Advanced Econometrics

Lecture 7: Asymptotic Theory for Least Square (Hansen Chapter 7)

Instructor: Ma, Jun

Renmin University of China

November 20, 2018

Introduction

(Y_i, X_i) $i=1,\cdots,n$ i.i.d.

$e_i \xrightarrow{\text{定义}} Y_i - X_i'\beta$

假设 ① $Y_i = X_i' \underset{\text{unknown}}{\beta} + \underset{\text{unobserved}}{e_i}$

② $E(e_i X_i) = 0$

- The model is

$$Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + e_i,\ i = 1, ..., n$$
$$\boldsymbol{\beta} = \left(\mathbb{E}\left(\boldsymbol{X}_i\boldsymbol{X}_i'\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{X}_iY_i\right).$$

Assumption

1. The obervations $(Y_i, \boldsymbol{X}_i)$, $i = 1, \ldots n$, *are independent and identically distributed.*

2.$\mathbb{E}\left(Y^2\right) < \infty.$

3.$\mathbb{E}\left\|\boldsymbol{X}^2\right\| < \infty.$

4.$\boldsymbol{Q}_{\boldsymbol{XX}} = \mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\right)$ *is positive definite.*

⇒ 可推最小二乘估计量的一致性.

Consistency of Least-Squares Estimator

- "$(Y_i, \boldsymbol{X}_i), i = 1, \ldots n$ are iid" implies that any function of $(Y_i, \boldsymbol{X}_i)$ is iid, including $\boldsymbol{X}_i \boldsymbol{X}_i'$ and $\boldsymbol{X}_i Y_i$.
- The LS estimator:

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{X}_i \boldsymbol{X}_i') \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{X}_i Y_i) \right) = \hat{\boldsymbol{Q}}_{\boldsymbol{XX}}^{-1} \hat{\boldsymbol{Q}}_{\boldsymbol{X}Y}$$

$$\hat{\boldsymbol{Q}}_{\boldsymbol{XX}} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{X}_i \boldsymbol{X}_i') \to_p \mathbb{E}(\boldsymbol{X}_i \boldsymbol{X}_i') = \boldsymbol{Q}_{\boldsymbol{XX}}$$

$$\hat{\boldsymbol{Q}}_{\boldsymbol{X}Y} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{X}_i Y_i') \to_p \mathbb{E}(\boldsymbol{X}_i Y_i) = \boldsymbol{Q}_{\boldsymbol{X}Y}.$$

- By Continuous Mapping Theorem,

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \hat{\boldsymbol{Q}}_{\boldsymbol{XX}}^{-1} \hat{\boldsymbol{Q}}_{\boldsymbol{X}Y} \\ &\to_p \boldsymbol{Q}_{\boldsymbol{XX}}^{-1} \boldsymbol{Q}_{\boldsymbol{X}Y} \\ &= \boldsymbol{\beta}. \end{aligned}$$

① $(X_i, Y_i) \perp\!\!\!\perp (X_j, Y_j)$. $i \neq j$

② $F(X_i, Y_i) = F(X_j, Y_j)$ 分布函数

$\Rightarrow h(X_i, Y_i) \perp\!\!\!\perp h(X_j, Y_j)$

$F_h(X_i, Y_i) = F_h(X_j, Y_j)$

$X_n \to_p a \in \mathbb{R}$ $Y_n \to_p b \in \mathbb{R}$. $b \neq 0$.

$h(x, y) = \frac{x}{y}$

$h(X_n, Y_n) \to_p h(a, b)$

连续 $\to g(A, b) = A^{-1} b$

$\Rightarrow \hat{\beta} = g(\hat{Q}_{xx}, \hat{Q}_{xy}) \to_p g(Q_{xx}, Q_{xy})$

$= Q_{xx}^{-1} Q_{xy}$

Consistency of Least-Squares Estimator

- A different approach:

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \hat{\boldsymbol{Q}}_{\boldsymbol{XX}}^{-1} \hat{\boldsymbol{Q}}_{\boldsymbol{X}e}$$

$$\hat{\boldsymbol{Q}}_{\boldsymbol{X}e} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{X}_i e_i) \, .$$

- The WLLN:

$$\hat{\boldsymbol{Q}}_{\boldsymbol{X}e} \to_p \mathbb{E}(\boldsymbol{X}_i e_i) = 0.$$

- Therefore,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \hat{\boldsymbol{Q}}_{\boldsymbol{XX}}^{-1} \hat{\boldsymbol{Q}}_{\boldsymbol{X}e} \to_p \boldsymbol{Q}_{\boldsymbol{XX}}^{-1} \boldsymbol{0} = \boldsymbol{0}.$$

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \underbrace{Y_i}_{X_i'\beta + e_i}\right)$$

$$= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} X_i e_i\right)}_{\to_p E(X_i e_i) = 0}$$

$X_1, \cdots, X_n$ iid.

