Problem Set 5

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

(Due: Send by March 6, 9pm)

[1] In the lecture we saw just two of the cases where the endogeneity problem might arise. This exercise covers two more such cases.

a) Suppose the model of interest is the following

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \varepsilon_t$$

and ε_t follows a first-order autoregressive process

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t$$

Show that the coefficients of this model $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ cannot be consistently estimated by OLS.

b) Consider the relation

$$y_t = \beta_1 + \beta_2 x_t + v_t \qquad (\star)$$

Suppose, however, that x_t can only be measured with some error, that is, if denote the measured value for x_t by $\tilde{x}_t$ we have

$$\tilde{x}_t = x_t + u_t$$

where u_t is a random measurement error that is mean zero and variance σ_u^2 and independent of x_t and v_t.

Now consider the OLS estimator b_2 of β_2. Show that

$$\text{plim } b_2 \neq \beta_2$$

to conclude that b_2 is not a consistent estimator of β_2.

c) Within the framework of part (b) show that β_1 cannot be consistently estimated either, i.e.

$$\text{plim } b_1 \neq \beta_1$$

Using this result conclude that inconsistency of one coefficient carries over to all other coefficients.

[2] Consider the simple model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

and suppose that we would like to use z_i as an instrument for x_i (but z might be neither exogenous nor strongly relevant).

a) Show that $b_{2,IV}$ is not unbiased.

b) Show that

$$\text{plim } b_{2,IV} = \beta_2 + \frac{\text{corr}(z, \varepsilon)}{\text{corr}(z, x)} \cdot \frac{\sigma_\varepsilon}{\sigma_x}$$

c) Argue that even if $\text{corr}(z, \varepsilon)$ is relatively low (so exogeneity assumption is almost satisfied), if $\text{corr}(z, x)$ is low (weak instrument problem) then the inconsistency in the IV estimator can be very large.

[3] Consider a linear model

$$y_{1i} = \alpha_2 y_{2i} + \alpha_3 y_{3i} + \beta_1 x_{1i} + \beta_2 x_{2i} + u_{1i}$$

where y_2 and y_3 are endogenous variables. Furthermore, let x_3 and x_4 be two other exogenous variables that are available to the econometrician.

a) Describe how one can obtain a consistent estimator for β_1 by a 2SLS procedure in this model.
In the first stage, obtain the fitted values

b) Show that a two-stage estimator which regresses y_2 on x_1, x_2 and x_3 to get $y_2 = \hat{y}_2 + \hat{v}_2$, y_3 on x_1, x_2 and x_4 to get $y_3 = \hat{y}_3 + \hat{v}_3$, and then (in the second stage) regresses y_1 on $\hat{y}_2$, $\hat{y}_3$, x_1 and x_2 does not necessarily yield consistent estimators. This part shows that if both y's are not regressed on the same set of x's, the resulting two stage regression estimates are not consistent.
(Hint: Show that the composite error is $\epsilon_1 = u_1 + \alpha_2 \hat{v}_2 + \alpha_3 \hat{v}_3$ and $\sum_{i=1}^N \hat{\epsilon}_{1i} \hat{y}_{2i} \neq 0$, because $\sum_{i=1}^N \hat{y}_{2i} \hat{v}_{3i} \neq 0$. The latter does not hold because $\sum_{i=1}^N x_{3i} \hat{v}_{3i} \neq 0$)

[4] Suppose that the market for a certain good is expressed by the following equations:

$$\begin{aligned} D_t &= \alpha_0 - \alpha_1 P_t + \alpha_2 X_t + u_{1t} \qquad (\alpha_1, \alpha_2 > 0) \\ S_t &= \beta_0 + \beta_1 P_t + u_{2t} \qquad (\beta_1 > 0) \\ D_t &= S_t = Q_t \end{aligned}$$

a) Discuss whether one can estimate α_1 consistently by OLS?

b) Suppose you wanted to estimate this system by OLS. Derive the asymptotic bias in the OLS estimator for β_1.

[5] Consider the single equation model

$$y_i = x_i \beta + \varepsilon_i$$

where y_i and x_i are both real-valued. Let b_{IV} denote the IV estimator of β using as an instrument a dummy variable z_i (takes only the values 0 and 1). Find a simple expression b_{IV} in this context.

Note. Example 8.2 in Greene is a similar example where he computes the IV estimator explicitly for a case when the instrument is a dummy variable. But it seems 7th edition doesn't have the computational details so I have posted these pages from my copy. It's a very good example, study that example together with Example 8.4 carefully. In general, Greene sections 8.1 through 8.4 are relevant to what we have done in class.

