EC402 (2023/24)

RAGVIR'S SPEAKING NOTES FOR CLASS

Vassilis PS 6 (WK 7, AT)

(Q1) $y = X\beta + \varepsilon$ where $(X'X) = \begin{bmatrix} x_1'x_1 & x_1'x_2 \\ x_2'x_1 & x_2'x_2 \end{bmatrix}$; $N = 11$, $K = 2$

Further $\varepsilon | X \sim N(0, \sigma^2 I_N)$

(a) We want a $100(1-\alpha)\%$ P.I. for $\begin{cases} y_{12} \\ y_{13} \end{cases}$ where $\alpha = 0.2$ & $x_{12}' = (5, -2)$
... $x_{13}' = (3, -7)$

- We have $(X'X) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $X'y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- So $(X'X)^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \frac{1}{3}$ and $\hat{\beta} := (X'X)^{-1}X'y = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}$

Thus, $\begin{bmatrix} \hat{y}_{12} \\ \hat{y}_{13} \end{bmatrix} = \begin{bmatrix} 5 & -2 \\ 3 & -7 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 1 \\ -4/3 \end{bmatrix}$

- let's consider the RVs:

$(\hat{y}_{12}-y_{12}) / \text{Var}^{1/2}(\hat{y}_{12}-y_{12}|X)$ & $(\hat{y}_{13}-y_{13}) / \text{Var}^{1/2}(\hat{y}_{13}-y_{13}|X)$

Both of these are conditionally $N(0,1)$ under A1, A2, A3'fi, A4GMiid, A5N.

- Of course, we don't know the denominators since σ^2 is unknown so we need to use

$(\hat{y}_{12}-y_{12}) / \widehat{\text{Var}}^{1/2}(\hat{y}_{12}-y_{12}|X)$ & $(\hat{y}_{13}-y_{13}) / \widehat{\text{Var}}^{1/2}(\hat{y}_{13}-y_{13}|X)$, which are conditionally $t_{11-2=9}$, as our pivotal functions instead.

- So what is $\widehat{\text{var}}(\hat{y}_{12}-y_{12}|X)$?

$$\text{var}(\hat{y}_{12}-y_{12}|X) = \text{var}(x_{12}'\hat{\beta} - x_{12}'\beta - \varepsilon_{12}|X) = \text{var}(x_{12}'(\hat{\beta}-\beta) - \varepsilon_{12}|X)$$

$$= x_{12}' \,\text{Var}(\hat{\beta}-\beta|X)\, x_{12} + \text{var}(\varepsilon_{12}|X)$$

$$= x_{12}' \sigma^2 (X'X)^{-1} x_{12} + \sigma^2.$$

$$\therefore \hat{\text{var}}(\hat{y}_{12} - y_{12} | X) = \left[1 + x_{12}'(X'X)^{-1} x_{12}\right] \hat{\sigma}^2$$

and we use an analogous expression for $\hat{y}_{13}$.

All that remains is to compute $\hat{\sigma}^2$:

$$(N-K)\hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon} \overset{\text{Appendix A}}{=} y'y - \hat{\beta}'X'X\hat{\beta} = \frac{4}{3} - \begin{bmatrix} 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix} = \frac{2}{3}, \text{ so}$$

$$\therefore \hat{\sigma}^2 = \frac{2}{27}.$$

Then, we can compute $\hat{\text{var}}(\hat{y}_{12} - y_{12} | X) = \left\{1 + \begin{bmatrix} 5 & -2 \end{bmatrix} \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 5 \\ -2 \end{bmatrix}\right\} \frac{2}{27}$

$$= (2/27)\left[1 + 26\right] = 2.$$

Given our pivotal functions, and with $\alpha = 0.2$, we know that

$$P\left(t_{9,\alpha/2} = -1.383 < \left(\frac{\hat{y}_{12} - y_{12}}{\hat{\text{var}}^{1/2}(\hat{y}_{12} - y_{12} | X)}\right) < t_{9,1-\frac{\alpha}{2}} = 1.383\right) = 1 - \alpha.$$

Indeed, $P\left(1-\sqrt{2}\,1.383 < y_{12} < 1+\sqrt{2}(1.383)\right) = 0.8$

So, an 80% P.I. for y_{12} is $1 \pm \sqrt{2}\,1.383$.