$E(X_1) = 0 \Rightarrow E|X_1| < \infty$

$\Rightarrow \bar{X}_n \to_p 0 = E(X_1)$ (WLLN)

Theorem

Consistency of Least-Squares

$\hat{\boldsymbol{Q}}_{\boldsymbol{XX}} \xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{XX}}$, $\hat{\boldsymbol{Q}}_{\boldsymbol{X}Y} \xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{X}Y}$, $\hat{\boldsymbol{Q}}_{\boldsymbol{XX}}^{-1} \xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{XX}}^{-1}$, $\hat{\boldsymbol{Q}}_{\boldsymbol{X}e} \xrightarrow{p} 0$, *and* $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$.

Asymptotic Normality

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i\boldsymbol{X}_i'\right)\right)^{-1}}_{\to_p Q_{xx}}\underbrace{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\boldsymbol{X}_ie_i\right)\right)}_{\to_d N(0,\Omega)}$$

- $\boldsymbol{X}_ie_i=\boldsymbol{X}_i\left(Y_i-\boldsymbol{X}_i'\boldsymbol{\beta}\right)$, $i=1,\ldots,n$ are iid and mean zero ($\mathbb{E}\boldsymbol{X}_ie_i=\mathbf{0}$).
- The covariance matrix: $\boldsymbol{\Omega}=\mathbb{E}\left(e_i^2\boldsymbol{X}_i\boldsymbol{X}_i'\right)$:

$$\begin{aligned}\|\boldsymbol{\Omega}\| &\leq \mathbb{E}\left\|\boldsymbol{X}_i\boldsymbol{X}_i'e_i^2\right\|=\mathbb{E}\left(\|\boldsymbol{X}_i\|^2e_i^2\right)\leq\mathbb{E}\left(\|\boldsymbol{X}_i\|^4\right)^{1/2}\left(\mathbb{E}\left(e_i^4\right)\right)^{1/2}\\ &<\infty.\end{aligned}$$

$\|X_iX_i'\|=\|X_i\|^2$

Asymptotic Normality

Theorem

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\boldsymbol{X}_i e_i \right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \mathbf{\Omega}\right).$$

Slutsky's theorem:

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} \boldsymbol{Q}_{\boldsymbol{XX}}^{-1} \mathrm{N}\left(\mathbf{0}, \mathbf{\Omega}\right)$$
$$= \mathrm{N}\left(\mathbf{0}, \boldsymbol{Q}_{\boldsymbol{XX}}^{-1} \mathbf{\Omega} \boldsymbol{Q}_{\boldsymbol{XX}}^{-1}\right).$$

正态分布的性质:

一个矩阵×一个正态随机变量 ⇒ 正态分布

Theorem

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\beta}}\right)$$

$\boldsymbol{\beta}$ 渐近斜方差矩阵

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{Q}_{\boldsymbol{XX}}^{-1} \mathbf{\Omega} \boldsymbol{Q}_{\boldsymbol{XX}}^{-1},$$

$\boldsymbol{Q}_{\boldsymbol{XX}} = \mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i'\right)$, *and* $\mathbf{\Omega} = \mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i' e_i^2\right)$.

$V_\beta = E(X_iX_i')^{-1}E(e_i^2X_iX_i')E(X_iX_i')^{-1}$

Asymptotic Normality

- $\boldsymbol{V}_{\boldsymbol{\beta}}$ is often referred to as the **asymptotic covariance matrix** of $\hat{\boldsymbol{\beta}}$.
- Distributional approximation: when n is large,

$$\hat{\boldsymbol{\beta}} \overset{a}{\sim} \mathrm{N}\left(\boldsymbol{\beta}, \frac{\boldsymbol{V}_{\boldsymbol{\beta}}}{n}\right).$$

- The finite-sample conditional variance:

精确的条件方差 ⇒
$$\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \mathrm{Var}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

  $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}$ is the exact conditional variance of $\hat{\boldsymbol{\beta}}$.
- We should expect $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} \approx \frac{\boldsymbol{V}_{\boldsymbol{\beta}}}{n}$.

$$n\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

  and $n\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} \to_p \boldsymbol{V}_{\boldsymbol{\beta}}$.

