OLS: Estimation and Standard Errors

Brandon Lee

15.450 Recitation 10

Ordinary Least Squares

- The model:

$$y = X\beta + \varepsilon$$

  where y and ε 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 β 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).

OLS Estimator

- We want to find $\hat{\beta}$ that solves

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

- The first order condition (in vector notation) is

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

and solving this leads to the well-known OLS estimator

$$\hat{\beta} = \left(X'X\right)^{-1}X'y$$

Geometric Interpretation

- 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.

Interest Rate Model

- Refer to pages 35-37 of Lecture 7.
- The model is

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

where

$$E\left[\varepsilon_{t+1}\right] = 0$$
$$E\left[\varepsilon_{t+1}^2\right] = b_0 + b_1 r_t$$

- One easy set of moment conditions:

$$0 = E\left[(1, r_t)'\left(r_{t+1} - a_0 - a_1 r_t\right)\right]$$
$$0 = E\left[(1, r_t)'\left(\left(r_{t+1} - a_0 - a_1 r_t\right)^2 - b_0 - b_1 r_t\right)\right]$$

Continued

- 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

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

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

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

Continued

- 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.

Standard Errors

- Let's suppose that $E\left[\varepsilon_i^2|X\right] = \sigma^2$ and $E\left[\varepsilon_i\varepsilon_j|X\right] = 0$ 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?

$$\begin{aligned} Var\left(\hat{\beta}|X\right) &= Var\left(\hat{\beta} - \beta|X\right) \\ &= Var\left(\left(X'X\right)^{-1}X'\varepsilon|X\right) \\ &= \left(X'X\right)^{-1}X'Var\left(\varepsilon|X\right)X\left(X'X\right)^{-1} \end{aligned}$$

- Under the above assumption,

$$Var\left(\varepsilon|X\right) = \sigma^2 I_n$$

and so

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

Continued

- We can estimate σ^2 by

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

and the standard error for the OLS estimator is given by

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

- 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.

Newey-West Standard Errors

- Again,

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

- The Newey-West procedure boils down to an alternative way of looking at $Var(X'\varepsilon|X)$.
- If we suspect that the error terms may be heteroskedastic, but still independent, then

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

and our standard error for the OLS estimate is

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

Continued

- If we suspect correlation between error terms as well as heteroskedasticity, then

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

and our standard error for the OLS estimator is

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

Continued

- 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

$$\begin{aligned}\widehat{Var}\left(\hat{\beta}|X\right) &= \left(X'X\right)^{-1}\left(\sum_{i=1}^{n}\hat{\varepsilon}_i^2\cdot x_i x_i'\right)\left(X'X\right)^{-1} \\ &= \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\cdot x_i x_i'\right)\left(\frac{X'X}{n}\right)^{-1} \\ &= \frac{1}{n}\hat{E}\left(x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\cdot x_i x_i'\right)\hat{E}\left(x_i x_i'\right)^{-1}\end{aligned}$$

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

$$\begin{aligned}\widehat{Var}\left(\hat{\beta}|X\right) &= \left(X'X\right)^{-1}\left(\sum_{j=-k}^{k}\frac{k-|j|}{k}\left(\sum_{t=1}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_{i+j}\cdot x_i x_{i+j}'\right)\right)\left(X'X\right)^{-1} \\ &= \frac{1}{n}\hat{E}\left(x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{j=-k}^{k}\frac{k-|j|}{k}\left(\sum_{t=1}^{n}\hat{\varepsilon}_i\hat{\varepsilon}_{i+j}\cdot x_i x_{i+j}'\right)\right)\hat{E}\left(x_i x_i'\right)^{-1}\end{aligned}$$

MIT OpenCourseWare
http://ocw.mit.edu

15.450 Analytics of Finance

Fall 2010

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.