Analogously, an 80% P.I. for y_{13} is ...

[Try this one yourself.]

(b) Now we need a $100(1-\alpha)\%$ P.I. for $m_{12} := E[y_{12}|X]$ and
$m_{13} := E[y_{13}|X]$.

The appropriate pivotal functions would be

$$(\hat{y}_{12} - m_{12}) / \hat{var}^{1/2}(\hat{y}_{12} - m_{12}|X) \;\&\; (\hat{y}_{13} - m_{13}) / \hat{var}^{1/2}(\hat{y}_{13} - m_{13}|X) \sim t_q.$$

As before, let's work out $\hat{var}(y_{12} - m_{12}|X)$.

$$var(\hat{y}_{12} - m_{12}|X) = var(x_{12}'(\hat{\beta} - \beta)|X) = x_{12}' \, var(\hat{\beta}|X) \, x_{12}$$

$$= x_{12}' \, \sigma^2 (X'X)^{-1} x_{12}.$$

$$\therefore \hat{var}^{1/2}(\hat{y}_{12} - m_{12}|X) = \hat{\sigma}^2 [x_{12}' (X'X)^{-1} x_{12}].$$

Plugging the given numbers in,

$$\hat{var}^{1/2}(\hat{y}_{12} - m_{12}|X) = \left[(2/27) 26\right]^{1/2} = \sqrt{1.926}$$

Then, an 80% P.I. for m_{12} is $1 \pm \sqrt{1.926}\, 1.383$

Analogously, an 80% P.I. for m_{13} is...

[Try this one yourself.]

(c) The 80% P.I.s for y_{12} and y_{13} are necessarily wider than those for m_{12} and m_{13} since the former also account for prediction error variance due to ε_{12} and ε_{13} respectively.

COMPARE CONDITIONAL VARIANCE EXPRESSIONS HERE, PLEASE:

APPENDIX A: SMALL COMMON QUESTION

"RAGUIR, WHY IS $\hat{\varepsilon}'\hat{\varepsilon} = y'y - \hat{\beta}'X'X\hat{\beta}$?"

Method 1 :

$$\hat{\varepsilon}'\hat{\varepsilon} = (y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

$$= y'y - \hat{\beta}'X'X\hat{\beta} - \hat{\beta}'(X'X)\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

$$= y'y - \hat{\beta}'X'X\hat{\beta} \,,$$

since $\hat{\beta}'X'y = \hat{\beta}'(X'X)(X'X)^{-1}X'y = \hat{\beta}'(X'X)\hat{\beta}$.

Method 2:

$$y'y = (\hat{y}+\hat{\varepsilon})'(\hat{y}+\hat{\varepsilon}) = \hat{y}'\hat{y} + \hat{\varepsilon}'\hat{y} + \hat{y}'\hat{\varepsilon} + \hat{\varepsilon}'\hat{\varepsilon}$$

Since $\hat{y}'\hat{\varepsilon} = \hat{\varepsilon}'\hat{y} = (M_X\varepsilon)'X\hat{\beta} = \varepsilon' M_X X\hat{\beta}$
$= 0$, it follows that

$$\hat{\varepsilon}'\hat{\varepsilon} = y'y - \hat{y}'\hat{y} = y'y - \hat{\beta}'X'X\hat{\beta}.$$

(There may exist other methods too.)

(Q2) Say $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$ for $t = 1, \dots, T$; where $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. Assume A3F.

(a) Construct a $100(1-\alpha)\%$ C.I. for β_2 given $\alpha \in (0,1)$.

Recall (from PS2 Q3) that the $(2,2)^{th}$ element of 2×2 matrix $\sigma^2 (X'X)^{-1}$ is $\sigma^2 T \left[T \sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2 \right] =: Var(\hat{\beta}_2)$, where $\hat{\beta}_2$ refers to the usual OLS estimator for β_2.

Note: If you cannot recall it then that's ok, but please in that case try to develop the ability to work out $Var(\hat{\beta}_2)$ (exactly as per PS2 Q3).

