Econometrics

Homework 5

Problem 1. Consider the following simple regression without an intercept:

$$Y_i = \beta X_i + U_i.$$

Assume the observations (Y_i, X_i), $i = 1, 2, ..., n$ are iid. Assume $E(X_i U_i) = 0$, $E(X_i^2 U_i^2) < \infty$ and $0 < E(X_i^2) < \infty$.
(a) Provide the expression of the OLS estimator $\widehat{\beta}_n$ and show it is a consistent estimator of β.
(b) Show that

$$\sqrt{n}\left(\widehat{\beta}_n - \beta\right) \to_d N(0, V), \text{ where } V = \frac{E(X_i^2 U_i^2)}{E(X_i^2)^2}.$$

(c) How to construct a consistent estimator of V? You do not need to show your estimator is consistent.

Problem 2. (a) Consider the following simple regression model:

$$Y_i = \alpha + \beta X_i + U_i.$$

Suppose the observations (Y_i, X_i), $i = 1, 2, ..., n$ are iid. Assume $E|U_i| < \infty$, $E|X_i| < \infty$ and $E(U_i) = 0$. Let $\widetilde{\beta}_n$ be any consistent estimator of β (not necessarily the OLS estimator). Define the following estimator for α:

$$\widetilde{\alpha}_n = \overline{Y}_n - \widetilde{\beta}_n \overline{X}_n,$$

where $\overline{Y}_n = n^{-1}\sum_{i=1}^n Y_i$ and $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$. Prove that $\widetilde{\alpha}_n$ is a consistent estimator of α. Hint: Show $\overline{Y}_n = \alpha + \beta\overline{X}_n + \overline{U}_n$, where $\overline{U}_n = n^{-1}\sum_{i=1}^n U_i$.
(b) Consider the following regression model without a regressor:

$$Y_i = \alpha + U_i.$$

Suppose the observations Y_i, $i = 1, 2, ..., n$ are iid and $E(Y_i^2) < \infty$. Assume $E(U_i) = 0$. What is the expression of the OLS estimator $\widehat{\alpha}_n$? Show that $\sqrt{n}(\widehat{\alpha}_n - \alpha) \to_d N(0, V)$ and find V.

Problem 3. Let Y be the face number showing when a die is rolled. Define X as

$$X = \begin{cases} Y & \text{if } Y \text{ is even,} \\ 0 & \text{if } Y \text{ is odd.} \end{cases}$$

Let $R(Y|X)$ denote the best linear approximation to the conditional expectation $E(Y|X)$. $R(Y|X) = \beta_0 + \beta_1 X$, where

$$(\beta_0, \beta_1) = \underset{b_0, b_1}{\operatorname{argmin}} E\left[(E(Y|X) - b_0 - b_1 X)^2\right].$$

Calculate $E\left[(Y - R(Y|X))^2\right]$ and $E\left[(Y - E(Y|X))^2\right]$.

Problem 4. Consider a simple model to estimate the effect of personal computer (PC) ownership on college grade point average for graduating seniors at a large public university:

$$GPA = \beta_0 + \beta_1 PC + u,$$

where PC is a binary variable indicating PC ownership.

(i) Why might PC ownership be correlated with u?

(ii) Explain why PC is likely to be related to parents' annual income. Does this mean parental income is a good IV for PC? Why or why not?

(iii) Suppose that, four years ago, the university gave grants to buy computers to roughly one-half of the incoming students, and the students who received grants were randomly chosen. Carefully explain how you would use this information to construct an instrumental variable for PC.

Problem 5. Suppose we observe the i.i.d. random sample $\{(Y_i, X_i)\}_{i=1}^n$. Denote $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$, $\overline{Y}_n = n^{-1}\sum_{i=1}^n Y_i$, $\mu_X = \mathrm{E}[X_i]$ $(\mu_X \neq 0)$ and $\mu_Y = \mathrm{E}[Y_i]$. We are interested in μ_Y/μ_X. Derive the asymptotic distribution of $\sqrt{n}\left(\overline{Y}_n/\overline{X}_n - \mu_Y/\mu_X\right)$. Hint: Write

$$\begin{aligned}
\frac{\overline{Y}_n}{\overline{X}_n} &= \frac{\overline{Y}_n}{\mu_X}\cdot\left(\frac{\mu_X}{\overline{X}_n} - 1\right) + \frac{\overline{Y}_n}{\mu_X} \\
&= -\frac{\overline{Y}_n}{\mu_X\overline{X}_n}\cdot\left(\overline{X}_n - \mu_X\right) + \frac{\overline{Y}_n}{\mu_X} \\
&= -\left(\frac{\overline{Y}_n}{\mu_X\overline{X}_n} - \frac{\mu_Y}{\mu_X^2} + \frac{\mu_Y}{\mu_X^2}\right)\cdot\left(\overline{X}_n - \mu_X\right) + \frac{\overline{Y}_n}{\mu_X}
\end{aligned}$$

You may use the following result: $W_n \to_d N(0, \sigma^2)$ and $\theta_n \to_p 0$, then $\theta_n W_n \to_p 0$.

Problem 6. Suppose that you wish to estimate the effect of class attendance on student performance. $stndfnl$ is the standardized outcome on a final exam. A basic model is

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + u,$$

where $atndrte$ is percentage of classes attended, $priGPA$ is prior college grade point average, and ACT is the achievement test score.

(i) Let $dist$ be the distance from the students' living quarters to the lecture hall. Do you think $dist$ is uncorrelated with u?

(ii) Assuming that $dist$ and u are uncorrelated, what other assumption must $dist$ satisfy to be a valid IV for $atndrte$?

(iii) Suppose we add the interaction term $priGPA \times atndrte$:

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA \times atndrte + u.$$

If $atndrte$ is correlated with u, then, in general, so is $priGPA \times atndrte$. What might be a good IV for $priGPA \times atndrte$? Hint: If $E(u|priGPA, ACT, dist) = 0$, as happens when $priGPA$, ACT, and $dist$ are all exogenous, then any function of $priGPA$ and $dist$ is uncorrelated with u.