Introduction to Econometrics II: Recitation 1

Seung-hun Lee*

January 24th, 2022

1 Logistics

1.1 Recitation

- 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.
- I will focus on reviewing the concepts covered on the two classes before the recitation - so my Monday recitations cover materials from 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.
- As you will notice in this semester, new methods arise in an attempt to fix drawbacks of the previous methods. I will attempt to establish the relationship among econometric methods by pointing out how one method makes up for flaws in the previous methods and so on.
- 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).

*Contact me at sl4436@columbia.edu if you spot any errors or have suggestions on improving this note.

1.2 Office Hours

- Location: Zoom (Click here to join)
- Time: Mondays 7:30PM - 8:30PM (If you can't make it on this time, send me an email)

1.3 What you should expect from me and the recitations

- Post recitation notes by 10:00PM on Sundays and suggested problem set solutions
- 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.

1.4 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.

2 Classical Linear Models

2.1 Ordinary Least Squares

Throughout this lecture (and possibly beyond), we will assume a data generating process that looks like

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

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

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

To demonstrate the consistency and the limiting distribution of the OLS estimators, I will use some of these assumptions

Assumption 2.1 (Assumptions for Classical Linear Models)**.** *The following assumptions are used in showing consistency and the limiting distribution of OLS estimators*

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

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***

The OLS estimator can be found by minimizing the sum of squared errors. In other words

$$\hat{\beta} = \min_b \sum_{i=1}^{n} (y_i - x_i'b)^2 = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\sum_{i=1}^{n} x_i y_i\right)$$

or in matrix notation, $(X'X)^{-1}(X'y)$. The key finite sample properties are as follows

Theorem 2.1 (Unbiasedness of $\hat{\beta}$)**.** *Under assumptions **A1-A2**,* $E[\hat{\beta}|X] = \beta$

Proof. Rewrite $\hat{\beta} = (X'X)^{-1}(X'Y)$ as $\beta + (X'X)^{-1}(X'e)$. We then obtain

$$\begin{aligned}
E[\hat{\beta}|X] &= E[\beta + (X'X)^{-1}(X'e)|X] \\
&= \beta + (X'X)^{-1}X'E[e|X] \\
&= \beta + 0 = \beta \ (\because \textbf{A2})
\end{aligned}$$

Thus, $\hat{\beta}$ is unbiased □

Theorem 2.2 (Variance of $\hat{\beta}$)**.** $Var[\hat{\beta}|X] = (X'X)^{-1}X'DX'(X'X)^{-1}$

Proof. Rewrite $\hat{\beta} = (X'X)^{-1}(X'Y)$ as $\beta + (X'X)^{-1}(X'e)$. We proceed as follows

$$\begin{aligned}
Var[\hat{\beta}|X] &= Var[\hat{\beta} - \beta|X] (\because \beta \text{ is nonrandom}) \\
&= Var[(X'X)^{-1}X'e|X] \\
&= (X'X)^{-1}X'(Var[e|X])X'(X'X)^{-1} \\
&= (X'X)^{-1}X'(E[ee'|X])X'(X'X)^{-1} (\because \textbf{A2}) \\
&= (X'X)^{-1}X'DX'(X'X)^{-1}
\end{aligned}$$

The final form of the variance depends on the assumptions on D. If **A5-A6** is satisfied, then $D = \sigma^2 I_n$ and $Var[\hat{\beta}|X] = \sigma^2(X'X)^{-1}$. (We can also show that in this case, OLS estimator is the best, linear, and unbiased estimator (BLUE) □

The consistency and the limiting distribution of OLS estimators can be demonstrated as follows

Theorem 2.3 (Consistency of $\hat{\beta}$)**.** *Under assumptions **A1-A3**,* $\hat{\beta} \xrightarrow{p} \beta$

Proof. Rewrite $\hat{\beta}$ as $\beta + \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\sum_{i=1}^{n} x_i e_i\right)$. To carry out the asymptotic analysis on the summation terms, multiply $\frac{1}{n}$ to both. By **A1**, we can deduce that x_i and e_i are IID. Then, we can apply weak law of large numbers and continuous mapping theorem to show that

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{p} E\left(x_i x_i'\right)$$
$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \xrightarrow{p} E\left(x_i x_i'\right)^{-1} (\because CMT)$$
$$\frac{1}{n}\sum_{i=1}^{n} x_i e_i' \xrightarrow{p} E\left(x_i e_i'\right)$$

By assumptions **A2,A3**, $\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \xrightarrow{p} Q^{-1}$ and $\frac{1}{n}\sum_{i=1}^{n} x_i e_i' \xrightarrow{p} 0$. By Slutsky's theorem, $\left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\sum_{i=1}^{n} x_i e_i\right) \xrightarrow{p} 0$. Thus, $\hat{\beta} \xrightarrow{p} \beta$. □

