Chapter 4 Large Sample Distribution Theory and Large Sample Properties of the Classical Regression Model

(5th , Appendix D & Chap.5)

4.1 Some basic concepts

4.1.1 σ-algebra, Probability Space and Distribution Function

Definition 4.1 σ-algebra

Let Ω be a sample space. A nonempty class F of subsets of Ω is an **algebra** if

(i) $\bar{A} \in F$ whenever $A \in F$,

(ii) $A_1 \cup A_2 \in F$ whenever $A_i \in F,\ i = 1, 2$.

Moreover, F is a **σ-algebra** if, in addition,

(iii) $\bigcup_{i=1}^{\infty} A_i \in F$ whenever $A_i \in F,\ i = 1, 2, \cdots$

Definition 4.2 Probability Space

Three elements for a probability space: Sample Space Ω, class of subsets of Ω, F and probability measure P.

Given a σ-algebra, a probability measure $P(A)$ is a real-valued set function defined on F, satisfying

(i) $P(\Omega) = 1$;

(ii) For any $A_j \in F$, $0 \le P(A_j) \le 1$;

(iii) For any $A_j \in F$ $(j = 1, 2, ...)$, if $A_j \cap A_i = \varnothing$, $i \neq j$, $P(\cup A_j) = \sum_j P(A_j)$.

The triplet $\{\Omega, F, P\}$ is a **probability space**.

Definition 4.3 Random variable

A real-valued measurable function X on a probability space $\{\Omega, F, P\}$ is called a **random variable** if for any real number x, $\{\omega | X(\omega) < x\} \in F$, where $\omega \in \Omega$.

Definition 4.4 Distribution Function

Associated with any random variable X on a probability space $\{\Omega, F, P\}$ is a real function $F(x)$ called the **distribution function** of the random variable X and defined by

$F(x) = P\{\omega | X(\omega) < x\},\ x \in (-\infty, +\infty)$.

4.1.2 Various modes of convergence

Definition 4.5 Convergence in Probability

Let $\{X_n\}$ be a sequence of random variables, if there exists a real number X such that for every $\varepsilon > 0$, $\lim_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1$ or $\lim_{n\to\infty} P\{|x_n - X| \geq \varepsilon\} = 0$, then X_n convergences in probability to X, written $X_n \xrightarrow{p} X$ or $\text{plim}\ X_n = X$.

This definition of convergence in probability can be extended to a sequence of random vectors (or random matrices). That is, a sequence of K-dimensional random vectors $\{X_n\}$ convergence in probability to a K-dimensional vector X if for any $\varepsilon > 0$, $\lim_{n\to\infty} P\{|X_{nk} - X_k| < \varepsilon\} = 1$ for all $k(k = 1, \cdots, K)$, where X_{nk} is the kth element of X_n, and X_k is the kth element of X .

Definition 4.6 Consistent Estimator

An estimator $\hat{\theta}_n$ of a parameter θ is a consistent estimator of θ if and only if $\hat{\theta}_n \xrightarrow{p} \theta$ or $\text{plim}\ \hat{\theta}_n = \theta$.

Theorem 4.1 Consistency of the Sample Mean

The mean of a random sample from any population with finite mean μ and finite variance σ^2 is a consistent estimator of μ .

Corollary 4.1 Consistency of a Mean of Functions

In random sampling, for any function $g(X)$, if $E[g(X)]$ and $Var[g(X)]$ are finite constants, then

$$\text{p}\lim \frac{1}{n}\sum g\left(X_i\right) = E\left[g\left(X\right)\right] \tag{4.1}$$

Definition 4.7　Convergence in Mean Squares

Let $\{X_n\}$ be a sequence of random variables, If $\lim_{n\to\infty} E\left(X_n - X\right)^2 = 0$, then X_n Convergences in Mean Squares to X, Written $X_n \xrightarrow{m.s} X$.

Definition 4.8　Almost Sure Convergence

Let $\{X_n\}$ be a sequence of random variables. We say that X_n converges almost surely to X, written $X_n \xrightarrow{a.s} X$ if there exists a real number X such that $P\{\lim_{n\to\infty} X_n = X\} = 1$.

Definition 4.8'　Almost Sure Convergence

The random variable X_n converges almost surely to the constant X if and only if

$\lim_{n\to\infty} P\left\{\left|X_k - X\right| > \varepsilon, \forall k \ge n\right\} = 0$ for all $\varepsilon > 0$.

