ECMA 31100: Solutions to Problem Set 2

Due Feb 23 by 11:59PM

Question 1 Consider the potential outcomes framework

$$Y = Dy_1 + (1 - D)\, y_0$$

and suppose that $(y_0, y_1) \perp D|X$, where X is a vector of covariates. Define $ATE\,(x) := \mathrm{E}\,(y_1 - y_0|X = x)$. You are given an iid sample of $\{Y_i, D_i, X_i\}_{i=1}^{N}$. Assume that all relevant moments exist. Assume $\mu_d\,(x) := \mathrm{E}\,(y_d|X = x) = \alpha_d + x'\beta_d$ for $d = 0, 1$.

a) Show that running a regression of Y on X and a constant for observations with $D = 1$ produces unbiased estimates of α_1, β_1. Argue likewise for $D = 0$ and α_0, β_0.

ANS: Consider running the regression

$$Y_i D = \alpha_1 D_i + \beta_1 D_i X_i + U.$$

For a particular choice of a_1, b_1, the sum of squared residuals is

$$\sum_{i=1}^{n} D_i\,(Y_i - a_1 - b_1 X_i)^2\,,$$

which is the SSR we would obtain if running the regression only for individuals with $D = 1$. Note that

$$\mathrm{E}\,(YD|D, X) = D\mathrm{E}\,(y_1|D, X) = \alpha_1 D + \beta_1 D \cdot X,$$

so $\mathrm{E}\,(U|D, X) = 0$ in the regression above, meaning the OLS estimates in this regression are unbiased estimates of α_1 and β_1.

b) Define $\hat{\mu}_d\,(x) := \hat{\alpha}_d + x'\hat{\beta}_d$ for $d = 0, 1$, where $\left\{\hat{\alpha}_d, \hat{\beta}_d\right\}_{d=0,1}$ denote the OLS estimates from part a). Construct an estimate of the ATE using

$$\widehat{ATE} = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_1\,(X_i) - \hat{\mu}_0\,(X_i)\,.$$

Show that

$$\widehat{ATE} = \bar{Y}_1 - \bar{Y}_0 - \left(\bar{X}_1 - \bar{X}_0\right)'\left(\frac{N_0}{N}\hat{\beta}_1 + \frac{N_1}{N}\hat{\beta}_0\right),$$

where N_d is the number of observations for which $D = d$, $\bar{Y}_d = \frac{1}{N_d}\sum_{i=1}^{N} Y_i \mathbf{1}\,(D_i = d)$ and $\bar{X}_d =$

$\frac{1}{N_d}\sum_{i=1}^{N} X_i \mathbf{1}(D_i = d)$. How does this estimate relate to your estimate of the ATE from problem 2 of problem set 1?

ANS: We know that

$$\widehat{ATE} = (\hat{\alpha}_1 - \hat{\alpha}_0) + \bar{X}_n' \left(\hat{\beta}_1 - \hat{\beta}_0\right)$$

which is exactly the estimate from PSET 1. Using the formulas for the intercept estimates gives

$$\begin{aligned}
\widehat{ATE} &= (\hat{\alpha}_1 - \hat{\alpha}_0) + \bar{X}_n' \left(\hat{\beta}_1 - \hat{\beta}_0\right) \\
&= \bar{Y}_1 - \bar{Y}_0 - \left(\bar{X}_1\hat{\beta}_1 - \bar{X}_0\hat{\beta}_0\right) + \bar{X}_n' \left(\hat{\beta}_1 - \hat{\beta}_0\right). \\
&= \bar{Y}_1 - \bar{Y}_0 - \frac{N_0}{N}\left(\bar{X}_1 - \bar{X}_0\right)' \hat{\beta}_1 - \frac{N_1}{N}\left(\bar{X}_1 - \bar{X}_0\right)' \hat{\beta}_0 \\
&= \bar{Y}_1 - \bar{Y}_0 - \left(\bar{X}_1 - \bar{X}_0\right)' \left(\frac{N_0}{N}\hat{\beta}_1 + \frac{N_1}{N}\hat{\beta}_0\right),
\end{aligned}$$

where the third equality follows because $N\bar{X}_n = N_1\bar{X}_1 + N_0\bar{X}_0$, so $\bar{X}_n - \bar{X}_1 = \frac{N_0}{N}\left(\bar{X}_0 - \bar{X}_1\right)$ and $\bar{X}_n - \bar{X}_0 = \frac{N_1}{N}\left(\bar{X}_1 - \bar{X}_0\right)$.

c) Estimate the standard deviation of $\widehat{ATE}$ using a bootstrap procedure:

- Draw N observations from your original dataset $\{Y_i, D_i, X_i\}_{i=1}^{N}$ at random and with replacement. Give each sample point equal probability of being drawn.
- Compute $\widehat{ATE}$ using your resampled data.
- Repeat 2,000 times to form a collection $\left\{\widehat{ATE}_j\right\}_{j=1}^{2000}$ of estimates.
- Compute the standard deviation of your estimates.

How does the bootstrap standard error compare to your consistent estimate of the asymptotic standard deviation from problem set 1?

ANS: It is highly likely that your bootstrap standard error is very similar to the asymptotic standard error you computed in PSET 1.

d) Under what conditions does $\widehat{ATE}$ converge in probability to the naive comparison $\mathrm{E}(Y|D=1) - \mathrm{E}(Y|D=0)$? Provide intuition for the validity of the naive comparison when D is independent of X. Is it necessarily the case that $(y_0, y_1) \perp D$?

ANS: The probability limit of $\widehat{ATE}$ is

$$\mathrm{E}(Y|D=1) - \mathrm{E}(Y|D=0) - [\mathrm{E}(X|D=1) - \mathrm{E}(X|D=0)]' (\mathrm{P}(D=0)\beta_1 + \mathrm{P}(D=1)\beta_0).$$

