SUMMARY

READING LIST

The course material is predominantly in the course notes that are included here, but the following books may be helpful.

References

[1] T. Amemiya, "Advanced Econometrics," *Blackwell.*
[2] E.L. Lehmann, "Elements of Large Sample Theory", *Springer.*
[3] C.R. Rao, "Linear Statistical Inference and its Applications," *Wiley.*
[4] R. Serfling, "Approximation Theorems for Mathematical Statistics," *Wiley.*
[5] H. White, "Asymptotic Theory for Econometricians," *Academic.*

1. **ASYMPTOTIC THEORY**

When we analyze the properties of estimators (or statistics), it often difficult to say something about their statistical properties for a given sample size "n".

However it might be the case that as $n \to \infty$, we could say something about their distribution or how it behaves.

So, one idea could be to behave as if $n = \infty$, that is we "extrapolate" the statistical properties when $n = \infty$ to the situation of finite sample sizes n.

This is the topic of this lectures.

2. SOME TYPES OF CONVERGENCE

In this section, we will see some types or modes of convergence as well as some basic results regarding functions of sequences of random variables.

Let $\{X_n, n = 1, 2, ...\}$ be a sequence of random variables and X another random variable defined on a common probability space, (X can be a constant).

Example 1. *For instance, in a linear regression model, $X_n = \widehat{\beta}_n$ i.e. the LSE of β based on n observations and $X = \beta_0$ (the true value).*

Definition 1. *(Convergence in Probability) We say that $X_n \xrightarrow{P} X$ ($p\lim X_n = X$) if $\forall\delta > 0$,*

$$\lim_{n\to\infty} \Pr\{|X_n - X| > \delta\} = 0;$$
$$\text{or } \lim_{n\to\infty} \Pr\{|X_n - X| \leq \delta\} = 1.$$

Alternative definition is: $\forall\delta > 0$ and $\varepsilon > 0$, $\exists n_0(\varepsilon, \delta)$ such that $\forall n \geq n_0$

$$\Pr\{|X_n - X| > \delta\} < \varepsilon.$$

Motivation:

Definition 2. *(Weak consistency) Suppose that we have an unknown parameter θ, and based on a sample of n observations we estimate it by $\widehat{\theta}_n$. Then, $\widehat{\theta}_n$ is (weakly) consistent if $\widehat{\theta}_n - \theta \xrightarrow{P} 0$.*

For instance, suppose that we wish to estimate $E\varepsilon = \theta$. Given a sample $\{\varepsilon_i; i = 1, ..., n\}$, we estimate μ by the sample mean. Then $\theta =: X$ and $\widehat{\theta}_n =: X_n = n^{-1} \sum_{i=1}^{n} \varepsilon_i$. Then if $\widehat{\theta}_n - \theta \xrightarrow{P} 0$, we say that the estimator is consistent.

Intuitively, consistency means that if n is large enough the probability that $\widehat{\theta}_n$ and θ differ at least by a quantity, say δ, is very small.

The concept of consistency is vital and a minimal requirement of any estimator.

Definition 3. *(Almost Sure Convergence) We say that* $X_n \xrightarrow{a.s.} X$ *if* $\forall\delta > 0$,

$$\lim_{n\to\infty} \Pr\{|X_m - X| > \delta \text{ for some } m \geq n\} = 0,$$
$$\text{or, } \lim_{n\to\infty} \Pr\{|X_m - X| \leq \delta \text{ for all } m \geq n\} = 1.$$

Alternatively, we can write this condition as saying that

$$\Pr\left\{\lim_{n\to\infty} |X_n - X| = 0\right\} = 1.$$

For instance if $\widehat{\theta}_n \xrightarrow{a.s.} \theta$, then we say that the estimator $\widehat{\theta}_n$ converges almost surely to θ or that $\widehat{\theta}_n$ is a strongly consistent estimator of θ.

Example 2. *Let* $\{X_n\}$ *be a sequence of r.v. defined as*

$$X_n = \begin{cases} X & 1 - \frac{1}{n} \\ 1 + X & \frac{1}{n}. \end{cases}$$

Then, $X_n \xrightarrow{p} X$*. Indeed, by definition,*

$$\Pr\{|X_n - X| > \delta\} = \Pr\{X_n = 1 + X\} = 1/n \to 0.$$

Example 3. *Let* $\{X_n\}$ *be a sequence of r.v. defined as*

$$X_n = \begin{cases} 0 & 1 - \frac{1}{n^2} \\ n & \frac{1}{n^2}. \end{cases}$$

Then, $X_n \xrightarrow{a.s.} 0$*. Indeed, all we need to show is that*

$$\Pr\{|X_m| \geq \delta, \text{ for some } m \geq n\} \to 0.$$

But, the last displayed expression is equal to

$$\begin{aligned} \Pr\left\{\bigcup_{m=n}^{\infty} \{|X_m| \geq \delta\}\right\} &\leq \sum_{m=n}^{\infty} \Pr\{|X_m| \geq \delta\} \\ &= \sum_{m=n}^{\infty} \frac{1}{m^2} < \varepsilon \end{aligned}$$

because $\sum_{m=1}^{\infty} m^{-2} < \infty$. ■

Up to now, we have seen two different types of statistical convergence, that is convergence in *Probability* and *almost surely.*

In the next type of convergence, we will assume that the sequence of r.v.'s are indexed by a set of parameters, say $\theta \in \Theta$.

Definition 4. *We will say that the sequence*

$$\{X_n(\theta); \quad n = 0, 1, 2, ...\}$$

*converges **uniformly** to a r.v. $X(\theta)$ in probability if*

$$\lim_{n\to\infty} \Pr\left\{\sup_{\theta\in\Theta} |X_n(\theta) - X(\theta)| < \varepsilon\right\} = 1.$$

Definition 5 (Complete convergence)**.** *We say that X_n converges **completely** to X ($X_n \xrightarrow{c} X$) if $\forall\delta > 0$,*

$$\sum_{n=0}^{\infty} \Pr\{|X_n - X| > \delta\} < \infty.$$

Proposition 1. *(Relation among the modes of convergence)*
(*a*) $X_n \xrightarrow{a.s.} X \Longrightarrow X_n \xrightarrow{P} X$.
(*b*) $X_n \xrightarrow{c} X \Longrightarrow X_n \xrightarrow{a.s.} X$.

Proof. We begin with part (a). By definition, we have convergence in probability $\Leftrightarrow \forall\varepsilon > 0, \forall\delta > 0, \exists n_0(\varepsilon, \delta)$ such that

$$\forall m \geq n_0(\varepsilon, \delta),\ \Pr\{\omega : |X_m(\omega) - X(\omega)| \leq \delta\} > 1 - \varepsilon,$$

whereas we have convergence almost surely $\Leftrightarrow \forall\varepsilon > 0, \exists n_0(\varepsilon)$ such that $\forall\delta > 0$,

$$1 - \varepsilon < \Pr\left\{\bigcap_{m=n_0}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| \leq \delta\}\right\}.$$

Now, using that if

$$A \subseteq B \Rightarrow \Pr(A) \leq \Pr(B), \tag{2.1}$$

we have that

$$\Pr\left\{\bigcap_{m=n_0}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| \leq \delta\}\right\} \leq \Pr\{\omega : |X_m(\omega) - X(\omega)| \leq \delta\}.$$

From here, the conclusion is now obvious because *almost sure* convergence yields that the right side is bounded from below by $1 - \varepsilon$.

Next we show part (b). Recall that $X_n \xrightarrow{a.s.} X \Longleftrightarrow \forall\delta > 0$, $\forall\varepsilon > 0$, $\exists n(\varepsilon) > 0$ such that:

$$\Pr\{|X_m - X| > \delta \text{ for some } m \geq n\} < \varepsilon.$$

But,

$$\begin{aligned}\Pr\{|X_m - X| > \delta \text{ for some } m \geq n\} &= \Pr\left\{\bigcup_{m=n+1}^{\infty} |X_m - X| > \delta\right\} \\ &\leq \sum_{m=n+1}^{\infty} \Pr\{|X_m - X| > \delta\}\end{aligned}$$

So, to complete the proof it suffices to show that the right side of the last displayed inequality is bounded by ε.

But because $X_n \xrightarrow{c} X$, we have that $\forall\delta > 0$,

$$\sum_{m=1}^{\infty} \Pr\{|X_m - X| > \delta\} < \infty,$$

which implies that $\forall\varepsilon > 0$, $\exists n(\varepsilon) > 0$ such that

$$\sum_{m=n+1}^{\infty} \Pr\{|X_m - X| > \delta\} < \varepsilon,$$

which concludes the proof. □

Amemiya's book provides an example of a sequence converging in probability but not almost surely.

Normally we are concerned with a vector or matrix in econometrics.

Definition 6. *Let X_n, X be $k \times 1$ vectors with j-th elements X_{nj} and X_j respectively. Then,*

$$X_n \xrightarrow{P} X \Longleftrightarrow X_{nj} \xrightarrow{P} X_j \ \forall j = 1, 2, ..., k.$$

(The same is true for $X_n \xrightarrow{c} X$ and $X_n \xrightarrow{a.s.} X$.)

We now give a couple of well known results.

Let A and B two disjoint sets, then

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \tag{2.2}$$

and, denoting by $\overline{C}$ the complementary set of C,

$$\begin{aligned} \Pr(A) &= \Pr\left(A \cap \left(C \cup \overline{C}\right)\right) = \Pr\left((A \cap C) \cup \left(A \cap \overline{C}\right)\right) \\ &= \Pr(A \cap C) + \Pr\left(A \cap \overline{C}\right) \qquad (2.3) \end{aligned}$$

using (2.2).

Theorem 1. *(Slutzky's Theorem) Let X_n, X be $k \times 1$ vectors. Assume that $X_n \xrightarrow{P} X$, and let $g(\circ)$ be a continuous function in the* ***domain*** *of definition of X. Then,*

$$g(X_n) \xrightarrow{P} g(X).$$

Proof. For any arbitrary small $\varepsilon > 0$, we can choose a compact set S such that

$$\Pr\{X \notin S\} \le \frac{\varepsilon}{2}.$$

Because $g(\circ)$ is continuous, it implies that $g(\circ)$ is uniformly continuous in S. Hence, for any given $\eta > 0$, we have that $\exists \delta(\eta) > 0$ (independent of $x \in S$) such that

$$\|x - y\| \le \delta \Longrightarrow \|g(x) - g(y)\| \le \eta.$$

Because $X_n \xrightarrow{P} X$, we can choose $n_0(\varepsilon, \delta) > 0$ such that $\forall n \ge n_0$,

$$\Pr\{\|X_n - X\| > \delta\} \le \frac{\varepsilon}{2}.$$

Alternatively

$$\begin{aligned}
1 - \frac{\varepsilon}{2} &\le \Pr\{\|X_n - X\| \le \delta\} \\
&= \Pr\{\|X_n - X\| \le \delta \ \text{ and } \ X \in S\} \\
&\quad + \Pr\{\|X_n - X\| \le \delta \ \text{ and } X \notin S\} \\
\text{by (2.1)} \quad &\le \Pr\{\|X_n - X\| \le \delta \ \text{ and } X \in S\} + \frac{\varepsilon}{2} \\
\text{by continuity and (2.1)} &\le \Pr\{\|g(X_n) - g(X)\| \le \eta \ \text{ and } X \in S\} + \frac{\varepsilon}{2} \\
\text{by (2.1)} \quad &\le \Pr\{\|g(X_n) - g(X)\| \le \eta\} + \frac{\varepsilon}{2},
\end{aligned}$$

where in the first equality we used (2.3).

Thus $1 - \varepsilon \le \Pr\{\|g(X_n) - g(X)\| \le \eta\}$ and the result follows since η and ε are arbitrary positive constants.

$\square$

Corollary 1. *Let $g(\circ)$ be a continuous function at c. Then,*

$$X_n \xrightarrow{P} c \Longrightarrow g(X_n) \xrightarrow{P} g(c).$$

Corollary 2.

$$X_n \xrightarrow{P} X \Longrightarrow \|X_n - X\| \xrightarrow{P} 0.$$

Example 4. *z_i ($k \times 1$ vector), $i = 1, ..., n$; $Z = (z_1, ..., z_n)'$ ($n \times k$ matrix). Then "second moment matrix" is*

$$\widehat{M}_n = \frac{1}{n} \sum_{i=1}^{n} z_i z_i' = \frac{1}{n} Z'Z.$$

Suppose that $\widehat{M}_n \xrightarrow{P} M > 0$ constant. Then

$$\widehat{M}_n^{-1} \xrightarrow{P} M^{-1} > 0$$

by Corollary 1, since if a matrix A is p.d., then $g(A) = A^{-1}$ is a continuous function.

Next suppose that

$$y_i = \beta' z_i + u_i$$

where u_i are iid $(0, \sigma^2)$ and independent of z_i. Now the LSE of β is

$$\widehat{\beta}_n = (Z'Z)^{-1} Z'Y$$

where $Y = (y_1, ..., y_n)'$. We will show in a later section that

$$n^{1/2} \left(\widehat{\beta}_n - \beta \right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 M^{-1}\right).$$

So, if we had an estimator of σ^2, say $\widehat{\sigma}_n^2$, such that $\widehat{\sigma}_n^2 \xrightarrow{P} \sigma^2$, then together with the fact that $\widehat{M}_n \xrightarrow{P} M$ we would conclude by Corollary 1 that

$$\widehat{\sigma}_n^2 \widehat{M}_n^{-1} \xrightarrow{P} \sigma^2 M^{-1}.$$

That is, we have a consistent estimator for the asymptotic variance of $\widehat{\beta}_n$.

3. ASYMPTOTIC UNBIASEDNESS

Definition 7. (1) *We say that $\widehat{\theta}_n$ is unbiased for θ if*

$$E_\theta\left(\widehat{\theta}_n\right) = \theta, \quad \forall\theta.$$

(2) *We say that $\widehat{\theta}_n$ is asymptotically unbiased (AU) for θ if*

$$\lim_{n\to\infty} E_\theta\left(\widehat{\theta}_n\right) = \theta, \quad \forall\theta.$$

Example 5.

$Y = Z\beta + U$

$Y = (y_1, ..., y_n)', Z = (z_1', ..., z_n')', U = (u_1, ..., u_n)', \beta = (\beta_1, ..., \beta_k)'$

If u_i are iid $(0, \sigma^2)$ and z_i are non-stochastic (deterministic), then $\widehat{\beta}_n$ is unbiased estimator for β if $\widehat{M}_n > 0$. Moreover,

$$\widehat{\sigma}_n^2 = \frac{1}{n-k}\sum_{i=1}^n \left(y_i - \widehat{\beta}_n' z_i\right)^2$$

is also an unbiased estimator for σ^2 if $\widehat{M}_n > 0$, while

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \left(y_i - \widehat{\beta}_n' z_i\right)^2$$

is asymptotic unbiased for σ^2 if $\widehat{M}_n \to M > 0$.

Theorem 2. *Consistency and asymptotic unbiasedness do not imply one to each other.*

Proof. (i) AU$\not\Rightarrow$consistency.

Consider $\widehat{\theta} = \widehat{\theta}_n \equiv X \sim \mathcal{N}(\theta, 1)$, $\forall n$.

Then, $E\widehat{\theta}_n = \theta$, but clearly $\Pr\left\{\left|\widehat{\theta}_n - \theta\right| > \delta\right\} > \varepsilon > 0$ and hence

$$\lim_{n\to\infty} \Pr\left\{\left|\widehat{\theta}_n - \theta\right| > \delta\right\} = \lim_{n\to\infty} \Pr\left\{|X - \theta| > \delta\right\} \nrightarrow 0.$$

(ii) Consistency $\not\Rightarrow$ AU.

Consider

$$\widehat{\theta}_n = \begin{cases} \theta & \text{with prob. } 1 - n^{-1} \\ n^{\alpha} & \text{with prob. } n^{-1}. \end{cases}$$

Then, for any $\delta > 0$, we obtain that

$$\Pr\left\{\left|\widehat{\theta}_n - \theta\right| > \delta\right\} = \Pr\left\{\widehat{\theta}_n = n^{\alpha}\right\} = \frac{1}{n} \to 0 \text{ as } n \to \infty.$$

On the other hand,

$$\begin{aligned} E\widehat{\theta}_n &= \theta\left(1 - \frac{1}{n}\right) + n^{\alpha} \cdot n^{-1} \\ &= \theta - \frac{\theta}{n} + n^{\alpha-1} \to \begin{cases} \theta & \text{if } \alpha < 1 \\ \theta + 1 & \text{if } \alpha = 1 \\ \infty & \text{if } \alpha > 1, \end{cases} \end{aligned}$$

which implies that $E\widehat{\theta}_n \nrightarrow \theta$ if $\alpha \geq 1$. □

As a rule, many econometric estimators are biased, or asymptotically biased, maybe because their expectation does not exist, or may be technically very difficult to prove their asymptotic unbiasedness. But at the same time, those estimators are consistent, and moreover after an appropriate normalization have a limiting distribution which is centered around the unknown true value of the parameter θ.

4. CONVERGENCE IN r-th MEAN

Definition 8. *(Indicator Function) The indicator function $\mathcal{I}(\bullet)$ is defined as*

$$\mathcal{I}(X \in A) = \begin{cases} 1 & \text{if} \quad X \in A \\ 0 & \text{if} \quad X \notin A. \end{cases}$$

Proposition 2.

$$E\{\mathcal{I}(X \in A)\} = \Pr\{X \in A\}.$$

Proof. Exercise. □

It is worth recalling the following fact. Let $f(\circ)$ and $g(\circ)$ be two nonnegative functions. Suppose that $f(\circ) \le g(\circ)$, then

$$\int f(\circ) \le \int g(\circ). \tag{4.1}$$

Theorem 3. *(Markov's inequality) For any $r > 0$ for which $E|X|^r < \infty$, we have that*

$$\Pr\{|X| > \delta\} \le \frac{E\{|X|^r\}}{\delta^r}, \qquad \forall \delta > 0.$$

Proof.

$$\begin{aligned} \Pr\{|X| > \delta\} &= E\{\mathcal{I}(|X| > \delta)\} \\ &\le E\left\{\mathcal{I}(|X| > \delta)\frac{|X|^r}{|\delta|^r}\right\} \\ &= \delta^{-r} E\{|X|^r \mathcal{I}(|X| > \delta)\} \\ &\le \delta^{-r} E\{|X|^r\}, \end{aligned}$$

where denoting by $p(\circ)$ the pdf of X, in the first inequality we used (4.1) with $f(\circ) =: \mathcal{I}(|X| > \delta)\, p(X)$ and $g(\circ) =: \mathcal{I}(|X| > \delta)\frac{|X|^r}{|\delta|^r} p(X)$ and in the last inequality $f(\circ) =: |X|^r \mathcal{I}(|X| > \delta)\, p(X)$ and $g(\circ) =: |X|^r p(X)$. □

Example 6. *When $r = 2$, the Markov's inequality is known as Chebyshev's inequality.*

Definition 9 (**Convergence in r-th mean**). *Assuming that* $E[|X_n|^r] < \infty$ *and* $E[|X|^r] < \infty$*, we say that* $X_n \xrightarrow{r-th} X$ *if*

$$\lim_{n\to\infty} E\left\{|X_n - X|^r\right\} = 0.$$

Example 7. *When* $r = 2$*, the convergence in* r*-th mean is known as "Mean square convergence".*

Theorem 4. *For any* $r > 0$*,*

$$X_n \xrightarrow{r-th} X \Longrightarrow X_n \xrightarrow{P} X.$$

Proof. By Markov's inequality, i.e. Theorem 3, we have that

$$\Pr\left\{|X_n - X| > \delta\right\} \leq \delta^{-r} E\left\{|X_n - X|^r\right\}.$$

But $X_n \xrightarrow{r-th} X$, so the right side of the last display inequality converges to 0 as $n \to \infty$. □

Theorem 5. *Let $s > r > 0$. Then,*

$$X_n \xrightarrow{s-th} X \Longrightarrow X_n \xrightarrow{r-th} X.$$

Proof. By Jensen's inequality, we know that if $g(\circ)$ is convex,

$$g(E\{X\}) \leq E\{g(X)\}.$$

Now, we know that for any $r > 1$, $f(z) =: z^r$ is a convex function.

So, choosing $g(z) = z^{s/r}$, we obtain that

$$\begin{aligned} g(E\{|X_n - X|^r\}) &= (E\{|X_n - X|^r\})^{s/r} \\ &\leq E\left\{(|X_n - X|^r)^{s/r}\right\} \\ &= E\{|X_n - X|^s\}. \end{aligned}$$

From here we conclude that $X_n \xrightarrow{r-th} X$ if $X_n \xrightarrow{s-th} X$. □

Definition 10. *(Consistency in rth-mean) $\widehat{\theta}_n$ is r-th mean consistent for θ if*

$$E_\theta\left\{\left|\widehat{\theta}_n - \theta\right|^r\right\} \underset{n\to\infty}{\longrightarrow} 0, \ \forall\theta.$$

Theorem 6. *r-th mean consistency ($r \geq 1$) implies asymptotic unbiasedness.*

Proof.

$$\left|E_\theta\left(\widehat{\theta}_n\right) - \theta\right| \leq E_\theta\left\{\left|\widehat{\theta}_n - \theta\right|\right\} \leq \left(E_\theta\left\{\left|\widehat{\theta}_n - \theta\right|^r\right\}\right)^{1/r} \to 0,$$

by Jensen's inequality, see also Theorem 5. □

Remark 1. *If $X_n \xrightarrow{r-th} c$ and $g(\circ)$ is a continuous function, then we have that $g(X_n) \xrightarrow{P} g(c)$. However, this does not imply that $g(X_n) \xrightarrow{r-th} g(c)$. The reason is because $E\{|g(X_n)|^r\}$ may not be even finite!!*

5. CONDITIONS FOR CONVERGENCE IN r-th MEAN

Many statistics of interest can be written as functions of sample means of random variables which they do not need to be iid.

