Large Sample Statistical Theories

G.J. LI

1 Introduction

The properties and inferences results we discussed about OLS estimator in the last lecture do no depend on the sample size as long as certain assumptions are satisfied, which could be somewhat impractical in reality. In econometrics, many important results only hold asymptotically (in a large sample sense). It would be helpful to study some large sample statistical theories. The main reference for this note is White (2001).

Question: Suppose we have a dataset $\{y_1, y_2, \ldots, y_N\}$, which are identically and independently distributed with finite population mean. If we use the first observation as our estimator for the population mean, would it be unbiased? Why do we usually use sample average as the mean estimator instead?

2 Limit

Definition 2.1. Let $\{b_n\}$ be a sequence of real numbers. If there exists a real number b and if for every real $\delta > 0$, there exists an integer $N(\delta)$ such that for all $n \geq N(\delta)$, $|b_n - b| < \delta$, then b is **the limit of the sequence** $\{b_n\}$, denoted as $b_n \to b$.

Definition 2.2. The sequence $\{b_n\}$ is **at most of order** n^λ, denoted as $b_n = O(n^\lambda)$, if for some finite real number $\Delta > 0$, there exists a finite integer N such that for all $n \geq N$, $|n^{-\lambda} b_n| < \Delta$.

bounded

Definition 2.3. The sequence $\{b_n\}$ is **of order smaller than** n^λ, denoted as $b_n = o(n^\lambda)$, if for every real number $\delta > 0$, there exists a finite integer $N(\delta)$ such that for all $n \geq N(\delta)$, $|n^{-\lambda} b_n| < \delta$, i.e. $n^{-\lambda} b_n \to 0$.

In other words, if $\{n^{-\lambda} b_n\}$ is eventually bouned, then $b_n = O(n^\lambda)$. If $b_n = o(n^\lambda)$, then $b_n = O(n^\lambda)$, but not vice versa. When $b_n = O(n^0) = O(1)$, it is simply bounded and may or may not have a limit.

Definition 2.4. Given $g : \mathbb{R}^k \to \mathbb{R}^l$ $(k, l \in \mathbb{N})$ and $b \in \mathbb{R}^k$, the function g is **continuous at** b if for any sequence $\{b_n\}$ such that $b_n \to b$ implies $g(b_n) \to g(b)$. Further if $B \subset \mathbb{R}^k$, then g is **continuous on** B if it is continuous at every point on B.

For example, the matrix inverse function is continuous at every point that represents a nonsingular matrix so that if $\frac{\sum_{i=1}^n x_i x_i'}{n} \to M$, where M is a finite nonsingular matrix, then $\left(\frac{\sum_{i=1}^n x_i x_i'}{n}\right)^{-1} \to M^{-1}$. Apart from matrix inverse function, the matrix determinant function, denoted as $det(\cdot)$, is also a continuous function of the matrix elements.

3 Consistency

Definition 3.1. Let $\{b_n(\cdot)\}$ be a sequence of real-valued random variables. $b_n(\cdot)$ **converges almost surely** to b, written as $b_n(\cdot) \overset{a.s.}{\to} b$, if $P\left(\omega : b_n(\omega) \to b\right) = 1$ for $n \to \infty$.

If an event happens with probability one, we can say that the event will happen almost surely. Because the set of ω's for which $b_n(\omega) \to b$ has probability one, b_n is sometimes said to **converge to** b **with probability** 1, or b_n **converges almost everywhere** or **strongly consistent** for b.

Definition 3.2. Let $\{b_n(\cdot)\}$ be a sequence of real-valued random variables. If there exists a real number b such that for every $\epsilon > 0$, $P\left(\omega : |b_n(\omega) - b| < \epsilon\right) \to 1$, then $b_n(\cdot)$ **converges in probability to b**, written as $b_n(\cdot) \overset{p}{\to} b$ or $plim b_n = b$.

Convergence in probability is also referred to as **weak consistency**, where "weak" is often dropped since it is the most common stochastic convergence concept in econometrics. If $b_n(\cdot) \overset{a.s.}{\to} b$, then $b_n(\cdot) \overset{p}{\to} b$. if $b_n(\cdot) \overset{p}{\to} b$, then there exists a subsequence $\{b_{n_j}(\cdot)\}$ such that $b_{n_j}(\cdot) \overset{a.s.}{\to} b$.

Definition 3.3. The random sequence $\{b_n\}$ is **at most of order** n^λ **almost surely**, denoted $b_n = O_{a.s.}(n^\lambda)$, if there exists $\Delta > 0$ and $N < \infty$ such that $P\left[|n^{-\lambda}b_n| < \Delta \text{ for all } n \geq N\right] = 1$. The sequence $\{b_n\}$ is of **order smaller than** n^λ **almost surely**, denoted $b_n = o_{a.s.}(n^\lambda)$ if $n^{-\lambda}b_n \overset{a.s.}{\to} 0$.

A sufficient condition that $b_n = O_{a.s.}(n^\lambda)$ is that $n^{-\lambda}b_n - a_n \overset{a.s.}{\to} 0$ where $a_n = O(1)$.

Definition 3.4. The random sequence $\{b_n\}$ is **at most of order** n^λ **in probability**, denoted $b_n = O_p(n^\lambda)$, if for every $\epsilon > 0$ there exists a finite $\Delta_\epsilon > 0$ and $N_\epsilon \in \mathbb{N}$, such that $P\left(|n^{-\lambda}b_n| > \Delta_\epsilon\right) < \epsilon$ for all $n \geq N_\epsilon$. The sequence $\{b_n\}$ is of **order smaller than** n^λ **in probability**, denoted $b_n = o_p(n^\lambda)$ if $n^{-\lambda}b_n \overset{p}{\to} 0$.

$G(\sqrt{x}) = \sqrt{G(x)}$

Note that if $b_n = O_{a.s.}(n^{\lambda})$ or $b_n = O_p(n^{\lambda})$, then $n^{-\lambda} b_n$ can be either random or deterministic at the limit.

Theorem 3.1. ***Continuous Mapping Theorem:*** *$g : \mathbb{R}^k \to \mathbb{R}^l$ be continuous on a compact set $C \subset \mathbb{R}^k$. Suppose that $\{b_n(\cdot)\}$ is a sequence of random vectors such that $b_n(\cdot) - c_n \overset{a.s.}{\to} 0$ ($b_n(\cdot) - c_n \overset{p}{\to} 0$) and for all n sufficiently large, c_n is interior to C uniformly in n. Then $g(b_n(\cdot)) - g(c_n) \overset{a.s.}{\to} 0$ ($g(b_n(\cdot)) - g(c_n) \overset{p}{\to} 0$).*

Note that c_n does not have to converge to a fixed point. In the case of $c_n = c$, we can have $g(b_n(\cdot)) \overset{a.s.}{\to} g(c)$ ($g(b_n(\cdot)) \overset{p}{\to} g(c)$).

