Econometrics 710 Problem Set #11

May 6, 2015 TA: Cristian Hernandez

cahernandez3@wisc.edu

Exercise 15.1

$$\hat{\beta} = \frac{\sum_{i=1}^{n} d_i y_i}{\sum_{i=1}^{n} d_i z_i} = \frac{\frac{1}{n_1}\sum_{d_i=1} y_i}{\frac{1}{n_1}\sum_{d_i=1} z_i} = \frac{\bar{y}_1}{\bar{z}_1}$$

where $n_1 = \sum_{i=1}^{n} d_i$, and $\bar{y}_1, \bar{z}_1$ are averages of y, z within the group of $d_i = 1$

Exercise 15.2

$\hat{\boldsymbol{\beta}}_{GLS} = (X'D^{-1}X)^{-1}X'D^{-1}y$ where $D = diag(\sigma_1^2, \cdots, \sigma_n^2)$. This is equivalent to the IV estimator with instruments $Z = D^{-1}X$, since $\hat{\boldsymbol{\beta}}_{IV} = (Z'X)^{-1}Z'y = (X'D^{-1}X)^{-1}X'D^{-1}y = \hat{\boldsymbol{\beta}}_{GLS}$. $Z = D^{-1}X \Leftrightarrow z_i = \sigma_i^{-2}x_i$

Exercise 15.3

By construction, the OLS estimator minimizes the sum of squared residuals. Therefore $\tilde{e}'\tilde{e} \geq \hat{e}'\hat{e}$ for any give sample size n. For large samples,

$$\begin{aligned}
\frac{1}{n}\tilde{e}'\tilde{e} &= \frac{1}{n}\sum_{i=1}^{n}\tilde{e}_i^2 = \frac{1}{n}\sum_{i=1}^{n}(e_i - x_i'(\tilde{\beta}-\beta))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} e_i^2 - 2(\tilde{\beta}-\beta)'\left(\frac{1}{n}\sum_{i=1}^{n} x_i e_i\right) + (\tilde{\beta}-\beta)'\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)(\tilde{\beta}-\beta) \\
&\xrightarrow{p} E(e_i^2)
\end{aligned}$$

where the consistency holds because $\tilde{\beta} \xrightarrow{p} \beta$ and $Ex_ie_i \neq 0 < \infty, Ex_ix_i' < \infty$. Similarly,

$$\begin{aligned}
\frac{1}{n}\hat{e}'\hat{e} &= \frac{1}{n}\sum_{i=1}^{n} e_i^2 - 2(\hat{\beta}-\beta)'\left(\frac{1}{n}\sum_{i=1}^{n} x_i e_i\right) + (\hat{\beta}-\beta)'\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)(\hat{\beta}-\beta) \\
&= \frac{1}{n}\sum_{i=1}^{n} e_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i e_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i e_i\right) \\
&\xrightarrow{p} E(e_i^2) - (Ex_ie_i)'(Ex_ix_i')^{-1}(Ex_ie_i) < E(e_i^2)
\end{aligned}$$

The last inequality holds because $(Ex_ie_i)'(Ex_ix_i')^{-1}(Ex_ie_i) > 0$ as $Ex_ie_i \neq 0$. Thus even in large samples, $\frac{1}{n}\hat{e}'\hat{e} < \frac{1}{n}\tilde{e}'\tilde{e}$ with probability approaching one.

Exercise 15.4

The moment conditions are $\mathbb{E}\boldsymbol{z}_i(x_i - \Gamma' z_i)' = \mathbb{E}\boldsymbol{z}_i(x_i' - z_i'\Gamma) = 0$. The method of moments estimator for Γ can be obtained from the sample moment equation, $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i(x_i' - \boldsymbol{z}_i'\hat{\Gamma}) = 0$. The solution is $\hat{\Gamma} = (\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{z}_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i') = (Z'Z)^{-1}(Z'X)$

Exercise 15.5

We know from the reduced form equations that

$$\lambda = \Gamma\beta$$

Then, given that λ and Γ are identified from the reduced form equations (if the instruments we use are relevant to identify Γ), β is identified if there is a unique vector $\beta \in \mathbb{R}^k$ such that $\lambda = \Gamma\beta$.

If $\text{rank}(\Gamma) < k$ (the matrix is not full rank) it follows that there exists $x \in \mathbb{R}^k$, $x \neq 0$, such that $\Gamma x = 0$. Suppose that β is identified and pick $\beta^* = \beta + x$. Then $\Gamma\beta^* = \Gamma\beta + \Gamma x = \lambda = \Gamma\beta$, which is a contradiction. Therefore, if $\text{rank}(\Gamma) < k$, then β is not identified.

