c.f) Question : the relationship between eigenvalues and testing power.

Answer) Testing $H_0 : C'\beta = a$.

$\hat{\beta} | \mathbb{X} \sim N(\beta, \sigma^2 (\mathbb{X}'\mathbb{X})^{-1})$

Then, $C'\hat{\beta} | \mathbb{X} \sim N(C'\beta, \sigma^2 C'(\mathbb{X}'\mathbb{X})^{-1} C)$

When we minimize the variance $\sigma^2 C'(\mathbb{X}'\mathbb{X})^{-1}C$, (i.e. increase the power)

$$\min_C \frac{C'(\mathbb{X}'\mathbb{X})^{-1}C}{C'C} \Rightarrow \text{The solution : } \arg\min_C \frac{C'(\mathbb{X}'\mathbb{X})^{-1}C}{C'C}$$

$\Rightarrow$ eigenvector of $(\mathbb{X}'\mathbb{X})^{-1}$ corresponding to the smallest eigen value of $(\mathbb{X}'\mathbb{X})^{-1}$

---

- Big $O_p(1)$, little $o_p(1)$

Note) $O_p(1)$ = stochastically bounded

$o_p(1)$ = $X_n \xrightarrow{P} 0$

① $o_p(1) + o_p(1) = o_p(1)$

If $X_n \xrightarrow{P} 0$, and $Y_n \xrightarrow{P} 0$, then $X_n + Y_n \xrightarrow{P} 0$

② $O_p(1) + O_p(1) = O_p(1)$

If $X_n = O_p(1)$ and $Y_n = O_p(1)$, then $X_n + Y_n = O_p(1)$

We know $|X_n + Y_n| \leq |X_n| + |Y_n|$

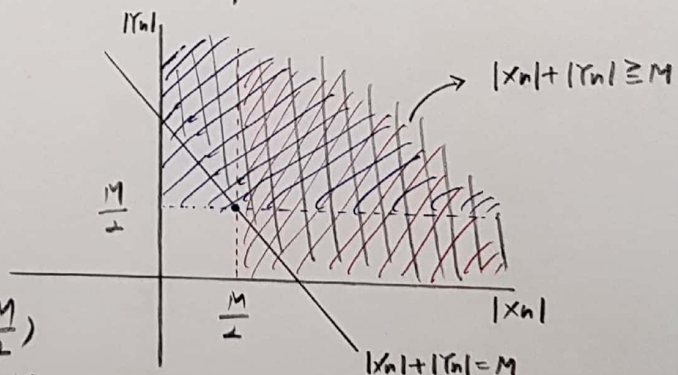
Then,

$Pr(|X_n + Y_n| \geq M) \leq Pr(|X_n| + |Y_n| \geq M)$

$\leq Pr(|X_n| \geq \frac{M}{2}) \cup Pr(|Y_n| \geq \frac{M}{2})$

$= Pr(|X_n| \geq \frac{M}{2}) + Pr(|Y_n| \geq \frac{M}{2})$

Both can be made small by choosing large M.



③ $\left\{\begin{array}{l} \underline{o_p(1) + O_p(1) = O_p(1)} \\ \underline{o_p(1) \cdot O_p(1) = o_p(1)} \end{array}\right.$

If $x_n = o_p(1)$, then $X_n = O_p(1)$ holds.

( Because $X_n \xrightarrow{P} 0$ implies $\exists M_\varepsilon > 0 : \lim_{n\to\infty} Pr\{|X_n| > M_\varepsilon\} < \varepsilon$, $\forall \varepsilon > 0$. )

(converges to 0 in probability) (stochastically bounded)

④ $\left\{\begin{array}{l} \underline{o_p(1) \cdot o_p(1) = o_p(1)} \\ \underline{O_p(1) \cdot O_p(1) = O_p(1)} \end{array}\right.$ (By continuous mapping theorem)

ex) Let $\underset{k\times k}{A_n} = O_p(1)$ and $\underset{k\times 1}{B_n} = o_p(1)$

$$\underset{k\times k\ k\times 1}{A_n B_n} = \begin{pmatrix} a_{n11} & \cdots & a_{n1k} \\ \vdots & \ddots & \vdots \\ a_{nk1} & \cdots & a_{nkk} \end{pmatrix} \begin{pmatrix} b_{n1} \\ \vdots \\ b_{nk} \end{pmatrix} = \begin{pmatrix} a_{n11} b_{n1} + \cdots + a_{n1k} b_{nk} \\ \vdots \\ a_{nk1} b_{n1} + \cdots + a_{nkk} b_{nk} \end{pmatrix}$$

$O_p(1) \cdot o_p(1) = o_p(1)$

and $o_p(1) + \cdots + o_p(1) = o_p(1)$.

$$= \begin{pmatrix} o_p(1) \\ \vdots \\ o_p(1) \end{pmatrix} = \underset{k\times 1}{o_p(1)} \quad \text{if } k \text{ is finite.}$$

exercise ) Let $X_{ni} = o_p(1)$.

