Basic Elements of Asymptotic Theory

Benedikt M. Pötscher[1] and Ingmar R. Prucha[2]

January 1999

Preliminary version; prepared for *Companion in Theoretical Econometrics.*

[1]Department of Statistics, University of Vienna, Universitätsstraße 5, A-1010 Vienna, Austria

[2]Department of Economics, University of Maryland, College Park, MD 20742, USA

Contents

1 Introduction

Consider the estimation problem where we would like to estimate a parameter vector θ from a sample $Y_1, \ldots, Y_n$. Let $\widehat{\theta}_n$ be an estimator for θ, i.e., let $\widehat{\theta}_n = h(Y_1, \ldots, Y_n)$ be a Borel measurable function of the sample. In the important special case where $\widehat{\theta}_n$ is a *linear* function of $Y_1, \ldots, Y_n$, i.e., $\widehat{\theta}_n = Ay$, where A is a nonrandom matrix and $y = (Y_1, \ldots, Y_n)'$, we can easily express the expected value and the variance-covariance matrix of $\widehat{\theta}_n$ in terms of the first and second moments of y (provided those moments exist). Also, if the sample is normally distributed, so is $\widehat{\theta}_n$. Well-known examples of linear estimators are the OLS- and the GLS-estimator of the classical linear regression model. Frequently, however, the estimator of interest will be a *nonlinear* function of the sample. In principle, the distribution of $\widehat{\theta}_n$ can then be found from the distribution of the sample by appealing to the transformation theorem for probability measures, if the model relating the parameter θ to the observables $Y_1, \ldots, Y_n$ fully specifies the distribution of the sample. E.g., in a linear regression model with independently and identically distributed errors this would require assuming a specific distribution for the errors. However, supposing the researcher feels comfortable with making such a specific assumption, for practical purposes it will then still often be impossible to obtain an exact expression for the distribution of $\widehat{\theta}_n$ because of the complexity of the necessary calculations. (Even if $\widehat{\theta}_n$ is linear, but the distribution of y is non-normal, it will typically be difficult to obtain the exact distribution of $\widehat{\theta}_n$.) Similarly, obtaining expressions for, say, the first and second moments of $\widehat{\theta}_n$ will, for practical purposes, typically be infeasible for nonlinear estimators; and even if it is feasible, the resulting expressions will usually depend on the entire distribution of the sample, and not only on the first and second moments as in the case of a linear estimator. A further complication arises in case the model relating θ to the observables $Y_1, \ldots, Y_n$ does not fully specify the distribution of $Y_1, \ldots, Y_n$. E.g., in a linear regression model the errors may only be assumed to be identically and independently distributed with zero mean and finite variance, without putting any further restrictions on the distribution function of the disturbances. In this case we obviously cannot get a handle on the distribution of $\widehat{\theta}_n$ (even if $\widehat{\theta}_n$ is linear), in the sense that this distribution will depend on the unknown distribution of the errors.

Given the above discussed difficulties in obtaining *exact* expressions for characteristics of estimators like their moments or distribution functions we will often have to be satisfied with *approximations* for these exact expressions. Ideally, these approximations should be easier to obtain than the exact expressions and they should hopefully be of a simpler form. Asymptotic theory is one way of obtaining such approximations by essentially asking what happens to the exact expressions as the sample size tends to infinity. For example, if we are interested in the expected value of $\widehat{\theta}_n$ and an exact expression for it is unavailable or unwieldy, we could then ask if the expected value of $\widehat{\theta}_n$ converges to θ as the sample size increases (i.e., if $\widehat{\theta}_n$ is "asymptotically unbiased"). One could try to verify this by first showing that the estimator $\widehat{\theta}_n$ itself "converges" to θ in an appropriate sense, and then by trying to obtain the convergence of the expected value of $\widehat{\theta}_n$ to θ from the "convergence" of the estimator. In order to properly pose and answer such questions we need to study various notions of convergence of random vectors.

The article is organized as follows: In Section 2 we define various modes of convergence of random vectors, and discuss the properties of and the relationships between these modes of convergence. Sections 3 and 4 provide results that allow us to deduce the convergence of certain important classes of random vectors from basic assumptions. In particular, in Section 3 we discuss laws of large numbers, including uniform laws of large numbers. A discussion of central limit theorems is given in Section 4. In Section 5 we suggest additional literature for further reading.

We emphasize that the article only covers material that lays the foundation for asymptotic theory. It does not provide results on the asymptotic properties of estimators for particular models; for references see Section 5. All of the material presented here is essentially textbook material. We provide proofs for some selected results for the purpose of practice and since some of the proofs provide interesting insights. For results given without a proof we provide references to widely available textbooks.