Theorem 2.4 (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 $\Omega = E(x_i x_i' e_i^2)$*

Proof. We can write $\sqrt{n}(\hat{\beta}-\beta) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i e_i\right)$. We know from the previous theorem that $\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \xrightarrow{p} Q^{-1}$. So we need to work on $\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i e_i\right)$. From the central limit theorem, we can obtain the limiting distribution of $\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i e_i\right)$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i e_i \xrightarrow{d} N(0, \Omega)$$

since $E(x_i e_i) = 0$ by **A2** and $var(x_i e_i) = E(x_i e_i e_i x_i') - (E(x_i e_i))^2 = E(x_i x_i' e_i^2) = \Omega$ (In using CLT, we need assumption **A4** so that the variance-covariance matrix obtained from here is finite.) Then, by Slutsky's theorem,

$$\sqrt{n}(\hat{\beta}-\beta) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i e_i\right) \xrightarrow{d} Q^{-1}N(0,\Omega) = N(0, \underbrace{Q^{-1}\Omega Q^{-1}}_{V})$$

□

If we are interested in a particular element of β, namely β_j, we will need to work on the following limiting distribution

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

2.2 Generalized least squares

In reality, the assumption of spherical variance may not be satisfied. This leads to the result where the OLS no longer gives us the efficient estimator. For this section, 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 change our data generating process $Y = X\beta + e$ a bit by multiplying P' to both sides.

$$Y = X\beta + e \implies \underbrace{P'Y}_{\tilde{Y}} = \underbrace{P'X}_{\tilde{X}}\beta + \underbrace{P'e}_{\tilde{e}}$$

You can see that the following two conditions hold

- $E[\tilde{e}|\tilde{X}] = E[P'e|X] = P'E[e|X] = 0$
- $E[\tilde{e}\tilde{e}'|\tilde{X}] = E[P'ee'P|X] = P'E[ee'|X]P = P'DP = I_n$

so we end up getting a spherical error variance by transforming e into $\tilde{e}$.

In general, 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. For the variance,

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

However, we were assuming that we knew what D looked like. That is not true in many cases, forcing us to take a stand on what the best estimate for D is. In doing so, D becomes a random variable and affects the distribution and the efficiency of the estimator.

3 Instrumental Variables Method

We still work with the data generating process $y_i = x_i'\beta + e_i$, but now with the possibility that $E(x_i e_i) \neq 0$. In other words, the error term and the regressor can now be correlated (or the regressors are endogenous). Since we required **A2** assumption in showing that OLS estimators are consistent, the fact that $E(x_i e_i)$ is not necessarily zero implies that OLS may no longer be consistent. In this section, we study cases in which regressors can become endogenous and how instrumental variables allow us to address this problem.

3.1 Sources of Endogeneity

- **Measurement Error in Regressors:** 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. So we have a classical measurement error in which x_i is unbiased, but noisy measure of x_i^*. If we use x_i instead,

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

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 $E(v_i v_i') = 0$, $E(x_i u_i) \neq 0$. When we use OLS on this context, the probability limit of the OLS estimator would be

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

The only time that $\frac{E(x_i^* x_i^{*\prime})}{E(x_i^* x_i^{*\prime})+E(v_i v_i')}\beta$ would equal β is when β itself is zero or when $E(v_i v_i') = 0$. The latter, however, implies that $var(v_i) = 0$ and the noise v_i has mean 0 and has a point mass at 0 - so no measurement error exists. In usual cases, the OLS estimator has a probability limit of something less than β. This is what is also known as **attenuation bias**.

Comment 3.1 (Comment on Measurement Errors)**.** *So how do we address the endogeneity problem?*

- *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 δ_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.*

- **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 $E(p_i v_i) \neq 0$. When naively applying OLS to this equation, the result is as follows.

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

Thus, OLS estimators does not converge to either one of β_1 or β_2, resulting in a **simultaneity bias**. Keys to identifying each curves is to model a reduced form using an exogenous shock that affects one of demand or supply but not the other.

Here is how it works, let z_i denote some exogenous shock to the demand equation (say, any preference shock). We write the two equations as

$$q_i = \beta_1 p_i + u_i \tag{Supply}$$
$$q_i = -\beta_2 p_i + \beta_3 z_i + v_i \tag{Demand}$$

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 Variable Bias (OVB):** 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

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

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

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

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 1)x_i and a_i are correlated and 2)$\beta_2 \neq 0$, x_i is endogenous with respect to u_i. On the flip side, if either one of the condition is not met, $E(x_i u_i) = 0$ again. 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 1) and 2) occurs, the above does not converge in probability to β_1. Also note that we can determine the direction of the bias by the sign of $E(x_i a_i)$ and β_2.