Asymptotic Normality

$\boxed{\text{非常重要}}$

- Under homoskedasticity, $\mathbb{E}\left(e_i^2|\boldsymbol{X}_i\right) = \sigma^2 = \text{constant}$,

$$\boldsymbol{\Omega} = \mathbb{E}\mathbb{E}\left(e_i^2\boldsymbol{X}_i\boldsymbol{X}_i'|\boldsymbol{X}_i\right) = \boldsymbol{Q}_{\boldsymbol{X}\boldsymbol{X}}\sigma^2$$

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{Q}_{\boldsymbol{X}\boldsymbol{X}}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}_{\boldsymbol{X}\boldsymbol{X}}^{-1} = \boldsymbol{Q}_{\boldsymbol{X}\boldsymbol{X}}^{-1}\sigma^2.$$

- We define $\boldsymbol{V}_{\boldsymbol{\beta}}^0 = \boldsymbol{Q}_{\boldsymbol{X}\boldsymbol{X}}^{-1}\sigma^2$ no matter $\mathbb{E}\left(e_i^2|\boldsymbol{X}_i\right) = \sigma^2$ is true or false. When it is true, $\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{V}_{\boldsymbol{\beta}}^0$. $\boldsymbol{V}_{\boldsymbol{\beta}}^0$ is called the homoskedastic asymptotic covariance matrix.

Asymptotic Normality

$\sqrt{n}(\hat{\beta}-\beta) \to_d N(0, V_\beta)$

n很大，$\sqrt{n}(\hat{\beta}-\beta) \overset{a}{\sim} N(0, V_\beta)$

$\Leftrightarrow \hat{\beta} \overset{a}{\sim} N(\beta, \frac{V_\beta}{n})$

$V_\beta = (EX_iX_i')^{-1}E(e_i^2X_iX_i')(EX_iX_i')^{-1}$

$V_\beta^0 = \sigma^2(EX_iX_i')^{-1}$, $\sigma^2 = Ee_i^2$

定义 $V_\beta^0 = E(X_iX_i')^{-1}\sigma^2$ $\quad \sigma^2 = E(e_i^2)$

当同方差假设成立时，$V_\beta^0 = V_\beta$.

$\Uparrow$

$E(e_i^2|X_i) = \sigma^2$

$\frac{1}{n}\sum_{i=1}^n e_i^2 \to_p E(e_i^2)$ 观测不到

$\frac{1}{n}\sum_{i=1}^n \hat{e}_i^2 \to_p E(e_i^2)$ 同样成立

$\{\hat{e}_i\}_{i=1}^n$ 不是 i.i.d.

a. $\hat{e}_i = Y_i - X_i'\hat{\beta}$

每个残差 $\hat{e}_i$ 与 $\hat{\beta}$ 有关，而 $\hat{\beta}$ 是根据所有观测值得到的.

b. $\sum_{i=1}^n \hat{e}_i = 0$ 要满足这样的约束，必然有相关性.

Consistency of Error Variance Estimators

- Write the residual $\hat{e}_i$ as the error e_i plus a deviation term:

$$\begin{aligned}\hat{e}_i &= Y_i - \boldsymbol{X}_i'\hat{\boldsymbol{\beta}} \\ &= e_i + \boldsymbol{X}_i'\boldsymbol{\beta} - \boldsymbol{X}_i'\hat{\boldsymbol{\beta}} \\ &= e_i - \boldsymbol{X}_i'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right).\end{aligned}$$

- Thus

$$\hat{e}_i^2 = e_i^2 - 2e_i\boldsymbol{X}_i'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\boldsymbol{X}_i'\boldsymbol{X}_i\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right).$$

- The estimator $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n \hat{e}_i^2$ of $\sigma^2 = \underline{\mathbb{E}e_i^2}$:

$\rightarrow e_i X_i'$ $\rightarrow_p 0$

$$\begin{aligned}\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n e_i^2 &- 2\left(\frac{1}{n}\sum_{i=1}^n e_i\boldsymbol{X}_i'\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \\ &+ \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i'\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right).\end{aligned}$$

$\rightarrow X_i X_i'$ $\rightarrow_p 0$

$\rightarrow_p \frac{1}{n}\sum_{i=1}^n e_i^2$

Consistency of Error Variance Estimators

- WLLN:

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 \xrightarrow{p} \sigma^2$$

$$\frac{1}{n}\sum_{i=1}^{n} e_i \boldsymbol{X}_i' \xrightarrow{p} \mathbb{E}\left(e_i^2 \boldsymbol{X}_i'\right) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \xrightarrow{p} \mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i'\right) = \boldsymbol{Q}_{\boldsymbol{XX}}.$$

- Another estimator $s^2 = (n-k)^{-1}\sum_{i=1}^{n} \hat{e}_i^2$. Since $n/(n-k) \to 1$ as $n \to \infty$,

$$s^2 = \left(\frac{n}{n-k}\right)\hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

调整过的估计量

Theorem

$\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ *and* $s^2 \xrightarrow{p} \sigma^2$.