Since $T \sum_{t=1}^{T} (x_t - \bar{x})^2 = T \sum_{t=1}^{T} (x_t^2 - 2x_t\bar{x} + \bar{x}^2) = T \sum_{t=1}^{T} x_t^2 - T^2 \bar{x}^2$

$= T \sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2$, we obtain that

$$Var(\hat{\beta}_2) = \sigma^2 \left(\sum_{t=1}^{T} (x_t - \bar{x})^2 \right)^{-1}.$$

- It follows given the model specification that the function

$$\frac{\hat{\beta}_2 - \beta_2}{\sigma^2 \left(\sum_{t=1}^{T} (x_t - \bar{x})^2 \right)^{-1}} \sim N(0,1)$$

would be pivotal for β_2 if σ^2 were known.

- However, since σ_ε^2 is unknown, we use instead

$$Q(\beta, x, y) := \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}^2 \left(\sum_{t=1}^{T} (x_t - \bar{x})^2 \right)^{-1}} \sim t_{T-2}$$

as our pivotal quantity, where $\hat{\sigma}^2 := (T-2)^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^2$.

- Given the above, we can always find $q_1, q_2 \in \mathbb{R}$ s.t. $P\left(q_1 \leq Q(\beta, x, y) \leq q_2 \right) = 1 - \alpha$.

- Indeed, suppose $q_2 = t_{T-2, 1-\alpha/2}$ and (by symmetry) $q_1 = -t_{T-2,1-\alpha/2}$.
- Let us also define $\hat{Var}(\hat{\beta}_2) := \hat{\sigma}^2 \left(\sum_{t=1}^{T} (x_t - \bar{x})^2 \right)^{-1}$ for convenience.
- Then, $1-\alpha = P\left(-t_{T-2,1-\alpha/2} \leq Q(\beta_1 x, y) \leq t_{T-2,1-\alpha/2} \right)$

$$= P\left(-t_{T-2,1-\alpha/2} \leq \frac{\hat{\beta}_2 - \beta_2}{\hat{Var}(\hat{\beta})} \leq t_{T-2,1-\alpha/2} \right)$$

$$= P\left(-\hat{\beta}_2 - t_{T-2,1-\alpha/2} \hat{Var}(\hat{\beta}) \leq -\beta_2 \leq -\hat{\beta}_2 + t_{T-2,1-\alpha/2} \hat{Var}(\hat{\beta}_2) \right)$$

$$= P\left(\hat{\beta}_2 - t_{T-2,1-\frac{\alpha}{2}} \hat{Var}(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{T-2,1-\alpha/2} \hat{Var}(\hat{\beta}) \right)$$

serves as a basis to define a $100(1-\alpha)\%$ C.I for β_2 given by

$$\left[\hat{\beta}_2 - t_{T-2,1-\frac{\alpha}{2}} \hat{Var}(\hat{\beta}_2),\ \hat{\beta}_2 + t_{T-2,1-\frac{\alpha}{2}} \hat{Var}(\hat{\beta}_2) \right].$$

(b) Show that the two-tailed test of the hypothesis $\beta_2 = 0$ at significance level α will fail to reject if and only if zero lies inside the $(1-\alpha)$ confidence interval for β_2.

- A test of the hypotheses $H_0: \beta_2 = 0$ versus $H_1: \beta_2 \neq 0$ at the $100\alpha\%$ significance level, under the given modelling assumptions, would be designed (using your TA's favourite recipe) as per the next slide.

- Once we see how the decision rule is defined, we can analyse how to use the C.I. from part (a) to conduct the exact same test.

(b) Hypotheses: $H_0: \beta_2 = 0$
$H_1: \beta_2 \neq 0$

Test Statistic: $V = \hat{\beta}_2 \Big/ \sqrt{\hat{\sigma}^2 \sum_{t=1}^{T}(x_t - \bar{x})^2}$

Distribution of V under H_0: $V \sim t_{T-2}$

Significance level: α

Critical value: $V_{crit,1-\frac{\alpha}{2}} = t_{T-2,1-\frac{\alpha}{2}}$

Decision rule: Reject H_0 iff $|V| > V_{crit,1-\frac{\alpha}{2}}$

That is, Reject H_0 iff $\left| \hat{\beta}_2 \Big/ \sqrt{\hat{\sigma}^2 / \sum_{t=1}^{T}(x_t - \bar{x})^2} \right| > t_{T-2,1-\frac{\alpha}{2}}$

i.e. $-t_{T-2,1-\frac{\alpha}{2}} < \dfrac{\hat{\beta}_2 - 0}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^{T}(x_t - \bar{x})^2}} < t_{T-2,1-\frac{\alpha}{2}}$ $\iff$ we fail-to-reject H_0

or indeed if the interval $\hat{\beta}_2 \pm t_{T-2,1-\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2 \sum_{t=1}^{T}(x_t - \bar{x})^2}$ contains 0, we fail-to-reject H_0 (and vice versa),

...but this is exactly the $100(1-\alpha)\%$ C.I. for β_2 (as derived in part (a))!

