OLS: Estimation and Standard Errors

Brandon Lee

15.450 Recitation 10
The model:

\[ y = X\beta + \varepsilon \]

where \( y \) and \( \varepsilon \) are column vectors of length \( n \) (the number of observations), \( X \) is a matrix of dimensions \( n \) by \( k \) (\( k \) is the number of parameters), and \( \beta \) is a column vector of length \( k \).

For every observation \( i = 1, 2, \ldots, n \), we have the equation

\[ y_i = x_{i1}\beta_1 + \cdots + x_{ik}\beta_k + \varepsilon_i \]

Roughly speaking, we need the orthogonality condition

\[ E[\varepsilon_i x_i] = 0 \]

for the OLS to be valid (in the sense of consistency).
We want to find $\hat{\beta}$ that solves

$$\min_\beta (y - X\beta)'(y - X\beta)$$

The first order condition (in vector notation) is

$$0 = X'(y - X\hat{\beta})$$

and solving this leads to the well-known OLS estimator

$$\hat{\beta} = (X'X)^{-1}X'y$$
The left-hand variable is a vector in the $n$-dimensional space. Each column of $X$ (regressor) is a vector in the $n$-dimensional space as well, and we have $k$ of them. Then the subspace spanned by the regressors forms a $k$-dimensional subspace of the $n$-dimensional space. The OLS procedure is nothing more than finding the orthogonal projection of $y$ on the subspace spanned by the regressors, because then the vector of residuals is orthogonal to the subspace and has the minimum length.

This interpretation is very important and intuitive. Moreover, this is a unique characterization of the OLS estimate.

Let’s see how we can make use of this fact to recognize OLS estimators in disguise as more general GMM estimators.
Refer to pages 35-37 of Lecture 7.
The model is

$$r_{t+1} = a_0 + a_1 r_t + \varepsilon_{t+1}$$

where

$$E[\varepsilon_{t+1}] = 0$$

$$E[\varepsilon_{t+1}^2] = b_0 + b_1 r_t$$

One easy set of moment conditions:

$$0 = E[(1, r_t)' (r_{t+1} - a_0 - a_1 r_t)]$$

$$0 = E[(1, r_t)' ((r_{t+1} - a_0 - a_1 r_t)^2 - b_0 - b_1 r_t)]$$
Solving these sample moment conditions for the unknown parameters is exactly equivalent to a two-stage OLS procedure.

Note that the first two moment conditions give us

$$E_T [(1, r_t)' (r_{t+1} - \hat{a}_0 - \hat{a}_1 r_t)] = 0$$

But this says that the estimated residuals are orthogonal to the regressors and hence $\hat{a}_0$ and $\hat{a}_1$ must be OLS estimates of the equation

$$r_{t+1} = a_0 + a_1 r_t + \epsilon_{t+1}$$
Now define

$$\hat{\varepsilon}_{t+1} = r_{t+1} - \hat{a}_0 - \hat{a}_1 r_t$$

then the sample moment conditions

$$E_T \left[ (1, r_t)' \left((r_{t+1} - \hat{a}_0 - \hat{a}_1 r_t)^2 - \hat{b}_0 - \hat{b}_1 r_t \right) \right] = 0$$

tell us that $\hat{b}_0$ and $\hat{b}_1$ are OLS estimates from the equation

$$\hat{\varepsilon}_{t+1}^2 = b_0 + b_1 r_t + u_{t+1}$$

by the same logic.
Let’s suppose that $E[ε_i^2|X] = σ^2$ and $E[ε_iε_j|X] = 0$ for $i \neq j$. In other words, we are assuming independent and homoskedastic errors.

What is the standard error of the OLS estimator under this assumption?

$$\text{Var}(\hat{β}|X) = \text{Var}(\hat{β} - β|X)$$

$$= \text{Var}((X'X)^{-1}X'|X)$$

$$= (X'X)^{-1}X'\text{Var}(ε|X)X(X'X)^{-1}$$

Under the above assumption,

$$\text{Var}(ε|X) = σ^2 I_n$$

and so

$$\text{Var}(\hat{β}|X) = σ^2 (X'X)^{-1}$$
We can estimate $\sigma^2$ by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

and the standard error for the OLS estimator is given by

$$\text{Var} \left( \hat{\beta} | X \right) = \hat{\sigma}^2 (X'X)^{-1}$$

This is the standard error that most (less sophisticated) statistical softwares report.

But it is rarely the case that it is safe to assume independent homoskedastic errors. The Newey-West procedure is a straightforward and robust method of calculating standard errors in more general situations.
Again,

\[
\text{Var} \left( \hat{\beta} | X \right) = \text{Var} \left( \hat{\beta} - \beta | X \right)
\]

\[
= \text{Var} \left( (X'X)^{-1} X' \varepsilon | X \right)
\]

\[
= (X'X)^{-1} \text{Var} \left( X' \varepsilon | X \right) (X'X)^{-1}
\]

The Newey-West procedure boils down to an alternative way of looking at \( \text{Var}(X' \varepsilon | X) \).

If we suspect that the error terms may be heteroskedastic, but still independent, then

\[
\hat{\text{Var}} \left( X' \varepsilon | X \right) = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \cdot x_i x_i'
\]

and our standard error for the OLS estimate is

\[
\hat{\text{Var}} \left( \hat{\beta} | X \right) = (X'X)^{-1} \left( \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \cdot x_i x_i' \right) (X'X)^{-1}
\]
If we suspect correlation between error terms as well as heteroskedasticity, then

$$\hat{V} ar \left( X' \epsilon | X \right) = \sum_{j=-k}^{k} \frac{k-|j|}{k} \left( \sum_{t=1}^{n} \hat{\epsilon}_i \hat{\epsilon}_{i+j} x_i x_{i+j} \right)$$

and our standard error for the OLS estimator is

$$\hat{V} ar \left( \hat{\beta} | X \right) = (X'X)^{-1} \left( \sum_{j=-k}^{k} \frac{k-|j|}{k} \left( \sum_{t=1}^{n} \hat{\epsilon}_i \hat{\epsilon}_{i+j} x_i x_{i+j} \right) \right) (X'X)^{-1}$$
We can also write these standard errors to resemble the general GMM standard errors (see page 23 of Lecture 8).

In the uncorrelated errors case, we have

\[
\widehat{\text{Var}} \left( \hat{\beta} \mid X \right) = (X'X)^{-1} \left( \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \cdot x_i x_i' \right) (X'X)^{-1}
\]

\[
= \frac{1}{n} \left( \frac{X'X}{n} \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \cdot x_i x_i' \right) \left( \frac{X'X}{n} \right)^{-1}
\]

\[
= \frac{1}{n} \hat{E} \left( x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \cdot x_i x_i' \right) \hat{E} \left( x_i x_i' \right)^{-1}
\]

and for the general Newey-West standard errors, we have

\[
\widehat{\text{Var}} \left( \hat{\beta} \mid X \right) = (X'X)^{-1} \left( \sum_{j=-k}^{k} \frac{k-|j|}{k} \left( \sum_{t=1}^{n} \hat{\varepsilon}_i \hat{\varepsilon}_{i+j} \cdot x_i x_{i+j}' \right) \right) (X'X)^{-1}
\]

\[
= \frac{1}{n} \hat{E} \left( x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{j=-k}^{k} \frac{k-|j|}{k} \left( \sum_{t=1}^{n} \hat{\varepsilon}_i \hat{\varepsilon}_{i+j} \cdot x_i x_{i+j}' \right) \right) \hat{E} \left( x_i x_i' \right)^{-1}
\]