Theorem 3.2. *Suppose that $\{a_n(\cdot)\}$ and $\{b_n(\cdot)\}$ are sequences of random vectors such that $a_n(\cdot) - d_n \overset{a.s.}{\to} 0$ ($a_n(\cdot) - d_n \overset{p}{\to} 0$) and $b_n(\cdot) - c_n \overset{a.s.}{\to} 0$ ($b_n(\cdot) - c_n \overset{p}{\to} 0$) and for all n sufficiently large. Then the following are true:*

1. ***Summation Rule:*** *$a_n(\cdot) + b_n(\cdot) - (c_n + d_n) \overset{a.s.}{\to} 0$ ($a_n(\cdot) + b_n(\cdot) - (c_n + d_n) \overset{p}{\to} 0$);*

2. ***Product Rule:*** *$a_n(\cdot) \times b_n(\cdot) - c_n \times d_n \overset{a.s.}{\to} 0$ ($a_n(\cdot) \times b_n(\cdot) - c_n \times d_n \overset{p}{\to} 0$);*

Definition 3.5. A sequence of matrices $\{A_n\}$ is said to be **uniformly nonsingular** if for some $\delta > 0$ and all n sufficiently large $|det(A_n)| > \delta$. If $\{A_n\}$ is positive semidefinite matrices, then $\{A_n\}$ is **uniformly positive definite** if $\{A_n\}$ is uniformly nonsingular. If $\{A_n\}$ is a sequence of $l \times k$ matrices, then $\{A_n\}$ has **uniformly full column rank** if there exists a sequence of $k \times k$ submatrices $\{A_n^*\}$ which is uniformly nonsingular.

4 Law of Large Numbers

Theorem 4.1. ***Independent Identically Distributed Observations*** *Let $\{Z_i\}$ be a sequence of i.i.d. random variables. Then $\bar{Z}_N = \frac{1}{N}\sum_i^N Z_i \overset{a.s.}{\to} \mu$ if and only if $E(|Z_i|) < \infty$ and $E(Z_i) = \mu$.*

If $E(|Z_i|) \leq \infty$, then by Jensen's inequality (see Proposition 5.3 below), $E(Z_i) = \mu \leq |E(Z_i)| \leq E(|Z_i|) \leq \infty$.

Theorem 4.2. ***Uncorrelated Identically Distributed Observations*** *Let $\{Z_i\}$ be a sequence of random variables. Then $\bar{Z}_N = \frac{1}{N}\sum_i^N Z_i \overset{p}{\to} \mu$ if and only if $E(Z_i) = \mu$, $Var(Z_i) = \sigma^2 < \infty$ and $Cov(Z_i, Z_j) = 0$ for $i \neq j$.*

Note that the above theorem states convergence in probability but not convergence almost surely.

Theorem 4.3. ***Independent Heterogeneously Distributed Observations*** *Let $\{Z_i\}$ be a sequence of independent random variables with finite means $E(Z_i) = \mu_i$. If $E(|Z_i|^{1+\delta}) < \infty$ for some $\delta > 0$ and all i, then $\bar{Z}_N - E(\bar{Z}_N) = \bar{Z}_N - \bar{\mu}_N \overset{a.s.}{\to} 0$.*

Strong

As we can see from above, if we relax the identical distribution assumption to uncorrelation assumption or heterogeneous assumption, we have to strengthen the moment restriction.

5 Consistency of OLS Estimator

For cross sectional data, we often assume the observations are independent. The following is a useful proposition.

Proposition 5.1. *Let $g : \mathbb{R}^k \to \mathbb{R}^l$ be a continuous function. If Z_t and Z_τ are independent, then $g(Z_t)$ and $g(Z_\tau)$ are independent. If Z_t and Z_τ are identically distributed, then $g(Z_t)$ and $g(Z_\tau)$ are also identically distributed.*

For the linear regression model discussed in the last lecture,

$$y = X\beta + \epsilon, \tag{1}$$

another way to represent the model is

$$y_i = x_i'\beta + \epsilon_i, \qquad i = 1, 2, \ldots, N, \tag{2}$$

where $x_i = (1, x_{i1}, x_{i2}, \ldots, x_{iK})' : (K+1) \times 1$. In other words, we have stacked up all the explanatory variables (including the intercept) of the ith observation. We have derived the OLS estimator as $\hat\beta_{OLS} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\epsilon$. Another way to write the estimator is

$$\begin{aligned} \hat\beta_{OLS} &= (\sum_{i=1}^N x_i x_i')^{-1} \sum_{i=1}^N x_i y_i \\ &= \beta + \left(\frac{\sum_{i=1}^N x_i x_i'}{N}\right)^{-1} \frac{\sum_{i=1}^N x_i \epsilon_i}{N} \quad = 0 \end{aligned} \tag{3}$$

where $\frac{X'X}{N} = \frac{\sum_{i=1}^N x_i x_i'}{N}$ and $\frac{X'\epsilon}{N} = \frac{\sum_{i=1}^N x_i \epsilon_i}{N}$. For the second equality of the above equation, the second term on the right hand side is a function of two sample averages.

Before we show that the OLS estimator is consistent, we first introduce the following three inequalities.

Proposition 5.2. ***Hölder's Inequality**: If $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$ and if $E(|y|^p) < \infty$ and $E(|z|^q) < \infty$, then $E(|yz|) \leq [E(|y|^p)]^{\frac{1}{p}} [E(|z|^q)]^{\frac{1}{q}}$.*

If $p = q = 2$, we have **Cauchy-Schwarz inequality**[1],

$$E(|yz|) \leq [E(|y|^2)]^{\frac{1}{2}} [E(|z|^2)]^{\frac{1}{2}} \tag{4}$$

Note that if $E(y) = E(z) = 0$, we have $Cov(y, z) = E(yz) \leq E(|yz|) \leq \sqrt{Var(y)Var(z)}$.

