Introduction to Econometrics 2: Recitation 3

Seung-hun Lee

Columbia University

February 5th, 2020

Limited Information Maximum Likelihood (LIML)

- Assume a data generating process

$$y_i = \beta_1' x_{1i} + \beta_2' x_{2i} + e_i, \ (x_{1i} \in \mathbb{R}^{k_1}, x_{1i} \in \mathbb{R}^{k_2})$$

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

- A **limited information maximum likelihood (LIML) estimator** derives the maximum likelihood estimator for the joint distribution of (y_i, x_{2i}) using structural equation of y_i and the reduced form equation for x_{2i}.
  - Full information maximum likelihood (FIML) requires structural equation for x_{2i} as well.

Instrumental Variables

So why do we want to use them?

- When the number of the instruments are fixed, 2SLS and LIML have the same asymptotic distribution
- When there is a problem of weak instrument variable or too many instrumental variables, it can be shown that 2SLS becomes biased towards OLS
    - Pischke's lecture slides (link to it in the recitation notes and here: `http://econ.lse.ac.uk/staff/spischke/ec533/Weak%20IV.pdf`) show that LIML performs better, whereas 2SLS bias approaches that of OLS in presence of weak IV or too many IV

Instrumental Variables

Derivation

- Assume that there is a $z_i = \begin{pmatrix} x_{1i} \\ z_{2i} \end{pmatrix} \in \mathbb{R}^l$ where $E(z_i e_i) = 0$.
- The reduced for for x_{2i} can be shown as

  $x_{2i} = \Gamma'_{12} x_{1i} + \Gamma'_{22} z_{2i} + u_{2i}$, $(\Gamma'_{12} \in \mathbb{R}^{k_2 \times k_1}, \Gamma'_{22} \in \mathbb{R}^{k_2 \times (l-k_1)}, z_{2i} \in \mathbb{R}^{l-k_1})$
- By putting the structural and reduced form equation together in a matrix form, we get

$$\underbrace{\begin{pmatrix} 1 & -\beta'_2 \\ 0 & I_{k_2} \end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} y_i \\ x_{2i} \end{pmatrix}}_{=w_i} = \underbrace{\begin{pmatrix} \beta'_1 & 0 \\ \Gamma'_{12} & \Gamma'_{22} \end{pmatrix}}_{=B} \underbrace{\begin{pmatrix} x_{1i} \\ z_{2i} \end{pmatrix}}_{z_i} + \underbrace{\begin{pmatrix} e_i \\ u_{2i} \end{pmatrix}}_{=\eta_i}$$

- To use the maximum likelihood approach, we need some assumptions on η_i. Specifically, we assume that η_i is conditionally IID normal and

$$\eta_i | z_i \sim N(0, \Sigma_\eta)$$

  If we apply these facts, then we can derive the log-likelihood function for η_i

Instrumental Variables

Step-by-step: 1.**Re-express the PDF**

- The PDF pf η_i is

$$(2\pi)^{-\frac{k_2+1}{2}} \det(\Sigma)^{-\frac{1}{2}} \times e^{-\frac{1}{2}(\eta_i)'\Sigma_\eta^{-1}(\eta_i)}$$

- Also, since $w_i|z_i \sim N(A^{-1}Bz_i, A^{-1}\Sigma_\eta A^{-1})$, note that $\chi_i = w_i - A^{-1}Bz_i$ is also another normal distribution. (conditional on z_i), we can write the pdf for χ_i
- Lastly, notice that $\eta_i = A\chi_i$, Then we can apply Jacobian transformation to write a PDF for w_i as

$$(2\pi)^{-\frac{k_2+1}{2}} \det(\Sigma)^{-\frac{1}{2}} \times e^{-\frac{1}{2}(Aw_i - Bz_i)'\Sigma_\eta^{-1}(Aw_i - Bz_i)}|A|$$

- Since $\det(A) = 1$, we can omit $|A|$
- By IID assumption, the log likelihood for $i = 1, ..., n$ becomes

$$C - \frac{n}{2}\log(\det(\Sigma_\eta)) - \frac{1}{2}\sum_{i=1}^{n}(Aw_i - Bz_i)'\Sigma_\eta^{-1}(Aw_i - Bz_i)$$

Step-by-step: 2. **Concentrating out**:

- To find A, B that maximizes the log - likelihood, we concentrate out the Σ_η term by using its estimator,

$$\widehat{\Sigma}_\eta = \frac{1}{n}\sum_{i=1}^{n}(Aw_i - Bz_i)(Aw_i - Bz_i)'$$

- Then plug this estimator into the log-likelihood function to get

$$C - \frac{n}{2}\log(\det(\widehat{\Sigma}_\eta)) - \frac{1}{2}\sum_{i=1}^{n}(Aw_i - Bz_i)'\widehat{\Sigma}_\eta^{-1}(Aw_i - Bz_i)$$

Instrumental Variables

Step-by-step: 2. **Concentrating out**:

- Since the last term is a scalar, we can take

$$(Aw_i - Bz_i)'\widehat{\Sigma}_\eta^{-1}(Aw_i - Bz_i) = tr\left((Aw_i - Bz_i)'\widehat{\Sigma}_\eta^{-1}(Aw_i - Bz_i)\right)$$

- Also, using the fact that $tr(AB) = tr(BA)$,

$$tr\left((Aw_i - Bz_i)'\widehat{\Sigma}_\eta^{-1}(Aw_i - Bz_i)\right) = tr\left(\widehat{\Sigma}_\eta^{-1}(Aw_i - Bz_i)(Aw_i - Bz_i)'\right)$$

- This reduced the likelihood function to

$$\begin{aligned} & C - \frac{n}{2}\log(\det(\widehat{\Sigma}_\eta)) - \frac{1}{2}tr\left(\widehat{\Sigma}_\eta^{-1}\sum_{i=1}^{n}(Aw_i - Bz_i)(Aw_i - Bz_i)'\right) \\ & = C - \frac{n}{2}\log(\det(\widehat{\Sigma}_\eta)) - \frac{n(k_2+1)}{2} \end{aligned}$$

Instrumental Variables

Step-by-step: 3. **Maximize**

- The only thing left to maximize over now is $-\frac{n}{2}\log(\det(\widehat{\Sigma}_\eta))$. This is equivalent to maximizing

$$-\det\left(\frac{1}{n}\sum_{i=1}^{n}(Aw_i - Bz_i)'(Aw_i - Bz_i)\right)$$
$$= -\det\left(\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} y_i - \beta_1' x_{1i} - \beta_2 x_{2i} \\ x_{2i} - \Gamma_{12}' x_{1i} - \Gamma_{22}' z_{2i}\end{pmatrix}\begin{pmatrix} y_i - \beta_1' x_{1i} - \beta_2 x_{2i} \\ x_{2i} - \Gamma_{12}' x_{1i} - \Gamma_{22}' z_{2i}\end{pmatrix}'\right)$$

  The combination of β and Γ that maximizes this is the LIML estimator.

Instrumental Variables

k-class Estimators

- Another way to compute the LIML estimator is to use a k-class estimator with a particular choice for k. In general, k-class estimator is defined as

$$\hat{\beta}_k = \arg\min_{\beta} (y - X\beta)'(I_n - kM_Z)(y - X\beta)$$

  where $M_Z = I - Z(Z'Z)^{-1}Z' = I - P_Z$.

- The other way to express this, after some matrix differentiation, is

$$\begin{aligned} -2X'y + 2kX'M_Z y + 2(X'X)\beta - 2k(X'M_Z X)\beta = 0 \\ \iff (X'(I_n - kM_Z)X)\beta = X'(I_n - kM_Z)y \\ \implies \hat{\beta}_k = (X'(I_n - kM_Z)X)^{-1}(X'(I_n - kM_Z)y) \end{aligned}$$

k-class Estimators

- In fact, we can show that OLS ($k = 0$) and 2SLS ($k = 1$) are k-class estimators.
- LIML is a k-class estimator with a parameter choice of some $k > 1$. What we need to do find the associated value of k
- To do so, define

$$W = (y \quad X_2) \in \mathbb{R}^{n\times(k_2+1)}, \; M_1 = I - X_1(X_1'X_1)^{-1}X_1'$$

Then we compute the minimum eigenvalue of

$$(W'M_1W)(W'M_ZW)^{-1}$$

k-class Estimators

- This minimum eigenvalue will be our choice of k, which will be denoted as $\hat{k}$, and the LIML estimator would be

$$\hat{\beta}_{LIML} = (X'(I_n - \hat{k}M_Z)X)^{-1}(X'(I_n - \hat{k}M_Z)y)$$

- From here, we can also find that LIML is also a type of an IV estimator. We can see that by rewriting above equation as

$$\hat{\beta}_{LIML} = (\tilde{X}'X)^{-1}(\tilde{X}'y)$$

where $\tilde{X} = (I_n - \hat{k}M_Z)X$. This also hints that the asymptotic properties of LIML and some other types of IV estimators are similar.