Homoskedastic Covariance Matrix Estimation

- For inference (confidence intervals and tests), we need a consistent estimate of $\boldsymbol{V}_{\boldsymbol{\beta}}$.
- Under homoskedasticity, $\boldsymbol{V}_{\boldsymbol{\beta}}$ simplifies to $\boldsymbol{V}^0_{\boldsymbol{\beta}} = \boldsymbol{Q}^{-1}_{\boldsymbol{XX}}\sigma^2$.
- A natural estimator of $\boldsymbol{V}^0_{\boldsymbol{\beta}} = \boldsymbol{Q}^{-1}_{\boldsymbol{XX}}\sigma^2$ is $\boxed{\hat{\boldsymbol{V}}^0_{\boldsymbol{\beta}} = \hat{\boldsymbol{Q}}^{-1}_{\boldsymbol{XX}} s^2}$. —默认计算标准误的公式.
- By CMT,

$$\hat{\boldsymbol{V}}^0_{\boldsymbol{\beta}} = \hat{\boldsymbol{Q}}^{-1}_{\boldsymbol{XX}} s^2 \to_p \boldsymbol{Q}^{-1}_{\boldsymbol{XX}}\sigma^2 = \boldsymbol{V}^0_{\boldsymbol{\beta}}.$$

- $\hat{\boldsymbol{V}}^0_{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{V}^0_{\boldsymbol{\beta}}$ regardless if the regression is homoskedastic or heteroskedastic.
- However, $\boldsymbol{V}^0_{\boldsymbol{\beta}} = \boldsymbol{V}_{\boldsymbol{\beta}}$, the asymptotic covariance matrix, only under homoskedasticity.

$\hat{V}^0_\beta$ 是 V^0_β 的一致估计量，不取决于同方差、异方差假设.

当同方差假设时，$V^0_\beta = V_\beta$.
于是 $\hat{V}^0_\beta$ 是 V_β 的一致估计量.

Heteroskedastic Covariance Matrix Estimation

- A method of moments estimator for $\boldsymbol{\Omega}$:

$$\hat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \hat{e}_i^2.$$

- The White covariance matrix estimator

reg y, x, robust $\longrightarrow$

$$\hat{\boldsymbol{V}}_{\boldsymbol{\beta}}^{W} = \hat{\boldsymbol{Q}}_{\boldsymbol{XX}}^{-1} \hat{\boldsymbol{\Omega}} \hat{\boldsymbol{Q}}_{\boldsymbol{XX}}^{-1}.$$

- Observe

$$\begin{aligned}\hat{\boldsymbol{\Omega}} &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \hat{e}_i^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' e_i^2 + \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \left(\hat{e}_i^2 - e_i^2\right).\end{aligned}$$

- By WLLN,

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' e_i^2 \xrightarrow{p} \mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{X}_i' e_i^2\right) = \boldsymbol{\Omega}.$$

$V_\beta = Q_{xx}^{-1} \Omega Q_{xx}^{-1}$

$\Omega = E(X_i X_i' e_i^2)$

$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2 X_i X_i'$

$\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \to_p E(X_i X_i')$

$\frac{1}{n}\sum_{i=1}^{n} e_i^2 X_i X_i' \to_p E(e_i^2 X_i X_i')$

↑ 관찰되지 않음)

$\hat{e}_i^2$

Heteroskedastic Covariance Matrix Estimation

- It remains to show

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \left(\hat{e}_i^2 - e_i^2\right) \rightarrow_p 0.$$

- Recall matrix norm: $\|\boldsymbol{A}\| = \operatorname{tr}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{1/2}$ and therefore,

$$\left\|\boldsymbol{X}_i \boldsymbol{X}_i'\right\| \cancel{= \operatorname{tr}\left(\boldsymbol{X}_i \boldsymbol{X}_i'\right)^{1/2} = \operatorname{tr}\left(\boldsymbol{X}_i' \boldsymbol{X}_i\right)^{1/2} = \|\boldsymbol{X}_i\|}.$$

$= tr(X_i' X_i X_i X_i')^{\frac{1}{2}} = tr(X_i' X_i X_i' X_i)^{\frac{1}{2}} = tr((X_i' X_i)^2)^{\frac{1}{2}} = \|X_i\|^2$