Proposition 5.3. ***Jensen's Inequality**: Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function on an interval $B \subset \mathbb{R}$ and let Z be a random variable such that $P(Z \in B) = 1$. Then $g(E(Z)) \leq E(g(Z))$. If g is concave in B, then $g(E(Z)) \geq E(g(Z))$.*

Proposition 5.4. ***Minkowski's Inequality**: Let $q \geq 1$. If $E(|y|^q) < \infty$ and $E(|z|^q) < \infty$, then $[E(|y + z|^q)]^{\frac{1}{q}} \leq [E(|y|^q)]^{\frac{1}{q}} + [E(|z|^q)]^{\frac{1}{q}}$.*

Theorem 5.5. *For the linear regression model in (2), suppose $\{(x_i, \epsilon_i)\}$ is an i.i.d. sequence with*

1. *$E(x_i \epsilon_i) = 0$, $i = 1, 2, \ldots, N$;*

2. *$E(|x_{ik} \epsilon_i|) < \infty$, for $k = 1, 2, \ldots, K + 1$ and $i = 1, 2, \ldots, N$;*

3. *$E(|x_{ik}|^2) < \infty$, $k = 1, \ldots, K + 1$;*

4. *$E(x_i x_i') = M$ is finite and positive definite.*

[1] More generally speaking, if we define some inner product, denoted $< x, y >$, Cauchy-Schwarz inequality states that $< x, y >^2 \leq < x, x >< y, y >$. In this case, the inner product is $E(|xy|)$.

Then $\hat{\beta}_{OLS}$ exist for all N sufficiently large a.s., and $\hat{\beta}_{OLS} \overset{a.s.}{\to} \beta$.

Proof. Note that the k,lth element in $x_i x_i'$ is $x_{ik}x_{il}$. Using Cauchy-Schwarz inequality, we can have

$$E(|x_{ik}x_{il}|) \leq \sqrt{E(|x_{ik}|^2)E(|x_{il}|^2)} \leq \infty.$$

Hence $\frac{\sum_{i=1}^N x_i x_i'}{N} \overset{a.s.}{\to} M$ from Theorem 4.1. Moreover, since matrix inverse function is continuous, we have $\left(\frac{\sum_{i=1}^N x_i x_i'}{N}\right)^{-1} \overset{a.s.}{\to} M^{-1}$, which is finite and positive definite. Similarly, we can have $\frac{\sum_{i=1}^N x_i \epsilon_i}{N} \overset{a.s.}{\to} 0$. Therefore, $\hat{\beta}_{OLS} \overset{a.s.}{\to} \beta + M^{-1}0 = \beta$. □

Theorem 5.6. *For the linear regression model in (2), suppose $\{(x_i, \epsilon_i)\}$ is a sequence of uncorrelated random vectors of the same distribution with*

1. *$E(x_i\epsilon_i) = 0$, for $i = 1, 2, \ldots, N$;*
2. *$E(\epsilon_i\epsilon_j x_i x_j') = 0$ for $i \neq j$;*
3. *$E(x_i x_j') = E(x_i)E(x_j')$ for $i \neq j$;*
4. *$E(\epsilon_i^2 x_{ik}^2) < \infty$, $E(x_{ik}^4) < \infty$ for $k = 1, 2, \ldots, K$;*
5. *$E(x_i\, x_i') = M$ is positive definite.*

Then $\hat{\beta}_{OLS} \overset{p}{\to} \beta$.

Proof. By Theorem 4.2, we can have $\frac{\sum_{i=1}^N x_i \epsilon_i}{N} \overset{p}{\to} 0$. The k, lth element in $x_i x_i'$ is $x_{ik}x_{il}$. Using Cauchy-Schwarz inequality, we can have $E(|x_{ik}x_{il}|^2) \leq \sqrt{E(|x_{ik}|^4)E(|x_{il}|^4)} \leq \infty$ and hence $\frac{\sum_{i=1}^N x_i x_i'}{N} \overset{p}{\to} M$. By continuous mapping theorem, $\hat{\beta}_{OLS} \overset{p}{\to} \beta + M^{-1}0 = \beta$. □

Theorem 5.7. *For the linear regression model in (2), suppose $\{(x_i, \epsilon_i)\}$ is an independent sequence with*

1. *$E(x_i\epsilon_i) = 0$, $i = 1, 2, \ldots, N$;*
2. *$E(|x_{ik}\epsilon_i|^{1+\delta}) < \infty$, for some $\delta > 0$, $k = 1, \ldots, K+1$ and $i = 1, 2, \ldots, N$;*
3. *$E(|x_{ik}^2|^{1+\delta}) < \Delta < \infty$, for some $\delta > 0$, $k = 1, \ldots, K+1$ and $i = 1, 2, \ldots, N$;*
4. *$E(\frac{1}{N}\sum_{i=1}^{N} x_i x_i') = M_N$ is uniformly positive definite.*

Then $\hat{\beta}_{OLS}$ exist for all N sufficiently large a.s., and $\hat{\beta}_{OLS} \overset{a.s.}{\to} \beta$.

Proof. Note that the k,lth element in $x_i x_i'$ is $x_{ik}x_{il}$. Using Cauchy-Schwarz inequality, we can have

$$E(|x_{ik}x_{il}|^{1+\delta}) \leq \sqrt{E(|x_{ik}^2|^{1+\delta})E(|x_{il}^2|^{1+\delta})} < \Delta < \infty.$$

Note that by Jensen's inequality, $|E(x_{ik}x_{il})| \leq \left(E(|x_{ik}x_{il}|^{1+\delta})\right)^{\frac{1}{1+\delta}} < \Delta^{\frac{1}{1+\delta}}$. Hence the k,lth element in M_N should be $\frac{1}{N}\sum_{i=1}^{N} E(x_{ik}x_{il}) \leq \frac{1}{N}\sum_{i=1}^{N} |E(x_{ik}x_{il})| < \Delta^{\frac{1}{1+\delta}} = O(1)$. By Theorem 4.3, $\frac{1}{N}\sum_{i=1}^{N} x_i x_i' - M_N \overset{a.s.}{\to} 0$. Similarly, we have $\frac{\sum_{i=1}^{N} x_i\epsilon_i}{N} \overset{a.s.}{\to} 0$. By continuous mapping theorem, $\hat{\beta}_{OLS} - (\beta + M_N^{-1}0) \overset{a.s.}{\to} 0$ or $\hat{\beta}_{OLS} \overset{a.s.}{\to} \beta$. □

If $\hat{\beta}_{OLS}$ is consistent for β, $\hat{\epsilon} = y - X\hat{\beta}_{OLS}$ will be consistent for ϵ and $\widehat{\sigma^2} = \frac{SSR}{N-k-1}$ will be consistent for σ^2.