[6] Consider the structural equation

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \tag{1}$$

with x_i treated as endogenous so that $E[x_i\varepsilon_i] \neq 0$. Assume y_i and x_i are scalar. Suppose we have a scalar instrument z_i which satisfies

$$E[x_i\varepsilon_i] \neq 0$$

so in particular $E[\varepsilon_i] = 0$, $E[z_i\varepsilon_i] = 0$ and $E[z_i^2\varepsilon_i] = 0$.

a) Should x_i^2 be treated as endogenous or exogenous?

b) Suppose we have a scalar instrument z_i which satisfies

$$x_i = \gamma_0 + \gamma_1 z_i + u_i \tag{2}$$

with u_i independent of z_i and mean zero.
Consider using $[1\ z_i\ z_i^2]$ as instruments. Is this a sufficient number of instruments? (Would this be just-identified, over-identified, or under-identified)?

c) Write out the reduced form equation for x_i^2. Under what condition on the reduced form parameters (2) are the parameters in (1) identified?

[7] Consider the structural equation and reduced form

$$\begin{aligned} y_i &= \beta x_i^2 + \varepsilon_i \\ x_i &= \gamma z_i + u_i \\ E[z_i\varepsilon_i] &= 0 \\ E[z_i u_i] &= 0 \end{aligned}$$

with x_i^2 treated as endogenous so that $E[x_i^2\varepsilon_i] \neq 0$. For simplicity assume there are no intercepts in the regressions. All variables are scalars and assume $\gamma \neq 0$.

Consider the following estimator. First estimate γ by OLS of x_i on z_i and obtain the fitted values $\hat{x}_i = \hat{\gamma} z_i$. Second, estimate β by OLS of y_i on $\hat{x}_i^2$.

a) Write out this estimator $\hat{\beta}$ explicitly as a function of the sample data.

b) Find its probability limit as $n \to \infty$

c) In general, is $\hat{\beta}$ consistent estimator for β? Is there a reasonable condition under which $\hat{\beta}$ is consistent?

[8] In the lecture, we derived IV estimator by using the method of moments (then also by using a GLS transformation). We can also obtain the OLS estimator using the method of moments (MoM). Consider a linear model (in matrix form)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$$

and assume that the standard Gauss-Markov assumptions hold. We would like to obtain an estimator for the coefficient vector $\boldsymbol{\beta}$.

a) Which assumption provides us a population moment condition that can be used to estimate $\boldsymbol{\beta}$?

b) Write the corresponding sample analog of the population moment condition?

c) Derive the MoM estimator using the sample moment condition and compare with the standard OLS estimator.

[9] Consider Hansen and Singleton (1982) model that we discussed in class. We showed that the intertemporal utility maximization problem gives the following Euler equation

$$E_t\left\{\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_{t+1})\right\}=1$$

where E_t denotes the expectation operator conditional upon time t information, C_t denotes consumption in period, C_t denotes consumption in period t, r_{t+1} is the return on financial wealth, δ is a discount rate and γ is the coefficient of relative risk aversion. Assume that we have a time series of observations on consumption levels, returns and instrumental variables $\mathbf{z}_t$.

a) Show how the above condition can be written as a set of *unconditional* moment conditions. Explain how we can estimate δ and γ consistently from these moment conditions.

b) What is the minimum number of moment conditions that is required? What do we (potentially) gain by having more moment conditions?

c) How can we improve the efficiency of the estimator for a given set of moment conditions? In which case does this not work?

d) Explain what we mean by *overidentifying restrictions.* Is this a good or a bad thing?

e) Explain how the overidentifying restrictions test is performed. What is the null hypothesis that is tested? What do you conclude if the test rejects?

[10] Consider again Hansen and Singleton (1982) model that we discussed in class. But now assume that the momentary utility function of the consumer is given as $U(C_t)=\ln C_t$.

a) Set up the consumer's optimization problem, derive the first order conditions and obtain Euler equation.

b) Write the population moment condition and the corresponding sample analog?

c) Describe how you would proceed to compute the GMM estimator for the population parameters δ?

d) Prove that given two random variables X_1 and X_2, $E[X_1|X_2]=0$ implies that $E[X_1 g(X_2)]=0$ for any function g. Where did we use this result in the lecture? Give an economic interpretation of this result.

[11] The sum of two independent uniform variables with support $(0,\theta)$ has the triangular distribution. The probability density function for this distribution is given by

$$f(x)=\begin{cases}\theta-\frac{\theta^2}{2}x & 0\le x\le\frac{2}{\theta}\\ 0 & elsewhere\end{cases}$$

a) Show that this is a proper density function.

b) Find $E[X]$

c) Find the method of moments estimator for θ.

[12] As an empirical exercise please read the paper *The Colonial Origins of Comparative Development- An Empirical Investigation* by Acemoglu, Johnson and Robinson (2001). Then by using the data `institutions.xls` replicate three tables presented in the lecture slides.