We adopt the following notation and conventions: Throughout the paper Z_1, Z_2, $\ldots$, and Z denote random vectors that take their values in a Euclidean space $\mathbf{R}^k$, $k \geq 1$. Furthermore, all random vectors involved in a particular statement are assumed to be defined on a common probability space $(\Omega,\mathfrak{F},P)$, except when noted otherwise. With $|.|$ we denote the absolute value and with $\|.\|$ the Euclidean norm. All matrices considered are real matrices. If A is a matrix, then A' denotes its transpose; if A is a square matrix, then A^{-1} denotes the inverse of A. The norm of a matrix A is denoted by $\|A\|$ and is taken to be $\|vec(A)\|$, where $vec(A)$ stands for the columnwise vectorization of A. If C_n is a sequence of sets, then $C_n \uparrow C$ stands for $C_n \subseteq C_{n+1}$ for all $n \in \mathbf{N}$ and $C = \bigcup_{n=1}^{\infty} C_n$. Similarly, $C_n \downarrow C$ stands for $C_n \supseteq C_{n+1}$ for all $n \in \mathbf{N}$ and $C = \bigcap_{n=1}^{\infty} C_n$.

2 Modes of Convergence for Sequences of Random Vectors

In this section we define and discuss various modes of convergence for sequences of random vectors taking their values in $\mathbf{R}^k$.

2.1 Convergence in Probability, Almost Surely, and in r-th Mean

We first consider the case where $k = 1$, i.e., the case of real valued random variables. Extension to the vector case are discussed later. We start by defining convergence in probability.

Definition 2.1 *(Convergence in probability) The sequence of random variables Z_n converges in probability (or stochastically) to the random variable Z if for every $\varepsilon > 0$*

$$\lim_{n\to\infty} P\left(|Z_n - Z| \leq \varepsilon\right) = 1. \tag{2.1}$$

We then write $\operatorname*{plim}_{n\to\infty} Z_n = Z$, or $Z_n \xrightarrow{p} Z$, or $Z_n \to Z$ i.p. as $n \to \infty$.

We next define almost sure convergence.

Definition 2.2 *(Almost sure convergence) The sequence of random variables Z_n converges almost surely (or strongly or with probability one) to the random variable Z if there exists a set $N \in \mathfrak{F}$ with $P(N) = 0$ such that $\lim_{n\to\infty} Z_n(\omega) = Z(\omega)$ for every $\omega \in \Omega - N$, or equivalently*

$$P\left(\{\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)\}\right) = 1. \tag{2.2}$$

We then write $Z_n \xrightarrow{a.s.} Z$, or $Z_n \to Z$ a.s., or $Z_n \to Z$ w.p.1 as $n \to \infty$.

The following theorem provides an alternative characterization of almost sure convergence.

Theorem 2.3 *The sequence of random variables Z_n converges almost surely to the random variable Z if and only if*

$$\lim_{n\to\infty} P\left(\{|Z_i - Z| \leq \varepsilon \text{ for all } i \geq n\}\right) = 1 \tag{2.3}$$

for every $\varepsilon > 0$.

Proof. Let

$$A = \{\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)\}$$

and

$$A_n^\varepsilon = \{\omega \in \Omega : |Z_i(\omega) - Z(\omega)| \leq \varepsilon \text{ for all } i \geq n\},$$

then (2.2) and (2.3) can be written equivalently as $P(A) = 1$ and $\lim_{n\to\infty} P(A_n^\varepsilon) = 1$. Next define $A^\varepsilon = \bigcup_{n=1}^\infty A_n^\varepsilon$ and observe that $A_n^\varepsilon \uparrow A^\varepsilon$. By construction A^ε is the set of all $\omega \in \Omega$ for which there exists some finite index $n_\varepsilon(\omega)$ such that $|Z_i(\omega) - Z(\omega)| \leq \varepsilon$ for all $i \geq n_\varepsilon(\omega)$. Consequently $A \subseteq A^\varepsilon$; in fact

$A = \bigcap_{\varepsilon>0} A^{\varepsilon}$. Now suppose (2.2) holds, i.e. $P(A) = 1$. Then, using the continuity theorem for probability measures, $P(A^{\varepsilon}) = \lim_{n\to\infty} P(A_n^{\varepsilon}) \geq P(A) = 1$, i.e., (2.3) holds. Conversely, suppose (2.3) holds, then $P(A^{\varepsilon}) = 1$. Observe that $A^{\varepsilon} \downarrow A$ as $\varepsilon \downarrow 0$. Choosing $\varepsilon = 1/k$ we have $A = \bigcap_{k=1}^{\infty} A^{1/k}$ and, using again the continuity theorem for probability measures, $P(A) = \lim_{k\to\infty} P(A^{1/k}) = 1$. ■

The above theorem makes it evident that almost sure convergence implies convergence in probability.

