Recitation 2: Various IV estimators and GMM

Seung-hun Lee

Columbia University
Introduction to Econometrics II Recitation

January 31st, 2022

Two stage least squares

2SLS teases out exogenous variation from X_2

- Suppose the structural equation and the first-stage regression is as follows.

$$
\begin{aligned}
y &= X\beta + e = X_1\beta_1 + X_2\beta_2 + e && \text{(Structural)} \\
X &= Z\Gamma + v && \text{(First Stage)}
\end{aligned}
$$

  where $Z \in \mathbb{R}^{n\times l}, \Gamma \in \mathbb{R}^{l\times k}$
- For X_2, write $X_2 = Z\pi_2 + v_2$
- Assume $E[z_i e_i] = 0$, $E[x_{1i}' e_i] = 0$, $E[z_i v_{2i}] = 0$ and $E[x_{2i} e_i] \neq 0$
- we can get

$$E[x_{2i} e_i] = E[z_i e_i]\pi_2 + E[v_{2i} e_i]$$

  - Endogeneity of the X_2 regressors come from the $E[v_{2i} e_i]$
  - X_2 is composed of the parts spanned by Z and the other that is orthogonal to Z.
  - We want to 'tease out' part of the X_2 variables that can be explained by Z and use a generated regressor from this process to derive the estimator of interest

But first, some useful tricks

...because you will use them a lot!

Projection matrix $P_Z = Z(Z'Z)^{-1}Z'$

- Symmetric: $P_Z' = (Z(Z'Z)^{-1}Z')' = Z(Z'Z)^{-1}Z' = P_Z$
- Idempotent: $P_Z^2 = Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z' = Z(Z'Z)^{-1}Z' = P_Z$

Residual matrix $M_Z = I - P_Z$

- Symmetric: $M_Z' = I' - (Z(Z'Z)^{-1}Z')' = I - Z(Z'Z)^{-1}Z' = I - P_Z = M_Z$
- Idempotent: $M_Z^2 = (I - P_Z)(I - P_Z) = I - P_Z - P_Z + P_Z'P_Z = I - P_Z = M_Z$

How to run 2SLS?

- First stage
  - Regress the first stage and obtain the first stage estimator $\widehat{\Gamma} = (Z'Z)^{-1}Z'X$.
  - Then, get the predicted value of X, denoted as $\widehat{X} = Z(Z'Z)^{-1}Z'X = P_Z X$.
  - By doing so, we provide a way to map the space of l-dimensional vectors to k-dimensional vectors. (so $l = k$ assumption can be relaxed here)
- Second stage
  - In the structural equation, replace X with $\widehat{X}$ and obtain

$$
\begin{aligned}
\hat{\beta}_{2SLS} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'y = (X'P_Z'P_Z X)^{-1}(X'P_Z'y) \\
&= (X'P_Z X)^{-1}X'P_Z y = (\widehat{X}'X)^{-1}\widehat{X}'y
\end{aligned}
$$

- Watch out for using the correct form of standard errors (a problem set question)

Assumptions for proving asymptotics of 2SLS estimators

2SLS assumptions

- T1 (y_i, x_i, z_i) are IID
- T2 Finite second moments: $E||y_i^2|| < \infty$, $E||x_i^2|| < \infty$, $E||z_i^2|| < \infty$
- T3 $E(z_i z_i') > 0$
- T4 $rank[E(z_i x_i')] = k$
- T5 $E(z_i e_i) = 0$
- T6 Finite fourth moments: $E||y_i^4|| < \infty$, $E||x_i^4|| < \infty$, $E||z_i^4|| < \infty$
- T7 $E(z_i z_i' e_i^2) = \Omega > 0$

Consistency and asymptotic normality of IV estmators

Consistency and normality

- Under assumptions **T1-T5**, $\hat{\beta}_{2SLS} \xrightarrow{p} \beta$
- Under assumptions **T1-T7**, the limiting distribution of the 2SLS estimator is characterized by $\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, V_\beta)$
  - Note that

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) = \left[\frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'X}{n}\right]^{-1}\left[\frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'e}{\sqrt{n}}\right]$$

By CLT, $\frac{Z'e}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i e_i \xrightarrow{d} N(0, \Omega)$. Then Slutsky theorem and CMT,

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, \underbrace{A_n \Omega A_n'}_{V_\beta})$$

where $A_n = \left[\frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'X}{n}\right]^{-1}\left[\frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\right]$

Control function methods

Endogeneity comes from correlation of two error terms!