This equals the naive comparison if $\mathrm{E}(X|D=1) = \mathrm{E}(X|D=0)$, or if it happens to be the case that $\mathrm{P}(D=0)\beta_1 + \mathrm{P}(D=1)\beta_0 = 0$. The former condition says that X is mean independent of D, and we know that if X is fully independent of D including it in a linear regression with D does not change the coefficient on D. It turns out that $D \perp X$ and unconfoundedness imply $(y_0, y_1) \perp D$, so the naive comparison is in fact the ATE. To show this, for any events A, B, let $\mathbf{1}((y_0, y_1) \in A) \equiv \mathbf{1}_A$ and $\mathbf{1}(D \in B) = \mathbf{1}_B$. Then:

$$\begin{aligned}
\mathrm{P}((y_0, y_1) \in A, D \in B) &= \mathrm{E}(\mathbf{1}_A \mathbf{1}_B) \\
&= \mathrm{E}(\mathrm{E}(\mathbf{1}_A \mathbf{1}_B | X)) \\
&= \mathrm{E}(\mathrm{E}(\mathbf{1}_A \mathrm{E}(\mathbf{1}_B | \mathbf{1}_A, X) | X)) \\
(\text{unconfoundedness}) &= \mathrm{E}(\mathrm{E}(\mathbf{1}_A \mathrm{E}(\mathbf{1}_B | X) | X)) \\
&= \mathrm{E}(\mathrm{E}(\mathbf{1}_A | X) \mathrm{E}(\mathbf{1}_B | X)) \\
(D \perp X) &= \mathrm{E}(\mathrm{E}(\mathbf{1}_A | X) \mathrm{E}(\mathbf{1}_B)) \\
&= \mathrm{P}((y_0, y_1) \in A) \mathrm{P}(D \in B).
\end{aligned}$$

Question 2 Let the set of possible treatments be $\mathcal{D}$. For each $d \in \mathcal{D}$ let $y(d)$ be the potential outcome under treatment d.

a) Suppose that $y(d) = x(d, w)' \beta + u$ where x is a known function, w is a set of covariates and $\mathrm{E}(u|w) = 0$. Show that for all $d, d_j \in \mathcal{D}$:

1. $y(d) - y(d_j) = j(d, w)' \beta_j$ for some function j. (Constant treatment effect)

2. $\mathrm{E}(y(d) | w) = x(d, w)' \beta$. (Correctly specified conditional mean)

ANS: We have

$$y(d) - y(d_j) = [x(d, w) - x(d_j, w)]' \beta,$$

so the representation in 1. holds with $j(d, w) = x(d, w) - x(d_j, w)$ and $\beta_j = \beta$ for all j. Also

$$\mathrm{E}(y(d) | w) = x(d, w)' \beta + \mathrm{E}(u|w) = x(d, w)' \beta$$

since $\mathrm{E}(u|d, w) = 0$, so 2. holds also.

b) Now suppose 1. and 2. hold from part a). Show that there exists a u such that for every d:

- $y(d) = x(d, w)' \beta + u$

- $\mathrm{E}(u|w) = 0$.

ANS: 1. gives us that for any d and d_j:

$$\begin{aligned}
y(d) &= j(d, w)' \beta_j + y(d_j) \\
&= j(d, w)' \beta_j + \mathrm{E}(y(d_j) | w) + \underbrace{y(d_j) - \mathrm{E}(y(d_j) | w)}_{u},
\end{aligned} \tag{1}$$

where the newly constructed u satisfies $\mathrm{E}(u|w) = 0$. Taking conditional expectations and applying 2. gives

$$\begin{aligned}\mathrm{E}(y(d)|w) &= j(d,w)'\beta_j + \mathrm{E}(y(d_j)|w) \\ \text{(by 2.)} &= x(d,w)'\beta.\end{aligned}$$

Substituting into (1) gives:

$$y(d) = x(d,w)'\beta + u.$$

where $\mathrm{E}(u|w) = 0$.

c) Consider the simultaneous equations example from Week 2. Suppose the model is in fact given by

$$\begin{aligned}q_D &= \beta_0 + \beta_1 p + \beta_2 r + u; \quad \mathrm{E}(u) = \mathrm{E}(ru) = 0, \\ q_S &= \gamma_0 + \gamma_1 p + \gamma_2 z + v; \quad \mathrm{E}(v) = \mathrm{E}(vz) = 0.\end{aligned}$$

where z is an "exogenous supply shifter" and r is an "exogenous demand shifter". You observe a sample of $\{q_i, p_i, r_i, z_i\}_{i=1}^n$ from n markets, where q, p are observed prices and quantities that occur in equilibrium. Give conditions under which an IV strategy would allow you to identify β_1 and γ_1 and detail the strategy.

ANS: Let's work backwards to see what we need for identification. Since (p, q) pairs appear in equilibrium, solve for price:

$$\beta_0 + \beta_1 p + \beta_2 r + u = \gamma_0 + \gamma_1 p + \gamma_2 z + v$$

implies

$$\begin{aligned}p &= \frac{1}{\beta_1 - \gamma_1}(\gamma_0 - \beta_0 + \gamma_2 z - \beta_2 r + v - u) \\ &= \frac{\gamma_0 - \beta_0}{\beta_1 - \gamma_1} + \frac{\gamma_2}{\beta_1 - \gamma_1} z + \frac{-\beta_2}{\beta_1 - \gamma_1} r + \frac{v - u}{\beta_1 - \gamma_1}\end{aligned}$$

which exists provided $\beta_1 \neq \gamma_1$. We see that z can be used as an instrument for p in the demand equation if $\mathrm{E}(zu) = 0$. Similarly, r can be used as an instrument for p in the supply equation if $\mathrm{E}(ru) = 0$. An IV strategy could make these assumptions:

- Exogeneity: The exogenous supply shifter z should be independent of each of the potential outcomes $q_D(p')$ conditional on covariates r, and the exogenous demand shifter r should be independent of each of the potential outcomes $q_S(p')$ conditional on covariates z.

- Exclusion: $q_D(p, z') = q_D(p, z'')$ for any two distinct values of the instrument z, so z has no direct impact on $q_D(p)$. $q_S(p, r') = q_S(p, r'')$ for any two distinct values of the instrument r, so r has no direct impact on $q_S(p)$.

- $q_D(p') = \beta_0 + \beta_1 p' + \beta_2 r + u \equiv x_D(p', r)' \beta + u$ for all prices p', and $q_S(p') = \gamma_0 + \gamma_1 p' + \gamma_2 r + u \equiv x_S(p', r)' \gamma + v$

- $\mathrm{E}(u|r) = 0$ and $\mathrm{E}(v|z) = 0$.