Theorem 2.4 *If $Z_n \overset{a.s.}{\to} Z$, then $Z_n \overset{p}{\to} Z$.*

Proof. Obviously the event $B_n^{\varepsilon} = \{\omega \in \Omega : |Z_n(\omega) - Z(\omega)| \leq \varepsilon\}$ contains the event $A_n^{\varepsilon} = \{\omega \in \Omega : |Z_i(\omega) - Z(\omega)| \leq \varepsilon \text{ for all } i \geq n\}$. Hence Theorem 2.3 implies that $\lim_{n\to\infty} P(B_n^{\varepsilon}) = 1$, i.e., that (2.1) holds . ■

The converse of the above theorem does not hold. That is, in general, convergence in probability does not imply almost sure convergence as is demonstrated by the following well-known example.

Example 2.5 *Let $\Omega = [0,1)$, let $\mathfrak{F}$ be the corresponding Borel σ-field, and let $P(.)$ be the uniform distribution on Ω, i.e., $P([a,b]) = b - a$. Define*

$$Z_n(\omega) = \begin{cases} 1 & \text{if } \omega \in \left[m_n 2^{-k_n}, (m_n+1)2^{-k_n}\right) \\ 0 & \text{otherwise} \end{cases}$$

where the integers m_n and k_n satisfy $n = m_n + 2^{k_n}$ and $0 \leq m_n < 2^{k_n}$. Let $Z = 0$ and let A_n^{ε} and B_n^{ε} be defined as above. Then for $\varepsilon < 1$ we have $B_n^{\varepsilon} = \Omega - \left[m_n 2^{-k_n}, (m_n+1)2^{-k_n}\right)$ and hence $P(B_n^{\varepsilon}) = 1 - 2^{-k_n} \to 1$ as $n \to \infty$. This establishes that Z_n converges to zero in probability. Observe further that $A_n^{\varepsilon} = \bigcap_{i=n}^{\infty} B_i^{\varepsilon} = \emptyset$. Consequently Z_n does not converge to zero almost surely. In fact, in this example $Z_n(\omega)$ does not converge to 0 for all $\omega \in \Omega$, although $Z_n \overset{p}{\to} 0$.

We next define convergence in r-th mean.

Definition 2.6 *(Convergence in r-th mean) The sequence of random variables Z_n converges in r-th mean (or in L^r) to the random variable Z, $0 < r < \infty$, if*

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

We then write $Z_n \overset{r-th}{\to} Z$ or $Z_n \to Z$ in L^r. For $r = 2$ we say the sequence converges in quadratic mean or mean square.

Remark 2.7 *For all three modes of convergence introduced above one can show that the limiting random variable Z is unique up to null sets. That is, suppose Z and Z^* are both limits of the sequence Z_n, then $P(Z = Z^*) = 1$.*

Lyapounov's inequality implies that $E|Z_n - Z|^s \leq \{E|Z_n - Z|^r\}^{s/r}$ for $0 < s \leq r$. As a consequence we have the following theorem, which tells us that the higher the value of r, the more stringent the condition for L^r convergence.

Theorem 2.8 *$Z_n \overset{r-th}{\to} Z$ implies $Z_n \overset{s-th}{\to} Z$ for $0 < s \leq r$.*

The following theorem gives conditions under which convergence in r-th mean implies convergence of the r-th moments.

Theorem 2.9 [1] *Suppose $Z_n \overset{r-th}{\to} Z$ and $E|Z|^r < \infty$. Then $E|Z|^r < \infty$ and $E|Z_n|^r \to E|Z|^r$. If, furthermore, Z_n^r and Z^r are well-defined for all n (e.g., if $Z_n \geq 0$ and $Z \geq 0$, or if r is a natural number), then also $EZ_n^r \to EZ^r$.*

By Chebyshev's inequality we have $P\{|Z_n - Z| \geq \varepsilon\} \leq E|Z_n - Z|^r / \varepsilon^r$ for $r > 0$. As a consequence, L^r-convergence implies convergence in probability, as stated in the following theorem.

Theorem 2.10 *If $Z_n \overset{r-th}{\to} Z$ for some $r > 0$, then $Z_n \overset{p}{\to} Z$.*

The corollary below follows immediately from Theorem 2.10 with $r = 2$ by utilizing the decomposition $E|Z_n - c|^2 = var(Z_n) + (EZ_n - c)^2$.

Corollary 2.11 *Suppose $EZ_n \to c$ and $var(Z_n) \to 0$, then $Z_n \overset{p}{\to} c$.*

The corollary is frequently used to show that for an estimator $\widehat{\theta}_n$ with $E\widehat{\theta}_n \to \theta$ (i.e., an asymptotically unbiased estimator) and with $var(\widehat{\theta}_n) \to 0$ we have $\widehat{\theta}_n \overset{p}{\to} \theta$.