On the other hand, if $\text{rank}(\Gamma) = k$ that means that there does not exist $x \in \mathbb{R}^k$, $x \neq 0$, such that $\Gamma x = 0$. Therefore, there is a unique β such that $\lambda = \Gamma\beta$, so β is identified.

Exercise 15.6

(a) $\mathbb{E}(x_i e_i) = \mathbb{E}(\mathbb{E}(x_i e_i|x_i)) = \mathbb{E}(x_i \mathbb{E}(e_i|x_i)) = 0$, and $\mathbb{E}(x_i^2 e_i) = \mathbb{E}(\mathbb{E}(x_i^2 e_i|x_i)) = \mathbb{E}(x_i^2 \mathbb{E}(e_i|x_i)) = 0$ by the law of iterated expectations and $\mathbb{E}(e_i|x_i) = 0$. Thus $z_i = (x_i \quad x_i^2)'$ is a valid instrumental variable since $\mathbb{E} z_i e_i = 0$ and also z_i explains x_i (it is relevant). The underlying exclusion restriction is that the quadratic terms x_i^2 are not included in the structural model.

(b) Let $P_Z = Z(Z'Z)^{-1}Z', \hat{X} = P_Z X$. Since Z includes X, we have perfect prediction of X on Z, that is $\hat{X} = P_Z X = X$. Therefore,

$$\hat{\beta}_{2SLS} = (X'P_Z X)^{-1} X' P_Z y = (\hat{X}'\hat{X})^{-1}\hat{X}'y = (X'X)^{-1}X'y = \hat{\beta}_{OLS}$$

In the scalar case, these are simply equal to $\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$.

(c) The efficient GMM estimator is

$$\hat{\beta}_{GMM} = (X'Z\hat{\Omega}^{-1}Z'X)^{-1}X'Z\hat{\Omega}^{-1}Z'y$$

where $\hat{\Omega}$ is a consistent estimator of $\Omega = \mathbb{E} z_i z_i' e_i^2$. If $\hat{\Omega}^{-1} \propto (Z'Z)^{-1}$ then $\hat{\beta}_{GMM} = \hat{\beta}_{2SLS} = \hat{\beta}_{OLS}$. This efficient weight matrix can be justified under the homoskedasticity assumption, $\mathbb{E}(e_i^2|z_i) = \sigma^2$(or $\mathbb{E}(e_i^2|x_i) = \sigma^2$) because under homoskedasticity, $\Omega \propto \mathbb{E} z_i z_i'$. However, the efficient GMM estimator is generally different from 2SLS (and OLS in this case), and 2SLS is inefficient under heteroskedasticity.

Exercise 15.7

Approach 1:

If Y and W are determined outside the market, that means that they are both orthogonal to e_1 and e_2. That means that $\mathbb{E}(Ye_1) = \mathbb{E}(We_1) = \mathbb{E}(e_1) = \mathbb{E}(Ye_2) = \mathbb{E}(We_2) = \mathbb{E}(e_2) = 0$. Therefore, we can write down the following GMM problem

$$\mathbb{E}(g_i(a,b)) = \mathbb{E}\begin{pmatrix} Y(Q - a_0 - a_1 P - a_2 Y) \\ W(Q - a_0 - a_1 P - a_2 Y) \\ Q - a_0 - a_1 P - a_2 Y \\ Y(Q - b_0 - b_1 P - b_2 W) \\ W(Q - b_0 - b_1 P - b_2 W) \\ Q - b_0 - b_1 P - b_2 W \end{pmatrix} = 0$$

Note that the problem has 6 equations and 6 unknowns, so as long as there is no redundancy in the equations, the parameters are identified. That means that,

$$\begin{pmatrix} \mathbb{E}(Y) & \mathbb{E}(YP) & \mathbb{E}(Y^2) & 0 & 0 & 0 \\ \mathbb{E}(W) & \mathbb{E}(WP) & \mathbb{E}(WY) & 0 & 0 & 0 \\ 1 & \mathbb{E}(P) & \mathbb{E}(Y) & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{E}(Y) & \mathbb{E}(YP) & \mathbb{E}(WY) \\ 0 & 0 & 0 & \mathbb{E}(W) & \mathbb{E}(WP) & \mathbb{E}(W^2) \\ 0 & 0 & 0 & 1 & \mathbb{E}(P) & \mathbb{E}(W) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \mathbb{E}(YQ) \\ \mathbb{E}(WQ) \\ \mathbb{E}(Q) \\ \mathbb{E}(YQ) \\ \mathbb{E}(WQ) \\ \mathbb{E}(Q) \end{pmatrix}$$