Since $x(p', r)' \beta$ is constant after conditioning on r, exogeneity implies that u is independent of z conditional on r. Hence,

$$\mathrm{E}(u|z, r) = \mathrm{E}(u|r) = 0.$$

In particular, $\mathrm{E}(zu) = 0$. Similarly, v is independent of r conditional on z, so $\mathrm{E}(v|r, z) = 0$. Finally, we must make sure the IV is relevant. This means that in the first stage

$$p = \pi_0 + \pi_1 r + \pi_2 z + \epsilon; \quad \mathrm{E}(\epsilon) = \mathrm{E}(r\epsilon) = \mathrm{E}(z\epsilon) = 0,$$

we need $\pi_1 \neq 0$ to identify supply and $\pi_2 \neq 0$ to identify demand. At a minimum, this requires that z and r are not perfectly correlated and that the variances of z and r are not zero (to avoid collinearity with the intercept). By comparison with the formula for p, we see that $\pi_1 = \frac{-\beta_2}{\beta_1 - \gamma_1}$ and $\pi_2 = \frac{\gamma_2}{\beta_1 - \gamma_1}$ because by our IV assumptions, $\mathrm{E}(u|z, r) = \mathrm{E}(v|z, r) = 0$ which implies $\mathrm{E}([v - u]\, r) = \mathrm{E}([v - u]\, z) = 0$ holds. We can ensure this is non-zero by requiring that β_2 and γ_2 are non-zero, so that the conditional mean of $q_S(p')$ depends on the level of z and the conditional mean of $q_D(p')$ depends on r. Intuitively, this allows us to shift either the supply or demand curve while holding the other fixed and thereby identify the effect of price on demand and supply separately. All of these assumptions ensure that the standard IV identification argument will hold equation by equation.

Question 3 Consider the linear model

$$y = \beta_0 + \beta_1 x + u;$$

where x is an endogenous scalar regressor and z is an excluded instrument. Suppose that $\mathrm{E}(u|x, z) = 0$ and $\mathrm{E}(u^2|x, z) = \sigma^2$, and that $Cov(x, z) \neq 0$.

a) Show that if z is used as an instrument for x,

$$\sqrt{n}\left(\hat{\beta}_1^{IV} - \beta_1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{Corr(x, z)^2\, Var(x)}\right)$$

It might be easier to work out the joint asymptotic distribution of $\left(\hat{\beta}_0^{IV}, \hat{\beta}_1^{IV}\right)'$ and then figure out the $(2, 2)$ entry of the limit variance matrix.

ANS: Let $w = (1, x)'$ and $q = (1, z)'$ denote the vectors of regressors and instruments. Then

$$\begin{aligned} \sqrt{n}\left(\hat{\beta}^{IV} - \beta\right) &= \left(\frac{1}{n}\sum_{i=1}^{n} q_i w_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_i u_i \\ &\xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}(qw')^{-1}\, \mathrm{E}\left(u^2 qq'\right) \mathrm{E}(qw')^{-1}\right) \end{aligned}$$

where the variance matrix V can be written as

$$\begin{aligned}
V &= \mathrm{E}\begin{pmatrix}1 & x\\ z & zx\end{pmatrix}^{-1}\mathrm{E}\begin{pmatrix}u^2 & u^2 z\\ u^2 z & u^2 z^2\end{pmatrix}\mathrm{E}\begin{pmatrix}1 & x\\ z & zx\end{pmatrix}^{-1}\\
&= \sigma^2\mathrm{E}\begin{pmatrix}1 & x\\ z & zx\end{pmatrix}^{-1}\mathrm{E}\begin{pmatrix}1 & z\\ z & z^2\end{pmatrix}\mathrm{E}\begin{pmatrix}1 & x\\ z & zx\end{pmatrix}^{-1}\\
&= \frac{\sigma^2}{Cov\left(z,x\right)^2}\mathrm{E}\begin{pmatrix}zx & -x\\ -z & 1\end{pmatrix}\mathrm{E}\begin{pmatrix}1 & z\\ z & z^2\end{pmatrix}\mathrm{E}\begin{pmatrix}zx & -x\\ -z & 1\end{pmatrix}\\
&= \frac{\sigma^2}{Cov\left(z,x\right)^2}\mathrm{E}\begin{pmatrix}zx & -x\\ -z & 1\end{pmatrix}\begin{pmatrix}\mathrm{E}\left(zx\right)-\mathrm{E}\left(z\right)^2 & \mathrm{E}\left(z\right)-\mathrm{E}\left(x\right)\\ \mathrm{E}\left(z\right)\mathrm{E}\left(zx\right)-\mathrm{E}\left(z\right)\mathrm{E}\left(z^2\right) & \mathrm{E}\left(z^2\right)-\mathrm{E}\left(z\right)\mathrm{E}\left(x\right)\end{pmatrix}\\
&= \frac{\sigma^2}{Cov\left(z,x\right)^2}\mathrm{E}\begin{pmatrix}\cdot & \cdot\\ \cdot & \mathrm{E}\left(z\right)\mathrm{E}\left(x\right)-\mathrm{E}\left(z\right)^2+\mathrm{E}\left(z^2\right)-\mathrm{E}\left(z\right)\mathrm{E}\left(x\right)\end{pmatrix},
\end{aligned}$$

so the $(2,2)$ entry is

$$V_{2,2} = \frac{\sigma^2 Var\left(z\right)}{Cov\left(z,x\right)^2} = \frac{\sigma^2 Var\left(z\right)}{Corr\left(z,x\right)^2 Var\left(z\right)Var\left(x\right)} = \frac{\sigma^2}{Corr\left(x,z\right)^2 Var\left(x\right)}.$$

b) Now derive the limit distribution of $\hat{\beta}_1^{OLS}$ and compare the asymptotic variances of the IV and OLS estimates. What do you conclude?

ANS: We have

$$\begin{aligned}
\sqrt{n}\left(\hat{\beta}^{IV}-\beta\right) &= \left(\frac{1}{n}\sum_{i=1}^{n}w_i w_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w_i u_i\\
&\xrightarrow{d}\mathcal{N}\left(0,\mathrm{E}\left(ww'\right)^{-1}\mathrm{E}\left(u^2 ww'\right)\mathrm{E}\left(ww'\right)^{-1}\right)\\
&\overset{d}{=}\mathcal{N}\left(0,\sigma^2\mathrm{E}\left(ww'\right)^{-1}\right).
\end{aligned}$$

The $(2,2)$ entry of this matrix is

$$\frac{\sigma^2}{Var\left(x\right)} \leq \frac{\sigma^2}{Corr\left(x,z\right)^2 Var\left(x\right)},$$

since $-1 \leq Corr\left(x,z\right) \leq 1$, so the OLS estimator is more efficient than IV under homoskedasticity, unless $z = a + bx$ for some constants a, b, in which case they are equally efficient, but this would effectively be using x as its own instrument.

c) Is your conclusion in part b) robust to heteroskedasticity? Prove it or find a counterexample.

