# Econometric Analysis of Cross Section and Panel Data

## Lecture 5: Asymptotic Theory for Least Squares

Zhian Hu

Central University of Finance and Economics

Fall 2024

# This Lecture

- Hansen (2022): Chapter 6 & 7
- The most widely-used tool in sampling theory is large sample asymptotics.
- By "asymptotics" we mean approximating a finite-sample sampling distribution by taking its limit as the sample size diverges to infinity.

# Modes of Convergence

依概率收敛

- **Definition:** A sequence of random vectors $Z_n \in \mathbb{R}^k$ converges in probability to $Z$ as $n \to \infty$, denoted $Z_n \underset{p}{\longrightarrow} Z$ or alternatively $\text{plim}_{n\to\infty} Z_n = Z$, if for all $\delta > 0$

$$\lim_{n\to\infty} \mathbb{P}\left[\|Z_n - Z\| \le \delta\right] = 1$$

We call $Z$ the probability limit (or plim) of $Z_n$

# Modes of Convergence

## 依分布收斂

- **Definition:** Let $Z_n$ be a sequence of random vectors with distributions $F_n(u) = \mathbb{P}[Z_n \leq u]$. We say that $Z_n$ converges in distribution to $Z$ as $n \to \infty$, denoted $Z_n \underset{d}{\longrightarrow} Z$, if for all $u$ at which $F(u) = \mathbb{P}[Z \leq u]$ is continuous, $F_n(u) \to F(u)$ as $n \to \infty$. We refer to $Z$ and its distribution $F(u)$ as the asymptotic distribution, large sample distribution, or limit distribution of $Z_n$.

# Weak Law of Large Numbers

大数定律

- **Weak Law of Large Numbers (WLLN):** If $Y_i \in \mathbb{R}^k$ are i.i.d. and $\mathbb{E}\|Y\| < \infty$, then as $n \to \infty$,

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[p]{} E[Y]$$

- **Theorem:** If $Y_i \in \mathbb{R}^k$ are i.i.d., $h(y) : \mathbb{R}^k \to \mathbb{R}^q$, and $\mathbb{E}\|h(Y)\| < \infty$, then $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} h(Y_i) \xrightarrow[p]{} \mu = \mathbb{E}[h(Y)]$ as $n \to \infty$.
- **Definition:** An estimator $\widehat{\theta}$ of $\theta$ is consistent if $\widehat{\theta} \xrightarrow[p]{} \theta$ as $n \to \infty$.

# Central Limit Theorem

- **Multivariate Lindeberg-Lévy Central Limit Theorem (CLT):** If $Y_i \in \mathbb{R}^k$ are i.i.d. and $\mathbb{E}\|Y\|^2 < \infty$, then as $n \to \infty$

$$\sqrt{n}(\bar{Y} - \mu) \xrightarrow[d]{} \mathrm{N}(0, V)$$

  where $\mu = \mathbb{E}[Y]$ and $\boldsymbol{V} = \mathbb{E}\left[(Y - \mu)(Y - \mu)'\right]$.
- The central limit theorem shows that the distribution of the sample mean is approximately normal in large samples.

# Continuous Mapping Theorem and Delta Method

- **Continuous Mapping Theorem (CMT):** Let $Z_n \in \mathbb{R}^k$ and $g(u) : \mathbb{R}^k \to \mathbb{R}^q$. If $Z_n \underset{p}{\longrightarrow} c$ as $n \to \infty$ and $g(u)$ is continuous at $c$ then $g\left(Z_n\right) \underset{p}{\longrightarrow} g(c)$ as $n \to \infty$.

# Continuous Mapping Theorem and Delta Method

三明治法则

- **Delta Method:** Let $\mu \in \mathbb{R}^k$ and $g(u) : \mathbb{R}^k \to \mathbb{R}^q$. If $\sqrt{n}(\widehat{\mu} - \mu) \underset{d}{\longrightarrow} \xi$, where $g(u)$ is continuously differentiable in a neighborhood of $\mu$, then as $n \to \infty$

$$\sqrt{n}(g(\widehat{\mu}) - g(\mu)) \underset{d}{\longrightarrow} G'\xi$$

where $\boldsymbol{G}(u) = \frac{\partial}{\partial u} g(u)'$ and $\boldsymbol{G} = \boldsymbol{G}(\mu)$. In particular, if $\xi \sim \mathrm{N}(0, \boldsymbol{V})$ then as $n \to \infty$

$$\sqrt{n}(g(\widehat{\mu}) - g(\mu)) \underset{d}{\longrightarrow} \mathrm{N}\left(0, \boldsymbol{G}'\boldsymbol{V}\boldsymbol{G}\right)$$

