Introduction to Econometrics 2: Recitation 2

Seung-hun Lee

Columbia University

January 29th, 2020

Classical Linear Models

Ordinary Least Squares

- Assume a data generating process

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

- To show consistency and limiting distributions, we need this set of assumptions

Assumption

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

A2 $E(x_i e_i) = 0$

A2' $E(e_i|x_i) = 0$

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

A4 $E||x_i^4|| < \infty, E||y_i^4|| < \infty$

Classical Linear Models

Consistency and Limiting Distributions

Theorem

- **Consistency**: Under assumptions **A1-A3**, $\hat{\beta} \xrightarrow{p} \beta$
- **Limiting Distribution**: 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 $\Omega = E(x_i x_i' e_i^2)$

Proof: On separate pages!

- When we are interested in the limiting distribution of a particular element of β, then we work with

$$\sqrt{n}(\hat{\beta}_j - \beta_j) \xrightarrow{d} N(0, V_{jj}) \text{ (V_{jj} is the (j,j)th element of V)}$$

where $V = Q^{-1}\Omega Q^{-1}$ and V_{jj} is the (j,j)th element of V

Classical Linear Models

Hypothesis tests

- **Single element**: Consider the following setting

$$H_0 : \beta_j = \beta_j^0, \quad H_1 : \beta_j \neq \beta_j^0$$

From the limiting distribution of β_j, we can show that the test statistic is distributed as a standard normal under H_0. It is characterized as

$$\frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\widehat{V}_{jj}/n}} \xrightarrow{d} N(0,1)$$

where $\widehat{V}$ is estimated version of the variance (more on that later).

Hypothesis tests

- **Multiple element**: We may be interested in the features of the linear combinations of the elements of β, or even multiple restrictions, Let $R \in \mathbb{R}^{k \times r}$ characterize such restrictions. Then we can write

$$H_0 : R'\beta = c, \quad H_1 : \neg H_0$$

Then from the limiting distribution of $\hat{\beta}$, we can apply Slutsky's theorem to get the necessary limiting distribution

$$\sqrt{n}(R'\hat{\beta} - R'\beta) = \sqrt{n}R'(\hat{\beta} - \beta) \xrightarrow{d} N(0, R'VR)$$

Since $R'\beta = c$ under H_0, we can obtain the following Wald test statistic. (For convenience, we also assume that V is known)

$$n(R'\hat{\beta} - c)'(R'VR)^{-1}(R'\hat{\beta} - c) \xrightarrow{d} \chi^2_r$$

Notes on $\widehat{V}$

- The definition of $\widehat{V}$ is $\widehat{V} \equiv \widehat{Q}^{-1}\widehat{\Omega}\widehat{Q}^{-1}$, where
  - $\widehat{Q}$ is a sample analogue of Q, written as $\frac{1}{n}\sum_{i=1}^{n} x_i x_i'$
  - $\widehat{\Omega}$ is a sample analogue of $E(x_i x_i' e_i^2)$, also with the consideration that the true value of e_i is replaced with the residual $\hat{e}_i$. Therefore, we can write $\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \hat{e}_i^2$

Instrumental Variables

Motivation and Sources of Endogeneity

- In showing consistency of $\hat{\beta}_{OLS}$, assumption **A2**: $E(x_i e_i) = 0$ was crucial
- However, if this assumption cannot be supported by the data for whatever reasons, OLS estimators may no longer be consistent.
- Sources of Endogeneity
  1. (Classical) Measurement Error
  2. Simultaneity Bias
  3. Omitted Variable Bias (OVB)

Measurement Error

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

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

  However, we cannot observe x_i^*.
- Instead, we can observe $x_i = x_i^* + v_i$, where v_i has mean zero and independent of both x_i^* and e_i.
- What we regress:

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

Instrumental Variables

Measurement Error

- Then, what happens to $E(x_i u_i)$?