Then $\frac{1}{n}\sum_{i=1}^{n} X_{ni} = o_p(1)$ ?

- What do we mean by "good estimator"?

So far conditional unbiasedness. ( $E(\hat{\theta} \mid \mathbb{X}) = \theta_0, \ \forall \theta_0 \in \Theta$ )

→ This is a kind of difficult one.

We can think of unbiasedness asymtotically.

① Asymtotic unbiasedness (weaker concept of unbiasedness)

$$E(\hat{\theta}_n \mid \mathbb{X}_n) \longrightarrow \theta_0, \quad \forall \theta_0 \in \Theta$$

⇑ assume $\mathbb{X}_n$ is a deterministic sequence.

ex)
$$\hat{\sigma}^2_{MLE} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - X_i' \hat{\beta}_{OLS})^2$$

$$E(\hat{\sigma}^2_{MLE} \mid \mathbb{X}_N) = E\left\{ \frac{N-K}{N} \cdot \frac{1}{N-K} \sum_{i=1}^{N} (Y_i - X_i' \hat{\beta}_{OLS})^2 \mid \mathbb{X}_N \right\}$$

$$= \frac{N-K}{N} E(\hat{\sigma}^2 \mid \mathbb{X}_N) = \frac{N-K}{N} \sigma^2 \longrightarrow \sigma^2 \text{ as } n \to \infty$$

⇒ $\hat{\sigma}^2_{MLE}$ : Not unbiased, but asymtotically unbiased.

$\hat{\sigma}^2_{OLS}$ : unbiased (It is shown as a specific parameter)

② Consistency ( Weak consistency $\xrightarrow{p}$, strong consistency $\xrightarrow{a.s}$ )

$\hat{\theta}_n$ is weakly consistent $\Leftrightarrow$ $\hat{\theta}_n \xrightarrow{p} \theta_0, \ \forall \theta_0 \in \Theta$

$\hat{\theta}_n$ is strongly consistent $\Leftrightarrow$ $\hat{\theta}_n \xrightarrow{a.s} \theta_0, \ \forall \theta_0 \in \Theta$

→ We need to check consistency in increasing sample size $n$.

→ Suppose two estimators are consistent.
In that case, their speeds to converge to $\theta_0$ would be different, so that the "convergence rate" is also an important property as an estimator.

⇓

## ③ Convergence rate

If $r_n \cdot (\hat{\theta}_n - \theta_0) = O_p(1)$, then an estimator has convergence rate $\frac{1}{r_n}$ ($r_n \to$

Typically, $r_n = \sqrt{n}$ (not always)

$\uparrow$

This is usually used for parametric / cross-section data cases.

→ Convergence rate only show how fast the convergence is.

If $O_p(1)$ has a distribution, then $\hat{\theta}_n - \theta_0$ also converges in the distribution.

$\boxed{\text{Note: If } r_n(\hat{\theta}_n - \theta_0) = O_p(1) \text{ and } r_n'(\hat{\theta}_n - \theta_0) = O_p(1), \text{ and } \frac{r_n}{r_n'} \to 0, \text{ then the rate is at least } r_n'.}$

## ④ Convergence in Distribution

$r_n(\hat{\theta}_n - \theta_0) \xrightarrow{d} Z$ (a random variable)

($r_n$ is the best rate)

→ Let $\hat{\theta}_n$ be consistent and $r_n(\hat{\theta}_n - \theta_0) = O_p(1)$.

Then, $\hat{\theta}_n \xrightarrow{p} \theta_0$ and $r_n$ is the convergence rate.

However, we cannot know to measure the inaccuracy of $\hat{\theta}_n$.

Thus, we can check its accuracy through the variance of $Z$.

→ Suppose $r_n(\hat{\theta}_n - \theta_0) \xrightarrow{d} Z$ hold. and $\exists \{r_n'\}$.

$$r_n'(\hat{\theta}_n - \theta_0) = \underbrace{\left(\frac{r_n'}{r_n}\right)}_{\Downarrow \; 0 \text{ or } \infty \text{ as } n \to \infty} \cdot \underbrace{r_n(\hat{\theta}_n - \theta_0)}_{\xrightarrow{d} Z}$$

(Because they are deterministic)

Therefore, this result is used to evaluate the errors we make by using $\hat{\theta}_n$.