# Assumptions

- Recall the model $Y = X'\beta + e$ with the linear projection coefficient $\beta = \left(\mathbb{E}\left[XX'\right]\right)^{-1}\mathbb{E}[XY]$.
- Assumptions:
  1. The variables $(Y_i, X_i)$, $i = 1, \ldots, n$, are i.i.d.
  2. $\mathbb{E}\left[Y^2\right] < \infty$.
  3. $\mathbb{E}\|X\|^2 < \infty$.
  4. $\boldsymbol{Q}_{XX} = \mathbb{E}\left[XX'\right]$ is positive definite.

# Consistency of Least Squares Estimator

- Observe that the OLS estimator

$$\widehat{\beta} = \left( \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} X_i Y_i \right) = \hat{\mathbf{Q}}_{XX}^{-1} \widehat{\mathbf{Q}}_{XY}$$

  is a function of the sample moments $\widehat{\boldsymbol{Q}}_{XX} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i'$ and $\widehat{\boldsymbol{Q}}_{XY} = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i$.
- Specifically, the fact that $(Y_i, X_i)$ are mutually i.i.d. implies that any function of $(Y_i, X_i)$ is i.i.d., including $X_i X_i'$ and $X_i Y_i$. These variables also have finite expectations.
- Under these conditions, the WLLN implies that as $n \to \infty$,

$$\widehat{\boldsymbol{Q}}_{XX} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \xrightarrow[p]{} \mathbb{E}\left[ XX' \right] = \boldsymbol{Q}_{XX}$$

  and

$$\widehat{\mathbf{Q}}_{XY} = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i \xrightarrow[p]{} \mathbb{E}[XY] = \boldsymbol{Q}_{XY}$$

# Consistency of Least Squares Estimator

- The CMT allows us to combine these equations to show that $\widehat{\beta}$ converges in probability to $\beta$.
- Specifically, as $n \to \infty$,

$$\widehat{\beta} = \widehat{\boldsymbol{Q}}_{XX}^{-1} \widehat{\boldsymbol{Q}}_{XY} \underset{p}{\longrightarrow} \boldsymbol{Q}_{XX}^{-1} \boldsymbol{Q}_{XY} = \beta$$

- We have shown that $\widehat{\beta} \underset{p}{\longrightarrow} \beta$ as $n \to \infty$. In words, the OLS estimator converges in probability to the projection coefficient vector $\beta$ as the sample size $n$ gets large.
- It states that the OLS estimator $\widehat{\beta}$ converges in probability to $\beta$ as $n$ increases and thus $\widehat{\beta}$ is consistent for $\beta$.

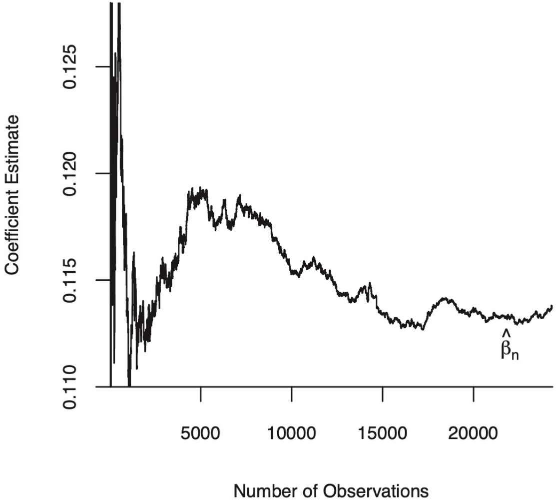
# Consistency of Least Squares Estimator

- To illustrate the effect of sample size on the least squares estimator consider the least squares regression.

$$\log(\text{ wage }) = \beta_1 \text{ education } + \beta_2 \text{ experience } + \beta_3 \text{ experience }^2 + \beta_4 + e$$

- We sequentially estimated the model by least squares starting with the first 5 observations and continuing until the full sample is used.