Asymptotics of LIML

- Given $\hat{\beta}_{LIML} = (X'(I_n - \hat{k}M_Z)X)^{-1}(X'(I_n - \hat{k}M_Z)y)$

$$\begin{aligned}
\hat{\beta}_{LIML} - \beta &= (X'(I_n - \hat{k}M_Z)X)^{-1}(X'(I_n - \hat{k}M_Z)e) \\
&= (X'(P_Z - (\hat{k}-1))M_Z)X)^{-1}(X'(P_Z - (\hat{k}-1))e) \\
&(\because I - M_Z = P_Z, \implies I - \hat{k}M_Z = P_Z - (\hat{k}-1)M_Z) \\
\sqrt{n}(\hat{\beta}_{LIML} - \beta) &= \left(\frac{X'P_ZX}{n} - (\hat{k}-1)\frac{X'M_ZX}{n}\right)^{-1}\left(\frac{X'P_ZX}{\sqrt{n}} - (\hat{k}-1)\frac{X'M_Ze}{\sqrt{n}}\right)
\end{aligned}$$

Instrumental Variables

Asymptotics of LIML

- Note that
  - Anderson and Rubin (1949) shows that $\hat{k} - 1 \xrightarrow{p} 0$, which we accept as given.
  - $\frac{X'M_Z X}{n} = \frac{X'X}{n} - \frac{X'P_Z X}{n} = \frac{X'X}{n} - \frac{X'P_Z' P_Z X}{n} = \frac{X'X}{n} - \frac{(P_Z X)'(P_Z X)}{n} \leq \frac{X'X}{n} \xrightarrow{p} E(x_i x_i')$. Therefore, $\frac{X'M_Z X}{n}$ is bounded (and thus $O_p(1)$)
  - $\frac{X'M_Z e}{\sqrt{n}}$ converges in distribution to a normal distribution (CLT), so it is $O_p(1)$
  - From the third and first points, we can infer that $(\hat{k} - 1)\frac{X'M_Z e}{\sqrt{n}} = o_p(1)$
- Combining these leads to the result that

$$\sqrt{n}(\hat{\beta}_{LIML} - \beta) = \left(\frac{X'P_Z X}{n}\right)^{-1} \frac{X'P_Z X}{\sqrt{n}} + o_p(1)$$

which is equivalent to $\sqrt{n}(\hat{\beta}_{2SLS} - \beta) + o_p(1)$

Instrumental Variables

Control Function Method

- This is another way to derive a 2SLS estimator.
- will assume $E(x_{1i}e_i) = 0, E(x_{2i}e_i) \neq 0$ and write the structural and reduced form regression as

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

- We have a $z_i \in (x_{1i} \; z_{2i})' \in \mathbb{R}^l$ that satisfies $E(z_ie_i) = 0$
- The key driving idea for the control function method is that we can write $E(x_{2i}e_i)$ differently. Specifically,

$$\begin{aligned} E[x_{2i}e_i] &= E[(\Gamma_{12}'x_{1i} + \Gamma_{22}'\beta_2 + u_{2i})e_i] \\ &= \Gamma_{12}'E(x_{1i}e_i) + \Gamma_{22}'E(z_{2i}e_i) + E(u_{2i}e_i) \\ &= E(u_{2i}e_i) \implies \therefore E[x_{2i}e_i] \neq 0 \iff E[u_{2i}e_i] \neq 0 \end{aligned}$$

Instrumental Variables

Control Function Method

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

$$e_i = u_{2i}'\alpha + \epsilon_i \tag{LP}$$

  where $E(u_{2i}\epsilon_i) = 0$ and the population analogue of $\alpha = E(u_{2i}u_{2i}')^{-1}E(u_{2i}e_i)$

- Substitute the e_i term in the (Structural) equation with the (LP) equation to get

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

  where the following are satisfied

$$E[x_{1i}\epsilon_i] = E[x_{2i}\epsilon_i] = E[u_{2i}\epsilon_i] = 0$$

  Why? Separate slide!