3.2 IV Estimators

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 $\dim(x_{1i}) = k_1, \dim(x_{2i}) = k_2,\ k_1 + k_2 = k$. In our case, x_{2i} is the collection of endogenous variables. It can be shown that consistency of the OLS estimators of β_2 and β_1 will not be guaranteed under this situation.

Example 3.1 (When $k_1 = k_2 = 1$)**.** *In this case, we can write $\hat{\beta}_1$ as*

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{1i}y_i - \sum_{i=1}^n x_{1i}x_{2i} \sum_{i=1}^n x_{2i}y_i}{\sum_{i=1}^n x_{1i}^2 \sum_{i=1}^n x_{2i}^2 - (\sum_{i=1}^n x_{1i}x_{2i})^2}$$

When we replace y_i with $x_{1i}\beta_1 + x_{2i}\beta_2 + e_i$, we end up with

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{1i}e_i - \sum_{i=1}^n x_{1i}x_{2i} \sum_{i=1}^n x_{2i}e_i}{\sum_{i=1}^n x_{1i}^2 \sum_{i=1}^n x_{2i}^2 - (\sum_{i=1}^n x_{1i}x_{2i})^2}$$

Since $E(x_{2i}e_i) \neq 0$, then $\frac{1}{n}\sum_{i=1}^n x_{2i}e_i$ converges to something that is not zero (whereas $\frac{1}{n}\sum_{i=1}^n x_{1i}e_i$ does converge in probability to 0). So the whole fraction term does not converge in probability to 0 and even $\hat{\beta}_1$ is not 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 $\dim(z_{2i}) = l - k_1$ and $l - k_1 \geq k_2$. For z_i to be a valid IV, the following conditions must be satisfied

Definition 3.1 (IV conditions)**.** z_i is a valid IV if

1. **Exogeneity**: $E(z_ie_i) = 0$
   (a) **Exclusion**: $E(z_iy_i) = \beta_1E(z_ix_{1i}) + \beta_2E(z_ix_{2i})$, in other words, z_i should impact y_i through x_{1i} and x_{2i} only
2. **Relevancy**: $rank[E(z_ix'_i)] = \dim(x_i) = k$
3. **PD**: $E(z_iz'_i) > 0$

We will derive IV estimators in several ways

3.2.1 Reduced form methods

Using the above setup (but in matrices), we can write the reduced form relationship for the X_2 variables as

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

You can think of π_2 as a linear projection of X_2 onto Z: $E[Z'Z]^{-1}E[Z'X_2]$. This implicitly implies that $E[Z'v_2] = 0$. Since the above formulation only involves exogenous terms, we can use this to create a reduced form for Y

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

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

From the above setup, we can also write

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

which sums up the exact relation between the l reduced form parameters and $k_1 + k_2$ 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: the ratio of the reduced form estimates of Γ and π_1.

3.3 Using moment conditions

We will focus on the case where we have equal number of reduced form and structural parameters, thus just-identified. The structural equation and the moment conditions we will use are

$$y_i = x_i'\beta + e_i \ (E[z_ie_i] = 0)$$

with $E[x_i e_i]$ not necessarily zero. 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

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

where $E[z_i x_i']$ satisfying relevancy condition and being a square matrix implies the existence of an inverse matrix. The IV estimator is a sample analogue of the above, or

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

We can obtain the consistency and the asymptotic distribution of the IV estimator in the following way

Theorem 3.1 (Consistency of $\hat{\beta}_{IV}$)**.** $\hat{\beta}_{IV} \xrightarrow{p} \beta$

Proof.

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

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

Theorem 3.2 (Limiting distribution of $\hat{\beta}_{IV}$)**.** *The limiting distribution of the IV estimator can be characterized by*

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} N(0, Q_{ZX}^{-1}\Omega Q_{ZX}^{-1})$$

where $\Omega = E[z_i z_i' e_i^2]$

Proof. 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. Also, $\frac{1}{n}\sum_{i=1}^{n} z_i x_i' \xrightarrow{p} Q_{ZX}^{-1}$ by weak law of large numbers. Using Slutsky theorem to combine the two, we get

$$\sqrt{n}(\hat{\beta}_{IV}-\beta) \xrightarrow{d} Q_{zx}^{-1}N(0,\Omega) = N(0, Q_{ZX}^{-1}\Omega Q_{ZX}^{-1})$$

Under homoskedasticity, we can get that $\Omega = \sigma^2 Q_{ZZ}$ and that the finite sample variance can be characterized as

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

□