Advanced Econometrics

Lecture 8: Asymptotic Theory for Least Square (Hansen Chapter 7)

Instructor: Ma, Jun

Renmin University of China

November 22, 2021

Introduction

- The model is

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

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.*$\mathbb{E}\left\Vert\boldsymbol{X}^2\right\Vert < \infty$. *4.*$\boldsymbol{Q}_{\boldsymbol{XX}} = \mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\right)$ *is positive definite.*

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:

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

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

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

- By Continuous Mapping Theorem,

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

- 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:

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

- Therefore,

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

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

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \left(\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i\boldsymbol{X}_i'\right)\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\boldsymbol{X}_i e_i\right)\right)$$

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

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

Theorem

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

Slutsky's theorem:

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

$$= \mathrm{N}\left(\mathbf{0}, \boldsymbol{Q}_{\boldsymbol{XX}}^{-1} \boldsymbol{\Omega} \boldsymbol{Q}_{\boldsymbol{XX}}^{-1} \right).$$

Theorem

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

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

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

- $\boldsymbol{V}_{\boldsymbol{\beta}}$ 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

$$\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \mathrm{Var}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = (\boldsymbol{X}'\boldsymbol{X})^{-1} (\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}) (\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

  $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}$ 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

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

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

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

- We define $\boldsymbol{V}_{\boldsymbol{\beta}}^0 = \boldsymbol{Q}_{\boldsymbol{XX}}^{-1} \sigma^2$ 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.

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 $\sigma^2 = \mathbb{E}e_i^2$:

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

- WLLN:

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

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

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

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

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

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 $\boldsymbol{V}_{\boldsymbol{\beta}}^0 = \boldsymbol{Q}_{\boldsymbol{XX}}^{-1}\sigma^2$.
- A natural estimator of $\boldsymbol{V}_{\boldsymbol{\beta}}^0 = \boldsymbol{Q}_{\boldsymbol{XX}}^{-1}\sigma^2$ is $\hat{\boldsymbol{V}}_{\boldsymbol{\beta}}^0 = \hat{\boldsymbol{Q}}_{\boldsymbol{XX}}^{-1}s^2$.
- By CMT,
$$\hat{\boldsymbol{V}}_{\boldsymbol{\beta}}^0 = \hat{\boldsymbol{Q}}_{\boldsymbol{XX}}^{-1}s^2 \to_p \boldsymbol{Q}_{\boldsymbol{XX}}^{-1}\sigma^2 = \boldsymbol{V}_{\boldsymbol{\beta}}^0.$$
- $\hat{\boldsymbol{V}}_{\boldsymbol{\beta}}^0$ is consistent for $\boldsymbol{V}_{\boldsymbol{\beta}}^0$ regardless if the regression is homoskedastic or heteroskedastic.
- However, $\boldsymbol{V}_{\boldsymbol{\beta}}^0 = \boldsymbol{V}_{\boldsymbol{\beta}}$, the asymptotic covariance matrix, only under homoskedasticity.

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

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

- 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,

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

- 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,

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

- Thus,

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

- By the triangle inequality and Cauchy-Schwarz inequality,

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

- Thus,

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

Theorem

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

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.

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

- 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

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

- 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

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

Asymptotic Standard Errors

- 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

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

- 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

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

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

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

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

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

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

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 $c = F_{|Z|}^{-1}(1-\alpha)$ 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* $c = 1.96$, $\Pr\left(\theta \in \hat{C}\right) \longrightarrow 0.95$.

- 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

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

  which is the same as the finite sample confidence interval.

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

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

- Since

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

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

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

Theorem

$$W(\boldsymbol{\theta}) \xrightarrow{d} \chi_q^2.$$

Confidence Regions

- 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

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

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

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