16.4 Asymptotic normality of LSE under heteroscedasticity

Repetition

(A0) $(Y_i, X_i^T)^T \overset{iid}{\sim} (Y, X^T)^T$, $E(Y|X) = X^T\beta$

(A1) $E|X_j X_l| < \infty$, $EXX^T = W > 0$

elements of $X = (X_0, \dots, X_{k-1})^T$ $\quad V := W^{-1} = (E(XX^T))^{-1}$

(A2 homoscedastic)

$$\sigma^2(X) := var(Y|X) = var(\varepsilon|X) = \sigma^2 < \infty$$

$$\Rightarrow var\,\varepsilon = \sigma^2$$

$$\mathbb{Y}_n = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbb{X}_n = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix}$$

$$\mathbb{X}_n^T \mathbb{X}_n = \sum_{i=1}^n X_i X_i^T$$

$$\mathbb{X}_n^T \mathbb{Y}_n = \sum_{i=1}^n X_i Y_i$$

$$\hat\beta_n := (\mathbb{X}_n^T \mathbb{X}_n)^{-1} \mathbb{X}_n^T \mathbb{Y}_n$$

$n \to \infty$

$$(A0) \,\&\, (A1) \Rightarrow \frac{1}{n} \mathbb{X}_n^T \mathbb{X}_n \xrightarrow{a.s.} W = EXX^T$$

$$n \cdot (\mathbb{X}_n^T \mathbb{X}_n)^{-1} \xrightarrow{a.s.} V = (EXX^T)^{-1}$$

& (A2 $\overset{\text{homo}}{\text{hetero}}$scedastic) $\Rightarrow \hat\beta_n \xrightarrow{a.s.} \beta$

& (A2 homoscedastic)
& $E|\varepsilon^2 X_j X_l| < \infty \;\forall j,l$ $\Rightarrow \sqrt{n}(\hat\beta_n - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 V)$

(A2 heteroscedastic), $\sigma^2(X) = var(Y|X) = var(\varepsilon|X)$

$\mathbb{E}\{\sigma^2(X)\} < \infty$, $\quad \mathbb{E}\{\sigma^2(X) X_j X_\ell\} < \infty \quad \forall j, \ell$

$$W^* := \mathbb{E}\{\sigma^2(X) X X^T\}$$

Theorem 16.5 Asymptotic normality of LSE in heteroscedastic case 12

Let assumptions (A0), (A1), (A2 heteroscedastic) hold.
Further, let $\mathbb{E}|\varepsilon^2 X_j X_\ell| < \infty \quad \forall j, \ell = 0, \dots, k-1$.

Then
$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}_k(0, V W^* V), \quad n \to \infty$$

$$\hat{\theta}_n = \ell^T \hat{\beta}_n, \quad \sqrt{n}(\hat{\theta}_n - \underbrace{\ell^T \beta}_{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}_1(0, \ell^T V W^* V \ell), \quad n \to \infty$$

$$\hat{\xi}_n = L \hat{\beta}_n, \quad \sqrt{n}(\hat{\xi}_n - \underbrace{L \beta}_{\xi}) \xrightarrow{\mathcal{D}} \mathcal{N}_m(0, L V W^* V L^T), \quad n \to \infty$$

Compare:

	homoscedastic	heteroscedastic
$var(Y\|X) = var(\varepsilon\|X)$	σ^2	$\sigma^2(X)$
Covariance matrix (asymptotic) for $\sqrt{n}(\hat{\beta}_n - \beta)$	$\sigma^2 V = \sigma^2 (EXX^T)^{-1}$	$VW^*V = (EXX^T)^{-1} E\{\sigma^2(X) XX^T\} (EXX^T)^{-1}$

$$\underbrace{(EXX^T)^{-1} E\{\sigma^2(X) XX^T\} (EXX^T)^{-1}}_{=}$$

if homoscedasticity

$(EXX^T)^{-1} \sigma^2 EXX^T \cdot (EXX^T)^{-1}$

$= \sigma^2 (EXX^T)^{-1} = \sigma^2 V$

Hence proof of Theorem 16.5 also proves Theorem 16.4.

$$V W^{*} V = \underbrace{(\mathbb{E} X X^T)^{-1}}_{\text{bread}} \underbrace{\mathbb{E}\{\sigma^2(X) X X^T\}}_{\text{MEAT}} \cdot \underbrace{(\mathbb{E} X X^T)^{-1}}_{\text{bread}}$$