# Consistency of Least Squares Estimator



# Asymptotic Normality

- Consistency is a good first step, but in itself does not describe the distribution of the estimator.
- Now we derive an approximation typically called **the asymptotic distribution**.

# Asymptotic Normality

$$\sqrt{n}(\widehat{\beta}-\beta) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i e_i\right)$$

- The random pairs $(Y_i, X_i)$ are i.i.d., so any function of $(Y_i, X_i)$ is also i.i.d. This includes $e_i = Y_i - X_i'\beta$ and the product $X_i e_i$. The latter is mean-zero ($\mathbb{E}[Xe] = 0$) and has $k \times k$ covariance matrix

$$\Omega = \mathbb{E}\left[(Xe)(Xe)'\right] = \mathbb{E}\left[XX'e^2\right]$$

# Asymptotic Normality

- $\Omega$ has finite elements under some assuptions:
  1. The variables $(Y_i, X_i)$, $i = 1, \ldots, n$, are i.i.d..
  2. $\mathbb{E}\left[Y^4\right] < \infty$.
  3. $\mathbb{E}\|X\|^4 < \infty$.
  4. $\boldsymbol{Q}_{XX} = \mathbb{E}\left[XX'\right]$ is positive definite.

# Asymptotic Normality

- Since $X_i e_i$ is i.i.d., mean zero, and finite variance, the central limit theorem implies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i e_i \underset{d}{\longrightarrow} \mathrm{N}(0, \Omega)$$

- So we can derive

$$\sqrt{n}(\widehat{\beta} - \beta) \underset{d}{\longrightarrow} Q_{XX}^{-1} \, \mathrm{N}(0, \Omega) = \mathrm{N}\left(0, \boldsymbol{Q}_{XX}^{-1} \Omega \boldsymbol{Q}_{XX}^{-1}\right)$$

as $n \to \infty$.

# Asymptotic Normality

- **Theorem: Asymptotic Normality of Least Squares Estimator** Under the above assumptions, as $n \to \infty$

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow[d]{} \mathrm{N}\left(0, V_\beta\right)$$

where $\boldsymbol{Q}_{XX} = \mathbb{E}\left[XX'\right], \Omega = \mathbb{E}\left[XX'e^2\right]$, and

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

# Asymptotic Normality

- Compare the variance of the asymptotic distribution and the finite-sample conditional variance in the CEF model:

$$\boldsymbol{V}_{\widehat{\beta}} = \operatorname{var}[\widehat{\beta} \mid \boldsymbol{X}] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right) \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

- Notice that $\boldsymbol{V}_{\widehat{\beta}}$ is the exact conditional variance of $\widehat{\beta}$ and $\boldsymbol{V}_{\beta}$ is the asymptotic variance of $\sqrt{n}(\widehat{\beta} - \beta)$. Thus $\boldsymbol{V}_{\beta}$ should be (roughly) $n$ times as large as $\boldsymbol{V}_{\widehat{\beta}}$, or $\boldsymbol{V}_{\beta} \approx n\boldsymbol{V}_{\widehat{\beta}}$.

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

- which looks like an estimator of $\boldsymbol{V}_{\beta}$. Indeed, as $n \to \infty$, $n\boldsymbol{V}_{\widehat{\beta}} \xrightarrow[p]{} \boldsymbol{V}_{\beta}$.

# Asymptotic Normality

- There is a special case where $\Omega$ and $\boldsymbol{V}_\beta$ simplify. Suppose that

$$\operatorname{cov}\left(XX', e^2\right) = 0$$

- It holds in the homoskedastic linear regression model but is somewhat broader. The asymptotic variance formulae simplify as

$$\Omega = \mathbb{E}\left[XX'\right]\mathbb{E}\left[e^2\right] = \boldsymbol{Q}_{XX}\sigma^2$$
$$\boldsymbol{V}_\beta = \boldsymbol{Q}_{XX}^{-1}\Omega\boldsymbol{Q}_{XX}^{-1} = \boldsymbol{Q}_{XX}^{-1}\sigma^2 \equiv \boldsymbol{V}_\beta^0$$

- We call $V_\beta^0$ the homoskedastic asymptotic covariance matrix.