Definition 4.9　Convergence in Distribution

The random variable X_n converges in distribution to a random variable X with cumulative distribution function $F(x)$ if $\lim_{n\to\infty} \left|F_n(x) - F(x)\right| = 0$ at all continuity points of $F(x)$. If X_n converges in distribution to X, where $F_n(x)$ is the cumulative distribution function of X_n, then $F(x)$ is the limiting distribution of $\{F_n(x)\}$. This is written $X_n \xrightarrow{d} X$.

The relations among the four convergence processes

(i) $X_n \xrightarrow{a.s} X \Rightarrow X_n \xrightarrow{p} X$;

(ii) $X_n \xrightarrow{m.s} X \Rightarrow X_n \xrightarrow{p} X$;

(iii) $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$.

4.1.3 Slutsky Theorem and its applications

Theorem 4.2　Slutsky Theorem

(i) Given a sequence of random variables $\{X_n\}$, such that $X_n \xrightarrow{p} X$, if $g(\cdot)$ is a continuous function of X, then

$$\text{plim } g(X_n) = g(X) \qquad (4.2).$$

(ii) Given two sequences of random variables $\{X_n\}$ and $\{Y_n\}$, if $X_n - Y_n \xrightarrow{p} 0$, and $Y_n \xrightarrow{d} Y$, then $X_n \xrightarrow{d} Y$.

Theorem 4.3 Rules for Probability Limits

If X_n and Y_n are random variables with $\text{plim } X_n = c_1$ and $\text{plim } Y_n = c_2$, then

(i) $\text{plim}(X_n + Y_n) = c_1 + c_2$;

(ii) $\text{plim}(X_n Y_n) = c_1 c_2$;

(iii) $\text{plim}(X_n / Y_n) = c_1 / c_2$ if $(c_2 \neq 0)$.

If W_n is a random matrix with $\text{plim } W_n = \Omega$, then $\text{plim } W_n^{-1} = \Omega^{-1}$.

If X_n and Y_n are random matrices with $\text{plim } X_n = A$ and $\text{plim } Y_n = B$, then $\text{plim}(X_n Y_n) = AB$.

Theorem 4.4 Rules for Limiting Distributions

(i) If $X_n \xrightarrow{p} \alpha$ and $Y_n \xrightarrow{d} Y$, then $X_n + Y_n \xrightarrow{d} Y + \alpha$, $X_n Y_n \xrightarrow{d} \alpha Y$ and $Y_n / X_n \xrightarrow{d} Y/\alpha$ if $\alpha \neq 0$;

(ii) If $X_n \xrightarrow{d} X$ and $g(X_n)$ is a continuous function of X_n, but not depend on n, then $g(X_n) \xrightarrow{d} g(X)$.

e.g., If $t_n \xrightarrow{d} N(0,1)$, then $t_n^2 \to \chi^2(1)$, $F(1,n) \to \chi^2(1)$, $J \cdot F(J,n) \to \chi^2(J)$.

4.2 Laws of Large Numbers and Central Limit Theorems

4.2.1 Laws of Large Numbers

For a sequence of random variables $\{Y_i\}$, the sample mean $\bar{Y}_n$ is defined as $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Consider the sequence $\{\bar{Y}_n\}$, Law of large numbers concern conditions under which $\{\bar{Y}_n\}$ converges either in probability or almost surely. An LLN is called “strong” if the convergence is almost surely and “weak” if the convergence is in probability.

Theorem 4.5　Chebychev’s Weak LLN: (Chebychev’s Inequality)

If $Y_i \quad (i=1,2,\cdots,n)$ is a sequence of random variables such that $E(Y_i)=\mu_i<\infty$ and $Var(Y_i)=\sigma_i^2<\infty$ such that $\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2=0$, then $\bar{Y}_n \xrightarrow{P} \bar{\mu}_n$, where $\bar{\mu}_n=\frac{1}{n}\sum_{i=1}^{n}\mu_i$.