$\equiv$ (theoretical) sandwich

practically eatable sandwich will follow

Proof:

$$\hat{\beta}_n = \underbrace{(X_n^T X_n)^{-1}}_{V_n} X_n^T Y_n = V_n \sum_{i=1}^{n} X_i Y_i = V_n \sum_{i=1}^{n} X_i (X_i^T \beta + \varepsilon_i)$$

$$= V_n \underbrace{\left(\sum_{i=1}^{n} X_i X_i^T\right)}_{V_n^{-1}} \beta + V_n \sum_{i=1}^{n} X_i \varepsilon_i = \beta + V_n \sum_{i=1}^{n} X_i \varepsilon_i$$

Pretend first that we do not know the rate of convergence ($\sqrt{n}$):

$$\hat{\beta}_n - \beta = V_n \sum_{i=1}^{n} X_i \varepsilon_i = \underbrace{n V_n}_{\downarrow \text{ a.s.}} \overbrace{\frac{1}{n} \sum_{i=1}^{n} X_i \varepsilon_i}^{\text{sample mean of iid random vectors, CLT ?}}$$

$\downarrow$ a.s. (in P would be sufficient, target is to use Cramer-Слуцкий)

W

Let us try to apply CLT on $\frac{1}{n} \sum_{i=1}^{n} X_i \varepsilon_i$

- $X_i \varepsilon_i$, $i = 1, 2, \ldots$ are iid (from assumpt. A0)
- $\mathbb{E} X_i \varepsilon_i = 0 \quad \forall i$ (see proof of theorem 16.2)
- $\operatorname{var} X_i \varepsilon_i < \infty$?

$var(X_i \varepsilon_i)$ = matrix with elements $var(X_{ij}\varepsilon_i)$ and $cov(X_{ij}\varepsilon_i, X_{il}\varepsilon_i)$.

all those elements are $<\infty$ because we are assuming $\mathbb{E}|\varepsilon^2 X_j X_l| < \infty \ \forall j, l$.

To apply CLT, we need $var(X_i \varepsilon_i)$:

$$var(X_i\varepsilon_i) = \mathbb{E}(var(X_i\varepsilon_i | X_i)) + var(\mathbb{E}(X_i\varepsilon_i | X_i))$$

already know that $<\infty$

$$= \mathbb{E}\Big(X_i \underbrace{var(\varepsilon_i | X_i)}_{\sigma^2(X_i)} X_i^T\Big) + var\Big(X_i \underbrace{\mathbb{E}(\varepsilon_i | X_i)}_{=0}\Big) =$$

$$= \mathbb{E}(\sigma^2(X_i) X_i X_i^T) = \mathbb{E}(\sigma^2(X) X X^T) = W^*$$

(if homoscedasticity $= \sigma^2 \mathbb{E} X X^T = \sigma^2 W$)

CLT (iid, multivariate, Theorem C.5)

$$\Longrightarrow \sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^n X_i \varepsilon_i = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \varepsilon_i \xrightarrow[n\to\infty]{\mathcal{D}} N(0, W^*)$$

We have (previous page)

$$(\hat{\beta}_n - \beta) = \underbrace{n \cdot V_n}_{\downarrow P \ V (=\mathbb{E}XX^T)^{-1}} \cdot \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\varepsilon_i}_{\downarrow \mathcal{D} \ N(0, W^*)} \cdot \frac{1}{\sqrt{n}}$$

$W^* = \mathbb{E}\{\sigma^2(X) X X^T\}$

Cramér-Слуцкий

$$\Longrightarrow \sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{\mathcal{D}} N(0, \underbrace{V W^* V^T}_{(\mathbb{E}XX^T)^{-1}\mathbb{E}\{\sigma^2(X)XX^T\}(\mathbb{E}XX^T)^{-1}})$$

Asymptotic normality of $l^T\hat{\beta}_n$ and $L\hat{\beta}_n$ is justified by Cramér-Wold theorem (Theorem C.6)

Summary & repetition

$(A0), (A1), \quad \mathbb{E}|\varepsilon^2 X_j X_l| < \infty \quad \forall j, l$

(A2 homoscedastic)