- For instance, the least squares $\widehat{\beta}_n$ satisfies that

$$\widehat{\beta}_n - \beta = \left(\frac{1}{n} \sum_{i=1}^{n} z_i^2 \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} z_i u_i.$$

  That is, a function of

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} z_i^2 &= : \frac{1}{n} \sum_{i=1}^{n} \xi_i \\ \frac{1}{n} \sum_{i=1}^{n} z_i u_i &= : \frac{1}{n} \sum_{i=1}^{n} \eta_i. \end{aligned}$$

  So, by Theorem 1, to show that $\widehat{\beta}_n$ is a consistent estimator for β, since product is a continuous function, it suffices to show that

$$\frac{1}{n} \sum_{i=1}^{n} \xi_i \xrightarrow{P} M > 0; \qquad \frac{1}{n} \sum_{i=1}^{n} \eta_i \xrightarrow{P} 0.$$

- You can check that, for instance, the MLE can also be written as a function of sample means.

The next theorem, although trivial, becomes a very useful tool to prove and/or check the convergence in <u>*2nd mean*</u>.

Theorem 7. *Let* $\{X_i \text{ } (m \times 1 \text{ vector}), \, i \geq 1\}$ *satisfy*

$$E(X_i) = \mu_i,$$
$$\text{Cov}(X_i, X_j) = E\left((X_i - \mu_i)(X_j - \mu_j)'\right) = R_{ij.}$$

Then,

$$\overline{X}_n - \overline{\mu} \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) \xrightarrow{2nd} 0 \Longleftrightarrow \frac{1}{n^2}\sum_{i,j=1}^{n} tr(R_{ij}) \to 0.$$

Proof. Using that $\|a\|^2 = a'a$, and that $tr(AB) = tr(BA)$, we have that

$$\begin{aligned} E\left\{\left\|\overline{X}_n - \overline{\mu}\right\|^2\right\} &= E\left\{tr\left(\left(\overline{X}_n - \overline{\mu}\right)\left(\overline{X}_n - \overline{\mu}\right)'\right)\right\} \\ &= trE\left\{\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(X_i - \mu_i)\tfrac{1}{n}\sum_{j=1}^{n}(X_j - \mu_j)'\right\} \\ &= tr\left\{\tfrac{1}{n^2}\textstyle\sum_{i,j=1}^{n} E\left\{(X_i - \mu_i)(X_j - \mu_j)\right\}'\right\} \\ &= \tfrac{1}{n^2}\textstyle\sum_{i,j=1}^{n} tr(R_{ij}). \end{aligned}$$

From here we conclude. □

Example 8. *(Chebyshev's Weak Law of Large Numbers)* x_i *scalar,* $Ex_i = \mu_i = 0$*,* $R_{ij} = \sigma_i^2 \mathcal{I}(i = j)$*. Then, we have that*

$$\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 \to 0 \Longrightarrow \overline{X}_n \overset{2-nd}{\to} 0 \Longrightarrow \overline{X}_n \overset{P}{\to} 0.$$

This is true under homoscedasticity, (i.e. $\sigma_i^2 = \sigma^2 \; \forall i = 1, ..., n$*); but it holds also true if* $\sigma_i^2 = \sigma^2 i^{\alpha}$ *for* $\alpha < 1$*. However, it would not be true if* $\sigma_i^2 = \sigma^2 i$*, (i.e. linear trend in variance).*

Definition 11 (*Linear process*)**.** *Let $\{e_i\}$ be a sequence of $m \times 1$ -random vectors where $E(e_i) = 0$ and $E\left(e_i e_j'\right) = S\mathcal{I}(i = j)$, $S > 0$. Let $\{A_j\}$ be a sequence of $m \times m$ -nonrandom matrices and*

$$v_i = \sum_{j=0}^{\infty} A_j e_{i-j};\ \sum_{j=0}^{\infty} \|A_j\| < \infty.$$

Then, we say that v_i is a linear process*.*

Example 9. *Some examples of linear processes*

(1) *Stationary Autoregressive of order p (AR(p))*

$$v_i = \sum_{j=1}^{p} C_j v_{i-j} + e_i,$$

where

$$\left| I_m - \sum_{j=1}^{p} C_j z^j \right| \neq 0 \text{ for } |z| \leq 1. \tag{5.1}$$

(Recall that the condition for stationarity for the scalar AR(1) model,

$$v_i = \alpha v_{i-1} + e_i,$$

is $|\alpha| < 1$.)

(2) *Moving Average of order q (MA(q))*

$$v_i = \sum_{j=0}^{q} D_j e_{i-j}.$$

(3) *Autoregressive Moving Average of order p,q (ARMA(p,q))*

$$v_i = \sum_{j=1}^{p} C_j v_{i-j} + \sum_{j=0}^{q} D_j e_{i-j},$$

where C_j satisfies the condition given in (5.1)*.*

Theorem 8. *(Autocovariance structure for linear processes)*

$$(a)\ \ v_i\ \ linear\ \Longrightarrow E\left(v_i v_j'\right) = \sum_{k=0}^{\infty} A_k S A_{k+j-i}',\ i \leq j,$$

$$(b)\qquad\qquad \lim_{n\to\infty}\sup \frac{1}{n}\sum_{i,j=1}^{n} \left\| E\left(v_i v_j'\right)\right\| < \infty.$$

Proof. We begin with (a).

$$\begin{aligned}
E(v_i v_j') &= E\left(\sum_{k=0}^{\infty} A_k e_{i-k}\left(\sum_{l=0}^{\infty} A_l e_{j-l}\right)'\right) \\
&= \sum_{k,l=0}^{\infty} E\left(A_k e_{i-k} e_{j-l}' A_l'\right) \\
&= \sum_{k=0}^{\infty} A_k S A_{k+j-i}'
\end{aligned}$$

because $E\left(e_{i-k} e_{j-l}'\right) = S\mathcal{I}\left(i-k=j-l\right)$.

On the other hand, $E(v_j v_i') = \left(E(v_i v_j')\right)'$ which implies that for $j < i$,

$$E(v_i v_j') = \sum_{k=0}^{\infty} A_{k+i-j} S A_k'.$$

This concludes the proof of part (a).

Next we show part (b). To that end, we notice that

$$\begin{aligned}
\frac{1}{n}\sum_{i,j=1}^{n} \left\|E(v_i v_j')\right\| &= \frac{1}{n}\sum_{i=1}^{n} \|E(v_i v_i')\| + \frac{2}{n}\sum_{i<j}^{n} \left\|E(v_i v_j')\right\| \\
&\leq \frac{2}{n}\sum_{i\leq j}^{n} \left\|E(v_i v_j')\right\| \\
&\leq \frac{2}{n}\sum_{i\leq j}^{n} \left\|\sum_{k=0}^{\infty} A_k S A_{k+j-i}'\right\|
\end{aligned}$$

$$\leq\ \ 2\|S\|\sum_{k=0}^{\infty}\|A_k\|\left\{\frac{1}{n}\sum_{i\leq j}^{n}\|A_{k+j-i}\|\right\} \leq 2\|S\|\sum_{k=0}^{\infty}\|A_k\|\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=0}^{\infty}\|A_j\|\right)$$

$$\leq\ \ 2\|S\|\left(\sum_{k=0}^{\infty}\|A_k\|\right)^2 < \infty.$$

From here the conclusion is straightforward. $\square$

Example 10. *(Multiple regression) Consider the linear regression model,*

$$y_i = z_i'\beta + u_i,$$

where y_i and u_i are scalars, z_i deterministic such that $\max_i \|z_i\| < \infty$, and u_i is homoscedastic although maybe a linear process. Then, we shall use Theorem 8 part (b) with $v_i = z_i u_i$ there. First, we notice that

$$R_{ij} = z_i z_j' E(u_i u_j).$$

Thus, recalling that $tr\,(AB) = tr\,(BA)$, we have that

$$\begin{aligned}
\frac{1}{n^2}\sum_{i,j=1}^{n} \operatorname{tr}(R_{ij}) &\le \frac{1}{n^2}\sum_{i,j=1}^{n} \left|\operatorname{tr}\left(z_i z_j'\right)\right| |E(u_i u_j)| \\
&= \frac{1}{n^2}\sum_{i,j=1}^{n} \left|z_j' z_i\right| |E(u_i u_j)| \\
&\le \frac{1}{n^2}\sum_{i,j=1}^{n} \|z_i\| \|z_j\| |E(u_i u_j)| \\
&\le \underbrace{\frac{\max_i \|z_i\|^2}{n}}_{\to 0} \cdot \underbrace{\frac{1}{n}\sum_{i,j=1}^{n} |E(u_i u_j)|}_{<\infty} \to 0,
\end{aligned}$$

by Theorem 8 part (b). Now use Theorem 7 to conclude that

$$\left.\begin{array}{r} \frac{1}{n}\sum_{i=1}^{n} z_i u_i \to 0 \text{ in 2nd mean} \\ \widehat{M}_n \to M > 0 \end{array}\right\} \Longrightarrow \widehat{\beta}_n \to \beta \text{ in 2nd mean.}$$

Example 11. *(Multiple regression)*

$$y_i = \beta' z_i + u_i$$

Now z_i is assumed to be stochastic and it satisfies

$$\textit{(1)}\ \widehat{M}_n \xrightarrow{P} M = E(z_i z_i') > 0.$$

Also, u_i are not assumed to be independent of z_i . But,

$$\textit{(2)}\ E\left(u_i|z_i\right) = 0,\ E\left(u_i u_j|z_i, z_j\right) = \sigma^2 \mathcal{I}\left(i = j\right),$$

which implies that $Eu_i = 0$, $Eu_i u_j = \sigma^2 \mathcal{I}\left(i = j\right)$.

Again we take $v_i = z_i u_i$, so that

$$\begin{aligned} R_{ij} &= E\left(z_i z_j' u_i u_j\right) \\ &= E\left(z_i z_j' E\left[u_i u_j | z_i, z_j\right]\right) = \sigma^2 E(z_i z_j') \cdot \mathcal{I}\left(i = j\right). \end{aligned}$$

Thus,

$$\begin{aligned} \frac{1}{n^2} \sum_{i,j=1}^{n} \operatorname{tr}\left(R_{ij}\right) &= \frac{\sigma^2}{n^2} \sum_{i=1}^{n} \operatorname{tr} E\left(z_i z_i'\right) \\ &= \frac{\sigma^2}{n} \frac{\operatorname{tr} \sum_{i=1}^{n} E(z_i z_i')}{n} \\ &= \frac{\sigma^2}{n} \operatorname{tr}\left(M\right) \to 0. \end{aligned}$$

Hence, we conclude that $n^{-1} \sum_{i=1}^{n} z_i u_i \xrightarrow{2nd} 0$ by Theorem 7, which implies convergence in probability by Theorem 4.

Also $\widehat{M}_n \xrightarrow{P} M > 0 \Longrightarrow \widehat{M}_n^{-1} \xrightarrow{P} M^{-1}$ (Theorem 1). Then, we have that

$$\widehat{\beta}_n - \beta = \widehat{M}_n^{-1} \frac{1}{n} \sum_{i=1}^{n} z_i u_i \xrightarrow{P} 0 \qquad \textit{(Theorem 1 again).}$$

Remark 2. (1) *The conditions "*$E(u_i|z_i)=0$*", and "*$E(u_iu_j|z_i,z_j)=\sigma^2\mathcal{I}(i=j)$*" are obviously satisfied if* $\{u_i\}$ *and* $\{z_i\}$ *are mutually independent sequences.*

(2) *The condition "*$\widehat{M}_n \xrightarrow{P} M$*" is satisfied if* z_iz_i' *is uncorrelated with the same finite 2nd moments. The latter is true if* z_i *are iid with* $E|z_i|^4<\infty$*. Why?*

Consider z_i *is scalar. Then, we need to show that*

$$\frac{1}{n}\sum_{i=1}^{n} z_i^2 \xrightarrow{P} E\left(z_i^2\right) \quad \text{or} \ \frac{1}{n}\sum_{i=1}^{n} z_i^2 - E\left(z_i^2\right) \xrightarrow{P} 0.$$

Denote $z_i^2-E(z_i^2)=:\eta_i$*. Then it suffices to show that*

$$\frac{1}{n}\sum_{i=1}^{n}\eta_i \xrightarrow{P} 0.$$

But η_i *is a zero mean uncorrelated sequence, so*

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\eta_i\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n}E\left(\eta_i^2\right) \le \frac{E\left(z_i^4\right)}{n} \to 0.$$

Use Theorem 7 and then Theorem 4 to conclude.

(3) *Under certain conditions, (1) and (2) still hold even if* z_i *includes lagged* y_i*, e.g.* $z_{ik}=y_{i-1}$ *for some* $k=1,...,p$*.*

(4) *Conditions* (1) *and* (2) *indicate that we can write the regression model in terms of conditional moments. That is,*

$$E(y_i|z_i) = \beta' z_i$$
$$\mathrm{Cov}(y_i,y_j|z_i,z_j) = \sigma^2\mathcal{I}(i=j).$$

So, we have given a very simple way to check convergence in *2nd mean* (Theorem 8) and hence in probability by Theorem 4.

However, if all we want is convergence in probability, then we know that convergence in **1st mean** is enough.

Also if we believe that the r.v.'s are long-tailed and only first moments (but not finite second ones) are finite, then when would we have convergence in *1st mean*?

To do that, we shall introduce a new concept, which plays an important role not only to show consistency, but it is also very useful to show Central Limit Theorems later on.

Definition 12. *(*Uniform Integrability*)* $\{X_i, i \geq 1\}$ *is a uniformly integrable (UI) sequence if*

$$\lim_{\delta \to \infty} \max_{i \geq 1} E \{ |X_i| \, \mathcal{I}(|X_i| > \delta) \} = 0.$$

Before we give some sufficient conditions for *UI*, we shall remember a standard result

Let X be a positive r.v. such that $EX < \infty$. Then $1 - F(x) \leq Cx^{-1-\zeta}$ for some $\zeta > 0$ and x large enough. Indeed, that $EX < \infty$ implies for all $\varepsilon > 0$

$$\left(\int_0^\infty x f(x) < \infty \right) \Rightarrow \int_M^\infty x f(x) < \varepsilon$$

for some M large enough. The latter is true if $f(x) \leq Cx^{-2-\zeta}$ for $x \geq M$. But

$$\begin{aligned} 1 - F(x) &= \int_x^\infty f(z) \leq C \int_x^\infty z^{-2-\zeta} \\ &\leq Cx^{-1-\zeta}. \end{aligned}$$

Theorem 9. *(*Sufficient conditions for UI*)*

(1) $\max_{i\geq 1} E\left\{|X_i|^{1+\eta}\right\} < \infty$ *for some* $\eta > 0 \Longrightarrow \{X_i, i \geq 1\}$ *UI.*

(2) $\{X_i, i \geq 1\}$ *identically distributed (I.D.) and* $E|X_i| < \infty \Longrightarrow \{X_i, i \geq 1\}$ *UI.*

(3) *If there exists a random variable* X *with finite first moment and such that* $\sup_{i\geq 1} \Pr\{|X_i| > k\} \leq C \Pr\{|X| > k\}$, *for all* k. *Then* $\{X_i, i \geq 1\}$ *is UI.*

Proof. We begin showing part (1). Using (4.1), we have that

$$\begin{aligned}\max_{i\geq 1} E\{|X_i| \mathcal{I}(|X_i| > \delta)\} &\leq \max_{i\geq 1} E\left\{\frac{|X_i|^{1+\eta}}{\delta^\eta} \mathcal{I}(|x_i| > \delta)\right\} \\ &\leq \delta^{-\eta} \max_{i\geq 1} E|x_i|^{1+\eta} \\ &\to 0 \quad \text{as} \quad \delta \to \infty.\end{aligned}$$

Next we show part (2). By definition,

$$\max_{i\geq 1} E\{|X_i| 1(|X_i| > \delta)\} = E\{|X_1| \mathcal{I}(|X_1| > \delta)\},$$

which obviously converges to zero as $\delta \to \infty$ because $E|X_1| < \infty$.

Finally part (3) follows because

$$\begin{aligned}\max_{i\geq 1} E\{|X_i| \mathcal{I}(|X_i| > \delta)\} &\leq C\left(\delta \Pr\{|X| > \delta\} + \int_\delta^\infty \Pr\{|X| > z\} dz\right) \\ &\to 0 \quad (\text{as } \delta \to \infty)\end{aligned}$$

since the first moment of x is finite. The last displayed equality comes from the fact that for a (positive) random variable with pdf $F(x) = 1 - G(x)$,

$$\begin{aligned}E(X\mathcal{I}(X > \delta)) &= \int_\delta^\infty x dF(x) \\ &= \int_\delta^\infty x d(1 - G(x)) \\ &= -\int_\delta^\infty x dG(x) \\ &= \delta G(\delta) + \int_\delta^\infty G(x) dx.\end{aligned}$$

□

Theorem 10. *(Necessary conditions for UI)*
$\{X_i, i \geq 1\}$ *UI* $\Longrightarrow \max_i E|X_i| < \infty$.

Proof. We have that by UI, for any $\varepsilon > 0$, there exists $\delta_0(\varepsilon) < \infty$ such that

$$\max_{i\geq 1} E\{|X_i| \mathcal{I}(|X_i| > \delta_0)\} < \varepsilon.$$

$$\begin{aligned}\max_{i\geq 1} E|X_i| &\leq \max_{i\geq 1} E\{|X_i| \mathcal{I}(|X_i| > \delta_0)\} + \max_{i\geq 1} E\{|X_i| \mathcal{I}(|X_i| \leq \delta_0)\} \\ &\leq \varepsilon + \delta_0 < \infty.\end{aligned}$$

□

The next example will illustrate the reason why the condition $\sup_{i\geq 1} \mathbb{E}(|X_i|) < \infty$ is not sufficient for UI.

Example 12. *The example shall illustrate why the condition of identically distributed is important for the necessity of UI.*

To that end, let's consider a sequence of random variables $\{X_i\}_{i\geq 1}$ *such that*

$$X_i = \begin{cases} i & i^{-1} \\ 0 & 1 - i^{-1}. \end{cases}$$

Then, it is clear that $\sup_{i\geq 1} \mathbb{E}(|X_i|) < \infty$ *since* $\mathbb{E}(|X_i|) = 1$ *for all* $i \geq 1$*. However,*

$$\mathbb{E}(|X_i| \mathcal{I}(|X_i| > \delta)) = \int_{|x_i|>\delta} x dF_i(x) = 1$$

for all $\delta > 0$*. So,*

$$\lim_{\delta\to\infty} \sup_{i\geq 1} \mathbb{E}(|X_i| \mathcal{I}(|X_i| > \delta)) = 1$$

and hence the sequence is not UI.

Remark 3. *The sufficient condition* (3) *is weaker than that in* (1)*. Indeed, Markov's inequality implies that*

$$\begin{aligned}\max_{i\geq 1}\Pr\left\{|X_i|>z\right\} &\leq \min\left(1,\max_{i\geq 1}\frac{E\left\{|X_i|^{1+\eta}\right\}}{z^{1+\eta}}\right)\\ &\leq \min\left(1,\frac{C}{z^{1+\eta}}\right). \qquad (5.2)\end{aligned}$$

Now consider the random variable X with probability density function

$$f(z)=\begin{cases} C_1, & 0<z<\varepsilon\\ \frac{C_2}{z^{2+\eta}}, & z\geq\varepsilon,\end{cases}$$

where the constants $C_1, C_2>0$ are chosen to guarantee that $f(\circ)$ integrates 1*. Then, we have that for $0<z<\varepsilon$*

$$\begin{aligned}\Pr\left\{X>z\right\} &= \int_z^\infty f(u)\,du = C_1\int_z^\varepsilon du + C_2\int_\varepsilon^\infty u^{-2-\eta}du\\ &= C_1(\varepsilon-z)+C_2\varepsilon^{-1-\eta}/(1+\eta).\end{aligned}$$

On the other hand, for $z\geq\varepsilon$, we have that

$$\Pr\left\{X>z\right\}=C_2 z^{-1-\eta}/(1+\eta).$$

Thus,

$$\begin{aligned}\Pr\left\{X>z\right\} &\geq \frac{C_2}{1+\eta}\min\left(\varepsilon^{-1-\eta},z^{-1-\eta}\right)\\ &\geq c\min\left(1,z^{-1-\eta}\right).\end{aligned}$$

So, using (5.2)*, we conclude that*

$$\max_{i\geq 1}\Pr\left\{|X_i|>z\right\}\leq C\Pr\left\{X>z\right\}.$$

This completes the proof.

Theorem 11. *(Weak Law of Large Numbers for Independent UI sequences.) Let $\{x_i, i \geq 1\}$ be a sequence of UI independent, random variables such that $Ex_i = 0$ for all $i \geq 1$. Then, we have that*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \overset{1st-mean}{\rightarrow} 0,$$

and as a consequence $\overline{X}_n$ converges to zero in probability by Theorem 4.

Proof. Fix any $\varepsilon > 0$ and split up x_i into x_i' and x_i'' as follows;

$$\left.\begin{array}{l} x_i' = x_i \mathcal{I}\left(|x_i| \leq \delta\right) \\ x_i'' = x_i \mathcal{I}\left(|x_i| > \delta\right) \end{array}\right\} \Rightarrow x_i = x_i' + x_i''.$$

Since x_i is UI $\Longrightarrow \exists \delta > 0$ such that $\max_{i \geq 1} E\left|x_i''\right| \leq \varepsilon$.

