ECON 31720: Problem Set 2

Nadia Lucas

November 10, 2020

Problem 1

Let $\mathcal{H}$ denote the set of all proper distribution function for a scalar random variable. We observe a non-negative random variable Y with distribution $G \in \mathcal{G}$ where $\mathcal{G}$ is the subset of $\mathcal{H}$ such that $P_G[Y \geq 0] = 1$. Y is determined by

$$Y = \max\{U, 0\}$$

where U is an unobserved random variable with distribution F. Assume $F \in \mathcal{F}$ where $\mathcal{F}$ is some subset of $\mathcal{H}$. This implies

$$G_F(y) \equiv P_F[\max\{U, 0\} \leq y]$$

We let $\mathcal{F}^*(G)$ denote the sharp identified set for F

$$\mathcal{F}^*(G) = \{F \in \mathcal{F} : G_F(y) = G(y) \quad \forall y \in \mathbb{R}\}$$

Consider the target parameter $\pi : \mathcal{F} \to \mathbb{R}$ defined as $\pi(F) = E_F[U]$. Let $\Pi^*(G)$ denote the sharp identified set for π.

Recall a model is falsifiable if there exists a known function $\tau : \mathcal{G}]\{0, 1\}$ such that whenver $\tau(G) = 1, \Theta^*(G) = \emptyset$ and $\tau(G) = 1$ for at least one $G \in \mathcal{G}$.

a

Suppose $\mathcal{F} = \mathcal{H}$. Then we determine $\Pi^*(G)$ for any $G \in \mathcal{G}$ and show it can never be empty. To cover the set of any $G \in \mathcal{G}$, if $G(0) > 0$, then $\Pi^*(G) = (-\infty, E[G]]$. If $G(0) = 0$, then $\Pi^*(G) = \{E[G]\}$. The model is not falsifiable because neither set is empty.

b

Suppose that $\mathcal{F} = \{F \in \mathcal{H} : F(-1) = 0 \text{ and } F(2) = 1\}$. This model is falsifiable because $G(2) < 1$ gives $\Pi^*(G) = \emptyset$ and would lead to rejecting the model.

c

Suppose $\mathcal{F} = \{F \in \mathcal{H} : F(0) = 1/2\}$. If $G(0) \neq 1/2$, then $\Pi^*(G) = \emptyset$ and the model is falsifiable.

d

Suppose $\mathcal{F} = \{F \in \mathcal{H} : E_F[U] = 0\}$. If $G(0) = 0$ and $E[G] \neq 0$, then $\Pi^*(G) = \emptyset$ and the model is falsifiable.

e

Changing the target parameter to

$$\pi(F) \equiv \text{med}_F(U) \equiv \inf\{u : F(u) \geq 1/2\}$$

Suppose that $\mathcal{F} = \mathcal{H}$. Similarly to part a, if $\text{med}(G) > 0$ then $\Pi^*(G) = \text{med}(U)$. If $\text{med}(G) = 0$, then $\Pi^*(G) = (-\infty, 0]$. This covers the space of G and none of the sets are empty. Therefore, this model is not falsifiable.

Problem 2

Consider the simple IV

$$Y = \alpha X + U$$

$\hat{F}$ is the sample first-stage F-statistic. The asymptotic bias of $\hat{F}$ as an estimator of the concentration parameter μ^2. From the supplement, the sample F-statistic is defined as

$$\hat{F} \equiv n\frac{\hat{\pi}^2}{\hat{\sigma}_\pi^2} = \left(\frac{\sqrt{n}\hat{\pi}}{\hat{\sigma}_\pi}\right)^2$$

where $\hat{\sigma}_\pi$ is a consistent estimator of $_\pi$. Under weak instrument asymptotics, we have

$$\begin{aligned}
\sqrt{n}\hat{\pi} &\equiv \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n Z_i(Z_i\pi_n + V_i)}{(\frac{1}{n}\sum_{i=1}^n Z_i^2)} \\
&= \sqrt{n}\pi_n + \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n Z_iV_i}{(\frac{1}{n}\sum_{i=1}^n Z_i^2)}
\end{aligned}$$