$var(Y|X) = var(\varepsilon|X) = \sigma^2 < \infty$

$W = \mathbb{E}XX^T$

$V = (\mathbb{E}XX^T)^{-1}$

$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}_k(0, \sigma^2 V)$

$\hat{\beta}_n \overset{\cdot}{\sim} \mathcal{N}_k(\beta, \underbrace{\sigma^2 \tfrac{1}{n} V}_{\approx MSe_{,n} \cdot (\mathbb{X}_n^T \mathbb{X}_n)^{-1}})$

since $\underbrace{MSe_{,n} \cdot n \cdot (\mathbb{X}_n^T \mathbb{X}_n)^{-1}}$ $\downarrow P$ $\sigma^2 V$

Cramér-Slutsky $\Rightarrow$ $\forall l \in \mathbb{R}^k$

$$\underbrace{\frac{l^T\hat{\beta}_n - l^T\beta}{\sqrt{MSe_{,n}\, l^T(\mathbb{X}_n^T\mathbb{X}_n)^{-1} l}}}_{\sim t_{n-k} \text{ (if normality)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

$\mathbb{L}_{m \times k}$ with linear. indep. rows

$$\underbrace{(\mathbb{L}\hat{\beta}_n - \mathbb{L}\beta)^T \{MSe_{,n} \mathbb{L}(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\mathbb{L}^T\}^{-1}(\mathbb{L}\hat{\beta}_n - \mathbb{L}\beta)}_{\frac{1}{m}\,\cdot\, \sim F_{m,n-k} \text{ (if normality)}} \xrightarrow{\mathcal{D}} \chi^2_m$$

(A2 heteroscedastic)

$var(Y|X) = var(\varepsilon|X) = \sigma^2(X)$

$\underbrace{\mathbb{E}\{\sigma^2(X) X X^T\}}_{W^*} < \infty$

$\xrightarrow{\mathcal{D}} \mathcal{N}_k(0, V W^* V)$

$\overset{\cdot}{\sim} \mathcal{N}_k(\beta, \tfrac{1}{n} V W^* V)$

$n \cdot \boxed{\ ?\ } \xrightarrow{P} V W^* V$

Cramér-Slutsky $\Rightarrow$ $\forall l \in \mathbb{R}^k$

$$\frac{l^T\hat{\beta}_n - l^T\beta}{\sqrt{l^T \boxed{\ \ } l}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

$\mathbb{L}_{m \times k}$ with linear. indep. rows

$$(\mathbb{L}\hat{\beta}_n - \mathbb{L}\beta)^T \{\mathbb{L} \boxed{\ \ } \mathbb{L}^T\}^{-1} (\mathbb{L}\hat{\beta}_n - \mathbb{L}\beta) \xrightarrow{\mathcal{D}} \chi^2_m$$

quantities related to model $M_n: Y_n | X_n \sim (X_n \beta, \sigma^2 I_n)$ 19

- $H_n = X_n (X_n^T X_n)^{-1} X_n = \begin{pmatrix} h_{n1} & & * \\ & \ddots & \\ * & & h_{nn} \end{pmatrix}$
- $M_n = I_n - H_n$

$$= \begin{pmatrix} m_{n,1} & & * \\ & \ddots & \\ * & & m_{n,n} \end{pmatrix}$$

- $U_n = M_n Y_n = (U_{n,1}, \dots, U_{n,n})^T$ (residuals)

We look for $\boxed{W_n}$ such that 20

$$n \boxed{W_n} \xrightarrow[a.s.]{P} V W^* V = (\mathbb{E} X X^T)^{-1} (\mathbb{E} \sigma^2(X) X X^T) (\mathbb{E} X X^T)^{-1}$$

$$\underbrace{n (X_n^T X_n)^{-1}}_{n \cdot V_n} \xrightarrow{a.s.} \quad \frac{1}{n} \boxed{W_n} \xrightarrow{a.s.} \quad \underbrace{n (X_n^T X_n)^{-1}}_{n \cdot V_n} \xrightarrow{a.s.}$$

Theorem 16.6 Sandwich estimator of the covariance matrix

Let assumptions (A0), (A1), (A2 heteroscedastic) hold.
Let additionally, for each $s, t, j, l = 0, \dots, k-1$
$\mathbb{E} |\varepsilon^2 X_j X_l| < \infty$, $\mathbb{E} |\varepsilon X_s X_j X_l| < \infty$, $\mathbb{E} |X_s X_t X_j X_l| < \infty$.

