Econometrics Problem Set 3

Gigi Lin

October 3, 2018

Question 1

Supply/demand framework for coffee

Part (a)

Demand equation is $Q = \alpha_0 + \alpha_p P + \alpha_M M + u$ Supply equation is $Q = \beta_0 + \beta_p P + \beta_R R + v$

Q: quantity P: price M: exogenous income R: exogenous rainfall

To solve for the reduced-form equations for P and Q

To solve for P:

$$\begin{aligned} Q =& \alpha_0 + \alpha_p P + \alpha_M M + u \qquad (1) \\ Q =& \beta_0 + \beta_p P + \beta_R R + v \qquad (2) \end{aligned}$$

Taking (1) and (2), we get $\alpha_0 + \alpha_p P + \alpha_M M + u = \beta_0 + \beta_p P + \beta_r R + v$, which gives

$$P = \frac{\alpha_0 + \alpha_M M + u - \beta_0 - \beta_r R}{\beta_p - \alpha_p}$$

To solve for Q, rearranging (1) and (2) gives:

$$\begin{aligned} P =& \frac{Q - \alpha_0 - \alpha_m M - u}{\alpha_p} \qquad (1') \\ P =& \frac{Q - \beta_0 - \beta_p P - v}{\beta_p} \qquad (2') \end{aligned}$$

Setting (1') and (2') equal gives:

$$Q = \frac{\alpha_p(-\beta_0 - \beta_R R - v) - \beta_p(-\alpha_0 - \alpha_M M - u)}{\beta_p - \alpha_p}$$

and setting it in terms of the covariates gives

$$Q = \frac{\beta_p \alpha_0 - \beta_0 \alpha_p}{\beta_p - \alpha_p} - \frac{\alpha_p \beta_R}{\beta_p - \alpha_p}(R) + \frac{\beta_p \alpha_M}{\beta_p - \alpha_p}(M) + \frac{\beta_p u - \alpha_p v}{\beta_p - \alpha_p}$$

Similarly, P can either be solved in a similar fashion, or by setting (1') and (2') equal to get:

$$P = \frac{\alpha_0 - \beta_0}{\beta_p - \alpha_p} + \frac{\alpha_M}{\beta_p - \alpha_p}(M) - \frac{\beta_R}{\beta_p - \alpha_p}(R) + \frac{u - v}{\beta_p - \alpha_p}$$

Compactly, we can represent as:
$P_i = \delta_{10} + \delta_{11} M - \delta_{12} R + \frac{1}{\beta_p - \alpha_p} * (u - v)$
$Q_i = \delta_{20} + \delta_{21} M - \delta_{22} R + \frac{1}{\beta_p - \alpha_p} * (\beta_p u - \alpha_p v)$

Given in the question is $Cov(u, M) = E(uM) = 0$ and $Cov(u, R) = E(uR) = 0$.

$$
\begin{aligned}
Cov(P_i,u) &= 0 + \frac{\alpha_M}{\beta_p-\alpha_p}Cov(M,u) - \frac{\beta_R}{\beta_p-\alpha_p}Cov(R,u) + \frac{1}{\beta_p-\alpha_p} * Cov(u-v,u) \\
&= 0+0-0+\frac{1}{\beta_p-\alpha_p} * [Cov(u,u) - Cov(v,u)] \\
&= \frac{1}{\beta_p-\alpha_p} * [Var(u) - 0] \\
&= \frac{\sigma_u^2}{\beta_p-\alpha_p} \neq 0
\end{aligned}
$$

Since $Cov(P_i,u) \neq 0$ (assuming $(\beta_p - \alpha_p) \neq 0$), Q cannot be estimated consistently by the OLS approach as there is bound to be bias. In this case, we can use IV because there is enough covariates to do so. We can use R as an IV for P in the demand equation as we can assume R is exogeneous to demand, and M as an IV for P in the supply equation as we can assume M is exogeneous to supply.

