Econometrics

Homework 6

Problem 1. Consider a simple regression model (with an intercept):

$$Y_i = \beta_0 + \beta_1 X_i + U_i,$$

and the IV estimator of β_1:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \left(Z_i - \bar{Z}\right) Y_i}{\sum_{i=1}^n \left(Z_i - \bar{Z}\right) X_i},$$

where

$$\bar{Z} = n^{-1} \sum_{i=1}^n Z_i.$$

Suppose that Z_i is a dummy variable. Show that $\hat{\beta}_1$ can be written as

$$\hat{\beta}_1 = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0},$$

where $\bar{X}_0$ and $\bar{Y}_0$ are the sample averages of X_i and Y_i over the part of the sample with $Z_i = 0$, and $\bar{X}_1$ and $\bar{Y}_1$ are the sample averages of X_i and Y_i over the part of the sample with $Z_i = 1$, by following the steps below. Let n_1 be the number of observations in the part of the sample with $Z_i = 1$. Let n_0 be the number of observations in the part of the sample with $Z_i = 0$. Hint: $n_1 = \sum_{i=1}^n Z_i$, $n_0 = n - \sum_{i=1}^n Z_i = \sum_{i=1}^n (1 - Z_i)$. $\bar{Y}_1 = \sum_{i=1}^n Z_i Y_i / n_1$ and $\bar{Y}_0 = \sum_{i=1}^n (1 - Z_i) Y_i / n_0$

(i) Show that $\sum_{i=1}^n \left(Z_i - \bar{Z}\right) Y_i = \sum_{i=1}^n Z_i \left(Y_i - \bar{Y}\right)$.

(ii) Show that $\sum_{i=1}^n Z_i \left(Y_i - \bar{Y}\right) = n_1 \left(\bar{Y}_1 - \bar{Y}\right)$.

(iii) Show that $\bar{Y} = \left(n_1 \bar{Y}_1 + n_0 \bar{Y}_0\right) / n$.

(iv) Show that $n_1 \left(\bar{Y}_1 - \bar{Y}\right) = (n_1 n_0 / n) \left(\bar{Y}_1 - \bar{Y}_0\right)$.

(v) Show how (i)-(iv) imply that $\hat{\beta}_1 = \left(\bar{Y}_1 - \bar{Y}_0\right) / \left(\bar{X}_1 - \bar{X}_0\right)$.

Problem 2. Consider the model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + U_i, \tag{1}$$

where X_{1i} is an exogenous regressor and X_{2i} is an endogenous regressor. Assume that data are iid and conditions required for LLNs hold. For each of the following statements, indicate true or false, and explain your answer.

(i) Let $\hat{\beta}_1$ denote the estimated coefficient on X_1 in the OLS regression of Y against a constant, X_1, and X_2. Since X_1 is exogenous, $\hat{\beta}_1$ consistently estimates β_1.

(ii) Let $\hat{\beta}_1$ denote the estimated coefficient on X_1 in the OLS regression of Y against a constant and X_1. If $Cov(X_{1i}, X_{2i}) = 0$, then $\hat{\beta}_1$ consistently estimates β_1.

(iii) Consider the following IV estimator of β_2 that uses X_1 as an IV:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)Y_i}{\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)X_{2i}}.$$

If $Cov(X_{1i}, X_{2i}) \neq 0$ and $\beta_1 = 0$, then $\hat{\beta}_2$ consistently estimates β_2.

Problem 3. Suppose that the linear model

$$\text{PS} = \beta_0 + \beta_1 \text{Funds} + \beta_2 \text{Risk} + U$$

satisfies $\text{E}[U] = \text{E}[U \cdot \text{Funds}] = \text{E}[U \cdot \text{Risk}] = 0$. PS is the percentage of a person's savings invested in the stock market, Funds is the number of mutual funds that the person can choose from, and Risk is some measure of risk tolerance (larger Risk means the person has a higher tolerance for risk).

(i) If Funds and Risk are positively correlated, does the slope coefficient in the simple regression of PS on Funds overestimate or underestimate β_1, in large samples?

(ii) We are unable to observe Risk directly, but we have data on the amount of life insurance a worker has, Insurance. Assume that Insurance is noisy measure of Risk, Insurance $=$ Risk $+ e$, with $\text{E}[e] = \text{E}[\text{Risk} \cdot e] = \text{E}[\text{Funds} \cdot e] = \text{E}[eU] = 0$. Will the OLS estimate of the coefficient on Funds in a regression of PS on Funds and Insurance be a consistent estimate of β_1?