Now,

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n}\left(x_i' - Ex_i'\right) + \frac{1}{n}\sum_{i=1}^{n}\left(x_i'' - Ex_i''\right). \tag{5.3}$$

Recall that $Ex_i = 0 = Ex_i' + Ex_i''$. Let's examine the first term on the right of (5.3). Its second moment is

$$\begin{aligned} E\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i' - Ex_i'\right)\right)^2 &= \frac{1}{n^2}\sum_{i=1}^{n} E\left(x_i' - Ex_i'\right)^2 \text{ by independence} \\ &\leq 2n^{-2}\delta^2 n = 2\delta^2/n \quad \text{since } |x_i'| \leq \delta. \end{aligned}$$

Then, the first term on the right of (5.3) converges in *2nd mean* to zero, and hence by Theorem 5, it converges in *1st mean.*

For the second term on the right of (5.3), we have that its first absolute moment is

$$\begin{aligned} E\left|\frac{1}{n}\sum_{i=1}^{n}\left(x_i'' - Ex_i''\right)\right| & \underset{\substack{\uparrow \\ \text{triangle inequality}}}{\leq} \frac{1}{n}\sum_{i=1}^{n} E\left|x_i'' - Ex_i''\right| \\ \leq \frac{2}{n}\sum_{i=1}^{n} E\left|x_i''\right| & \\ \leq \frac{2}{n}\sum_{i=1}^{n}\max_{i\geq 1} E\left|x_i''\right| = 2\max_{i\geq 1} E\left|x_i''\right| < \varepsilon \quad \text{by} \quad \text{UI.} & \end{aligned}$$

Now, because we can take ε arbitrarily small, the second term on the right of (5.3) converges to zero in <u>*1st mean*</u> also. So, we conclude that

$$\overline{X}_n \overset{1st-mean}{\rightarrow} 0,$$

and by Theorem 4, $\overline{X}_n \xrightarrow{P} 0$.

$\square$

Remark 4. *The assumption of independence can be relaxed to the assumption that* $\{X_i\}$ *is a martingale difference sequence, i.e.,*

$$E\left(X_i|X_j, j < i\right) = 0.$$

This is important for time series applications.

Remark 5. *The assumption of* $EX_i = 0$ *is not important, we just replace* X_i *by* $Z_i = X_i - EX_i$*.*

Example 13. *(Khinchine's Weak Law of Large Numbers) If* $\{X_i, i \geq 1\}$ *is iid with* $E\left|X_i\right| < \infty$*,* $EX_i = 0$*, then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} 0.$$

In fact, under the same conditions, Kolmogorov showed that we have almost sure convergence and it is known as the Strong Law of Large Numbers (SLLN), that is

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} 0.$$

If we relax the condition of identically distributed, then to achieve the *SLLN* we need to strengthen the moment conditions. In particular we have that if the sequence X_i is independent and satisfies that (a) for some $\delta > 0$, $\sum_{i=1}^{\infty} E\left|X_i\right|^{1+\delta}/i^{1+\delta} < \infty$. Then

$$\frac{1}{n}\sum_{i=1}^{n} \left(X_i - \mu_i\right) \xrightarrow{a.s.} 0.$$

Observe that a sufficient condition for condition (a) is that $\sup_i E\left|X_i\right|^{1+\delta} \leq C < \infty$.

Example 14. *(Multiple regression with deterministic z_i)*

Assume $\max_{i\geq 1}\|z_i\| < \infty$ *(i.e.,* $\max_{i\geq 1}|z_{ji}| < \infty$, $\forall j = 1,...,k$*) and* $\widehat{M}_n = n^{-1}Z'Z \to M > 0$. *Let* $\{u_i, i \geq 1\}$ *be a UI sequence of independent random variables with* $Eu_i = 0$. *Denote by* $\widehat{\beta}_n$ *the LSE of* β. *Then,*

$$\begin{aligned} E\left\|\widehat{\beta}_n - \beta\right\| &= E\left\|\widehat{M}_n^{-1} n^{-1} Z'U\right\| \\ &\leq \left\|\widehat{M}_n^{-1}\right\| E\left\|n^{-1}Z'U\right\| \text{ where } \left\|\widehat{M}_n^{-1}\right\| \to \left\|M^{-1}\right\| < \infty. \end{aligned}$$

Denote $X_i = z_i u_i$. *Because* $\{u_i\}$ *is independent and* $\{z_i\}$ *is deterministic, we have that* $\{X_i\}$ *is also an independent sequence such that*

$$\begin{aligned} EX_i &= z_i E u_i = 0 \\ \|X_i\| &\leq \|z_i\|\,|u_i|\,. \end{aligned}$$

Now, using the inequality in (4.1) *when it is needed, we have that*

$$\begin{aligned} E\left(\|X_i\|\,\mathcal{I}\left(\|X_i\| > \delta\right)\right) &\leq E\left(\|z_i\|\,|u_i|\,\mathcal{I}\left(|u_i| > \frac{\delta}{\|z_i\|}\right)\right) \\ &\leq \max_i \|z_i\|\, E\left(|u_i|\,\mathcal{I}\left(|u_i| > \frac{\delta}{\max_i \|z_i\|}\right)\right) \\ &\leq \max_i \|z_i\| \max_i E\left(|u_i|\,\mathcal{I}\left(|u_i| > \frac{\delta}{c}\right)\right) \\ &\to 0 \text{ as } \delta \to \infty \text{ since } u_i \text{ is UI.} \end{aligned}$$

So, $\{X_i, i \geq 1\}$ *is UI and thus Theorem 11 implies that*

$$\frac{1}{n}Z'U \xrightarrow{1st} 0 \;\Rightarrow \widehat{\beta}_n \xrightarrow{1st} \beta.$$

Definition 13. *(Generalized Linear Processes) Let $\{e_i, -\infty < i < \infty\}$ be an independent, UI sequence of $m \times 1$ -random vectors, (i.e., $\forall j = 1, ..., m$, e_{ij} is UI), where $Ee_i = 0$ $\forall i \geq 1$. Let $\{A_j\}$ be a sequence of $m \times m$ -nonrandom matrices and*

$$v_i = \sum_{j=0}^{\infty} A_j e_{i-j}, \ \sum_{j=0}^{\infty} \|A_j\| < \infty.$$

Then v_i is a generalized linear process (GLP).

Example 15 (cont. example 9)**.** *MA(q), AR(p), ARMA(p,q) satisfying*
$|I_m - \sum_1^p C_j z^j| \neq 0$ *for all $|z| \leq 1$, are GLP if the e_i are iid with finite first moment.*

Theorem 12. $\{X_i, i \geq 1\}$ *GLP* $\Longrightarrow n^{-1} \sum_{i=1}^n X_i \stackrel{1st}{\to} 0.$

Proof. Take $m = 1$ for simplicity. Fix any $\varepsilon > 0$, then there exists a finite N such that $\sum_{j=N+1}^{\infty} \|A_j\| < \varepsilon$. Now,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n X_i &= \frac{1}{n}\sum_{i=1}^n \left(\sum_{j=0}^{\infty} A_j e_{i-j}\right) \\
&= \frac{1}{n}\sum_{i=1}^n \sum_{j=0}^{N} A_j e_{i-j} + \frac{1}{n}\sum_{i=1}^n \sum_{j=N+1}^{\infty} A_j e_{i-j} \\
&= \sum_{j=0}^{N} A_j \left(\frac{1}{n}\sum_{i=1}^n e_{i-j}\right) + \sum_{j=N+1}^{\infty} A_j \left(\frac{1}{n}\sum_{i=1}^n e_{i-j}\right).
\end{aligned}$$

Hence
(5.4)
$$E\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \leq \sum_{j=0}^{N} \|A_j\| \, E\left|\frac{1}{n}\sum_{i=1}^n e_{i-j}\right| + \sum_{j=N+1}^{\infty} \|A_j\| \, E\left|\frac{1}{n}\sum_{i=1}^n e_{i-j}\right|.$$

By Theorem 11, $n^{-1}\sum_{i=1}^n e_{i-j} \stackrel{1st}{\to} 0$. So, $\exists n_0$ such that $\forall n \geq n_0$, $E\left|n^{-1}\sum_{i=1}^n e_{i-j}\right| < \varepsilon$ uniformly over $j = 0, 1, ..., N$.

Thus, the first term on the right of (5.4) is

$$\sum_{j=0}^{N} \|A_j\| \, E\left|\frac{1}{n}\sum_{i=1}^n e_{i-j}\right| \leq \varepsilon \sum_{j=0}^{\infty} \|A_j\| = \varepsilon'.$$

Because $E\left|n^{-1}\sum_{i=1}^n e_{i-j}\right| \leq \max_{i \geq 1} E|e_i|$, we have that the second term of (5.4) is

$$\sum_{j=N+1}^{\infty} \|A_j\| \, E\left|\frac{1}{n}\sum_{i=1}^n e_{i-j}\right| \leq \max_{i\geq 1} E|e_i| \quad \sum_{j=N+1}^{\infty} \|A_j\| = \varepsilon''.$$

The result now follows since $\varepsilon' + \varepsilon'' = \varepsilon$ which can be chosen arbitrarily small. $\square$

6. CONDITIONS FOR CONSISTENCY IN REGRESSION WITH TRENDING REGRESSORS

Consider the linear regression model

$$y_i = \beta' z_i + u_i, \qquad i = 1, ..., n,$$

where $E(u_i) = 0$, $E(u_i^2) = \sigma^2$ and $E(u_i u_j) = 0$ for $i \neq j$. Put $Q_n = Z'Z$ and let z_i be deterministic. We have seen that under these conditions the LSE of β, $\widehat{\beta}_n$, converges in second mean if

$$\lim_{n\to\infty} \frac{Q_n}{n} = M, \tag{6.1}$$

where M is finite and non-singular. The question is how important is condition (6.1) for such a result. For instance what would it happen if I do not assume that Q_n/n has a limit which is either finite or greater than zero? That is, if we allow for trends.

Example 16. *If $z_i = i$, then $\frac{Q_n}{n} = \frac{1}{n}\sum_{i=1}^{n} i^2 \to \infty$.*

Theorem 13. *Let $Eu_i = 0$, $E(u_i u_j) = \sigma^2 \mathcal{I}(i = j)$. Then $\widehat{\beta}_n \overset{2nd}{\to} \beta \Leftrightarrow \underline{\lambda}(Q_n) \to \infty$ (i.e., the smallest eigenvalue of Q_n increases to infinity with the sample size).*

Proof. Recall $\widehat{\beta}_n = \beta + (Z'Z)^{-1} Z'U$. So,

$$\begin{aligned} E\left\{\left\|\widehat{\beta}_n - \beta\right\|^2\right\} &= \operatorname{tr} E\left\{\left(\widehat{\beta}_n - \beta\right)\left(\widehat{\beta}_n - \beta\right)'\right\} \\ &= \operatorname{tr} E\left\{Q_n^{-1}\left(\sum_i z_i u_i\right)\left(\sum_i z_i u_i\right)' Q_n^{-1}\right\} \\ &= \sigma^2 \operatorname{tr}\left(Q_n^{-1}\right). \end{aligned}$$

Then, the right side of the last displayed equality converges to zero if and only if $\underline{\lambda}(Q_n) \to \infty$. □

Remark 6. *$\underline{\lambda}(Q_n) \leq \underline{q}_{ii,n}$, where $\underline{q}_{ii,n}$ is the i-th diagonal element of Q_n. So $\underline{\lambda}(Q_n) \to \infty \Rightarrow \underline{q}_{ii,n} \to \infty \;\; \forall i = 1, ..., k$.*

Another way to discuss when (and why) the *LSE* $\widehat{\beta}_n$ is consistent is as follows. Let $h(\eta_n)$ be a sequence of positive random variables. Then, for any $\kappa > 0$, we have that Markov's inequality implies

$$\Pr\{h(\eta_n) > \kappa\} \leq \frac{Eh(\eta_n)}{\kappa}.$$

Hence, if we take $h(\eta_n) = (\eta_n - \mu)^2$, where $E(\eta_n) = \mu$ and $var(\eta_n) = \sigma_n^2$, the Markov's inequality implies that

$$\Pr\{|\eta_n - \mu| \geq \kappa\} \leq \frac{\sigma_n^2}{\kappa^2}.$$

From here it follows that a sufficient (and necessary condition **as** the second moments are finite) for $\eta_n \xrightarrow{P} \mu$ is

$$\lim_{n\to\infty} \sigma_n^2 = 0.$$

So, if $\eta_n = \widehat{\theta}_n$, an estimator of a parameter $\theta = \mu$, a sufficient condition for $\widehat{\theta}_n$ to be consistent is that $E\widehat{\theta}_n = \theta$ and $V\left(\widehat{\theta}_n\right) \to 0$. Again, I should emphasize that it is sufficient because for consistency we do not need any moment to be finite.

In our case we have that $E\left(\widehat{\beta}_n\right) = \beta$ and

$$V_n = \sigma^2 \left(Z'Z\right)^{-1}.$$

Thus, we can conclude that a sufficient condition for consistency is that $\lim_{n\to\infty} \left(Z'Z\right)^{-1} = 0$. From here we can see that the condition $Z'Z/n \to M$ is sufficient. Indeed,

$$\begin{aligned}\lim_{n\to\infty} V_n &= \sigma^2 \lim_{n\to\infty} n^{-1} \left(\frac{Z'Z}{n}\right)^{-1} \\ &= \sigma^2 M \lim_{n\to\infty} n^{-1} = 0.\end{aligned}$$

However, we might have that

$$\frac{Z'Z}{n^{1/2}} \to M$$

and $\widehat{\beta}_n$ still being consistent. The reason is because

$$\begin{aligned}\lim_{n\to\infty} V_n &= \sigma^2 \lim_{n\to\infty} \left(Z'Z\right)^{-1} \\ &= \sigma^2 \lim_{n\to\infty} n^{-1/2} \left(\frac{Z'Z}{n^{1/2}}\right)^{-1} \\ &= \sigma^2 M \lim_{n\to\infty} n^{-1/2} = 0.\end{aligned}$$

Of course, we can have also the situation where, for some $\delta > 0$,

$$\frac{Z'Z}{n^{1+\delta}} \to M.$$

Example 17. *Consider the regression model*

$$y_i = \alpha + \beta i + u_i.$$

Then,

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^{n} i \\ \sum_{i=1}^{n} i & \sum_{i=1}^{n} i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n} u_i \\ \sum_{i=1}^{n} iu_i \end{pmatrix}$$

which is consistent as both n and $\sum_{i=1}^{n} i^2$ increase to infinity. Recall our previous remark.

7. STOCHASTIC ORDER OF MAGNITUDE

These are very useful in proving theorems on large sample behaviour and also enable us to concisely make stronger statements than $X_n \xrightarrow{P} c$ when necessary.

Let f_n be a nonstochastic sequence.

Definition 14. *(X_n is deterministic)*

(1) *We say that* $X_n = O(f_n)$ *if* $\left|\frac{X_n}{f_n}\right| \to c < \infty$ *(i.e.,* $\left|\frac{X_n}{f_n}\right|$ *is bounded for all sufficiently large* n*).*

(2) *We say that* $X_n = o(f_n)$ *if* $\left|\frac{X_n}{f_n}\right| \to 0$ *as* $n \to \infty$*.*

Definition 15. *(X_n is stochastic)*

(1) We say that $X_n = O_p(f_n)$ (or X_n/f_n bounded in probability) if $\forall\varepsilon > 0$, $\exists c \geq 0$ and $n_0 > 0$ such that

$$\Pr\{|X_n| > cf_n\} < \varepsilon, \quad \forall n \geq n_0.$$

(2) We say that $X_n = o_p(f_n)$ if $X_n/f_n \xrightarrow{P} 0$.

Remark 7. $X_n = o_p(1)$ *means that* $X_n \xrightarrow{P} 0$*.*

Theorem 14. *For a finite constant* c*,*

$$X_n \xrightarrow{P} c \Rightarrow X_n = O_p(1).$$

When $c = 0$*, we have* $X_n = o_p(1) \Leftrightarrow X_n \xrightarrow{P} 0$*.*

Proof. For any $\delta > 0$, there exists $C = |c| + \delta$ such that

$$\begin{aligned} \Pr(|X_n| > C) &= \Pr(|X_n| - |c| > \delta) \\ &\leq \Pr\{|X_n - c| > \delta\}. \end{aligned}$$

But, we have that $X_n \xrightarrow{P} c$, so the right hand side of the last displayed inequality converges to zero as $n \to \infty$. $\square$

Remark 8. *The difference between* $O_p(\cdot)$ *and* $o_p(\cdot)$ *is that: for any* $\varepsilon > 0$*,*

$$O_p(1) \text{ means } \Pr\{|X_n| > c\} < \varepsilon \text{ for } \underline{\textbf{some}}\ c > 0,$$

whereas

$$o_p(1) \text{ means } \Pr\{|X_n| > c\} < \varepsilon \text{ for } \underline{\textbf{all}}\ c > 0.$$

Theorem 15.

(*i*) $X_n = o_p(f_n) \Rightarrow X_n = O_p(f_n)$.

(*ii*) $X_n = O_p(f_n) \Rightarrow X_n = o_p(g_n)$ *if* $\frac{f_n}{g_n} \to 0$.

(*iii*) $X_n = O_p\left((E|X_n|^r)^{1/r}\right)$ *for* $r > 0$.

Proof. We begin with part (*i*). Use Theorem 14 but replacing X_n by X_n/f_n there.

Next part (*ii*). For any $c > 0$,

$$\begin{aligned}\Pr(|X_n| > cg_n) &= \Pr(|X_n| > cf_n g_n/f_n) \\ &\leq \Pr(|X_n| > Cf_n)\end{aligned}$$

as $n \to \infty$ for all $c > 0$, we can choose a finite positive C large enough, (since $g_n/f_n \to \infty$, $cg_n/f_n > C$ for n large enough). Now by assumption, the right side is less than ε.

Finally part (*iii*). By Markov's inequality (Theorem 3), for any $\varepsilon > 0$, there exists a $C = C(\varepsilon) > 0$ such that:

$$\Pr\left\{|X_n| > C(E|X_n|^r)^{1/r}\right\} \leq \frac{E|X_n|^r}{C^r E|X_n|^r} = C^{-r} < \varepsilon.$$

$\square$

Theorem 16.

(1) $X_n = O_p(f_n)$, $Y_n = O_p(g_n)$ *then,*
  (a) $X_n Y_n = O_p(f_n g_n)$
  (b) $X_n + Y_n = O_p(\max(f_n, g_n))$
(2) *We can replace "O" by "o" everywhere*
(3) $X_n = O_p(f_n)$, $Y_n = o_p(g_n) \Longrightarrow X_n Y_n = o_p(f_n g_n)$.

Proof. We begin with part $(1.a)$. For any $\varepsilon > 0$, $\exists C, D, n_1, n_2 > 0$ such that

$$\Pr\{|X_n| > Cf_n\} < \frac{\varepsilon}{2}, \quad \forall n \geq n_1$$
$$\Pr\{|Y_n| > Dg_n\} < \frac{\varepsilon}{2}, \quad \forall n \geq n_2.$$

Hence for all $n \geq \max\{n_1, n_2\} > 0$, we have that

$$\Pr(|X_n Y_n| > CDf_n g_n) \leq \Pr\{|X_n| > Cf_n\} + \Pr\{|Y_n| > Dg_n\} < \varepsilon,$$

where we have used the fact that

$$|cd| > ef \Rightarrow |c| > e \text{ or } |d| > f \tag{7.1}$$

and then (2.1) and that $\Pr\{A \cup B\} \leq \Pr\{A\} + \Pr\{B\}$.

The proof of part $(1.b)$ is similar to that of $(1.a)$ except that we employ the fact:

$$\begin{aligned} &\Pr\{|X_n + Y_n| > (C+D)\max(f_n, g_n)\} \\ &\leq \Pr\{|X_n| + |Y_n| > (C+D)\max(f_n, g_n)\} \text{ triangle inequality} \\ &\leq \Pr\{|X_n| > Cf_n\} + \Pr\{|Y_n| > Dg_n\} \end{aligned}$$

because

$$\Pr\{|A+B| > a+b\} \leq \Pr\{|A| > a\} + \Pr\{|B| > b\}.$$

Part (2) is similar to (1).

Finally part (3). For any $\varepsilon > 0$, $\exists C, n_1 > 0$ such that

$$\Pr\{|X_n| > Cf_n\} < \frac{\varepsilon}{2}, \quad \forall n \geq n_1.$$

For any $\varepsilon > 0, \delta > 0$, $\exists n_2 > 0$ such that

$$\Pr\{|Y_n| > (\delta/C)\, g_n\} < \frac{\varepsilon}{2}, \quad \forall n \geq n_2.$$

Now the result follows by using $\delta f_n g_n = \left(\frac{\delta}{C} g_n\right)(Cf_n)$ and

$$\Pr\{|X_n Y_n| > \delta f_n g_n\} \leq \Pr\{|X_n| > Cf_n\} + \Pr\left\{|Y_n| > \frac{\delta}{C} g_n\right\},$$

using (7.1) again. $\square$

Example 18 (Regression with stochastic z_i)**.** *We already know that* $\widehat{M}_n \xrightarrow{P} M > 0 \Rightarrow \widehat{M}_n^{-1} \xrightarrow{P} M^{-1} < \infty$ *by Theorem 1. Also,*

$$\begin{aligned} R_{ij} &= E\left(z_i z_j'\right)\sigma^2 \mathcal{I}\left(i = j\right) \\ v_i &= z_i u_i. \end{aligned}$$

Thus,

$$\frac{1}{n^2}\sum_{i,j=1}^{n} \operatorname{tr} R_{ij} = \sigma^2 \frac{\operatorname{tr}(M)}{n} = O\left(n^{-1}\right)$$

which implies that

$$\frac{1}{n}\sum_{i=1}^{n} v_i = \frac{1}{n} Z'U = O_p\left(n^{-1/2}\right)$$

by Theorem 15 part (iii)*. So,* $\widehat{\beta}_n - \beta = \widehat{M}_n^{-1}\left(Z'U/n\right)$ *will be* $O_p(1)\, O_p\left(n^{-1/2}\right) = O_p\left(n^{-1/2}\right)$ *by Theorem 16 part* (1.*a*)*.*

8. **CONVERGENCE IN DISTRIBUTION**

We denote by $F_X(x) = \Pr(X \leq x)$ the probability distribution function (DF) of X, and let $\phi_X(t) = E[e^{it'X}]$ be its characteristic function (CF).

Definition 16. *We say that X_n converges in distribution to X, denoted as $X_n \xrightarrow{d} X$, if*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at every ***continuity*** *point of $F_X(\cdot)$.*

Remark 9. *$F(\cdot)$ can have at most a countable number of discontinuous points or jumps such as in discrete or mixed distributions.*

Example 19. $\{X_i, i \geq 1\}$ *are* iid (μ, σ^2) *then,*

$$Y_n =: \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right) \xrightarrow{d} \mathcal{N}(0,1)$$

by Lindeberg-Levy's central limit theorem (CLT) (see later).