Instrumental Variables

Control Function Method

- So the key takeaway is that now x_{2i} is exogenous with ϵ_i. In words, x_{2i} is correlated with e_i through u_{2i} and ϵ_i is the error after e_i has been projected onto u_{2i}.
- Usual 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 on the REWRITTEN (CFA) equation.
- Once this is done, we can show that the estimates from control function approach is numerically identical to 2SLS

Control Function Method

- One thing to note is that with this setup, we can conduct a test of endogeneity. Formally we want to test

$$H_0 : E(x_{2i}e_i) = 0,\ H_1 : E(x_{2i}e_i) \neq 0$$

- If it is the case that $E(x_{2i}e_i) = 0$, then $E(u_{2i}e_i) = 0$. Then, by how we constructed α, this implies that $\alpha = 0$. We can then show that the Wald statistics, under H_0, is distributed as

$$\hat{\alpha}'(var(\hat{\alpha}))^{-1}\hat{\alpha} \xrightarrow{d} \chi^2_{k_2}$$

- To test the null against the alternative hypothesis, define $C_{1-\alpha}$ as $\Pr(\chi^2_{k_2} \leq C_{1-\alpha}) = 1 - \alpha$. Then, we can reject the H_0 hypothesis if $\hat{\alpha}'(var(\hat{\alpha}))^{-1}\hat{\alpha} > C_{1-\alpha}$

Topics on Instrumental Variables

Hausmann Tests

- Assume the following data generating process

$$y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + e_i$$

  and we are interested in checking the exogeneity / endogeneity of x_{2i}.
- So we test $H_0 : E(x_{2i}e_i) = 0$ against $H_1 : E(x_{2i}e_i) \neq 0$. Consider these properties of 2SLS and OLS estimators
  - $\hat{\beta}_{OLS}$: Consistent and minimal variance under H_0, inconsistent under H_1
  - $\hat{\beta}_{2SLS}$: Consistent in either H_0 or H_1. Inefficient under H_0.
- Under H_0, $\sqrt{n}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})$ converges in distribution to $N(0, var(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}))$
- Hausmann also shows that in H_0, $var(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}) = var(\hat{\beta}_{2SLS}) - var(\hat{\beta}_{OLS})$ holds in general

Hausmann Tests

- Given this, we can write the Hausmann test statistic as

$$(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})'(var(\hat{\beta}_{2SLS}) - var(\hat{\beta}_{OLS}))^{-}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}) \xrightarrow{d} \chi^2_{k_2}$$

  where $(var(\hat{\beta}_{2SLS}) - var(\hat{\beta}_{OLS}))^{-}$ indicates a generalized inverse.

- We can use the test statistic, construct a critical value and reject/not reject H_0 based on comparison.
- Note that generalized inverse is required since $(var(\hat{\beta}_{2SLS}) - var(\hat{\beta}_{OLS}))$ is usually not a full (column) rank matrix. I leave further explanation to my recitation notes.

Subset Endogeneity Tests

- We break down x_{2i} into two parts - one that is 'potentially' endogenous (x_{2i}) and one that is endogenous for sure (x_{3i})
- Then, we again test for $H_0 : E(x_{2i}e_i) = 0$ against $H_1 : E(x_{2i}e_i) \neq 0$
- Using a control function approach, with following setup

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + x'_{3i}\beta_3 + e_i \qquad \text{(Structural)}$$
$$x_{2i} = \Gamma'_2 z_i + u_{2i} \qquad \text{(Reduced Form 2)}$$
$$x_{3i} = \Gamma'_3 z_i + u_{3i} \qquad \text{(Reduced Form 3)}$$

- We then project e_i onto u_{2i}, u_{3i} to obtain

$$e_i = u'_{2i}\alpha_2 + u'_{3i}\alpha_3 + \epsilon_i \qquad \text{(LP2)}$$

Subset Endogeneity Tests

- We can show that the following is satisfied

$$\begin{pmatrix} E(u_{2i}u_{2i}') & E(u_{2i}u_{3i}') \\ E(u_{3i}u_{2i}') & E(u_{3i}u_{3i}') \end{pmatrix} \begin{pmatrix} \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} E(u_{2i}e_i) \\ E(u_{3i}e_i) \end{pmatrix}$$

- Control function approach would require us to apply an OLS to

$$y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + x_{3i}'\beta_3 + u_{2i}'\alpha_2 + u_{3i}'\alpha_3 + \epsilon_i \tag{CFA2}$$

- As before, $E(x_{2i}e_i) = 0$ implies $E(u_{2i}e_i) = 0$. This allows us to replace the H_0 with

$$H_0 : E[u_{2i}u_{2i}']\alpha_2 + E[u_{2i}u_{3i}']\alpha_3 = 0$$

- This is a special case of $R\beta = c$ type of hypothesis test. However, since we cannot usually know what the true value of u_{2i}, u_{3i} are, we use the residuals, and replace R with $\widehat{R}$ constructed from the sample analogue using residuals.