Proof: $P\left\{\left|\bar{Y}_n-\bar{\mu}_n\right|>\varepsilon\right\}=P\left\{\left|\bar{Y}_n-E\left(\bar{Y}_n\right)\right|>\varepsilon\right\}\le\frac{\operatorname{var}\left(\bar{Y}_n\right)}{\varepsilon^2}=\frac{1}{\varepsilon^2}\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2$

$\therefore \lim_{n\to\infty}P\left\{\left|\bar{Y}_n-\bar{\mu}_n\right|>\varepsilon\right\}\le\lim_{n\to\infty}\frac{1}{\varepsilon^2}\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2=0$

$\therefore p\lim\bar{Y}_n=\bar{\mu}_n$.

Collorary 4.5 If $Y_i \quad (i=1,2,\cdots,n)$ is a sequence of random variables such that $\lim_{n\to\infty}E\left(\bar{Y}_n\right)=\mu$ and $\lim_{n\to\infty}\operatorname{var}\left(\bar{Y}_n\right)=0$, then $\bar{Y}_n \xrightarrow{P} \mu$.

Proof: $\lim_{n\to\infty}\operatorname{var}\left(\bar{Y}_n\right)=\lim_{n\to\infty}E\left(\bar{Y}_n-E\left(\bar{Y}_n\right)\right)^2=\lim_{n\to\infty}E\left(\bar{Y}_n-\mu\right)^2=0$

$\therefore \bar{Y}_n \xrightarrow{m.s.} \mu \Rightarrow \bar{Y} \xrightarrow{P} \mu$.

Theorem 4.6　Khinchine’s Weak LLN

If $Y_1,\cdots,Y_n$ are a random (*i.i.d.*) sample from a distribution with finite mean $E(Y_i)=\mu$, then $\bar{Y}_n \xrightarrow{p} \mu$.

Theorem 4.7　Kolmogorov’s Strong LLN

① If $Y_i(i=1,2,\cdots,n)$ is a sequence of independent and identically distributed random variables such that $E(Y_i)=\mu<\infty$, then $\bar{Y}_n \xrightarrow{a.s.} \mu$.

② If $Y_i (i=1,2,\cdots,n)$ is a sequence of independently distributed random variables such that $E(Y_i)=\mu_i<\infty$ and $Var(Y_i)=\sigma_i^2<\infty$ such that $\sum_{i=1}^{\infty}\sigma_i^2/i^2<\infty$ as $n\to\infty$, then $\bar{Y}_n \xrightarrow{a.s.} \bar{\mu}_n$.

Theorem 4.8 Markov LLN

If $\{Y_i\}$ is a sequence of independent random variables with $E(Y_i)=\mu_i<\infty$ and if for some $\delta>0$, $\sum_{i=1}^{\infty}\left(E\left|Y_i-\mu_i\right|^{1+\delta}\right)/i^{1+\delta}<\infty$, then $\bar{Y}_n-\bar{\mu}_n$ converges almost surely to 0, which we denote $\bar{Y}_n-\bar{\mu}_n \xrightarrow{a.s} 0$.

4.2.2 Central Limit Theorems (Greene 5th, P908)

Theorem 4.9 Lindberg-Levy CLT (Univariate)

If $X_1,\cdots,X_n$ are a random sample from a probability distribution with finite mean μ and finite variance σ^2 and $\bar{X}_n=(1/n)\sum_{i=1}^{n}X_i$, then

$$\sqrt{n}\left(\bar{X}_n-\mu\right)\xrightarrow{d}N\left[0,\sigma^2\right].$$

A proof appears in Rao(1973, P_{127}).

Theorem 4.10 Lindberg-Feller CLT (with Unequal Variances)

Suppose that $\{X_i\}, i=1,\cdots n,$ is a sequence of independent random variables with finite means μ_i and finite positive variances σ_i^2. Let

$\bar{\mu}_n=\frac{1}{n}(\mu_1+\mu_2+\cdots+\mu_n)$ and $\bar{\sigma}_n^2=\frac{1}{n}\left(\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2\right)$.

If no single term dominates this average variance, which we could state as $\lim_{n\to\infty}\max\left(\sigma_i^2\right)/\left(n\bar{\sigma}_n^2\right)=0$, and if the average variance converges to a finite constants, $\bar{\sigma}^2=\lim_{n\to\infty}\bar{\sigma}_n^2$, then

$$\sqrt{n}\left(\bar{X}_n-\bar{\mu}_n\right)\xrightarrow{d}N\left[0,\bar{\sigma}^2\right].$$

Theorem 4.11 Multivariate Lindberg-Levy CLT

If $X_1,\cdots,X_n$ are a random sample from a multivariate distribution with finite mean vector μ and finite positive definite covariance matrix Q, then

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} N[0, Q],$$

Where $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$.