We see that in the last expression, if we can get subtract off $\frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n Z_iV_i}{(\frac{1}{n}\sum_{i=1}^n Z_i^2)}$ from $\sqrt{\hat{F}}$, then we converge in distribution to π by Slutsky and the CLT.

$$\begin{aligned}
\sqrt{n}\hat{\pi} - \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n Z_iV_i}{(\frac{1}{n}\sum_{i=1}^n Z_i^2)} &\equiv \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n Z_i(Z_i\pi_n + V_i)}{(\frac{1}{n}\sum_{i=1}^n Z_i^2)} - \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n Z_i(Z_i\pi_n + V_i)}{(\frac{1}{n}\sum_{i=1}^n Z_i^2)} \\
&= \sqrt{n}\pi_n
\end{aligned}$$

Thus we have the unbiased estimator:

$$\begin{aligned}
\hat{F}_{adj} &\equiv \left(\frac{\sqrt{n}\hat{\pi} - \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n Z_iV_i}{(\frac{1}{n}\sum_{i=1}^n Z_i^2)}}{\hat{\sigma}_\pi}\right)^2 \\
&= \left(\frac{\sqrt{n}\pi_n}{\hat{\sigma}_\pi}\right)^2 \\
&\xrightarrow{d} \left(\frac{\pi E[Z^2]^{1/2}}{\sigma_V}\right) \\
&\equiv \mu^2
\end{aligned}$$

Problem 3

Consider the binary treatment potential outcomes model. Maintain monotonicity assumption (no such thing as defiers) and assume $P[D = 1 \mid Z = 1] - P[D = 1 \mid Z = 0] > 0$.

a

$F_1(y \mid a)$ is point identified for any y. This is the probability distribution of always-takers which is given by

$$F_1(y \mid a) = P[Y \leq y \mid D = 1, Z = 0]$$

Which is observed in the data

b

Similarly,

$$F_0(y \mid n) = P[Y \leq y \mid D = 0, Z = 1]$$

which is the probability distribution of never-takers and is also observed in data.

c

Now we have to weight observed distributions to get complier distributions. Denote

$$\begin{aligned} p_a &= P[D = 1 \mid Z = 0] \\ p_n &= P[D = 0 \mid Z = 1] \\ p_c &= 1 - P[D = 1 \mid Z = 0] - P[D = 0 \mid Z = 1] \end{aligned}$$

And we know from the assumption that p_c is strictly positive. From data, we know the distribution, $P[Y \leq y \mid D = 1, Z = 1]$ which is a combination of both always takers and compliers. But given the probability weights on compliers and always takers, along with the distribution of the always takers, we can point identify $F_1(y \mid c)$

$$P[Y \leq y \mid D = 1, Z = 1] = F_1(y \mid c)\frac{p_c}{p_c + p_a} + F_1(y \mid a)\frac{p_a}{p_c + p_a}$$

Plugging in from part a and rearranging

$$F_1(y \mid c) = \left(P[Y \leq y \mid D = 1, Z = 1] - P[Y \leq y \mid D = 1, Z = 0])\frac{p_a}{p_c + p_a}\right)\frac{p_c + p_a}{p_c}$$

which is point identified for any y. Similarly,

$$\begin{aligned} P[Y \leq y \mid D = 0, Z = 0] &= F_0(y \mid c)\frac{p_c}{p_c + p_n} + F_0(y \mid n)\frac{p_n}{p_c + p_n} \\ \implies F_0(y \mid c) &= \left(P[Y \leq y \mid D = 0, Z = 0] - P[Y \leq y \mid D = 0, Z = 1])\frac{p_n}{p_c + p_n}\right)\frac{p_c + p_n}{p_c} \end{aligned}$$

which is also point identified.

d

Define $U(0) \equiv F_0(Y(0))$ and $U(1) \equiv F_1(Y(1))$. Assume $U(0) = U(1) = U$ with probability 1. We want to find

$$F_0(y) = P[Y(0) \leq y] = P[G = a] * F_0(y \mid a) + P[G = c]F_0(y \mid c) + P[G = n]F_0(y \mid n)$$

The only thing that is not observed if $F_0(y \mid a)$. We can find any y's that overlap between the always takers and the compliers in the F_1 distribution

$$F_1(y^* \mid a) = F_1(y \mid c)$$

And with rank invariance, it must be that using these y^*s

$$F_0(y \mid a) = F_0(y^* \mid c)$$

so $F_0(y)$ is point identified.