Example 20. *Let $\{X_i, i \geq 1\}$ be a sequence of $\mathcal{U}[0,\theta]$ random variables. Consider $Y_n = X_{(n)} = \max_{1\leq i\leq n} X_i$. Then, $Y_n \xrightarrow{d} Y$ whose DF is $F(y) = \mathcal{I}(\theta \leq y)$. To that end, observe that we already know that*

$$f_n(y) = \begin{cases} 0 & \text{if} \quad y < 0 \\ \frac{ny^{n-1}}{\theta^n} & \text{if} \quad 0 \leq y \leq \theta \\ 0 & \text{if} \quad \theta < y. \end{cases}$$

Then,

$$F_n(y) = \begin{cases} 0 & \text{if} \quad y < 0 \\ \left(\frac{y}{\theta}\right)^n & \text{if} \quad 0 \leq y \leq \theta \\ 1 & \text{if} \quad \theta < y, \end{cases}$$

which implies that $\lim_{n\to\infty} F_n(y) = F(y)$ *for all $y < \theta$ or $y > \theta$. That is, the points y where $F(y)$ is a continuous function.*

8.1. **Relation between convergence in distribution and in** rth**-mean.**

We can be inclined to believe that if $X_n \xrightarrow{d} X$ and $E|X_n|$ and $E|X|$ are finite, then $E|X_n| \to E|X|$.

However, this is not true as the next example illustrates.

Example 21. *Let X_n be a sequence of random variables whose DF is given by*

$$F_n(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 - \frac{1}{n} & \text{if} \quad 0 \leq x \leq n \\ 1 & \text{if} \quad n < x. \end{cases}$$

Then it is clear that $X_n \xrightarrow{d} X$ where the DF of X is $F(x) = \mathcal{I}(0 \leq x)$.

However, $E|X_n| \nrightarrow E|X|$ because $EX_n^k = n^k(1/n) = n^{k-1}$, whereas $EX^k = 0$.

So, $EX_n^k \nrightarrow 0$ if $k \geq 1$.

Remark 10. *However, if the sequence $\{X_n^k, n \geq 1\}$ is UI, we would then have*

$$E|X_n|^k \to E|X|^k.$$

This is an important result. You can see the proof in Serfling's book.

Theorem 17. *For all bounded continuous functions $g(\circ)$,*

$$\begin{aligned} X_n \xrightarrow{d} X \quad &\Leftrightarrow \quad \phi_{X_n}(t) \to \phi_X(t) \qquad \forall t \in \mathbb{R} \\ &\Leftrightarrow \quad E\{g(X_n)\} \to E\{g(X)\}. \end{aligned}$$

Proof. See Rao's (1973) Chapter 2. □

Theorem 18. $X_n \xrightarrow{d} X \Leftrightarrow \lambda' X_n \xrightarrow{d} \lambda' X,\ \ \forall\lambda$*-vectors.*

Proof.

$$\begin{aligned} \lambda' X_n \xrightarrow{d} \lambda' X,\ \forall\lambda &\Rightarrow E\left(e^{iv\lambda' X_n}\right) \to E\left(e^{iv\lambda' X}\right)\ \forall v,\ \forall\lambda \\ &\Rightarrow \phi_{X_n}(u) \to \phi_X(u)\ \ \forall u = \lambda v \\ &\Rightarrow X_n \xrightarrow{d} X. \end{aligned}$$

The implication $X_n \xrightarrow{d} X \Rightarrow \lambda' X_n \xrightarrow{d} \lambda' X$ is shown similarly. Indeed,

$$\left.\begin{array}{c} E\left(e^{iu' X_n}\right) \to E\left(e^{iu' X}\right) \\ \text{but } \forall u \in \mathbb{R}^p \ \ u' = v\lambda' \end{array}\right\} \Rightarrow E\left(e^{iv\lambda' X_n}\right) \to E\left(e^{iv\lambda' X}\right).$$

□

Definition 17. *$X \sim \mathcal{N}_k(\mu, \Sigma)$ means that X is a k-variate normal with mean μ and variance-covariance matrix $\Sigma > 0$.*

Theorem 19. *$X \sim \mathcal{N}_k(\mu, \Sigma) \Leftrightarrow \lambda' X \sim \mathcal{N}_1(\lambda'\mu, \lambda'\Sigma\lambda)$ $\forall \lambda \neq 0, \lambda \in \mathbb{R}^k$.*

Proof. $X \sim \mathcal{N}_k(\mu, \Sigma) \Leftrightarrow \phi_X(u) = \exp\{i\mu' u - \frac{1}{2}u'\Sigma u\}$, $\forall u \in \mathbb{R}^k$. But,

$$\begin{aligned} \lambda' X \sim \mathcal{N}_1(\lambda'\mu, \lambda'\Sigma\lambda) \quad &\Leftrightarrow \phi_{\lambda' X}(v) = \exp\{iv\lambda'\mu - \tfrac{1}{2}v^2\lambda'\Sigma\lambda\} \\ \forall \lambda \neq 0, \lambda \in \mathbb{R}^k, \quad &\phantom{\Leftrightarrow \phi_{\lambda' X}(v)} = \phi_X(v\lambda), \ \forall v \in \mathbb{R}, \\ &\Leftrightarrow X \sim \mathcal{N}_k(\mu, \Sigma) \end{aligned}$$

The converse follows from the fact that linear combinations of k-variate normals are normal. □

This theorem is very useful, since if we want to show that a vector X_n is k-variate normal, it suffices to show is that **any linear** combination is univariate normal.

In particular, this fact together with Theorem 18 implies that to show that $X_n \xrightarrow{d} \mathcal{N}_k(\mu, \Sigma)$, it suffices to show that $\lambda' X_n \xrightarrow{d} \mathcal{N}_1(\lambda'\mu, \lambda'\Sigma\lambda)$ for **all** $\lambda \in \mathbb{R}^k$ with $|\lambda| = 1$.

Theorem 20. *For scalar X_n, X and c (constant)*

(**a**) $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$.

(**b**) $X_n \xrightarrow{d} c \Leftrightarrow X_n \xrightarrow{P} c$.

Proof. We begin with part (a). $\forall \varepsilon > 0$, let $x, x-\varepsilon, x+\varepsilon$ be continuity points of $F_X(\cdot)$. Then, using (2.3), we have that

$$\begin{aligned} F_{X_n}(x) &= \Pr\{X_n \leqslant x\} \\ &= \Pr\{X_n \leqslant x, \quad X - X_n \leqslant \varepsilon\} + \Pr\{X_n \leqslant x, \quad X - X_n > \varepsilon\} \\ &\leqslant F_X(x+\varepsilon) + \Pr(|X - X_n| > \varepsilon) \end{aligned}$$

using (2.1). So, we have shown that

$$\limsup_{n\to\infty} F_{X_n}(x) \leqslant F_X(x+\varepsilon). \tag{8.1}$$

However, we have also that

$$\begin{aligned} F_X(x-\varepsilon) &= \Pr\{X \leqslant x-\varepsilon\} \\ &= \Pr\{X \leq x-\varepsilon, \quad X_n - X \leq \varepsilon\} + \Pr\{X \leq x-\varepsilon, \quad X_n - X > \varepsilon\} \\ &\leq F_{X_n}(x) + \Pr\{|X_n - X| > \varepsilon\} \end{aligned}$$

using again (2.1). Thus,

$$F_X(x-\varepsilon) \leq \liminf_{n\to\infty} F_{X_n}(x) \tag{8.2}$$

together with (8.1) implies that

$$\begin{aligned} F_X(x-\varepsilon) \leqslant \liminf_{n\to\infty} F_{X_n}(x) \leqslant \limsup_{n\to\infty} F_{X_n}(x) \leqslant F_X(x+\varepsilon) \\ \Rightarrow F_X(x) = \lim_{n\to\infty} F_{X_n}(x) \end{aligned}$$

since ε is arbitrarily small. This concludes the proof of part (a).

Next, we examine part (b). $X_n \xrightarrow{d} c \Rightarrow X_n \xrightarrow{P} c$ (the other implication has already been proved in (a)). For all $\varepsilon > 0$,

$$\begin{aligned} P(|X_n - c| > \varepsilon) &= 1 - F_{X_n}(c+\varepsilon) + F_{X_n}(c-\varepsilon) \\ &\to 1 - F_X(c+\varepsilon) + F_X(c-\varepsilon) \\ &= 0 \end{aligned}$$

because $F_X(\cdot)$ is a continuous function at $c+\varepsilon$ and $c-\varepsilon$,
(in fact, we have that $F_X(c+\varepsilon) = 1$ and $F_X(c-\varepsilon) = 0$). $\square$

Theorem 21 (for vectors)**.**

(**a**) $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$.

(**b**) $X_n \xrightarrow{d} c \Leftrightarrow X_n \xrightarrow{P} c$.

Proof. We shall begin with part (a).

$$\begin{aligned} X_n \xrightarrow{P} X &\Rightarrow \lambda' X_n \xrightarrow{P} \lambda' X, \ \forall\lambda \qquad \text{(Theorem 1)} \\ &\Rightarrow \lambda' X_n \xrightarrow{d} \lambda' X, \ \forall\lambda \qquad \text{(Theorem 20)} \\ &\Rightarrow X_n \xrightarrow{d} X \qquad \text{(Theorem 18).} \end{aligned}$$

Part (b) follows because

$$\begin{aligned} X_n \xrightarrow{d} c &\Rightarrow \lambda' X_n \xrightarrow{d} \lambda' c, \ \forall\lambda \qquad \text{(Theorem 18)} \\ &\Rightarrow \lambda' X_n \xrightarrow{P} \lambda' c, \ \forall\lambda \qquad \text{(Theorem 20)} \end{aligned}$$

Hence $X_n \xrightarrow{P} c$ by taking $\lambda = [1, 0, ..., 0]$ or $[0, 1, 0, ..., 0]$ and so on. □

Theorem 22. *Let X_n, X be $k \times 1$ vectors. Let g be a continuous function in the* ***domain*** *of X. Then,*

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

Proof. Because $e^{iu'g(x)}$ is continuous and bounded in x for each u, Theorem 17 implies that

$$\begin{aligned} X_n \xrightarrow{d} X &\Rightarrow E[e^{iu'g(X_n)}] \to E[e^{iu'g(X)}] \\ &\Leftrightarrow \phi_{g(X_n)}(u) \to \phi_{g(X)}(u) \\ &\Leftrightarrow g(X_n) \xrightarrow{d} g(X), \end{aligned}$$

which concludes the proof. □

Remark 11. *The previous theorem is known as the* ***Continuous Mapping Theorem****.*

Example 22. $X_n \xrightarrow{d} X \sim \mathcal{N}_k(0, I_k) \Rightarrow X_n' X_n \xrightarrow{d} \mathcal{X}_k^2$.

Theorem 23. *Suppose that* $Y_n \xrightarrow{d} Y$, $Z_n \xrightarrow{d} c$ *and* $g(\circ)$ *is a continuous function in* $(y, z)'$. *Then,* $g(Y_n, Z_n) \xrightarrow{d} g(Y, c)$.

Proof. Apply Theorems 21 and 22 to the vector $(y_n, z_n)'$. $\square$

Corollary 3. (1) $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{P} c \Rightarrow X_n + Y_n \xrightarrow{d} X + c$.
(2) $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{P} c \Rightarrow Y_n X_n \xrightarrow{d} cX$.
(3) $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{P} c \neq 0 \Rightarrow Y_n^{-1} X_n \xrightarrow{d} c^{-1} X$.

Example 23. $Y = Z\beta + U$ *with stochastic* Z*'s. If*

$$\widehat{M}_n = \frac{Z'Z}{n} \xrightarrow{P} M > 0$$

$$n^{1/2}\widehat{m}_n = \frac{Z'U}{n^{1/2}} \xrightarrow{d} X \sim \mathcal{N}_k\left(0, \sigma^2 M\right).$$

Then, we have that

$$\begin{aligned} n^{1/2}\left(\widehat{\beta}_n - \beta\right) &= \left(\frac{Z'Z}{n}\right)^{-1} \frac{Z'U}{n^{1/2}} = \widehat{M}_n^{-1} n^{1/2} \widehat{m}_n \\ &\xrightarrow{d} \mathcal{N}_k\left(0, \sigma^2 M^{-1}\right). \end{aligned}$$

Example 24 (Hypothesis testing in multiple regression with stochastic z_i').

Assuming $Y = Z\beta + U$, consider the null hypothesis $H_0 : W\beta = w$ for given known $q \times k$ -matrix W and $q \times 1$ -vector w. Further assume that rank $(W) = q \leq k$. Consider the F-test:

$$F_n = \frac{n\left(w - W\widehat{\beta}_n\right)'\left(W\widehat{M}_n^{-1}W'\right)^{-1}\left(w - W\widehat{\beta}_n\right)}{q\widehat{\sigma}^2}$$

$$\widehat{\sigma}^2 = \frac{\widehat{U}'\widehat{U}}{n-k},\ \widehat{U} = Y - Z\widehat{\beta}_n,$$

where $\widehat{\beta}_n$ denotes the LSE.

If (*) $u_i \sim iid\ \mathcal{N}(0, \sigma^2)$ and z_i are fixed, then $F_n \sim F_{q,n-k}$ under H_0, (which is a finite sample exact distribution).

Here we do not assume (*), so we shall obtain a large sample approximation to the finite sample distribution under milder conditions.

Assuming (these can be checked under more primitive conditions):

$$\begin{aligned}
(i)\ n^{1/2}\widehat{m}_n &= \frac{Z'U}{n^{1/2}} \xrightarrow{d} X \sim \mathcal{N}_k\left(0, \sigma^2 M\right) \\
(ii)\ \widehat{M}_n &= \frac{Z'Z}{n} \xrightarrow{p} M > 0 \\
(iii)\ s_n^2 &= \frac{U'U}{n} \xrightarrow{p} \sigma^2.
\end{aligned}$$

Recall $n^{1/2}\left(\widehat{\beta}_n - \beta\right) = \widehat{M}_n^{-1}n^{1/2}\widehat{m}_n$ and

$$\begin{aligned}
\widehat{U}'\widehat{U} &= -(\widehat{\beta}_n - \beta)'Z'Z(\widehat{\beta}_n - \beta) + U'U \\
&= -n^{1/2}\widehat{m}_n'\widehat{M}_n^{-1}n^{1/2}\widehat{m}_n + ns_n^2.
\end{aligned}$$

Now under $H_0 : W\beta = w$, we have:

$$\begin{aligned}
F_n &= \frac{n^{1/2}\left(\widehat{\beta}_n - \beta\right)'W'\left(W\widehat{M}_n^{-1}W'\right)^{-1}Wn^{1/2}\left(\widehat{\beta}_n - \beta\right)}{q\frac{\widehat{U}'\widehat{U}}{n-k}} \\
&= \frac{n^{1/2}\widehat{m}_n'\widehat{M}_n^{-1\prime}W'\left(W\widehat{M}_n^{-1}W'\right)^{-1}W\widehat{M}_n^{-1}n^{1/2}\widehat{m}_n}{q\left[\left(\frac{n}{n-k}\right)s_n^2 - \frac{1}{n-k}n^{1/2}\widehat{m}_n'\widehat{M}_n^{-1}n^{1/2}\widehat{m}_n\right]}.
\end{aligned}$$

That is, $F_n =: g\left(n^{1/2}\widehat{m}_n, \widehat{M}_n, s_n^2\right)$, where $g(\circ)$ is a continuous function.

So by Theorem 22, and under Assumptions (i), (ii) and (iii):

$$F_n \xrightarrow{d} g(X, M, \sigma^2),$$

so that,

$$F_n \xrightarrow{d} \frac{X'M^{-1}W'\left(WM^{-1}W'\right)^{-1}WM^{-1}X}{q\sigma^2}.$$

The latter is true because

$$\operatorname*{plim}_{n\to\infty}\left(\frac{n}{n-k}\right)s_n^2 = \sigma^2$$

$$X'M^{-1}X = O_p(1) \text{ since } X \sim \mathcal{N}_k\left(0, \sigma^2 M\right) \text{ and } M > 0$$

and hence

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n-k}X'M^{-1}X = 0.$$

Finally since $X \sim \mathcal{N}_k\left(0, \sigma^2 M\right)$ and $rank\left(W\right) = q \leq k$, we have:

$$\frac{WM^{-1}X}{\sigma} \sim \mathcal{N}_q\left(0, WM^{-1}W'\right)$$

$$\left(WM^{-1}W'\right)^{-1/2}\frac{WM^{-1}X}{\sigma} \sim \mathcal{N}_q\left(0, I_q\right).$$

Hence, we conclude that

$$(\dotplus)\quad F_n \xrightarrow{d} \frac{\left[\left(WM^{-1}W'\right)^{-1/2}\frac{WM^{-1}X}{\sigma}\right]'\left[\left(WM^{-1}W'\right)^{-1/2}\frac{WM^{-1}X}{\sigma}\right]}{q} =: \frac{\mathcal{X}_q^2}{q}.$$

Remark 12. *The $F_{q,n-k}$ distribution approaches the $\mathcal{X}_q^2/q$ distribution as $n \to \infty$, but this fact is NOT a proof of $(\dotplus)$.*

We now give the relationship among the different types or modes of statistical convergence.

The relationship among the types of convergence is as follows:

$$\begin{array}{ccccc} a.s. & \Rightarrow & PROB \Rightarrow DIST & & \\ & & \Uparrow & & \\ \circ & \mapsto & r-th\ mean & \Leftarrow & \left\{ \begin{array}{c} \text{If the limit is a constant, e.g.} \\ \text{it is degenerate r.v.} \end{array} \right. \end{array}$$

We have already mentioned that

$$PROBAB. \nRightarrow a.s.$$

and also that

$$PROBAB \nRightarrow rth - mean,$$

the reason for the latter being that for convergence in *rth-mean* we need that the rth-moments of both X_n and X to be finite, e.g.

$$E\left|X_n\right|^r < \infty \text{ and } E\left|X\right|^r < \infty,$$

whereas for convergence in *Probability*, we have made no reference to the existence of moments of the r.v. X_n nor of X at all.

Theorem 24. *(Relationship between* $\xrightarrow{d}$ *and* $O_p(1)$*)*

$$X_n \xrightarrow{d} X \Rightarrow X_n = O_p(1).$$

Proof. For any $\varepsilon > 0$, we can choose points of continuity $x > 0$ of $F_X(\cdot)$ such that $\Pr\{|X| > x\} < \varepsilon$. Then, choose $n_0 > 0$ such that $\forall n \geq n_0$,

$$|F_{X_n}(x) - F_X(x)| < \varepsilon \ \text{ and } |F_{X_n}(-x) - F_X(-x)| < \varepsilon.$$

Now,

$$\begin{aligned}
\Pr\{|X_n| > x\} &= \Pr\{X_n > x\} + \Pr\{X_n < -x\} \\
&= \Pr\{X_n > x\} - \Pr\{X > x\} \\
&\quad + \Pr\{X > x\} + \Pr\{X < -x\} \\
&\quad + \Pr\{X_n < -x\} - \Pr\{X < -x\} \\
&= \Pr\{|X| > x\} \\
&\quad + \Pr\{X_n > x\} - \Pr\{X > x\} \\
&\quad + \Pr\{X_n < -x\} - \Pr\{X < -x\}
\end{aligned}$$

$$\begin{aligned}
= \quad & \Pr\{|X| > x\} + (1 - \Pr\{X_n \leq x\}) \\
& - (1 - \Pr\{X \leq x\}) \\
& + \Pr\{X_n < -x\} - \Pr\{X < -x\} \\
\leq \quad & \varepsilon - (F_{X_n}(x) - F_X(x)) + (F_{X_n}(-x) - F_X(-x)) \\
\leq \quad & 3\varepsilon.
\end{aligned}$$

This concludes the proof. □

Theorem 25.

$$X_n \xrightarrow{d} X, \quad Y_n \xrightarrow{P} 0 \Rightarrow X_n Y_n \xrightarrow{P} 0.$$

Proof.

$$X_n \xrightarrow{d} X \Rightarrow X_n = O_p(1) \qquad \text{(Theorem 24)}$$

$$Y_n \xrightarrow{P} 0 \Rightarrow Y_n = o_p(1).$$

Therefore, by Theorem 16 part (3), we have that $X_n Y_n = O_p(1)\, o_p(1) = o_p(1)$. □

9. CENTRAL LIMIT THEOREMS

In Examples 18 and 22, we have assumed that certain sums of random variables were asymptotically normal.

The purpose of this section is to provide primitive conditions which will guarantee to obtain this asymptotic distribution.

To allow for sufficient generality, so that we can apply them to many important cases, we will consider a triangular array of random variables.

