ECMA31000: Introduction to Empirical Analysis
Solutions to Exam 2 2020

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

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

ANS: By Chebyshev's inequality, for any $\epsilon > 0$:

$$P(|X_n| > \epsilon) \leq \frac{E(|X_n|)}{\epsilon} = \frac{1}{n\epsilon} \to 0,$$

so $X_n \xrightarrow{p} 0$.

b) Is it true that $E\left(|X_n|^{\frac{3}{2}}\right) \to 0$? Prove it or find a counterexample. (5 points)

ANS: This is false. Let $\delta > 0$ and let

$$X_n = \begin{cases} n^{1/\delta} & \text{with probability } \frac{1}{n^{1+1/\delta}} \\ 0 & \text{with probability } 1 - \frac{1}{n^{1+1/\delta}}. \end{cases}$$

Then $E(|X_n|) = \frac{1}{n} \to 0$, but $E\left(|X_n|^{1+\delta}\right) = 1$ for all n. Take $\delta = \frac{1}{2}$.

c) Is it true that $X_n \xrightarrow{a.s.} 0$? Prove it or find a counterexample. (10 points)

ANS: This is false. Let X_n be independent random variables such that

$$X_n = \begin{cases} 1 & \text{with probability } \frac{1}{n} \\ 0 & \text{with probability } 1 - \frac{1}{n}. \end{cases}$$

Then for $0 < \epsilon < 1$:

$$
\begin{aligned}
&P\left(\cup_{k\geq n} \{|X_k| > \epsilon\}\right) \\
=&1 - P\left(\cap_{k\geq n} \{|X_k| \leq \epsilon\}\right) \\
=&1 - \Pi_{k=n}^{\infty} P(X_k = 0) \\
=&1 - \Pi_{k=n}^{\infty} \left(\frac{k-1}{k}\right) \\
=&1 - \lim_{N\to\infty} \Pi_{k=n}^{N} \left(\frac{k-1}{k}\right) \\
=&1 - \lim_{N\to\infty} \left(\frac{n-1}{N}\right) = 1
\end{aligned}
$$

for all n. So we do not get convergence $a.s.$

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: $X_i \sim U[-\theta, \theta]$, 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)

ANS: First note that $E(X) = 0$, so this does not produce a valid MoM estimator. We compute the second moment instead:

$$E\left(X^2\right) = \int_{-\theta}^{\theta} \frac{x^2}{2\theta} dx = \frac{\theta^2}{3}.$$

The sample analog principle suggests we set

$$\frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{\hat{\theta}_{MOM}^2}{3} \implies \hat{\theta}_{MOM} = \sqrt{\frac{3}{n} \sum_{i=1}^n X_i^2}.$$

Since $\{X_i^2\}_{i \geq 1}$ is an iid sequence with finite mean, the strong law of large numbers implies

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \overset{a.s.}{\to} E\left(X^2\right) = \frac{\theta^2}{3}.$$

The continuous mapping theorem with $g(x) = \sqrt{3x}$ now implies that

$$\sqrt{\frac{3}{n} \sum_{i=1}^n X_i^2} \overset{a.s.}{\to} \sqrt{3E(X^2)} = \theta.$$

The estimator is biased downward, because by Jensen's inequality:

$$E\left(\sqrt{\frac{3}{n} \sum_{i=1}^n X_i^2}\right)^2 \leq E\left(\frac{3}{n} \sum_{i=1}^n X_i^2\right) = \theta^2,$$

where equality is attained iff the X_i are constant, which is not the case here.

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)

ANS: Our goal is to apply the CLT to a sample average and then get at the square root via the delta method. Note that

$$\sqrt{n}\left(\hat{\theta}_{MOM}^2 - \theta^2\right) = 3\sqrt{n}\left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{\theta^2}{3}\right)$$
$$\overset{d}{\to} \mathcal{N}\left(0, 9 \cdot Var\left(X_i^2\right)\right).$$

by the CLT and continuous mapping theorem with $g(x) = 3x$. Note that

$$
\begin{aligned}
Var\left(X_i^2\right) &= \mathrm{E}\left(X_i^4\right) - \mathrm{E}\left(X_i^2\right)^2 \\
&= \frac{\theta^4}{5} - \frac{\theta^4}{9} \\
&= \frac{4\theta^4}{45},
\end{aligned}
$$

so

$$
\sqrt{n}\left(\hat{\theta}_{MOM}^2 - \theta^2\right) \overset{d}{\to} \mathcal{N}\left(0, \frac{4\theta^4}{5}\right).
$$