6 Convergence in Distribution

Definition 6.1. Let $\{b_n\}$ be a sequence of random finite-dimensional vectors with joint distribution function $\{F_n\}$. If $F_n(z) \to F(z)$ as $n \to \infty$ for every continuity point z (points where F is continuous, pointwise convergence), where F is the distribution function of a random variable Z, then b_n **converges in distribution** to the random variable Z, denoted $b_n \xrightarrow{d} Z$.

When $b_n \xrightarrow{d} Z$, we also say that b_n converges in law to Z ($b_n \xrightarrow{L} Z$), or that b_n is asymptotically distributed as F, denoted as $b_n \overset{A}{\sim} F$. F is called the limiting distribution of b_n.

Lemma 6.1. *The following are true*

standard distribution

1. *If* $b_n \xrightarrow{d} Z$*, then* $b_n = O_p(1)$*;*

apporch distribution

2. *If* $A_n \xrightarrow{p} 0$ *and* $b_n \xrightarrow{d} Z$*, then* $A_n b_n \xrightarrow{p} 0$*;*
3. *If* $a_n - b_n \xrightarrow{p} 0$ *and* $b_n \xrightarrow{d} Z$*, then* $a_n \xrightarrow{d} Z$*. We sat* a_n *is* ***asymptotically equivalent*** *to* b_n*.*

Definition 6.2. Let U be a positive (semi) definite symmetric matrix. Then there exists a positive (semi) definite symmetric **square root** $U^{\frac{1}{2}}$ such that $U^{\frac{1}{2}}U^{\frac{1}{2}} = U$.[2]

If $z \sim N(0, U)$, then $U^{-\frac{1}{2}}z \sim N(0, I)$, where U is positive definite and $U^{-\frac{1}{2}} = (U^{\frac{1}{2}})^{-1}$.

[2] $U^{\frac{1}{2}} = QD^{\frac{1}{2}}Q'$ where Q is an orthonormal matrix ($Q' = Q^{-1}$) and D is diagonal with the eigenvalues of U along the diagonal.

Proposition 6.2. ***Cramér-Wold device:*** *Let $\{b_n\}$ be a sequence of random $k \times 1$ vectors and suppose for* ***any*** *real $k \times 1$ vector λ such that $\lambda'\lambda = 1$, $\lambda' b_n \xrightarrow{d} \lambda' Z$, where Z is a $k \times 1$ vector with joint distribution F. Then the limiting distribution of b_n exists and equals F, i.e. $b_n \overset{A}{\sim} F$.*

When $Z \sim N(0, I_k)$, note that $\lambda' Z \sim N(0, 1)$.

7 Central Limit Theorem

Theorem 7.1. ***Lindeberg-Lévy, for i.i.d. observations:*** *Let $\{Z_i\}$ be a sequence of i.i.d. random scalars, with $\mu = E(Z_i)$ and $0 < \sigma^2 = Var(Z_i) < \infty$. Then*

need add a sqrt（n）for setoff 分母的N

$$\sqrt{n}\frac{\overline{Z_n} - \mu}{\sigma} \overset{A}{\sim} N(0, 1) \qquad (5)$$

$$Var\left(\frac{\sum X_i}{N}\right) = \frac{Var(X)}{N} \to 0$$

Theorem 7.2. ***Independent Heterogeneously Distributed Observations:*** *Let $\{Z_i\}$ be a sequence of independent random scalars with $\mu_i = E(Z_i)$ and $\sigma_i^2 = Var(Z_i)$, and $E(|Z_i|^{2+\delta}) < \Delta < \infty$ for some $\delta > 0$ and all i. If $\overline{\sigma_n^2} = \frac{\sum_{i=1}^n \sigma_i^2}{n} > 0$[3] for all n sufficiently large, then*

$$\sqrt{n}\frac{\overline{Z_n} - \bar{\mu}}{\sqrt{\overline{\sigma_n^2}}} \overset{A}{\sim} N(0, 1). \qquad (6)$$

8 Asymptotic Normality of OLS Estimator

Theorem 8.1. *For the linear regression model in (2), suppose*

1. *$\{(x_i, \epsilon_i)\}$ is an i.i.d. sequence;*

[3] That is $\frac{1}{\overline{\sigma_n^2}}$ has upper bound.

2. *$E(x_i\epsilon_i) = 0$, $i = 1, 2, \ldots, N$;*

3. *$E(|x_{ik}\epsilon_i|^2) < \infty$, for $k = 1, 2, \ldots, K+1$ and $i = 1, 2, \ldots, N$;*

4. *$Var(x_i\epsilon_i) = U$ is positive definite;*

5. *$E(|x_{ik}|^2) < \infty$, $k = 1, \ldots, K+1$;*

6. *$E(x_i x_i') = M$ is finite and positive definite.*

Then $\sqrt{N}(\hat{\beta}_{OLS} - \beta) \overset{A}{\sim} N(0, M^{-1}UM^{-1})$, where $M^{-1}UM^{-1}$ is the asymptotic variance of $\sqrt{N}(\hat{\beta}_{OLS} - \beta)$.

Proof. Note that from (3) $\sqrt{N}(\hat{\beta}_{OLS} - \beta) = \left(\frac{\sum_{i=1}^{N} x_i x_i'}{N}\right)^{-1} \frac{\sum_{i=1}^{N} x_i\epsilon_i}{\sqrt{N}}$. For any $(K+1)\times 1$ vector λ with $\lambda'\lambda = 1$, $\lambda' U^{-\frac{1}{2}} x_i\epsilon_i$ is i.i.d. with $E(\lambda' U^{-\frac{1}{2}} x_i\epsilon_i) = 0$ and $Var(\lambda' U^{-\frac{1}{2}} x_i\epsilon_i) = 1$. By Lindeberg-Lévy theorem, we have $\frac{\sum_{i=1}^{N} \lambda' U^{-\frac{1}{2}} x_i\epsilon_i}{\sqrt{N}} \overset{A}{\sim} N(0,1)$. By Proposition 6.2, we know $\frac{\sum_{i=1}^{N} U^{-\frac{1}{2}} x_i\epsilon_i}{\sqrt{N}} \overset{A}{\sim} N(0, I)$ and hence $M^{-1}\frac{\sum_{i=1}^{N} x_i\epsilon_i}{\sqrt{N}} \overset{A}{\sim} N(0, M^{-1}UM^{-1})$. Moreover, since $\left(\frac{\sum_{i=1}^{N} x_i x_i'}{N}\right)^{-1} - M^{-1} \overset{a.s.}{\to} 0$ (see the proof for Theorem 5.5), we have $\left(\frac{\sum_{i=1}^{N} x_i x_i'}{N}\right)^{-1} \frac{\sum_{i=1}^{N} x_i\epsilon_i}{\sqrt{N}} - M^{-1}\frac{\sum_{i=1}^{N} x_i\epsilon_i}{\sqrt{N}} \overset{a.s.}{\to} 0$ By Lemma 6.1, $\left(\frac{\sum_{i=1}^{N} x_i x_i'}{N}\right)^{-1} \frac{\sum_{i=1}^{N} x_i\epsilon_i}{\sqrt{N}}$ and $M^{-1}\frac{\sum_{i=1}^{N} x_i\epsilon_i}{\sqrt{N}}$ are asymptotically equivalent and $\sqrt{N}(\hat{\beta}_{OLS} - \beta) \overset{A}{\sim} N(0, M^{-1}UM^{-1})$. □