Then, if there is a unique vector $(a, b) = (a_0, a_1, a_2, b_0, b_1, b_2)$ that solves this system of equations, then (a, b) is identified. Note that in this case, since the system is block diagonal, we only need that,

$$\text{rank}\begin{pmatrix} \mathbb{E}(Y) & \mathbb{E}(YP) & \mathbb{E}(Y^2) \\ \mathbb{E}(W) & \mathbb{E}(WP) & \mathbb{E}(WY) \\ 1 & \mathbb{E}(P) & \mathbb{E}(Y) \end{pmatrix} = 3$$

and

$$\text{rank}\begin{pmatrix} \mathbb{E}(Y) & \mathbb{E}(YP) & \mathbb{E}(WY) \\ \mathbb{E}(W) & \mathbb{E}(WP) & \mathbb{E}(W^2) \\ 1 & \mathbb{E}(P) & \mathbb{E}(W) \end{pmatrix} = 3$$

This holds true if both Y and W are correlated with P and we also need that Y and W to be linearly independent. So basically we are saying that the instruments have to be relevant (explain the endogenous variable) and that we cannot have redundancy in the instruments.

Approach 2:

Let's solve for the endogenous variables (Q, P) in terms of the exogenous variables $(1, Y, W)$ and the error terms. In matrix notation,

$$\begin{pmatrix} 1 & -a_1 \\ 1 & -b_1 \end{pmatrix} \begin{pmatrix} Q \\ P \end{pmatrix} = \begin{pmatrix} a_0 & a_2 & 0 \\ b_0 & 0 & b_2 \end{pmatrix} \begin{pmatrix} 1 \\ Y \\ W \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

Assume $a_1 \neq b_1$(one sufficient condition for this is that their sign is different which is reasonable in the supply-demand model.), then we can rewrite endogenous variable (Q, P) as a function of parameters, exogenous variables and the error terms.

$$\begin{pmatrix} Q \\ P \end{pmatrix} = \begin{pmatrix} \frac{a_1 b_0 - a_0 b_1}{a_1 - b_1} & \frac{-a_2 b_1}{a_1 - b_1} & \frac{a_1 b_2}{a_1 - b_1} \\ \frac{b_0 - a_0}{a_1 - b_1} & \frac{-a_2}{a_1 - b_1} & \frac{b_2}{a_1 - b_1} \end{pmatrix} \begin{pmatrix} 1 \\ Y \\ W \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

Let $X = \begin{pmatrix} Q \\ P \end{pmatrix}, Z = \begin{pmatrix} 1 \\ Y \\ W \end{pmatrix}, \Gamma' = \begin{pmatrix} \frac{a_1 b_0 - a_0 b_1}{a_1 - b_1} & \frac{-a_2 b_1}{a_1 - b_1} & \frac{a_1 b_2}{a_1 - b_1} \\ \frac{b_0 - a_0}{a_1 - b_1} & \frac{-a_2}{a_1 - b_1} & \frac{b_2}{a_1 - b_1} \end{pmatrix} u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$. Because of the exogeneity of Z, we have $\mathbb{E}Zu' = 0$

Then, the equation above can be written as a projection model

$$X = \Gamma' Z + u \quad \mathbb{E}Zu' = 0.$$

$\Gamma = (\mathbb{E}ZZ')^{-1}\mathbb{E}(ZX')$ are identified. Let Γ_{ij} as (i,j) elements of 3×2 matrix Γ.

If $a_2 \neq 0, b_2 \neq 0, a_1 \neq 0, b_1 \neq 0$ (This is exactly the relevancy condition of the instruments Y, W) then a_1, b_1 is identified as $\frac{\Gamma_{31}}{\Gamma_{32}}, \frac{\Gamma_{21}}{\Gamma_{22}}$. Also b_2, a_2 can be identified as $b_2 = (a_1 - b_1)\Gamma_{32}, a_2 = -(a_1 - b_1)\Gamma_{22}$. Moreover, $b_0 = \Gamma_{11} - b_1\Gamma_{21}, a_0 = b_0 - (a_1 - b_1)\Gamma_{12}$. So every structural parameter $(a_0, a_1, a_2, b_0, b_1, b_2)$ can be identified.

Exercise 15.8

(a)-(e) Let $X = [1, Educ, Exper, Exper^2, South, Black]$ as variables included in the wage equation, and let Z as instrumental variables. We use $Z = [1, Exper, Exper^2, South, Black, near4]$ for question (b) and $Z = [1, Exper, Exper^2, South, Black, near4, near2, fatheduc, mothereduc]$ for (c), (d).