Now we apply the delta method, with $h(x) = x^{1/2}$. Since

$$
h'(x) = \frac{1}{2\sqrt{x}};
$$

$$
\left(h'\left(\theta^2\right)\right)^2 = \frac{1}{4\theta^2};
$$

we conclude that

$$
\begin{aligned}
\sqrt{n}\left(h\left(\hat{\theta}_{MOM}^2\right) - h\left(\theta^2\right)\right) &= \sqrt{n}\left(\hat{\theta}_{MOM} - \theta\right) \\
&\overset{d}{\to} \mathcal{N}\left(0, \left(h'\left(\theta^2\right)\right)^2 \frac{4\theta^4}{5}\right) \\
&\overset{d}{=} \mathcal{N}\left(0, \frac{\theta^2}{5}\right).
\end{aligned}
$$

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)

ANS: Since $\hat{\theta}_{MOM}$ is consistent, by Slutsky's Theorem,

$$
\frac{\sqrt{n}\left(\hat{\theta}_{MOM} - \theta\right)}{\sqrt{\hat{\theta}_{MOM}^2/5}} \overset{d}{\to} \mathcal{N}(0, 1).
$$

It follows that for any θ,

$$
\mathrm{P}_\theta\left(-z_{1-\alpha/2} < \frac{\sqrt{n}\left(\hat{\theta}_{MOM} - \theta\right)}{\sqrt{\hat{\theta}_{MOM}^2/5}} < z_{1-\alpha/2}\right) \to 1 - \alpha.
$$

It follows that $\mathrm{P}\left(\theta \in C_n\right) \to 1 - \alpha$, where

$$
C_n = \left[\hat{\theta}_{MOM} - z_{1-\alpha/2}\sqrt{\frac{\hat{\theta}_{MOM}^2}{5n}}, \hat{\theta}_{MOM} + z_{1-\alpha/2}\sqrt{\frac{\hat{\theta}_{MOM}^2}{5n}}\right]
$$

is the confidence set. If the null hypothesis is true, then $P(1 \in C_n) \to 1 - \alpha$. So rejecting the null hypothesis iff $1 \notin C_n$ yields a test with null rejection probability converging to α.

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)

ANS: Note that the likelihood given an iid sample $X^n = \{X_i\}_{i=1}^n$ is:

$$L_\theta(X^n) = \Pi_{i=1}^n \left[\frac{1}{2\theta} \mathbf{1}(-\theta \leq X_i \leq \theta) \right]$$

$$= \frac{1}{(2\theta)^n} \mathbf{1}(|X_i| \leq \theta \text{ for all } i).$$

This function is decreasing in θ, but is 0 if any $|X_i|$ is larger than θ, so the best that can be achieved is

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

Note that the distribution of $|X_i|$ is

$$P(|X_i| \leq x) = P(-x \leq X_i \leq x)$$

$$= \begin{cases} 0 & x < 0 \\ \frac{x}{\theta} & 0 \leq x \leq \theta \\ 1 & x > \theta \end{cases},$$

so $|X_i| \sim U[0, \theta]$. Now since $\hat{\theta}_{ML} \leq \theta$ with probability 1, consistency can be established by showing that for any $\epsilon > 0$:

$$\lim_{n \to \infty} P\left(\theta - \max_{i \leq n} |X_i| > \epsilon \right) = 0.$$

To see this, note that

$$P\left(\theta - \max_{i \leq n} |X_i| > \epsilon \right) = P\left(\max_{i \leq n} |X_i| < \theta - \epsilon \right)$$

$$= P\left(\bigcap_{i=1}^n \{|X_i| < \theta - \epsilon\} \right)$$

$$= \Pi_{i=1}^n P(|X_i| < \theta - \epsilon)$$

$$\leq \left| \frac{\theta - \epsilon}{\theta} \right|^n \to 0.$$

4

e) Show that $n \left(\theta - \hat{\theta}_{ML} \right) \overset{d}{\to} X$ 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? Explain. It may help to note that $\lim_{n \to \infty} \left(1 - \frac{k}{n} \right)^n = \exp \left(-k \right)$. (5 points)