Similarly, we can use rank invariance and $F_1(y \mid c)$ to get $F_y(y \mid n)$ on the proper support and solve for $F_1(y)$

e

If G is the unconditional distribution function of $Y(1)-Y(0)$, then we can define it by differentiating the point-identified F_1 and F_0 distributions found in part c.

$$G(y) = P[Y(1) - Y(0) \leq y] = \int_{-\infty}^{\infty} f_1(x)\mathbb{1}\{F_1^{-1}(x) - y \leq F_0^{-1}(x)\}dx$$

Problem 4

a

The interpretation of the monotonicity condition when using the same-sex instrument separately is that parents who have two children of the same sex are more likely to have a third child to try for a different sex. The monotonicity further implies that there are no defiers, or in other words, there are no parents who want a third child if the first two are different sex but not if they are the same sex. This seems possible, say parents who prefer to have children of one sex over another. I'm less inclined to believe the monotonicity condition is entirely credible.

b

The twin birth instrument claims that parents expecting twins from the second pregnancy are induced into having three children when otherwise could have stopped at two. The credibility of the monotonicity assumption holds because in this case there are no defiers.

The exogeneity of the instrument more generally is questionable because twins are more common among older women or women who use IVF who might be more career-oriented.

c

In the multiple instruments case, we cannot rule out defiers from the same-sex instrument, and therefore the monotonicity assumption does not hold for the combined instrument.

Problem 5

Consider $D \in \{0, 1\}$ and $Z \in \{0, 1\}$. $X \in \{1, ..., K\}$ is a vector of covariate that has a discrete distribution with K points of support. β_{tsls} denotes population coefficient on D following 2SLS specification.

a

The monotonicity condition here is conceptually the same except we restrict $D(1) \geq D(0)$

b

Show that

$$\beta_{\text{tsls}} = E\left[\frac{Cov(D, Z \mid X)}{E[Cov(D, Z \mid X)]} E[Y(1) - Y(0) \mid D(1) = 1, D(0) = 0, X]\right]$$

We start with the model

$$Y = \alpha_0 + \hat{\beta} D + \gamma X + \epsilon$$

In a switching regression framework

$$\hat{\beta} = E[Y(1) - Y(0) \mid D(1) = 1, D(0) = 0]$$

This gives

$$\begin{aligned}
\beta_{\text{tsls}} &= \frac{Cov(Y, Z)}{Cov(D, Z)} \\
&= \frac{E[Cov(Y, Z \mid X)]}{E[Cov(D, Z \mid X)]} \\
&= E\left[\frac{Cov(Y, Z \mid X)}{E[Cov(D, Z \mid X)]}\right] \\
&= E\left[\frac{Cov(\hat{\beta} D, Z \mid X)}{E[Cov(D, Z \mid X)]}\right]
\end{aligned}$$

By the conditional independence assumption

$$\begin{aligned}
&= E\left[\frac{Cov(D, Z \mid X) E[Y(1) - Y(0) \mid D(1) = 1, D(0) = 0, X]}{E[Cov(D, Z \mid X)]}\right] \\
&= E\left[\frac{Cov(D, Z \mid X)}{E[Cov(D, Z \mid X)]} E[Y(1) - Y(0) \mid D(1) = 1, D(0) = 0, X]\right]
\end{aligned}$$

c

The result shows that β_{tsls} captures only the effect from the support of X that are affected by the treatment. Note that $E[Y(1) - Y(0) \mid D(1) = 1, D(0) = 0, X = x] = LATE(x)$ so the IV estimator is a weighted sum of LATEs.