Then $n V_n W_n^* V_n \xrightarrow{a.s.} V W^* V$ as $n \to \infty$,

where for $n = 1, 2, \dots$ $W_n^* = \sum_{i=1}^{n} U_{n,i}^2 X_i X_i^T = X_n^T \Omega_n X_n$,

$\Omega_n = \mathrm{diag}(\omega_{n,1}, \dots, \omega_{n,n})$, $\omega_{n,i} = U_{n,i}^2$,
$i = 1, 2, \dots, n$.

Proof: $V W^* V = (\mathbb{E} XX^T)^{-1} \, \mathbb{E}\{\sigma^2(X) XX^T\} \, (\mathbb{E} XX^T)^{-1}$

$$n V_n W_n^* V_n = \underbrace{n V_n}_{\uparrow \text{a.s.}} \cdot \frac{1}{n} W_n^* \cdot \underbrace{n V_n}_{\uparrow \text{a.s.}} \quad \text{(Lemma 16.1)}$$

It remains to be shown that

$$\frac{1}{n} W_n^* = \frac{1}{n} \sum_{i=1}^{n} U_{n,i}^2 X_i X_i^T \xrightarrow{a.s.} W^* = \mathbb{E}\{\sigma^2(X) XX^T\}$$

Remember $\sigma^2(X) = \mathrm{var}(\varepsilon | X) = \mathbb{E}(\varepsilon^2 | X)$ (since $\mathbb{E}(\varepsilon | X) = 0$)

Perhaps $\mathbb{E}(\varepsilon^2 XX^T) = W^*$ in which case $\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 X_i X_i^T \xrightarrow{a.s.} W^*$ by SLLN?

Since $U_{n,i}^2 \approx \varepsilon_i^2$, perhaps also

$$\frac{1}{n} \sum_{i=1}^{n} U_{n,i}^2 X_i X_i^T \xrightarrow{a.s.} W^* \ ?$$

Let us first apply SLLN on a sequence $\varepsilon_i^2 X_i X_i^T$, $i = 1, 2, \ldots$

- $\varepsilon_i^2 X_i X_i^T$ are iid random vectors (matrices)
- For each element of $\varepsilon^2 XX^T$ $(\forall j, l)$

  $\mathbb{E} | \varepsilon^2 X_j X_l | < \infty$ (is assumed)
- $\mathbb{E}(\varepsilon^2 X_j X_l) = \mathbb{E}(\mathbb{E}(\varepsilon^2 X_j X_l | X)) =$

$$= \mathbb{E}\big(X_j X_l \underbrace{\mathbb{E}(\varepsilon^2 | X)}_{= \mathrm{var}(\varepsilon | X)}\big) = \mathbb{E}(\sigma^2(X) X_j X_l) \quad \forall j, l$$

Hence $\mathbb{E}(\varepsilon^2 XX^T) = \mathbb{E}(\sigma^2(X) XX^T)$.

<u>SLLN</u>: $\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 X_i X_i^T \xrightarrow[n\to\infty]{a.s.} \underbrace{\mathbb{E}\{\sigma^2(X)\, X X^T\}}_{= W^*}$

Can we replace (unobservable) ε_i^2 $(i=1,2,\ldots)$ by $U_{n,i}^2 = (Y_i - X_i^T \hat{\beta}_n)^2$?

<u>Let's see</u>:

$$\frac{1}{n}\sum_{i=1}^{n} U_{n,i}^2 X_i X_i^T = \frac{1}{n}\sum_{i=1}^{n} (Y_i - X_i^T\hat{\beta}_n)^2 X_i X_i^T =$$

$$= \frac{1}{n}\sum_{i=1}^{n} (\underbrace{Y_i - X_i^T\beta}_{\varepsilon_i} + \underbrace{X_i^T\beta - X_i^T\hat{\beta}_n}_{X_i^T(\beta-\hat{\beta}_n)})^2 X_i X_i^T =$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 \underbrace{X_i X_i^T}_{\text{matrix}}}_{A_n} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} \{\underbrace{(\beta-\hat{\beta}_n)^T X_i X_i^T (\beta-\hat{\beta}_n)}_{\text{scalar}}\} \underbrace{X_i X_i^T}_{\text{matrix}}}_{B_n}$$

$$+ \underbrace{\frac{2}{n}\sum_{i=1}^{n} \{\underbrace{(\beta-\hat{\beta}_n)^T X_i \varepsilon_i}_{\text{scalar}}\} \underbrace{X_i X_i^T}_{\text{matrix}}}_{C_n}$$