If $E(\epsilon_i^2|x_i) = \sigma^2$, we have $U = E(\epsilon_i^2 x_i x_i') = E(E(\epsilon_i^2 x_i x_i'|x_i)) = \sigma^2 M$ and $\sqrt{N}(\hat{\beta}_{OLS} - \beta) \overset{A}{\sim} N(0, \sigma^2 M^{-1})$. Under the assumption of $E(\epsilon\epsilon'|X) = \sigma^2 I_N$, OLS estimator is asymptotically efficient.

Theorem 8.2. *For the linear regression model in (2), suppose*

1. *$\{(x_i, \epsilon_i)\}$ is an independent sequence;*

2. *$E(x_i\epsilon_i) = 0$, $i = 1, 2, \ldots, N$;*

3. *$E(|x_{ik}\epsilon_i|^{2+\delta}) < \Delta < \infty$, for some $\delta > 0$, $k = 1, \ldots, K+1$ and $i = 1, 2, \ldots, N$;*

4. *$Var(\frac{\sum_{i=1}^N x_i\epsilon_i}{\sqrt{N}}) = U_N$ is uniformly positive definite;*

5. *$E(|x_{ik}^2|^{1+\delta}) < \infty$, for some $\delta > 0$, $k = 1, \ldots, K+1$ and $i = 1, 2, \ldots, N$;*

6. *$E(\frac{1}{N}\sum_{i=1}^N x_i x_i') = M_N$ is uniformly positive definite.*

Then $\sqrt{N}(\hat{\beta}_{OLS} - \beta) \overset{A}{\sim} N(0, M_N^{-1}U_N M_N^{-1})$.

Proof. First note that $\sqrt{N}(\hat{\beta}_{OLS} - \beta) = \left(\frac{\sum_{i=1}^N x_i x_i'}{N}\right)^{-1} \frac{\sum_{i=1}^N x_i\epsilon_i}{\sqrt{N}}$. For any $(K+1)\times 1$ vector λ with $\lambda'\lambda = 1$, $\lambda' U_N^{-\frac{1}{2}} x_i\epsilon_i$ is independent with $E(\lambda' U_N^{-\frac{1}{2}} x_i\epsilon_i) = 0$ and $Var(\frac{\sum_{i=1}^N \lambda' U_N^{-\frac{1}{2}} x_i\epsilon_i}{\sqrt{N}}) = 1 > 0$. Since U_N is uniformly positive definite, the elements in $U_N^{-\frac{1}{2}}$ should be bouneded. Denote $q = (q_1, q_2, \ldots, q_{K+1})' = U_N^{-\frac{1}{2}}\lambda$ and q should be finite. We can have $\lambda' U_N^{-\frac{1}{2}} x_i\epsilon_i = \sum_{k=1}^{K+1} q_k x_{ik}\epsilon_i$. By Minkowski's inequality, we obtain:

$$\begin{aligned} E\left(|\lambda' U_N^{-\frac{1}{2}} x_i\epsilon_i|^{2+\delta}\right) &= E\left(|\sum_{k=1}^{K+1} q_k x_{ik}\epsilon_i|^{2+\delta}\right) \\ &\leq \left[\sum_{k=1}^{K+1}\left(E(|q_k x_{ik}\epsilon_i|^{2+\delta})\right)^{\frac{1}{2+\delta}}\right]^{2+\delta} \\ &= \left[\sum_{k=1}^{K+1}|q_k|\left(E(|x_{ik}\epsilon_i|^{2+\delta})\right)^{\frac{1}{2+\delta}}\right]^{2+\delta} \\ &< \left(\sum_{k=1}^{K+1}|q_k|\Delta^{\frac{1}{2+\delta}}\right)^{2+\delta} < \infty. \end{aligned} \tag{7}$$

Hence by Theorem 7.2, it follows $\frac{\sum_{i=1}^N \lambda' U_N^{-\frac{1}{2}} x_i\epsilon_i}{\sqrt{N}} \overset{A}{\sim} N(0, 1)$ and $\frac{\sum_{i=1}^N x_i\epsilon_i}{\sqrt{N}} \overset{A}{\sim} N(0, U_N)$. The remaining proof is similar to that in Theorem 8.1. □

Note that $Var(\frac{\sum_{i=1}^{N} x_i\epsilon_i}{\sqrt{N}}) = \frac{1}{N}\sum_{i=1}^{N} Var(\epsilon_i x_i) = \frac{1}{N}\sum_{i=1}^{N} E(\epsilon_i^2 x_i x_i').(\{(x_i,\epsilon_i)\}$ is an independent sequence.) Under the condition in Theroem 8.2, $\frac{1}{N}\sum_{i=1}^{N} \epsilon_i^2 x_i x_i' - \frac{1}{N}\sum_{i=1}^{N} E(\epsilon_i^2 x_i x_i') \xrightarrow{a.s} 0$.(Why?) law of large number Hence if we can estimate ϵ consistently, by e.g. $\hat{\epsilon} = y - X\hat{\beta}_{OLS}$, we can estimate U_N consistently by $\hat{U}_N = \frac{1}{N}\sum_{i=1}^{N} \underline{\widehat{\epsilon_i}^2 x_i x_i'}$. Under homoskedasticity ($E(\epsilon_i^2|x_i) = \sigma^2$), we can also estimate U_N by $\tilde{U}_N = \widehat{\sigma^2}\frac{1}{N}\sum_{i=1}^{N} x_i x_i' = \boxed{\frac{SSR}{N-K-1}\frac{X'X}{N}}$.

One important implication for $\hat{\beta}_{OLS}$ to have asymptotic normal distribution is that the t statistic[4] will be asymptotically valid even if ϵ_i does not follow a normal distribution.

