# Econometrics HW2

Michael B. Nattinger*

February 2, 2021

## 1 Question 1

### 1.1 Part A

Note that, by the law of large numbers and continuous mapping theorem, $\hat{Cov}(Y, Z) \to_p Cov(Y, Z), \hat{Cov}(Z, X) \to_p Cov(Z, X)$. Then,

$$\begin{aligned}
\hat{\beta}_1^{iv} &\to_p \frac{Cov(Z, Y)}{Cov(Z, X)} \\
&= \frac{Cov(Z, \beta_0 + X\beta_1 + U)}{Cov(Z, X)} \\
&= \frac{Cov(Z, \beta_0 + X\beta_1 + U)}{Cov(Z, X)} \\
&= \frac{Cov(Z, \beta_0) + Cov(Z, X\beta_1) + Cov(Z, U)}{Cov(Z, X)} \\
&= \frac{0 + \beta_1 Cov(Z, X) + Cov(Z, U)}{Cov(Z, X)} \\
&= \beta_1 + \frac{Cov(Z, U)}{Cov(Z, X)}
\end{aligned}$$

Note that $Cov(U, Z) = E(UZ) - EUEZ =$
$E[ZE[U|Z]] - E[Z]E[U] = E[Z(2)] - EZE[E[U|Z]] = 2E[Z] - 2E[Z] = 0$. Therefore, $\hat{\beta}_1^{iv} \to_p \beta_1$.

### 1.2 Part B

By LLN and CMT,

$$\begin{aligned}
\hat{\beta}_0^{iv} &\to_p E[Y] - E[X]\beta_1 \\
&= E[\beta_0 + X\beta_1 + U] - E[X]\beta_1 \\
&= \beta_0 + \beta_1 E[X] + E[U] - E[X]\beta_1 \\
&= \beta_0 + E[E[U|Z]] \\
&= \beta_0 + 2 \neq \beta_0.
\end{aligned}$$

Therefore, $\hat{\beta}_0^{iv} \to_p \beta_0 + 2 \neq \beta_0$.

*I worked on this assignment with my study group: Alex von Hafften, Andrew Smith, and Ryan Mather. I have also discussed problem(s) with Emily Case, Sarah Bass, Katherine Kwok, and Danny Edgel.

# 2 Question 2

## 2.1 Part A

Z is a valid instrument for X so long as Z satisfies exogeneity and relevance conditions. We are given that exogeneity is satisfied because we know that $E[U, V|Z] = 0$, and Z is only in the second equation of the triangular form, and not the first. We are not given sufficient information to know with certainty that relevance is satisfied. This will be true if $\pi_1 \neq 0$. This is true because of the following:

$$\begin{aligned}
Cov(X, Z) &= Cov(\pi_0 + Z\pi_1 + V, Z) \\
&= Cov(\pi_0, Z) + Cov(Z\pi_1, Z) + Cov(V, Z) \\
&= \pi_1 + E[VZ] + EVEZ \\
&= \pi_1 + E[ZE[V|Z]] + EZE[E[V|Z]] \\
&= \pi_1.
\end{aligned}$$

## 2.2 Part B

$$\begin{aligned}
Y &= \beta_0 + X\beta_1 + U \\
&= \beta_0 + (\pi_0 + Z\pi_1 + V)\beta_1 + U \\
&= \beta_0 + \pi_0\beta_1 + Z\pi_1\beta_1 + V\beta_1 + U \\
&= \gamma_0 + Z\gamma_1 + \epsilon,
\end{aligned}$$

where $\gamma_0 = \beta_0 + \pi_0\beta_1, \gamma_1 = \pi_1\beta_1, \epsilon = V\beta_1 + U$.

## 2.3 Part C

From partitioned regression we have the following:

$$\begin{aligned}
\hat{\gamma}_1/\hat{\pi}_1 &= \left(\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2\right)^{-1}\left(\sum_{i=1}^{n}(Z_i - \bar{Z}_n)(Y_i - \bar{Y}_n)\right)\left(\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2\right)\left(\sum_{i=1}^{n}(Z_i - \bar{Z}_n)(X_i - \bar{X}_n)\right)^{-1} \\
&= \left(\sum_{i=1}^{n}(Z_i - \bar{Z}_n)(Y_i - \bar{Y}_n)\right)\left(\sum_{i=1}^{n}(Z_i - \bar{Z}_n)(X_i - \bar{X}_n)\right)^{-1} \\
&= \hat{Cov}(Z, Y)/\hat{Cov}(Z, X).
\end{aligned}$$

This is the form of $\hat{\beta}_{iv}$.

