Econ 710A Problem Set 2

Sarah Bass

2/4/21

Problem set 2

1) Covariance estimator $\to_p$ covariance by WLLN and CMT

a) $$\frac{\widehat{\text{cov}}(z, Y)}{\widehat{\text{cov}}(z, X)} = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(Y_i - \bar{Y})}{\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(X_i - \bar{X})}$$

$$\to_p \frac{\text{cov}(z, Y)}{\text{cov}(z, X)}$$

$$= \frac{\text{cov}(z, B_0 + XB_1 + U)}{\text{cov}(z, X)}$$

$$= \frac{\text{cov}(z, B_0) + \text{cov}(z, XB_1) + \text{cov}(z, U)}{\text{cov}(z, X)}$$

$$= \frac{0 + B_1 \text{cov}(z, X) + \text{cov}(z, U)}{\text{cov}(z, X)}$$

$$= 0 + B_1 + \frac{\text{cov}(z, U)}{\text{cov}(z, X)}$$

Note $\text{cov}(z, U) = \mathbb{E}[z\mathbb{E}[U|z]] - \mathbb{E}[z]\mathbb{E}[\mathbb{E}[U|z]]$

$= \mathbb{E}[z \cdot 2] - \mathbb{E}[z]\mathbb{E}[2]$

$= 2\mathbb{E}[z] - 2\mathbb{E}[z]$

$= 0.$

Thus $\hat{B}_1^{IV} \to_p B_1$.

b) $\hat{\beta}_0^{IV} = \bar{Y} - \bar{X}\hat{\beta}_1^{IV}$

$= \mathbb{E}[Y] - \mathbb{E}[X]\,\hat{\beta}_1^{IV}$

$\to \mathbb{E}[Y] - \mathbb{E}[X]\,\beta_1$

$= \mathbb{E}[\beta_0 + X\beta_1 + U] - \mathbb{E}[X]\beta_1$

$= \mathbb{E}[\beta_0] + \mathbb{E}[X\beta_1] + \mathbb{E}[U] - \mathbb{E}[X]\beta_1$

$= \beta_0 + \beta_1\mathbb{E}[X_1] + \mathbb{E}[\mathbb{E}[U|Z]] - \beta_1\mathbb{E}[X]$

$= \beta_0 + \mathbb{E}[2]$

$= \beta_0 + 2$

$\neq \beta_0$

Thus $\hat{\beta}_0^{IV} \not\to_p \beta_0$.

2) $Y = \beta_0 + X\beta_1 + U$

$X = \pi_0 + Z\pi_1 + V$

a) - Exogeneity: $(\mathbb{E}[U|Z] = 0,\ \mathbb{E}[V|Z] = 0]$

Z only in second eq'n of triangle form$)$

- Relevance: $\mathrm{cov}(X, Z) \neq 0$, $\pi_1 \neq 0$

b) $Y = \gamma_0 + Z\gamma_1 + \varepsilon$ $\quad \mathbb{E}[\varepsilon|Z] = 0$

$Y = \beta_0 + X\beta_1 + U$

$= \beta_0 + [\pi_0 + Z\pi_1 + V]\beta_1 + U$

$= \beta_0 + \beta_1\pi_0 + \beta_1 Z\pi_1 + \beta_1 V + U$

$\gamma_0 = \beta_0 + \beta_1\pi_0$

$\gamma_1 = \beta_1\pi_1$

$\varepsilon = \beta_1 V + U$

c) $$\frac{\hat{\gamma}_1}{\hat{\pi}_1} = \frac{\left[\frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})(z_i-\bar{z})\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})(y_i-\bar{y})\right]}{\left[\frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})(z_i-\bar{z})\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})(x_i-\bar{x})\right]}$$

$$= \frac{\left[\frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})(y_i-\bar{y})\right]}{\left[\frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})(x_i-\bar{x})\right]}$$

$$= \hat{\beta}_{IV}$$

sorry, I can't draw xi right ∴

d) Let $u = v\delta_2 + \xi$, where $\delta_2 = \dfrac{cov(v,u)}{Var(v)}$.

Then
$$\begin{aligned}
cov(v,\xi) &= cov(v, u - v\delta_2) \\
&= cov(v,u) - cov(v, v\delta_2) \\
&= cov(v,u) - \delta_2\, cov(v,v) \\
&= cov(v,u) - \frac{cov(v,u)}{var(v)}\, var(v) \\
&= 0.
\end{aligned}$$

Also,
$$\begin{aligned}
cov(x\ \xi) &= cov(\pi_0 + z\pi_1 + v, \xi) \\
&= cov(\pi_0, \xi) + cov(z\pi_1, \xi) + cov(v,\xi) \\
&= \pi_1\, cov(z,\xi) \\
&= \pi_1\, cov(z, u - v\delta_2) \\
&= \pi_1\, cov(z,u) - \pi_1\, cov(z, v\delta_2) \\
&= \pi_1\, cov(z,u) - \pi_1\delta_2\, cov(z,v) \\
&= 0.
\end{aligned}$$