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

- So unless $\beta = 0$ or $E(v_i v_i') = 0$, $E(x_i u_i) \neq 0$
  - $\beta = 0$ implies that x_i^* has no role in determining y_i to begin with
  - $E(v_i v_i') = 0$ implies that $var(v_i) = 0$, so that v_i has mean 0 and has point mass at 0 - no measurement error!
- The probability limit of the OLS estimator is no longer β.
  - It converges in probability to $\frac{E(x_i^* x_i^{*'})}{E(x_i^* x_i^{*'}) + E(v_i v_i')}\beta$. This is an **attenuation bias**.
  - Proof: separate page!

Some Comments on Measurement errors

- If there exists another noisy, but unbiased measure of x_i^*, namely $w_i = x_i^* + \delta_i$, we can use w_i to instrument for x_i. The condition is that η_i has mean zero and uncorrelated with $(x_i^*, e_i.v_i)$. Try verifying that this satisfies all IV conditions.
- If there is a measurement error in y_i, the only this it does is to change the component of e_i. Assuming all the old assumptions hold, this does not pose as much problem as having a measurement error in the regressor.

Instrumental Variables

Simultaneity Bias

- A classic example of this would be a supply and demand system type of setting:

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

I will assume $e_i = (u_i \ v_i)'$ is IID, $E(e_i) = 0, E(e_i e_i') = I_2$ When you do some algebra, 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 $E(p_i v_i) \neq 0$

Instrumental Variables

Simultaneity Bias

- When naively applying OLS to this equation, the result is as follows.

$$q_i = \beta^* p_i + \eta_i,\ E(p_i \eta_i) = 0$$

$$\Longrightarrow\ \hat{\beta}^* = \frac{E(p_i q_i)}{E(p_i^2)} = \frac{\beta_1 - \beta_2}{2}$$

Thus, OLS estimators does not converge to either one of β_1 or β_2, resulting in a **simultaneity bias**.

Omitted Variable Bias

- Suppose that we are interested in the determinant of wages (y_i). Also assume that education, x_i, and innate ability, a_i, determine wages in the following manner

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

- However, instead of observing (y_i, x_i, a_i), we can only observe (y_i, x_i). the best we can do at the moment is to estimate the following equation

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

Instrumental Variables

Omitted Variable Bias

- Then $E(x_i u_i)$ becomes

$$E(x_i u_i) = E(x_i(a_i\beta_2 + e_i)) = E(x_i a_i)\beta_2 + 0 = E(x_i a_i)\beta_2$$

- Therefore, when
  - x_i and a_i are correlated and
  - $\beta_2 \neq 0$,

  x_i is endogenous with respect to u_i.
- Moreover, the OLS estimator acquired here has a probability limit of

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

  So if both conditions occur, the above does not converge in probability to β_1.
- we can determine the direction of the bias by the sign of $E(x_i a_i)$ and β_2.

Instrumental Variables

IV Estimator

- Assume that the data generating process is as follows

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

  where $E(x_{1i}e_i) = 0, E(x_{2i}e_i) \neq 0$
- OLS estimators of β_2 and β_1 will not be consistent.
- Let $z_i \in \mathbb{R}^l = \begin{pmatrix} z_{1i} \\ z_{2i} \end{pmatrix} = \begin{pmatrix} x_{1i} \\ z_{2i} \end{pmatrix}$, where $\dim(z_{2i}) = l - k_1$.
- Valid IV should satisfy

IV Conditions

1. **Exogeneity**: $E(z_i e_i) = 0$
   1. **Exclusion**: $E(z_i y_i) = \beta_1 E(z_i x_{1i}) + \beta_2 E(z_i x_{2i})$, in other words, z_i should impact y_i through x_{1i} and x_{2i}
2. **Relevancy**: $rank[E(z_i x_i')] = \dim(x_i) = k$
3. **PD**: $E(z_i z_i') > 0$

Instrumental Variables

Reduced Form

- In this approach, we assume that z_i is a least squares projection. So we can write

$$
\begin{aligned}
x_i &= \Gamma' z_i + u_i \ (\Gamma \in \mathbb{R}^{l \times k}) \\
\implies z_i x_i' &= z_i z_i' \Gamma + z_i u_i' \\
\implies E(z_i x_i') &= E(z_i z_i')\Gamma + E(z_i u_i')
\end{aligned}
$$

