking expectations gives

$$E[b] = \beta + (X'X)^{-1}X'E[\epsilon] = \beta$$

because $\beta$ and $X$ are not random variables (recall $X$ is assumed to be fixed in repeated sampling). Therefore, $b$ is an unbiased estimator of $\beta$.

Second, we need to show that $b$ has the smallest variance (among all linear unbiased estimators). Begin by noting that $b$ is a linear estimator because it is linear in $Y$ (or alternatively $\epsilon$). Now consider all other possible linear unbiased estimators $b_0 = CY$, where $C$ is a fixed $(k \times n)$ matrix. For $b_0$ to be unbiased, it must be that $CX = I$ because

$$E[b_0] = E[CY\beta + C\epsilon] = CX\beta.$$
The variance of $b$ is

$$\text{var}(b) = E[(b - \beta)(b - \beta)']$$

$$= E[(X'X)^{-1}X'e'X(X'X)^{-1}]$$

$$= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1}.$$

The variance of $b_0$ is

$$\text{var}(b_0) = E[(b_0 - \beta)(b_0 - \beta)']$$

$$= E[C\epsilon'C']$$

$$= \sigma^2CC'.$$

The question is now whether $(X'X)^{-1}$ or $CC'$ is bigger (in a matrix sense). Toward that end, define $D \equiv C - (X'X)^{-1}X'$ so that $DX = 0$. Using this, we can write

$$\text{var}(b_0) = \sigma^2(D + (X'X)^{-1}X')(D + X'X)^{-1}X')'$$

$$= \sigma^2(X'X)^{-1} + \sigma^2DD'$$

$$= \text{var}(b) + \sigma^2DD'.$$

Finally, we note that $DD'$ is a non-negative definite matrix (Greene A-114) so that the variance of $b$ is no larger than the variance $b_0$. ▼

### 2.2 Estimating the Variance of the LS Estimator

We know $\text{var}(b) = \sigma^2(X'X)^{-1}$, but $\sigma^2$ is an unknown parameter. Therefore in order to find $\text{var}(b)$, we need to find a good estimator for $\sigma^2$.

Start by defining $M = I - X(X'X)^{-1}X'$, which is symmetric and idempotent. This matrix can be used to relate the residuals $e$ to the errors $\epsilon$

$$e = MY = M(X\beta + \epsilon) = M\epsilon.$$

Using this relation, we can then find an unbiased estimator for $\sigma^2$. Begin by finding the expectation of the inner product of $e$

$$E[e'e] = E[e'M\epsilon] = E[tr(e'M\epsilon)].$$
The last equality uses the fact that the trace \((tr)\) of a scalar is simply the scalar. This can be further manipulated to give

\[ E[tr(\epsilon' M\epsilon)] = E[tr(\epsilon\epsilon' M)] \]

using (Greene A-94). Taking expectations through the trace operator then gives

\[ E[tr(\epsilon\epsilon' M)] = \sigma^2 tr(M) = \sigma^2[tr(I_n) - tr(X(X'X)^{-1}X')] \]

which after using (Greene A-94) again produces

\[ \sigma^2[tr(I_n) - tr(X(X'X)^{-1}X')] = \sigma^2[n - tr((X'X)^{-1}(X'X))] = \sigma^2[n - k]. \]

Therefore, if we define \( s^2 = e'\epsilon/(n - k) \), we know it will be an unbiased estimator for \( \sigma^2 \) and \( \text{var}(b) = s^2(X'X)^{-1} \). The square root of the estimated variance of \( b \) is often called the **standard error of \( b \)**.

### 3 Large-Sample Properties

In many cases, we cannot calculate the exact distribution of our estimators. This is generally true when we relax Classical assumption #6, which we do here. Fortunately, however, we can often calculate approximate distributions that hold when the sample size is large.

#### 3.1 Consistency of \( b \)

Recall, a **consistent estimator** has the following property

\[ \lim_{n \to \infty} \Pr(|b - \beta| < \delta) = 1 \]

for any positive \( \delta \). It is said that the probability limit of \( b \) is \( \beta \), that is \( \text{plim}(b) = \beta \). Next, we are going to establish the consistency of \( b \).

Continue to assume that \( X \) is nonstochastic and

\[ \lim_{n \to \infty} \frac{1}{n}(X'X) = Q, \]

is a positive-definite finite matrix. This condition is fairly restrictive (less restrictive assumptions can be used) and guarantees that the explanatory data are "well-behaved" in the sense that their variance does not get too large. Here is an example where the condition is not satisfied, but the LS estimator is still consistent.
Example. Consider the time-series model

\[ y_t = \beta_1 + \beta_2 t + \epsilon_t \]

where \( t = 1, \ldots, n \). In this case,

\[
X'X = \begin{bmatrix}
\sum_{t=1}^{n} t \\
\sum_{t=1}^{n} t^2
\end{bmatrix} = \begin{bmatrix}
n & \frac{n(n+1)}{2} \\
\frac{n(n+1)}{2} & \frac{n(n+1)(2n+1)}{6}
\end{bmatrix} \implies \lim_{n \to \infty} \frac{1}{n} (X'X) = \begin{bmatrix}
1 & \infty \\
\infty & \infty
\end{bmatrix}.
\]

To show consistency, rewrite \( b \) as

\[ b = \beta + \left( \frac{1}{n} X'X \right)^{-1} \left( \frac{1}{n} X' \epsilon \right). \]

Taking the probability limit gives

\[
plim(b - \beta) = plim\left( \frac{1}{n} X'X \right)^{-1} plim\left( \frac{1}{n} X' \epsilon \right) = (plim\left( \frac{1}{n} X'X \right)^{-1} plim\left( \frac{1}{n} X' \epsilon \right) = Q^{-1} \times 0 = 0
\]

where \( plim\left( \frac{1}{n} X'X \right)^{-1} = (plim\left( \frac{1}{n} X'X \right))^{-1} \) via Slutsky’s Theorem (Greene Theorem D.12) and \( plim\left( \frac{1}{n} X' \epsilon \right) = 0 \) because \( \frac{1}{n} X' \epsilon \) converges in mean square to zero (Greene Theorem D.11). As a result, \( plim(b) = \beta \) or \( b \) is a consistent estimator of \( \beta \).

3.2 Asymptotic Distribution of \( b \)

Continue to assume that \( X \) is nonstochastic, \( \lim_{n \to \infty} \frac{1}{n} (X'X) = Q \) and \( \epsilon \sim (0, \sigma^2 I) \). Because \( b \) is a consistent estimator of \( \beta \), the limiting distribution of \( b \) is degenerate (i.e., a spike at \( \beta \)). However, using the Central Limit Theorem, we can take a stabilizing transformation of \( b \) to produce a non-degenerate limiting distribution

\[
\sqrt{n}(b - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1}).
\]

This result suggests that, in large samples, we can approximate the distribution of \( b \) as \( N(\beta, \frac{\sigma^2}{n} Q^{-1}) \). We call this the asymptotic distribution of \( b \) or \( b^* \sim N(\beta, \frac{\sigma^2}{n} Q^{-1}) \).