Problem Set #2

Danny Edgel
Econ 710: Economic Statistics and Econometrics II
Spring 2021

February 3, 2021

Collaborated with Sarah Bass, Emily Case, Michael Nattinger, and Alex Von Hafften

Question 1

(i) Yes, $\hat{\beta}_1^{IV} \to_p \beta_1$. By the Weak Law of Large Numbers (WLLN) and the recognition that the law of iterated expectation (LIE) implies $\mathbb{E}[U] = \mathbb{E}[\mathbb{E}[U|Z]] = 2$,

$$\begin{aligned}
\hat{\beta}_1^{IV} \to_p & \frac{\mathbb{E}[(Z - \mathbb{E}[Z])(Y - \mathbb{E}[Y])]}{\mathbb{E}[(Z - \mathbb{E}[Z])(X - \mathbb{E}[X])]} \\
= & \frac{\mathbb{E}[(Z - \mathbb{E}[Z])(\beta_0 + X\beta_1 + U - \mathbb{E}[\beta_0 + X\beta_1 + U])]}{\mathbb{E}[(Z - \mathbb{E}[Z])(X - \mathbb{E}[X])]} \\
= & \frac{\mathbb{E}[(Z - \mathbb{E}[Z])(\beta_0 + X\beta_1 + U - \beta_0 - \beta_1\mathbb{E}[X] - \mathbb{E}[U])]}{\mathbb{E}[(Z - \mathbb{E}[Z])(X - \mathbb{E}[X])]} \\
= & \frac{\mathbb{E}[\beta_1(Z - \mathbb{E}[Z])(X - \mathbb{E}[X]) + (Z - \mathbb{E}[Z])(U - \mathbb{E}[U])]}{\mathbb{E}[(Z - \mathbb{E}[Z])(X - \mathbb{E}[X])]} \\
= & \frac{\beta_1\mathbb{E}[(Z - \mathbb{E}[Z])(X - \mathbb{E}[X])] + \mathbb{E}[ZU - U\mathbb{E}[Z] - Z\mathbb{E}[U] + \mathbb{E}[Z]\mathbb{E}[U]]}{\mathbb{E}[(Z - \mathbb{E}[Z])(X - \mathbb{E}[X])]} \\
= & \beta_1 + \frac{2\mathbb{E}[Z] - 2\mathbb{E}[Z] + 2\mathbb{E}[Z] - 2\mathbb{E}[Z]}{\mathbb{E}[(Z - \mathbb{E}[Z])(X - \mathbb{E}[X])]} \\
= & \beta_1
\end{aligned}$$

(ii) Yes, $\hat{\beta}_0^{IV} \to_p \beta_0$. Given (i), we can calculate:

$$\hat{\beta}_0^{IV} = \overline{Y} - \overline{X}\hat{\beta}_1 \to_p \mathbb{E}[Y] - \mathbb{E}[X]\beta_1 = \beta_0$$

Question 2

(i) Z is a valid instrument if $Cov(Z, X) \neq 0$, i.e., if $\pi_1 \neq 0$.

(ii) We can derive γ_0, γ_1 and ε as functions of the structural parameters by first deriving the reduced form of the model:

$$Y = \beta_0 + (\pi_0 + Z\pi_1 + V)\,\beta_1 + U$$
$$Y = \beta_0 + \pi_0\beta_1 + Z\pi_1\beta_1 + V\beta_1 + U$$
$$Y = \gamma_0 + Z\gamma_1 + \varepsilon$$

Where:

$$\gamma_0 = \beta_0 + \pi_0\beta_1,\ \gamma_1 = \pi_1\beta_1,\ \varepsilon = V\beta_1 + U$$

(iii) The IV estimator of β_1 is

$$\hat{\beta}_1^{IV} = \frac{\widehat{Cov(Z,Y)}}{\widehat{Cov(Z,X)}}$$

And the two OLS estimators, of γ_1 and π_1 respectively, are

$$\hat{\pi}_1 = \frac{\sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X})}{\sum_{i=1}^n (Z_i - \overline{Z})^2},\ \hat{\gamma}_1 = \frac{\sum_{i=1}^n (Z_i - \overline{Z})(Y_i - \overline{Y})}{\sum_{i=1}^n (Z_i - \overline{Z})^2}$$