## 2.4 Part D

We will begin from the least squares projection of $U$ onto $V$. Let $U = \delta_2 V + \xi$, where $\delta_2 = \frac{E[VU]}{E[V^2]}$. Now, note the following:

$$\begin{aligned}
Var(V) &= E[V^2] - E[V]^2 = E[V^2] - E[E[V|Z]]^2 \\
&= E[V^2]. \\
Cov(V,U) &= E[VU] - E[V]E[U] = E[VU] - E[E[V|Z]]E[E[U|Z]] \\
&= E[VU] \\
\Rightarrow \delta_2 &= \frac{Cov(V,U)}{Var(V)}. \\
Cov(V,\xi) &= Cov(V, U - \delta_2 V) \\
&= Cov(V,U) - \delta_2 Cov(V,V) \\
&= Cov(V,U) - \frac{Cov(V,U)}{Var(V)} Var(V) \\
&= 0. \\
Cov(X,\xi) &= Cov(\pi_o + Z\pi_1 + V, \xi) \\
&= Cov(Z\pi_1, U - V\delta_2) + Cov(V,\xi) \\
&= \pi_1 Cov(Z,U) - \pi_1\delta_2 Cov(Z,V) \\
&= 0.
\end{aligned}$$

Therefore, if we define $\delta_0 := \beta_0, \delta_1 := \beta_1$ :

$$Y = \delta_0 + X\delta_1 + V\delta_2 + \xi$$

where $\delta_2 = \frac{Cov(V,U)}{Var(V)}, \xi = U - \delta_2 V$, and $Cov(X,\xi) = Cov(V,\xi) = 0$.

## 2.5 Part E

As in Part C, I appeal to partitioned regression:

$$\begin{aligned}
c_i &= 1 - \hat{V}_i \left(\sum_{i=1}^{n} \hat{V}_i\right)\left(\sum_{i=1}^{n} \hat{V}_i^2\right)^{-1} = 1 \\
\tilde{X}_i &= X_i - \hat{V}_i \left(\sum_{i=1}^{n} \hat{V}_i X_i\right)\left(\sum_{i=1}^{n} \hat{V}_i^2\right)^{-1} \\
&= X_i - \hat{V}_i \\
&= \hat{\pi}_0 + Z_i \hat{\pi}_1.
\end{aligned}$$

We can now calculate our OLS estimate $\hat{\delta}_2$ as a simple regression result including the constant (as our residualized constant term remains exactly a constant term) and $\tilde{X} = \hat{\pi}_0 + Z\hat{\pi}_1$:

$$\begin{aligned}
\hat{\delta}_1 &= \frac{\hat{Cov}(\tilde{X}, Y)}{\hat{Var}(\tilde{X})} \\
&= \frac{\hat{Cov}(\hat{\pi}_0 + Z\hat{\pi}_1, Y)}{\hat{Var}(\hat{\pi}_0 + Z\hat{\pi}_1)} \\
&= \frac{\hat{Cov}(Z\hat{\pi}_1, Y)}{\hat{Var}(Z\hat{\pi}_1)} \\
&= \frac{1}{\hat{\pi}_1} \frac{\hat{Cov}(Z, Y)}{\hat{Var}(Z)} \\
&= \frac{\hat{\gamma}_1}{\hat{\pi}_1} \\
&= \hat{\beta}_1^{iv}
\end{aligned}$$

The control variable estimator is equivalent to the IV estimator.

# 3 Question 3

## 3.1 Part A

$\beta_1$ is the expected change in the mother's probability of working caused by having more than 2 children in the household.

## 3.2 Part B

$X_1$ is likely to be endogenously determined. Mothers that have better jobs and work more may be more likely to prioritize their career and less likely to have more children. Additionally, mothers with more children may have less time to work. The result is a simultaneous system of endogeneous equations, and the OLS estimate is likely to overstate $X_1$'s negative effect on mother's labor supply.

## 3.3 Part C

In this case, $\beta_1$ would be the expected change in the probability that the husband worked during the year caused by having more than 2 children in the household.

As before, $X_1$ is likely to be endogenous for the same reason described in Part B, with the resulting bias in the estimated OLS coefficient being an overstatement on the negative impact on labor supply hours coming from $X_1$.

## 3.4 Part D

The two conditions $Z_1$ must satisfy to be a valid instrument for $X_1$ are relevance and exogeneity. $Z_1$ seems to me to be relevant, as it is certainly conceivable to me that parents who have two sons may want at least one daughter, and have another child to try for a daughter, or vice-versa. So, the first two children being the same sex may have some nonzero correlation with having more than two children.

In this case, one would argue that the exogeneity comes from the fact that the sex of the first two children is determined by nature. This seems hard to disagree with.

## 3.5 Part E

(matlab)

### 3.6 Part F

(matlab)