That is, our sequence of random variables is denoted as $\{X_{in}, i = 1, ..., n;\ n \geq 1\}$. That is, we have the following sequence

$$\begin{array}{l} X_{11} \\ X_{12}, X_{22} \\ X_{13}, X_{23}, X_{33} \\ \ldots\ldots\ldots\ldots \\ X_{1n}, X_{2n}, ..., X_{nn}. \end{array}$$

Example 25. *Consider the sequence* $\{X_i,\ i = 1, 2, ...\}$, *such that* $EX_i = \mu,\ \mathrm{Var}\,(X_i) = \sigma^2$ *Then*

$$X_{in} = \left(\sigma^2 n\right)^{-1/2} (X_i - \mu)\,.$$

$$\frac{1}{\sqrt{n}} \sum x_i = \sum \boxed{\frac{1}{\sqrt{n}} x_i}$$

Let us introduce the following assumptions:

***Assumption 1*:** $EX_{in} = 0$, $\forall i, n$.

***Assumption 2*:** $\sum_{i=1}^{n} \operatorname{Var} X_{in} = 1, \quad \forall n$.

***Assumption 3*:** X_{in} and X_{jn} are independent $\forall i \neq j \; \forall n$.

That is, there is independence **within** rows, but not necessarily **between** rows

Theorem 26. *(*Lindeberg-Feller's CLT for triangular arrays*) Under Assumptions 1 to 3, it follows that*

$$\sum_{i=1}^{n} X_{in} \xrightarrow{d} X \sim \mathcal{N}(0,1)$$

if and only if

***Assumption 4*:**

$$\sum_{i=1}^{n} E\left\{X_{in}^2 \mathcal{I}\left(|X_{in}| > \varepsilon\right)\right\} \underset{n\to\infty}{\longrightarrow} 0 \quad \forall \varepsilon > 0. \tag{9.1}$$

Remark 13. *Assumption 4 given in* (9.1) *is known as the*

$LINDEBERG'S\ \ CONDITION.$

Example 26. *(*Liapunov's CLT*) If for some $\delta > 0$,*

Assumption 5: $\sum_{i=1}^{n} E\left|X_{in}\right|^{2+\delta} \to 0,$

which is known as Liapunov's condition, then (9.1) *holds true.*

Example 27. *We can relax even further the previous condition to $E\left|X_i\right|^{2+\delta} \leq c$. Take $\mu = 0$ for simplicity, so*

$$X_{in} = \left(\sigma^2 n\right)^{-1/2} X_i.$$

Assume that $\{X_i^2, i \geq 1\}$ is UI, then:

$$\begin{array}{lcl} \textit{LHS of (9.1)} & = & (\sigma^2 n)^{-1} \sum_{i=1}^{n} E\left\{X_i^2 \mathcal{I}\left(|X_i| \geq \varepsilon\sigma n^{1/2}\right)\right\} \\ & \leq & \sigma^{-2} \max\limits_{i\geq 1} E\left\{X_i^2 \mathcal{I}\left(|X_i| \geq \varepsilon\sigma n^{1/2}\right)\right\} \\ \textit{UI of } \{X_i^2\} & \to & 0 \quad \textit{as } n \to \infty, \forall\varepsilon > 0. \end{array}$$

Theorem 27. *(*Lindeberg-Levy's CLT for triangular arrays*) Let* $\{X_{in},\ i = 1, ..., n\}$ *be a sequence of identically distributed (across* i*) random variables although the distribution can vary across* n*. Assumption 4 in this case is*

$$\begin{aligned} LHS \text{ of } (9.1) &= nE\left\{X_{in}^2 \mathcal{I}1\left(|X_{in}| \geq \varepsilon\right)\right\} \\ &= E\left\{nX_{in}^2 \mathcal{I}\left(n^{1/2}|X_{in}| \geq n^{1/2}\varepsilon\right)\right\} \end{aligned}$$

Thus, a sufficient condition for (9.1)*, that is Assumption 4, is*

$$E\left\{nX_{in}^2\right\} \leq c < \infty, \forall n.$$

For instance when

$$X_{in} = \left(\sigma^2 n\right)^{-1/2}(X_i - \mu)$$

a sufficient condition is that $EX_i^2 < \infty$*.*

The proof of Theorem 26 is beyond the scope of this course. However, in the case of *iid* data (Example 19), as in the previous example, the proof becomes an exercise of Theorem 17.

Indeed, denote

$$S_n = \sum_{k=1}^{n} X_{kn}; \qquad X_{kn} = \left(\sigma^2 n\right)^{-1/2} X_k,$$

where X_k is an iid$(0, \sigma^2)$ sequence. Then,

Claim 1. $S_n \to_d \mathcal{N}(0,1)$.

Proof. The CF of S_n is

$$\begin{aligned} \phi_{S_n}(t) &= E\left(e^{itS_n}\right) = E\left(\prod_{k=1}^{n} e^{itX_{kn}}\right) \\ &= \prod_{k=1}^{n} E\left(e^{itX_{kn}}\right) = \left(E\left(e^{itX_{kn}}\right)\right)^n \\ &= \left(\phi_{X_1}\left(t/\left(\sigma n^{1/2}\right)\right)\right)^n, \end{aligned}$$

where $\phi_{X_1}(u) = E\left(e^{iuX_1}\right)$ denotes the CF of X_1.

Now, we know that X_1 has second moments, so that the CF is twice continuously differentiable around the origin. So,

$$\begin{aligned} \phi_{X_1}(u) &= 1 + E(X_1)\, iu + \frac{1}{2}(iu)^2 E\left(X_1^2\right) + o\left(u^2\right) \\ &= 1 - \frac{1}{2}u^2\sigma^2 + o\left(u^2\right) \end{aligned}$$

when $u \to 0$.

Hence,

$$\begin{aligned} \phi_{S_n}(t) &= \left(1 - \frac{1}{2}\frac{t^2}{n} + o\left(\frac{1}{n}\right)\right)^n \\ &\to e^{-\frac{1}{2}t^2}. \end{aligned}$$

But the latter is the characteristic function of a $\mathcal{N}(0,1)$. We now employ Theorem 17 to conclude. $\square$

Now we will illustrate how these and earlier results can be used to establish a CLT for a useful parameter estimate under comprehensive conditions.

In particular, we would like to study the asymptotic behaviour of the LSE of β in the multiple regression model,

$$\begin{aligned} y_i &= \beta' z_i + u_i \\ &= \beta_1 z_{1i} + \beta_2 z_{2i} + ... + \beta_k z_{ki} + u_i, \quad i = 1, ..., n. \end{aligned} \tag{9.2}$$

We know that the LSE of β is

$$\widehat{\beta}_n = \beta + \left(Z'Z\right)^{-1} Z'U$$

and the question is, what is its asymptotic distribution?

How to proceed?

First, which normalization to use?

Suppose that there exists a positive-definite diagonal $k \times k$ -matrix D such that

$$\left(D^{-1}\left(Z'Z\right)D^{-1}\right) \to R > 0 \text{ and finite,}$$

that is $\overline{\lambda}\left(R\right) < \infty$ and $\underline{\lambda}\left(R\right) > 0$.

This implies that the required normalization will be D.

Thus, we will examine the behaviour of

$$\begin{aligned} D\left(\widehat{\beta}_n - \beta\right) &= D\left(Z'Z\right)^{-1} Z'U \\ &= \left(D^{-1}\left(Z'Z\right)D^{-1}\right)^{-1} D^{-1}Z'U, \end{aligned}$$

since it appears that

$$D\left(\widehat{\beta}_n - \beta\right) = O_p\left(1\right).$$

Now, $D\left(\widehat{\beta}_n - \beta\right)$ can be written as

$$D\left(\widehat{\beta}_n - \beta\right) = \widehat{R}_n^{-1}\widehat{r}_n,$$

where

$$\widehat{R}_n = D^{-1}\left(Z'Z\right)D^{-1} \text{ and } \widehat{r}_n = D^{-1}Z'U.$$

Thus,

$$\begin{aligned} D\left(\widehat{\beta}_n - \beta\right) &= \sum_{i=1}^{n} \widehat{R}_n^{-1} D^{-1} z_i u_i \\ &= \sum_{i=1}^{n} \widetilde{w}_{in} u_i = \sum_{i=1}^{n} \widetilde{X}_{in} \end{aligned}$$

and hence we have been able to write $D\left(\widehat{\beta}_n - \beta\right)$ into the same framework of the *LINDEBERG-FELLER CLT*, i.e.,

$$\sum_{i=1}^{n} \widetilde{X}_{in},\ n = 1, 2, ..., (n \text{ sample size}).$$

Our next question is:

What type of assumptions do we need to impose on the model (9.2) to guarantee that Assumptions $1-4$ for the *LINDEBERG-FELLER's CLT* are satisfied?

In particular, we need to multiply $D\left(\widehat{\beta}_n-\beta\right)$, or $\lambda' D\left(\widehat{\beta}_n-\beta\right)$, by something in order to obtain the required "normalization" in Assumption 2.

Recall that by Theorem 18, we know:

$$\sum_{i=1}^{n}\widetilde{X}_{in}\xrightarrow{d}X\Leftrightarrow\sum_{i=1}^{n}\lambda'\widetilde{X}_{in}\xrightarrow{d}\lambda' X,\qquad\forall\lambda\neq0.$$

Hence, we will focus on the convergence of

$$\begin{aligned}\lambda' D\left(\widehat{\beta}_n-\beta\right) &= \sum_{i=1}^{n}\lambda'\widehat{R}_n^{-1}D^{-1}z_i u_i\\ &= \sum_{i=1}^{n}\lambda'\widetilde{w}_{in}u_i=\sum_{i=1}^{n}X_{in}.\end{aligned}$$

(1) $EX_{in} = 0$? trivially true if $\underline{Eu_i = 0}$ and $\underline{z_i \text{ non-stochastic}}$.
It looks like the previous two conditions are needed on model (9.2).

(2) $\sum_{i=1}^{n} \operatorname{Var}(X_{in}) = 1$, $\forall n = 1, 2, ...$ Suppose $Eu_i^2 = \sigma^2$, will this be enough?

We start by looking at EX_{in}^2, which by definition is

$$\begin{aligned} E\left(\lambda' \widetilde{w}_{in} u_i\right)^2 &= E\left(\lambda' \widetilde{w}_{in} u_i^2 \widetilde{w}_{in}' \lambda\right) \\ &= \lambda' \widetilde{w}_{in} \widetilde{w}_{in}' \lambda \sigma^2 \\ &= \sigma^2 \lambda' \widehat{R}_n^{-1} D^{-1} z_i z_i' D^{-1} \widehat{R}_n^{-1} \lambda \end{aligned}$$

Hence

$$\sum_{i=1}^{n} \operatorname{Var}(X_{in}) = \lambda' \widehat{R}_n^{-1} D^{-1} \left(\sum_{i=1}^{n} z_i z_i' \right) D^{-1} \widehat{R}_n^{-1} \lambda \sigma^2 = \lambda' \widehat{R}_n^{-1} \lambda \sigma^2,$$

which implies that the asymptotic variance will be

$$\lim_{n \to \infty} \lambda' \widehat{R}_n^{-1} \lambda \sigma^2$$

But, we want that $\sum_{i=1}^{n} \operatorname{Var}(X_{in}) = 1$, $\forall n = 1, 2, ...$ so, what shall we do?

(1) Suppose that we multiply $D\left(\widehat{\beta}_n-\beta\right)$ by $\widehat{R}_n^{1/2}/\left(\sigma\left\|\lambda\right\|\right)$. In this case, we obtain that

$$\lambda'\frac{\widehat{R}_n^{1/2}}{\sigma\left\|\lambda\right\|}D\left(\widehat{\beta}_n-\beta\right)=\sum_{i=1}^{n}\lambda'\frac{\widehat{R}_n^{-1/2}}{\sigma\left\|\lambda\right\|}D^{-1}z_iu_i,$$

which implies that

$$\begin{aligned}\lambda'\frac{\widehat{R}_n^{1/2}}{\sigma\left\|\lambda\right\|}D\left(\widehat{\beta}_n-\beta\right)&=\sum_{i=1}^{n}\frac{\lambda'\widehat{R}_n^{-1/2}}{\sigma\left\|\lambda\right\|}D^{-1}z_iu_i\\&=\sum_{i=1}^{n}X_{in}.\end{aligned}$$

Observe that this new definition of X_{in} still satisfies Assumption 1, that is that $EX_{in}=0$. Moreover, by construction, we have now that

$$\begin{aligned}\sum_{i=1}^{n}EX_{in}^2&=\frac{\lambda'\widehat{R}_n^{-1/2}D^{-1}\left(\sum_{i=1}^{n}z_iz_i'\right)D^{-1}\widehat{R}_n^{-1/2}\lambda\sigma^2}{\sigma^2\left\|\lambda\right\|^2}\\&=\frac{\lambda'\widehat{R}_n^{-1/2}\widehat{R}_n\widehat{R}_n^{-1/2}\lambda}{\left\|\lambda\right\|^2}\\&=\frac{\lambda'\lambda}{\left\|\lambda\right\|^2}=1,\ \forall n.\end{aligned}$$

From now on we will assume that $\left\|\lambda\right\|=1$, and hence Assumption 2 is now satisfied with this new definition of X_{in}.

(2) $\{X_{in}\}$ is independent across i? True if $\{u_i\}$ is independent given that the regressors z_i are assumed to be non-stochastic.

(3) Is the Lindeberg's condition satisfied? that is,

$$\sum_{i=1}^{n} E\left\{|X_{in}|^2 \mathcal{I}\left(|X_{in}| > \varepsilon\right)\right\} \to 0 \quad as \quad n \to \infty \quad \forall \varepsilon > 0?$$

Since

$$X_{in} = \frac{\lambda' \widehat{R}_n^{-1/2} D^{-1} z_i u_i}{\sigma} = w_{in} u_i,$$

we have that

$$\begin{aligned} \sum_{i=1}^{n} E\left\{|X_{in}|^2 \mathcal{I}\left(|X_{in}| > \varepsilon\right)\right\} &= \sum_{i=1}^{n} w_{in}^2 E\left[|u_i|^2 \mathcal{I}\left(w_{in}^2 |u_i|^2 > \varepsilon^2\right)\right] \\ &= \sum_{i=1}^{n} w_{in}^2 E\left[|u_i|^2 \mathcal{I}\left(|u_i|^2 > \varepsilon^2 / w_{in}^2\right)\right]. \end{aligned}$$

However, we already know that

$$\sum_{i=1}^{n} w_{in}^2 = \frac{1}{\sigma^2} < \infty,$$

which suggests that to show the Lindeberg's Assumption 4, that is (9.1), it suffices to show that

$$\max_{1 \le i \le n} E\left[|u_i|^2 1\left(|u_i|^2 > \varepsilon^2 / w_{in}^2\right)\right] \to 0. \tag{9.3}$$

But (9.3) looks similar to the definition of UI for the sequence $\{u_i^2\}$.

Indeed, because

$$\mathcal{I}\left(|u_i|^2 > \varepsilon^2/w_{in}^2\right) \leqslant \mathcal{I}\left(|u_i|^2 > \frac{\varepsilon^2}{\max\limits_{1\le i\le n} w_{in}^2}\right),$$

then if $\{u_i^2\}$ is UI and in addition

$$\max_{1\le i\le n} |w_{in}| \to 0, \tag{9.4}$$

we have that (9.1) will hold true.

But as $w_{in} = \sigma^{-1}\lambda'\widehat{R}_n^{-1/2}D^{-1}z_i$, it suffices to check what conditions we need to impose on $\{z_i\}$ to guarantee that $\max_{1\le i\le n}|w_{in}| \to 0$. Notice that

$$\begin{aligned}
\max_{1\le i\le n} w_{in}^2 &= \max_{1\le i\le n} \frac{\lambda'\widehat{R}_n^{-1/2}D^{-1}z_iz_i'D^{-1}\widehat{R}_n^{-1/2}\lambda}{\sigma^2} \\
&\leqslant \frac{1}{\sigma^2}\lambda'\lambda \max_{1\le i\le n}\left|\widehat{R}_n^{-1/2}D^{-1}z_iz_i'D^{-1}\widehat{R}_n^{-1/2}\right| \\
&= \frac{1}{\sigma^2}\max_{1\le i\le n}\left|\widehat{R}_n^{-1/2}D^{-1}z_iz_i'D^{-1}\widehat{R}_n^{-1/2}\right| \\
&\le \frac{1}{\sigma^2}\left\|\widehat{R}_n^{-1}\right\| \max_{1\le i\le n}\left\|D^{-1}z_i\right\|^2.
\end{aligned}$$

Recall we have assumed that $D = Diag[d_1, ..., d_k]$ is a positive finite matrix such that

$$\widehat{R}_n = D^{-1}\left(Z'Z\right)D^{-1} \to R > 0.$$

So, we have that $\left\|\widehat{R}_n^{-1}\right\| = O(1)$ and

$$D^{-1}z_i = \left(\frac{z_{1i}}{d_1}, ..., \frac{z_{ki}}{d_k}\right)'.$$

The latter implies that

$$\begin{aligned}
\max_{1\le i\le n} w_{in}^2 &\le \frac{1}{\sigma^2}\left\|\widehat{R}_n^{-1}\right\| \max_{1\le i\le n}\sum_{j=1}^{k}\left(\frac{z_{ji}}{d_j}\right)^2 \\
&= \frac{1}{\sigma^2}\left\|\widehat{R}_n^{-1}\right\| \sum_{j=1}^{k}\left(\frac{\max_{1\le i\le n}|z_{ji}|}{d_j}\right)^2 \\
&\to 0 \text{ if } \max_{1\le j\le k}\frac{\max_{1\le i\le n}|z_{ji}|}{d_j} \to 0.
\end{aligned}$$

Summing up, the assumptions that we need (or better say sufficient conditions) are:

(v) *$\{u_i\}$ is independent with $Eu_i = 0$ and $Eu_i^2 = \sigma^2$ for all $i \geq 1$, and $\{u_i^2\}$ is UI.*

(vi) *$\{z_i\}$ is deterministic and satisfies the* "Grenander's conditions", *i.e.,*

(vi1) *for all $j = 1, ..., k$,*

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \frac{|z_{ji}|}{d_j} \to 0, \text{ where } d_j = \left(\sum_{i=1}^{n} z_{ji}^2\right)^{1/2} \to \infty;$$

*and (**vi2**)*

$$\widehat{R}_n = D^{-1}\left(Z'Z\right)D^{-1} \to R > 0, \text{ where } D = Diag(d_1, ..., d_k).$$

Therefore,

$$\frac{\widehat{R}_n^{1/2}}{\sigma} D\left(\widehat{\beta}_n - \beta\right) \xrightarrow{d} \mathcal{N}_k\left(0, I_k\right).$$

But, because $\widehat{R}_n \to R$, we have that

$$\left(\frac{\widehat{R}_n^{1/2}}{\sigma}\right)^{-1} \to \left(\frac{R^{1/2}}{\sigma}\right)^{-1}$$

and hence

$$D\left(\widehat{\beta}_n - \beta\right) \xrightarrow{d} \mathcal{N}_k\left(0, \sigma^2 R^{-1}\right).$$

So, we have shown the following theorem.

Theorem 28. *Suppose model* (9.2) *satisfies Assumptions (*v*) and (*vi*). Then,*

$$D\left(\widehat{\beta}_n - \beta\right) \xrightarrow{d} \mathcal{N}_k\left(0, \sigma^2 R^{-1}\right).$$

We now provide a simple, but useful, theorem due to Cramér.

Theorem 29. (*Cramér*)

Let $\{X_n\}_{n\in\mathbb{Z}}$ be a sequence of $(k \times 1)$ vectors of rv's, and assume that $X_n = A_n Z_n$, such that (a) $A_n \xrightarrow{p} A$ (psd) and (b)

$$Z_n \xrightarrow{d} \mathcal{N}(\mu, \Sigma)$$

that is, $\Sigma^{-1/2}(Z_n - \mu) \xrightarrow{d} \mathcal{N}(0, I)$. Then,

$$A_n Z_n \xrightarrow{d} \mathcal{N}(A\mu, A\Sigma A').$$

Remark 14. (**a**) *Assumption (*v*) is satisfied if* $\{u_i\}$ *is iid.* $(0, \sigma^2)$, *since iid with finite second moments* $\Rightarrow \{u_i^2\}$ *is UI.*

(**b**) *Assumption (*vi*) is true if* $z_{ij} = i^p$ *for* $p > -1/2$, *since*

$$\begin{aligned} d_{jn}^2 &= \sum_{i=1}^{n} i^{2p} = n^{2p+1} \sum_{i=1}^{n} \left(\frac{i}{n}\right)^{2p} \left(\frac{1}{n}\right) \\ &\simeq n^{2p+1} \int_{1/n}^{1} x^{2p} dx = \frac{n^{2p+1}}{2p+1} \left\{ 1 - \left(\frac{1}{n}\right)^{2p+1} \right\} \\ &\simeq cn^{2p+1} \to \infty \end{aligned}$$

$$\max_{1 \le i \le n} \frac{|z_{ij}|}{d_{jn}} \simeq Cn^{-1/2} \to 0.$$

*However (*vi*) is not satisfied if* $z_{ij} = e^i$, *since*

$$d_{jn}^2 = \sum_{i=1}^{n} e^{2i} = e^2 \frac{e^{2n} - 1}{e^2 - 1} \simeq ce^{2n}$$

$$\max_{1 \le i \le n} \frac{|z_{ij}|}{d_{jn}} \simeq \frac{e^n}{ce^n} \nrightarrow 0.$$

Is this something to be expected?

We agree that the asymptotic distribution of the LSE is governed by the behaviour of

$$\sum_{i=1}^{n} \frac{e^i}{d_n} u_i$$

where

$$d_n^2 = \sum_{i=1}^{n} e^{2i} = e^2 \frac{e^{2n}-1}{e^2-1} \simeq c e^{2n}.$$

So except a multiplicative constant, it suffices to examine the behaviour of

$$X_n = \sum_{i=1}^{n} \frac{e^i}{e^n} u_i =: u_n + \sum_{i=1}^{n-1} \frac{e^i}{e^n} u_i.$$

Let u be a random variable independent of u_i but with the same DF of u_n.

Then

$$\widetilde{X}_n = u + \sum_{i=1}^{n-1} \frac{e^i}{e^n} u_i,$$

has the same DF as X_n, which implies that both sequences, X_n and $\widetilde{X}_n$, would converge to the same distribution, say X.

But clearly X **cannot** be a normal random variable unless u is, which is not the case.

9.1. CENTRAL LIMIT THEOREMS FOR TIME SERIES DATA.

This section will discuss how to show *CLT*'s for time series data.

We will start with a *Lemma* which plays a very important role, albeit being a typical trick in asymptotic theory.

