Econometrics

Homework 7

Problem 1. that (Y_i, X_i, Z_i), $i = 1, ..., n$ is a sequence of i.i.d. discrete random vectors and $Y_i \in \{0, 1, 2\}$, $Z_i \in \{0, 1\}$ and $X_i \in \{0, 1\}$.

(i) Show that for any $a \in \{0, 1\}$, we have

$$\begin{aligned} \mathrm{E}\left[Y_i | X_i = a\right] = & \mathrm{E}\left[Y_i | X_i = a, Z_i = 0\right] \mathrm{P}\left[Z_i = 0 | X_i = a\right] \\ & + \mathrm{E}\left[Y_i | X_i = a, Z_i = 1\right] \mathrm{P}\left[Z_i = 1 | X_i = a\right]. \end{aligned}$$

(ii) Show $\mathrm{E}\left[Z_i X_i\right] = \mathrm{P}\left[Z_i = 1, X_i = 1\right]$.

(iii) Show $\mathrm{E}\left[\mathrm{E}\left[Z_i | X_i = 1\right] X_i\right] = \mathrm{E}\left[Z_i X_i\right]$.

(iv) Show that $\hat{\theta} = \frac{\sum_{i=1}^n Z_i X_i}{\sum_{i=1}^n X_i}$ is a consistent estimator of $\theta = \mathrm{P}\left[Z_i = 1 | X_i = 1\right]$.

(v) Find a formula for σ^2 such that

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \to_d \mathrm{N}\left(0, \sigma^2\right).$$

Problem 2. Let $\{(Y_i, X_i, D_i)\}_{i=1}^n$ be a sequence of i.i.d. observations. D_i is a dummy variable. Consider the following binary choice model:

$$Y_i = 1\left(\beta_0 + \beta_1 X_i + \beta_2 X_i D_i \geqslant U_i\right),$$

where the conditional CDF of U_i is given by

$$\mathrm{P}\left[U_i \leqslant t | X_i, D_i\right] = \frac{\exp(t)}{1 + \exp(t)}.$$

(i) Define and derive the expression of the log-likelihood function for the i.i.d. observations $\{(Y_i, X_i, D_i)\}_{i=1}^n$.

(ii) Derive the average derivative (or average partial effect) with respect to X_i in terms of the observations and the parameters.

(iii) Let the MLE's for β_0, β_1 and β_2 be denoted by $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$. Provide an estimator of the average derivative in (ii).

Problem 3. In this question, you will derive the asymptotic distribution of the OLS estimator under endogeneity. Consider the usual linear regression model (without intercept) $Y_i = \beta X_i + U_i$. Assume, however, that X_i is endogenous:

$$E\left(X_i U_i\right) = \mu \neq 0,$$

where μ is unknown. Let $\hat{\beta}_n$ denote the OLS estimator of β. Make the following additional assumptions:

A1. Data are iid.
A2. $0 < Q = E\left(X_i^2\right) < \infty$.
A3. $0 < E\left(U_i - \delta X_i\right) X_i^2 < \infty$, where $\delta = Q^{-1}\mu$.

(i) Find the probability limit of $\hat{\beta}_n$.

(ii) Re-write the model as $Y_i = (\beta + \delta)X_i + (U_i - \delta X_i)$ and find $E\left(X_i(U_i - \delta X_i)\right)$.

(iii) Using the result in (ii), derive the asymptotic distribution of $\hat{\beta}_n$ and find its asymptotic variance. Explain how this result differs from the asymptotic normality of OLS with exogenous regressors.

(iv) Can $\hat{\beta}_n$ and its asymptotic distribution be used for constructing a confidence interval about β? Explain why or why not.

(v) Suppose that the errors U_i's are homoskedastic:

$$E\left(U_i^2 | X_i\right) = \sigma^2 = constant.$$

Consider the usual estimator of the asymptotic variance of OLS designed for a model with homoskedastic errors and exogenous regressors:

$$\left(n^{-1}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_n X_i\right)^2\right)\left(n^{-1}\sum_{i=1}^{n} X_i^2\right)^{-1}.$$

Is it consistent for the asymptotic variance of the OLS estimator if X_i's are in fact endogenous? Explain why or why not.

Problem 4. 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.