We know: $A_n = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 X_i X_i^T \xrightarrow{a.s.} \mathbb{E}\{\sigma^2(X)\, X X^T\} = W^*$

Hopefully: $B_n \xrightarrow{a.s.} 0 \quad , n\to\infty$

$C_n \xrightarrow{a.s.} 0 \quad , n\to\infty$

$$B_n = \frac{1}{n}\sum_{i=1}^{n} \underbrace{(\beta-\hat{\beta}_n)^T X_i}_{\text{scalar}}\ \underbrace{X_i^T(\beta-\hat{\beta}_n)}_{\text{scalar}}\ \underbrace{X_i X_i^T}_{\text{matrix}} =$$

$$= \frac{1}{n}\sum_{i=1}^{n} \underbrace{\{(\beta-\hat{\beta}_n)^T X_i\}}_{\text{scalar}}\ \underbrace{X_i X_i^T}_{\text{matrix}}\ \underbrace{\{X_i^T(\beta-\hat{\beta}_n)\}}_{\text{scalar}}$$

element (j,l) of matrix B_n:

$$B_n(j,l) = \frac{1}{n}\sum_{i=1}^{n} (\beta-\hat{\beta}_n)^T X_i\ \underbrace{X_{ij} X_{il}}_{\text{scalar}}\ X_i^T(\beta-\hat{\beta}_n) =$$

$$= \underbrace{(\beta-\hat{\beta}_n)^T}_{\xrightarrow{a.s.}\ 0} \Big\{ \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_{ij} X_{il} X_i X_i^T}_{\xrightarrow{a.s.}} \Big\} \underbrace{(\beta-\hat{\beta}_n)}_{\xrightarrow{a.s.}\ 0}$$

some matrix with finite elements since we assume $E|X_j X_l X_s X_t| < \infty$

$$\Rightarrow B_n \xrightarrow{a.s.} 0, \quad n\to\infty$$

$$C_n = \frac{2}{n}\sum_{i=1}^{n} \underbrace{(\beta-\hat{\beta}_n)^T X_i \varepsilon_i}_{\text{scalar}}\ \underbrace{X_i X_i^T}_{\text{matrix}}$$

element (j,l) of matrix C_n:

$$C_n(j,l) = \frac{2}{n}\sum_{i=1}^{n} (\beta-\hat{\beta}_n)^T X_i \varepsilon_i X_{ij} X_{il} =$$

$$= 2\cdot \underbrace{(\beta-\hat{\beta}_n)^T}_{\xrightarrow{a.s.}\ 0} \Big\{ \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i \varepsilon_i X_{ij} X_{il}}_{\xrightarrow{a.s.}} \Big\}$$

some vector with finite elements since we assume $E|\varepsilon X_s X_j X_l| < \infty$

$$\Rightarrow C_n \xrightarrow{a.s.} 0, \quad n\to\infty$$

In summary:

$$n V_n W_n^* V_n = n\left(X_n^T X_n\right)^{-1} \cdot \frac{1}{n}\sum_{i=1}^{n} U_{n,i}^2 X_i X_i^T \cdot n\left(X_n^T X_n\right)^{-1},$$

where $n\left(X_n^T X_n\right)^{-1} \xrightarrow{a.s.} \mathbb{V} = \left(\mathbb{E}\, X X^T\right)^{-1}$

$$\frac{1}{n}\sum_{i=1}^{n} U_{n,i}^2 X_i X_i^T = \underbrace{A_n}_{\downarrow a.s.} + \underbrace{B_n}_{\xrightarrow{a.s.} 0} + \underbrace{C_n}_{\xrightarrow{a.s.} 0} \xrightarrow{a.s.} W^*$$

$$W^* = \mathbb{E}\left\{\sigma^2(X) X X^T\right\}$$

$$\Rightarrow n V_n W_n^* V_n \xrightarrow{a.s.} \mathbb{V} \cdot W^* \cdot \mathbb{V}$$

$$V_n W_n^* V_n = \underbrace{\left(X_n^T X_n\right)^{-1} X_n^T}_{\text{bread}} \underbrace{\Omega_n}_{\text{MEAT}} \underbrace{X_n \left(X_n^T X_n\right)^{-1}}_{\text{bread}}$$ [21]

$$\Omega_n = \operatorname{diag}\left(U_{n,1}^2, \dots, U_{n,n}^2\right)$$

For info only: $U_{n,i}^2$ is replacing ε_i^2,

- with homoscedasticity: $\operatorname{var}(\varepsilon_i \mid X_n) = \mathbb{E}(\varepsilon_i^2 \mid X_n) = \sigma^2$,

whereas $\operatorname{var}(U_{n,i} \mid X_n) = \mathbb{E}(U_{n,i}^2 \mid X_n) = \sigma^2 m_{n,i}$