- Since z_i is a least squares projection, $E(z_i u_i) = 0$
- As a result, $\Gamma = E(z_i z_i')^{-1} E(z_i x_i')$ and the estimator for Γ would be its sample analogue,

$$
\widehat{\Gamma} = \left(\frac{1}{n} \sum_{i=1}^{n} z_i z_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} z_i x_i' \right) = (Z'Z)^{-1}(Z'X)
$$

Instrumental Variables

Reduced Form

- Then we get to the structural equation

$$y_i = x_i'\beta + e_i \iff y_i = (z_i'\Gamma + u_i')\beta + e_i \iff y_i = z_i' \underbrace{\Gamma\beta}_{=\lambda} + \underbrace{u_i'\beta + e_i}_{=v_i}$$

- From the assumptions, we can show that $E(z_i v_i) = 0$. As such,

$$\lambda = E(z_i z_i')^{-1} E(z_i y_i)$$

with its sample analogue being

$$\hat{\lambda} = \left(\frac{1}{n} \sum_{i=1}^{n} z_i z_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} z_i y_i \right) = (Z'Z)^{-1} Z'y$$

Reduced Form

- In case where $k = l$, Z itself becomes invertible. Then we can show that $\beta = \Gamma^{-1}\lambda$ and thus

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

- If otherwise, We note that $\widehat{\Gamma}\beta + \text{error} = \hat{\lambda}$ and show that

$$\hat{\beta}_{IV} = (\widehat{\Gamma}'\widehat{\Gamma})^{-1}\widehat{\Gamma}'\hat{\lambda}$$

2SLS Estimator

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

$$y = X\beta + e \qquad \text{(Structural)}$$
$$X = Z\Gamma + u \qquad \text{(First Stage)}$$

  where $Z \in \mathbb{R}^{n\times l}, \Gamma \in \mathbb{R}^{l\times k}$
- We still maintain the least square projection assumption
- We proceed as follows
  1. Regress the first stage and obtain $\widehat{\Gamma} = (Z'Z)^{-1}Z'X$. Then the predicted value of X, denoted as $\widehat{X} = Z(Z'Z)^{-1}Z'X = P_Z X$.
  2. In the structural equation, replace X with $\widehat{X}$ and obtain

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

  which is effectively replacing Z in the previous approach with $\widehat{X}$.

Consistency and the Limiting Distribution of 2SLS Estimator

- To proceed, we need these assumptions

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 $E(z_i z_i' e_i^2) = \Omega > 0$

Instrumental Variables

Consistency and the Limiting Distribution of 2SLS Estimator

Theorem

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

$$V_\beta = (Q'_{ZX} Q_{ZZ}^{-1} Q_{ZX})^{-1} Q'_{ZX} Q_{ZZ}^{-1} \Omega Q_{ZZ}^{-1} Q_{ZX} (Q'_{ZX} Q_{ZZ}^{-1} Q_{ZX})^{-1}$$

Proof: On separate pages!

Remarks

- **Correct Standard Errors** When estimating $\Omega = E(z_i z_i' e_i^2)$, we are not aware of the true error. So we need to estimate this as well. Note that we need to use a proper residual. Namely, we must use

$$\hat{e}_i = y_i - x_i'\hat{\beta}_{2SLS}$$

This is correct, as $\hat{\beta}_{2SLS} \xrightarrow{p} \beta$. However, we frequently make a mistake of using

$$\tilde{e}_i = y_i - \hat{x}_i'\hat{\beta}_{2SLS}$$

Since $\hat{x}_i$ is not exactly x_i, this converges in probability to something else.

  - Try this on STATA

Instrumental Variables

Remarks

- **Nonlinear Extension**: If we are instead interested in the properties of $g(\hat{\beta}_{2SLS})$, where $g(\cdot)$ is not necessarily nonlinear, we can apply delta method here.
- **Too Many IV?**: In practice, when we have too many IV's (large dimension for z_i), it may be possible for the instruments to 'perfectly' predict x_i. In other words, $\hat{x}_i$ becomes nearly identical to x_i. This can become a problem because the endogeneity bias that plagued original structural equation may not be mitigated.
  One way of getting around this is to use a regression method that involves penalty mechanism - like LASSO, for instance. We will also formally treat this issue when we learn over-identification tests in the near future.