Lemma 1. *(Bernstein) Let X_n be a sequence of random variables with zero mean such that for every $\varepsilon > 0, \zeta > 0, \eta > 0$ there exist two sequences $Y_n(\varepsilon)$ and $Z_n(\varepsilon)$ such that*

$$X_n = Y_n(\varepsilon) + Z_n(\varepsilon),$$

where $Y_n(\varepsilon) \xrightarrow{d} \mathcal{N}(0, \Omega(\varepsilon))$ and

$$\lim_{\varepsilon \to 0} \Omega(\varepsilon) = \Omega, \quad \Pr\{|Z_n(\varepsilon)| > \zeta\} < \eta.$$

Then

$$X_n \xrightarrow{d} \mathcal{N}(0, \Omega).$$

Proof. We will not give the proof, but you can see it, in a more general framework, in Anderson's (1971, *p*.425) Theorem 7.7.1., The Statistical Analysis of Time Series. John Wiley. □

Theorem 30. *Let x_t be a sequence of random variables such that*

$$x_t = \sum_{j=0}^{m} a_j \varepsilon_{t-j}, \quad t = 1, ..., n$$

and ε_t is a sequence of iid $(0,1)$. Then,

$$\frac{1}{n^{1/2}} \sum_{t=1}^{n} x_t \xrightarrow{d} \mathcal{N}(0, \Omega) \tag{9.5}$$

where $\Omega = \gamma_0 + 2\sum_{j=1}^{m} \gamma_j$, with $\gamma_j = \sum_{p=0}^{m-j} a_p a_{p+j}$, that is $2\pi f(0)$ where $f(0)$ is the spectral density of x_t at the zero frequency.

Proof. The proof is done by the so-called **blocking method**.

The idea comes from the fact that, for instance, x_t and x_{t-m-1} are independent or more generally that $z_t = (x_t, ..., x_{t-k+1})$ and $z_s = (x_s, ..., x_{s-k+1})$, with $t < s$, are independent if $|s - k - t| > m$.

That is, if the blocks are separated by more than m observations.

So, let $k > 2m$ be an integer and $n = kN$.

That is we assume, without loss of generality, that nN^{-1} is integer.

Define

$$z_\ell = \frac{1}{k^{1/2}} \sum_{j=1}^{k-m} x_{(\ell-1)k+j}, \quad \ell = 1, ..., N$$

and

$$w_\ell = \frac{1}{k^{1/2}} \sum_{j=1}^{m} x_{\ell k-m+j}, \quad \ell = 1, ..., N.$$

Clearly

$$\frac{1}{n^{1/2}} \sum_{t=1}^{n} x_t = \frac{1}{N^{1/2}} \sum_{\ell=1}^{N} z_\ell + \frac{1}{N^{1/2}} \sum_{\ell=1}^{N} w_\ell. \tag{9.6}$$

Let k be a fixed integer. Then,

$$E\left(\frac{1}{N^{1/2}}\sum_{\ell=1}^{N} w_\ell\right)^2 = \frac{1}{N}\sum_{\ell=1}^{N} Ew_\ell^2 = Ew_1^2$$

because w_ℓ is a sequence of independent random variables and stationarity. Observe that $k > 2m$, so that the distance between the blocks is more than m periods. But, since x_t is a linear process with finite second moments, as ε_t does, we know that

$$w_1 = O_p\left((m/k)^{1/2}\right).$$

So, we conclude that $\left|N^{-1/2}\sum_{\ell=1}^{N} w_\ell\right| = O_p\left(m^{1/2}k^{-1/2}\right)$.

Next, we examine the behaviour of the first term on the right of (9.6). As we argued with the second term on the right of (9.6), the variables z_ℓ are independent with mean zero and variance $Ez_1 = \Omega(k)$.

So, *Lindeberg-Lévy's CLT* yields that

$$\frac{1}{N^{1/2}}\sum_{\ell=1}^{N} z_\ell \xrightarrow{d} \mathcal{N}(0, \Omega(k)).$$

But, now if $k \to \infty$, we have that $\left|N^{-1/2}\sum_{\ell=1}^{N} w_\ell\right| = o_p(1)$, whereas $\Omega(k) \to \Sigma$. Now Lemma 1 implies that (9.5) holds true. □

$\sum z_{\ell N}$ $\quad z_{\ell N} = \dfrac{z_\ell}{N^{1/2}}$

Theorem 31. *Let x_t be a sequence of random variables such that*

$$x_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} |a_j| < L < \infty$$

$a_j = \frac{1}{j \lg j}$

and ε_t is a sequence of iid $(0,1)$. Then,

$$\frac{1}{n^{1/2}} \sum_{t=1}^{n} x_t \xrightarrow{d} \mathcal{N}(0, \Omega)\,. \tag{9.7}$$

Proof. The idea in this case is to decompose x_t into two sequences, one of them behaving as an $MA(k)$, say, for which we already know a *CLT Theorem* and the remaining sequence, which for large k should be small in probability.

So, let

$$x_t' = \sum_{j=0}^{k} a_j \varepsilon_{t-j}, \; x_t'' = \sum_{j=k+1}^{\infty} a_j \varepsilon_{t-j}.$$

$\sum_{k0}^{\infty} a_j^2 -$

Because $\sum_{j=0}^{\infty} |a_j| < L$ we should expect that x_t'' is "small" since $\sum_{j=k+1}^{\infty} |a_j| = O\left(\log^{-1} k\right)$. In fact, $E\left|x_t''\right| = O\left(\log^{-1} k\right)$. Now

$E|\,| \le \sum_{k} |a_j| \, E|\varepsilon_{-j}|$

$$\begin{aligned}
E\left(\frac{1}{n^{1/2}} \sum_{t=1}^{n} x_t''\right)^2 &= \frac{1}{n} \sum_{t,s=1}^{n} E\left(x_t'' x_s''\right) \\
&= \frac{1}{n} \sum_{t,s=1}^{n} \sum_{j,r=k+1}^{\infty} a_j a_r E\left(\varepsilon_{t-j} \varepsilon_{s-r}\right) \\
&= \frac{\sigma^2}{n} \sum_{t,s=1}^{n} \sum_{j=\max(0,t-s)+k+1} |a_j| \, |a_{j+s-t}| \\
&\le \sigma^2 \sum_{q=k+1}^{\infty} \sum_{j=k+1}^{\infty} |a_j| \, |a_q| \\
&\le \sigma^2 \left(\sum_{j=k+1}^{\infty} |a_j|\right)^2 = O\left(\log^{-2} k\right).
\end{aligned}$$

$\le C \sum_{j=k+1}^{\infty} |a_j|$

$\le C \sum_{k} \frac{1}{j \lg j}$

$O\left(\lg(k)^{-1}\right)$

Now, let's examine x_t'. This is a $MA(k)$ process, so Theorem 30 yields that

$$\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t' \xrightarrow{d} \mathcal{N}\left(0, \Omega\left(k\right)\right),$$

where $\Omega\left(k\right) = \gamma_0\left(k\right) + 2\sum_{j=1}^{k}\gamma_j\left(k\right)$, with $\gamma_j\left(k\right) = \sum_{p=0}^{k-j} a_p a_{p+j}$. Now, let use Lemma 1 again to conclude that (9.7) holds true. First, as $k \to \infty$ it is clear that $n^{-1/2}\sum_{t=1}^{n} x_t'' \xrightarrow{P} 0$ by *Markov's inequality.* On the other hand, summability of a_j implies that

$$\lim_{k\to\infty}\Omega\left(k\right) = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j$$

where $\gamma_j = \lim_{k\to\infty}\gamma_j\left(k\right)$. This concludes the proof. $\square$

The next theorem is heuristic (the proof) and details are missing. However, the main ingredients will be given.

Theorem 32. *Let x_t be a sequence of random variables such that*

$$|E(g(x_1,...,x_a)\,g(x_{a+b},...,x_{2a+b-1})) - Eg(x_1,...,x_a)\,Eg(x_{a+b},...,x_{2a+b-1})|$$

$$\leq Cb^{-\alpha} \tag{9.8}$$

for some $\alpha > 0$ and any function $g(\circ)$ such that its second moments is finite. Then,

$$\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t \xrightarrow{d} \mathcal{N}(0,\Omega) \tag{9.9}$$

where $\Omega = 2\pi f(0)$.

Proof. As we have mentioned, we are not going to give a complete and formal proof of the theorem. The details are rather complicated. However, I will give the main ideas of the proof. First, as in Theorems 30 or 31, let k be an integer such that $n = Nk$, being N also an integer. Define

$$z_\ell = \frac{1}{k^{1/2}}\sum_{j=1}^{k-m} x_{(\ell-1)k+j}, \quad \ell = 1,...,N$$

and

$$w_\ell = \frac{1}{k^{1/2}}\sum_{j=1}^{m} x_{\ell k-m+j}, \quad \ell = 1,...,N.$$

Clearly

$$\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t = \frac{1}{N^{1/2}}\sum_{\ell=1}^{N} z_\ell + \frac{1}{N^{1/2}}\sum_{\ell=1}^{N} w_\ell. \tag{9.10}$$

Let k and m be now fixed but large, with the condition that $k^{-1}m \to 0$ when we let k and m to increase to infinity. Clearly, the former conditions on k and m imply that

$$E\left(\frac{1}{N^{1/2}}\sum_{\ell=1}^{N} w_\ell\right)^2 = o(1)$$

so that $N^{-1/2}\sum_{\ell=1}^{N} w_\ell = o_p(1)$, and Lemma 1 implies that the asymptotic distribution of $n^{-1/2}\sum_{t=1}^{n} x_t$ will be that of the first term on the right of (9.10).

Next, we examine the properties of z_ℓ, $\ell = 1, ..., N$. The idea is to replace z_ℓ by $\widetilde{z}_\ell$ with the same distribution as z_ℓ, but they are independent. Now, we know that

$$\frac{1}{N^{1/2}} \sum_{\ell=1}^{N} \widetilde{z}_\ell \xrightarrow{d} \mathcal{N}(0, \Omega)$$

by *Lindeberg-Lévy's CLT*, so that (9.9) holds true. So, it remains to show that we can do the latter.

To that end, we are going to make use of (9.8). Observing that z_ℓ and $z_{\ell+1}$ are separated by m periods, then (9.8) implies that

$$\left| E\left(e^{itz_\ell} e^{itz_{\ell+1}}\right) - E\left(e^{itz_\ell}\right) E\left(e^{itz_{\ell+1}}\right) \right| \leq Cm^{-\alpha}.$$

So, applying this we have that

$$\left| E\left(e^{it\frac{1}{N^{1/2}}\sum_{\ell=1}^{N} z_\ell}\right) - E\left(e^{it\frac{1}{N^{1/2}}\sum_{\ell=1}^{N-1} z_\ell}\right) E\left(e^{it\frac{1}{N^{1/2}} z_N}\right) \right| \leq Cm^{-\alpha}$$

Repeating this step several times, we obtain that

$$\left| E\left(e^{it\frac{1}{N^{1/2}}\sum_{\ell=1}^{N} z_\ell}\right) - \prod_{\ell=1}^{N} E\left(e^{it\frac{1}{N^{1/2}} z_\ell}\right) \right| \leq CNm^{-\alpha}.$$

Now, let $\widetilde{z}_\ell$ be independent but with the same probability distribution as z_ℓ, for $\ell = 1, ..., N$. Then, we have that

$$\left| E\left(e^{it\frac{1}{N^{1/2}}\sum_{\ell=1}^{N} z_\ell}\right) - \prod_{\ell=1}^{N} E\left(e^{it\frac{1}{N^{1/2}} \widetilde{z}_\ell}\right) \right| \leq CNm^{-\alpha}.$$

However, we have assumed that $Nm^{-\alpha} \to 0$, and thus together with the last displayed inequality we have that

$$\frac{1}{N^{1/2}} \sum_{\ell=1}^{N} \widetilde{z}_\ell \overset{d}{\simeq} \frac{1}{N^{1/2}} \sum_{\ell=1}^{N} z_\ell.$$

So, this concludes the proof. $\square$

We shall give an extension of Theorem 28 to time series. For that purpose, let's assume that

*(*vii*)* $u_i = \sum_{j=0}^{\infty} A_j e_{i-j}$, $\sum_{j=0}^{\infty} |A_j| < \infty$, $\{e_i\}$ *is independent with* $Ee_i = 0$*,* $Ee_i^2 = \sigma^2$*, and* $\{e_i^2, i \geq 1\}$ *is UI.*

*(*viii*) for all* j, there exists a matrix D such that

$$\lim_{n\to\infty} D^{-1} \sum_{i=1}^{n-j} z_i z_{i+j}' D^{-1} = R(j),$$
$$\text{with } R(0) = R.$$

Theorem 33. *(Anderson's Th 10.2.11) If the model* (9.2) *satisfies Assumptions (vi)-(viii), then*

$$D\left(\widehat{\beta}_n - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, R^{-1}\left(\sum_{j=-\infty}^{\infty} \gamma_j R(j)\right) R^{-1}\right)$$

where

$$\gamma_j = Eu_i u_{i+j} = \sigma^2 \sum_{\ell=0}^{\infty} A_\ell A_{\ell+j}, \quad j \geq 0$$
$$= \gamma_{-j}, \quad j < 0.$$

Remark 15. (1) *(vii) can be modified to allow for martingale differences* $\{e_i, i \geq 1\}$.

(2) *Many alternative "weak dependence" conditions to (vii) can be used.*

Example 28. *Consider the linear regression model*

$$
\begin{aligned}
y_t &= \beta x_t + u_t \\
u_t &= \rho \varepsilon_{t-1} + \varepsilon_t
\end{aligned}
$$

where the sequence $\{x_t\}_{t=1}^n$ *has finite eighth moments and satisfies* (9.8) *in Theorem 32 and* $\{\varepsilon_t\}_{t=1}^n$ *are iid random variables with* $E\varepsilon_t = 0$, $\sigma_\varepsilon^2 = E\varepsilon_t^2$ *and* $E\varepsilon_t^4 < \infty$. *Assume that the sequences* $\{x_t\}_{t=1}^n$ *and* $\{\varepsilon_t\}_{t=1}^n$ *are mutually independent. Then*

(a) The LSE of β is consistent and its asymptotic distribution is $\mathcal{N}(0,1)$ after appropriate normalization.

(b) Obtain the asymptotic distribution of the unfeasible GLS estimator of β, when

$$u_t = \rho u_{t-1} + \varepsilon_t.$$

Part (a). By definition of LSE, we have that

$$\widehat{\beta}_n - \beta = \left(\frac{1}{n} \sum_{t=1}^{n} x_t^2 \right)^{-1} \frac{1}{n} \sum_{t=1}^{n} x_t u_t. \tag{9.11}$$

Now, because the x_t have finite second moments we shall show first that $n^{-1} \sum_{t=1}^{n} x_t^2$ converges in $2nd$ mean to $Ex_t^2 =: \zeta^2 > 0$. Indeed, by definition

$$\begin{aligned} E\left(\frac{1}{n} \sum_{t=1}^{n} \left(x_t^2 - \zeta^2 \right) \right)^2 &= \frac{1}{n^2} \sum_{t,s=1}^{n} E\left\{ \left(x_t^2 - \zeta^2 \right) \left(x_s^2 - \zeta^2 \right) \right\} \\ &= \frac{1}{n} \sum_{t=1-n}^{n-1} \left(1 - \frac{|t|}{n} \right) E\left\{ \left(x_t^2 - \zeta^2 \right) \left(x_0^2 - \zeta^2 \right) \right\} \\ &\leq \frac{C}{n} \sum_{t=1-n}^{n-1} \left(1 - \frac{|t|}{n} \right) |t|_+^{-\alpha} \qquad (9.12) \\ &= o(1), \end{aligned}$$

where $|a|_+ = \max(1, |a|)$ and using (9.8). So, Slutzky's theorem implies that

$$\left(\frac{1}{n} \sum_{t=1}^{n} x_t^2 \right)^{-1} \xrightarrow{P} \zeta^{-2}$$

and hence by Cramér's theorem, it suffices to examine the behaviour of the second term on the right of (9.11).

That is the asymptotic limit of the sequence

$$\frac{1}{n^{1/2}} \sum_{t=1}^{n} x_t u_t.$$

To that end, we first notice that the last display expression can be written as

$$\frac{1}{n^{1/2}}\sum_{t=1}^{n-1} x_t^* \varepsilon_t + \frac{1}{n^{1/2}} x_n \varepsilon_n, \tag{9.13}$$

where $x_t^* = x_t + \rho x_{t+1}$.

Because the second term of (9.13) converges to zero in probability, we need to examine the first term.

To that end, we shall look first at the behaviour of

$$X_n =: \sum_{t=1}^{n-1} \widetilde{x}_{tn}^* \varepsilon_t,$$

where $\widetilde{x}_{tn}^* =: x_t^* / \left(\sum_{t=1}^{n-1} x_t^{*2}\right)^{1/2}$. Conditionally on x_t, we have that $\widetilde{x}_{tn}^* \varepsilon_t$ is a sequence of independent random variables with finite fourth moments, as $E\varepsilon_t^4 < \infty$. So, we shall employ Lyapunov's CLT.

Theorem 34. (Lyapunov) *Let* $\{x_i\}_{i\in\mathbb{N}}$ *be an independent sequence of random variables with zero mean and variance* σ_i^2*. Suppose that for some* $\upsilon > 2$ *and* $\forall\delta > 0$ *arbitrary small*

$$\frac{1}{B_n^\upsilon}\sum_{i=1}^{n} E\left|x_i\right|^\upsilon < \delta,$$

where

$$B_n^2 = \sum_{i=1}^{n}\sigma_i^2;\, Var\left(x_i\right) = \sigma_i^2.$$

Then,

$$\frac{1}{B_n}\sum_{i=1}^{n} x_i \xrightarrow{d} \mathcal{N}\left(0,1\right).$$

Now,

$$X_n =: \sum_{t=1}^{n-1} \widetilde{x}^*_{tn} \varepsilon_t =: \sum_{t=1}^{n-1} w_{tn}$$

satisfies that

$$\sum_{t=1}^{n-1} Ew^2_{tn} = \left(E\varepsilon^2_t\right) \sum_{t=1}^{n-1} \widetilde{x}^{*2}_{tn} =: B^2_n.$$

Notice that $B^2_n =: E\varepsilon^2_t$.

Next,

$$\begin{aligned} \frac{1}{B^4_n} \sum_{t=1}^{n-1} Ew^4_{tn} &= \frac{1}{\left(E\varepsilon^2_t\right)^2} \sum_{t=1}^{n-1} x^{*4}_t \frac{E\varepsilon^4_t}{\left(\sum_{t=1}^{n-1} x^{*2}_t\right)^2} \\ &= \frac{E\varepsilon^4_t}{\left(E\varepsilon^2_t\right)^2 \left(\sum_{t=1}^{n-1} x^{*2}_t\right)^2} \sum_{t=1}^{n-1} x^{*4}_t \\ &= \frac{E\varepsilon^4_t}{\left(E\varepsilon^2_t\right)^2 n} \frac{\frac{1}{n} \sum_{t=1}^{n-1} x^{*4}_t}{\left(n^{-1} \sum_{t=1}^{n-1} x^{*2}_t\right)^2} \\ &< \delta \end{aligned}$$

for some n large enough.

So, Lyapunov's theorem yields that

$$X_n =: \sum_{t=1}^{n-1} w_{tn} \xrightarrow{d} \mathcal{N}\left(0, E\varepsilon^2_t\right)$$

because

$$\sum_{t=1}^{n-1} Ew^2_{tn} = E\varepsilon^2_t.$$

Now, we notice that

$$\begin{aligned}
\widehat{\beta}_n - \beta &= \left(\frac{1}{n}\sum_{t=1}^{n} x_t^2\right)^{-1} \frac{1}{n}\sum_{t=1}^{n} x_t u_t \\
&= \left(\frac{1}{n}\sum_{t=1}^{n} x_t^2\right)^{-1} \left\{\frac{1}{n}\sum_{t=1}^{n-1} x_t^* \varepsilon_t + \frac{1}{n} x_n \varepsilon_n\right\}
\end{aligned}$$

$$X_n =: \frac{1}{\left(\sum_{t=1}^{n-1} x_t^{*2}\right)^{1/2}} \sum_{t=1}^{n-1} x_t^* \varepsilon_t.$$

So, we conclude that, conditional on x_t,

$$\begin{aligned}
&\left(\sum_{t=1}^{n-1} x_t^2\right)\left(\sum_{t=1}^{n-1} x_t^{*2}\right)^{-1/2}\left(\widehat{\beta}_n - \beta\right) \\
&= \left(\sum_{t=1}^{n-1} x_t^{*2}\right)^{-1/2} \sum_{t=1}^{n-1} x_t u_t \\
&= \left(\sum_{t=1}^{n-1} x_t^{*2}\right)^{-1/2} \left\{\sum_{t=1}^{n-1} x_t^* \varepsilon_t + x_n \varepsilon_n\right\}.
\end{aligned}$$

Now by (9.13) and using (9.12), the left side of the last displayed expression is

$$\begin{aligned}
\left(\sum_{t=1}^{n-1} x_t^{*2}\right)^{-1/2} \sum_{t=1}^{n-1} x_t^* \varepsilon_t + o_p(1) &= : X_n + o_p(1) \\
&\xrightarrow{d} \mathcal{N}\left(0, E\varepsilon_t^2\right).
\end{aligned}$$

But this is the case also for the unconditional distribution as the limit is independent of x_t, so is the unconditional one, that is

$$\left(\sum_{t=1}^{n-1} x_t^2\right)\left(\sum_{t=1}^{n-1} x_t^{*2}\right)^{-1/2}\left(\widehat{\beta}_n - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, E\varepsilon_t^2\right).$$

Part (b). It is well known that the infeasible $GLSE$ is the LSE on the transformed model, which in our case it becomes the LSE in the regression model

$$y_t - \rho y_{t-1} = \beta \left(x_t - \rho x_{t-1} \right) + \varepsilon_t,$$

$$\begin{aligned} \widehat{\beta}_n - \beta &= \left(\frac{1}{n} \sum_{t=2}^{n} x_t^{*2} \right)^{-1} \frac{1}{n} \sum_{t=2}^{n} x_t^{*} \varepsilon_t, \\ x_t^{*} &= x_t - \rho x_{t-1}. \end{aligned}$$

To show the asymptotic normality of the estimator, we shall proceed similarly as in part (a). We first examine the behaviour of

$$\begin{aligned} \frac{1}{n} \sum_{t=2}^{n} x_t^{*2} &= \frac{1}{n} \sum_{t=2}^{n} x_t^2 + \frac{\rho^2}{n} \sum_{t=2}^{n} x_{t-1}^2 - \frac{2\rho}{n} \sum_{t=2}^{n} x_t x_{t-1} \\ \xrightarrow{P} \left(1 + \rho^2 \right) E x_t^2 + 2\rho E \left(x_t x_{t-1} \right) &= : \zeta^2 > 0 \end{aligned}$$

because x_t is a sequence satisfying (9.8) with finite four moments, so that for instance

$$E \left(\frac{1}{n} \sum_{t=2}^{n} \left(x_t x_{t-1} - E \left(x_t x_{t-1} \right) \right) \right)^2 \to 0$$

which implies that, say,

$$\frac{1}{n} \sum_{t=2}^{n} x_t x_{t-1} \xrightarrow{P} E x_t x_{t-1}.$$

Next, the previous arguments imply that we only need to focus on the behaviour of the sequence $n^{-1}\sum_{t=2}^{n} x_t^* \varepsilon_t$, and in particular that of

$$\sum_{t=2}^{n} \frac{x_t^*}{\left(\sum_{t=2}^{n} x_t^{*2}\right)^{1/2}} \varepsilon_t =: \frac{1}{n^{1/2}} \sum_{t=2}^{n} w_{tn}.$$

Again we shall use Lyapunov's theorem. First, conditionally on x_t, we have that w_t is independent sequence and $Ew_t = 0$. On the other hand,

$$\begin{aligned}\sum_{t=2}^{n} Ew_{tn}^2 &= \sum_{t=2}^{n} \frac{x_t^{*2}}{\left(\sum_{t=2}^{n} x_t^{*2}\right)} E\varepsilon_t^2 \\ &= E\varepsilon_t^2.\end{aligned}$$

Recall that we are doing everything conditionally on x_t and that the distributional properties of ε_t are INDEPENDENT of those of x_t. Now, in our case $B_n =: E\varepsilon_t^2$.