QUESTION 3

Notes on Delta method.

2018

let $y = X\beta + \varepsilon$ where $\underset{N\times K}{X} = [x_1\ x_2 \dots x_K]$ and $\hat{\beta} := (X'X)^{-1}X'y$

Let's assume: A1, A2, A3R SRU, A4GM.

NON-LINEAR RESTRICTIONS RECAP:

- let $g(\beta) \in \mathbb{R}^r$ be a set of r non-linear and continuously differentiable functions of $\beta \in \mathbb{R}^K$.

Say we wish to test $H_0: g(\beta) = 0$
$H_1: g(\beta) \neq 0$

- Using the WALD PRINCIPLE of testing based on $\hat{\beta} := (X'X)^{-1}X'y$ (i.e. from the unrestricted model), we would ideally wish to construct a statistic of the form:

$$[g(\hat{\beta})]' \left[\hat{Var}(g(\hat{\beta}))\right]^{-1} [g(\hat{\beta})]$$

as the non-linear analogue to the usual $(R\hat{\beta}-q)' [R\, \hat{Var}(\hat{\beta}) R']^{-1} (R\hat{\beta}-q) \overset{approx.}{\sim} \chi^2_r$ statistic as in the case of r linear restrictions contained in the matrix R.

The problem is that we need $\hat{var}(g(\hat{\beta}))$ where $g(\cdot)$ is a NON-LINEAR function.
(i.e. you can't just "pull out" the "R" happily "with a square"!)

2018

- We use the result (from the "delta method") that:

For "large" N (under A1, A2, A3R sru, A4GM),

$$g(\hat{\beta}) \overset{\text{approx}}{\sim} N\left(g(\beta), \underbrace{[\partial g(\beta)/\partial \beta']}_{\text{Kinda like an "R"}} Var(\hat{\beta}) \underbrace{[\partial g(\beta)/\partial \beta']'}_{\text{..."R'"}} \right).$$

- We can proceed then with test statistic V:

$$V := g(\hat{\beta})' \left[[\partial g(\hat{\beta})/\partial \hat{\beta}'] [\hat{\sigma}^2 (X'X)^{-1}] [\partial g(\hat{\beta})/\partial \hat{\beta}']' \right]^{-1} g(\hat{\beta})$$

Under H_0: $g(\beta) = 0$, $V \overset{\text{approx.}}{\sim} \chi^2_r$ for "large" N (assuming A1, A2, A3R sru, A4GM).

- For Q[illegible], we have in particular:

$$g(\beta) := \begin{bmatrix} \beta_1 \beta_2 - 1 \\ \beta_3 - 4\beta_4 + 2 \end{bmatrix} \quad \text{and} \quad \partial g(\beta)/\partial \beta' := \begin{bmatrix} \beta_2 & \beta_1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 \end{bmatrix}$$

Thus,

$$V = [\hat{\beta}_1 \hat{\beta}_2 - 1, \hat{\beta}_3 - 4\hat{\beta}_4 + 2] \left[\begin{pmatrix} \hat{\beta}_2 & \hat{\beta}_1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 \end{pmatrix} \hat{\sigma}^2 (X'X)^{-1} \begin{pmatrix} \hat{\beta}_2 & 0 \\ \hat{\beta}_1 & 0 \\ 0 & 1 \\ 0 & -4 \\ 0 & 0 \end{pmatrix} \right]^{-1} \begin{bmatrix} \hat{\beta}_1 \hat{\beta}_2 - 1 \\ \hat{\beta}_3 - 4\hat{\beta}_4 + 2 \end{bmatrix}$$

$\overset{\text{approx}}{\sim} \chi^2_2$ under H_0 for "large N" (assuming A1, A2, A3R sru, A4GM)