ANS: No, consider the following counterexample:

$$\begin{aligned}
x &\sim Unif\left[1,2\right]\\
y &= \sqrt{x}\cdot\epsilon
\end{aligned}$$

$$\epsilon \sim \mathcal{N}(0,1),$$

where ϵ is independent of x. We have

$$\mathrm{E}(y|x) = \sqrt{x} \cdot \mathrm{E}(\epsilon|x) = 0,$$

so

$$y = 0 + 0 \cdot x + u,$$

where $\mathrm{E}(u|x) = 0$. Note that $\mathrm{E}(u^2|x) = Var(u|x) = Var(y|x) = x$. Now consider using the instrument $z = \frac{1}{x}$. Notice that the order of the instruments listed in the instrument vector doesn't matter. (It shouldn't, we're using all the included and excluded instruments together to estimate all the parameters in the main equation). To see this, let Z be the $n \times 2$ matrix of observations of $(1, 1/x)$ and Z^* the $(n \times 2)$ matrix of observations of $(1/x, 1)$. Then $Z^* = ZT$ where

$$T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is a matrix that switches the columns of Z. (T is symmetric and its own inverse, since applying T twice leaves the columns where they started). The IV estimator using Z^* is

$$\begin{aligned}
\hat{\beta}^*_{IV} &= \left((Z^*)' X\right)^{-1} (Z^*)' Y \\
&= (T'Z'X)^{-1} TZ'Y \\
&= (Z'X)^{-1} (T')^{-1} TZ'Y \\
&= (Z'X)^{-1} T^{-1}TZ'Y \\
&= \hat{\beta}_{IV}.
\end{aligned}$$

The limit variance of the IV estimator is easier to derive if we switch the order of the instruments to $(z, 1)$ rather than $(1, z)$, (which we now know won't affect the result):

$$\begin{aligned}
V &= \mathrm{E}\begin{pmatrix} z & zx \\ 1 & x \end{pmatrix}^{-1} \mathrm{E}\begin{pmatrix} u^2 z^2 & u^2 z \\ u^2 z & u^2 \end{pmatrix} \left[\mathrm{E}\begin{pmatrix} z & zx \\ 1 & x \end{pmatrix}^{-1}\right]^T \\
&= \begin{pmatrix} \ln(2) & 1 \\ 1 & 3/2 \end{pmatrix}^{-1} \begin{pmatrix} \ln(2) & 1 \\ 1 & 3/2 \end{pmatrix} \begin{pmatrix} \ln(2) & 1 \\ 1 & 3/2 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \ln(2) & 1 \\ 1 & 3/2 \end{pmatrix}^{-1} \\
&= \frac{1}{\frac{3}{2}\ln(2) - 1} \begin{pmatrix} 3/2 & -1 \\ -1 & \ln(2) \end{pmatrix}
\end{aligned}$$

so $V_{22} = \ln(2) \cdot \left(\frac{3}{2}\ln(2) - 1\right)^{-1} \approx 17.5$.

How did I know this would be easier? The matrices cancel because we've used so called 'optimal instruments' to construct the example. If we know that $\mathrm{E}(u|x) = 0$ and $\mathrm{E}(u^2|x) = \sigma^2(x)$ (arbitrary form of heteroskedasticity), there are many method of moments estimators we could use to identify β_0 and β_1 since $\mathrm{E}(f(x)u) = 0$ for any x. We want the best two moments in terms of the limiting variance, and it turns out that the optimal instrument vector is $z^* = \frac{1}{\sigma^2(x)}(1, x)$, which is a scaled version of the original vector of regressors. I needed to choose $\sigma^2(x) = x$ so that the resulting instrument vector still contains a 1, yielding $z^* = (1/x, 1)$. Since the order of inclusion doesn't matter, I now have the instrument set $(1, 1/x)$ which will do better than OLS in terms of the limiting variance and still be an IV estimator using 1 as the included instrument and $1/x$ as an IV for x. This optimality result holds because the limiting variance of resulting the IV estimate is

$$V = \mathrm{E}\left(\frac{1}{\sigma^2(x)}\begin{pmatrix}1 & x\\ x & x^2\end{pmatrix}\right)^{-1},$$

which is the same as the limiting variance of the OLS estimate in the model

$$\frac{y}{\sigma(x)} = \beta_0\frac{1}{\sigma(x)} + \beta_1\frac{x}{\sigma(x)} + \frac{u}{\sigma(x)},$$

which satisfies MLR.1-5. We know that under conditional homoskedasticity, the optimal GMM estimator is 2SLS, which in this case is OLS because $\frac{x}{\sigma(x)}$ is an included instrument. We can verify our answer above using the formula for the limiting variance with optimal instruments:

$$V = \mathrm{E}\begin{pmatrix}\frac{1}{x} & 1\\ 1 & x\end{pmatrix}^{-1} = \frac{1}{\left(\frac{3}{2}\ln(2) - 1\right)}\begin{pmatrix}\frac{3}{2} & -1\\ -1 & \ln(2)\end{pmatrix}.$$

As for the OLS estimator:

$$V = \mathrm{E}\begin{pmatrix}1 & x\\ x & x^2\end{pmatrix}^{-1}\mathrm{E}\begin{pmatrix}u^2 & u^2x\\ u^2x & u^2x^2\end{pmatrix}\mathrm{E}\begin{pmatrix}1 & x\\ x & x^2\end{pmatrix}^{-1}.$$

Note that $\mathrm{E}(x) = \frac{3}{2}$, $\mathrm{E}(x^2) = \int_1^2 x^2\mathrm{d}x = \frac{7}{3}$, and

$$\mathrm{E}\left(u^2x^2\right) = \mathrm{E}\left(x^3\right) = \int_1^2 x^3\mathrm{d}x = \frac{15}{4}.$$

Therefore,

$$\begin{aligned}V &= \begin{pmatrix}1 & 3/2\\ 3/2 & 7/3\end{pmatrix}^{-1}\cdot\begin{pmatrix}3/2 & 7/3\\ 7/3 & 15/4\end{pmatrix}\cdot\begin{pmatrix}1 & 3/2\\ 3/2 & 7/3\end{pmatrix}^{-1}\\ &= (12)^2\begin{pmatrix}7/3 & -3/2\\ -3/2 & 1\end{pmatrix}\cdot\begin{pmatrix}3/2 & 7/3\\ 7/3 & 15/4\end{pmatrix}\cdot\begin{pmatrix}7/3 & -3/2\\ -3/2 & 1\end{pmatrix}\end{aligned}$$

$$
\begin{aligned}
&= (12)^2 \begin{pmatrix} 7/3 & -3/2 \\ -3/2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1/12 \\ \frac{49}{9} - \frac{45}{8} & \frac{1}{4} \end{pmatrix} \\
&= (12)^2 \begin{pmatrix} \cdot & \cdot \\ \cdot & 1/8 \end{pmatrix}
\end{aligned}
$$

where $V_{22} = 18 > \ln(2) \cdot \left(\frac{3}{2} \ln(2) - 1\right)^{-1}$.