9 Hypothesis Testing

Let us first consider testing linear restrictions: $R\beta = r$, where R is a $q \times (K+1)$ matrix with rank equal to $q \leq (K+1)$ and r is a $q \times 1$ vector. We know what R and r are.

Lemma 9.1. *Let $g : \mathbb{R}^k \to \mathbb{R}^l$ be continuous on $\mathbb{R}^k$. If $b_n \xrightarrow{d} Z$, a $k \times 1$ random vector, then $g(b_n) \xrightarrow{d} g(Z)$.*

Corollary 9.2. *If $b_n \overset{A}{\sim} N(0, U_n)$, then $b_n' U_n^{-1} b_n \overset{A}{\sim} \chi_k^2$.*

Consider the restricted model,

$$\min_{\beta}(y - X\beta)'(y - X\beta), \qquad s.t. \quad R\beta = r, \tag{8}$$

which is equivalent to finding the stationary point of the Lagrangian

[4] Note that for student t distribution, if its degrees of freedom is sufficiently large, it will be very similar to standard normal distribution.

$$\mathcal{L} = (y - X\beta)'(y - X\beta) + \lambda'(R\beta - r), \tag{9}$$

where λ is the Lagrangian multiplier, a parameter associated with the constraint. The first order conditions are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \beta} &= 2(X'X\beta - X'y) + R'\lambda = 0, && (10)\\
\frac{\partial \mathcal{L}}{\partial \lambda} &= R\beta - r = 0. && (11)
\end{aligned}$$

Premultiplying (10) by $R(X'X)^{-1}$ and substituting out $R\beta$ by r gives

$$\begin{aligned}
\tilde{\lambda} &= 2[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) && \text{如果beta是真值，lamada就接近0}\\
&= 2[R(X'X)^{-1}R']^{-1}R(\hat{\beta} - \tilde{\beta}), && (12)\\
\tilde{\beta} &= \hat{\beta} - \frac{1}{2}(X'X)^{-1}R'\tilde{\lambda}, && (13)\\
SSR_R - SSR_U &= (\hat{\beta} - \tilde{\beta})'R'[R(X'X)^{-1}R']^{-1}R(\hat{\beta} - \tilde{\beta}), && (14)\\
SSR_R &= (y - X\tilde{\beta})'(y - X\tilde{\beta}), && (15)\\
SSR_U &= (y - X\hat{\beta})'(y - X\hat{\beta}). && (16)
\end{aligned}$$

Note that $R\tilde{\beta} = r$.

Quesion: What is the asymptotic variance of $\tilde{\beta}$? How is compared to that of $\hat{\beta}$? *hint:* $\tilde{\beta} = A\hat{\beta} + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}r$, *where* $A = I - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R = (X'X)^{-1}[X'X - R'(R(X'X)^{-1}R')^{-1}R]$ *with rank equal to* $K + 1 - q$ *and* $RA = 0$.

Theorem 9.3. ***Wald Test Under Linear Restrictions*** *Let the condi-*

tions of Theorem 8.2 hold. Then under $H_0 : R\beta = r$,

1. $\sqrt{N}(R\hat{\beta} - r) \overset{A}{\sim} N\left(0, RM_N^{-1}U_NM_N^{-1}R'\right)$, *where* $\hat{\beta} = (X'X)^{-1}X'y$*;*

2. *The Wald statistic is defined as* $\frac{1}{N}(R\hat{\beta}-r)'(R(X'X)^{-1}\hat{U}_N(X'X)^{-1}R')^{-1}(R\hat{\beta}-r) \overset{A}{\sim} \chi^2_q$*, where* $\hat{U}_N$ *is a symmetrical positive definite matrix computed from the unconstrained regresson such that* $\hat{U}_N - U_N \xrightarrow{p} 0$.

Proof. From Theorem 8.2, we know $\sqrt{N}(\hat{\beta}_{OLS} - \beta) \overset{A}{\sim} N(0, M_N^{-1}U_NM_N^{-1})$ and hence $\sqrt{N}(R\hat{\beta}_{OLS} - R\beta) \overset{A}{\sim} N(0, RM_N^{-1}U_NM_N^{-1}R')$. Subsituting $R\beta = r$ yields the first result.

Under the conditions of Theorem 8.2, we have $\left(\frac{X'X}{N}\right)^{-1} - (M_N)^{-1} \xrightarrow{a.s.} 0$, therefore $N^2R(X'X)^{-1}\hat{U}_N(X'X)^{-1}R' - RM_N^{-1}U_NM_N^{-1}R' \xrightarrow{a.s.} 0$. Using Corollary 9.2, we obtain the second result. □

Under homoskedasticity, we can estimate U_N by $\widehat{\sigma^2}\frac{X'X}{N}$, where $\widehat{\sigma^2} = \frac{SSR_U}{N-K-1}$, and the Wald statistic can be simplified as $\frac{1}{\widehat{\sigma^2}}(R\hat{\beta}-r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r) = \frac{SSR_R - SSR_U}{\widehat{\sigma^2}}$. In other words, the relevant F statistic multiplied by q follows chi-squared distribution with degrees of freedom q asymptotically.

Theorem 9.4. ***Lagrange Multiplier Test Under Linear Restrictions:*** *Let the conditions of Theorem 8.2 hold. Then under $H_0 : R\beta = r$,*

1. $\frac{1}{\sqrt{N}}\tilde{\lambda} \overset{A}{\sim} N\left(0, \Lambda = 4(RM_N^{-1}R')^{-1}RM_N^{-1}U_NM_N^{-1}R'(RM_N^{-1}R')^{-1}\right)$*;*

2. *The Lagrange multiplier test statistic is defined as* $\frac{1}{N}\tilde{\lambda}'\tilde{\Lambda}^{-1}\tilde{\lambda} = \frac{1}{N}(R\hat{\beta}-r)'(R(X'X)^{-1}\tilde{U}_N(X'X)^{-1}R')^{-1}(R\hat{\beta}-r) \overset{A}{\sim} \chi^2_q$*, where* $\tilde{U}_N$ *is a symmetrical positive definite matrix computed from the constrained regresson such that* $\tilde{U}_N - U_N \xrightarrow{p} 0$.

Note that the Wald and the Lagrange multiplier statistics would be the same if $\hat{U}_N$ is replaced by $\tilde{U}_N$. Under H_0, the Wald and the Lagrange multiplier statistics are asymptotically equivalent.