# Consistency of Error Variance Estimators

- We can show that the estimators $\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^n \widehat{e}_i^2$ and $s^2 = (n-k)^{-1}\sum_{i=1}^n \widehat{e}_i^2$ are consistent for $\sigma^2$
- The trick is to write the residual $\widehat{e}_i$ as equal to the error $e_i$ plus a deviation

$$\widehat{e}_i = Y_i - X_i'\widehat{\beta} = e_i - X_i'(\widehat{\beta} - \beta)$$

- Thus the squared residual equals the squared error plus a deviation

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

- So when we take the average of the squared residuals we obtain the average of the squared errors, plus two terms which are (hopefully) asymptotically negligible. This average is:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n e_i^2 - 2\left(\frac{1}{n}\sum_{i=1}^n e_i X_i'\right)(\widehat{\beta} - \beta) + (\widehat{\beta} - \beta)'\left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)(\widehat{\beta} - \beta)$$

# Consistency of Error Variance Estimators

- The WLLN implies that

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

$$\frac{1}{n}\sum_{i=1}^{n} e_i X_i' \underset{p}{\longrightarrow} \mathbb{E}\left[eX'\right] = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \underset{p}{\longrightarrow} \mathbb{E}\left[XX'\right] = Q_{XX}$$

- Since $\widehat{\beta} \underset{p}{\longrightarrow} \beta$, $\widehat{\sigma}^2$ converges in probability to $\sigma^2$ as desired.
- Since $n/(n-k) \to 1$ as $n \to \infty$, it follows that $s^2 = \left(\frac{n}{n-k}\right)\widehat{\sigma}^2 \underset{p}{\longrightarrow} \sigma^2$. Thus both estimators are consistent.

# Homoskedastic Covariance Matrix Estimation

- We have shown that $\sqrt{n}(\widehat{\beta} - \beta)$ is asymptotically normal with asymptotic covariance matrix $\boldsymbol{V}_\beta$.
- For asymptotic inference (confidence intervals and tests) we need a consistent estimator of $\boldsymbol{V}_\beta$.
- Under homoskedasticity $\boldsymbol{V}_\beta$ simplifies to $\boldsymbol{V}_\beta^0 = \boldsymbol{Q}_{XX}^{-1}\sigma^2$ and now we consider the simplified problem of estimating $\boldsymbol{V}_\beta^0$.

# Homoskedastic Covariance Matrix Estimation

- The standard moment estimator of $\boldsymbol{Q}_{XX}$ is $\widehat{\mathbf{Q}}_{XX}$ and thus an estimator for $\boldsymbol{Q}_{XX}^{-1}$ is $\widehat{\boldsymbol{Q}}_{XX}^{-1}$.
- The standard estimator of $\sigma^2$ is the unbiased estimator $s^2$.
- Thus a natural plug-in estimator for $\boldsymbol{V}_{\boldsymbol{\beta}}^0 = \boldsymbol{Q}_{XX}^{-1}\sigma^2$ is $\widehat{\boldsymbol{V}}_{\beta}^0 = \widehat{\boldsymbol{Q}}_{XX}^{-1} \; \mathrm{s}^2$.
- Consistency of $\widehat{\boldsymbol{V}}_{\beta}^0$ for $\boldsymbol{V}_{\beta}^0$ follows from consistency of the moment estimators $\widehat{\boldsymbol{Q}}_{XX}$ and $s^2$ and an application of the continuous mapping theorem.

$$\widehat{\boldsymbol{V}}_{\beta}^0 = \widehat{\boldsymbol{Q}}_{XX}^{-1} s^2 \underset{p}{\longrightarrow} \boldsymbol{Q}_{XX}^{-1}\sigma^2 = \boldsymbol{V}_{\beta}^0$$

# Heteroskedastic Covariance Matrix Estimation

- We have established the asymptotic covariance matrix of $\sqrt{n}(\widehat{\beta} - \beta)$ is $V_\beta = \boldsymbol{Q}_{XX}^{-1} \Omega \boldsymbol{Q}_{XX}^{-1}$.
- We now consider estimation of this covariance matrix without imposing homoskedasticity.
- The standard approach is to use a plug-in estimator which replaces the unknowns with sample moments.
- A natural estimator for $\boldsymbol{Q}_{XX}^{-1}$ is $\widehat{\boldsymbol{Q}}_{XX}^{-1}$.
- The moment estimator for $\Omega$ is