ANS: Note that for $x \geq 0$:

$$\begin{aligned} \mathrm{P} \left(n \left[\theta - \max_{i \leq n} |X_i| \right] \leq x \right) &= \mathrm{P} \left(\theta - \frac{x}{n} \leq \max_{i \leq n} |X_i| \right) \\ &= 1 - \mathrm{P} \left(\max_{i \leq n} |X_i| < \theta - \frac{x}{n} \right) \\ &= 1 - \left(1 - \frac{x}{n\theta} \right)^n \\ &\to 1 - \exp \left(-\frac{x}{\theta} \right). \end{aligned}$$

For $x < 0$, the probability equals 0 for all n. We cannot use the CLT in this derivation because we are studying the distribution of the maximum of iid random variables, not a sum of iid random variables.

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

ANS: Studying the form of the asymptotic distribution reveals that for any x^*

$$\mathrm{P} \left(n \left[\theta - \hat{\theta}_{ML} \right] \leq x^* \right) \to 1 - \exp \left(-\frac{x^*}{\theta} \right).$$

The RHS equals $1 - \alpha$ iff $\exp \left(-\frac{x^*}{\theta} \right) = \alpha$, which implies

$$x^* = -\theta \ln \left(\alpha \right).$$

Therefore,
$$\mathrm{P} \left(n \left[\theta - \hat{\theta}_{ML} \right] \leq -\theta \ln \left(\alpha \right) \right) \to 1 - \alpha.$$

Rearranging yields
$$\mathrm{P} \left(\theta \leq \hat{\theta}_{ML} \left(1 + \frac{\ln \left(\alpha \right)}{n} \right)^{-1} \right) \to 1 - \alpha.$$

We construct the confidence set
$$C_n = \left[\hat{\theta}_{ML}, \hat{\theta}_{ML} \left(1 + \frac{\ln \left(\alpha \right)}{n} \right)^{-1} \right]$$

To check that this indeed yields a confidence set with the desired asymptotic coverage probability, observe that $\theta \geq \hat{\theta}_{ML}$ holds automatically because we cannot draw a value of X with larger absolute

value than θ.

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 $\mathrm{E}\,(x_i^* u_i) = \mathrm{E}\,(u_i) = \mathrm{E}\,(v_i) = 0$. 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 $\mathrm{E}\,(f(a) \cdot g(b)) = \mathrm{E}\,(f(a)) \cdot \mathrm{E}\,(g(b))$.

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

ANS: Consider the model $y = x'\beta + u$, where $(y, x', u)'$ is a random vector with $y, u \in \mathbb{R}$ and $x \in \mathbb{R}^{k+1}$, and only y, x are observed. Endogeneity occurs when $\mathrm{E}\,(ux_j) \neq 0$ for some $1 \leq j \leq k+1$.

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)

ANS: The model estimated is

$$y_i = \beta x_i + q_i,$$

where $q_i = u_i - \beta x_i^* v_i$. x_i is endogenous because

$$
\begin{aligned}
\mathrm{E}\,(x_i q_i) &= \mathrm{E}\,(x_i^* (1 + v_i) u_i) - \beta \mathrm{E}\,(x_i^* (1 + v_i) x_i^* v_i) \\
&= -\beta \left(\mathrm{E}\,\left([x_i^*]^2 v_i \right) + \mathrm{E}\,\left([x_i^*]^2 v_i^2 \right) \right) \\
&= -\beta \mathrm{E}\,\left([x_i^*]^2 \right) \mathrm{E}\,\left(v_i^2 \right) \neq 0.
\end{aligned}
$$

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)

ANS: The OLS estimator is given by

$$\hat{\beta}_{OLS} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \beta + \frac{\frac{1}{n} \sum_{i=1}^n x_i q_i}{\frac{1}{n} \sum_{i=1}^n x_i^2}.$$