- If we have two estimators that converge with different rate, then the one that converges faster is the better estimator.
- If two estimators converge at the same rate, then the one which has the smaller inaccuracy measure (typically variance) is the better estimator.

ex) Typically, $\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{d} N(0, \sigma_1^2)$

$\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{d} N(0, \sigma_2^2)$

then if $\sigma_1^2 < \sigma_2^2$, then we say $\hat{\theta}_{1n}$ is more efficient then $\hat{\theta}_{2n}$.

and $\frac{\sigma_2^2}{\sigma_1^2}$ is the relative efficiency of the first estimator.

ex2)

Consider estimation of $E(X)$ by $\frac{1}{N}\sum_{i=1}^{N} X_i$

$Var\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N}\sigma_X^2$

$Var(\hat{\theta}) = \frac{1}{N}\sigma_1^2$

Suppose $\frac{\sigma_1^2}{\sigma_X^2} = 2$. and assume that $\exists$ sample size $\frac{N}{2}$

Then $Var\left(\frac{1}{(\frac{N}{2})}\sum_{i=1}^{\frac{N}{2}} X_i\right) = \frac{1}{2N}\cdot\sigma_X^2$

At the viewpoint of sample size ploblem,

"2" means sample size of $\hat{\theta}$ is double more than $\frac{1}{N}\sum_{i=1}^{N} X_i$.

• Asymtotic Properties of the OLS estimator

$$\hat{\beta} = (X'X)^{-1}X'Y = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i y_i = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i (x_i'\beta + u_i)$$

$$= \cancel{\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i x_i'}\beta + \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i u_i$$

$$= \beta + \underbrace{\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}}_{:=A_N} \underbrace{\sum_{i=1}^{N} x_i u_i}_{:=B_N}$$

By Kolmogorov's SLLN.

① $k$-$l$ element of $A_N$ is $\frac{1}{N}\sum_{i=1}^{N} x_{ki} x_{li} \xrightarrow{a.s.} E(x_{ki} x_{li})$ Under iid

Therefore, $\underline{A_N \xrightarrow{a.s} E\{x_i x_i'\}}$ ( Because each element converges to $E(x_{ki} x_{li})$ for each $k, l$ )

$(\Rightarrow)$ $A_N^{-1} \xrightarrow{a.s} E\{x_i x_i'\}^{-1})$ $\cdots$ ⊛

② $k$th element of $B_N$ is $\frac{1}{N}\sum_{i=1}^{N} x_{ki} u_i \xrightarrow{a.s.} E(x_{ki} u_i)$

$$= E(E(x_{ki} u_i \mid x_i))$$

$$= E(x_{ki} \cdot \underbrace{E(u_i \mid x_i)}_{=0}) = 0$$

(Since $x_{ki}$ is constant)

Therefore, $\frac{1}{N}\sum_{i=1}^{N} x_{ki} u_i \xrightarrow{a.s.} 0$

From ⊛, $A_N^{-1} = O_p(1)$, and $B_N = o_p(1)$. Therefore, $O_p(1) \cdot o_p(1) = o_p(1)$

$\hat{\beta} = \beta + o_p(1)$, therefore, this shows strong consistency of OLS estimator.

※ Important assumptions

① iid sampling

② $E(x_i x_i')$ is invertible

③ $E(u_i \mid x_i) = 0$

(We need $E(x_i u_i) = 0$)

• To show asym. property of OLS estimator, We don't need $X'X$ is invertible, but need $E(x_i x_i')$ is invertible

• For consistency, Homoskedasticity is not important.

• Asymtotic Properties of the OLS estimator (continued..)

$\hat{\beta} \xrightarrow{a.s} \beta \qquad \sqrt{N}(\hat{\beta}-\beta) = \underbrace{\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}}_{①} \underbrace{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i u_i}_{②}$

① $\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \xrightarrow{a.s.} E\{x_i x_i'\}$ by kolmogrov's SLLN

Note < Kolmogrov's SLLN >

If $\{x_i\}$ is an independence sequence of random variables and $E(x_i)=\mu_i$, $Var(x_i)=\sigma_i^2$ and $\sum_{i=1}^{\infty}\frac{\sigma_i^2}{i^2}<\infty$

then $\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_i) \xrightarrow{a.s} 0$ as $n\to\infty$

② $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \underbrace{\underset{K\times1}{x_i}\,\underset{1\times1}{u_i}}_{K\times1} \xrightarrow{d} N(0, E(u_i^2 x_i x_i'))$

proof sketch ) Let $\lambda$: $K\times1$ vector. Then, $\lambda'\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i u_i\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\lambda' x_i)u_i$ : Scalar.