Part (b) Now the case is that the data for rainfall R is not available and the supply equation is $Q = \beta_0 + \beta_P P + v$. Solving in a similar way, we will get

$$
P = \frac{\alpha_0 + \alpha_M M + u - \beta_0 - v}{\beta_p - \alpha_p} = \frac{\alpha_0 - \beta_0}{\beta_p - \alpha_p} + \frac{\alpha_M}{\beta_p-\alpha_p}(M) + \frac{1}{\beta_p-\alpha_p} * (u - v)
$$

and

$$
Q = \frac{\beta_p \alpha_m}{\beta_p - \alpha_p}(M) + \frac{\beta_p u - \alpha_p v}{\beta_p - \alpha_p}
$$

Given in the question is $Cov(v,M) = E(vM) = 0$ and $Cov(u,M) = E(uM) = 0$, and assumption satisfied that $Cov(v,u) = 0$:

$$
\begin{aligned}
Cov(P_i,u) &= 0 + \frac{\alpha_M}{\beta_p-\alpha_p}Cov(M,u) + \frac{1}{\beta_p-\alpha_p} * Cov(u-v,u) \\
&= 0+0+\frac{1}{\beta_p-\alpha_p}[Cov(u,u) - Cov(v,u)] \\
&= \frac{\sigma_u^2}{\beta_p-\alpha_p} \neq 0
\end{aligned}
$$

In this case, OLS estimation approach will still produce biased parameters. However, in this case, as the demand equation has enough covariates, we can still use M as an IV for P in the supply equation (exogeneous to supply), but there is not enough covariates for the IV approach in the demand equation.

Question 2

Using state-level data,

$$
gEMP_t = \beta_0 + \beta_1 gMINt + \beta_2 POP_t + \beta_3 gGSP_t + \beta_4 gGDP_4 + u_t
$$

MIN_t: minimum wage (real dollars)
POP_t: population
GSP_t: gross state product
GDP_t: US GDP
g prefix: difference in logs
$USMIN_t$: US minimum wage (real dollars)

(i) Problems with the OLS estimation of this model include:

- There may be simultaneity bias between $gGSP_t$ and $gEMP_t$ as the logged difference in gross state product may be jointly determined with the logged difference in the youth unemployment level. This means we cannot use OLS.
- There may also be reverse causality as the variables $gEMP_t$ may have influence on GSP_t, which in turn may also have a direct relationship with GDP_t.

(ii) Yes. We can assume that after controlling for the other variables in the model, $gUSMIN_t$ is uncorrelated with u_t.

(iii) $gUSMIN_t$ can be a potential IV candidate for $gMIN_t$ because it satisfies the "exogeneity" condition; it is uncorrelated with u_t so $Cov(gUSMIN_t, u_t) = 0$. This is true because of the other controls in the model. The "relevance" condition also holds so $Cov(gUSMIN_t, gMIN_t) \neq 0$. This is because MIN_t depends on the US minimum wage, so there is a relationship there.

Question 3: dataset *wage2.dta*

Q3.1 *educ* and *brthord* might be negatively correlated as the parents may have higher expectations on their first-born than their last-born child. This may be due to influence from the older generations to ensure the grandchildren receive adequate education. Therefore, *brthord* may be a relevant variable in this regression model.

Q3.2

```r
# regress educ on brthord
lm1 <- lm(educ ~ brthord, data = wage2)
summary(lm1)
```

```
##
## Call:
## lm(formula = educ ~ brthord, data = wage2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8668 -1.5842 -0.7362  2.1332  6.1117
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.14945    0.12868 109.962  < 2e-16 ***
## brthord     -0.28264    0.04629  -6.106 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.155 on 850 degrees of freedom
##   (83 observations deleted due to missingness)
## Multiple R-squared:  0.04202,    Adjusted R-squared:  0.04089
## F-statistic: 37.29 on 1 and 850 DF,  p-value: 1.551e-09
```

Running a linear regression of *educ* on *brthord* gives a negative coefficient of -0.28264, with the statistical significance being very high as the p-value is very low.

Q3.3 Using *brthord* as an IV for *educ*

```
##
## Call:
## lm(formula = lnwage ~ educ, data = wage2)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.94620 -0.24832  0.03507  0.27440  1.28106
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.973062   0.081374   73.40   <2e-16 ***
## educ        0.059839   0.005963   10.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4003 on 933 degrees of freedom
## Multiple R-squared:  0.09742,    Adjusted R-squared:  0.09645
## F-statistic: 100.7 on 1 and 933 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## ivreg(formula = lnwage ~ educ | brthord, data = wage2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8532 -0.2557  0.0435  0.2970  1.3033
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.03040    0.43295  11.619  < 2e-16 ***
## educ         0.13064    0.03204   4.078 4.97e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4215 on 850 degrees of freedom
## Multiple R-Squared: -0.02862,    Adjusted R-squared: -0.02983
## Wald test: 16.63 on 1 and 850 DF,  p-value: 4.975e-05
```

The results show that *educ* now has a positive relationship with $ln(wage)$. This is plausible as higher education plays a large factor in the attainment of a higher-paying job. Studies have shown that lower education levels can lead to lower paying jobs. *brthord* is exogeneous from the error term u as it should not matter whether one is born first.