Now note that

$$\frac{1}{n} \sum_{i=1}^n x_i^2 \xrightarrow{a.s.} \mathrm{E}\,\left(x_i^2 \right) = \mathrm{E}\,\left([x_i^*]^2 \right) \left(1 + \mathrm{E}\,\left(v_i^2 \right) \right)$$

6

$$\frac{1}{n} \sum_{i=1}^{n} x_i q_i \overset{a.s.}{\to} \mathrm{E}\left(x_i q_i\right) = -\beta \mathrm{E}\left(\left[x_i^*\right]^2\right) \mathrm{E}\left(v_i^2\right).$$

Therefore, by the continuous mapping theorem,

$$\hat{\beta}_{OLS} \overset{a.s.}{\to} \beta \left(1 - \frac{\mathrm{E}\left(\left[x_i^*\right]^2\right) \mathrm{E}\left(v_i^2\right)}{\mathrm{E}\left(\left[x_i^*\right]^2\right)\left(1 + \mathrm{E}\left(v_i^2\right)\right)}\right) = \beta \left(1 - \frac{\mathrm{E}\left(v_i^2\right)}{1 + \mathrm{E}\left(v_i^2\right)}\right).$$

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)

ANS: The probability limit of the OLS estimator is smaller than β in magnitude, because

$$1 - \frac{\mathrm{E}\left(v_i^2\right)}{1 + \mathrm{E}\left(v_i^2\right)} \in (0, 1),$$

so

$$\left|\hat{\beta}_{OLS}\right| \overset{a.s.}{\to} |\beta| \left(1 - \frac{\mathrm{E}\left(v_i^2\right)}{1 + \mathrm{E}\left(v_i^2\right)}\right) < |\beta|.$$

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

ANS: The IV estimator is given by

$$\hat{\beta}_{IV} = \frac{\sum_{i=1}^{n} z_i y_i}{\sum_{i=1}^{n} z_i x_i}.$$

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)

ANS: We need relevance and validity. That is $\mathrm{E}\left(z_i q_i\right) = 0$ must hold. But this is trivially true because $\mathrm{E}\left(z_i q_i\right) = \mathrm{E}\left(q_i\right) = 0$. Relevance holds if $\mathrm{E}\left(z_i x_i\right) = \mathrm{E}\left(x_i\right) = \mathrm{E}\left(x_i^*\right) \neq 0$. Under these conditions,

$$\hat{\beta}_{IV} = \frac{\sum_{i=1}^{n} z_i y_i}{\sum_{i=1}^{n} z_i x_i} = \beta + \frac{\sum_{i=1}^{n} z_i q_i}{\sum_{i=1}^{n} z_i x_i} \overset{a.s.}{\to} \beta + \frac{\mathrm{E}\left(q_i\right)}{\mathrm{E}\left(x_i^*\right)} = \beta.$$

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 $\mathrm{E}\left(u|x, z\right) = 0$ and $Var\left(u|x, z\right) = \sigma^2$. Assume $\mathrm{E}\left(xx'\right), \mathrm{E}\left(zx'\right), \mathrm{E}\left(zz'\right)$ exist and are invertible, and assume $\mathrm{E}\left(u^2 xx'\right)$ and $\mathrm{E}\left(u^2 zz'\right)$ exist.

a) Write the model in the form

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

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 $\mathrm{E}\left(U|X,Z\right) = 0$ and $Var\left(U|X,Z\right) = \mathrm{E}\left(UU'|X,Z\right) = \sigma^2 I_n$ under the stated assumptions. (4 points)

ANS: We have

$$
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} - & x_1' & - \\ - & x_2' & - \\ - & \vdots & - \\ - & x_n' & - \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} ; Z = \begin{pmatrix} - & z_1' & - \\ - & z_2' & - \\ - & \vdots & - \\ - & z_n' & - \end{pmatrix},
$$
$$\underbrace{}_{Y} \quad \underbrace{}_{X} \quad \underbrace{}_{\beta} \quad \underbrace{}_{U}$$

where for each $i = 1, \ldots, n$, $x_i = (x_{0i}, \ldots, x_{ki})$ are the observed covariates for individual i, and $z_i = (z_{0i}, \ldots, z_{ki})$ are the observed instruments for individual i. $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ is a constant vector of unknown parameters. For each element u_i of U, note that