- Thus,

$$\left\|\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \left(\hat{e}_i^2 - e_i^2\right)\right\| \leq \frac{1}{n}\sum_{i=1}^{n} \left\|\boldsymbol{X}_i \boldsymbol{X}_i' \left(\hat{e}_i^2 - e_i^2\right)\right\|$$
$$= \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{X}_i\|^2 \left|\hat{e}_i^2 - e_i^2\right|.$$

$\|x+y\| \leq \|x\| + \|y\|$

$\|c\cdot x\| = |c|\cdot\|x\|$

$\left\|\frac{1}{n}\sum_{i=1}^{n} x_i\right\| = \frac{1}{n}\left\|\sum_{i=1}^{n} x_i\right\| \leq \frac{1}{n}\sum_{i=1}^{n}\|x_i\|$

Heteroskedastic Covariance Matrix Estimation

$\hat{e}_i^2 - e_i^2 = -2X_i'(\hat{\beta}-\beta)e_i + (\hat{\beta}-\beta)XX_i'(\hat{\beta}-\beta)$

- By the triangle inequality and Cauchy-Schwarz inequality,

$$
\begin{aligned}
\left|\hat{e}_i^2 - e_i^2\right| &\le 2\left|e_i \boldsymbol{X}_i'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right| + \left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\boldsymbol{X}_i\boldsymbol{X}_i'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \\
&= 2\left|e_i\right|\left|\boldsymbol{X}_i'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right| + \left|\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\boldsymbol{X}_i\right|^2 \\
&\le 2\left|e_i\right|\left\|\boldsymbol{X}_i\right\|\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\| + \left\|\boldsymbol{X}_i\right\|^2\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2.
\end{aligned}
$$

- Thus,

概率极限是有限的　　概率极限为0.

$$
\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\left(\hat{e}_i^2 - e_i^2\right)\right\| &\le 2\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{X}_i\right\|^3\left|e_i\right|\right)\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\| \\
&\quad + \left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{X}_i\right\|^4\right)\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2.
\end{aligned}
$$

Heteroskedastic Covariance Matrix Estimation

Cauchy-Schwarz: $E|XY| \le (EX^2)^{\frac{1}{2}}(EY^2)^{\frac{1}{2}}$

Hölder: if $p>1$, $q>1$, $\frac{1}{p}+\frac{1}{q}=1$

then $E|XY| \le (E|X|^p)^{\frac{1}{p}}(E|Y|^q)^{\frac{1}{q}}$

$Y_i = X_i'\beta + e_i$

$\left.\begin{array}{l} E e_i^4 < \infty \\ E\|X_i\|^4 < \infty \end{array}\right\} \Rightarrow E(\|X_i\|^3 |e_i|) < \infty$

$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} \|X_i\|^3 e_i \xrightarrow{p} E(\|X_i\|^3 |e_i|)$

- By Hölder's inequality,

$$\mathbb{E}\left(\|\boldsymbol{X}_i\|^3 |e_i|\right) \le \left(\mathbb{E}\left(\|\boldsymbol{X}_i\|^3\right)^{4/3}\right)^{3/4} \left(\mathbb{E}\left(e_i^4\right)\right)^{1/4}$$
$$= \left(\mathbb{E}\left(\|\boldsymbol{X}_i\|^4\right)\right)^{3/4} \left(\mathbb{E}\left(e_i^4\right)\right)^{1/4} < \infty.$$

$Y_i = X_i'\beta + e_i$

① 有限样本：假设 $e \sim$ Normal + 同方差. $\hat{V}_\beta^0$

② 大样本：$\hat{V}_\beta^W$

③ 有限：假设 $e_i \perp\!\!\!\perp X_i$ rank some-test.

Thus WLLN applies to $n^{-1}\sum_{i=1}^{n} \|\boldsymbol{X}_i\|^3 |e_i|$.

Theorem

$\hat{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega}$ *and* $\hat{\boldsymbol{V}}_{\boldsymbol{\beta}}^{W} \xrightarrow{p} \boldsymbol{V}_{\boldsymbol{\beta}}$.

怀特是稳健的一致估计量

大样本情况下 reg y x, robust 加 robust, 用怀特估计量. ← 主流观点

严格来说样本量只有20–30时，统计学里有一套有限样本的分析方法.

Functions of Parameters

- The parameter of interest $\boldsymbol{\theta}$ is a function of the coefficients, $\boldsymbol{\theta} = \boldsymbol{r}(\boldsymbol{\beta})$ for some function $\boldsymbol{r} : \mathbb{R}^k \to \mathbb{R}^q$. The estimate of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{r}\left(\hat{\boldsymbol{\beta}}\right).$$

Theorem
If $\boldsymbol{r}(\cdot)$ is continuous at the true value of $\boldsymbol{\beta}$, then $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$.