Question 4 Consider the linear model

$$
y = \beta x + u;
$$

where x is an endogenous scalar regressor and z is an excluded instrument. Suppose that $\mathrm{E}(u|z) = 0$ and $\mathrm{E}(u^2|z) = \sigma_u^2$, and that $\mathrm{E}(xz) \neq 0$.

a) Show under weak instrument asymptotics that the OLS estimator is inconsistent and find its probability limit. That is, suppose that at sample size n the first stage is given by

$$
x_n = \pi_n z + v; \quad \mathrm{E}(zv) = 0, \; \pi_n = \frac{\pi}{\sqrt{n}}, \; \pi \in \mathbb{R}.
$$

Does the result depend on π? If not, why?

ANS: Under weak instrument asymptotics we have

$$
\begin{aligned}
y_n &= \beta x_n + u; \\
x_n &= \pi_n z + v.
\end{aligned}
$$

Hence:

$$
\begin{aligned}
\hat{\beta} = \frac{\frac{1}{n}\sum x_{i,n} y_{i,n}}{\frac{1}{n}\sum_{i=1}^n x_{i,n}^2} &= \beta + \frac{\frac{1}{n}\sum_{i=1}^n x_{i,n} u_i}{\frac{1}{n}\sum_{i=1}^n (\pi_n z_i + v_i)^2} \\
&= \beta + \frac{\pi_n \cdot \frac{1}{n}\sum_{i=1}^n z_i u_i + \frac{1}{n}\sum_{i=1}^n v_i u_i}{\pi_n^2 \cdot \frac{1}{n}\sum_{i=1}^n z_i^2 + 2 \cdot \pi_n \cdot \frac{1}{n}\sum_{i=1}^n z_i v_i + \frac{1}{n}\sum_{i=1}^n v_i^2} \\
&\xrightarrow{p} \beta + \frac{0 \cdot \mathrm{E}(zu) + \mathrm{E}(vu)}{0 \cdot \mathrm{E}(z^2) + 2 \cdot 0 \cdot 0 + \mathrm{E}(v^2)} \\
&= \beta + \frac{\mathrm{E}(vu)}{\mathrm{E}(v^2)},
\end{aligned}
$$

where the convergence in probability follows as a result of applying the LLN and continuous mapping theorem. This result does not depend on π because the $\frac{1}{n}\sum_{i=1}^n v_i^2$ term in the denominator dominates the behaviour of the denominator since $\pi_n \to 0$, and the same is true for $\frac{1}{n}\sum_{i=1}^n v_i u_i$ in the numerator. There's no need to rescale these terms because the error v gets squared, so

$\frac{1}{n}\sum_{i=1}^{n} v_i^2 \xrightarrow{p} \mathrm{E}\left(v^2\right) \neq 0$, whereas with the IV estimator, the denominator is

$$\frac{1}{n}\sum_{i=1}^{n} z_i x_{i,n} = \pi_n \cdot \frac{1}{n}\sum_{i=1}^{n} z_i^2 + \frac{1}{n}\sum_{i=1}^{n} z_i v_i \xrightarrow{p} 0.$$

b) Do we really think the first stage is changing with the sample size? How do you interpret this result?

ANS: We don't, we're letting π_n vary with n so that the asymptotic approximation to the behaviour of the IV estimator is more like the finite sample distribution of the IV estimator. Using this statistical tool we can approximate (with the weak instrument limit) this finite sample distribution. The figure in Stock, Wright and Yogo (2002) is produced by noting that in a finite sample with the given simulation parameters, we have (exactly):

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i v_i \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}\right)$$

To see this, note that because the pairs (u_i, v_i) are iid, the $2n \times 2$ vector formed by stacking them: $(u_1, v_1, u_2, v_2, \ldots, u_n, v_n)'$ is multivariate normal (since any linear combination of the elements of this vector is univariate normal), so adding iid normal random vectors and dividing by $\sqrt{n}$ will give the same distribution as (u, v). Since the z_i are nonstochastic, we have:

$$\hat{\beta}_{IV} - \beta \sim \frac{u}{\pi^2 \sum_{i=1}^{n} z_i^2 + v} = \frac{u}{n\pi^2 + v}; \qquad \begin{pmatrix} u \\ v \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}\right)$$

We don't get such a finite sample result in the case that (u, v) is non-normal, but the weak instrument limit would have a distribution like this after applying the usual normal approximation. The way we control the strength of the instrument in this simulation (since $z_i = 1$ is fixed and n is fixed) is by varying π. The concentration parameter in a finite sample is defined in Stock Wright and Yogo (2002) as: $\mu^2 = \pi^2 \left(\sum_{i=1}^{n} z_i^2\right)/\sigma_v^2$, which will diverge unless we use weak instrument asymptotics, in which case it will converge to $\pi^2 \mathrm{E}\left(z^2\right)/\sigma_v^2$. In our simluation, $\mu^2 = 500\pi^2$.

c) Find the limit distribution of the IV estimator once assuming $\pi_n = \pi \neq 0$ for all n, and again under weak instrument asymptotics, assuming homoskedasticity. Are your homoskedasticity assumptions covered by the information given in the question?