Theorem 4.12 Multivariate Lindberg-Feller CLT

Suppose that $X_1, \cdots, X_n$ are a sample of random vectors such that $E(X_i) = \mu_i$, $Var(X_i) = Q_i$, and all mixed third moments of the multivariate distribution are finite. Let

$$\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i,$$

$$\bar{Q}_n = \frac{1}{n}\sum_{i=1}^{n} Q_i.$$

We assume that $\lim_{n\to\infty} \bar{Q}_n = Q$.

where Q is a finite, positive definite matrix, and that for every i,

$$\lim_{n\to\infty}\left(n\bar{Q}_n\right)^{-1} Q_i = \lim_{n\to\infty}\left(\sum_{i=1}^{n} Q_i\right)^{-1} Q_i = 0 .$$

We allow the means of the random vectors to differ, although in the cases that we will analyze, they will generally be identical. The second assumption states that individual components of the sum must be finite and diminish in significance. There is also an implicit assumption that the sum of matrices is nonsingular. Since the limiting matrix is nonsingular, the assumption must hold for large enough n, which is all that concerns us here. With these in place, the result is

$$\sqrt{n}\left(\bar{X}_n - \bar{\mu}_n\right) \xrightarrow{d} N[0, Q]..$$

4.2.3 The Delta Method

Suppose $\{X_n\}$ is a sequence of k-dimensional random vectors such that $X_n \xrightarrow{P} \mu$, and $\sqrt{n}(X_n - \mu) \xrightarrow{d} X$, where $X \sim N(0, \Sigma)$, and suppose $c(X_n)$ is a set of J continuous functions of X_n not involving n, $c(X_n)$ has continuous first derivatives evaluated at μ:

$$C(\mu) = \frac{\partial c(\mu)}{\partial \mu'} = \begin{pmatrix} \frac{\partial c_1}{\partial X_1} & \cdots & \frac{\partial c_1}{\partial X_K} \\ \cdots & \cdots & \cdots \\ \frac{\partial c_J}{\partial X_1} & \cdots & \frac{\partial c_J}{\partial X_K} \end{pmatrix}_{evaluated\ at\ \mu} \quad , \quad c(X_n) = \begin{pmatrix} c_1(X_n) \\ \vdots \\ c_J(X_n) \end{pmatrix}$$

Then $\sqrt{n}\left[c(X_n) - c(\mu)\right] \xrightarrow{d} N\left[0, C(\mu)\Sigma C(\mu)'\right]$ (4.3)

Proof: By the mean-value theorem from calculus, there exists a k-dimensional vector Y_n between X_n and μ such that

$$c(X_n) - c(\mu) = C(Y_n)(X_n - \mu)$$

Multiplying both sides by $\sqrt{n}$, we obtain

$$\sqrt{n}\left[c(X_n) - c(\mu)\right] = C(Y_n)\sqrt{n}(X_n - \mu)$$

Since Y_n is between X_n and μ, and $X_n \xrightarrow{P} \mu$, we have $Y_n \xrightarrow{P} \mu$, so by Slutsky theorem, we have

$$C(Y_n) \xrightarrow{P} C(\mu), \quad C(Y_n)\sqrt{n}(X_n - \mu) \xrightarrow{d} C(\mu)X$$

$$C(\mu)X \sim N\left[0, C(\mu)\Sigma C(\mu)'\right]$$

So,

$$\sqrt{n}\left[c(X_n) - c(\mu)\right] \xrightarrow{d} N\left[0, C(\mu)\Sigma C(\mu)'\right].$$

In particular, if $\{X_n\}$ is a sequence of random variables and $\sqrt{n}(X_n - \mu) \xrightarrow{d} N(0, \sigma^2)$, then

$$\sqrt{n}\left[g(X_n) - g(\mu)\right] \xrightarrow{d} N\left[0, \{g'(\mu)\}^2 \sigma^2\right].$$

Where $g(X_n)$ is a continuous function not involving n.