Thus, possible to apply CLT.

By CLT & under iid assumption,

$E(\lambda' x_i u_i) = E(E[\lambda' x_i u_i \mid x_i]) = E(\lambda x_i u_i \underbrace{E(u_i \mid x_i)}_{=0}) = 0$

$Var(\lambda' x_i u_i) = E[(\lambda' x_i u_i)^2]$ ($\because \underset{1\times K}{\lambda'}\underset{K\times1}{x_i}$ is scalar)

$= E[(\lambda' x_i)^2 u_i^2]$

$= E[(\lambda' x_i) u_i^2 (x_i'\lambda)]$

$= \lambda' E[x_i u_i^2 x_i']\lambda$ ($\because \lambda$ is fixed)

$= \underset{1\times K}{\lambda'} \underbrace{E[u_i^2 x_i x_i']}_{K\times K} \underset{K\times1}{\lambda}$ ($\because u_i^2$ is scalar)

If assumed that $E\{u_i^2 x_{ji} x_{ki}\}<\infty$, $\forall j,k\in\{1,\dots,K\}$

then $Var(\lambda' x_i u_i)<\infty$

$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\lambda' x_i)u_i \xrightarrow{d} N(0, Var(\lambda' x_i u_i)) = N(0, \lambda' E[u_i^2 x_i x_i']\lambda)$

$= \lambda' N(0, E[u_i^2 x_i x_i'])$

Therefore, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(x_i u_i) \xrightarrow{d} N(0, E[u_i^2 x_i x_i'])$ (Asymtotically Normal)

**Remark**

$$E(u_i^2 x_i x_i') = E(E(u_i^2 x_i x_i' \mid x_i))$$

$$= E(E(u_i^2 \mid x_i) x_i x_i')$$

$\downarrow$ under Homoskedasticity ( $E(u_i^2 \mid x_i) = \sigma^2, \forall i$ )

$$= \sigma^2 E(x_i x_i')$$

→ Until now, we watched the infinite sample case so that we used the CLT. The result is "$\sqrt{n}(\hat{\beta}-\beta)$ is asymptotically normal".

$$\sqrt{N}(\hat{\beta}-\beta) = \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{\sqrt{N}}X'u$$

asym. normal by CLT

$$= \underbrace{E(x_i x_i')^{-1}}_{\text{constant}} \cdot \underbrace{\frac{1}{\sqrt{N}}X'u}_{\xrightarrow{d} N(0,V)} + \underbrace{\left[\left(\frac{1}{N}X'X\right)^{-1} - E(x_i x_i')^{-1}\right]}_{\xrightarrow{p} 0} \underbrace{\frac{1}{\sqrt{N}}X'u}_{\xrightarrow{d} N(0,V)}$$

$V = E(u_i^2 x_i x_i')$

$\xrightarrow{d} E(x_i x_i')^{-1} N(0, E(u_i^2 x_i x_i'))$

(By continuous mapping thm)

asym. normal by CLT

Thus, $= o_p(1)$

Thus, $\frac{1}{\sqrt{N}}X'u = O_p(1)$

$o_p(1) \cdot O_p(1) = o_p(1)$

By slutsky, $\square \xrightarrow{d} \star$ & $\Delta = o_p(1)$

$$\xrightarrow{d} E(x_i x_i')^{-1} N(0, E(u_i^2 x_i x_i'))$$

$$= N(0, E(x_i x_i')^{-1} E(u_i^2 x_i x_i') E(x_i x_i')^{-1'})$$

Asymptotically Normal

**Remark** Under Homoskedasticity,

$$\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{d} N(0, \sigma^2 E(x_i x_i')^{-1})$$

→ Compare the result with the finite sample case.

In finite sample analysis <u>with Normality assumption</u>,

$$\hat{\beta} - \beta \sim N\left(0,\ (X'X)^{-1}(X'\Omega X)(X'X)^{-1}\right)$$

$$\underline{\sqrt{N}(\hat{\beta}-\beta)} \sim N\left(0,\ N(X'X)^{-1}(X'\Omega X)(X'X)^{-1}\right)$$

$$= \underline{N\left(0,\ \left(\tfrac{1}{N}X'X\right)^{-1}\underbrace{\left(\tfrac{1}{N}X'\Omega X\right)}_{=\frac{1}{N}\sum_{i=1}^{N}\sigma^2(x_i)x_ix_i'}\left(\tfrac{1}{N}X'X\right)^{-1}\right)}$$

$$\& \quad \Omega = \begin{pmatrix} \sigma^2(x_1) & 0 & \cdots & 0 \\ 0 & \sigma^2(x_2) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2(x_N) \end{pmatrix}$$

※ Infinite sample case, that is, CLT is applied
Finite sample case & Normality assumption
) The result is same!

