EC402 and FM437: additional handout.

Timothée Carayol

December 5, 2008

This is a (very) detailed proof of the unbiasedness of the s^2 estimator of σ_ϵ

Assume the model is:

$$Y = X\beta + \epsilon$$

(Where X is $N \times K$.) Then under $A1, A2, A3, A4GM$ ($A4GM$ is crucial here so let me spell it out: $Var\left(\boldsymbol{\epsilon}|X\right) = \sigma_\epsilon^2 I_n$), the following holds:

$$E\left(\sum_i (\hat{\epsilon_i}^2)|X\right) = (N-K)\sigma_\epsilon^2$$

A proof goes like this:

$$\begin{aligned} E\left(\sum_i (\hat{\epsilon_i}^2)|X\right) &= E\left(\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}|X\right) = E\left((M_X\boldsymbol{\epsilon})'(M_X\boldsymbol{\epsilon})|X\right) & (1) \\ &= E\left(\boldsymbol{\epsilon}' M_X M_X \boldsymbol{\epsilon}|X\right) \\ &= E\left(\boldsymbol{\epsilon}' M_X \boldsymbol{\epsilon}|X\right) & (2) \end{aligned}$$

In (1), I use the identity $M_X\boldsymbol{\epsilon} = \hat{\boldsymbol{\epsilon}}$; in (2) I use the idempotence of M_X.
Here is the property that I am going to need next:

Lemma 0.1. *$\forall\ (k,n) \in \mathbb{N}^2$, if A is a $(N \times K)$ matrix and B is a $(K \times N)$ matrix, then $tr\left(AB\right) = tr\left(BA\right)$.*

Proof. $C = AB \Rightarrow c_{ij} = \sum_{k=1}^K (a_{ik}b_{kj})$, hence $tr\left(AB\right) = \sum_{n=1}^N c_{nn} = \sum_{n=1}^N \sum_{k=1}^K (a_{nk}b_{kn}) = \sum_{k=1}^K \sum_{n=1}^N (b_{kn}a_{nk}) = \sum_{k=1}^K d_{kk}$ where $D = BA$, so $tr\left(AB\right) = tr\left(BA\right)$ □

Back to the main proof:

$$
\begin{aligned}
E\left(\sum_i (\hat{\epsilon_i}^2)|X\right) &= E\left(tr(\boldsymbol{\epsilon}' M_X \boldsymbol{\epsilon})|X\right) = E\left(tr(M_X \boldsymbol{\epsilon}\boldsymbol{\epsilon}')|X\right) \\
&= tr\left(E\left(M_X \boldsymbol{\epsilon}\boldsymbol{\epsilon}'|X\right)\right) \\
&= tr\left(M_X E\left(\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|X\right)\right) && (3) \\
&= tr\left(M_X(\sigma_\epsilon^2 I_n)\right) && (4) \\
&= \sigma_\epsilon^2 tr\left(M_X\right) \\
&= (N-K)\sigma_\epsilon^2 && (5)
\end{aligned}
$$

In (3), I use the linearity of conditional expectations (remember that the only stochastic thing in M_X is X, which is conditioned on in the expected value). In (4), I use the homoskedasticity assumption. In (5), I use the following lemma:

Lemma 0.2. *If P_X $(n \times n)$ is an orthogonal projection matrix on a vector space of dimension k (with basis given by the k columns of the $(n \times k)$ matrix X, so that $P_X = X(X'X)^{-1}X'$), then $tr(P_X) = k$.*

Proof. $tr(P_X) = tr\left(X(X'X)^{-1}X'\right) = tr\left(X'X(X'X)^{-1}\right) = tr\left(I_k\right)$ (I used *Lemma* (0.1) to commute within the trace operator.) □

Corollary 0.3. *If M_X $(n \times n)$ is the annihilator matrix of X, so that $M_X = I_n - P_X$, then $tr(M_X) = tr(I_n) - tr(P_X) = n - k$.*

Note that another way to think about *Corollary* (0.3) is to understand that M_X is also the orthogonal projection matrix on the orthogonal to the vector space spanned by the columns of X (i.e. the set of all vectors that are orthogonal to the columns of X), which is a vectorial space of dimension $n-k$: then *Corollary* (0.3) is a direct application of *Lemma* (0.2).

A useful reference on the geometry of OLS, if you are curious, would be chapter 2 of Davidson and McKinnon's "Econometric Theory and Methods".