Lemma 9.5. *For the partitioned regression* $y = X_1\beta_1 + X_2\beta_2 + \epsilon$, *under* $H_0: \underset{q\times 1}{\beta_2} = 0$, *i.e.* $R = [\underset{q\times(K+1-q)}{0}, I_q]$ *and* $r = 0$, *the following are true:*

1. $\tilde{\beta}' = (\tilde{\beta}_1', 0')$, *where* $\tilde{\beta}_1 = (X_1'X_1)^{-1}X_1'y$;

2. $\tilde{\lambda} = 2X_2'M_{X_1}y = 2X_2'M_{X_1}\epsilon = 2X_2'M_{X_1}\tilde{\epsilon}$;

Homo

3. *If* $\widehat{\sigma^2}\frac{X'X}{N} - U_N \xrightarrow{p} 0$, *the Lagrange multiplier test statistic can be calculated as* $N\frac{\tilde{\epsilon}'P_X\tilde{\epsilon}}{\tilde{\epsilon}'\tilde{\epsilon}} = NR_*^2$, *where* R_*^2 *is obtained by regressing the estimated residual from the restricted model on the regressors of the unrestricted model.*

Next let us consider testing nonlinear restrictions: $H_0: s(\beta) = 0$, where $s: \mathbb{R}^{K+1} \to \mathbb{R}^q$ is a continuously differentiable function of β with $q \leq (K+1)$. We need the following theorem to help construnt test statistics.

Theorem 9.6. ***Mean Value Theorem****: Let* $s: \mathbb{R}^k \to \mathbb{R}$ *be a continuously differentiable function on an open convex set* $\Theta \subset \mathbb{R}^k$, *Then for any points* θ *and* θ_0 *in* Θ, *there exists* $\bar{\theta}$ *on the segment connecting* θ *and* θ_0 *such that* $s(\theta) = s(\theta_0) + \frac{\partial s(\bar{\theta})}{\partial \theta'}(\theta - \theta_0)$.

Theorem 9.7. ***Wald Test Under Nonlinear Restrictions****: Let the conditions of Theorem 8.2 hold and the* $q \times (K+1)$ *gradient matrix,* $\frac{\partial s(\beta)}{\partial \beta'}$ *has rank* q. *Then under* $H_0: s(\beta) = 0$,

1. $\sqrt{N}s(\hat{\beta}) \overset{A}{\sim} N\left(0, \frac{\partial s(\beta)}{\partial \beta'}M_N^{-1}U_NM_N^{-1}\frac{\partial s(\beta)}{\partial \beta'}'\right)$;

2. *The Wald statistic is calculated as* $\frac{1}{N}s(\hat{\beta})'\left[\frac{\partial s(\hat{\beta})}{\partial \beta'}(X'X)^{-1}\hat{U}_N(X'X)^{-1}\frac{\partial s(\hat{\beta})}{\partial \beta'}'\right]^{-1}s(\hat{\beta}) \overset{A}{\sim} \chi^2_q$, *where* $\hat{U}_N$ *is a symmetrical positive definite matrix computed from the unconstrained regresson such that* $\hat{U}_N - U_N \overset{p}{\to} 0$.

Proof. Note that $s(\cdot)$ is a vector function. We can apply mean value theorem to each of its element around β to get $s_i(\hat{\beta}) = s_i(\beta) + \frac{\partial s_i(\bar{\beta}^{(i)})}{\partial \beta'}(\hat{\beta}-\beta)$ with $s_i(\beta) = 0$ under the restriction for $i = 1, 2, \ldots, q$. Since $\hat{\beta} \overset{a.s.}{\to} \beta$, we have $\bar{\beta}^{(i)} \overset{a.s.}{\to} \beta$. Hence $s_i(\hat{\beta})$ is asymptotically equivalent to $\frac{\partial s_i(\beta)}{\partial \beta'}(\hat{\beta}-\beta)$ and $s(\hat{\beta})$ is asymptotically equivalent to $\frac{\partial s(\beta)}{\partial \beta'}(\hat{\beta}-\beta)$. Given the result in Theorem 8.2, we can have $\sqrt{N}s(\hat{\beta}) \overset{A}{\sim} N\left(0, \frac{\partial s(\beta)}{\partial \beta'}M_N^{-1}U_N M_N^{-1}\frac{\partial s(\beta)}{\partial \beta'}'\right)$. The proof of the second result is similar to that in Thereom 9.3. □

Theorem 9.8. ***Lagrange Multiplier Test Under Nonlinear Restrictions****: Let the conditions of Theorem 8.2 hold. Then under* $H_0 : s(\beta) = 0$,

1. $\frac{1}{\sqrt{N}}\tilde{\lambda} \overset{A}{\sim} N\left(0, \Lambda = 4(\frac{\partial s(\beta)}{\partial \beta'}M_N^{-1}\frac{\partial s(\beta)}{\partial \beta'}')^{-1}\frac{\partial s(\beta)}{\partial \beta'}M_N^{-1}U_N M_N^{-1}\frac{\partial s(\beta)}{\partial \beta'}'(\frac{\partial s(\beta)}{\partial \beta'}M_N^{-1}\frac{\partial s(\beta)}{\partial \beta'}')^{-1}\right)$, *where* $\tilde{\lambda} = 2\left[\frac{\partial s(\bar{\beta})}{\partial \beta'}(X'X)^{-1}\frac{\partial s(\beta)}{\partial \beta'}'\right]^{-1}\frac{\partial s(\bar{\beta})}{\partial \beta'}(\hat{\beta}-\beta) = 2\left[\frac{\partial s(\bar{\beta})}{\partial \beta'}(X'X)^{-1}\frac{\partial s(\beta)}{\partial \beta'}'\right]^{-1}s(\hat{\beta})$;

2. *The Lagrange multiplier test statistic is defined as*

$$\frac{1}{N}s(\hat{\beta})'\left[\frac{\partial s(\tilde{\beta})}{\partial \beta'}(X'X)^{-1}\tilde{U}_N(X'X)^{-1}\frac{\partial s(\tilde{\beta})}{\partial \beta'}'\right]^{-1}s(\hat{\beta}) \overset{A}{\sim} \chi^2_q,$$

where $\tilde{U}_N$ *is a symmetrical positive definite matrix computed from the constrained regresson such that* $\tilde{U}_N - U_N \overset{p}{\to} 0$.