( In short, asymtotically normal distribution can be made by CLT, with No assumption, if the sample size $\to \infty$ )

- Estimation of the variance - covariance matrix.

→ we cannot see $u_i$ so that we should check variance.

$E(x_i x_i')$ can be estimated consistently by $\frac{1}{N}\sum_{i=1}^{N} x_i x_i'$

SLLN implies $\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \xrightarrow{a.s.} E(x_i x_i')$ Under iid and $E(x_i x_i') < \infty$

$E(u_i^2 x_i x_i') = \sigma^2 E(x_i x_i')$ Under Homoskedasticity.

Under the assumption, the only parameter to be estimated is $\sigma^2$

$$\hat{\sigma}^2 = \frac{1}{N-K}\sum_{i=1}^{N}\hat{u}_i^2 = \frac{1}{N-K}\sum_{i=1}^{N}(y_i - x_i'\hat{\beta})^2$$

(Note) $y_i = x_i'\beta + u_i$
$y_i = x_i'\hat{\beta} + \hat{u}_i$
$\hat{y}_i = x_i'\hat{\beta}$
$\bar{y} = \bar{x}'\hat{\beta}$

$$= \frac{1}{N-K}\sum_{i=1}^{N}(x_i'\beta + u_i - x_i'\hat{\beta})^2$$

$$= \frac{1}{N-K}\sum_{i=1}^{N}(u_i - x_i'(\hat{\beta}-\beta))^2 = \frac{1}{N-K}\sum_{i=1}^{N}\left[u_i^2 - 2u_i x_i'(\hat{\beta}-\beta) + [x_i'(\hat{\beta}-\beta)]^2\right]$$

$$= \underbrace{\frac{1}{N-K}\sum_{i=1}^{N}u_i^2}_{①} - 2\underbrace{\frac{1}{N-K}\sum_{i=1}^{N}u_i x_i'(\hat{\beta}-\beta)}_{②} + \underbrace{\frac{1}{N-K}\sum_{i=1}^{N}[x_i'(\hat{\beta}-\beta)]^2}_{③}$$

① $\frac{1}{N-K}\sum_{i=1}^{N}u_i^2 = \underbrace{\frac{N}{N-K}}_{\to 1}\underbrace{\boxed{\frac{1}{N}\sum_{i=1}^{N}u_i^2}}_{\xrightarrow{a.s} E(u_i^2)=\sigma^2 \text{ (by SLLN)}} \xrightarrow{a.s.} \sigma^2$

② $\frac{1}{N-K}\sum_{i=1}^{N}u_i x_i'(\hat{\beta}-\beta) = \underbrace{\frac{N}{N-K}}_{\to 1}\underbrace{\boxed{\frac{1}{N}\sum_{i=1}^{N}u_i x_i'}}_{\xrightarrow{a.s} E(u_i x_i')}\underbrace{(\hat{\beta}-\beta)}_{\xrightarrow{a.s} 0} \xrightarrow{a.s} 0$

$= 0$
↑
(By Iterated expectation)

③ $\frac{1}{N-K}\sum_{i=1}^{N}[x_i'(\hat{\beta}-\beta)]^2 = \frac{N}{N-K}\cdot\frac{1}{N}\sum_{i=1}^{N}[(\hat{\beta}-\beta)'x_i][x_i'(\hat{\beta}-\beta)] = o_p(1)\cdot O_p(1)\cdot o_p(1) = o_p(1)$.

$$= (\hat{\beta}-\beta)'\frac{N}{N-K}\boxed{\frac{1}{N}\sum_{i=1}^{N}x_i x_i'}(\hat{\beta}-\beta)$$

$(\hat{\beta}-\beta)' \downarrow a.s$ 0, Thus, $o_p(1)$

$\frac{1}{N}\sum x_i x_i' \downarrow a.s$ $E(x_i x_i')$ Thus, this is $O_p(1)$

$(\hat{\beta}-\beta) \downarrow a.s$ 0 Thus, $= o_p(1)$

$\xrightarrow{a.s.} \sigma^2$ : unbiased & consistent. Therefore, $\hat{\sigma}^2$ is strongly consistent estimator of $\sigma^2$