Notice that the mean and variance of the limiting distribution are the mean and variance of the linear Taylor series approximation:

$$g(X_n) \simeq g(\mu) + g'(\mu)(X_n - \mu). \quad (4.4)$$

Chapter 4 Large Sample Distribution Theory and Large Sample Properties of the Classical Regression Model （part 2）

(5th , Appendix D & Chap.5)

4.3 Asymptotic Distributions

4.3.1 Asymptotic Distribution

Definition 4.10 Asymptotic Distribution

An asymptotic distribution is a distribution that is used to approximate the true finite sample distribution of a random variable.

e.g., If $\sqrt{n}\left[(\bar{x}_n-\mu)/\sigma\right]\xrightarrow{d}N\left[0,1\right]$, then approximately, or asymptotically, $X_n \sim N\left(\mu,\sigma^2/n\right)$, which we write as $X_n \overset{a}{\sim} N\left(\mu,\sigma^2/n\right)$.

Large-Sample Properties of the Estimator

Definition 4.11　Consistency

An estimator $\hat{\theta}_n$ of a parameter θ is a consistent estimator of θ if and only if $P\lim\hat{\theta}_n=\theta$.

Definition 4.12　Consistent and Asymptotic Normal Estimator

For an unknown parameter θ, if an estimator $\hat{\theta}_n$ satisfies $\sqrt{n}\left(\hat{\theta}_n-\theta\right)\overset{a}{\sim}N\left(0,V\right)$, then $\hat{\theta}_n$ is the CAN estimator of θ. The asymptotic covariance matrix of $\hat{\theta}_n$ is $\frac{V}{n}$. Written $Asy.\text{var}\left(\hat{\theta}_n\right)=\frac{V}{n}$ or $A\text{var}\left(\hat{\theta}_n\right)=\frac{V}{n}$.

Definition 4.13　Asymptotic Efficiency

The estimator is asymptotically efficient if it is consistent and has an asymptotic covariance matrix that is not larger than the asymptotic covariance matrix of any other consistent estimator.

Discussion　about　Asymptotic Expectation

Suppose that the estimator $\hat{\theta}_n$ satisfies $\sqrt{n}\left(\hat{\theta}_n-\theta\right)\overset{a}{\sim}N\left(0,V\right)$, then asymptotic expectation of $\hat{\theta}_n$ is θ. Written as $AE\left(\hat{\theta}_n\right)=\theta$.

There are at least three possible definitions of asymptotic unbiasedness:

(A) $\lim_{n\to\infty}E\left(\hat{\theta}_n\right)=\theta$;

(B) $AE\left(\hat{\theta}_n\right)=\theta$;

(C) $p\lim\hat{\theta}_n=\theta$.

In most cases encountered in practice, the estimator in hand will have all three properties, so there is no ambiguity. It is not difficult to construct cases in which the left-hand sides of all three definitions are different, however.

4.4 Asymptotic Distribution of the LS Estimator

4.4.1 Assumptions

A.1 linear model $y_i = x_i'\beta+\varepsilon_i \quad (i=1,\cdots,n)$

A.2 X is an $n\times K$ matrix, $P\{rank(X)=K\}=1$

A.3 $E(\varepsilon_i|X)=0$

A.4 $E(\varepsilon\varepsilon'|X)=\sigma^2 I$

A.5 $(x_i,\varepsilon_i) \quad (i=1,\cdots,n)$ is a sequence of independent observations.

$p\lim\frac{1}{n}X'X=Q$, (Q is a finite positive matrix)

4.4.2 Asymptotic properties of the LS Estimators

1) Consistency of $\hat{\beta}$

$$\hat{\beta}=\beta+(X'X)^{-1}X'\varepsilon=\beta+\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}$$

$$p\lim\hat{\beta}=\beta+p\lim\left(\frac{X'X}{n}\right)^{-1}\cdot p\lim\frac{X'\varepsilon}{n}$$

Where $\frac{X'\varepsilon}{n}=\frac{1}{n}\sum_{i=1}^{n}x_i\varepsilon_i=\frac{1}{n}\sum_{i=1}^{n}w_i=\overline{w}_n$, $x_i=\begin{pmatrix}1\\x_{i1}\\\vdots\\x_{iK}\end{pmatrix}$.

$$E(w_i)=E_X\left[E(w_i|X)\right]=E_X\left[E(x_i\varepsilon_i|X)\right]=E_X\left[x_iE(\varepsilon_i|X)\right]=0 \quad \text{……………….(4.5)}$$

