Recitation 1: Logistics and IV estimation

Seung-hun Lee

Columbia University
Introduction to Econometrics II Recitation

January 24th, 2022

Logistics of the recitation

About me

- Name: Seung-hun Lee (이승헌)
- 4th-year student in economics PhD program
- Research: Development, Labor, Political economy (all of these involve applied econometrics and treatment effects taught in this part of Econometrics sequence)
- Feel free to reach out to me for things related to this class and more!

How recitations will be organized

- Location: 520 Mathematics Building
- Time: Mondays 11:00AM-12:30PM
- Note that I intend to run the recitation for 80 - 90 minutes, depending on the rigor of the material.
- Depending on the room availability, I will stay an extra 10-20 minutes to answer your questions about problem sets, concepts covered, and etc.

My approach to recitations

- Focus on reviewing the concepts covered on the Tuesday and Thursday regular classes in the previous week.
- In particular, I will attempt to give you an intuition on what various econometric methods aim to achieve, discuss key results and proofs, and mention how such methods are applied in various literatures in Economics.
  (If there is demand, I am willing to incorporate different methods into the recitation.)
- Key is to establish how one method stands in relation to other methods (how does one complement the other?)
- I TA-ed the same class two years ago. The materials are expected to be similar. For those who want to take a look, go to my Github repository (click here).

What you should expect from me and the recitations

- Post recitation notes by 10:00PM on Sundays (slides will be posted after recitation)
- Suggested problem set solutions will be posted (about 3 days after deadline)
- Goal: Help you get through Econometrics sequence. This means helping you achieve high grades to avoid certs (for Economics Ph.D. students) or passing the course with a sufficient grade (other Ph.D. students).
- While I am the one teaching the course, it will not be complete without you. Do not hesitate to ask questions, make suggestions at any time. I am here to help you all in any way I can.

Office hour logistics

- Location: Zoom (Click here to join)
- Time: Mondays 7:30PM - 8:30PM
- If you can't make it on this time for whatever reasons, send me an email
- I tend to use office hours to answer your questions on problem-solving (problem sets, past exams, etc) but you can bring in any questions.

Here are some references!

- Primary resources are the lecture notes of the professors and Hansen (2021)
- I also make use of Angrist and Pischke (2009), Arellano (2003), Baltagi (2005), Cameron and Trivedi (2005), Baltagi (1999), Hayashi (2000), Imbens and Rubin (2015), and Wooldridge (2010) for additional references.
- For Statistics, I usually refer to Casella and Berger(2002) and Hogg et al.(2014).
- For Linear Algebra, I rely on Gockenbach(2010), Lang(1987), and Strang(2009).
- From time to time, I may use papers published in various journals to show how the methods are applied in research. Those papers will be cited as I go by.

Instrumental variables

Setting up our notation

- Assume that data generating process (DGP) looks like

$$y_i = x_i'\beta + e_i, \ x_i = \begin{pmatrix} x_{i1} \\ \dots \\ x_{ik} \end{pmatrix}, \ i = 1, \dots, n$$

  where x_i and β are both in $\mathbb{R}^k$ and y_i and e_i are scalars.
- In a matrix notation, this can be written as

$$y = X\beta + e, \ y = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix} \in \mathbb{R}^n, X = \begin{pmatrix} x_1' \\ \dots \\ x_n' \end{pmatrix} \in \mathbb{R}^{n \times k}, e = \begin{pmatrix} e_1 \\ \dots \\ e_n \end{pmatrix} \in \mathbb{R}^n$$

- OLS estimator: $\hat{\beta} = \left(\sum_{i=1}^n x_i x_i'\right)^{-1} \left(\sum_{i=1}^n x_i y_i\right)$ or $(X'X)^{-1}(X'y)$

Following assumptions are useful in proving key properties

A1 (y_i, x_i) are IID across i's

A2 **Strict exogeneity**: $E(e_i|x_i) = 0$. This implies **orthogonality** ($E(x_i e_i) = 0$)

A3 **Identification**: $E(x_i x_i') = Q$ is a positive definite matrix (hereafter PD matrix)

A4 **Bounded moments**: $E||x_i^4|| < \infty$, $E||y_i^4|| < \infty$

A5 **Homoskedasticity**: Let $D = E(ee'|X) = \begin{pmatrix} E(e_1^2|X) & E(e_1 e_2|X) & \dots \\ E(e_2 e_1|X) & E(e_2^2|X) & \dots \\ \dots & \dots & \dots \\ \dots & \dots & E(e_n^2|X) \end{pmatrix}$.