(c.f.) $\hat{\sigma}^2_{mle}$ : Not unbiased, but consistent)

- Single equation estimation / testing

$$C'\hat{\beta} \sim N(C'\beta,\ C'\sigma^2(X'X)^{-1}C)$$

$$\frac{C'\hat{\beta} - C'\beta}{\sqrt{\sigma^2 C(X'X)^{-1}C}} \sim N(0,1)$$

We do not know $\sigma^2$. Thus,

$$\frac{C'\hat{\beta} - C'\beta}{\sqrt{\hat{\sigma}^2 C'(X'X)^{-1}C}} \sim t_{(N-K)}$$

Apply CLT

asymtotically (By CLT)

$$\frac{C'\sqrt{N}(\hat{\beta}-\beta)}{\sqrt{\hat{\sigma}^2 N C'(X'X)^{-1}C}} = \frac{C'\sqrt{N}(\hat{\beta}-\beta)}{\sqrt{\hat{\sigma}^2 C'(\frac{1}{N}(X'X))^{-1}C}} \overset{a}{\sim} \frac{C'N(0,\sigma^2 E(x_i x_i')^{-1}) = N(0, C'\sigma^2 E(x_i x_i')^{-1}C)}{\sqrt{\hat{\sigma}^2 C' E(x_i x_i')^{-1}C}}$$

In fact, this follows t-statistics.

When we see the asymtotical result (?), as $n \to \infty$, this is approximately Normal.

---

- Multiple equations and testing

$\underset{r\times k}{C'}(\hat{\beta}-\beta)$  $r \geq k$, $\text{rank}(C) = r$   "2/1 note"

$$[C'(\hat{\beta}-\beta)]'[\sigma^2 C'(X'X)^{-1}C]^{-1}[C'(\hat{\beta}-\beta)] \sim \chi^2_{(r)} \text{ Under Normality}$$

$$[C'(\hat{\beta}-\beta)]'[\hat{\sigma}^2 C'(X'X)^{-1}C]^{-1}[C'(\hat{\beta}-\beta)]/r \sim F(r, N-K) \text{ Under Normality}$$

Apply CLT

$$N[C'(\hat{\beta}-\beta)]'\frac{1}{N}[\hat{\sigma}^2 C'(X'X)^{-1}C]^{-1}[C'(\hat{\beta}-\beta)]/r$$

$$= [C'\sqrt{N}(\hat{\beta}-\beta)]'[\hat{\sigma}^2 C'(\tfrac{1}{N}X'X)^{-1}C]^{-1}[C'\sqrt{N}(\hat{\beta}-\beta)]/r \xrightarrow{d} \chi^2_{(r)}/r$$

$\downarrow d$ $N(0, C'\sigma^2 E(x_i x_i')^{-1}C)$ (By CLT)

$\downarrow$ a.s $\sigma^2 C' E(x_i x_i')^{-1}C$ (By $(\frac{1}{N}X'X)^{-1} \xrightarrow{a.s} E(x_i x_i')^{-1}$)

$\downarrow d$ $N(0, C'\sigma^2 E(x_i x_i')^{-1}C)$ (By CLT)

※ Both of the above cases show that, with No Normality assumption,

CLT makes $C'(\hat{\beta}-\beta)$ asymtotically Normal distribution.

- Analysis without assuming homoskedasticity.
  - Estimation of $E(x_i x_i')$ is the same.
  - Main issue : $E(u_i^2 x_i x_i')$

$\downarrow$

Estimation of $E(u_i^2 x_i x_i')$ by $\frac{1}{N}\sum_{i=1}^{N} \hat{u}_i^2 x_i x_i'$

we don't know $u_i$. Thus, use $\hat{u}_i$

$$\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i^2 x_i x_i' = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\hat{\beta})^2 x_i x_i' = \frac{1}{N}\sum_{i=1}^{N}(x_i'\beta + u_i - x_i'\hat{\beta})^2 x_i x_i'$$

$$= \frac{1}{N}\sum_{i=1}^{N}[u_i - x_i'(\hat{\beta}-\beta)]^2 x_i x_i' = \frac{1}{N}\sum_{i=1}^{N}[u_i^2 - 2u_i x_i'(\hat{\beta}-\beta) + \{x_i'(\hat{\beta}-\beta)\}^2] x_i x_i'$$

$$= \underbrace{\frac{1}{N}\sum_{i=1}^{N} u_i^2 x_i x_i'}_{①} - \underbrace{\frac{2}{N}\sum_{i=1}^{N} u_i x_i'(\hat{\beta}-\beta) x_i x_i'}_{②} + \underbrace{\frac{1}{N}\sum_{i=1}^{N}[x_i'(\hat{\beta}-\beta)]^2 x_i x_i'}_{③}$$

① $\frac{1}{N}\sum_{i=1}^{N} u_i^2 x_i x_i' \xrightarrow{a.s} E(u_i^2 x_i x_i')$ By SLLN.