For (b), (c), 2SLS estimates and standard errors are calculated as follows;

$$\begin{aligned}
&\hat{\beta}_{2SLS} = (X'P_ZX)^{-1}(X'P_ZX) \\
&\hat{V}_{\beta,2SLS} = n \cdot (X'P_ZX)^{-1}(X'P_ZDP_ZX)(X'P_ZX)^{-1} \\
&s.e.(\hat{\beta}_{2SLS,i}) = \sqrt{[\frac{1}{n}\hat{V}_{\beta,2SLS}]_{ii}}
\end{aligned}$$

where $D = \text{diag}\{\hat{e}_1^2, \cdots, \hat{e}_n^2\}$ and $\hat{e}_i = y_i - x_i'\hat{\beta}_{2SLS}$. $\hat{V}_{\beta,2SLS}$ is consistent estimator of asymptotic variance of 2SLS estimator $V_\beta = (Q'WQ)^{-1}(Q'W\Omega WQ)(Q'WQ)^{-1}$, where $Q = \mathbb{E}z_ix_i', W = \mathbb{E}(z_iz_i')^{-1}, \Omega = \mathbb{E}z_iz_i'e_i^2$

For (d), estimates and standard errors for efficient GMM are specified as follows;

$$\begin{aligned}
&\hat{\beta}_{GMM} = (X'Z\hat{\Omega}^{-1}Z'X)^{-1}(X'Z\hat{\Omega}^{-1}Z'X) \\
&\hat{\Omega} = \left(\frac{1}{n}\sum_{i=1}^n z_iz_i'\hat{e}_i^2 - (\frac{1}{n}\sum_{i=1}^n z_i\hat{e}_i)(\frac{1}{n}\sum_{i=1}^n z_i\hat{e}_i)'\right) \\
&\hat{V}_{\beta,GMM} = n^2 \cdot \left(X'Z\hat{\Omega}^{-1}Z'X\right)^{-1}
\end{aligned}$$

where $\hat{e}_i = y_i - x_i'\hat{\beta}_{2SLS}$ is residuals from 2SLS in (c).

All the estimation results for OLS, 2SLS, GMM are as follows;

For (e), we report J statistic

$$J_n(\hat{\beta}_{GMM}) = n(\frac{1}{n}\sum_{i=1}^n z_i\hat{e}_i)'\hat{\Omega}^{-1}(\frac{1}{n}\sum_{i=1}^n z_i\hat{e}_i)$$

where $\hat{e}_i = y_i - x_i'\hat{\beta}_{GMM}$ We reject the null since $J_n = 17.0002$ exceeds 5 % critical value $\chi^2_{0.95}(3) = 7.815$.

	OLS	2SLS(1)	2SLS(2)	GMM
Education	0.0802	0.2226	0.1176	0.1170
	(0.0043)	(0.0577)	(0.0128)	(0.0128)
Experience	0.0893	0.1507	0.1054	0.1054
	(0.0082)	(0.0270)	(0.0098)	(0.0098)
Experience squared	−0.0025	−0.0027	−0.0025	−0.0025
	(0.0004)	(0.0004)	(0.0001)	(0.0004)
South	−0.1359	−0.0856	−0.1227	−0.1234
	(0.0177)	(0.0294)	(0.0185)	(0.0185)
Black	−0.1591	−0.0359	−0.1267	−0.1288
	(0.0238)	(0.0562)	(0.0263)	(0.0261)
Constant	4.7351	2.2659	4.0867	4.0945
	(0.0839)	(0.9999)	(0.2235)	(0.2231)
# of instruments	-	1	4	4
J statistic	-	-	2.6437	17.0002

Table 1: Estimates and standard error of linear equation for Log(Wage) using (1) OLS, (2) 2SLS with 1 instrument (3) 2SLS with 4 instruments, and (4) Efficient Two Step GMM. Sample Size $n = 2215$.

(f) Because of possible endogeneity problem, we expect OLS estimates are unreliable, but 2SLS and GMM estimates are not. Parameter estimates of 2SLS using $near4$ instrument is significantly different from the results of 2SLS using 4 instruments and GMM, which casts doubt on the use of $near4$ as a valid instrument.

From the first-stage reduced form F statistic (or R^2), weak instrument problem seems not severe. Thus standard asymptotic theory are more likely to be valid for estimation and construction of confidence interval of the structural parameter.

Even though J-test rejects the model specification in the over identified GMM setup, it is unclear what is wrong. It is possible that model is mispecificied or instruments are likely to be invalid. However, J-test itself is well-known for it's over-rejection property in many finite sample evidences.