Alastair Hall

ECON61001:
Econometric Methods

Autumn 2020

Solutions to Problem Set for Tutorial 2

1.(a) Note that $rank(X) = k$ implies $rank(X_1) = k_1$ (why?), and so $(X_1'X_1)^{-1}$ exists. By definition,

$$\hat{\gamma}_T \;=\; (X_1'X_1)^{-1}X_1'y.$$

Substituting for y, it follows that

$$\begin{aligned}\hat{\gamma}_T &= (X_1'X_1)^{-1}X_1'(X\beta_0 + u)\\ &= (X_1'X_1)^{-1}X_1'(X_1\beta_{0,1} + X_2\beta_{0,2} + u)\\ &= \beta_{0,1} + (X_1'X_1)^{-1}X_1'X_2\beta_{0,2} + (X_1'X_1)^{-1}X_1'u\end{aligned}$$

Using *CA2*, it follows that

$$E[\hat{\gamma}_T] \;=\; \beta_{0,1} + (X_1'X_1)^{-1}X_1'X_2\beta_{0,2} + (X_1'X_1)^{-1}X_1'E[u],$$

and so from *CA4*, it follows that

$$E[\hat{\gamma}_T] \;=\; \beta_{0,1} + (X_1'X_1)^{-1}X_1'X_2\beta_{0,2}.$$

$\hat{\gamma}_T$ is an unbiased estimator of $\beta_{0,1}$ if $\beta_{0,2} = 0$ and so X_2 does not belong in the model or $X_1'X_2 = 0$ that is, if the included regressors (X_1) and excluded regressors (X_2) are orthogonal. The latter condition implies that the part of y that can be linearly predicted by X_1 is linearly unrelated to the part of y that can be linearly predicted using X_2.

1.(b) If $k_2 = 1$ then we have

$$(X_1'X_1)^{-1}X_1'x_2 \;=\; \hat{\delta}_T$$

where $\hat{\delta}_T$ is the OLS estimator of the regression coefficient from the regression of x_2 on X_1. Therefore it follows from part (a) that $E[\hat{\gamma}_T] \;=\; \beta_{0,1} + \hat{\delta}_T\beta_{0,2}$.

1.(c) Let $\beta_{0,1} = [\alpha_1, \alpha_2]'$. Given this model, the true value of returns to education is α_2 and the estimated returns to education is given by the second element of $\hat{\gamma}_T$, $\hat{\gamma}_{T,2}$ say. From part (b) we have,

$$E[\hat{\gamma}_{T,2}] \;=\; \alpha_2 + \hat{\delta}_{T,2}\beta_{0,2} \tag{1}$$

where $\hat{\delta}_{T,2}$ is the slope coefficient from the regression of ability on education. Therefore, the bias of the estimator of the returns to education is

$$bias[\hat{\gamma}_{T,2}] \;=\; E[\hat{\gamma}_{T,2}] - \alpha_2 \;=\; \hat{\delta}_{T,2}\beta_{0,2}.$$

It would be anticipated that $\hat{\delta}_{T,2} > 0$ and $\beta_{0,2} > 0$, and so $bias[\hat{\gamma}_{T,2}] > 0$. Therefore in this simple setting, we expect the omission of innate ability from the wage equation to cause the estimator of the returns to education to be upward biased. (In this question, we maintained the assumption that X is fixed in repeated samples so that $\hat{\delta}_T$ is a constant. This is, of course, unrealistic in this example. In lecture 4, we extend our analysis to stochastic regressors in which case we could repeat the analysis here subject to certain conditions only replacing $\hat{\delta}_T$ by its expectation.)

2. Recall that for an $\ell \times \ell$ matrix $tr(C) = \sum_{i=1}^{\ell} C_{i,i}$ where $C_{i,j}$ is the $i-j^{th}$ element of C. Now consider $C = AB$ and $D = BA$ where A and B are $m \times n$ and $n \times m$, so that C is $m \times m$ and D is $n \times n$. By definition

$$C_{i,i} \;=\; \sum_{j=1}^{n} A_{i,j} B_{j,i}$$

and so

$$tr(C) \;=\; \sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j} B_{j,i}$$

Similarly,

$$D_{j,j} \;=\; \sum_{i=1}^{m} B_{j,i} A_{i,j}$$

and so

$$tr(D) \;=\; \sum_{j=1}^{n} \sum_{i=1}^{m} B_{j,i} A_{i,j}.$$

Since we can reverse the order of the summations and of scalars, $tr(D) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j} B_{j,i} = tr(C)$.