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

- leading to the plug-in covariance matrix estimator and the estimator is consistent

$$\widehat{\boldsymbol{V}}_\beta^{\mathrm{HC0}} = \widehat{\boldsymbol{Q}}_{XX}^{-1} \widehat{\Omega} \widehat{\boldsymbol{Q}}_{XX}^{-1}$$

# Summary of Covariance Matrix Notation

- The exact variance of $\widehat{\beta}$ (under the assumptions of the linear regression model) and the asymptotic variance of $\sqrt{n}(\widehat{\beta} - \beta)$ (under the more general assumptions of the linear projection model) are

$$\boldsymbol{V}_{\widehat{\beta}} = \mathsf{var}[\widehat{\beta} \mid \boldsymbol{X}] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right) \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

$$\boldsymbol{V}_{\beta} = \mathsf{avar}[\sqrt{n}(\widehat{\beta} - \beta)] = \boldsymbol{Q}_{XX}^{-1} \Omega \boldsymbol{Q}_{XX}^{-1}$$

- The HC0 estimators of these two covariance matrices are

$$\widehat{\boldsymbol{V}}_{\widehat{\beta}}^{\mathrm{HC0}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\sum_{i=1}^{n} X_i X_i' \widehat{e}_i^2\right) \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

$$\widehat{\boldsymbol{V}}_{\beta}^{\mathrm{HC0}} = \widehat{\boldsymbol{Q}}_{XX}^{-1} \widehat{\Omega} \widehat{\boldsymbol{Q}}_{XX}^{-1}$$

and satisfy the simple relationship $\widehat{\boldsymbol{V}}_{\beta}^{\mathrm{HC0}} = n \widehat{\boldsymbol{V}}_{\widehat{\beta}}^{\mathrm{H}}$.

# Summary of Covariance Matrix Notation

- Similarly, under the assumption of homoskedasticity the exact and asymptotic variances simplify to

$$\boldsymbol{V}^0_{\widehat{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sigma^2$$
$$\boldsymbol{V}^0_{\beta} = \boldsymbol{Q}^{-1}_{XX}\sigma^2$$

- Their standard estimators are

$$\widehat{\boldsymbol{V}}^0_{\widehat{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}s^2$$
$$\widehat{\boldsymbol{V}}^0_{\beta} = \widehat{\boldsymbol{Q}}^{-1}_{XX}s^2$$

which also satisfy the relationship $\widehat{\boldsymbol{V}}^0_{\beta} = n\widehat{\boldsymbol{V}}^0_{\widehat{\beta}}$.

# Functions of Parameters

- In most serious applications a researcher is actually interested in a specific transformation of the coefficient vector $\beta = (\beta_1, \ldots, \beta_k)$.
- For example, the researcher may be interested in a single coefficient $\beta_j$ or a ratio $\beta_j / \beta_l$.
- In any of these cases we can write the parameter of interest $\theta$ as a function of the coefficients, e.g. $\theta = r(\beta)$ for some function $r : \mathbb{R}^k \to \mathbb{R}^q$. The estimate of $\theta$ is

$$\widehat{\theta} = r(\widehat{\beta})$$

- By the continuous mapping theorem and the fact $\widehat{\beta} \xrightarrow[p]{} \beta$ we can deduce that $\widehat{\theta}$ is consistent for $\theta$ if the function $r(\cdot)$ is continuous.

# Functions of Parameters

- **Assumption:** $r(\beta) : \mathbb{R}^k \to \mathbb{R}^q$ is continuously differentiable at the true value of $\beta$ and $\boldsymbol{R} = \frac{\partial}{\partial \beta} r(\beta)'$ has rank $q$.
- **Theorem: Asymptotic Distribution of Functions of Parameters** Under the above assumptions, as $n \to \infty$,

$$\sqrt{n}(\widehat{\theta} - \theta) \underset{d}{\longrightarrow} \mathrm{N}\left(0, V_\theta\right)$$

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

## Functions of Parameters

- In many cases the function $r(\beta)$ is linear:

$$r(\beta) = \boldsymbol{R}'\beta$$

  for some $k \times q$ matrix $R$. In particular if $R$ is a "selector matrix"