- Write the structural and reduced form regression as

$$y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + e_i$$
$$x_{2i} = \Gamma_{12}' x_{1i} + \Gamma_{22}' z_{2i} + u_{2i}$$

  where $E[z_i e_i] = 0$, $E(x_{1i} e_i) = 0$, $E(x_{2i} e_i) \neq 0$

- We can write $E(x_{2i} e_i)$ as

$$\begin{aligned} E[x_{2i} e_i] &= E[(\Gamma_{12}' x_{1i} + \Gamma_{22}' \beta_2 + u_{2i}) e_i] \\ &= \Gamma_{12}' E(x_{1i} e_i) + \Gamma_{22}' E(z_{2i} e_i) + E(u_{2i} e_i) \\ &= E(u_{2i} e_i) \ (\because \text{The first by exogeneity, the second by IV conditions}) \end{aligned}$$

- So correlation comes from $E(u_{2i} e_i)$

Idea: What part of e_i is u_{2i} vs something exogenous?

- consider a linear projection of e_i onto u_{2i}, which we write as

$$e_i = u_{2i}'a + \epsilon_i \; (E(u_{2i}\epsilon_i) = 0)$$

  where the population analogue of $a = E(u_{2i}u_{2i}')^{-1}E(u_{2i}e_i)$
- Substitute the e_i term in the structural equation with the above to get

$$y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + u_{2i}'a + \epsilon_i$$

- Then we can show

$$\begin{aligned} E[x_{2i}\epsilon_i] &= E[x_{2i}(e_i - u_{2i}'a)] = E[x_{2i}e_i] - E[x_{2i}u_{2i}']a \\ &= E[u_{2i}e_i] - \Gamma_{12}'E[x_{1i}u_{2i}']a - \Gamma_{22}'E[z_{2i}u_{2i}']a - E[u_{2i}u_{2i}']a \\ &= E[u_{2i}e_i] - 0 - 0 - E[u_{2i}e_i] = 0 \end{aligned}$$

#metricstotheface: How and when to implement them

- We usually don't know u_{2i}
- We work in these steps
  1. **Obtain** $\hat{u}_{2i}$: This is done by regressing (Reduced Form) equation.
  2. **Work with (CFA)**: However, instead of u_{2i}, use $\hat{u}_{2i}$. Then we can run an OLS
- Use Frisch-Waugh-Lovell to show this is equivalent to 2SLS
- When to use control function?
  - Endogeneity test: $H_0 : a = 0 vs. H_1 : \neg H_0$
  - (Wooldridge 2015): Study the self-selection to treatment, flexibly applicable in nonlinear setups, does not rely on assumptions of MLE being correct

Other topics in IV estmation

Generated regressors: First-stage sampling uncertainties carry over!

- Suppose we have a latent variable W, and $y = W\beta + e$
- We do know $W = Z\gamma + u$ and assume it is estimable in that $\widehat{W} = Z\hat{\gamma}$
- We can write main regression as

$$y = W\beta + e \iff y = \widehat{W}\beta + (e + (W - \widehat{W})\beta)$$

- So $\hat{\beta} - \beta = (\widehat{W}'\widehat{W})^{-1}\widehat{W}'(e + (W - \widehat{W})\beta)$
- Consistency is not a problem, but inference could be affected
  - Rewrite $\hat{\beta} - \beta = (\hat{\gamma}'Z'Z\hat{\gamma})^{-1}\hat{\gamma}'Z'[\underbrace{e - u\beta}_{\epsilon}]$
  - Then $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N(0, V_{\beta}^{*})$ where $V_{\beta}^{*} = \left(\gamma' E[Z'Z]\gamma\right)^{-1}\left(\gamma' E[Z'\epsilon\epsilon' Z]\gamma\right)\left(\gamma' E[Z'Z]\gamma\right)^{-1}$, which is inflated compared to true variance $\left(\gamma' E[Z'Z]\gamma\right)^{-1}\left(\gamma' E[Z'ee'Z]\gamma\right)\left(\gamma' E[Z'Z]\gamma\right)^{-1}$

Nonlinear IV: $(\widehat{X}_2)^2$ and $\widehat{X_2^2}$ is not the same!