Then, $E(e_i^2|X) = \sigma^2 \ \forall i$

A6 **No autocorrelation**: From the D matrix above, $E(e_i e_j|X) = 0 \ \forall i \neq j$

- Note: **A5-A6** collectively is referred to as **spherical error variance**

Small and large sample properties

- **Unbiasedness**: Under assumptions **A1-A2**, $E[\hat{\beta}|X] = \beta$
- **Variance of** $\hat{\beta}$: $Var[\hat{\beta}|X] = (X'X)^{-1}X'DX'(X'X)^{-1}$
- **Consistency**: Under assumptions **A1-A3**, $\hat{\beta} \xrightarrow{p} \beta$
- **Limiting distribution of** $\hat{\beta}$: Under assumptions **A1-A4**, the limiting distribution of $\hat{\beta}$ is characterized by $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, Q^{-1}\Omega Q^{-1})$, where $\Omega = E(x_i x_i' e_i^2)$

$\rightarrow$ I put the detailed proof in my notes

Generalized least squares: Our variances may not be spherical!

- We will write $D = \sigma^2\Sigma$ where Σ is an n-dimensional matrix where off-diagonals may not be zero and diagonal elements may vary.
- The matrix is still symmetric and positive definite, so there exists a $P \in \mathbb{R}^{n\times n}$ (not necessarily unique) satisfying
$$D^{-1} = PP'$$
- We can 'transform' our DGP
$$Y = X\beta + e \implies \underbrace{P'Y}_{\tilde{Y}} = \underbrace{P'X}_{\tilde{X}}\beta + \underbrace{P'e}_{\tilde{e}}$$

  You can see that the following two conditions hold (back to spherical variance!)
  - $E[\tilde{e}|\tilde{X}] = E[P'e|X] = P'E[e|X] = 0$
  - $E[\tilde{e}\tilde{e}'|\tilde{X}] = E[P'ee'P|X] = P'E[ee'|X]P = P'DP = I_n$

Generalized least squares: Estimation and properties

- GLS estimators can be obtained by minimizing a weighted sum squred of residuals

$$\hat{\beta}_{GLS} = \min_b (Y - Xb)' D^{-1} (Y - Xb) \implies \hat{\beta}_{GLS} = (X'D^{-1}X)^{-1}(X'D^{-1}Y)$$

- The unbiasedness and variances of GLS can be shown in a similar manner.

Variance of the GLS

$$\begin{aligned}
Var[\hat{\beta}_{GLS}|X] &= Var[\hat{\beta}_{GLS} - \beta|X](\because \beta \text{ is nonrandom}) \\
&= Var[(X'D^{-1}X)^{-1}X'D^{-1}e|X] \\
&= (X'D^{-1}X)^{-1}X'D^{-1}(E[ee'|X])D^{-1}X(X'D^{-1}X)^{-1} \\
&= (X'D^{-1}X)^{-1}X'D^{-1}DD^{-1}X(X'D^{-1}X)^{-1} = (X'D^{-1}X)^{-1}
\end{aligned}$$

- Feasible GLS? We were assuming that we knew what D looked like.
  - Not true in many cases (need stance on what the best estimate for D is)
  - D becomes a random variable and affects the distribution and the efficiency

If measurement error exist, OLS estimate has attenuation bias

- Suppose that the linear model we want to estimate is as follows

$$y_i = x_i^{*'}\beta + e_i \text{ (We assume } E(x_i^* e_i) = 0\text{)}$$

- However, we only observe $x_i = x_i^* + v_i$, where $E[v_i] = 0$, $E[x_i^* v_i] = 0$, $E[e_i v_i] = 0$
- We end up with

$$y_i = (x_i - v_i)'\beta + e_i \quad = x_i'\beta \underbrace{-v_i'\beta + e_i}_{=u_i}$$

- Then $E(x_i u_i)$ is as follows

$$E(x_i u_i) = E[x_i(-v_i'\beta + e_i)] = E[(x_i^* + v_i)(-v_i'\beta + e_i)] = -E(v_i v_i')\beta$$

So unless $\beta = 0$, or $E(v_i v_i') = 0$, $E(x_i u_i) \neq 0$

How does the attenuation bias look like?