Thus from $E(\varepsilon_i|X)=0$,we can get $E(w_i|X)=0$, $E(\overline{w}_n|X)=0$, $E(\overline{w}_n)=0$

$$\text{var}\left(\overline{w}_n\right) = E_X\left[\text{var}\left(\overline{w}_n \middle| X\right)\right] + \text{var}\left[E\left(\overline{w}_n \middle| X\right)\right] = E_X\left[\text{var}\left(\overline{w}_n \middle| X\right)\right]$$

$$= E_X\left[E\left(\overline{w}_n \overline{w}_n' \middle| X\right)\right] = E_X\left[E\left(\frac{1}{n} X'\varepsilon\varepsilon' X \frac{1}{n} \middle| X\right)\right]$$

$$= E_X\left[\frac{1}{n} X' E\left(\varepsilon\varepsilon' \middle| X\right)\frac{1}{n} X\right] = \frac{\sigma^2}{n} E\left(\frac{X'X}{n}\right) \text{..(4.6)}$$

The variance will collapse to zero if the expectation in parentheses is (or converges to) a constant matrix, so that the leading scalar will dominate the product as n increases. Assumption A.5 should be sufficient. It then follows that $\lim_{n\to\infty} \text{var}\left(\overline{w}_n\right) = 0 \cdot Q = 0$.

By Chebychev's weak LLN, $p\lim \frac{X'\varepsilon}{n} = 0$ ………………………………… ..(4.7)

$\therefore \quad p\lim \hat{\beta} = \beta + Q^{-1} \cdot 0 = \beta$ ……………………………………………………(4.8)

2) Consistency of s^2

$$s^2 = \hat{\sigma}^2 = \frac{e'e}{n-K} = \frac{\varepsilon' M \varepsilon}{n-K} = \frac{n}{n-K}\left[\frac{\varepsilon'\varepsilon}{n} - \frac{\varepsilon' X}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}\right]$$

$$p\lim s^2 = \lim \frac{n}{n-K} \cdot p\lim\left[\frac{\varepsilon'\varepsilon}{n} - \frac{\varepsilon' X}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}\right]$$

$$\Rightarrow p\lim s^2 = p\lim \frac{\varepsilon'\varepsilon}{n} = p\lim \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 = p\lim \overline{\varepsilon}^2 .$$

$$E\left(\varepsilon_i^2\right) = E_X\left[E\left(\varepsilon_i^2 \middle| X\right)\right] = \sigma^2 \text{ ……………………………………………………..(4.9)}$$

This is a narrow case in which the random variables ε_i^2 are independent with the same finite mean σ^2, so not is required to get the mean to converge almost surely to $\sigma^2 = E\left(\varepsilon_i^2\right)$. By the Markov Theorem (Th 4.8), what is need is for $E\left[\left|\varepsilon_i^2\right|^{1+\delta}\right]$ to be finite, so the minimal assumption thus far is that ε_i have finite moments up to slightly greater than 2. Indeed, if we further assume that every ε_i has the same distribution, then by Khinchine Theorem (Th 4.6), finite moments (of ε_i) up to 2 is sufficient. So, under fairly weak condition, the first term in brackets converges to σ^2, which gives our result,

$$p\lim s^2 = \sigma^2 \text{ ……………………………………………………………………(4.10)}$$

4.4.3 Asymptotic Normality of the LS Estimator

$$\sqrt{n}\left(\hat{\beta}-\beta\right)=\sqrt{n}\left[\left(X'X\right)^{-1}X'\varepsilon\right]=\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{\sqrt{n}} \text{ ………………………………(4.11)}$$

We must establish the limiting distribution of $\frac{1}{\sqrt{n}}X'\varepsilon$,since $\overline{w}_n = \frac{1}{n}\sum_{i=1}^{n} x_i\varepsilon_i$,

from(4.5),we have

$$E\left(x_i\varepsilon_i\right)=E_X\left[E\left(x_i\varepsilon_i \middle| X\right)\right]=E_X\left[x_i E\left(\varepsilon_i \middle| X\right)\right]=0 \text{ ……………………………..(4.12)}$$

$$\therefore \quad E\left(\frac{X'\varepsilon}{\sqrt{n}}\right)=E\left(\sqrt{n}\overline{w}_n\right)=0$$

$$\operatorname{var}\left(x_i\varepsilon_i\right)=E_X\left[E\left(x_i\varepsilon_i\varepsilon' x_i' \middle| X\right)\right]=\sigma^2 E\left(x_i x_i'\right)=\sigma^2 Q_i \text{ …………………………(4.13)}$$