Then the indirect least squares estimator of β_1 is

$$\frac{\hat{\gamma}_1}{\hat{\pi}_1} = \frac{\sum_{i=1}^n (Z_i - \overline{Z})(Y_i - \overline{Y}) \sum_{i=1}^n (Z_i - \overline{Z})^2}{\sum_{i=1}^n (Z_i - \overline{Z})^2 \sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X})} = \frac{\sum_{i=1}^n (Z_i - \overline{Z})(Y_i - \overline{Y})}{\sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X})} = \hat{\beta}_1^{IV}$$

(iv) We can show that $Y = \delta_0 + X\delta_1 + V\delta_2 + \xi$ by first regressing U on V with the specification $U = \delta_2 V + \xi$. Then,

$$\delta_2 = \frac{Cov(U,V)}{Var(V)}$$

And:

$$Y = \beta_0 + X\beta_1 + U$$
$$Y = \delta_0 + X\delta_1 + \delta_2 V + \xi$$

Where $\delta_1 = \beta_1$ and:

$$\begin{aligned}
Cov(X,\xi) &= \frac{Cov(X, U - \delta_2 V)}{Var(\xi)} = \frac{Cov(\pi_0 + Z\pi_1 + V, U) - \delta_2 Cov(\pi_0 + Z\pi_1 + V, V)}{Var(\xi)} \\
&= \frac{\pi_1 Cov(Z,U) + Cov(V,U) - \delta_2\pi_1 Cov(Z,V) - \delta_2 Var(V)}{Var(\xi)} = \frac{Cov(V,U) - Cov(U,V)}{Var(\xi)} \\
&= 0 \\
Cov(V,\xi) &= \frac{Cov(V, U - \delta_2 V)}{Var(\xi)} = \frac{Cov(V,U) - \delta_2 Var(V)}{Var(\xi)} = \frac{Cov(V,U) - Cov(U,V)}{Var(\xi)} \\
&= 0
\end{aligned}$$

(v) Since $\hat{V}_i$ is an OLS residual, $\sum_{i=1}^n \hat{V}_i = 0$ and $\sum_{i=1}^n \hat{V}_i X_i = \sum_{i=1}^n \hat{V}_i^2$. Then, to calculate $\hat{\delta}_1$ using the paritition formula, we first want to obtain the residuals from a regression of 1 and X_i on $\hat{V}_i$ separately (say, $\hat{e}_1$ and $\hat{e}_X$), which we can then use to calculate the slope of a single linear regression:

$$\begin{aligned}
\hat{e}_1 &= 1 - \hat{V}_i\left(\frac{\sum_{i=1}^n \hat{V}_i}{\sum_{i=1}^n \hat{V}_i^2}\right) = 1 \\
\hat{e}_X &= X_i - \hat{V}_i\left(\frac{\sum_{i=1}^n \hat{V}_i X_i}{\sum_{i=1}^n \hat{V}_i^2}\right) = X_i - \hat{V}_i \\
\Rightarrow \hat{e}_X &= \hat{\pi}_0 + \hat{\pi}_1 Z_i \\
\Rightarrow \hat{\delta}_1 &= \frac{\widehat{Cov}(\hat{e}_X, Y)}{\widehat{Var}(\hat{e}_X)} = \frac{\sum_{i=1}^n(\hat{\pi}_0 + \hat{\pi}_1 Z_i - \hat{\pi}_0 - \hat{\pi}_1 \overline{Z}_n)(Y_i - \overline{Y}_n)}{\sum_{i=1}^n(\hat{\pi}_0 + \hat{\pi}_1 Z_i - \hat{\pi}_0 - \hat{\pi}_1 \overline{Z}_n)^2} \\
&= \frac{\hat{\pi}_1 \sum_{i=1}^n (Z_i - \overline{Z}_n) Y_i}{\hat{\pi}_1^2 \sum_{i=1}^n (Z_i - \overline{Z}_n)^2} = \frac{1}{\hat{\pi}_1}\left(\frac{\sum_{i=1}^n (Z_i - \overline{Z}_n)}{\sum_{i=1}^n (Z_i - \overline{Z}_n)^2}\right) Y_i \\
\hat{\delta}_1 &= \frac{\hat{\gamma}_1}{\hat{\pi}_1} = \hat{\beta}_1^{IV}
\end{aligned}$$