So if we define $\delta_0 = B_0$, $\delta_1 = B_1$, we have
$$y = \delta_0 + x\delta_1 + v\delta_2 + \xi.$$

e) $\hat{V}_i = X_i - \hat{\pi}_0 - Z_i\hat{\pi}_1$ Using partitioned regression:

$$C_i = 1 - \hat{V}_i \underbrace{\left(\sum_{i=1}^{n} \hat{V}_i\right)}_{=0} \left(\sum_{i=1}^{n} \hat{V}_i^2\right)^{-1} = 1$$

$$\tilde{X}_i = X_i - \hat{v}_i \underbrace{\left(\sum_{i=1}^{n} \hat{V}_i X_i\right)\left(\sum_{i=1}^{n} \hat{V}_i^2\right)}_{=1}$$

$$= X_i - \hat{V}_i$$

$$= X_i - (X_i - \hat{\pi}_0 - Z_i\hat{\pi}_1)$$

$$= \hat{\pi}_0 + Z_i\hat{\pi}_1$$

Using these, we can solve for $\hat{\delta}_1$ as follows:

$$\hat{\delta}_1 = \frac{\widehat{\text{cov}}(\tilde{X}, Y)}{\widehat{\text{var}}(\tilde{X})}$$

$$= \frac{\widehat{\text{cov}}(\hat{\pi}_0 + Z\hat{\pi}_1, Y)}{\widehat{\text{var}}(\hat{\pi}_0 + Z\hat{\pi}_1)}$$

$$= \frac{\widehat{\text{cov}}(\hat{\pi}_0, Y) + \text{cov}(Z\hat{\pi}_1, Y)}{\widehat{\text{var}}(\hat{\pi}_0 + Z\hat{\pi}_1)}$$

$$= \frac{\hat{\pi}_1 \text{cov}(Z, Y)}{\hat{\pi}_1^2 \text{var}(Z)}$$

$$= \frac{1}{\hat{\pi}_1} \frac{\text{cov}(Z, Y)}{\text{var}(Z)}$$

$$= \frac{\hat{\gamma}_1}{\hat{\pi}_1} \quad \text{from 2b.}$$

$$= \hat{\beta}_1$$

Thus the control variable estimator is equal to the IV estimator of β_1.

3) a) β_1 is the change in the probability of the mother working in the last year that results from having more than 2 children in the household.

b) X_1 could be endogenous. It could be the case that women with "better" jobs (higher pay, greater utility from work) prefer to have fewer children so that they can focus on their career. Similarly, women who would like to have more children may select careers where they can more easily step in and out of the workforce.

c) In this case B_1 would be the change in probability of the husband working in the last year that results from having more than 2 children in the household. X_1 is likely to be endogenous for the same reasons described in part b, with the OLS coefficient overstating the reduction in labor supply caused by having >2 children.

d) In order for z_1 to be a good instrument, it needs to be relevant and exogenous. z_1 is relevant to X_1 because parents whose first 2 children are the same sex may be more inclined to have a 3rd child with the hope of having a child of the opposite sex. z_1 satisfies exogeneity because sex of a child is determined by nature.

e) The regression output from Stata is shown below. We can see that z_1 is a relevant instrument for X_1 because samesex has a statistically significant non-zero value.

Source	SS	df	MS
Model	7974.62958	8	996.828698
Residual	86948.2882	394,831	.220216468
Total	94922.9177	394,839	.240409174

Number of obs	=	394,840
F(8, 394831)	=	4526.59
Prob > F	=	0.0000
R-squared	=	0.0840
Adj R-squared	=	0.0840
Root MSE	=	.46927

morekids	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
samesex	.0611486	.0014944	40.92	0.000	.0582195	.0640777
agem	.0302059	.0002335	129.39	0.000	.0297483	.0306634
agefstm	-.0451303	.0002821	-159.99	0.000	-.0456832	-.0445775
boy1st	-.007932	.0014944	-5.31	0.000	-.0108611	-.0050029
boy2nd	-.0086896	.0014945	-5.81	0.000	-.0116187	-.0057605
blackm	.071419	.0023633	30.22	0.000	.066787	.0760511
hispm	.1562174	.0043981	35.52	0.000	.1475972	.1648377
othracem	.0721126	.0044892	16.06	0.000	.063314	.0809113
_cons	.3633578	.0072817	49.90	0.000	.349086	.3776296

f) Below are the replication results. As we can see, the results varied slightly across models, however we can see consistently that there is a negative relationship between labor supply and children.

	All W., OLS	All W., 2SLS	M. W., 2SLS	H. of M.W., OLS	H. of M.W., 2SLS
Worked for pay	-.176	-.117	-.117	-.007	.004
	(.002)	(.025)	(.028)	(.001)	(.009)
Weeks worked	-8.978	-5.559	-5.272	-.741	.613
	(.072)	(1.118)	(1.218)	(.044)	(.598)
Hours per week	-6.647	-4.547	-4.784	.254	.539
	(.062)	(.954)	(1.023)	(.052)	(.702)