$$
\mathrm{E}\left(u_i|X,Z\right) = \mathrm{E}\left(u_i|x_i,z_i\right)
$$
$$
= 0,
$$

where the first equality holds because the observations are iid, so (u_i, x_i, z_i) is independent of $\cup_{j \neq i} \{u_j, x_j, z_j\}$, so, in particular, the conditional expectation of u_i does not depend on these observations. Furthermore,

$$
Var\left(U|X,Z\right) := \mathrm{E}\left([U - \mathrm{E}\left(U|X,Z\right)][U - \mathrm{E}\left(U|X,Z\right)]' |X,Z\right)
$$
$$
= \mathrm{E}\left(UU'|X,Z\right).
$$

Now for $i = j$, we have

$$
\mathrm{E}\left(u_i^2|X,Z\right) = \mathrm{E}\left(u_i^2|x_i,z_i\right)
$$
$$
= Var\left(u|x,z\right)
$$
$$
= \sigma^2,
$$

while if $i \neq j$, we have

$$
\mathrm{E}\left(u_i u_j|X,Z\right) = \mathrm{E}\left(u_i u_j|x_i,z_i,x_j,z_j\right)
$$
$$
= \mathrm{E}\left(u_j \mathrm{E}\left(u_i|x_i,z_i,u_j,x_j,z_j\right)|x_i,z_i,x_j,z_j\right)
$$
$$
= \mathrm{E}\left(u_j \mathrm{E}\left(u_i|x_i,z_i\right)|x_i,z_i,x_j,z_j\right) = 0.
$$

It follows that only the diagonal elements of $\mathrm{E}\left(UU'|X,Z\right)$ are equal to σ^2 while the off diagonal elements are 0, yielding
$$
\mathrm{E}\left(UU'|X,Z\right) = \sigma^2 I_n.
$$

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

ANS: The OLS estimator is

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y$$
$$= \beta + (X'X)^{-1} X'U.$$

The expectation of $\hat{\beta}_{OLS}$ conditional on X,Z is

$$\mathrm{E}\left(\hat{\beta}_{OLS}|X,Z\right) = \beta + (X'X)^{-1} X'\mathrm{E}\left(U|X,Z\right) = \beta.$$

The variance of $\hat{\beta}_{OLS}$ conditional on X,Z is

$$Var\left(\hat{\beta}_{OLS}|X,Z\right) := \mathrm{E}\left[\left(\hat{\beta}_{OLS} - \mathrm{E}\left(\hat{\beta}_{OLS}|X,Z\right)\right)\left(\hat{\beta}_{OLS} - \mathrm{E}\left(\hat{\beta}_{OLS}|X,Z\right)\right)'|X,Z\right]$$
$$= \mathrm{E}\left((X'X)^{-1} X'UU'X (X'X)^{-1}|X,Z\right)$$
$$= (X'X)^{-1} X\mathrm{E}\left(UU'|X,Z\right) X (X'X)^{-1}$$
$$= \sigma^2 (X'X)^{-1} X'X (X'X)^{-1}$$
$$= \sigma^2 (X'X)^{-1}.$$

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 $\mathrm{E}\left(\hat{\beta}_{IV}|X,Z\right)$ and $Var\left(\hat{\beta}_{IV}|X,Z\right)$. (4 points)

ANS: The IV estimator is

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'Y$$
$$= \beta + (Z'X)^{-1} Z'U.$$

The expectation of $\hat{\beta}_{IV}$ conditional on X,Z is

$$\mathrm{E}\left(\hat{\beta}_{IV}|X,Z\right) = \beta + (Z'X)^{-1} Z'\mathrm{E}\left(U|X,Z\right) = \beta.$$