Overidentification Test

- Consider the following setup

$$y_i = x_i'\beta + e_i,\ (\dim(x_i) = k < \dim(z_i) = l)$$

- We can test for the validity of the instruments by testing $H_0 : E(z_i e_i) = 0$ vs. $H_1 : E(z_i e_i) \neq 0$.
- To do this, we construct a linear projection equation by projecting e_i onto z_i, obtaining

$$e_i = z_i'\alpha + \epsilon_i$$

  Thus, $\alpha = E(z_i z_i')^{-1} E(z_i e_i)$
- Therefore, $H_0 : \alpha = 0$ vs. $H_1 : \alpha \neq 0$

Overidentification Test

- This leaves us with the challenge of estimating α
  1. **Obtain** $\hat{e}_i$: This can be done using 2SLS estimates of β. As a result,
  $$\hat{e}_i = y_i - x_i'\hat{\beta}_{2SLS} \implies \hat{e} = y - X\hat{\beta}_{2SLS}$$
  2. **Obtain** $\hat{\alpha}$: Replace e with $\hat{e}$ to get
  $$\hat{\alpha} = (Z'Z)^{-1}Z'\hat{e}$$
  3. **Sargan Test**: We will assume homoskedasticity (otherwise, the test statistic does not converge to χ^2 distribution). Then we make use of the following test statistic
  $$S = \hat{\alpha}'(var(\hat{\alpha}))^{-}\hat{\alpha} = \frac{\hat{e}'Z(Z'Z)^{-1}Z'\hat{e}}{\hat{\sigma}^2}$$
  where $\hat{\sigma}^2$ can be obtained from two paths - one using $\frac{1}{n}\hat{e}'\hat{e}$ and the other using $\frac{1}{n}\hat{\epsilon}'\hat{\epsilon} = \frac{1}{n}(\hat{e} - Z\hat{\alpha})'(\hat{e} - Z\hat{\alpha})$. While they are slightly different in that the first estimate has larger variances, they are asymptotically equal. Under the null, $S \xrightarrow{d} \chi^2_{l-k}$

Weak IV Test

- Assume that the structural and reduced form equations are (we are working with a scalar regressors)

$$y_i = x_i\beta + e_i$$
$$x_i = z_i\gamma + u_i$$

- We say that there is a problem of weak instrument if $\gamma \simeq 0$.
- This can cause the 2SLS estimates to be biased and the distribution to be affected.
- Let $\gamma = \frac{\mu}{\sqrt{n}}$. For simplicity, I will assume
  - $var\left(\begin{pmatrix} e_i \\ u_i \end{pmatrix} |z_i\right) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \Sigma$
  - $E(z_i^2) = 1$
  - $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} z_i e_i \\ z_i u_i \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = N(0, \Sigma)$

Weak IV Test

- Now I will show that neither OLS nor IV estimator is not consistent.
  - OLS: Note that $\hat{\beta}_{OLS} - \beta$ can be written as $\frac{n^{-1}\sum_{i=1}^n x_i e_i}{n^{-1}\sum_{i=1}^n x_i^2}$, equivalent to

$$\begin{aligned}\frac{n^{-1}\sum_{i=1}^n x_i e_i}{n^{-1}\sum_{i=1}^n x_i^2} &= \frac{n^{-1}\sum_{i=1}^n (z_i\gamma + u_i)e_i}{n^{-1}\sum_{i=1}^n (z_i\gamma + u_i)^2} \\ &= \frac{n^{-1}\sum_{i=1}^n u_i e_i}{n^{-1}\sum_{i=1}^n u_i^2} + o_p(1) \\ &\xrightarrow{p} E(u_i e_i)/E(u_i^2) = \rho\end{aligned}$$

  - IV: For $\hat{\beta}_{IV} - \beta$, we can write $\frac{n^{-1/2}\sum_{i=1}^n z_i e_i}{n^{-1/2}\sum_{i=1}^n z_i x_i}$ or equivalently

$$\begin{aligned}\hat{\beta}_{IV} - \beta &= \frac{n^{-1/2}\sum_{i=1}^n z_i e_i}{n^{-1/2}\sum_{i=1}^n z_i(z_i\gamma + u_i)} \\ &= \frac{n^{-1/2}\sum_{i=1}^n z_i e_i}{n^{-1/2}\sum_{i=1}^n z_i u_i + n^{-1}\sum_{i=1}^n z_i^2 \mu} \xrightarrow{d} \frac{\xi_1}{\xi_2 + \mu}\end{aligned}$$

    which is not centered at 0, making it inconsistent.