ANS: If $\pi_n = \pi \neq 0$ for all n, the usual asymptotics provide

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

where the homoskedasticity assumption $\mathrm{E}\left(u^2|z\right) = \sigma_u^2$ was used to simplify the variance. Under weak instrument asymptotics (where $\sqrt{n}\pi_n = \pi \in \mathbb{R}$), as in the slides:

$$\begin{aligned}\hat{\beta}_{IV} - \beta &= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i u_i}{\pi_n \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^n z_i^2 + \frac{1}{\sqrt{n}}\sum_{i=1}^n z_i v_i} \\ &\xrightarrow{d} \frac{A}{\pi \mathrm{E}\left(z^2\right) + B},\end{aligned}$$

where

$$\begin{pmatrix} A \\ B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{E}\left(u^2 z^2\right) & \mathrm{E}\left(z^2 uv\right) \\ \mathrm{E}\left(z^2 uv\right) & \mathrm{E}\left(z^2 v^2\right) \end{pmatrix}\right).$$

The homoskedasticity assumptions we used to simplify this joint distribution are $\mathrm{E}\left(u^2|z\right) = \sigma_u^2$, $\mathrm{E}\left(v^2|z\right) = \sigma_v^2$, $\mathrm{E}\left(uv|z\right) = \sigma_{uv}$; the latter two are not covered by the information in the question, but are not relevant for the standard asymptotic argument, because

$$\mathrm{E}\left(zx\right) = \pi \mathrm{E}\left(z^2\right) + \underbrace{\mathrm{E}\left(zv\right)}_{=0} = \pi \mathrm{E}\left(z^2\right).$$

The distribution of $\frac{1}{n}\sum_{i=1}^n z_i v_i$ vanishes in the asymptotic approximation, though $\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i v_i$ has a non-degenerate limiting distribution in the weak instrument asymptotics.

d) If you drop homoskedasticity, is the IV estimator still centered at the probability limit of the OLS estimator asymptotically when $\pi = 0$? Prove it or find a counterexample.

ANS: When $\pi = 0$, the limiting distribution was shown in the slides to be

$$\hat{\beta}_{IV} - \beta \xrightarrow{d} \frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} + \frac{U}{B}.$$

Under homoskedasticity this simplifies to

$$\hat{\beta}_{IV} - \beta \xrightarrow{d} \frac{\mathrm{E}\left(uv\right)}{\mathrm{E}\left(v^2\right)} + \frac{U}{B},$$

where σ_{uv}/σ_v^2 is the probability limit of $\hat{\beta}_{OLS} - \beta$ under homoskedasticity. Since U/B is a scaled ratio of independent standard normals, it is a scaled standard cauchy, which is symmetric and centered at 0, so the distributional limit of the IV estimator is centered at the probability limit of the OLS estimator. Under heteroskedasticity, this will fail provided

$$\frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} \neq \frac{\mathrm{E}\left(uv\right)}{\mathrm{E}\left(v^2\right)}.$$

Set $z = v^2$, $u = v^3$ and let $v \sim \mathcal{N}\left(0, 1\right)$. Then $\mathrm{E}\left(u^2|z\right) = z^3, \mathrm{E}\left(v^2|z\right) = z, \mathrm{E}\left(uv|z\right) = z^2$, so we have heteroskedasticity, and

$$7 = \frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} \neq \frac{\mathrm{E}\left(uv\right)}{\mathrm{E}\left(v^2\right)} = 3.$$

An older version of this question neglected to include the condition $\pi = 0$ stated in class, but the asymptotic distribution can be derived as follows. Thanks to Reigner Kane for providing the following derivation under weak instrument asymptotics which shows why adding the condition $\pi = 0$ is necessary:

Suppose
$$\begin{pmatrix} A \\ B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{E}\left(u^2 z^2\right) & \mathrm{E}\left(z^2 uv\right) \\ \mathrm{E}\left(z^2 uv\right) & \mathrm{E}\left(z^2 v^2\right) \end{pmatrix} \right)$$
and consider the limiting distribution
$$\hat{\beta}_{IV} - \beta \xrightarrow{d} \frac{A}{\pi \mathrm{E}\left(z^2\right) + B}.$$

We saw in the slides that
$$\mathrm{E}\left(A | B = b\right) = \frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} b,$$

so $A = \frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} B + U$ where $\mathrm{E}\left(U|B\right) = 0$, so:

$$\begin{aligned} \frac{A}{\pi \mathrm{E}\left(z^2\right) + B} &= \frac{\frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} B + \frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} \pi \mathrm{E}\left(z^2\right) + U}{\pi \mathrm{E}\left(z^2\right) + B} - \frac{\frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} \pi \mathrm{E}\left(z^2\right)}{\pi \mathrm{E}\left(z^2\right) + B} \\ &= \frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} - \frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} \cdot \frac{\pi \mathrm{E}\left(z^2\right)}{\pi \mathrm{E}\left(z^2\right) + B} + \frac{U}{\pi \mathrm{E}\left(z^2\right) + B}. \end{aligned}$$

Since U is independent of B, the final term has median 0 because

$$\begin{aligned} \mathrm{P}\left(\frac{U}{\pi \mathrm{E}\left(z^2\right) + B} < 0 \right) &= \mathrm{P}\left(U < 0 \cap B > -\pi \mathrm{E}\left(z^2\right) \right) + \mathrm{P}\left(U > 0 \cap B < -\pi \mathrm{E}\left(z^2\right) \right) \\ &= \mathrm{P}\left(U < 0\right) \mathrm{P}\left(B > -\pi \mathrm{E}\left(z^2\right) \right) + \mathrm{P}\left(U > 0\right) \mathrm{P}\left(B < -\pi \mathrm{E}\left(z^2\right) \right) \\ &= \frac{1}{2}. \end{aligned}$$

Reigner's derivation shows that, under homoskedasticity, the limit distribution of IV equals the probability limit of OLS plus a term with median 0, minus a term that, for large π, is centered away from zero. Indeed, if π is large, the denominator is greater than zero with very high probability, and on this set of high probability we can say roughly that

$$\frac{\pi \mathrm{E}\left(z^2\right)}{\pi \mathrm{E}\left(z^2\right) + B} < 1 \iff \frac{\pi \mathrm{E}\left(z^2\right) + B}{\pi \mathrm{E}\left(z^2\right)} > 1,$$

where the latter occurs with probability $\frac{1}{2}$ because $\frac{\pi \mathrm{E}\left(z^2\right) + B}{\pi \mathrm{E}\left(z^2\right)}$ has mean 1. For large π the distribution is very tightly concentrated around 1, so, with high probability, its reciprocal is too, so intuitively the limit distribution is centered more towards zero for large π.

