XVI. Asymptotic Properties of the LSE and Sandwich Estimator

0

16.1 Assumptions and setup

1

2

(A0) • $(Y_1, X_1^T)^T, (Y_2, X_2^T)^T, \dots \overset{iid}{\sim} (Y, X^T)^T$

$X_i = (X_{i0}, \dots, X_{ik-1})^T$ $\qquad X = (X_0, \dots, X_{k-1})^T$

• $\beta = (\beta_0, \dots, \beta_{k-1})^T \in \mathbb{R}^k$: unknown parameter

• $\mathbb{E}(Y|X) = X^T\beta \qquad (\mathbb{E}(Y_i|X_i) = X_i^T\beta \ \forall i)$

$\varepsilon := Y - X^T\beta$ generic error term

$\varepsilon_i := Y_i - X_i^T\beta \quad , i = 1, 2, \dots$

Notes

• $\varepsilon_1, \varepsilon_2, \dots$ are iid $\sim \varepsilon$

• $\mathbb{E}(\varepsilon|X) = \mathbb{E}(Y - X^T\beta|X) = 0$

$\Rightarrow \mathbb{E}(\varepsilon) = \mathbb{E}(\mathbb{E}(\varepsilon|X)) = \mathbb{E}(0) = 0$

• $\mathrm{var}(\varepsilon|X) = \mathrm{var}(Y - X^T\beta|X) = \mathrm{var}(Y|X) =: \sigma^2(X)$

$\Rightarrow \mathrm{var}(\varepsilon) = \mathbb{E}(\mathrm{var}(\varepsilon|X)) + \mathrm{var}(\mathbb{E}(\varepsilon|X)) =$

$= \mathbb{E}(\sigma^2(X)) + \mathrm{var}\,0 = \mathbb{E}(\sigma^2(X))$

$\mathbb{E}$ w.r.t. distribution of X

(A1) The covariate vector $X = (X_0, \dots, X_{k-1})^T$ ~~stati~~ satisfies [3]

- $\mathbb{E}|X_j X_l| < \infty$, $j,l = 0, \dots, k-1$
- $\mathbb{E}(XX^T) = W$, W: positive definite matrix

- $W = (w_{j,l})_{j,l=0,\dots,k-1}$, $w_j^2 := w_{j,j} = \mathbb{E}X_j^2$, $j = 0, \dots, k-1$

  $w_{j,l} = \mathbb{E}X_j X_l$, $j \neq l$

- $V := W^{-1} = (v_{j,l})_{j,l=0,\dots,k-1}$

---

[4]

NOTATION: data of size $n \geq 1$:

$Y_n = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, $\mathbb{X}_n = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix}$ $W_n := \mathbb{X}_n^T \mathbb{X}_n = \sum_{i=1}^{n} X_i X_i^T$

$V_n := (\mathbb{X}_n^T \mathbb{X}_n)^{-1}$ (if it exits)

→ ASYMPTOTIC THEOREMS (well known...) in Appendix

Lemma 16.1 Consistent estimator of the second and first mixed moments of the covariates

Let assumptions (A0) and (A1) hold. Then

$$\frac{1}{n}\mathbb{W}_n \xrightarrow{a.s.} \mathbb{W} \quad \text{as } n\to\infty$$

$$n\mathbb{V}_n \xrightarrow{a.s.} \mathbb{V} \quad \text{as } n\to\infty.$$

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

$$\frac{1}{n}\mathbb{W}_n = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T \underset{?}{\xrightarrow{a.s.}} \mathbb{W} = \mathbb{E}XX^T$$

Element (j,l) of above matrices:

$$\frac{1}{n}\sum_{i=1}^{n} X_{ij}X_{il} \underset{?}{\xrightarrow{a.s.}} w_{j,l} = \mathbb{E}X_jX_l$$

SLLN (Theorem C.2): $? \iff \mathbb{E}|X_jX_l| < \infty$ (which is assumed by A1)

Hence indeed $\frac{1}{n}\mathbb{W}_n \xrightarrow{a.s.} \mathbb{W}$ as $n\to\infty$.

$\mathbb{W} > 0$ and hence with probab. 1 $\exists n_0\ \forall n \geq n_0\ \mathbb{W}_n > 0$

$$\Rightarrow n\mathbb{W}_n^{-1} = n\mathbb{V}_n \xrightarrow{a.s.} \mathbb{V} \quad \text{as } n\to\infty$$

$$n(\mathbb{X}_n^T\mathbb{X}_n)^{-1} \qquad = (\mathbb{E}XX^T)^{-1}$$

$$\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T \xrightarrow[n\to\infty]{a.s.} \mathbb{E}XX^T = W > 0$$

That is, $P(\exists n_0 > k \ \forall n \geq n_0 \ \text{rank}(\mathbb{X}_n) = k) = 1$.

for $n \geq n_0$ $\hat{\beta}_n := (\mathbb{X}_n^T\mathbb{X}_n)^{-1}(\mathbb{X}_n^T\mathbb{Y}_n) = \left(\sum_{i=1}^{n} X_i X_i^T\right)^{-1}\left(\sum_{i=1}^{n} X_i Y_i\right)$

$$MSe_n := \frac{1}{n-k}\|\mathbb{Y}_n - \mathbb{X}_n\hat{\beta}_n\|^2$$

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

$\equiv$ LSE of β and σ^2 in a model

M_n: $\mathbb{Y}_n | \mathbb{X}_n \sim (\mathbb{X}_n\beta, \sigma^2 I_n)$.