- By the Delta Method, $\hat{\boldsymbol{\theta}}$ is asymptotically normal.

Assumption
$\boldsymbol{r} : \mathbb{R}^k \to \mathbb{R}^q$ is continuously differentiable at the true value of $\boldsymbol{\beta}$ and $\boldsymbol{R} = \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{r}(\boldsymbol{\beta})'$ has rank q.

Functions of Parameters

$\sqrt{n}(\hat{\beta}-\beta) \rightarrow_d N(0, V_\beta)$

Delta Method $\Longrightarrow$ $\sqrt{n}(\hat{\theta}-\theta) \rightarrow_d N(0, R' V_\beta R)$

$\hat{\theta} = r(\hat{\beta})$, $\theta = r(\beta)$

Theorem

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\theta}}\right)$$

where

$$\boldsymbol{V}_{\boldsymbol{\theta}} = \boldsymbol{R}'\boldsymbol{V}_{\boldsymbol{\beta}}\boldsymbol{R}$$

- $\boldsymbol{r}$ can be linear: $\boldsymbol{r}(\boldsymbol{\beta}) = \boldsymbol{R}'\boldsymbol{\beta}$, for some $k \times q$ matrix $\boldsymbol{R}$.
- An even simpler case is when $\boldsymbol{R}$ is of the form $\boldsymbol{R} = \begin{pmatrix} \boldsymbol{I} \\ \mathbf{0} \end{pmatrix}$.
- Then we can partition $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\right)'$ so that $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{\beta}_1$. Then

$$\boldsymbol{V}_{\boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{I} & \mathbf{0} \end{pmatrix} \boldsymbol{V}_{\boldsymbol{\beta}} \begin{pmatrix} \boldsymbol{I} \\ \mathbf{0} \end{pmatrix} = \boldsymbol{V}_{11},$$

where $\boldsymbol{V}_{\boldsymbol{\beta}}$ is partitioned:

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix}.$$

Functions of Parameters

- Take the example $\theta = \beta_j / \beta_l$ for $j \neq l$. Then

$$\boldsymbol{R} = \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{r}(\boldsymbol{\beta}) = \begin{pmatrix} \frac{\partial}{\partial \beta_1}\left(\beta_j/\beta_l\right) \\ \vdots \\ \frac{\partial}{\partial \beta_j}\left(\beta_j/\beta_l\right) \\ \vdots \\ \frac{\partial}{\partial \beta_l}\left(\beta_j/\beta_l\right) \\ \vdots \\ \frac{\partial}{\partial \beta_k}\left(\beta_j/\beta_l\right) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 1/\beta_l \\ \vdots \\ -\beta_j/\beta_l^2 \\ \vdots \\ 0 \end{pmatrix}.$$

- So

$$\boldsymbol{V_\theta} = \boldsymbol{V}_{jj}/\beta_l^2 + \boldsymbol{V}_{ll}\beta_j^2/\beta_l^4 - 2\boldsymbol{V}_{jl}\beta_j/\beta_l^3.$$

Functions of Parameters

- For inference, we need an estimate of $\boldsymbol{V}_{\boldsymbol{\theta}} = \boldsymbol{R}'\boldsymbol{V}_{\boldsymbol{\beta}}\boldsymbol{R}$. The natural estimator of $\boldsymbol{R}$ is

$$\hat{\boldsymbol{R}} = \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{r}\left(\hat{\boldsymbol{\beta}}\right)'.$$

- The estimate of $\boldsymbol{V}_{\boldsymbol{\theta}}$ is

$$\hat{\boldsymbol{V}}_{\boldsymbol{\theta}} = \hat{\boldsymbol{R}}'\hat{\boldsymbol{V}}_{\boldsymbol{\beta}}\hat{\boldsymbol{R}}.$$

实证研究中，r是线性函数，R就是已知的一个矩阵，根本不用估计.

Asymptotic Standard Errors

标准误是 $\hat{\beta}$ 的抽样分布的标准差的估计量.