In fact, we can show that any quantile (in $(0, 1)$) of the limiting distribution converges to zero as the concentration parameter tends to infinity. To see this, write the limiting distribution of the IV estimator as

$$\hat{\beta}_{IV} - \beta \xrightarrow{d} \frac{A}{\pi \mathrm{E}\left(z^2\right) + B},$$

and note that for any $t > 0$:

$$\begin{aligned}
\mathrm{P}\left(\frac{A}{\pi \mathrm{E}\left(z^2\right) + B} > t\right) &\leq \mathrm{P}\left(\left|\frac{A}{\pi \mathrm{E}\left(z^2\right) + B}\right| > t\right) \\
&\leq \mathrm{P}\left(|A| > t\left|\pi \mathrm{E}\left(z^2\right) + B\right|\right) \\
&\leq \mathrm{P}\left(|A| > t\left|\pi \mathrm{E}\left(z^2\right)\right| - t\,|B|\right) \\
&\leq \mathrm{P}\left(|A| + t\,|B| > t\left|\pi \mathrm{E}\left(z^2\right)\right|\right) \\
&\to 0
\end{aligned}$$

as $|\pi| \to \infty$. The third inequality follows from $|a - b| \geq |a| - |b|$. This means we can always find π large enough such that the q-th quantile of the limiting distribution is less than t.

For the rest of this question, retain the homoskedasticity assumptions you stated in part c).

e) Reproduce figure 1a in Stock, Wright and Yogo (2002) (here) by simulating 10,000 draws of the IV estimator and plotting a kernel density estimate (Set $n = 500$, $z_i = 1$ for all $i \leq n$, and choose π_n to fix the concentration parameter accordingly).

f) Derive the t-test statistic based on the normal approximation of the IV estimator for the null hypothesis $\beta = \beta_0$. Find its limit distribution under H_0 assuming $\pi_n = \pi \neq 0$.

ANS: With $\pi_n = \pi \neq 0$,

$$\sqrt{n}\left(\hat{\beta}_{IV} - \beta\right) = \mathcal{N}\left(0, \frac{\sigma_u^2 \mathrm{E}\left(z^2\right)}{\mathrm{E}\left(zx\right)^2}\right).$$

A consistent estimate of σ_u^2 is

$$\hat{\sigma}_u^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_{IV} x_i\right)^2,$$

so by Slutsky's Theorem and under H_0 the t-stat satisfies

$$T_n = \frac{\sqrt{n}\left(\hat{\beta}_{IV} - \beta_0\right)}{\sqrt{\frac{\hat{\sigma}_u^2 \frac{1}{n}\sum_{i=1}^{n} z_i^2}{\frac{1}{n}\sum_{i=1}^{n} z_i x_i}}} \xrightarrow{d} \mathcal{N}\left(0, 1\right).$$

g) Suppose $\beta = \beta_0$ is true. Simluate (using the same design as in part e)) the rejection probability of the t-test for different values π under weak instrument asymptotics. Show that the anderson rubin test of the same null hypothesis has asymptotically correct rejection probability no matter the strength of the instruments and confirm this with a simulation.

ANS: The AR test is an F test of the significance of γ in

$$y_i - \beta_0 x_i = \gamma z_i + u_i, \quad \mathrm{E}\left(zu\right) = 0.$$

Under H_0, $y_i - \beta_0 x_i = \gamma z_i + u_i$ with $\gamma = 0$ and where $\mathrm{E}(zu) = 0$ holds because of the assumption $\mathrm{E}(u|z) = 0$. We have

$$\sqrt{n}\hat{\gamma} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i^2\right)^{-1} \frac{1}{\sqrt{n}} \sum z_i u_i \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_u^2}{\mathrm{E}(z)^2}\right)$$

The test statistic under H_0 satisfies

$$T_n = \frac{n \cdot (\hat{\gamma} - 0)^2}{\frac{\hat{\sigma}_u^2}{\frac{1}{n}\sum_{i=1}^{n} z_i^2}} \xrightarrow{d} \chi_1^2,$$

where

$$\hat{\sigma}_u^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \beta_0 x_i - \hat{\gamma} z_i)^2 .$$

For completeness, note that this is the same as the formula for the AR test Conroy posted in his TA slides up to degrees of freedom corrections:

$$\begin{aligned}
(Y - X\beta_0) M_Z (Y - X\beta_0) &= n\hat{\sigma}_u^2; \\
(Y - X\beta_0) P_Z (Y - X\beta_0) &= (Z\hat{\gamma})' (Z\hat{\gamma}) \\
&= \hat{\gamma}^2 \sum_{i=1}^{n} z_i^2 .
\end{aligned}$$

We find

$$\frac{(Y - X\beta_0) P_Z (Y - X\beta_0)}{(Y - X\beta_0) M_Z (Y - X\beta_0)} = \frac{\hat{\gamma}^2 \sum_{i=1}^{n} z_i^2}{n\hat{\sigma}_u^2} = \frac{n \cdot (\hat{\gamma} - 0)^2}{\frac{\hat{\sigma}_u^2}{\frac{1}{n}\sum_{i=1}^{n} z_i^2}} .$$

None of this derivation depended on π, so regardless of the strength of the instruments the Anderson Rubin test has asymptotically correct null rejection probability. Your simulation should confirm this. For the t-test, however, the strength of the instrument does impact the limiting distribution under weak instrument asymptotics. We saw from part e) that the pdf of $\hat{\beta}_{IV}$ has more of its mass centered close to 0.99 the weaker the instrument is, so we might expect false rejection more often with weak instruments as the distribution shifts further from being centered at 0. We can simulate the rejection probability of the t-test by specifying a value of π in the first stage, generating the data and checking whether the t-stat exceeds $z_{0.975}$. We can also derive the limiting distribution of the t-stat under weak instrument asymptotics. Under H_0:

$$\begin{aligned}
T_n &= \frac{\sqrt{n}\left(\hat{\beta}_{IV} - \beta_0\right)}{\sqrt{\frac{\hat{\sigma}_u^2 \frac{1}{n}\sum_{i=1}^{n} z_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i\right]^2}}} \\
&= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i}{\sqrt{\hat{\sigma}_u^2 \cdot \frac{1}{n}\sum_{i=1}^{n} z_i^2}} .
\end{aligned}$$