Next,

$$\begin{aligned}
\frac{1}{B_n^4}\sum_{t=2}^{n} Ew_{tn}^4 &= \frac{1}{\left(E\varepsilon_t^2\right)^2}\sum_{t=2}^{n} x_t^{*4}\frac{E\varepsilon_t^4}{\left(\sum_{t=2}^{n} x_t^{*2}\right)^2} \\
&= \frac{E\varepsilon_t^4}{\left(E\varepsilon_t^2\right)^2 n}\frac{n^{-1}\sum_{t=2}^{n} x_t^{*4}}{\left(n^{-1}\sum_{t=2}^{n} x_t^{*2}\right)^2} \\
&< \delta
\end{aligned}$$

for n large enough. So, we have that

$$\sum_{t=2}^{n} w_{tn} \xrightarrow{d} \mathcal{N}\left(0, E\varepsilon_t^2\right)$$

because

$$\sum_{t=2}^{n} Ew_{tn}^2 = E\varepsilon_t^2.$$

From here then we have that

$$\begin{aligned}
\left(\frac{1}{n}\sum_{t=2}^{n} x_t^{*2}\right)^{1/2}\left(\widehat{\beta}_n - \beta\right) &= \sum_{t=2}^{n}\frac{x_t^{*}}{\left(\sum_{t=2}^{n} x_t^{*2}\right)^{1/2}}\varepsilon_t \\
&\xrightarrow{d} \mathcal{N}\left(0, E\varepsilon_t^2\right).
\end{aligned}$$

10. CONSISTENCY OF EXTREMUM ESTIMATORS

Given a sample $X = (x_1, ..., x_n)$, suppose that we estimate the unknown parameter $\theta \in \Theta \subset \mathbb{R}^p$, by using the distance generated by the loss function $Q_n(\theta; X)$, and which we shall abbreviate as $Q_n(\theta)$. Then, our estimator of θ_0, the true value of the parameters, will be given by that value which minimizes $Q_n(\theta)$. That is

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta). \tag{10.1}$$

Example 29. *Nonlinear least squares regression. Let $f_i(\theta)$ be given, scalar, nonlinear functions of θ, varying with i.*

$y_i = f_i(\theta_0) + u_i$

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_i(\theta))^2.$$

$\frac{1}{n}\sum \rho(z_i;\theta)$

For example we can have a regression model where

$$f_i(\theta) = \alpha + \beta e^{\gamma' z_i},$$

with $\theta = (\alpha, \beta, \gamma')'$.

When $\widehat{\theta}_n$ is explicitly defined, we can study its properties directly. However, in our present context, $\widehat{\theta}_n$ is only implicitly defined, making much harder to examine its statistical properties.

$Q_n(\theta) - Q_n(\theta_0) = S(\theta) - T_n(\theta)$

$\frac{1}{n}\sum_i (u_i + (f_i(\theta_0) - f_i(\theta)))^2 - \frac{1}{n}\sum u_i^2$

$\cancel{\frac{1}{n}\sum u_i^2} + \frac{1}{n}\sum (f_i(\theta) - f_i(\theta_0))^2 - \frac{2}{n}\sum u_i f_i(\ )$

$- \cancel{\frac{1}{n}\sum u_i^2}$

$\frac{1}{n}\sum_i (f_i(\theta;\theta_0))^2 - \frac{2}{n}\sum u_i f_i(\theta;\theta_0)$

$\frac{1}{n}\sum\{f_i^2(\theta;\theta_0) - E(f_i^2(\theta,\theta_0)) - 2u_i f_i(\theta,\theta_0)\}$

$+ \frac{1}{n}\sum E(f_i^2(\theta,\theta_0))$

$y - f(x_i,\theta_0)$

$E(f_i^2(\theta,\theta_0)) \to S(\theta)$

$$\frac{1}{n}\sum_i g(\underbrace{y_i, x_i}_{z_i}, \theta) \xrightarrow{P} 0$$

Let us introduce the following conditions:

A: The parameter space Θ is a compact subset of $\mathbb{R}^p$.
B: $\theta_0 \in \Theta$.
C:

$$Q_n(\theta) - Q_n(\theta_0) = S(\theta) - T_n(\theta)$$

such that $S(\theta)$ is deterministic and constant over n, and for all $\varepsilon > 0$, there exists $\eta > 0$ such that

$$\inf_{\|\theta - \theta_0\| \geq \varepsilon} S(\theta) > \eta$$

and $T_n(\theta)$ satisfies

$$\sup_{\theta \in \Theta} |T_n(\theta)| \xrightarrow{P} 0.$$

Theorem 35. *Assuming Conditions A, B and C then*

$$\widehat{\theta}_n \xrightarrow{p} \theta_0.$$

Proof. For $\varepsilon > 0$, let

$$\mathcal{N} = \{\theta : \|\theta - \theta_0\| < \varepsilon\},$$

with $\mathcal{N}^c$ denoting its complimentary set, that is $\Theta - \mathcal{N}$. Write

$$Q = Q_n(\theta), \qquad Q_0 = Q_n(\theta_0).$$

Then,

$$\begin{aligned} \left\|\widehat{\theta}_n - \theta_0\right\| &\geq \varepsilon \Leftrightarrow \widehat{\theta}_n \in \mathcal{N}^c \\ &\Rightarrow \inf_{\mathcal{N}^c} Q \leq \inf_{\mathcal{N}} Q \\ &\Rightarrow \inf_{\mathcal{N}^c} Q \leq Q_0 \end{aligned}$$

by Condition B. Now,

$$\begin{aligned} \Pr\left\{\left\|\widehat{\theta}_n - \theta_0\right\| \geq \varepsilon\right\} &\leq \Pr\left\{\inf_{\mathcal{N}^c}(Q - Q_0) \leq 0\right\} \\ &= \Pr\left\{\inf_{\mathcal{N}^c}(S(\theta) - T_n(\theta)) \leq 0\right\} \end{aligned}$$

$$\begin{aligned} &\leq \Pr\left\{\inf_{\mathcal{N}^c} S(\theta) - \sup_{\theta \in \Theta} |T_n(\theta)| \leq 0\right\} \\ &\leq \Pr\left\{\sup_{\theta \in \Theta} |T_n(\theta)| > \eta\right\} \to 0 \end{aligned}$$

by Condition C.

Observe that the previous inequality comes from the fact that

$$\begin{aligned} \inf(S - T_n) &= \inf\{S - \sup|T_n| + (\sup|T_n| - T_n)\} \\ &\geq \inf\{S - \sup|T_n|\} \\ &= \inf S - \sup|T_n|. \end{aligned}$$

□

What we now should do is try to give more primitive conditions on z_t, and the function $Q_n(\cdot)$ under which Conditions A, B and C hold true. Conditions A and B are easy to check, so we do not need to give any further conditions. The only condition which is a bit more problematic is Condition C. Therefore, we are going to give a set of conditions that guarantees Condition C. This will be done in the form of a Lemma.

Lemma 2. *Let $g(z;\theta)$ a function on z and θ where $\theta \in \Theta$. Assuming that*

(i) Θ is a compact set in $\mathbb{R}^p$.
(ii) $g(z;\theta)$ is continuous in θ for all z.
(iii) $E(g(z;\theta)) = 0$.
(iv) $z_1, ..., z_n$ are iid such that

$$E\left(\sup_{\theta\in\Theta} |g(z_t;\theta)|\right) < \infty$$

that is, $\sup_{\theta\in\Theta} |g(z;\theta)| \leq L(z)$ with finite expectation.
Then,

$$\frac{1}{n}\sum_{t=1}^{n} g(z_t;\theta) \xrightarrow{p} 0 \qquad \textit{uniformly in } \theta.$$

Proof. We are going to use very much assumptions (i) and (ii), and the fact that those two assumptions imply that $g(z;\theta)$ is uniformly continuous in θ. That Θ is a compact set implies that $\exists$ a partition of Θ, say $\Theta^1, ..., \Theta^q$ such that

$$\Theta = \bigcup_{i=1}^{q} \Theta^i \text{ and } \Theta^i \cap \Theta^j = \emptyset \text{ for all } i \neq j.$$

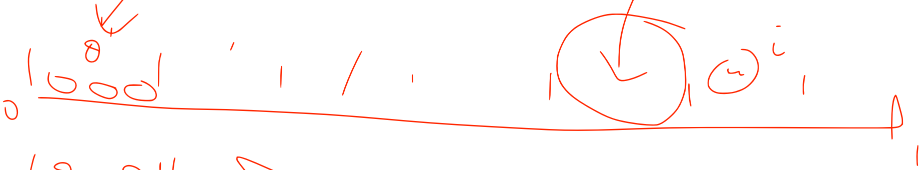
Also we can choose the partition such that $\forall\, \theta^1$ and $\theta^2 \in \Theta^i$, $i = 1, ..., q$,

(10.2) $$\left|\theta^1 - \theta^2\right| < \delta$$

for any arbitrarily small $\delta > 0$.

$q = \frac{1}{\|\theta - \theta^i\|}$

$\theta,\ \theta^2$



$q\,\delta = 1$

$|\theta - \theta^i| < \delta \to 0$

$q \cdot \delta = 1$

$q = 1/\delta$

Let $\theta^1, ..., \theta^q$ be a sequence of $\theta' s$ such that $\theta^i \in \Theta^i$. We need to show that for any arbitrary $\varepsilon > 0$,

$$\Pr\left\{\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n} g\left(z_t;\theta\right)\right| > \varepsilon\right\} \underset{n\to\infty}{\to} 0.$$

But the left side of the last displayed expression is bounded by

$$\begin{aligned} (10.3) \qquad & \Pr\left\{\cup_{i=1}^{q}\left\{\sup_{\theta\in\Theta^i}\left|\frac{1}{n}\sum_{t=1}^{n} g\left(z_t;\theta\right)\right| > \varepsilon\right\}\right\} \\ & \leq \sum_{i=1}^{q}\Pr\left\{\sup_{\theta\in\Theta^i}\left|\frac{1}{n}\sum_{t=1}^{n} g\left(z_t;\theta\right)\right| > \varepsilon\right\} \end{aligned}$$

by (2.1). Now, adding and subtracting $g\left(z_t;\theta^i\right)$ we have that, by the triangle inequality,

$$\begin{aligned} \text{sup} \quad & \left|\frac{1}{n}\sum_{t=1}^{n} g\left(z_t;\theta\right)\right| \\ & \leq \left|\frac{1}{n}\sum_{t=1}^{n} g\left(z_t;\theta^i\right)\right| + \left|\frac{1}{n}\sum_{t=1}^{n}\left(g\left(z_t;\theta\right) - g\left(z_t;\theta^i\right)\right)\right|. \end{aligned}$$

So, using the last displayed inequality, we have that the right side of (10.3) is bounded by

$\delta \to 0 \qquad \exists\, q \to \infty$

$$\sum_{i=1}^{q} \Pr\left\{\left|\frac{1}{n}\sum_{t=1}^{n} g\left(z_t;\theta^i\right)\right| > \varepsilon/2\right\} + \sum_{i=1}^{q} \Pr\left\{\sup_{\theta\in\Theta^i}\left|\frac{1}{n}\sum_{t=1}^{n}\left(g\left(z_t;\theta\right) - g\left(z_t;\theta^i\right)\right)\right| > \varepsilon/2\right\}. \tag{10.4}$$

From here, the proof is completed if we show that both terms of (10.4) converge to zero.

We begin showing that the first term of (10.4) converges to zero. But this is the case since it is equal to

$$\sum_{i=1}^{q} \Pr\left\{\left|\frac{1}{n}\sum_{t=1}^{n} g\left(z_t;\theta^i\right)\right| > \varepsilon/2\right\}$$

and since z_t are *iid* it implies that $g\left(z_t;\theta^i\right)$ also is an *iid* sequence of random variables with finite first moments. So, by Khinchine's (or Kolmogorov) theorem, see Theorem 11, see also Example 13 that follows, since $Eg\left(z_t;\theta^i\right) = 0$ implies that the last expression converges to zero.

To finish the proof, it suffices to show that the second term of (10.4) converges also to zero.

To that end, denoting $h_t\left(\theta,\theta^i\right) =: g\left(z_t;\theta\right) - g\left(z_t;\theta^i\right)$,

$$\sup_{\theta\in\Theta^i}\left|\frac{1}{n}\sum_{t=1}^{n} h_t\left(\theta,\theta^i\right)\right| \leq \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right|.$$

Now for all z, by continuity of $g\left(z;\theta\right)$ and (10.2) we have that $\sup_{\theta\in\Theta^i}\left|g\left(z;\theta\right) - g\left(z;\theta^i\right)\right| \to 0$, whereas by (iii)

$$\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right| \leq 2\sup_{\theta\in\Theta}\left|g\left(z_t;\theta\right)\right| \leq 2L\left(z\right).$$

So, the dominated convergence theorem implies that for any $\varepsilon > 0$ we can choose q large enough such that

$$E\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right| \leq \varepsilon/4$$

which implies that

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right| \leq \frac{1}{n}\sum_{t=1}^{n}\left\{\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right| - E\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right|\right\} + \frac{\varepsilon}{4}.$$

So, the last displayed expression yields that

$$\Pr\left\{\sup_{\theta\in\Theta^i}\left|\frac{1}{n}\sum_{t=1}^{n}h_t\left(\theta,\theta^i\right)\right|>\varepsilon/2\right\}$$

$\varepsilon/4$

$$\leq\quad\Pr\left\{\frac{1}{n}\sum_{t=1}^{n}\left\{\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right|-E\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right|\right\}>\varepsilon/4\right\}$$

because recall if $A\subset B$ then $\Pr(A)\leq\Pr(B)$.

But z_t are iid, so then it is $\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right|$, which yields that

$$\frac{1}{n}\sum_{t=1}^{n}\left\{\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right|-E\sup_{\theta\in\Theta^i}\left|h_t\left(\theta,\theta^i\right)\right|\right\}\xrightarrow{P}0.$$

From here that the second term of (10.4) converges to zero is standard and hence the proof of the lemma.

$\varepsilon/2 \leq \left| \boxed{\quad - E\sup} + E\sup \right| \leq \left| \quad - \quad \right| + E|| \quad | \leq \varepsilon_n$

□

Theorem 36. *Suppose that $g_n(\theta)$ converges in probability uniformly in $\theta \in \Theta$ to a nonstochastic function $g(\theta)$ in an open neighbourhood $\mathcal{N}(\theta_0)$ of θ_0. Then*

$$p\lim g_n\left(\widehat{\theta}_n\right) = g(\theta_0)$$

if $\widehat{\theta}_n - \theta_0 \xrightarrow{P} 0$ and $g(\theta)$ is continuous at θ_0.

Proof. First by the uniform convergence of $g_n(\theta)$ to $g(\theta)$ in $\mathcal{N}(\theta_0)$ and that with high probability $\widehat{\theta}_n \in \mathcal{N}(\theta_0)$, by $\widehat{\theta}_n - \theta_0 \xrightarrow{P} 0$, we have that there exists n_1 such that for $n > n_1$,

$$\Pr\left\{\left|g_n\left(\widehat{\theta}_n\right) - g\left(\widehat{\theta}_n\right)\right| > \delta/2\right\} < \varepsilon/2. \tag{10.5}$$

Moreover, by continuity of $g(\cdot)$ at θ_0, by Slutzky's theorem, see Theorem 1, we obtain that there exists n_2 such that for $n > n_2$,

$$\Pr\left\{\left|g\left(\widehat{\theta}_n\right) - g(\theta_0)\right| > \delta/2\right\} < \varepsilon/2, \tag{10.6}$$

since $\widehat{\theta}_n - \theta_0 \xrightarrow{P} 0$. So, there exists $n_3 > \max\{n_1, n_2\}$ such that for $n > n_3$,

$$\begin{aligned}
\Pr\left\{\left|g_n\left(\widehat{\theta}_n\right) - g(\theta_0)\right| > \delta\right\} &\leq \Pr\left\{\left|g_n\left(\widehat{\theta}_n\right) - g\left(\widehat{\theta}_n\right)\right| > \delta/2\right\} \\
&\quad + \Pr\left\{\left|g\left(\widehat{\theta}_n\right) - g(\theta_0)\right| > \delta/2\right\} \\
&< \varepsilon
\end{aligned}$$

by (10.5) and (10.6). This concludes the proof. □

$$\to \frac{1}{n}\sum (y_i - f_i(\theta))^2$$

CLT

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = -\left(\frac{1}{n}\sum \frac{\partial^2}{\partial\theta^2} f_i(\tilde{\theta})\right)^{-1} \frac{1}{\sqrt{n}}\sum u_i f_i(\theta_0)$$

$$0 = \frac{\partial}{\partial\theta}\cancel{Q_n(\hat{\theta})} = \frac{2}{n}\sum u_i(\hat{\theta}_n) = 0$$

$$= \frac{\partial}{\partial\theta} Q_n(\theta_0) + (\hat{\theta}_n - \theta_0)\frac{\partial^2}{\partial\theta^2} Q_n(\tilde{\theta}_n)$$

$$\cancel{\frac{1}{n}\sum \partial f_i^2} \qquad \boxed{\frac{1}{n}\sum\left(\frac{\partial}{\partial\theta} f_i(\tilde{\theta}_n)\right)^2} \xrightarrow{P} \; ? > 0$$

$\tilde{\theta}_n \xrightarrow{p} \theta_0$ $\qquad \frac{1}{n}\sum \frac{\partial}{\partial\theta} f_i(\theta_0) \longrightarrow$

$g(\tilde{\theta}_n) \longrightarrow g(\theta_0)$

$g_n(\tilde{\theta}_n)$

10.0.1. *CENTRAL LIMIT THEOREM FOR EXTREMUM ESTIMATORS.*

We finish this section given a Central Limit Theorem for extreme estimators. That is, let $\widehat{\theta}_n$ be such that

$$\widehat{\theta}_n = \underset{\theta\in\Theta}{\arg\min}\, Q_n\left(\theta; \underset{\sim}{x}\right) = \underset{\theta\in\Theta}{\arg\min}\, Q_n(\theta).$$

iid $\qquad \sqrt{n}\,\theta_n \geq 0$

$\hat{\theta}_n \geq 0$

To that end, we impose the following (not the weakest) conditions

(1) θ_0 is an interior point of Θ, a compact subset in $\mathbb{R}^p$.

(2)

$$Q_n(\theta) \xrightarrow{p} Q(\theta)$$

uniformly in $\theta \in \Theta$ where $Q(\theta)$ is a nonstochastic function which attains a unique minimum at $\theta = \theta_0$.

(3) $Q(\theta)$ is a continuous function in $\theta \in \Theta$.

(4) $Q_n\left(\theta; \underset{\sim}{x}\right)$ is twice continuously differentiable with respect to θ in a neighbourhood of θ_0.

(5)

$$\frac{\partial^2}{\partial\theta\partial\theta'} Q_n\left(\theta; \underset{\sim}{x}\right)\Big|_{\theta_n^*} \xrightarrow{p} A(\theta_0) = \lim_{n\to\infty} E \frac{\partial^2}{\partial\theta\partial\theta'} Q_n\left(\theta_0; \underset{\sim}{x}\right) > 0,$$

for every sequence $\theta_n^* - \theta_0 = o_p(1)$.

(6)

$$n^{1/2} \frac{\partial}{\partial\theta} Q_n\left(\theta_0; \underset{\sim}{x}\right) \xrightarrow{d} \mathcal{N}\left(0, B(\theta_0)\right).$$

Then, we have the following theorem.