Problem 6

Table 1: Monte Carlo simulations

N = 100				
	Median	Bias	Standard deviation	95% coverage
OLS	1.589389	0.588851	0.061693	0.248727
IV (1)	0.998982	-0.020507	0.172868	0.681267
IV (20)	1.281195	0.279585	0.106936	0.420430
Jackknife (1)	0.949941	-0.080003	0.212403	0.827507
Jackknife(20)	0.971610	-0.237782	2.954564	1.640808

N = 200				
	Median	Bias	Standard deviation	95% coverage
OLS	1.587365	0.588432	0.044854	0.174787
IV (1)	1.002676	-0.011123	0.128450	0.500091
IV (20)	1.169593	0.163110	0.098306	0.386372
Jackknife (1)	0.978459	-0.036636	0.139038	0.540751
Jackknife(20)	0.979480	-0.050990	0.193649	0.756578

N = 400				
	Median	Bias	Standard deviation	95% coverage
OLS	1.588613	0.588458	0.030367	0.120940
IV (1)	1.002973	-0.001156	0.083088	0.324190
IV (20)	1.096655	0.092101	0.069536	0.274073
Jackknife (1)	0.991590	-0.012854	0.086069	0.334817
Jackknife(20)	0.989813	-0.017040	0.096835	0.372660

N = 800				
	Median	Bias	Standard deviation	95% coverage
OLS	1.589672	0.589337	0.021706	0.085684
IV (1)	1.000504	-0.002855	0.061853	0.242319
IV (20)	1.049020	0.047672	0.055965	0.226555
Jackknife (1)	0.994826	-0.008630	0.062931	0.247020
Jackknife(20)	0.993550	-0.009163	0.066243	0.269509

We see that the OLS estimate is biased, this makes sense because we've constructed our DGP to have unobserved heterogeneity that is positively correlated with the observable heterogeneity, thus the estimate is biased upwards. We also have a slight bias upwards with the many instruments (IV20) specifications which tracks with the analytical upward bias from many instruments we went through in class. The true IV does quite well, and so does the jackknife for both 1 and many instruments. Although we do see that the coverage rate for jackknife can get quite large if N is too small.

Problem 7

Table 2: Table II Replication

	OLS	First Stage	Second Stage	Least Squares Treated
Participation in 401(k)	13,527.05		9,418.828	11,209.898
	(1,8095.9)		(1,808.471)	(2,674.93)
Constant	-23,549.003	-.030572	-23,298.7	-26,435.22
	(2,1772.58)	(.008702)	(2,175.5)	(2,889.27)
Family income (thousand $)	976.93	.001334	997.19	999.60
	(83.335)	(.00139)	(83.229)	(123.28)
Age (minus 25)	-376.165	-.0021605	-345.955	181.13
	(236.888)	(.00101)	(236.667)	(359.20)
Age (minus 25) squared	38.699	.0000532	37.852	27.16
	(7.663)	(.00003)	(7.657)	(11.91)
Married	-8,369.471	-.00047	-8,355.87	-8,186.567
	(1,829.238)	(.00788)	(1,829.11)	(2,792.03)
Family size	-785.65	.00006	-818.963	-1,140.051
	(410.623)	(.0024)	(410.796)	(638.54)
Eligibility for 401(k)		.6883		
		(.0080)		

Standard errors in parentheses

1

The errors are all heteroskedasticity-robust standard errors. The exception is Least Squares Treated where the errors are bootstrapped.

2

The IV estimator is 9.418.83 on the parameter of interest. The Anderson Rubin 95% confidence interval is more conservative than the robust standard error calculated and is shown to be: (6,470, 12,422). This was done using the method from the notes, cycling through all possible betas until we can no longer reject from zero. More details given in code. This is a more conservative estimate of errors because it does not require there to be a coefficient on Z, namely that it is $\pi(\beta - \beta_0)$ which will cause the estimate to have less power (Marmer, Vadim, *Econometrics with Weak Instruments: Consequences, Detections, and Solutions*, 2017).

3

The Jackknife estimator gives: 9,418.52 with bootstrapped standard error (2,492.15) which is almost identical to the IV estimator. This makes sense because we don't have the bias from the many instruments problem that led to our jackknife being useful in part 6.