Alastair Hall

ECON61001:
Econometric Methods

Autumn 2020

Problem Set for Tutorial 2

The first question considers the consequences for OLS of omitting relevant regressors from the model.

1. Consider the linear regression model

$$y = X\beta_0 + u$$

and assume it satisfies Assumptions *CA1-CA6* discussed in lecture 1. Let $X = [X_1, X_2]$, $\beta_0 = [\beta_{0,1}', \beta_{0,2}']'$ and X_ℓ, $\beta_{0,\ell}$ are $T \times k_\ell$, $k_\ell \times 1$ for $\ell = 1, 2$. Suppose that a researcher estimates the following model by OLS,

$$y = X_1\gamma_* + \text{error}.$$

Let $\hat{\gamma}_T$ be the OLS estimator of γ_*.

(a) Show that $E[\hat{\gamma}_T] = \beta_{0,1} + (X_1'X_1)^{-1}X_1'X_2\beta_{0,2}$. Under what condition is $\hat{\gamma}_T$ an unbiased estimator of $\beta_{0,1}$. Interpret this condition in terms of the regression model.

(b) Now consider the case where $k_2 = 1$ with $X_2 = x_2$. Use part (a) to show that

$$E[\hat{\gamma}_T] = \beta_{0,1} + \hat{\delta}_T\beta_{0,2}$$

where $\hat{\delta}_T$ is the OLS estimator of the regression coefficients in $x_2 = X_1\delta +$ "error". (Here, to match our assumptions about X we treat $\hat{\delta}_T$ as a constant.)

(c) Consider the case where $k_1 = 2$ with $X_1 = [\iota_T, x_1]$ and the variables are defined as follows: y is log wage, ι_T is a $T \times 1$ vector of ones, x_1 is the number of years of education and x_2 denotes innate (intellectual) ability. Use the result in part(b) to suggest the likely direction of bias, if any, in the estimator of the return to education.

Question 2 establishes a result used in the proof that $\hat{\sigma}_T^2$ is an unbiased estimator of σ_0^2.

2. Show that, for two conformable matrices A and B, $trace(AB) = trace(BA)$.

Question 3 establishes a result quoted in Lectures as part of our discussion of prediction intervals

3. Consider the linear model

$$y_t = x_t'\beta_0 + u_t,$$

for $t = 1, 2, \ldots T+1$. Assume *CA1*, *CA2*, *CA4*-*CA6* are satisfied. Let X be the $T \times k$ matrix whose t^{th} row is x_t', and assume $rank(X) = k$. Define $\hat{\beta}_T$ to be the OLS estimator of β_0 based on sample $\{y_t, x_t;\ t = 1, 2, \ldots T\}$. Consider the prediction error

$$e_{T+1}^p \;=\; u_{T+1} \;-\; x_{T+1}'(\hat{\beta}_T \;-\; \beta_0),$$

discussed in Lecture 2. Show that

$$e_{T+1}^p \;\sim\; N\left(0,\, \sigma_0^2(1 \,+\, x_{T+1}'(X'X)^{-1}x_{T+1})\right).$$

In question 1, you establish that the omission of relevant regressors (most likely) leads to bias. In this question you consider the consequences of the inclusion of irrelevant regressors.

4. Consider the case where the true regression model is

$$y \;=\; X_1\beta_{1,0} \,+\, u \tag{1}$$

where y is a $(T \times 1)$ vector of observations on the dependent variable, X_1 is a $(T \times k_1)$ matrix of observed explanatory variables that are fixed in repeated samples, $rank(X_1) = k_1$, u is $(T \times 1)$ vector containing the unobservable error term, $E[u] = 0$ and $Var[u] = \sigma_0^2 I_T$, where $\sigma_0^2 > 0$. Let $\hat{\beta}_1$ be the OLS estimator of $\beta_{0,1}$ based on (1). Now consider $\tilde{\beta}_1$, the OLS estimator of β_1, based on estimation of the model

$$y \;=\; X_1\beta_1 \,+\, X_2\beta_2 \,+\, u \;=\; X\beta \,+\, u \tag{2}$$

where X_2 is a $(T \times k_2)$ matrix of observed explanatory variables that are fixed in repeated samples and $rank(X) = k$ where $X = (X_1, X_2)$ and $k = k_1 + k_2$. Thus, $\tilde{\beta}_1$ is defined via

$$\tilde{\beta} \;=\; \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{bmatrix} \;=\; (X'X)^{-1}X'y.$$

(a) Using the partitioned matrix inversion formula, show that $Var[\tilde{\beta}_1] = \sigma_0^2(X_1'M_2X_1)^{-1}$ where $M_2 = I_T - X_2(X_2'X_2)^{-1}X_2'$.

(b) Show that $\hat{\beta}_1$ is at least as efficient as $\tilde{\beta}_1$ that is, $Var[\tilde{\beta}_1] - Var[\hat{\beta}_1]$ is a positive semi-definite (psd) matrix. *Hint: For two conformable nonsingular matrices A and B, if $B^{-1} - A^{-1}$ is psd then $A - B$ is psd.*