Proof. The Lagrangian is now

$$\mathcal{L} = (y - X\beta)'(y - X\beta) + \lambda' s(\beta). \tag{17}$$

The first order conditions are

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \beta} &= 2(X'X\beta - X'y) + \frac{\partial s(\beta)}{\partial \beta'}' \lambda = 0, && (18)\\
\frac{\partial \mathcal{L}}{\partial \lambda} &= s(\beta) = 0. && (19)
\end{aligned}
$$

Taking a mean value expansion of $s(\beta)$ around $\hat{\beta}$ gives

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \beta} &= 2X'X(\beta - \hat{\beta}) + \frac{\partial s(\beta)}{\partial \beta'}' \lambda = 0, && (20)\\
\frac{\partial \mathcal{L}}{\partial \lambda} &= s(\hat{\beta}) + \frac{\partial s(\bar{\beta})}{\partial \beta'}(\beta - \hat{\beta}) = 0. && (21)
\end{aligned}
$$

Premultiplying (20) by $\frac{\partial s(\bar{\beta})}{\partial \beta'}(X'X)^{-1}$ gives $\tilde{\lambda} = 2\left[\frac{\partial s(\bar{\beta})}{\partial \beta'}(X'X)^{-1}\frac{\partial s(\beta)}{\partial \beta'}'\right]^{-1} \frac{\partial s(\bar{\beta})}{\partial \beta'}(\hat{\beta} - \beta)$. Under H_0, $\tilde{\beta}$ is asymptotically equivalent to $\bar{\beta}$ and β. Using the result from Theorem 8.2, and replacing U_N and M_N by $\tilde{U}_N$ and $\frac{X'X}{N}$ yield the results. □

The Lagrange statistic differs from the Wald statistic only in that $\tilde{U}_N$ is used in place of $\hat{U}_N$ and $\tilde{\beta}$ replaces $\hat{\beta}$ and $\bar{\beta}$ in the derivative matrix. Also note that $\frac{\partial s(\bar{\beta})}{\partial \beta'}$ and $\frac{\partial s(\beta)}{\partial \beta'}$ under nonlinear restrictions are analogical to R under linear restrictions.

From (20) and (21), we can obtain the following solutions,

$$
\begin{aligned}
\tilde{\lambda} &= 2\left[\frac{\partial s(\bar{\beta})}{\partial \beta'}(X'X)^{-1}\frac{\partial s(\beta)}{\partial \beta'}'\right]^{-1} s(\hat{\beta}), && (22)\\
\beta &= \hat{\beta} - (X'X)^{-1}\frac{\partial s(\beta)}{\partial \beta'}'\left[\frac{\partial s(\bar{\beta})}{\partial \beta'}(X'X)^{-1}\frac{\partial s(\beta)}{\partial \beta'}'\right]^{-1} s(\hat{\beta}), && (23)
\end{aligned}
$$

The solution for β in (23) does not have closed form. Since we do not know $\bar{\beta}$ (the point between the true value of β and $\hat{\beta}$), it further complicates the problem. A computationally practical and asymptotically equivalent result can be obtained by replacing β and $\bar{\beta}$ by $\hat{\beta}$ on the right hand side of (23), which yields,

$$\tilde{\beta} = \hat{\beta} - (X'X)^{-1}\frac{\partial s(\hat{\beta})}{\partial \beta'}' \left[\frac{\partial s(\hat{\beta})}{\partial \beta'}(X'X)^{-1}\frac{\partial s(\hat{\beta})}{\partial \beta'}'\right]^{-1} s(\hat{\beta}). \qquad (24)$$

After we obtain $\tilde{\beta}$, we can plug it back in place of $\hat{\beta}$ on the right hand side of (24) to get another estimator. This process could continue until the change in the resulting estimator was sufficiently small. Nevertheless, the iteration process has no effect on the asymptotic variance matrix of the resulting estimator. Since $\hat{\beta}$ converges to β asymptotically, if $s(\beta) = 0$ is true, $\tilde{\beta}$ will also converge to β. Now if we get another estimator $\tilde{\tilde{\beta}}$ based on $\tilde{\beta}$ by one interation, we have

$$\sqrt{N}(\tilde{\beta} - \tilde{\tilde{\beta}}) = (\frac{X'X}{N})^{-1}\frac{\partial s(\tilde{\beta})}{\partial \beta'}' \left[\frac{\partial s(\tilde{\beta})}{\partial \beta'}(\frac{X'X}{N})^{-1}\frac{\partial s(\tilde{\beta})}{\partial \beta'}'\right]^{-1} \sqrt{N}s(\tilde{\beta})$$

where

$$\sqrt{N}s(\tilde{\beta}) = \sqrt{N}s(\hat{\beta}) + \sqrt{N}\frac{\partial s(\bar{\beta})}{\partial \beta'}(\tilde{\beta} - \hat{\beta})$$

with $\bar{\beta}$ being a point between $\tilde{\beta}$ and $\hat{\beta}$. Substituting (24) into the right hand side of the above equation produces

$$\sqrt{N}s(\tilde{\beta}) = \left(I - \frac{\partial s(\bar{\beta})}{\partial \beta'}(X'X)^{-1}\frac{\partial s(\hat{\beta})}{\partial \beta'}' \left[\frac{\partial s(\hat{\beta})}{\partial \beta'}(X'X)^{-1}\frac{\partial s(\hat{\beta})}{\partial \beta'}'\right]^{-1}\right)\sqrt{N}s(\hat{\beta})$$

From Theorem 9.7, we know $\sqrt{N}s(\hat{\beta}) = O_p(1)$. Since $\bar{\beta} - \hat{\beta} \overset{a.s.}{\rightarrow} 0$ and hence $\frac{\partial s(\bar{\beta})}{\partial \beta'} - \frac{\partial s(\hat{\beta})}{\partial \beta'} = o_p(1)$, we have $I - \frac{\partial s(\bar{\beta})}{\partial \beta'}(X'X)^{-1}\frac{\partial s(\hat{\beta})}{\partial \beta'}' \left[\frac{\partial s(\hat{\beta})}{\partial \beta'}(X'X)^{-1}\frac{\partial s(\hat{\beta})}{\partial \beta'}'\right]^{-1} = \sqrt{N}s(\tilde{\beta}) = o_p(1)$ and $\sqrt{N}(\tilde{\beta} - \tilde{\tilde{\beta}}) \overset{p}{\rightarrow} 0$. Therefore $\sqrt{N}(\tilde{\beta} - \beta)$ and $\sqrt{N}(\tilde{\tilde{\beta}} - \beta)$ are asymptotically equivalent and have the same asymptotic distribution.

References

[1] W.H. Greene, *Econometric Analysis*, 7th ed., Pearson Education, Appendix D, 2011.

[2] H. White, *Asymptotic Theory for Econometricians*, Prentice Hall, 2001.

[3] J.M. Wooldridge, *Introductory Econometrics (a modern approach)*, 4th ed., South Western College, Appendix C3, 2008.