- Consider this structural equation (supply and demand)

$$y_1 = \gamma_{12} y_2 + \gamma_{13} y_2^2 + \delta_{11} z_1 + u_1 \; (E[u_1|\mathbf{z}] = 0)$$
$$y_1 = \gamma_{12} y_2 + \delta_{22} z_2 + u_2 \; (E[u_2|\mathbf{z}] = 0)$$

- Let $y_3 = y_2^2$. Then the natural candadiates for an IV would be all of $z_1, z_2, z_1^2, z_2^2, z_1 z_2$
- Frequent mistake: Including $x_{1i}, \hat{x}_{2i}, (\hat{x}_{2i})^2$ in the second stage regression
  - This ignores the fact that the linear projection of the squared is not equal to the square of the linear projection, leading to an inconsistent estimation.
  - Consistent estimation must project the original and squared values separately (this is a problem set question!)

Correlated random coefficients: How to do 2SLS and control function

- Write this model as

$$y_i = x_{2i}\beta_i + e_i, \ (E[x_{2i}e_i] \neq 0, \beta_i = \beta + w_i, E[\beta_i] = \beta)$$

- This can be written as $y_i = x_{2i}\beta + e_i + x_{2i}w_i$
- Problem: Even if we assume that $E[e_i|z_i] = E[w_i|z_i] = 0$ IV estimator is not consistent if $E[x_{2i}w_i|z_i]$ depends on z_i
- 2SLS: Find z_i s.t. $cov(x_{2i}, w_i|z_i) = cov(x_{2i}, w_i)$
- Control function: If the model for x_2 is

$$x_{2i} = z_i\pi_2 + v_{2i} \ (E[v_{2i}|z_i] = 0, E[e_i|v_{2i}] = \rho_e v_{2i}, E[w_i|v_{2i}] = \rho_w v_{2i})$$

Then $E[y_i|x_{2i}, v_{2i}] = x_{2i}\beta + \rho_e v_{2i} + \rho_w v_{2i}x_{2i}$. So regress y_i on x_{2i}, $\hat{v}_{2i}$, and $\hat{v}_{2i}x_i$

Generalized method of moments

GMM: Extension of method of moments approach

- GMM methods utilize the method of moments estimators to identify the values of the parameters of interest
- It can be generalized in the sense that the number of moment conditions can be greater than the number of unknown parameters.
- Let w_i be IID across $i = 1, .., n$, $g_i(w_i, \beta)$ be a $l \times 1$ function of the ith observation, and $\beta \in \mathbb{R}^{k\times 1}$ be the parameter of interest. ($l \geq k$). Then, the **moment equation model** is characterized by
$$E[g(w_i, \beta)] = 0$$
- We say β is identified if there is a unique β satisfying $E[g(w_i, \beta)] = 0$
  - When $l = k$, then we are in a just-identified case
  - If $l > k$, then we are in the over-identified case
  - If $l < k$, we are in an under-identified case

GMM estimator

- Define $J(\beta)$ as
$$J(\beta) = n\bar{g}_n(\beta)' W \bar{g}_n(\beta)$$
where $W \in \mathbb{R}^{l \times l}$ is a positive definite weight matrix that is given.
  - n does not really affect our estimation, but it makes the analysis of the asymptotic features much easier
- The **generalized method of moments estimator** is defined as the minimizer of the GMM criterion above, or

$$\hat{\beta}_{GMM} = \arg\min_{\beta} J_n(\beta)$$
$$\implies \frac{\partial J_n(\beta)}{\partial \beta} = 2n \frac{\partial \bar{g}(\beta)'}{\partial \beta} W \bar{g}(\beta) = 0$$

- We can show that OLS, MLE, and IV are all part of GMM (check the note)

Limiting Distribution of GMM: Building block

- Given that $\hat{\beta}_{GMM} = (X'ZWZ'X)^{-1}(X'ZWZ'y)$ for overidentified IV model, we can rewrite this by replacing y with $X\beta + e$
- As a result, $\sqrt{n}(\hat{\beta}_{GMM} - \beta) = \left(\frac{X'Z}{n} W \frac{Z'X}{n}\right)^{-1} \left(\frac{X'Z}{n} W \frac{Z'e}{\sqrt{n}}\right)$

