Regression #3: Properties of OLS Estimator

Econ 671

Purdue University
In this lecture, we establish some desirable properties associated with the OLS estimator.

These include proofs of *unbiasedness* and *consistency* for both $\hat{\beta}$ and $\hat{\sigma}^2$, and a derivation of the conditional and unconditional variance-covariance matrix of $\hat{\beta}$. 
Unbiasedness

\[ y_i = x_i \beta + \epsilon_i. \]

\[ \hat{\beta} = (X'X)^{-1}X'y. \]

We continue with our standard set of regression assumptions, including 
\[ E(\epsilon|X) = 0 \] and \[ E(\epsilon\epsilon'|X) = \sigma^2 I_n. \]

Theorem

What does this actually mean? Can you think of a situation where an unbiased estimator might not be preferred over a biased alternative?
Unbiasedness

Proof.

First, consider $E(\hat{\beta}|X)$. To this end, we note:

Therefore, by the law of iterated expectations,
Variance-Covariance Matrix

We now seek to obtain the variance-covariance matrix of the OLS estimator. To this end, we note:
Variance-Covariance Matrix

Another way to get this same result is as follows:

So, what do the elements of this $k \times k$ matrix represent? Why are they useful?
**Variance-Covariance Matrix**

To obtain an *unconditional* variance-covariance matrix, i.e., $\text{Var}(\hat{\beta})$, we note that, in general,

- 

Thus, (why?)

- 

In practice, we evaluate this at the observed $X$ values:

- 

Variance-Covariance Matrix

Another issue that arises is that the variance parameter, $\sigma^2$ is also unknown and must be estimated. A natural estimator arises upon considering its definition:

Replacing the population expectation with its sample counterpart, and using $\hat{\beta}$ instead of $\beta$, we obtain an intuitive estimator:
Variance-Covariance Matrix

Though this estimator is widely used, it turns out to be a biased estimator of $\sigma^2$. An unbiased estimator can be obtained by incorporating the degrees of freedom correction:

$$
\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^{n} (y_i - \bar{y})^2
$$

where $k$ represents the number of explanatory variables included in the model. In the following slides, we show that $\hat{\sigma}^2$ is indeed unbiased.
We seek to show

\[ E(\hat{\sigma}^2|X) = \sigma^2. \]

Proof.

where the last result follows since \( X'M = MX = 0 \).
Proof.

It follows that

... is an unbiased estimator of $\sigma^2$, as claimed.
**Consistency**

Recall the definition of a *consistent* estimator, $\hat{\theta}(x_n) = \hat{\theta}_n$ of $\theta$. We say $\hat{\theta}_n$ is consistent if for any $\epsilon > 0$,

$$\lim_{n \to \infty} \Pr \left\{ |\hat{\theta}_n - \theta| > \epsilon \right\} = 0.$$

Relatedly, we say that $\hat{\theta}_n$ converges in *mean square* to $\theta$ if:

$$\lim_{n \to \infty} E(\hat{\theta}_n - \theta)^2 = 0.$$

The MSE criterion can also be written as the Bias squared plus the variance, whence

$$\hat{\theta}_n \xrightarrow{m.s.} \theta \quad \text{iff} \quad \text{Bias}(\hat{\theta}_n) \to 0 \quad \text{and} \quad \text{Variance} \ (\hat{\theta}_n) \to 0.$$
**Consistency**

We will prove that MSE can be written as the square of the bias plus the variance:

\[
E([\hat{\theta}_n - \theta]^2) = E([\hat{\theta}_n - E(\hat{\theta}_n) + E(\hat{\theta}_n) - \theta]^2)
\]

\[
= E([\hat{\theta}_n - E(\hat{\theta}_n)]^2) + 2E([\hat{\theta}_n - E(\hat{\theta}_n)][E(\hat{\theta}_n) - \theta])
\]

\[
+ E([E(\hat{\theta}_n) - \theta]^2)
\]

\[
= E([\hat{\theta}_n - E(\hat{\theta}_n)]^2) + [E(\hat{\theta}_n) - \theta]^2
\]

\[
= \text{Variance} + \text{Bias}^2
\]
**Consistency**

Convergence in mean square is also a *stronger* condition than convergence in probability:

**Proof.**

Fix $\epsilon > 0$ and note:

$$E \left[ (\hat{\theta}(x_n) - \theta)^2 \right] = \int_{X_n} (\hat{\theta}(x_n) - \theta)^2 f_n(x_n) dx_n$$

$$\geq \int_{\{x_n:|\hat{\theta}(x_n) - \theta| > \epsilon\}} (\hat{\theta}(x_n) - \theta)^2 f_n(x_n) dx_n$$

$$\geq \epsilon^2 \int_{\{x_n:|\hat{\theta}(x_n) - \theta| > \epsilon\}} f_n(x_n) dx_n$$

$$= \epsilon^2 \Pr \left\{ |\hat{\theta}(x_n) - \theta| > \epsilon \right\}.$$
Consistency

Thus,

\[ 0 \leq \Pr \left\{ |\hat{\theta}(x_n) - \theta| > \epsilon \right\} \leq \frac{1}{\epsilon^2} E \left[ (\hat{\theta}(x_n) - \theta)^2 \right]. \]

For fixed \( \epsilon \) and taking limits as \( n \to \infty \) gives the result.

The assumption of convergence in mean square therefore guarantees that the estimator converges in probability.
Consistency

Now, let us revisit $\hat{\beta}$.

To show that $\hat{\beta} \xrightarrow{p} \beta$ [or $\text{plim}(\hat{\beta}) = \beta$], it is enough to show that the bias and variance of $\hat{\beta}$ go to zero.

The estimator has already been demonstrated to be unbiased. As for the variance, note:
Consistency

Consider the matrix $X'X/n$.

A typical element of this matrix is a sample average of the form:

$$n^{-1} \sum_{i=1}^{n} x_{ij} x_{il}.$$ 

Provided these averages settle down to finite population means, it is reasonable to assume

where $Q$ has finite elements and is nonsingular.
Consistency

Since the inverse is a continuous function, we have:

Thus,

whence

\[ \text{plim}(\hat{\beta}) = \beta, \]

as needed.
Consistency

Let us now investigate the consistency of $\hat{\sigma}^2$. From before, we can write:

- 

We can now use some properties of plim’s to simplify this result. First, note that:

- 

by Chebyshev’s LLN. Similarly, note
Consistency

By assumption, we have \((X'X/n)^{-1} \overset{p}{\rightarrow} Q^{-1}\) and we also note

- 

given that \(E(\epsilon|X) = 0\). Putting all of this together, we have

- 

as needed.