- A standard error is an estimate of the standard deviation of the distribution of an estimator.
- Since $\hat{\boldsymbol{\beta}} \overset{a}{\sim} \mathrm{N}\left(\boldsymbol{\beta}, \frac{\boldsymbol{V}_{\boldsymbol{\beta}}}{n}\right)$ and $\hat{\beta}_j \overset{a}{\sim} \mathrm{N}\left(\beta_j, \frac{[\boldsymbol{V}_{\boldsymbol{\beta}}]_{jj}}{n}\right)$, the standard error takes the form

$$\text{标准误} \Rightarrow s\left(\hat{\beta}_j\right) = \sqrt{\frac{\left[\hat{\boldsymbol{V}}_{\boldsymbol{\beta}}^{W}\right]_{jj}}{n}}.$$

- Suppose the parameter of interest is $\theta = r(\boldsymbol{\beta})$ ($r : \mathbb{R}^k \to \mathbb{R}$, $q = 1$), the standard error for $\hat{\theta} = r\left(\hat{\boldsymbol{\beta}}\right)$ is

$$s\left(\hat{\theta}\right) = \sqrt{\frac{\hat{\boldsymbol{R}}' \hat{\boldsymbol{V}}_{\boldsymbol{\beta}} \hat{\boldsymbol{R}}}{n}}.$$

t-statistic

严格来说可以叫 t ratio

- $\theta = r(\beta)$ is the parameter of interest. Consider

θ是个参数真值.

$$T(\theta) = \frac{\hat{\theta} - \theta}{s(\hat{\theta})}.$$

$s(\hat{\theta}) = \sqrt{\frac{\hat{V}_\theta}{n}}$

- Since $\sqrt{n}\left(\hat{\theta} - \theta\right) \to_d \mathrm{N}(0, V_\theta)$ and $\hat{V}_\theta \to_p V_\theta$,

$$\begin{aligned} T(\theta) &= \frac{\hat{\theta} - \theta}{s(\hat{\theta})} \\ &= \frac{\sqrt{n}(\hat{\theta} - \theta)}{\sqrt{\hat{V}_\theta}} \\ &\xrightarrow{d} \frac{N(0, V_\theta)}{\sqrt{V_\theta}} \\ &= Z \sim N(0, 1). \end{aligned}$$

正态随机分布乘一个数还是正态的.

t-statistic

- Since $T(\theta) \to_d Z$, CMT yields $|T(\theta)| \to_d |Z|$.

$$
\begin{aligned}
\Pr(| Z |\leq u) &= \Pr(-u \leq Z \leq u) \\
&= \Pr(Z \leq u) - \Pr(Z < -u) \\
&= \Phi(u) - \Phi(-u) \\
&= 2\Phi(u) - 1.
\end{aligned}
$$

Φ 标准正态的分布函数.

Theorem
$T(\theta) \xrightarrow{d} Z \sim N(0,1)$ *and* $|T(\theta)| \xrightarrow{d} | Z |$.

Confidence Intervals

- A conventional confidence interval takes the form

$$\hat{C} = \left[\ \hat{\theta} - c \cdot s(\hat{\theta}), \quad \hat{\theta} + c \cdot s(\hat{\theta})\ \right],$$

where $c = F^{-1}_{|Z|}(1-\alpha)$ or $2\Phi(c) - 1 = 1 - \alpha$.

- Equivalently,

$$\hat{C} = \{\theta\colon |T(\theta)| \leq c\} = \left\{\theta\colon\ -c \leq \frac{\hat{\theta} - \theta}{s(\hat{\theta})} \leq c\right\}.$$

The coverage probability:

$$\Pr\left(\theta \in \hat{C}\right) = \Pr\left(|T(\theta)| \leq c\right) \longrightarrow \Pr\left(|Z| \leq c\right) = 1 - \alpha.$$

Theorem

With $c = \Phi^{-1}(1-\alpha/2)$, $\Pr\left(\theta \in \hat{C}\right) \longrightarrow 1 - \alpha$. *For* $c = 1.96$, $\Pr\left(\theta \in \hat{C}\right) \longrightarrow 0.95.$

$P(|Z| \leq c) = 1-\alpha \Leftrightarrow 2\Phi(c) - 1 = 1-\alpha$

$Z \sim N(0,1)$ $\quad \alpha = 0.05$

$c = 1.96$

$\hat{C} = [\hat{\theta} - c \cdot s(\hat{\theta}), \hat{\theta} + c \cdot s(\hat{\theta})]$

$\Pr(\theta \in \hat{C}) = \Pr(\hat{\theta} - c \cdot s(\hat{\theta}) \leq \theta \leq \hat{\theta} + c \cdot s(\hat{\theta}))$

$= \Pr\left(-c \leq \frac{\hat{\theta} - \theta}{s(\hat{\theta})} \leq c\right)$

$= \Pr\left(\left|\frac{\hat{\theta} - \theta}{s(\hat{\theta})}\right| \leq c\right)$