$Q_n(\theta) - Q(\theta_0)$

$Q(\theta) - Q(\theta_0)$

$S(\theta) = 0$

$\frac{1}{n}\sum \ell_i(\theta)$

$\hat{\theta}_n \quad \frac{1}{n}\sum x_i(y_i - \hat{\beta}_i)$

$\frac{1}{n}\sum \frac{\partial}{\partial\theta}\ell_i(\theta_0) - g(z_i) \qquad \frac{1}{n}\sum (g(z_i)) \longrightarrow N(\cdot,\cdot)$

$\hat{\beta}_n \to \frac{1}{n}\sum x_i(y_i - \hat{\beta}_n) = 0$

$\frac{1}{n}\sum \ell_i = x_i(y_i - \beta_0 x_i) \neq 0$

$Q_n(\theta) = \sum (y_i - \theta x_i)^2$

$Q_n(\theta) = \frac{1}{n}\sum \ell_i(\theta) \qquad g(z_i, \theta)$

$= \frac{1}{n} \sum \underbrace{(y_i - f_i(\theta))^2}_{\rho(z_i, \theta)}$

Theorem 37. *Under the previous conditions*

$$n^{1/2}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, A^{-1}(\theta_0)\, B(\theta_0)\, A^{-1}(\theta_0)'\right).$$

Proof. The proof follows by a simple use of Taylor's expansion. First, under our assumptions we know that $\widehat{\theta}_n - \theta_0 = o_p(1)$. So, because θ_0 is an interior point, it implies that for n large enough $\widehat{\theta}_n$ is an interior point of Θ with probability approaching one. Thus, because $\widehat{\theta}_n$ is an extreme value and the first derivative of $Q_n(\theta)$ exists then

$\frac{\partial}{\partial \theta} Q_n(\theta_0) = 0$

$$\begin{aligned} \boxed{\frac{\partial}{\partial \theta} Q_n\left(\widehat{\theta}_n\right) = 0} \\ = \frac{\partial}{\partial \theta} Q_n(\theta_0) + \frac{\partial^2}{\partial \theta \partial \theta'} Q_n\left(\widetilde{\theta}_n\right)\left(\widehat{\theta}_n - \theta_0\right), \end{aligned}$$

where $\widetilde{\theta}_n$ is an intermediate point between $\widehat{\theta}_n$ and θ_0, and hence it satisfies that $\widetilde{\theta}_n - \theta_0 = o_p(1)$. So, from the previous equation, we have that

$$(10.7) \qquad n^{1/2}\left(\widehat{\theta}_n - \theta_0\right) = -\left(\frac{\partial^2}{\partial \theta \partial \theta'} Q_n\left(\widetilde{\theta}_n\right)\right)^{-1} n^{1/2} \frac{\partial}{\partial \theta} Q_n(\theta_0).$$

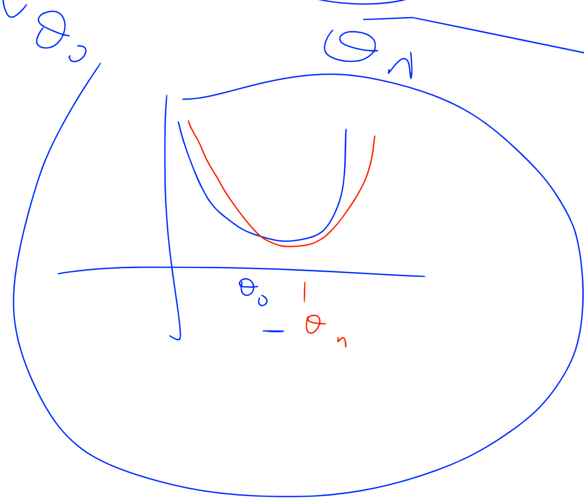
But we have assumed that the argument in the first factor on the right of (10.7) converges in probability to $A(\theta_0) > 0$. Then, we have that the first factor converges to $A^{-1}(\theta_0)$ in probability. On the other hand, the second factor on the right of (10.7) converges in distribution to $\mathcal{N}(0, B(\theta_0))$.

So, by Cramér Theorem (see Theorem 29), we conclude the proof of this theorem. □

$\hat{\Theta}_n \downarrow \theta_0$

$Q_n(\theta) \to \boxed{Q(\theta)}$

θ_0



$\sqrt{n}(\hat{\beta}_n - \beta)$

Normally, our loss function $Q_n(\theta)$ takes the simple form of being the average of some functions. That is

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} q_t(\theta),$$

where $q_t(\theta) = q(\theta; x_t)$. In these circumstances and based on our results given in Lemma 2 and Theorem 37, it is easy to check, under primitive conditions on the properties of x_t and q_t, whether the previous conditions hold or not. Moreover, regarding the CLT condition, we have already give conditions under which the average of some random variables satisfies or not the CLT. Observe that $q_t(\theta)$ are at θ_0.

Another point is that if $Q_n(\theta)$ corresponds to the *Maximum Likelihood*, then in that case we have that $B(\theta_0) = A(\theta_0)$. This is the case because remember that

$$\begin{aligned} Var\left(\frac{\partial}{\partial\theta}q_t(\theta)\right) &= E\left(\frac{\partial}{\partial\theta}q_t(\theta)\frac{\partial}{\partial\theta'}q_t(\theta)\right) \\ &= -E\left(\frac{\partial^2}{\partial\theta\partial\theta'}q_t(\theta)\right) \end{aligned}$$

and if we assume that x_t are iid, then

$$Var\left(\frac{1}{n^{1/2}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}q_t(\theta)\right) = E\left(\frac{\partial}{\partial\theta}q_t(\theta)\frac{\partial}{\partial\theta'}q_t(\theta)\right).$$

So, in this case we have as a corollary that The *Maximum Likelihood achieves the Cramér-Rao efficiency bound.*

$\sqrt{n}(\hat{\beta}_n - \underbrace{E[\hat{\beta}]}_{\beta}) \to$

$\sqrt{n}(\hat{\theta}_n - \theta_0) \; ? \to \infty$

$d: (\sqrt{n}(\hat{\theta}_n - E(\hat{\theta}_n))) + (\sqrt{n}(E(\hat{\theta}_n) - \theta_0))$

11. ASYMPTOTICS WHEN THE TRUE VALUE OF THE PARAMETER IS AT THE BOUNDARY

The purpose of this section is to present the basic ideas and results of extreme estimators when the true value, θ_0 say, may lie at the boundary of the compact parameter space $\Theta \subset \mathbb{R}^p$. We shall begin with the simplest of the models, see for instance Gourieroux, Holy and Monfort (1982) Econometrica paper. However, we shall proceed quite differently than them.

Suppose that we have the following linear regression model

$$y_i = \beta x_i + u_i; \quad i = 1, ..., n.$$

It is assumed that $\Theta \equiv [0, M]$, for some $M < \infty$. If $\{u_i\}_{i\in\mathbb{Z}}$ were a sequence of *iid* random variables, the obvious objective function would be

$$Q(\beta) = \sum_{i=1}^{n} (y_i - \beta x_i)^2,$$

although even if $\{u_i\}_{i\in\mathbb{Z}}$ were heteroscedastic and/or autocorrelated we may still employ $Q(\beta)$ as objective function to estimate β. Obviously, our estimator of β is given by

$$\widehat{\beta}_n = \arg\min_{\beta\in\Theta} Q(\beta), \tag{11.1}$$

and we know that under suitable regularity conditions, $\widehat{\beta}_n \to_P \beta_0$. In particular, we need that

$$\Pr\left\{\sup_{\beta\notin\mathcal{N}(\beta_0)} Q(\beta_0) - Q(\beta) > 0\right\} \to 0, \tag{11.2}$$

where $\mathcal{N}(\beta_0) = \{\beta : |\beta - \beta_0| \le \varepsilon\}$. That (11.2) holds true comes from the (obvious) observation that

$$Q(\beta_0) - Q(\beta) = -(\beta - \beta_0)^2 \frac{1}{n}\sum_{i=1}^{n} x_i^2 + 2(\beta_0 - \beta)\frac{1}{n}\sum_{i=1}^{n} x_i u_i$$
$$\xrightarrow{P} -(\beta - \beta_0)^2 E x_i^2 \tag{11.3}$$

uniformly in β and that $-(\beta - \beta_0)^2 E x_i^2 < 0$.

Next, we shall see the rate of convergence, that is the value of α for which (11.2) still true but where now $\mathcal{N}(\beta_0) = \{\beta : |\beta - \beta_0| \leq Ln^{-\alpha}\}$ for some $L > 0$. To that end and denoting $\beta_0 - \beta$ by $\widetilde{\beta}$, we need to show that, using the first equality in (11.3),

$$\Pr\left\{\sup_{\mathcal{A}_n(\varepsilon)} -\widetilde{\beta}^2 \frac{X'X}{n} + \frac{2\widetilde{\beta}}{n} X'U > 0\right\} < \eta \tag{11.4}$$

for some arbitrarily small η, and where $\mathcal{A}_n(\varepsilon)$ denotes the set $\varepsilon > |\beta - \beta_0| \geq Ln^{-1/2}$.

The left side of (11.4) is

$$\begin{aligned}
& \Pr\left\{\sup_{\mathcal{A}_n(\varepsilon)} -\widetilde{\beta}^2 + 2\widetilde{\beta}\left(\frac{X'X}{n}\right)^{-1}\frac{X'U}{n} > 0\right\} \\
\leq\ & \Pr\left\{\sup_{\mathcal{A}_n(\varepsilon)} 2\left(\frac{X'X}{n}\right)^{-1}\left|\frac{X'U}{n}\right| > \inf_{\mathcal{A}_n(\varepsilon)}\left|\widetilde{\beta}\right|\right\} \\
\leq\ & \Pr\left\{\sup_{\mathcal{A}_n(\varepsilon)} \left(\frac{X'X}{n}\right)^{-1}\left|\frac{X'U}{n^{1/2}}\right| > \inf_{\mathcal{A}_n(\varepsilon)} n^{1/2}\left|\widetilde{\beta}\right|\right\} \\
\leq\ & \Pr\left\{\sup_{\mathcal{A}_n(\varepsilon)} \left(\frac{X'X}{n}\right)^{-1}\left|\frac{X'U}{n^{1/2}}\right| > L\right\}
\end{aligned}$$

which is bounded by η choosing L large enough, because $n^{-1/2}\sum_{i=1}^{n} x_i u_i = O_p(1)$ and $n^{-1}\sum_{i=1}^{n} x_i^2$ converges in probability to $Ex_i^2 > 0$.

Next, the previous result indicates that our estimator of β_0 given in (11.1) can be written as

$$\widehat{\beta}_n = \beta_0 + \frac{\widehat{v}}{n^{1/2}} \tag{11.5}$$

where

$$\widehat{v} = \arg\min_{|v|<L} Q\left(\beta_0 + \frac{v}{n^{1/2}}\right).$$

Also, the previous arguments indicate that, when minimizing $Q(\beta)$, we only need to consider $\beta's$ of the type

$$\beta = \beta_0 + \frac{v}{n^{1/2}},$$

with $|v| < L \in (0, \infty)$.

Let us consider

$$n\left(Q\left(\beta_0\right)-Q\left(\beta\right)\right)=n\left(Q\left(\beta_0\right)-Q\left(\beta_0+\frac{v}{n^{1/2}}\right)\right).$$

From (11.3), we have that the right side of the last displayed equation is

$$-v^2\frac{1}{n}\sum_{i=1}^{n}x_i^2-2v\frac{1}{n^{1/2}}\sum_{i=1}^{n}x_iu_i:=\Lambda_n\left(v\right).$$

So, we can consider $\Lambda_n\left(v\right)$ as a process indexed by v, where v belongs to a compact set. Moreover, the process is continuous in v, that is $\Lambda_n\left(v\right)\in\mathbb{C}\left[-L,L\right]$. We shall now see where $\Lambda_n\left(v\right)$ converges.

To that end, we need to check two things. (a) The convergence of the finite-dimensional distributions and (b) that the process is tight. We shall begin with (a). It is clear that for any finite collection $v_1, ..., v_q$,

$$\begin{pmatrix} \Lambda_n(v_1) \\ . \\ . \\ . \\ \Lambda_n(v_q) \end{pmatrix} \xrightarrow{d} - \begin{pmatrix} v_1^2 \\ . \\ . \\ . \\ v_q^2 \end{pmatrix} \sigma_x^2 + 2\Omega^{1/2}\mathcal{N}\left(0, \sigma_u^2\sigma_x^2\right)$$

where $\sigma_x^2 = E\left(x_i^2\right)$ and the $(i,j)\,th$ element of Ω is $v_i v_j$. In particular if $q = 1$, we have that

$$\Lambda_n(v) \xrightarrow{d} -v^2\sigma_x^2 + 2v\sigma_u\sigma_x\mathcal{Z}$$

where $\mathcal{Z}$ denotes the standard normal random variable.

Next, we need to show tightness, that is (b). Consider $\overline{\Lambda}_n(v_2, v_1) = \Lambda_n(v_2) - \Lambda_n(v_1)$ for any $v_2 > v_1$. By Theorem 12.6 of Billingsley (1968), a sufficient condition for tightness is that

$$E\left|\overline{\Lambda}_n(v_2, v_1) - E\overline{\Lambda}_n(v_2, v_1)\right|^{\xi} < K\left(F(v_2) - F(v_1)\right)^{1+\delta} \tag{11.6}$$

for some $\xi > 0$, $\delta > 0$ and where $F(\cdot)$ is a monotonic nondecreasing and continuous function. By definition of $\Lambda_T(v)$,

$$\begin{aligned} & E\left|\overline{\Lambda}_n(v_2, v_1) - E\overline{\Lambda}_n(v_2, v_1)\right|^{\xi} \\ = \; & E\left|v_2 \frac{1}{n^{1/2}} \sum_{i=1}^{n} x_i u_i - v_1 \frac{1}{n^{1/2}} \sum_{i=1}^{n} x_i u_i\right|^{\xi} \\ \leq \; & (v_2 - v_1)^{\xi} E\left|\frac{1}{n^{1/2}} \sum_{i=1}^{n} x_i u_i\right|^{\xi} \\ \leq \; & D(v_2 - v_1)^{\xi}, \end{aligned}$$

which implies that (11.6) holds true choosing $\xi > 2$. It goes without saying that I am assuming that the ξth moment of $x_i u_i$ is finite.

So, we have shown that

$$\Lambda_n(v) \overset{weakly}{\Rightarrow} \Lambda(v) \equiv -v^2 \sigma_x^2 + 2v\sigma_u \sigma_x \mathcal{Z}.$$

Now, the limit process $\Lambda(v)$ is a parabola with fixed second derivatives in v. So, if this is the case, it turns out that the functional "arg max" is continuous, so that by the continuous mapping theorem we can conclude that

$$\widehat{v} = \arg\max_{|v|<L} \Lambda_n(v) \xrightarrow{d} v^* = \arg\max_{|v|<L} \Lambda(v). \tag{11.7}$$

But, by definition of $\Lambda_n(v)$, we have that

$$n^{1/2}\left(\widehat{\beta}_n - \beta_0\right) = \widehat{v}.$$

Now, what about v^*? Suppose that $\beta_0 > 0$. Then the true value is an interior point of the parameter space Θ. In this case, we have that all the β of the type $\beta = \beta_0 + n^{-1/2}v$ belong to Θ for all v, so that there is no constraints in v and thence

$$v^* = \sigma_u \sigma_x^{-1} \mathcal{Z}$$

is the maximum of $\Lambda(v)$, and we conclude that

$$n^{1/2}\left(\widehat{\beta}_n - \beta_0\right) \xrightarrow{d} \sigma_u \sigma_x^{-1} \mathcal{Z}.$$

This is the case when $\beta_0 > 0$. Now what about if $\beta_0 = 0$? It is obvious that we cannot expect the same type of asymptotic distribution because $\widehat{\beta}_n - \beta_0 \geq 0$, so that the only "admissible" values of v are ≥ 0. So,

$$v^* = \sigma_u \sigma_x^{-1} \mathcal{Z} \mathbb{I}(\mathcal{Z} \geq 0)$$

which implies that

$$n^{1/2}\left(\widehat{\beta} - \beta_0\right) \xrightarrow{d} \sigma_u \sigma_x^{-1} \mathcal{Z} \mathbb{I}(\mathcal{Z} \geq 0)$$

by (11.7).

11.1. **NONLINEAR MODELS.**

In this section we discuss what happens with nonlinear models such a nonlinear regression models. Consider the nonlinear regression model

$$y_i = f(x_i;\beta) + u_i; \quad i = 1, ..., n.$$

Again we assume that $\Theta \equiv [0, M]$, for some $M < \infty$. If $\{u_i\}_{i\in\mathbb{Z}}$ were a sequence of *iid* random variables, the obvious objective function would be

$$Q(\beta) = \sum_{i=1}^{n} (y_i - f_i(\beta))^2,$$

where we have abbreviated $f(x_i;\beta)$ by $f_i(\beta)$. Obviously, our estimator of β is given by

$$\widehat{\beta}_n = \arg\min_{\beta\in\Theta} \; Q(\beta). \tag{11.8}$$

We know that under suitable regularity conditions, $\widehat{\beta}_n \xrightarrow{p} \beta_0$. In particular we need that

$$\Pr\left\{ \sup_{\beta\notin\mathcal{N}(\beta_0)} Q(\beta_0) - Q(\beta) > 0 \right\} \to 0, \tag{11.9}$$

where $\mathcal{N}(\beta_0) = \left\{\beta : \left|\widetilde{\beta}\right| \leq \varepsilon\right\}$. That (11.9) holds true comes from the observation that

$$\begin{aligned} Q(\beta_0) - Q(\beta) &= -\widetilde{\beta}^2 \frac{1}{n}\sum_{i=1}^{n} \left(f_i'\left(\overline{\beta}\right)\right)^2 + 2\widetilde{\beta}\frac{1}{n}\sum_{i=1}^{n} f_i'\left(\overline{\beta}\right) u_i \\ &\xrightarrow{p} -\widetilde{\beta}^2 E\left(f_i'\left(\overline{\beta}\right)\right)^2 \end{aligned} \tag{11.10}$$

uniformly in β since as usual we can give several conditions under which

$$\frac{1}{n}\sum_{i=1}^{n} (f_i'(\beta))^2 \xrightarrow{P} E(f_i'(\beta))^2$$

uniformly in β.

Next, we shall see the rate of convergence, that is the value of α for which (11.9) still true but where now $\mathcal{N}(\beta_0) = \left\{\beta : \left|\widetilde{\beta}\right| \leq Ln^{-\alpha}\right\}$ for some $L > 0$. To that end, we need to show that, using the first equality in the last displayed expression,

$$\Pr\left\{\sup_{\mathcal{A}_n(\varepsilon)} -\frac{\widetilde{\beta}^2}{n}F'\left(\overline{\beta}\right)F\left(\overline{\beta}\right) + 2\frac{\widetilde{\beta}}{n}F'\left(\overline{\beta}\right)U > 0\right\} < \eta \tag{11.11}$$

for some arbitrarily small η, where $F\left(\overline{\beta}\right)$ is the matrix out of $f_i'\left(\overline{\beta}\right)$. The proof of (11.11) proceeds as that of (11.4) after we note that

$$\sup_{\beta\in\Theta}\left|\frac{1}{n^{1/2}}\sum_{i=1}^{n} f_i'(\beta)\,u_i\right| = O_p(1)\,. \tag{11.12}$$

So, we shall show (11.12). The way we are going to do it is as follows.

Denoting

$$\Upsilon_n(\beta) = \frac{1}{n^{1/2}}\sum_{i=1}^{n} f_i'(\beta)\,u_i,$$

we are going to show that

$$\Upsilon_n(\beta) \overset{weakly}{\Rightarrow} \Upsilon(\beta)\,,$$

actually to a Gaussian process.

As we did in the previous section, we need to show (a) the convergence of the finite dimensional distributions and (b) tightness condition.

The proof of (a) is trivial. This is not more than to show a CLT property of $\Upsilon_n\left(\beta_j\right)$ for $j=1,...,q$ with q finite.

Next (b). As we proceed above we shall show that

$$E\left|\Upsilon_n\left(\beta_2\right)-E\Upsilon_n\left(\beta_1\right)\right|^{\xi}<K\left(H\left(\beta_2\right)-H\left(\beta_1\right)\right)^{1+\delta} \tag{11.13}$$

for some $\xi>0$, $\delta>0$ and where $H(\cdot)$ is a monotonic nondecreasing and continuous function. Take $\xi=4$. Then the left side of (11.13) is

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^{n}E\left(f_i'\left(\beta_2\right)-f_i'\left(\beta_1\right)\right)^4 Eu_i^4\\
&+\frac{6}{n^2}\sum_{i<j}E\left\{\left(f_i'\left(\beta_2\right)-f_i'\left(\beta_1\right)\right)^2\left(f_j'\left(\beta_2\right)-f_j'\left(\beta_1\right)\right)^2\right\}\\
&\qquad\times Eu_i^2Eu_j^2.
\end{aligned}$$

Now assuming that $\left|f_i'\left(\beta_2\right)-f_i'\left(\beta_1\right)\right|\le\left|\beta_2-\beta_1\right|^{\varsigma}g_i\left(\beta\right)$ with $E\sup_{\beta\in\Theta}\left|g_i\left(\beta\right)\right|^4<K$, then it is easily shown that the last displayed expression is bounded by

$$\frac{K}{n}\left|\beta_2-\beta_1\right|^{4\varsigma}+K\left|\beta_2-\beta_1\right|^{4\varsigma}$$

which implies that if $\varsigma>\frac{1}{4}$, then (11.13) holds true with $\delta=4\varsigma-1$.

So, we have that

$$\Upsilon_n(\beta) \overset{weakly}{\Longrightarrow} \Upsilon(\beta)$$

and since the "sup" is a continuous function we have that (11.12) holds true. From here the proof proceeds as in the case of the linear regression model but replacing x_i by $f_i'(\beta)$. The only difference is to notice that

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} f_i' \left(\beta_0 + \tau \frac{v}{n^{1/2}} \right) u_i$$

where $\tau \in (0,1)$, satisfies that its difference with

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} f_i'(\beta_0) u_i,$$

that is,

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(f_i' \left(\beta_0 + \tau \frac{v}{n^{1/2}} \right) - f_i'(\beta_0) \right) u_i$$

satisfies

$$\sup_{|v|<L} \left| \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(f_i' \left(\beta_0 + \tau \frac{v}{n^{1/2}} \right) - f_i'(\beta_0) \right) u_i \right| = o_p(1).$$

The latter displayed equality implies that instead of

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} f_i' \left(\beta_0 + \tau \frac{v}{n^{1/2}} \right) u_i$$

we only need to consider $n^{-1/2} \sum_{i=1}^{n} f_i'(\beta_0) u_i$.