$\rightarrow$ better properties of sandwich ($\equiv$ faster convergence)

with $\Omega_n = \operatorname{diag}\left(\dfrac{n}{\nu_n} \cdot \dfrac{U_{n,i}^2}{m_{n,i}^{\delta_{n,i}}},\ i = 1, \dots, n\right)$

- $\delta_n = (\delta_{n,1}, \dots, \delta_{n,n})^T$: suitable seq.
- $\nu_1, \nu_2, \dots$ sequence such that $\dfrac{\nu_n}{n} \to 1,\ n \to \infty$

still for info only:

22

Alternative sorts of meat for sandwich

→ see slide

→ labels $\underbrace{HC}$0, ..., HC4

heteroscedasticity consistent (estimator)

are used in the R package sandwich

Repetition, with HOMOscedasticity

23

$\theta = \ell^T \beta$, $\hat{\theta}_n = \ell^T \hat{\beta}_n$

$$T_n = \frac{\ell^T \hat{\beta}_n - \ell^T \beta}{\sqrt{\ell^T (MSe_{,n} (X_n^T X_n)^{-1}) \ell}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

$\sim t_{n-k}$ (under normality)

$$MSe_{,n} (X_n^T X_n)^{-1} = \widehat{var}(\hat{\beta}_n | X_n)$$

such that $n \cdot MSe_{,n} (X_n^T X_n)^{-1} \xrightarrow[P]{a.s.} \underbrace{\sigma^2 V}$

asymptotic covariance m. of $\sqrt{n}(\hat{\beta}_n - \beta)$

---

with HETEROscedasticity

$$MSe_{,n} (X_n^T X_n)^{-1} \xrightarrow{\text{replace}} \underbrace{(X_n^T X_n)^{-1} X_n^T \Omega_n^{HC} X_n (X_n^T X_n)^{-1}}_{V_n^{HC}}$$

such that

$$n \cdot V_n^{HC} \xrightarrow[P]{a.s.} \underbrace{V W^* V}$$

asymptotic covariance matrix of $\sqrt{n}(\hat{\beta}_n - \beta)$

→ statistic T_n^{HC}

$T_n^{HC} = T_n$ where $MSE_n (X_n^T X_n)^{-1}$ replaced by sandwich $V_n^{HC} = (X_n^T X_n)^{-1} X_n^T \Omega_n^{HC} X_n (X_n^T X_n)^{-1}$

$$T_n^{HC} = \frac{l^T \hat{\beta}_n - l^T \beta}{\sqrt{l^T V_n^{HC} l}} \xrightarrow[?]{\mathcal{D}} \mathcal{N}(0,1)$$

$$T_n^{HC} = \underbrace{\frac{\sqrt{n}(l^T \hat{\beta}_n - l^T \beta)}{\sqrt{l^T V W^+ V l}}}_{\downarrow \mathcal{D} \; \mathcal{N}(0,1)} \cdot \underbrace{\sqrt{\frac{l^T V W^+ V l}{n \cdot l^T V_n^{HC} l}}}_{\downarrow a.s. (P) \; 1}$$

Cramér-Слуцкий $\Rightarrow$ $T_n^{HC} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$ $\boxed{24}$

again $T_\nu \sim t_\nu$, $T_\nu \xrightarrow[\nu\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$ $\boxed{25}$

Hence with

$I_n^N := \left(l^T \hat{\beta}_n \pm u(1-\frac{\alpha}{2}) \sqrt{l^T V_n^{HC} l}\right)$

$I_n^t := \left(l^T \hat{\beta}_n \pm t_{n-k}(1-\frac{\alpha}{2}) \sqrt{l^T V_n^{HC} l}\right)$

$\forall \theta^0 \in \mathbb{R}$, $\theta^0 = l^T \beta^0$, $\theta = l^T \beta$
$(\forall \beta^0 \in \mathbb{R}^k)$