By A.5, $\{x_i\varepsilon_i\}$ is a sequence of independent vectors.

$$\operatorname{var}\left(\sqrt{n}\overline{w}_n\right)=\operatorname{var}\left(\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}x_i\varepsilon_i\right)\right)=\frac{1}{n}\sum_{i=1}^{n}\operatorname{var}\left(x_i\varepsilon_i\right)=\frac{\sigma^2}{n}\sum_{i=1}^{n}Q_i \text{ ………………(4.14)}$$

As long as the sum is not dominated by any particular term and the regressors are well behaved, which in this case means that A.5 holds,

$$\lim_{n\to\infty}\operatorname{var}\left(\sqrt{n}\overline{w}_n\right)=\lim_{n\to\infty}\sigma^2\overline{Q}_n=\sigma^2 Q \text{ ……………………………………………(4.15)}$$

So, according to Lindberg-Feller CLT (multivariate), we have

$$\sqrt{n}\overline{w}_n=\frac{X'\varepsilon}{\sqrt{n}}\xrightarrow{d}N\left(0,\sigma^2 Q\right) \text{ ………………………………………………..(4.16)}$$

$$p\lim\left(\frac{X'X}{n}\right)^{-1}=Q^{-1},$$

$$\therefore \quad \sqrt{n}\left(\hat{\beta}-\beta\right)\xrightarrow{d}N\left(0,\sigma^2 Q^{-1}\right) \text{ ……………………………………………..(4.17)}$$

i.e., $\hat{\beta}\overset{a}{\sim}N\left(\beta,\frac{\sigma^2}{n}Q^{-1}\right)$ …………………………………………………..(4.18)

The appropriate estimator of the asymptotic covariance matrix of $\hat{\beta}$ is

$$Est.A\operatorname{var}\left(\hat{\beta}\right)=s^2\left(X'X\right)^{-1}\xrightarrow{P}\frac{\sigma^2}{n}Q^{-1}=A\operatorname{var}\left(\hat{\beta}\right).$$

4.4.4 Asymptotic Behavior of the Standard Test Statistics

$H_0 : R\beta = q$

1) When R is a $1 \times K$ matrix, we have $\dfrac{R\left(\hat{\beta}-\beta\right)}{\sqrt{s^2 R\left(X'X\right)^{-1} R'}} \xrightarrow{d} N\left(0,1\right)$.

Proof:

$$\because \quad \frac{R\left(\hat{\beta}-\beta\right)}{\sqrt{s^2 R\left(X'X\right)^{-1} R'}} = \frac{R\sqrt{n}\left(\hat{\beta}-\beta\right)}{\sqrt{s^2 R\left(\frac{X'X}{n}\right)^{-1} R'}}$$

$$R\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \sigma^2 R Q^{-1} R'\right)$$

$$s^2 \xrightarrow{P} \sigma^2$$

$$\sqrt{s^2 R\left(\frac{X'X}{n}\right)^{-1} R'} \xrightarrow{P} \sqrt{\sigma^2 R Q^{-1} R'}$$

$$\therefore \quad \frac{R\left(\hat{\beta}-\beta\right)}{\sqrt{s^2 R\left(X'X\right)^{-1} R'}} \xrightarrow{d} = N\left(0,1\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(4.19)}$$

As a particular case, $t_k = \dfrac{\sqrt{n}\left(\hat{\beta}_k - \beta_k\right)}{\sqrt{s^2\left(\frac{X'X}{n}\right)^{-1}_{kk}}} \xrightarrow{d} N\left(0,1\right) \quad \left(k = 1, \cdots, K\right)$……….(4.20)

2) When R is a $J \times K$ matrix, we have

$$\frac{\left(R\hat{\beta}-q\right)'\left[R\left(X'X\right)^{-1} R'\right]^{-1}\left(R\hat{\beta}-q\right)\big/J}{e'e\big/\left(n-K\right)} \xrightarrow{d} \frac{1}{J}\chi^2\left(J\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(4.21)}$$

Proof:

$\because \quad R\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \sigma^2 R Q^{-1} R'\right)$,

Let $Z = R\sqrt{n}\left(\hat{\beta}-\beta\right)$, $P = \sigma^2 R Q^{-1} R'$,

$\therefore \quad Z \xrightarrow{d} N\left(0, P\right)$, P is a $J \times J$ symmetry matrix with $rank(P) = J$.