3. Using $\hat{\beta}_T - \beta_0 = (X'X)^{-1} X'u$, we have

$$e^p_{T+1} \;=\; u_{T+1} \,-\, x'_{T+1}(X'X)^{-1}X'u.$$

where u is the $T \times 1$ vector with t^{th} element u_t. To establish the desired result, it is easiest to define the $(T+1) \times 1$ vector v whose t^{th} element is u_t, and a $1 \times (T+1)$ vector n' given by

$$n' \;=\; [-\, x'_{T+1}(X'X)^{-1}X', 1].$$

Note that: $e^p_{T+1} = n'v$. We are given that $v \sim N(\,0,\, \sigma_0^2 I_{T+1})$, and so the prediction error is a linear combination of normal random variables, and so from Lemma 2.1 in the Lecture Notes it follows that

$$e^p_{T+1} \;\sim\; N(\,0,\, \sigma_0^2 n'n).$$

Multiplying out, we obtain

$$n'n \;=\; 1 \,+\, x'_{T+1}(X'X)^{-1}x_{T+1}.$$

4.(a) Notice that equation (2) on the problem set represents the true model - that is, it reduces to equation (1) - if we put $\beta_1 = \beta_{1,0}$ and $\beta_2 = 0_{k_2\times 1}$ (equal to a $k_2 \times 1$ vector of zeros). Therefore, equation (2) can be viewed as a correctly specified model (with the aforementioned values for the parameters), and so from lectures we have $Var[\tilde{\beta}] = \sigma_0^2 (X'X)^{-1}$. Using the partitions of β and X we have

$$Var[\tilde{\beta}] \;=\; \begin{bmatrix} Var[\tilde{\beta}_1] & Cov[\tilde{\beta}_1, \tilde{\beta}_2] \\ Cov[\tilde{\beta}_2, \tilde{\beta}_1] & Var[\tilde{\beta}_2] \end{bmatrix}$$

and so $Var[\tilde{\beta}_1] = \sigma_0^2 V_1$ where V_1 is the $k_1 \times k_1$ matrix defined by

$$(X'X)^{-1} \;=\; \begin{bmatrix} V_1 & C \\ C' & V_2 \end{bmatrix}.$$

Using the partition of $X = (X_1, X_2)$, we have

$$(X'X)^{-1} \;=\; \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1},$$

and so using the partitioned matrix inversion formula (Lemma 2.3 in the Lecture Notes), it follows that $V_1 = (X_1'M_2X_1)^{-1}$ which gives the desired result.

4.(b) By similar arguments to our analysis of OLS in lectures, $Var[\hat{\beta}_1] = \sigma_0^2 (X_1'X_1)^{-1}$. Since $\sigma_0^2 > 0$, $Var[\tilde{\beta}_1] - Var[\hat{\beta}_1]$ is psd iff $(X_1'M_2X_1)^{-1} - (X_1'X_1)^{-1}$. Using the hint, with $A = (X_1'M_2X_1)^{-1}$ and $B = (X_1'X_1)^{-1}$, it suffices to consider $B^{-1} - A^{-1}$. We have

$$\begin{aligned} B^{-1} - A^{-1} &= X_1'X_1 \;-\; X_1'M_2X_1 \\ &= X_1'X_1 \;-\; X_1'(I_T \;-\; X_2(X_2'X_2)^{-1}X_2')X_1 \\ &= X_1'X_1 \;-\; X_1'X_1 \;+\; X_1'X_2(X_2'X_2)^{-1}X_2'X_1 \\ &= X_1'X_2(X_2'X_2)^{-1}X_2'X_1 \;=\; C, \text{ say.} \end{aligned}$$

Notice that $C = DD'$ where $D = X_2(X_2'X_2)^{-1}X_2'X_1$ and so is psd by construction (see Tutorial 1 Question 3). Therefore, $Var[\tilde{\beta}_1] - Var[\hat{\beta}_1]$ is psd.

Note: this result implies that the inclusion of irrelevant regressors does not improve the efficiency of OLS estimators. Combining the results in Questions 1 and 4 we have the following:

- *the inclusion of irrelevant regressors in general leads to less efficient OLS estimators;*
- *the exclusion of relevant regressors in general leads to biased OLS estimators.*

The exception to these "in general" rules is where $X_1'X_2 = 0$ *- when* X_1 *and* X_2 *are linearly unrelated - because then* $Var[\tilde{\beta}_1] = Var[\hat{\beta}_1]$ *and* $E[\hat{\gamma}_T] \;=\; \beta_{0,1}$.