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{I} \\ 0 \end{pmatrix}$$

- then we can partition $\beta = (\beta_1', \beta_2')'$ so that $\boldsymbol{R}'\beta = \beta_1$. Then

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

- Thus

$$\sqrt{n}\left(\widehat{\beta}_1 - \beta_1\right) \xrightarrow[d]{} \mathrm{N}\left(0, \boldsymbol{V}_{11}\right)$$

# Functions of Parameters

- To illustrate the case of a nonlinear transformation take the example $\theta = \beta_j/\beta_l$ for $j \neq l$. Then

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

- so

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

where $\boldsymbol{V}_{ab}$ denotes the $ab^{\text{th}}$ element of $\boldsymbol{V}_\beta$.

# Asymptotic Standard Errors

- A standard error is an estimator of the standard deviation of the distribution of an estimator. Thus
$$s\left(\widehat{\beta}_j\right) = \sqrt{\widehat{\boldsymbol{V}}_{\widehat{\beta}_j}} = \sqrt{\left[\widehat{\boldsymbol{V}}_{\widehat{\beta}}\right]_{jj}}$$

- Standard errors for $\widehat{\theta}$ are constructed similarly. Supposing that $\theta = h(\beta)$ is real-valued then the standard error for $\widehat{\theta}$ is
$$s(\widehat{\theta}) = \sqrt{\widehat{\boldsymbol{R}}'\widehat{\boldsymbol{V}}_{\widehat{\beta}}\widehat{\boldsymbol{R}}} = \sqrt{n^{-1}\widehat{\boldsymbol{R}}'\hat{\boldsymbol{V}}_{\beta}\widehat{\boldsymbol{R}}}$$

- When the justification is based on asymptotic theory we call $s(\widehat{\beta}_j)$ or $s(\widehat{\theta})$ an asymptotic standard error for $\widehat{\beta}_j$ or $\hat{\theta}$.

## t-statistic

- Let $\theta = r(\beta) : \mathbb{R}^k \to \mathbb{R}$ be a parameter of interest, $\widehat{\theta}$ its estimator, and $s(\widehat{\theta})$ its asymptotic standard error. Then the t-statistic is

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

- We know that $\sqrt{n}(\widehat{\theta} - \theta) \underset{d}{\longrightarrow} \mathrm{N}\left(0, V_\theta\right)$ and $\widehat{V}_\theta \underset{p}{\longrightarrow} V_\theta$. Thus

$$\begin{aligned} T(\theta) &= \frac{\widehat{\theta} - \theta}{s(\hat{\theta})} \\ &= \frac{\sqrt{n}(\widehat{\theta} - \theta)}{\sqrt{\widehat{V}_\theta}} \\ &\underset{d}{\longrightarrow} \frac{\mathrm{N}\left(0, V_\theta\right)}{\sqrt{V_\theta}} \\ &= Z \sim \mathrm{N}(0, 1) \end{aligned}$$

# t-statistic

- It is also useful to consider the distribution of the absolute t-ratio $|T(\theta)|$.
- Since $T(\theta) \underset{d}{\longrightarrow} Z$ the continuous mapping theorem yields $|T(\theta)| \underset{d}{\longrightarrow} |Z|$.
- Letting $\Phi(u) = \mathbb{P}[Z \leq u]$ denote the standard normal distribution function we calculate that the distribution of $|Z|$ is

$$\begin{aligned}
\mathbb{P}[|Z| \leq u] &= \mathbb{P}[-u \leq Z \leq u] \\
&= \mathbb{P}[Z \leq u] - \mathbb{P}[Z < -u] \\
&= \Phi(u) - \Phi(-u) \\
&= 2\Phi(u) - 1
\end{aligned}$$