② j-kth element is $\frac{2}{N}\sum_{i=1}^{N} u_i x_i'(\hat{\beta}-\beta) x_{ji} x_{ki} = 2\cdot \boxed{\frac{1}{N}\sum_{i=1}^{N} u_i x_{ji} x_{ki} x_i'}(\hat{\beta}-\beta) \xrightarrow{a.s.} 0$

$\downarrow a.s.$

$E(u_i x_{ji} x_{ki} x_i') = 0$

In ②, $(\hat{\beta}-\beta)$ is inside the equation.
but, in this j-k element case, those are scalars.
Thus, this approach is possible to apply SLLN.

③ j-kth element is

$$\frac{1}{N}\sum_{i=1}^{N} x_{ji} x_{ki} [x_i'(\hat{\beta}-\beta)]^2 = \frac{1}{N}\sum_{i=1}^{N} x_{ji} x_{ki} [(\hat{\beta}-\beta)' x_i][x_i'(\hat{\beta}-\beta)]$$

$$= \underbrace{(\hat{\beta}-\beta)'}_{\downarrow a.s.\ 0} \boxed{\frac{1}{N}\sum_{i=1}^{N} x_{ji} x_{ki} x_i x_i'} \underbrace{(\hat{\beta}-\beta)}_{\xrightarrow{a.s} 0} \xrightarrow{a.s} 0$$

(Under the assumption that elements of $x_i$ have 4th moment, this will be finite.

Therefore, $\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i^2 x_i x_i' \xrightarrow{a.s.} E(u_i^2 x_i x_i')$ //

PS #4

#1. (a) $y_{it} = \beta_{1t}\, d_{June} + \beta_{2t} \cdot d_{July} + \varepsilon_{it}$

where $d_{June} = \begin{cases} 1 & \text{if give birth in June 1990} \\ 0 & \text{if in July} \end{cases}$

→ $\hat{\beta}_1, \hat{\beta}_2$ then $\hat{\beta}_2 - \hat{\beta}_1 = [-1, 1]\begin{bmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{bmatrix}$ and use $Var(c \cdot \hat{\beta}) = c' Var \hat{\beta} c$

Or

$y_{it} = (\beta_{11} d_{June,1} + \beta_{21} d_{June,1})\, 1\{t=1\} + (\beta_{12} d_{June,2} + \beta_{22} d_{July,2}) \cdot 1\{t=2\} + \varepsilon_{it}$

(b) Base : $y_i = \alpha + \beta \cdot July + \varepsilon$

Control : $y_i = \alpha + \beta \cdot July + X\gamma + \varepsilon$

$y = \alpha + \beta_1 x_1 + \varepsilon$

$= \alpha + \beta_1 x_1^* + \varepsilon$, $x_1^* := x_1 - \bar{x}_1$

$$\begin{pmatrix}\hat{\alpha} \\ \hat{\beta}_1\end{pmatrix} = \begin{bmatrix} 1'1 & 1'x_1^* \\ x_1^{*\prime}1 & x_1^{*\prime}x_1^* \end{bmatrix}^{-1} \begin{bmatrix} 1' \\ x_1^{*\prime} \end{bmatrix} Y$$

$$= \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & (x_1^{*\prime}x_1^*)^{-1} \end{bmatrix} \begin{bmatrix} 1' \\ x_1^{*\prime} \end{bmatrix} Y$$

Thus, $\hat{\beta}_1 = (x_1^{*\prime}x_1^*)^{-1}x_1^{*\prime}Y$

$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

$= \alpha + \beta_1 x_1^* + \beta_2 x_2^* + \varepsilon$

$$\begin{pmatrix}\tilde{\alpha} \\ \tilde{\beta}_1 \\ \tilde{\beta}_2\end{pmatrix} = \begin{pmatrix} 1'1 & \underbrace{1'x_1^*}_{=0} & \underbrace{1'x_2^*}_{=0} \\ \underbrace{x_1^{*\prime}1}_{=0} & x_1^{*\prime}x_1 & \underbrace{x_1^{*\prime}x_2}_{\approx 0} \\ \underbrace{x_2^{*\prime}1}_{=0} & \underbrace{x_2^{*\prime}x_1}_{\approx 0} & x_2^{*\prime}x_2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 1'Y \\ x_1^{*\prime}Y \\ x_2^{*\prime}Y \end{pmatrix}$$

