Econometrics 522A
Spring, 2019
Hidehiko Ichimura, TA: Zejin Shi

PS #7 (Due 6 March, 2019)

1. Consider the following regression model with a measurement error problem:

$$y_i = 1 + x_i^* + z_i + u_i$$

where

$$\begin{pmatrix} x_i^* \\ z_i \\ u_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$$

and $x_i = x_i^* + v_i$ where $v_i \sim N(0, \sigma_v^2)$ where $\sigma_v^2 = 1$ and $\rho = 0$.

(a) Generate 1000 i.i.d. data from the model above and regress y_i on the constant term, x_i and z_i to verify that there is the attenuation bias using R and Python. State a way you can simulate data to verify the direction of inconsistency on the coefficient on x_i proved in class.

(b) Is the coefficient on z_i consistently estimated? Explain your answer.

(c) Theoretically, what will happen to the size of inconsistency on the OLS estimator of the coefficients on x_i and z_i, if σ_v^2 increases? Verify this by simulating the data with $\sigma_v^2 = 2$.

(d) What will happen to the size of inconsistency if ρ increases? Verify this by simulating the data with $\rho = 0.5$.

2. Consider the following simultaneous equations model

$$\begin{aligned} y_i^S &= \beta_1^S + \beta_2^S p_i + u_i^S \\ y_i^D &= \beta_1^D + \beta_2^D p_i + u_i^D \end{aligned}$$

where

$$\begin{pmatrix} u_i^S \\ u_i^D \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_S^2 & 0 \\ 0 & \sigma_D^2 \end{pmatrix} \right)$$

and p_i is determined via $y_i^S = y_i^D$. Let $y_i = y_i^S = y_i^D$.

(a) What is the equilibrium p_i in terms of u_i^S and u_i^D? State clearly the conditions under which the solution exists.

(b) Obtain the probability limit of the OLS estimator when observed y_i is regressed on the constant term and p_i under the standard assumptions.

(c) Discuss conditions under which the probability limit of the OLS estimator of the coefficient on p_i equals β_2^S or β_2^D.

3. Consider the following linear regression model:

$$y_i = 1 + x_i + z_i + u_i$$

where

$$\begin{pmatrix} x_i \\ z_i \\ u_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$$

where $\rho = 0.1$. Assume that the sampling is i.i.d.

(a) Explain why this model satisfies all the assumptions we made on the linear regression model including homoskedasticity.

(b) Using R and Python, generate 1000 i.i.d. data from the model above and select observations that satisfy $x_i > 0$. Using the selected observations, regress y_i on the constant term, x_i and z_i to verify, by simulation, that this type of sample selection still yields a consistent estimator.

(c) Explain why (b) is the case under correct model specification. Explain why this result is misleading when the model is misspecified.

(d) Using R and Python, generate 1000 i.i.d. data from the model above and select observations that satisfy $y_i > \epsilon_i$, where ϵ_i is the standard normal random variable. Using the selected observations, regress y_i on the constant term, x_i and z_i to verify, by simulation, that this type of sample selection yields an inconsistent estimator.

(e) Explain why (d) is the case.

4. Consider the following linear regression model:

$$\begin{aligned} y_t &= 1 + 0.3y_{t-1} + u_t \\ u_t &= 0.5u_{t-1} + v_t \end{aligned}$$

where v_t is i.i.d and y_0, u_0 and v_t for any t are mutually independent. Assume that y_0, u_0, and v_t all have the standard normal distribution.

(a) Using R and Python, generate 1000 observations from the model above. Using the generated observations, regress y_t on the constant term and y_{t-1} to verify, by simulation, that the OLS estimator is inconsistent under this model.

(b) Explain why (a) is the case.

(c) Show by simulation that if there is no serial correlation, then the OLS estimator is consistent. Explain why.

5. Consider the same model as in problem 3 without the sample selection problem.

(a) Show by simulation, using R and Python, that if we use the OLS estimator to estimate the coefficient on x_i by regressing y_i on a constant term and x_i, the estimator is inconsistent.

(b) Explain the direction and the size of the inconsistency by using auxiliary regression analysis.

(c) Do the same by setting $\rho = 0$. Explain the result in this case.

6. Suppose you have two omitted variables and two included variables in addition to the constant term. Extend the auxiliary regression analysis to discuss the directions of inconsistencies.