Now note that

$$
\begin{aligned}
\hat{\sigma}_u^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_{IV}x_i\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\beta_0 x_i + u_i - \hat{\beta}_{IV}x_i\right)^2 \\
&= \left(\beta_0 - \hat{\beta}_{IV}\right)^2 \frac{1}{n}\sum_{i=1}^{n} x_i^2 + 2\left(\beta_0 - \hat{\beta}_{IV}\right)\frac{1}{n}\sum_{i=1}^{n} x_i u_i + \frac{1}{n}\sum_{i=1}^{n} u_i^2 \qquad (2) \\
&\xrightarrow{d} \left[\frac{A}{\pi \mathrm{E}\left(z^2\right) + B}\right]^2 \sigma_v^2 - 2\cdot \frac{A}{\pi \mathrm{E}\left(z^2\right) + B}\cdot \sigma_{uv} + \sigma_u^2 \\
&= \sigma_u^2 \left(\left[\frac{A/\sigma_u}{\frac{\pi \mathrm{E}(z^2)}{\sigma_v} + B/\sigma_v}\right]^2 + 1 - 2\cdot \frac{A/\sigma_u}{\frac{\pi \mathrm{E}(z^2)}{\sigma_v} + B/\sigma_v}\cdot \rho\right)
\end{aligned}
$$

where $\rho = Corr\,(u, v)$. Note that

$$
\hat{\beta}_{IV} - \beta_0 = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i}{\pi_n \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i^2 + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i v_i},
$$

so (2) reveals that the t-statistic is a continuous function of

$$
\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i v_i \end{pmatrix}
$$

and other quantities that converge in probability. We conclude that

$$
\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i}{\sqrt{\hat{\sigma}_u^2 \cdot \frac{1}{n}\sum_{i=1}^{n} z_i^2}} \xrightarrow{d} \frac{\frac{A}{\sigma_u\sqrt{\mathrm{E}(z^2)}}}{\sqrt{1 + \left[\frac{\frac{A}{\sigma_u\sqrt{\mathrm{E}(z^2)}}}{\frac{\pi\sqrt{\mathrm{E}(z^2)}}{\sigma_v} + \frac{B}{\sigma_v\sqrt{\mathrm{E}(z^2)}}}\right]^2 - 2\rho\frac{\frac{A}{\sigma_u\sqrt{\mathrm{E}(z^2)}}}{\frac{\pi\sqrt{\mathrm{E}(z^2)}}{\sigma_v} + \frac{B}{\sigma_v\sqrt{\mathrm{E}(z^2)}}}}}.
$$

Let's simplify this by noting that

$$
\begin{pmatrix} \frac{A}{\sigma_u\sqrt{\mathrm{E}(z^2)}} \\ \frac{B}{\sigma_v\sqrt{\mathrm{E}(z^2)}} \end{pmatrix} := \begin{pmatrix} A' \\ B' \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{uv}/\sigma_u\sigma_v \\ \sigma_{uv}/\sigma_u\sigma_v & 1 \end{pmatrix}\right),
$$

so

$$
T_n \xrightarrow{d} \frac{A'}{\sqrt{1 - 2\rho\frac{A'}{\mu + B'} + \left[\frac{A'}{\mu + B'}\right]^2}},
$$

where the limiting distribution depends only on $\rho = Corr\,(u, v)$ and μ. Using this result and the given parameters we can simulate and plot the asymptotic approximation to $\mathrm{P}\left(|T_n| > z_{1-\alpha}|\mu\right)$:

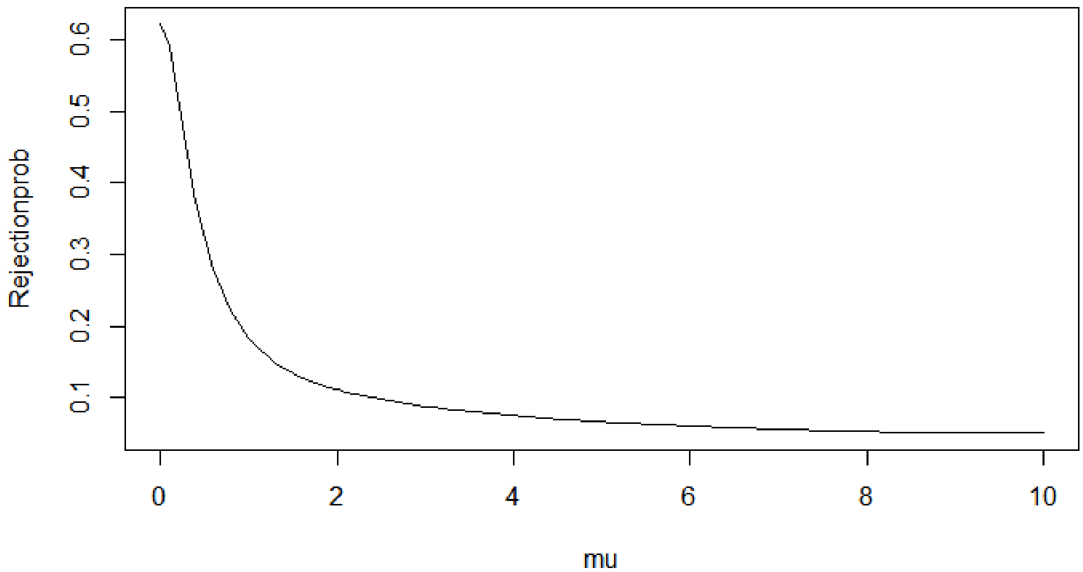

Figure 1: Asymptotic approximation to $\mathrm{P}\left(|T_n| > z_{0.975}|\mu\right)$

h) Suppose you are interested in testing the validity condition $\mathrm{E}\left(zu\right) = 0$, but do not make any assumption about β. What is the value of the test statistic used for a test of overidentifying restrictions from the slides in this setting with $\dim\left(x\right) = \dim\left(z\right) = K$? Can you use this test statistic to construct a test with desirable properties under H_0 and H_1?

ANS: A test of overidentifying restrictions would be derived from the OLS estimate of γ in the regression

$$y_i - x_i'\hat{\beta}_{IV} = z'\gamma + \epsilon.$$

The OLS estimate would be

$$\hat{\gamma} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \frac{1}{n}\sum z_i\left(y_i - x_i'\hat{\beta}_{IV}\right) = 0,$$

since the IV estimator is constructed by solving

$$\frac{1}{n}\sum z_i\left(y_i - x_i'\hat{\beta}_{IV}\right) = 0.$$

The F-statistic for the joint hypothesis that $\gamma = 0$ is therefore identically zero, so any test which rejects with probability α under H_0 based on this statistic will also reject with probability α under H_1, so the test statistic isn't useful. Intuitively, we can't test whether the instruments are exogenous if we only have as many instruments as regressors, since it's always possible to solve $\mathrm{E}\left(z\left(y - x'b\right)\right) = 0$ by setting $b = \mathrm{E}\left(zx'\right)^{-1}\mathrm{E}\left(zy\right)$, even though $\mathrm{E}\left(zu\right) = 0$ may fail.