Assumptions

Assume that

1. $E(z_i x_i') = Q$, and that $\frac{Z'X}{n} \xrightarrow{p} Q$
2. $\frac{Z'e}{\sqrt{n}} \xrightarrow{d} N(0, \Omega)$, where $\Omega = E(z_i z_i' e_i^2)$
3. (If we are willing to assume W depends on n, thus W_n): $W_n \xrightarrow{p} W$, where W is a positive definite weight matrix

Limiting Distribution of GMM

- If the above assumptions are satisfied, the limiting distribution of the GMM estimator can be characterized by

$$\sqrt{n}(\hat{\beta}_{GMM} - \beta) \xrightarrow{d} N(0, (Q'WQ)^{-1}(Q'W\Omega W'Q)(Q'WQ)^{-1})$$

- Even if we suppose that W depends on n somehow, the above theorem still holds, provided that W_n converges in probability to W
- Question: What is the best selection for W?

Efficient GMM uses $W = \Omega^{-1}$

- To select an optimal W matrix, the resulting variance should be the smallest.
- If we let $W = \Omega^{-1}$ and work with $(Q'WQ)^{-1}(Q'W\Omega W'Q)(Q'WQ)^{-1} - (Q'\Omega Q)^{-1}$, we can see that it is positive semidefinite
- When we recalculate the variance, we get that the efficient GMM has a limiting distribution characterized by

$$\sqrt{n}(\hat{\beta}_{GMM} - \beta) \xrightarrow{d} N(0, (Q'WQ)^{-1})$$

- Try using the approach suggested in the problem set

Implementing efficient GMM: Two-step Optimal GMM

- We have no idea what Ω truly is.
- Therefore, we require a consistent estimator, denoted as $\widehat{W}$, for $W = \Omega^{-1}$

Two-step Optimal GMM

We can compute Optimal Two-step GMM in these steps

1. Compute a preliminary, but consistent estimator for the true β. Denote this as $\tilde{\beta}$. In this step, you can use any W, say $(Z'Z)^{-1}$
2. Using $\Omega = E[g(w_i, \beta)g(w_i, \beta)']$, create a sample analogue of this, defined as $\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} g(w_i.\tilde{\beta})g(w_i.\tilde{\beta})'$. We can find our $\widehat{\Omega}^{-1}$ here.
3. Using this $\widehat{\Omega}^{-1}$, construct an efficient GMM estimator $\hat{\beta}_{GMM}$

Alternative to the $\widehat{\Omega}$ that is (slightly) better

- There is another way to come up with a $\widehat{\Omega}^{-1}$ in this context.
- Define $\bar{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g(w_i, \tilde{\beta})$.
- Then, an alternative definition of $\widehat{\Omega}$ can be written as

$$\widehat{\Omega}^{+} = \frac{1}{n}\sum_{i=1}^{n}(g(w_i, \tilde{\beta}) - \bar{g}(\beta))(g(w_i, \tilde{\beta}) - \bar{g}(\beta))'$$

- Both $\widehat{\Omega}$ and $\widehat{\Omega}^{+}$ converge in probability to $E[g(w_i, \beta)g(w_i, \beta)']$
- However, if $E[g(w_i, \beta)] \neq 0$, we view $\widehat{\Omega}^{+}$ as a robust estimator. $\widehat{\Omega}$ is inconsistent in case where $E[g(w_i, \tilde{\beta})] = 0$ is not guaranteed.
- For tests such as overidentification tests, it is much more desirable to use $\widehat{\Omega}^{+}$

Estimator for the asymptotic variance?

- Since we know how to find the optimal $\widehat{W}$, we can estimate the asymptotic variance of the GMM estimators
- This can be done by replacing matrices in the original variance with their sample counterparts. In general, we can estimate by

$$\widehat{V}_{GMM} = \left(\widehat{Q}'\widehat{W}\widehat{Q}\right)^{-1}\left(\widehat{Q}'\widehat{W}\widehat{\Omega}\widehat{W}\widehat{Q}\right)\left(\widehat{Q}'\widehat{W}\widehat{Q}\right)^{-1}$$

  where $\widehat{Q} = \frac{1}{n}\sum_{i=1}^{n} z_i x_i' = \frac{Z'X}{n}$, $\widehat{W}$ is expressed by either $\widehat{\Omega}$ or $\widehat{\Omega}^{+}$.
- The residuals used in this estimation is defined as $\hat{e}_i = y_i - x_i'\hat{\beta}_{GMM}$