$P(I_n^N \ni \theta^0;\ \theta = \theta^0) \to 1-\alpha$, $n\to\infty$

$P(I_n^t \ni \theta^0;\ \theta = \theta^0) \longrightarrow 1-\alpha$, $n\to\infty$

Repetition, with HOMOscedasticity

$\xi = \mathbb{L}\beta$, $\mathbb{L}_{m\times k}$ matrix with linearly indep. rows, $m \le k$

$\hat{\xi}_n = \mathbb{L}\hat{\beta}_n$

$$Q_n = \frac{1}{m}(\mathbb{L}\hat{\beta}_n - \mathbb{L}\beta)^T\{MSe_n \mathbb{L}(X_n^T X_n)^{-1}\mathbb{L}^T\}^{-1}(\mathbb{L}\hat{\beta}_n - \mathbb{L}\beta)$$

$\sim F_{m,n-k}$ (under normality)

$$m \cdot Q_n \xrightarrow{\mathcal{D}} \chi^2_m$$

$$MSe_n (X_n^T X_n)^{-1} = \widehat{var}(\hat{\beta}_n | X)$$

such that $n \, MSe_n (X_n^T X_n)^{-1} \xrightarrow[P]{a.s.} \underbrace{\sigma^2 V}_{\text{asymptotic covariance m. of } \sqrt{n}(\hat{\beta}_n - \beta)}$

---

with HETEROscedasticity

$MSe_n (X_n^T X_n)^{-1}$ $\xrightarrow{\text{replace}}$ $\underbrace{(X_n^T X_n)^{-1} X_n^T \Omega_n^{HC} X_n (X_n^T X_n)^{-1}}_{V_n^{HC}}$

such that $n \cdot V_n^{HC} \xrightarrow[P]{a.s.} \underbrace{V W^* V}_{\text{asymptotic covariance matrix of } \sqrt{n}(\hat{\beta}_n - \beta)}$

$\xrightarrow{\text{statistic}}$ Q_n^{HC}

$Q_n^{HC} = Q_n$ where $MSe_n (X_n^T X_n)^{-1}$ replaced by sandwich $V_n^{HC} = (X_n^T X_n)^{-1} X_n^T \Omega_n X_n (X_n^T X_n)^{-1}$

$$Q_n^{HC} = \frac{1}{m} (L\hat{\beta}_n - L\beta)^T (L V_n^{HC} L^T)^{-1} (L\hat{\beta}_n - L\beta)$$

$$m \cdot Q_n^{HC} \xrightarrow[?]{\mathcal{D}} \chi^2_m$$

$$m \cdot Q_n^{HC} = \underbrace{\sqrt{n} (L\hat{\beta}_n - L\beta)^T}_{\xrightarrow{\mathcal{D}} N(0,\, LVW^+VL^T)} \underbrace{(L\, n V_n^{HC} L^T)^{-1}}_{\xrightarrow[(P)]{a.s.} (LVW^+VL^T)^{-1}} \underbrace{(L\hat{\beta}_n - L\beta)\sqrt{n}}_{\xrightarrow{\mathcal{D}} N(0,\, LVW^+VL^T)}$$

Cramér-Слуцкий $\Rightarrow$

$$m \cdot Q_n^{HC} \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_m$$ [24]

again $Q_\nu \sim F_{m,\nu}$, $m \cdot Q_\nu \xrightarrow[\nu\to\infty]{\mathcal{D}} \chi^2_m$ [26]

Hence with

$$K_n^{\chi} = \{ \xi \in \mathbb{R}^m : (\xi - L\hat{\beta}_n)^T (L V_n^{HC} L^T)^{-1} (\xi - L\hat{\beta}_n) < \chi^2_m(1-\alpha) \}$$

$$K_n^F = \{ \xi \in \mathbb{R}^m : (\xi - L\hat{\beta}_n)^T (L V_n^{HC} L^T)^{-1} (\xi - L\hat{\beta}_n) < m \cdot F_{m,n-k}(1-\alpha) \}$$

$\forall \xi^0 \in \mathbb{R}^m$, $\xi^0 = L\beta^0$, $\xi = L\beta$
$(\forall \beta^0 \in \mathbb{R}^k)$

$$P(K_n^{\chi} \ni \xi^0;\ \xi = \xi^0) \longrightarrow 1-\alpha, \quad n\to\infty$$

$$P(K_n^{F} \ni \xi^0;\ \xi = \xi^0) \longrightarrow 1-\alpha, \quad n\to\infty.$$