The variance of $\hat{\beta}_{IV}$ conditional on X,Z is

$$Var\left(\hat{\beta}_{IV}|X,Z\right) := \mathrm{E}\left[\left(\hat{\beta}_{IV} - \mathrm{E}\left(\hat{\beta}_{IV}|X,Z\right)\right)\left(\hat{\beta}_{IV} - \mathrm{E}\left(\hat{\beta}_{IV}|X,Z\right)\right)'|X,Z\right]$$
$$= \mathrm{E}\left((Z'X)^{-1} Z'UU'Z (X'Z)^{-1}|X,Z\right)$$
$$= (Z'X)^{-1} Z'\mathrm{E}\left(UU'|X,Z\right) Z (X'Z)^{-1}$$
$$= \sigma^2 (Z'X)^{-1} Z'Z (X'Z)^{-1}.$$

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)

ANS: A linear estimator of β is of the form $\tilde{\beta} = AY$, where A is a $(k+1) \times n$ matrix depending only on X, Z, and satisfies

$$\mathrm{E}\left(AY|X, Z\right) = A\mathrm{E}\left(Y|X, Z\right) = AX\beta + A\mathrm{E}\left(U|X, Z\right) = AX\beta.$$

To be unbiased, it must be the case that $AX\beta = \beta$ <u>for any β</u>, so we must have $AX = I_k$. Now note that

$$
\begin{aligned}
Var\left(\tilde{\beta}|X, Z\right) &= AVar\left(Y|X, Z\right)A' \\
&= AVar\left(U|X, Z\right)A' \\
&= \sigma^2 AA'.
\end{aligned}
$$

Let $C = A - \left(X'X\right)^{-1} X'$, and notice that $CX = AX - I_k = 0$, so

$$
\begin{aligned}
Var\left(\tilde{\beta}|X, Z\right) - Var\left(\hat{\beta}|X, Z\right) &= \sigma^2 \left[AA' - \left(X'X\right)^{-1}\right] \\
&= \sigma^2 \left[\left(C + \left(X'X\right)^{-1} X'\right)\left(C + \left(X'X\right)^{-1} X'\right)' - \left(X'X\right)^{-1}\right] \\
&= \sigma^2 CC',
\end{aligned}
$$

which is positive semidefinite, so $Var\left(\hat{\beta}_{OLS}|X, Z\right) \leq Var\left(\tilde{\beta}|X, Z\right)$ for any other linear unbiased estimator $\tilde{\beta}$ of β.

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)

ANS: Yes, the model is overidentified, because we have at least $2k + 2$ instruments (from x, z, though any function of them is also a valid instrument) and only $k + 1$ unknown parameters. The random vector x contains the included instruments since they are present in the model we wish to estimate. The rank condition holds because $\mathrm{E}\left(xx'\right)$ has rank $k + 1$ by assumption.

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)

ANS: In the first stage, we regress all the covariates on all of the included and excluded instruments. In matrix notation, we run OLS on

$$X = X\Pi + Z\delta + V.$$

We obtain a perfect fit simply by setting $\hat{\Pi} = I_k$ and $\hat{\delta} = 0$, so the first stage returns fitted values

of X which are simply X:

$$\hat{X} = X\hat{\Pi} + Z\hat{\delta} = X.$$

In the second stage, we regress Y on $\hat{X}$ by OLS, and the 2SLS estimator is then

$$\hat{\beta}_{2SLS} = \left(\hat{X}'\hat{X}\right)^{-1} \hat{X}'Y = (X'X)^{-1} X'Y = \hat{\beta}_{OLS}.$$

Under conditional homoskedasticity, the optimal GMM estimator is the 2SLS estimator, which is the OLS estimator in this case. Therefore, the OLS estimator attains the lowest asymptotic variance among GMM estimators despite not incorporating additional information from Z, so we can effectively ignore all the moment conditions other than $\mathrm{E}(ux) = 0$ in terms of constructing an asymptotically optimal estimate. In this sense, it is unsurprising that in a finite sample, the OLS estimator attains the lowest variance among all linear estimators incorporating information from X, Z, since we are essentially choosing the estimate to solve the sample moments

$$X'Y = X'X\hat{\beta}.$$

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)

