Economics 522A, Spring 2007

Lecture Note 8: Estimating Variance of OLS

First, some more OLS Algebra:

Consider the OLS fitted values:

$$\hat{y} := X\hat{\beta} = X(X'X)^{-1}X'y.$$

The matrix $X(X'X)^{-1}X'$ is worth studying in a bit more detail.

As notation, let

$$P_X := X(X'X)^{-1}X'.$$

We can think of P_X as defining a transformation $z \mapsto P_X z$. It is called a *projection matrix*, because it turns out to project z onto the space spanned by the columns of X. To see this, note that if z is already in the column space of X, so that we can write

$$z = [\xi_1 \cdots \xi_k]\alpha = X\alpha,$$

then

$$P_X z = X(X'X)^{-1}X'X\alpha = X\alpha = z.$$

The projection matrix has some useful properties:

- Symmetry: it is easy to see that $X'X$ is symmetric (write it out as a sum), so that $(X'X)^{-1}$ is symmetric, so:

$$\begin{aligned} P_X' &= \left(X(X'X)^{-1}X'\right)' \\ &= X\left[(X'X)^{-1}\right]'X' \\ &= X(X'X)^{-1}X' \\ &= P_X. \end{aligned}$$

- Idempotence: $P_X P_X = P_X$, since

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

Also, we can write

$$e = y - X\hat{\beta} = y - \hat{y} = (I_n - P_X)y.$$

The matrix $(I_n - P_X)$ is also symmetric and idempotent — try showing this yourself.

Intuitively, we are taking the column vector y, and decomposing it into two parts:

$$y = \hat{y} + e,$$

where $\hat{y} = P_X y$, and $e = (I_n - P_X)y$. As we have seen, the decomposition is such that $P_X y$ lies in the column space of X. As for the residual e, note that

$$\begin{aligned} X'e &= X'(I_n - P_X)y \\ &= X'(I_n - X(X'X)^{-1}X')y \\ &= \left(X' - X'X(X'X)^{-1}X'\right)y \\ &= (X' - X')y \\ &= 0. \end{aligned}$$

So e is *orthogonal* to X, and more generally, the transform $z \mapsto (I_n - P_X)\,z$ isolates the part of z orthogonal to the column space of X.

Now, on to properties of s^2 and $\hat{\sigma}^2$:

As before, we assume that $y|X \sim N(X\beta, \sigma^2 I_n)$, and that X has full column rank. Let $s^2 = (e'e)/(n-k)$.

Proposition 1

$$E[s^2|X] = \sigma^2.$$

Proof: There is a proof in Ruud, Ch.8, but we provide a slightly different proof, which might be easier to follow. First, recall that the *trace* of a square matrix is the sum of its diagonal elements: for

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, \qquad tr(A) = \sum_{i=1}^{n} a_{ii}.$$

Useful properties of the trace:

$$tr(A+B) = tr(A) + tr(B), \qquad tr(cA) = c \cdot tr(A) \text{ if } c \text{ scalar},$$

and if AB is square,

$$tr(AB) = tr(BA).$$

Note that

$$\begin{aligned} E(e|X) &= E[y - X\hat{\beta}|X] \\ &= E[y|X] - XE[\hat{\beta}|X] \\ &= E[y|X] - X\beta \\ &= 0. \end{aligned}$$

So

$$Var(e|X) = E[ee'|X].$$

We can also calculate the variance as

$$\begin{aligned}
Var(e|X) &= Var((I_n - P_X)y|X) \\
&= (I_n - P_X)Var(y|X)(I_n - P_X) \\
&= (I_n - P_X)\sigma^2 I_n (I_n - P_X) \\
&= \sigma^2 (I_n - P_X).
\end{aligned}$$

Next, consider $E(e'e|X)$. Notice that

$$e'e = \sum_{i=1}^{n} e_i^2.$$

So

$$\begin{aligned}
E(e'e|X) &= E[\sum_i e_i^2|X] \\
&= \sum_i E[e_i^2|X] \\
&= tr(V(e|X)) \\
&= tr(\sigma^2(I_n - P_X)) \\
&= \sigma^2[tr(I_n - P_X)] \\
&= \sigma^2[tr(I_n) - tr(P_X)]
\end{aligned}$$

Notice that

$$\begin{aligned}
tr(P_X) &= tr(X(X'X)^{-1}X') \\
&= tr((X'X)(X'X)^{-1}) \\
&= tr(I_k)
\end{aligned}$$

So we have

$$\begin{aligned}
E(e'e|X) &= \sigma^2[tr(I_n) - tr(I_k)] \\
&= \sigma^2(n-k).
\end{aligned}$$

Therefore,

$$E[s^2|X] = E[(e'e)/(n-k)|X] = \sigma^2.$$

$\square$

Using iterated expectations,

$$E[s^2] = E\left[E[s^2|X]\right] = \sigma^2.$$

By similar reasoning,

$$E[\hat{\sigma}^2|X] = \frac{n-k}{n}\sigma^2,$$

so $\hat{\sigma}^2$ is slightly biased, although the bias will be small if n is large relative to k.

Joint distribution of e and $\hat{\beta}$:

In order to say a bit more about the distribution of s^2, which is a function of e, it's useful to work out the distribution of $e = y - X\hat{\beta}$ in more detail.

Recall that for a multivariate normal variable $y \sim N(\mu, \Omega)$, a linear transform of y is also multivariate normal:

$$Ay \sim N(A\mu, A\Omega A').$$

Now consider

$$\begin{pmatrix} \hat{\beta} \\ e \end{pmatrix} = \begin{pmatrix} (X'X)^{-1}X'y \\ (I_n - P_x)y \end{pmatrix} = \begin{pmatrix} (X'X)^{-1}X' \\ I_n - P_X \end{pmatrix} y.$$

Since $y|X \sim N(X\beta, \sigma^2 I_n)$,

$$\begin{pmatrix} \hat{\beta} \\ e \end{pmatrix} \mid X \sim N(m, V),$$

where

$$m = \begin{pmatrix} (X'X)^{-1}X' \\ I_n - P_X \end{pmatrix} X\beta = \begin{pmatrix} (X'X)^{-1}X'X\beta \\ (I_n - P_X)X\beta \end{pmatrix} = \begin{pmatrix} \beta \\ 0 \end{pmatrix},$$

and

$$\begin{aligned} V &= \begin{bmatrix} (X'X)^{-1}X' \\ I_n - P_X \end{bmatrix} \sigma^2 I_n \left[X(X'X)^{-1}, (I_n - P_X)' \right] \\ &= \sigma^2 \begin{pmatrix} (X'X)^{-1} & (X'X)^{-1}X'(I_n - P_X)' \\ (I_n - P_X)X(X'X)^{-1} & (I_n - P_X)(I_n - P_X)' \end{pmatrix}. \end{aligned}$$

Using the symmetry and idempotence of $(I_n - P_X)$, and the fact that

$$(X'X)^{-1}X'(I_n - P_X) = (X'X)^{-1}X' - (X'X)^{-1}X'X(X'X)^{-1}X' = 0,$$

we can write

$$V = \sigma^2 \begin{pmatrix} (X'X)^{-1} & 0 \\ 0 & (I_n - P_X) \end{pmatrix}.$$

Since zero covariance implies independence for multivariate normal random variables, we see that e is independent of $\hat{\beta}$ conditional on X.

So, since $s^2 = (e'e)/(n-k)$ is a function of e, it follows that s^2 is independent of $\hat{\beta}$ conditional on X.

By similar arguments, we can show that e (and therefore s^2) is independent of $\hat{y} = P_X y$ conditional on X.