ECMA31000: Introduction to Empirical Analysis
Exam 2 2020

Question 1: (20 points) Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $\mathrm{E}\left(|X_n|\right) = \frac{1}{n}$ for all n.

a) Prove that $X_n \xrightarrow{p} 0$. **(5 points)**

b) Is it true that $\mathrm{E}\left(|X_n|^{\frac{3}{2}}\right) \to 0$? Prove it or find a counterexample. **(6 points)**

c) Is it true that $X_n \xrightarrow{a.s.} 0$? Prove it or find a counterexample. **(9 points)**

Question 2: (35 points) Suppose you observe an iid sample $\{X_i\}_{i=1}^n$, where the X_i have a continuous uniform distribution on $[-\theta, \theta]$. That is: $X_i \sim U\left[-\theta, \theta\right]$, for some $\theta > 0$.

a) Construct a consistent method of moments estimator of θ, denote it by $\hat{\theta}_{MOM}$, and prove its consistency. Is your estimator unbiased? **(6 points)**

b) Derive the asymptotic distribution of $\hat{\theta}_{MOM}$. Write out all of the steps in detail, and justify your use of any theorems discussed in class. **(6 points)**

c) Construct an asymptotic $1 - \alpha$ confidence set for θ using $\hat{\theta}_{MOM}$. Explain how to use your confidence set to construct a test of asymptotic size α of the null hypothesis $H_0 : \theta = 1$ vs. the alternative $H_1 : \theta \neq 1$. **(5 points)**

d) Show that the Maximum Likelihood estimator of θ is given by

$$\hat{\theta}_{ML} = \max_{1\leq i\leq n} |X_i|,$$

and prove that $\hat{\theta}_{ML} \xrightarrow{p} \theta$. **(6 points)**

e) Show that $n\left(\theta - \hat{\theta}_{ML}\right) \xrightarrow{d} X$ where X has an exponential distribution with CDF

$$F_X\left(x\right) = \begin{cases} 0 & x < 0, \\ 1 - \exp\left(-\frac{x}{\theta}\right) & x \geq 0. \end{cases}$$

Did you use the CLT in your argument? Why or why not? You may use the fact that $\lim_{n\to\infty}\left(1 - \frac{k}{n}\right)^n = \exp\left(-k\right)$. **(5 points)**

f) Construct an asymptotic $1 - \alpha$ confidence set for θ using $\hat{\theta}_{ML}$. Prove your claims. **(7 points)**

Question 3: (25 points) Consider the model

$$y_i = \beta x_i^* + u_i,$$

where $(y_i, x_i^*, u_i)'$ is a random vector with y_i, x_i^*, u_i all scalar random variables, β is an unobserved scalar constant, and only y_i is observed. You wish to estimate β, but you observe a noisy measurement of x_i^*, given by

$$x_i = x_i^* (1 + v_i),$$

for some unobserved error term v_i, where $\{v_i\}_{i \geq 1}$ is iid and independent of $\{x_i^*\}_{i \geq 1}$ and $\{u_i\}_{i \geq 1}$. Assume $E(x_i^* u_i) = E(u_i) = E(v_i) = 0$. You observe an iid sample $\{y_i, x_i\}_{i=1}^n$, and you regress y on x. There is no intercept in this regression. For what follows, assume all moments you need exist. You may use the fact that if a and b are independent random variables, then for any functions $f, g : \mathbb{R} \to \mathbb{R}$, $f(a)$ is independent of $g(b)$, and $E(f(a) \cdot g(b)) = E(f(a)) \cdot E(g(b))$.

a) Define endogeneity. Write out as much notation as you need to give a good definition. **(3 points)**

b) You wish to estimate β. Write out the regression model that is being estimated in this question, clearly define the error term, and show that there is an endogeneity problem in this regression. **(5 points)**

c) Write a formula for the OLS estimator, $\hat{\beta}_{OLS}$, and find its probability limit. Is $\hat{\beta}_{OLS}$ a consistent estimator of β? Explain. **(6 points)**

d) Let the probability limit of $\hat{\beta}_{OLS}$ be $\bar{\beta}$. Is $\left| \bar{\beta} \right| < |\beta|$, or is it not possible to determine this? Explain. **(3 points)**

e) Suppose you plan to use a constant random variable $z = 1$ as an instrument. Write down the IV estimator using this instrument. **(2 points)**

f) What conditions must hold for $z = 1$ to produce a consistent estimate of β using your IV estimator from part e)? Prove your claims. **(6 points)**

Question 4: (40 points): Let y be a scalar random variable, x be a $(k+1) \times 1$ random vector, and z be a $(k+1) \times 1$ random vector which has no elements in common with x. Consider the model

$$y = x'\beta + u.$$

Suppose you observe an iid sample of $\{y_i, x_i, z_i\}_{i=1}^n$ and assume that $E(u|x, z) = 0$ and $Var(u|x, z) = \sigma^2$. Assume $E(xx'), E(zx'), E(zz')$ exist and are invertible, and assume $E(u^2 xx')$ and $E(u^2 zz')$ exist.

a) Write the model in the form

$$Y = X\beta + U; \quad E(U|X, Z) = 0, Var(U|X, Z) = \sigma^2 I_n.$$

where Y and U are $n \times 1$ vectors, X is an $n \times (k+1)$ matrix, Z is an $n \times (k+1)$ matrix, and β

is a $(k+1) \times 1$ vector. Describe each of these objects explicitly. Show that $E(U|X, Z) = 0$ and $Var(U|X, Z) = E(UU'|X, Z) = \sigma^2 I_n$ under the stated assumptions. **(4 points)**

b) Let $\hat{\beta}_{OLS}$ be the OLS estimate of β in a regression of y on x (and <u>not</u> z). Derive $E\left(\hat{\beta}_{OLS}|X, Z\right)$ and $Var\left(\hat{\beta}_{OLS}|X, Z\right)$. **(4 points)**

c) Let $\hat{\beta}_{IV}$ be the IV estimate of β which uses the vector z (and <u>not</u> x) as the set of instruments. Derive $E\left(\hat{\beta}_{IV}|X, Z\right)$ and $Var\left(\hat{\beta}_{IV}|X, Z\right)$. **(4 points)**

d) Define a linear estimator of β to be any estimator of the form $\tilde{\beta} = AY$, for some $(k+1) \times n$ matrix A which depends only on X, Z. Show that, despite observing Z, the best linear unbiased estimator of β is still the OLS estimator from b). **(8 points)**

e) Is the model overidentified? Which are the included instruments? Does the rank condition hold using only the included instruments as the set of instruments? Explain. **(4 points)**

f) Explain the two stage construction of the two stage least squares estimator of β using the vector $z^* = (x', z')'$ as the set of instruments. How does it relate to the OLS estimator in this question? Explain why your answer to part d) may be unsurprising given what you know about asymptotically optimal GMM estimators under conditional homoskedasticity. **(5 points)**

g) Find the asymptotic distribution of the IV and OLS estimators from b) and c) separately. Which has a larger asymptotic variance? Explain. **(6 points)**

h) How does your answer to g) change if you drop conditional homoskedasticity? Explain. <u>Hint</u>: Let x be a scalar random variable and let $z = x^2$. Choose $E(u^2|x)$ wisely. **(5 points)**