ANS: First note that the sequences $\{x_i x_i'\}_{i\geq 1}$, $\{z_i x_i'\}_{i\geq 1}$, $\{z_i z_i'\}_{i\geq 1}$ are iid with finite first moment. The sequence $\{z_i u_i\}_{i\geq 1}$ is iid with finite second moments. Rewriting using summations yields

$$\sqrt{n}\left(\hat{\beta}_{OLS} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i$$
$$\xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}(xx')^{-1} \mathrm{E}\left(u^2 xx'\right) \mathrm{E}(xx')^{-1}\right),$$
$$\sqrt{n}\left(\hat{\beta}_{IV} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i$$
$$\xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}(zx')^{-1} \mathrm{E}\left(u^2 zz'\right) \mathrm{E}(xz')^{-1}\right),$$

using the CLT, SLLN and Slutsky's Theorem. Using the fact that $Var(u|x,z) = \mathrm{E}(u^2|x,z) = \sigma^2$ yields, by the law of iterated expectation:

$$\mathrm{E}(xx')^{-1} \mathrm{E}\left(u^2 xx'\right) \mathrm{E}(xx')^{-1} = \sigma^2 \mathrm{E}(xx')^{-1},$$
$$\mathrm{E}(zx')^{-1} \mathrm{E}\left(u^2 zz'\right) \mathrm{E}(xz')^{-1} = \sigma^2 \mathrm{E}(zx')^{-1} \mathrm{E}(zz') \mathrm{E}(xz')^{-1}.$$

One can also derive the asymptotic efficiency of OLS, the asymptotic analog of the finite sample result, by noting that if the entire vector of instruments is (x, z), under conditional homoskedasticity, the asymptotically optimal estimator of $\hat{\beta}$ is the 2SLS estimator. In the first stage, X is regressed on X, Z, which just returns the fitted values $\hat{X} = X$ and results in OLS. A more direct argument

relies on Gauss-Markov, and notes that for any $c \neq 0$, and any matrices X, Z,

$$c' \left[\left(\frac{X'X}{n} \right)^{-1} - \left(\frac{Z'X}{n} \right)^{-1} \frac{Z'Z}{n} \left(\frac{X'Z}{n} \right)^{-1} \right] c \leq 0.$$

By the continuous mapping theorem and the strong law of large numbers, we have that almost surely,

$$\lim_{n \to \infty} c' \left[\left(\frac{X'X}{n} \right)^{-1} - \left(\frac{Z'X}{n} \right)^{-1} \frac{Z'Z}{n} \left(\frac{X'Z}{n} \right)^{-1} \right] c \leq 0,$$

since inequalities preserve limits. However, the left hand side is almost surely equal to

$$c' \left[\mathrm{E}\left(xx' \right) - \mathrm{E}\left(zx' \right) \mathrm{E}\left(zz' \right) \mathrm{E}\left(xz' \right) \right] c.$$

This shows that the asymptotic variance of OLS dominates that of the IV estimator.

h) How does your answer to g) change if you drop conditional homoskedasticity? Explain. (5 points)

<u>Hint</u>: Let x be a scalar random variable and let $z = x^2$. Choose $\mathrm{E}\left(u^2 | x \right)$ wisely.

ANS: If we drop conditional homoskedasticity, the IV estimator can actually be more efficient than the OLS estimator. To see this, consider Question 6 of problem set 7. The limiting joint distribution from question 5c is

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_{OLS} - \beta \\ \hat{\beta}_{IV} - \beta \end{pmatrix} \xrightarrow{d} \mathcal{N} \left(0, \begin{pmatrix} \frac{\mathrm{E}\left(u^2 x^2 \right)}{\mathrm{E}(x^2)^2} & \frac{\mathrm{E}\left(u^2 xz \right)}{\mathrm{E}(x^2)\mathrm{E}(zx)} \\ \frac{\mathrm{E}\left(u^2 xz \right)}{\mathrm{E}(zx)\mathrm{E}(x^2)} & \frac{\mathrm{E}\left(u^2 z^2 \right)}{\mathrm{E}(zx)^2} \end{pmatrix} \right).$$

Let $z = x^2$. If $\mathrm{E}\left(u^2 | x \right) = \frac{1}{x^2}$, then the IV estimator is more efficient iff

$$\frac{\mathrm{E}\left(x^2 \right)}{\mathrm{E}\left(x^3 \right)^2} \leq \frac{1}{\mathrm{E}\left(x^2 \right)^2},$$

which is true by Jensen's inequality with $g\left(x \right) = x^{3/2}$.