$\rightarrow \Pr(|Z| \leq c) = 1-\alpha$

Confidence Intervals

$\Pr\left(\beta_j \in \left[\hat{\beta}_j - z_{1-\frac{\alpha}{2}} \cdot s(\hat{\beta}),\ \hat{\beta}_j + z_{1-\frac{\alpha}{2}} \cdot s(\hat{\beta})\right]\right) = 1-\alpha$

大样本置信区间

- Under homoskedasticity,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) \to_d N\left(0, \sigma^2\left(\mathbb{E}\left(\boldsymbol{X}_1\boldsymbol{X}_1'\right)\right)^{-1}\right).$$

- We estimate the asymptotic variance by $s^2\left(n^{-1}\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i'\right)^{-1}$.
- The confidence interval for β_j is given by

$$\left[\widehat{\beta}_j \pm z_{1-\alpha/2}\sqrt{\left[s^2\left(n^{-1}\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i'\right)^{-1}\right]_{jj}/n}\right]$$

$$= \left[\widehat{\beta}_j \pm z_{1-\alpha/2}\sqrt{\left[s^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{jj}}\right]$$

which is the same as the finite sample confidence interval.

(有限样本)

t统计量构造的置信区间依赖于正态分布的假设. $e|X \sim$ Normal

大样本置信区间即使正态假设不成立, 也是稳健的.

Wald Statistic

- The parameter of interest is $\boldsymbol{\theta} = \boldsymbol{r}(\boldsymbol{\beta})$. $\boldsymbol{r} : \mathbb{R}^k \to \mathbb{R}^q$. Consider the Wald statistic

$$W(\boldsymbol{\theta}) = n\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)' \hat{\boldsymbol{V}}_{\boldsymbol{\theta}}^{-1} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right).$$

$$= \sqrt{n}(\hat{\theta}-\theta)' \hat{V}_\theta^{-1} \sqrt{n}(\hat{\theta}-\theta)$$

$$= \sqrt{n}(\hat{\theta}-\theta)' V_\theta^{-1} \sqrt{n}(\hat{\theta}-\theta)$$

$$+ \underbrace{\sqrt{n}(\hat{\theta}-\theta)'}_{\to_d N(0,V_\theta)} \underbrace{(\hat{V}_\theta^{-1} - V_\theta^{-1})}_{\to_p 0} \underbrace{\sqrt{n}(\hat{\theta}-\theta)}_{\to_d N(0,V_\theta)}$$

$$= \underbrace{\sqrt{n}(\hat{\theta}-\theta)'}_{\to_d N(0,V_\theta)} V_\theta^{-1} \underbrace{\sqrt{n}(\hat{\theta}-\theta)}_{\to_d N(0,V_\theta)} = h(\sqrt{n}(\hat{\theta}-\theta))$$

- Since

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{d} \boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{V}_{\boldsymbol{\theta}})$$

and $\hat{\boldsymbol{V}}_{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{V}_{\boldsymbol{\theta}}$,

$$W(\boldsymbol{\theta}) = n\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)' \hat{\boldsymbol{V}}_{\boldsymbol{\theta}}^{-1} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \to_d \boldsymbol{Z}'\boldsymbol{V}_{\boldsymbol{\theta}}^{-1}\boldsymbol{Z} \sim \chi^2_q.$$

$$\to_d h(Z) = Z' \underbrace{V_\theta^{-1}}_{= V_\theta^{-\frac{1}{2}} V_\theta^{-\frac{1}{2}}} Z \sim \chi^2_q$$

> **Theorem**
>
> $$W(\boldsymbol{\theta}) \xrightarrow{d} \chi^2_q.$$

Confidence Regions 置信域

- A confidence region $\hat{C}$ is a set estimator for $\boldsymbol{\theta} \in \mathbb{R}^q$ when $q > 1$. Ideally, we hope $\Pr\left(\boldsymbol{\theta} \in \hat{C}\right) = 1 - \alpha$.
- A natural confidence region is

$$\hat{C} = \left\{\boldsymbol{\theta} : W\left(\boldsymbol{\theta}\right) \leq c_{1-\alpha}\right\},$$

  with $c_{1-\alpha}$ being the $1 - \alpha$ quantile of the χ_q^2 distribution: $F_{\chi_q^2}\left(c_{1-\alpha}\right) = 1 - \alpha$.
- Thus,

$$\Pr\left(\boldsymbol{\theta} \in \hat{C}\right) \to \Pr\left(\chi_q^2 \leq c_{1-\alpha}\right) = 1 - \alpha.$$