(∵ $x_i^*$'s are data.)

$$= \begin{pmatrix} \frac{1}{N} & 0 & 0 \\ 0 & (x_1^{*\prime}x_1^*)^{-1} & 0 \\ 0 & 0 & (x_2^{*\prime}x_2^*)^{-1} \end{pmatrix} \begin{pmatrix} 1'Y \\ x_1^{*\prime}Y \\ x_2^{*\prime}Y \end{pmatrix}$$

Thus, $\tilde{\beta}_1 = (x_1^{*\prime}x_1^*)^{-1}x_1^{*\prime}Y$

※ ① Constant term does not affect $\hat{\beta}$.

② Controls are independent with the variable of interest.

- $E(X_iX_i')^{-1} \cdot N(0, E(u_i x_i x_i')) = N(0, E(x_ix_i')^{-1} E(u_i x_i x_i') E(x_ix_i')^{-1\prime})$

WTS : $\underset{k\times k}{\mathcal{X}} \underset{k\times 1}{V}$, $\mathcal{X}$ is symmetric. $V \sim N(0, \Sigma)$

Then, $\mathcal{X}V \sim N(0, \mathcal{X}\Sigma\mathcal{X}')$

① $\mathcal{X}V$ is normal

$\underset{k\times k}{\mathcal{X}} = \begin{bmatrix} x_1' \\ \vdots \\ \underbrace{x_k'}_{1\times k} \end{bmatrix}_{k\times k}$ where $x_i'$ is a $k\times 1$ vector, $i = 1, \dots, k$. and $\underset{k\times 1}{V} = \begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix}$

$\underbrace{\underset{k\times k}{\mathcal{X}} \underset{k\times 1}{V}}_{k\times 1} = \begin{bmatrix} x_1'V \\ \vdots \\ x_k'V \end{bmatrix}$ $\therefore$ $x_i'V$s are linear combinations of $\{v_i\}$.

Therefore, $\mathcal{X}V$ is normal by $V \sim N(0, \Sigma)$

② $E(\mathcal{X}V) = 0$

$E(\mathcal{X}V) = E\begin{bmatrix} x_1'V \\ \vdots \\ x_k'V \end{bmatrix} = \begin{bmatrix} E(x_1'V) \\ \vdots \\ E(x_k'V) \end{bmatrix} = 0$ ( Because $E(x_i'V)$
$= E[x_i' \underbrace{E(V|x_i')}_{=0}] = 0$

③ $Var(\mathcal{X}V) = \mathcal{X}\Sigma\mathcal{X}'$

$\underset{1\times k\ k\times 1}{Var(x_i'V)} = E[\underbrace{(x_i'V)^2}_{\text{Scalar}}] = E[\underset{1\times k\ k\times 1\ 1\times k\ k\times 1}{x_i'VV'x_i}] = x_i'E[VV']x_i = x_i'\Sigma x_i$, $\forall i = 1, \dots, k$

$cov(x_i'V, x_j'V) = E[x_i'V x_j'V] = E[x_i'VV'x_j] = x_i'E[VV']x_j = x_i'\Sigma x_j$, $\forall i = 1, \dots, k$, $j = 1, \dots, k$

Thus, $Var(\mathcal{X}V) = Var\begin{pmatrix} x_1'V \\ \vdots \\ x_2'V \end{pmatrix} = \begin{pmatrix} Var(x_1'V) & cov(x_1'V, x_2'V) & \cdots & cov(x_1'V, x_k'V) \\ \vdots & & \ddots & \vdots \\ cov(x_k'V, x_1'V) & \cdots & \cdots & Var(x_k'V) \end{pmatrix}$

$= \begin{pmatrix} x_1'\Sigma x_1 & x_1'\Sigma x_2 & \cdots & x_1'\Sigma x_k \\ x_2'\Sigma x_1 & x_2'\Sigma x_2 & & \vdots \\ \vdots & & \ddots & \\ x_k'\Sigma x_1 & x_k'\Sigma x_2 & \cdots & x_k'\Sigma x_k \end{pmatrix} = \begin{pmatrix} x_1' \\ \vdots \\ x_k' \end{pmatrix} \Sigma (x_1 \cdots x_k)$

$= \mathcal{X}\Sigma\mathcal{X}'$ //