(iii) Suppose we also have data on how often a worker gambles, Gamble. Assume that Gamble is an independent noisy measure of Risk, Gamble $=$ Risk $+ v$, with $\text{E}[v] = \text{E}[vU] = \text{E}[ve] = \text{E}[\text{Risk} \cdot v] = \text{E}[\text{Funds} \cdot v] = 0$. Explain how we can consistently estimate β_1 using our data on PS, Funds, Insurance, and Gamble.

Problem 4. Aggregate demand Q_D for a certain commodity is determined by its price P, aggregate income Y, and population, POP,

$$Q_D = \beta_1 + \beta_2 P + \beta_3 Y + \beta_4 POP + U^D$$

and aggregate supply is given by

$$Q_S = \alpha_1 + \alpha_2 P + U^S$$

where U_D and U_S are independently distributed error terms: U_D and U_S are independent from all other variables and they are also independent from each other. Remember that the quantity and the price are determined simultaneously in the equilibrium $Q_S = Q_D = Q$. We observe only the equilibrium values Q so that the observed price must satisfy the equation (demand $=$ supply):

$$\beta_1 + \beta_2 P + \beta_3 Y + \beta_4 POP + U^D = \alpha_1 + \alpha_2 P + U^S.$$

(i) Show that the OLS (ordinary least squares) estimator of α_2 will be inconsistent if OLS is used to fit the supply equation.

(ii) Show that a consistent estimator of α_2 is

$$\widetilde{\alpha}_2 = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(Q_i - \overline{Q}\right)}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(P_i - \overline{P}\right)}.$$

$(\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i,\ \overline{Q} = n^{-1}\sum_{i=1}^{n} Q_i,\ \overline{P} = n^{-1}\sum_{i=1}^{n} P_i.)$

Problem 5. In an econometric model, we say that a parameter is identified if we can recover its value perfectly given the joint distribution of the observable variables. Suppose that (Y, X) is the observable variables and U is the unobservable variable.

(i) Suppose that $Y = \beta_0 + \beta_1 X + U$ and $\mathrm{E}[U] = \mathrm{E}[XU] = 0$. Show that β_1 is identified. I.e., if you know the joint distribution of (Y, X), how do you determine the value of the parameter β_1?

(ii) Suppose that Y is binary and $Y = 1(\beta_0 + \beta_1 X \geq U)$ and U is a standard normal $(\mathrm{N}(0,1))$ random variable that is independent of X. If you know the joint distribution of (Y, X), how do you determine the value of the parameter β_1? Hint: $\mathrm{E}[Y \mid X] = \mathrm{E}[1(\beta_0 + \beta_1 X \geq U) \mid X] = \Phi(\beta_0 + \beta_1 X)$, where Φ is the standard normal CDF.

Problem 6. Let (X_i, Y_i), $i = 1, ..., n$ be an i.i.d. random sample where $Y_i \geqslant 0$ and $X_i \geqslant 0$ is a discrete random variable for all i. The conditional density of Y given X belong to the family:

$$f_{Y|X}(y|x, \lambda) = \frac{\lambda \exp(-\lambda y)(\lambda y)^x}{x!},$$

$y \geqslant 0$, $\lambda > 0$, i.e., the conditional density of Y given X is $f_{Y|X}(\cdot|\cdot, \lambda_*)$ for some $\lambda_* > 0$. Write the likelihood function for estimating λ_*. Provide the maximum likelihood estimator for λ_* as a solution of an equation. Give the asymptotic distribution for the maximum likelihood estimator, i.e. find the asymptotic variance V_{ML} of

$$\sqrt{n}\left(\hat{\lambda}_{ML} - \lambda_*\right) \to_d \mathrm{N}(0, \mathrm{V}_{ML}).$$

Suggest a consistent estimator of V_{ML}.

Problem 7. Define a density function

$$f(x \mid \theta) = \begin{cases} \left(1 + \frac{1-2\theta}{\theta-1}\right) x^{\frac{1-2\theta}{\theta-1}} & x \in (0,1) \\ 0 & x \notin (0,1), \end{cases}$$

where $0 < \theta < 1$ is a parameter. $X_1, ..., X_n$ is an independent and identically distributed sample with true density $f(\cdot \mid \theta_*)$ for some θ_*.

(i) Show that $f(\cdot \mid \theta)$ is a probability density function, for all $0 < \theta < 1$.

(ii) Show that $\theta_* = \int_0^1 x f(x \mid \theta_*)\,\mathrm{d}x$. I.e., in this parametrization, θ_* is also the population mean.