# Confidence Intervals

- An interval estimator $\widehat{C}$ is called a confidence interval when the goal is to set the coverage probability to equal a pre-specified target such as 90% or 95% .
- $\widehat{C}$ is called a $1-\alpha$ confidence interval if $\inf_\theta \mathbb{P}_\theta[\theta \in \widehat{C}] = 1-\alpha$.
- When $\widehat{\theta}$ is asymptotically normal with standard error $s(\widehat{\theta})$ the conventional confidence interval for $\theta$ takes the form

$$\widehat{C} = [\widehat{\theta} - c \times s(\widehat{\theta}), \quad \widehat{\theta} + c \times s(\widehat{\theta})]$$

  where $c$ equals the $1-\alpha$ quantile of the distribution of $|Z|$.
- Using (7.34) we calculate that $c$ is equivalently the $1-\alpha/2$ quantile of the standard normal distribution. Thus, $c$ solves

$$2\Phi(c) - 1 = 1 - \alpha$$

# Confidence Intervals

- Equivalently, it is the set of parameter values for $\theta$ such that the t-statistic $T(\theta)$ is smaller (in absolute value) than $c$, that is

$$\widehat{C} = \{\theta : |T(\theta)| \leq c\} = \left\{\theta : -c \leq \frac{\widehat{\theta} - \theta}{s(\widehat{\theta})} \leq c\right\}$$

- The coverage probability of this confidence interval is

$$\mathbb{P}[\theta \in \widehat{C}] = \mathbb{P}[|T(\theta)| \leq c] \to \mathbb{P}[|Z| \leq c] = 1 - \alpha$$

  where the limit is taken as $n \to \infty$, and holds because $T(\theta)$ is asymptotically $|Z|$ .
- We call the limit the asymptotic coverage probability and call $\widehat{C}$ an asymptotic $1 - \alpha\%$ confidence interval for $\theta$.

# Confidence Intervals

- The standard coverage probability for confidence intervals is 95%, leading to the choice $c = 1.96$.
- Rounding 1.96 to 2 , we obtain the most commonly used confidence interval in applied econometric practice

$$\widehat{C} = [\widehat{\theta} - 2s(\widehat{\theta}), \quad \widehat{\theta} + 2s(\widehat{\theta})]$$

- This is a useful rule-of thumb.

# Regression Intervals

- In the linear regression model the conditional expectation of $Y$ given $X = x$ is

$$m(x) = \mathbb{E}[Y \mid X = x] = x'\beta$$

- In some cases we want to estimate $m(x)$ at a particular point $x$. Notice that this is a linear function of $\beta$.
- Letting $r(\beta) = x'\beta$ and $\theta = r(\beta)$ we see that $\hat{m}(x) = \widehat{\theta} = x'\widehat{\beta}$ and $\boldsymbol{R} = x$ so $s(\widehat{\theta}) = \sqrt{x'\widehat{\boldsymbol{V}}_{\widehat{\beta}}x}$. Thus an asymptotic 95% confidence interval for $m(x)$ is

$$\left[x'\widehat{\beta} \pm 1.96\sqrt{x'\hat{\boldsymbol{V}}_{\widehat{\beta}}x}\right]$$

- It is interesting to observe that if this is viewed as a function of $x$ the width of the confidence interval is dependent on $x$.

# Regression Intervals

- To illustrate we return to the log wage regression. The estimated regression equation is

$$\log(wage) = x'\widehat{\beta} = 0.155x + 0.698$$

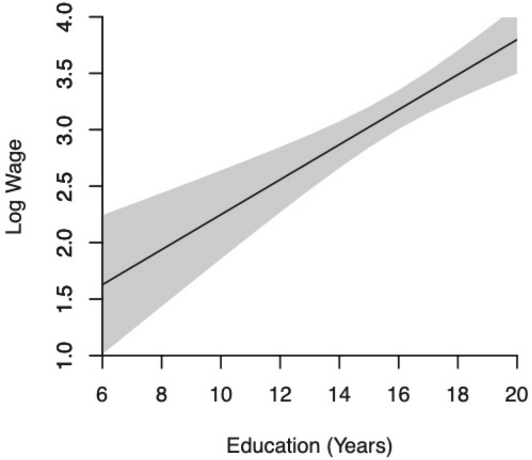
where $x =$ education. The covariance matrix estimate is

$$\widehat{\boldsymbol{V}}_{\widehat{\beta}} = \begin{pmatrix} 0.001 & -0.015 \\ -0.015 & 0.243 \end{pmatrix}$$

- Thus the 95% confidence interval for the regression is

$$0.155x + 0.698 \pm 1.96\sqrt{0.001x^2 - 0.030x + 0.243}$$

# Regression Intervals




(a) Wage on Education