Interest: Behavior of $\hat{\beta}_n$, MSe_n as $n\to\infty$

+ behavior of $l^T\hat{\beta}_n =: \hat{\theta}_n$ for given $l \in \mathbb{R}^k$

$L\hat{\beta}_n =: \hat{\xi}_n$ for given $L_{m\times k}$

under two "truths" $\longrightarrow$ homoscedasticity

($\text{var}(\varepsilon|X) = \sigma^2 \equiv$ const)

heteroscedasticity

($\text{var}(\varepsilon|X) = \sigma^2(X) \neq$ const)

Remember: $\varepsilon = Y - X^T\beta$

$$\mathrm{var}(\varepsilon|X) = \mathrm{var}(Y|X) =: \sigma^2(X)$$

$$\Rightarrow \mathrm{var}(\varepsilon) = \mathbb{E}\,\sigma^2(X)$$

(A2 homoscedastic)

$$\sigma^2(X) := \mathrm{var}(Y|X) = \sigma^2 \quad (= \text{const})$$

$0 < \sigma^2 < \infty$: unknown parameter

(A2 heteroscedastic)

$\sigma^2(X) := \mathrm{var}(Y|X)$ satisfies

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

NOTATION: $W^* := \mathbb{E}(\sigma^2(X) X X^T) = \left(\mathbb{E}(\sigma^2(X) X_j X_\ell)\right)_{j,\ell = 0,\dots,k-1}$

Remember that (A1) is $\mathbb{E}|X_j X_\ell| < \infty \quad \forall j,\ell$

Hence

(A0), (A1), (A2 homoscedastic)

$\Longrightarrow$ (A0), (A1), (A2 heteroscedast.)

Everything proved with assumption of heteroscedasticity holds in the homoscedastic case as well.

16.2 Consistency of LSE

→ read slide

6

7

8

Theorem 16.2 Strong consistency of LSE

Let (A0), (A1), (A2 heteroscedastic) hold.
Then

$$\hat\beta_n \xrightarrow{a.s.} \beta, \quad n\to\infty$$

$$l^T\hat\beta_n = \hat\theta_n \xrightarrow{a.s.} l^T\beta, \quad n\to\infty$$

$$L\hat\beta_n = \hat\xi_n \xrightarrow{a.s.} L\beta, \quad n\to\infty.$$

Proof: It is sufficient to show consistency of $\hat\beta_n$, the rest follows from properties of a.s. convergence & linearity.

$$\hat\beta_n = (X_n^TX_n)^{-1}(X_n^TY_n) = \underbrace{\left(\frac{1}{n}\sum_{i=1}^n X_iX_i^T\right)^{-1}}_{\downarrow \text{a.s.}\; W^{-1}=V \text{ (Lemma 16.1)}}\left(\frac{1}{n}\sum_{i=1}^n X_iY_i\right)$$

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

$$= \underbrace{\frac{1}{n}\sum_{i=1}^n X_i\varepsilon_i}_{?\ \downarrow \text{a.s.}\ 0} + \underbrace{\frac{1}{n}\sum_{i=1}^n X_iX_i^T\beta}_{?\ \downarrow \text{a.s}\ W\cdot\beta}$$

(a) $\frac{1}{n}\sum_{i=1}^{n} X_i X_i^T \beta = \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i^T\right)}_{\substack{n\to\infty \downarrow a.s. \\ W = V^{-1} \text{ (Lemma 16.1)}}} \cdot \beta \xrightarrow[n\to\infty]{a.s.} W \cdot \beta$

(b) $\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i \cdot \varepsilon_i}_{} \xrightarrow[n\to\infty]{a.s.} 0$?

vector of length k $\to$ its element j $(j = 0, \dots, k-1)$ is

$\frac{1}{n}\sum_{i=1}^{n} X_{ij} \cdot \varepsilon_i$ Can we apply SLLN?

- $X_{ij} \cdot \varepsilon_i$ are iid for $i = 1, 2, \dots$
- $\mathbb{E}|X_{ij} \cdot \varepsilon_i| < \infty$?

$$\mathbb{E}|X_{ij}\varepsilon_i| \underset{\text{Cauchy-Schwarz}}{\leq} \sqrt{\mathbb{E}X_{ij}^2 \, \mathbb{E}\varepsilon_i^2} = \sqrt{\underbrace{\mathbb{E}X_{ij}^2}_{\substack{<\infty \\ \text{by (A1)}}} \underbrace{\mathbb{E}(\sigma^2(X_i))}_{\substack{<\infty \\ \text{by (A2)}}}}$$

Hence indeed $\mathbb{E}|X_{ij} \cdot \varepsilon_i| < \infty$.