Attenuation bias

With the above setup, the probability limit of the OLS estimator is

$$
\begin{aligned}
\hat{\beta}_{OLS} &= \beta + E(x_i x_i')^{-1} E(x_i u_i) \\
&= \beta - E(x_i x_i')^{-1} E(v_i v_i) \beta \\
&= \beta - E[(x_i^* + v_i)(x_i^* + v_i)']^{-1} E(v_i v_i) \beta \\
&= \beta - [E(x_i^* x_i^{*'}) + E(v_i v_i')]^{-1} E(v_i v_i) \beta \\
&= \frac{E(x_i^* x_i^{*'})}{E(x_i^* x_i^{*'}) + E(v_i v_i')} \beta \; (\leq \beta)
\end{aligned}
$$

OLS estimator is unestimable in simultaneous equations

- Suppose you have ($e_i = (u_i \ v_i)'$ is IID, $E(e_i) = 0$, $E(e_i e_i') = I_2$)

$$q_i = \beta_1 p_i + u_i \qquad \text{(Supply)}$$
$$q_i = -\beta_2 p_i + v_i \qquad \text{(Demand)}$$

- The equilibrium of this system is

$$p_i = \frac{v_i - u_i}{\beta_1 + \beta_2}, q_i = \frac{\beta_1 v_i + \beta_2 u_i}{\beta_1 + \beta_2}$$

So for both supply and demand equations, we have $E(p_i u_i) \neq 0$ and $E(p_i v_i) \neq 0$

- **Simultaneity bias**: When naively applying OLS to this equation, the result is as follows.

$$q_i = \beta^* p_i + \eta_i, \ E(p_i \eta_i) = 0 \implies \hat{\beta}^* = \frac{E(p_i q_i)}{E(p_i^2)} = \frac{\beta_1 - \beta_2}{2}$$

Workaround: Find exogenous shock affecting one but not the other

- Let z_i denote some exogenous shock to the demand equation (preference shock).
- We write the two equations as

$$q_i = \beta_1 p_i + u_i \qquad \text{(Supply)}$$
$$q_i = -\beta_2 p_i + \beta_3 z_i + v_i \qquad \text{(Demand)}$$

- Using a similar approach we employed, we can write

$$p_i = \frac{v_i - u_i}{\beta_1 + \beta_2} + \frac{\beta_3}{\beta_1 + \beta_2} z_i$$
$$q_i = \frac{\beta_1 v_i + \beta_2 u_i}{\beta_1 + \beta_2} + \frac{\beta_3 \beta_1}{\beta_1 + \beta_2} z_i$$

we have the endogenous variables in terms of exogenous variables (the reduced form).

Omitted variables may make OLS estimates inconsistent

- Let y: wage, x: education, a: innate ability and the DGP be

$$y_i = x_i\beta_1 + a_i\beta_2 + e_i, \quad E(x_i e_i) = 0, E(a_i e_i) = 0$$

- IRL, we can only observe (y_i, x_i) (what is innate ability anyway?), leaving us with

$$y_i = x_i\beta_1 + u_i, \text{ where } u_i = a_i\beta_2 + e_i$$

- Then $E(x_i u_i) = E(x_i(a_i\beta_2 + e_i)) = E(x_i a_i)\beta_2 + 0 = E(x_i a_i)\beta_2$
- Therefore, when 1)x_i and a_i are correlated and 2)$\beta_2 \neq 0$, $E[x_i u_i] \neq 0$
- The probability limit of OLS estimate of x_i would be

$$\hat{\beta}_{OLS} = \beta_1 + E(x_i^2)^{-1}E(x_i u_i) = \beta_1 + E(x_i^2)^{-1}E(x_i a_i)\beta_2$$

Setting up for the IV estimation

- Assume that the data generating process is as follows

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

where $E(x_{1i}e_i) = 0$, $E(x_{2i}e_i) \neq 0$, and $\dim(x_{1i}) = k_1, \dim(x_{2i}) = k_2,\ k_1 + k_2 = k$

- Let $z_i \in \mathbb{R}^l = \begin{pmatrix} z_{1i} \\ z_{2i} \end{pmatrix} = \begin{pmatrix} x_{1i} \\ z_{2i} \end{pmatrix}$, where $\dim(z_{2i}) = l - k_1$ and $l - k_1 \geq k_2$.
- z_i is a valid IV if
  - **Exogeneity**: $E(z_i e_i) = 0$
    - **Exclusion**: $E(z_i y_i) = \beta_1 E(z_i x_{1i}) + \beta_2 E(z_i x_{2i})$ (z_i should impact y_i through x_{1i} and x_{2i} only)
  - **Relevancy**: $rank[E(z_i x_i')] = \dim(x_i) = k$
  - **PD**: $E(z_i z_i') > 0$ (for inverse matrices in 2SLS to be defined)