$$P^{-1/2} Z \xrightarrow{d} N\left(0, P^{-1/2} P \left(P^{-1/2}\right)'\right) = N\left(0, I\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(4.22)}$$

$Rank(I) = Rank(P) = J$,

$$\therefore \quad \left(P^{-1/2}Z\right)'\left(P^{-1/2}Z\right) \xrightarrow{d} \chi^2(J)$$

i.e. $Z'P^{-1}Z = \left[\sqrt{n}R\left(\hat{\beta}-\beta\right)\right]'\left[R\sigma^2Q^{-1}R'\right]^{-1}\left[\sqrt{n}R\left(\hat{\beta}-\beta\right)\right] \xrightarrow{d} \chi^2(J)$

$$\left[s^2R\left(\frac{X'X}{n}\right)^{-1}R'\right]^{-1} \xrightarrow{P} \left[R\sigma^2Q^{-1}R'\right]^{-1}$$

$$\therefore \quad \left[\sqrt{n}R\left(\hat{\beta}-\beta\right)\right]'\left[s^2R\left(\frac{X'X}{n}\right)^{-1}R'\right]^{-1}\left[\sqrt{n}R\left(\hat{\beta}-\beta\right)\right] \xrightarrow{d} \chi^2(J)$$

$$\frac{\left[\sqrt{n}R\left(\hat{\beta}-\beta\right)\right]'\left[R\left(\frac{X'X}{n}\right)^{-1}R'\right]^{-1}\left[\sqrt{n}R\left(\hat{\beta}-\beta\right)\right]\Big/J}{e'e\big/(n-K)}$$

$$= \frac{\left(R\hat{\beta}-q\right)'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}-q\right)\Big/J}{e'e\big/(n-K)} \xrightarrow{d} \frac{\chi^2(J)}{J} \quad \text{...(4.23)}$$

$$Wald = JF \xrightarrow{d} \chi^2(J) \quad \text{...(4.24)}$$

For the more general cases, the asymptotic behavior of test statistics under conditional heteroskedasticity and autocoerelation , we will discuss in Chap.11 and Chap.12.

4.5 Sequences and the Order of a Sequence

An important characteristic of a sequence is the rate at which it converges (or diverges). We will define the rate at which a sequence converges or diverges in terms of the order of the sequence.

Definition 4.14 Order n^{λ}

A sequence $\{b_n\}$ is at most of order n^{λ} in probability, denoted $b_n = O_p\left(n^{\lambda}\right)$,

If for every $\varepsilon > 0$ there exist a finite $\Delta_{\varepsilon} > 0$ and $N_{\varepsilon} \in \mathbb{N}$, such that $P\left\{\left|n^{-\lambda} b_n\right| > \Delta_{\varepsilon}\right\} < \varepsilon$ for all $n \geq N_{\varepsilon}$.

When $b_n = O_p(1)$, we say $\{b_n\}$ is bounded in probability.

Definition 4.15: Order Less than n^{λ}

A sequence $\{b_n\}$ is of order smaller than n^{λ} in probability, denoted $b_n = o_p\left(n^{\lambda}\right)$, if $p\lim\left(n^{-\lambda} b_n\right) = 0$.

When $b_n = o_p(1)$, we have $b_n \xrightarrow{P} 0$.

Theorem 4.13

1) If $a_n = O_p\left(n^{\lambda}\right)$ and $b_n = O_p\left(n^{\delta}\right)$, then $a_n b_n = O_p\left(n^{\lambda+\delta}\right)$ and $a_n + b_n = O_p\left(n^{k}\right)$, where $k = \max(\lambda, \delta)$.

2) If $a_n = o_p\left(n^{\lambda}\right)$ and $b_n = o_p\left(n^{\delta}\right)$, then $a_n b_n = o_p\left(n^{\lambda+\delta}\right)$ and $a_n + b_n = o_p\left(n^{k}\right)$, where $k = \max(\lambda, \delta)$.

3) If $a_n = O_p\left(n^{\lambda}\right)$, $b_n = o_p\left(n^{\delta}\right)$, then $a_n b_n = o_p\left(n^{\lambda+\delta}\right)$, $a_n + b_n = O_p\left(n^{k}\right)$, where $k = \max(\lambda, \delta)$.