- $\mathbb{E}(X_{ij}\varepsilon_i) \overset{\mathbb{E}|X_{ij}\varepsilon_i|<\infty}{=} \mathbb{E}(\mathbb{E}(X_{ij}\cdot\varepsilon_i | X_i)) = \mathbb{E}(X_{ij} \underbrace{\mathbb{E}(\varepsilon_i | X_i)}_{=0}) = 0$

That is, by SLLN for iid rand. variables

$$\frac{1}{n}\sum_{i=1}^{n} X_{ij}\varepsilon_i \xrightarrow[n\to\infty]{a.s.} 0 \quad \forall j \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i \cdot \varepsilon_i \xrightarrow[n\to\infty]{a.s.} 0$$

In summary: $\hat{\beta}_n = \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i^T\right)^{-1}}_{\xrightarrow{a.s.} W^{-1} = V} \left(\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i \varepsilon_i}_{\xrightarrow{a.s.} 0} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i X_i^T \beta}_{\xrightarrow{a.s.} W \cdot \beta}\right)$

Hence $\hat{\beta}_n \xrightarrow[n\to\infty]{a.s.} W^{-1} W \beta = \beta$

Theorem 16.3 Strong consistency of the mean squared error [9]

Let assumptions (A0), (A1), (A2 homoscedastic) hold.

Then $MSe_n \xrightarrow{a.s.} \sigma^2$, $n \to \infty$.

Proof: $$MSe_n = \frac{1}{n-k} SSe_n = \underbrace{\frac{n}{n-k}}_{\substack{\downarrow n\to\infty \\ 1}} \cdot \underbrace{\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i^T \hat{\beta}_n)^2}_{\substack{\downarrow a.s. ? \\ \sigma^2}}$$

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

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

(a) $$A_n := \frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i^T\beta)^2 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 \xrightarrow[n\to\infty]{a.s.} \sigma^2$$

by using SLLN for iid random variables ε_i^2, $i = 1, 2, \dots$

$$\mathbb{E}|\varepsilon_i^2| = \mathbb{E}\varepsilon_i^2 = \operatorname{var}\varepsilon_i \overset{\text{from assumption (A2)}}{=} \sigma^2$$

(b) $$B_n := \frac{1}{n}\sum_{i=1}^{n}\{X_i^T(\beta - \hat{\beta}_n)\}^2 \xrightarrow[n\to\infty]{a.s.} 0 \quad ?$$

(c) $$C_n := \frac{2}{n}\sum_{i=1}^{n}(Y_i - X_i^T\beta)X_i^T(\beta - \hat{\beta}_n) \xrightarrow[n\to\infty]{a.s.} 0 \quad ?$$

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

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

$$= \underbrace{(\beta-\hat{\beta}_n)^T}_{\substack{\text{Theorem 16.2}\ \downarrow\ a.s.\\ 0}} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i^T\right)}_{\substack{=\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\\ \xrightarrow{a.s.} W>0\\ \text{Lemma 16.1}}} \underbrace{(\beta-\hat{\beta}_n)}_{\substack{\downarrow\ a.s.\\ 0}} \xrightarrow[n\to\infty]{a.s.} 0$$

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

$$= 2\cdot\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i X_i^T\right)}_{\substack{\text{Proof of Theorem 16.2}\ \downarrow\ a.s.\\ 0}} \underbrace{(\beta-\hat{\beta}_n)}_{\substack{\text{Th. 16.2}\ \downarrow\ a.s.\\ 0}} \xrightarrow[n\to\infty]{a.s.} 0$$

<u>In summary</u>:

$$MSe_{,n} = \underbrace{\frac{n}{n-k}}_{\to 1}\Bigg[\underbrace{\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i^T\beta)^2}_{\xrightarrow{a.s.}\sigma^2} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\{X_i^T(\beta-\hat{\beta}_n)\}^2}_{\xrightarrow{a.s.} 0}$$

$$+ \underbrace{\frac{2}{n}\sum_{i=1}^{n}(Y_i - X_i^T\beta)X_i^T(\beta-\hat{\beta}_n)}_{\xrightarrow{a.s.} 0}\Bigg]$$

Hence $MSe_{,n} \xrightarrow[n\to\infty]{a.s.} \sigma^2$

$\boxed{\text{W}}$

Overall summary

Even without normality and homoscedast.

$$\hat{\beta}_n \xrightarrow{a.s.} \beta \quad (\Rightarrow \hat{\beta}_n \xrightarrow{P} \beta)$$

Even without normality (if homoscedasticity assumed)

$$MSe_{,n} \xrightarrow{a.s.} \sigma^2 \quad (\Rightarrow MSe_{,n} \xrightarrow{P} \sigma^2)$$

What consistency means?