IV in reduced form method

- We can write the reduced form relationship for the X_2 variables as

$$X_2 = X_1\pi_{21} + Z_2\pi_{22} + v_2 = Z\pi_2 + v_2$$

π_2 is a linear projection of X_2 onto Z $\left(E[Z'Z]^{-1}E[Z'X_2]\right)$ and we have $E[Z'v_2] = 0$

- With this, the reduced form for Y is

$$\begin{aligned}
Y &= X_1\beta_1 + X_2\beta_2 + e \\
&= X_1\beta_1 + (X_1\pi_{21} + Z_2\pi_{22} + v_2)\beta_2 + e \\
&= X_1(\beta_1 + \pi_{21}\beta_2) + Z_2\pi_{22}\beta_2 + e + v_2\beta_2 \\
&= Z_1\pi_{11} + Z_2\pi_{12} + v_1 = Z\pi_1 + v_1
\end{aligned}$$

where $E[Z'v_1] = 0$ based on our IV conditions and reduced form setup for X_2.

Bridging the structural and reduced forms

- From the above setup, we can also write

$$\pi_1 = \begin{pmatrix} \pi_{11} \\ \pi_{12} \end{pmatrix} = \underbrace{\begin{pmatrix} I_{k_1} & \pi_{21} \\ 0 & \pi_{22} \end{pmatrix}}_{=\bar{\Gamma}} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{=\beta}$$

  This is the exact relation between the *l* reduced form and *k* structural parameters.
- If $rank(\bar{\Gamma}) = k$ ($\pi_{22} \neq 0$), we can solve for β using the least squares

$$\beta = (\bar{\Gamma}'\bar{\Gamma})^{-1}\bar{\Gamma}'\pi_1$$

- In practice, IV estimators in this context can be obtained by the indirect least squares

In $l = k$, using moment condition is straightforward

- The structural equation and the moment conditions we will use are

$$y_i = x_i'\beta + e_i \; (E[z_i e_i] = 0, E[x_i e_i] \neq 0)$$

- We replace e_i in the moment condition using the structural equation and get

$$E[z_i e_i] = 0 \iff E[z_i(y_i - x_i'\beta)] = 0 \iff E[z_i y_i] - E[z_i x_i'\beta] = 0$$

- This gives us the result that (relevancy and square matrix!)

$$\beta = \left(E[z_i x_i']\right)^{-1} E[z_i y_i]$$

- The IV estimator is a sample analogue of the above, or

$$\hat{\beta}_{IV} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i y_i = (Z'X)^{-1}Z'y$$

IV estimators are consistent!

Consistency

Assume that Z is a valid IV

$$\begin{aligned}
(Z'X)^{-1}Z'y &= (Z'X)^{-1}Z'(X\beta + e) \\
&= \beta + (Z'X)^{-1}Z'e \\
&= \beta + \underbrace{\left(\frac{Z'X}{n}\right)^{-1}}_{\xrightarrow{p} Q_{ZX}^{-1}} \underbrace{\left(\frac{Z'e}{n}\right)}_{\xrightarrow{p} 0}
\end{aligned}$$

Thus, $\hat{\beta}_{IV} \xrightarrow{p} \beta$

IV estimators are asymptotically normal!

Asymptotic distribution: $\sqrt{n}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} N(0, Q_{ZX}^{-1} \Omega Q_{ZX}^{-1})$ with $\Omega = E[z_i z_i' e_i^2]$

Note that

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) = \left(\frac{1}{n} \sum_{i=1}^{n} z_i x_i' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i e_i \right)$$

- We can obtain $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i e_i \xrightarrow{d} N(0, \Omega)$ using the central limit theorem.
- $\frac{1}{n} \sum_{i=1}^{n} z_i x_i' \xrightarrow{p} Q_{ZX}^{-1}$ by weak law of large numbers.
- By Slutsky theorem $\sqrt{n}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} Q_{zx}^{-1} N(0, \Omega) = N(0, Q_{ZX}^{-1} \Omega Q_{ZX}^{-1})$
- Under homoskedasticity, we can get that $\Omega = \sigma^2 Q_{ZZ}$ and that the finite sample variance can be characterized as

$$\frac{1}{n} \widehat{V}_{\hat{\beta}_{IV}} = \hat{\sigma}^2 (Z'X)^{-1} (Z'Z)(Z'X)^{-1}$$