Question 3

(i) If X_1 is exogenous, then, since $\beta_1 = \frac{\partial Y}{\partial X_1}$, β_1 is the marginal effect that having more than 2 children in the household has on a family's labor supply. This would be presumed negative for the secondary working parent, since *ex ante*, we would expect that the presence of more children would lead require the heads of household to spend more time on childcare, chores, etc. It would be presumed positive for the primary working parent, as the marginal child adds new expenses in clothing, food, medical and dental care, etc.

(ii) One could convincingly argue that X_1 is endogenous because families who rear many children have likely selected into such a circumstance, with one or more of the parents having a preference for child-rearing, non-wage labor over wage labor.[1] This would bias $\hat{\beta}_1^{OLS}$ downward for the secondary working parent and upward for the primary working parent, as the OLS estimate would capture both the fixed effect of the type of family who chooses to have more than 2 children in addition to the causal effect of having more than 2 children.

(iii) My answers to (i) and (ii) were worded flexibly for a reason. I think that the exogeneity of X_1 is in question in either scenario, though the case for exogeneity is stronger if the labor supply of the male parent (assuming a two-parent household with one male and one female parent) is being

[1] As the youngest of eight children, I have some opinions on this matter that are better witheld from an econometrics assignment.

estimated. This is especially true for older sample periods, when legal and cultural constraints on the salaries and labor supply of women (especially married women, especially with children) were stronger.

(iv) I do not believe Z_1 is a valid instrument for X_1 because it is likely to be a weak instrument, though I think it satisfies the exclusion restriction. While I do not doubt that there are cases where parents decide to have a third child because they do not want two children of the same gender, I think that this correlation is too weak, and relevant to too narrow of a set of families, to provide meaningful exogenous variation in X_1.

(v) The table displays the results of the reduced-form regression of X_1 on Z_1 and X_2. the coefficient on Z_1 is highly statistically significant and leads to a 6% increase in the likelihood of having greater than 2 children. This is strong evidence in favor of the relevance of Z_1.

	(1)
VARIABLES	$\mathbb{1}\{3+\text{kids}\}$
first two kids are of same sex	0.0611***
	(0.00150)
age of mom at first birth	-0.0463***
	(0.000282)
age in years of mom	0.0300***
	(0.000234)
first birth boy	-0.00795***
	(0.00150)
boy2nd	-0.00906***
	(0.00150)
=1 of black	0.0687***
	(0.00305)
=1 if hispanic	0.159***
	(0.00494)
=1 if other race (white is ref)	-0.0198***
	(0.00201)
Constant	0.404***
	(0.00731)
Observations	394,840
R-squared	0.083

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(vi) Assuming that the difference between the 2SLS coefficient and that of the OLS coefficient is due not to noise but to the validity of Z_1 as an instrument, the results suggest that X_1 is indeed endogenous and biased downward for mothers. The point estimate is biased downward for fathers, as well, but the primary difference for fathers is that the OLS estimate is significant, while the 2SLS estimate is not.

	All women		Husbands of married women	
	(1)	(2)	(7)	(8)
Esimation method	OLS	2SLS	OLS	2SLS
Instrument for *More than 2 children*	—	*Same sex*	—	*Same sex*
Dependent variable:				
Worked for pay	-.175	-.114	.003	-.001
	(.002)	(.025)	(.001)	(.01)
Weeks worked	-8.882	-5.386	-.91	.481
	(.072)	(1.111)	(.04)	(.592)
Hours/Week	-6.537	-4.407	.104	.791
	(.061)	(.9460)	(.046)	(.668)
Labor income	-3678.391	-1810.531	-1378.654	-761.801
	(35.08)	(542.1)	(86.175)	(1261.)
ln(*Family income*)	-.182	-.042	—	—
	(.004)	(.065)	—	—