Example 2.12 *Let y_t be a sequence of i.i.d. distributed random variables with $Ey_t = \theta$ and $var(y_t) = \sigma^2 < \infty$. Let $\widehat{\theta}_n = n^{-1} \sum_{t=1}^n y_t$ denote the sample mean. Then $E\widehat{\theta}_n = \theta$ and $var(\widehat{\theta}_n) = \sigma^2/n \to 0$, and hence $\widehat{\theta}_n \overset{p}{\to} \theta$.*

Theorem 2.10 and Corollary 2.11 show how convergence in probability can be implied from the convergence of appropriate moments. The converse is not true in general, and in particular $Z_n \overset{p}{\to} Z$ does not imply $Z_n \overset{r-th}{\to} Z$. In fact even $Z_n \overset{a.s.}{\to} Z$ does not imply $Z_n \overset{r-th}{\to} Z$. These claims are illustrated by the following example.

[1] For a proof see, e.g., Serfling (1980), p. 15.

Example 2.13 *Let Ω, $\mathfrak{F}$ and P be as in Example 2.5 and define*

$$Z_n(\omega) = \begin{cases} 0 & \text{for } \omega \in [0, 1-1/n), \\ n & \text{for } \omega \in [1-1/n, 1). \end{cases}$$

Then $Z_n(\omega) \to 0$ for all $\omega \in \Omega$ and hence $Z_n \xrightarrow{a.s.} 0$. However, $E|Z_n| = 1$ for all n and hence Z_n does not converge to 0 in L^1.

The above example shows in particular that an estimator that satisfies $\widehat{\theta}_n \xrightarrow{p} \theta$ (or $\widehat{\theta}_n \xrightarrow{a.s.} \theta$) need not satisfy $E\widehat{\theta}_n \to \theta$, i.e., need not be asymptotically unbiased. Additional conditions are needed for such a conclusion to be proper. Such conditions are given in the following theorem. The theorem states that convergence in probability implies convergence in r-th mean, given that the convergence is dominated.

Theorem 2.14 *(Dominated convergence theorem) Suppose $Z_n \xrightarrow{p} Z$, and there exists a random variable Y satisfying $|Z_n| \leq Y$ a.s. for all n and $EY^r < \infty$. Then $Z_n \xrightarrow{r-th} Z$ and $E|Z|^r < \infty$. (Of course, the theorem also holds if $Z_n \xrightarrow{p} Z$ is replaced by $Z_n \xrightarrow{a.s.} Z$, since the latter implies the former.)*

Under the assumptions of the above theorem also convergence of the r-th moments follows in view of Theorem 2.9. We also note that the existence of a random variable Y satisfying the requirements in Theorem 2.14 is certainly guaranteed if there exists a real number M such that $|Z_n| \leq M$ a.s. for all n (choose $Y = M$).

Now let Z_n be a sequence of random vectors taking their values in $\mathbf{R}^k$. Convergence in probability, almost surely and in the r-th mean are then defined exactly as in the case $k = 1$ with the only difference that the absolute value $|.|$ has to be replaced by $\|.\|$, the Euclidean norm on $\mathbf{R}^k$. Upon making this replacement all of the results presented in this subsection generalize to the vector case with two obvious exceptions: First, in Corollary 2.11 the condition $var(Z_n) \to 0$ has to be replaced by the conditions that the variances of the components of Z_n converge to zero, or equivalently, that the variance covariance matrix of Z_n converges to zero. Second, the last claim in Theorem 2.9 continues to hold if the symbol Z_n^r is interpreted as to represent the vector of the r-th power of the components of $Z_{n.}$ Instead of extending the convergence notions to the vector case by replacing the absolute value $|.|$ by the norm $\|.\|$, we could have defined convergence in probability, almost surely and in r-th mean for sequences of random vectors by requiring that each component of Z_n satisfies Definition 2.1, 2.2, or 2.6, respectively. That this leads to an equivalent definition is shown in the following theorem.

Theorem 2.15 *Let Z_n and Z be random vectors taking their values in $\mathbf{R}^k$, and let $Z_n^{(i)}$ and $Z^{(i)}$ denote their i-th component, respectively. Then $Z_n \xrightarrow{p} Z$ if and only if $Z_n^{(i)} \xrightarrow{p} Z^{(i)}$ for $i = 1, \ldots, k$. An analogous statement holds for almost sure convergence and for convergence in r-th mean.*

The theorem follows immediately from the following simple inequality:

$$\left| Z_n^{(i)} - Z^{(i)} \right| \le \| Z_n - Z \| \le \sqrt{k} \max_{i=1,\ldots,k} \left(\left| Z_n^{(i)} - Z^{(i)} \right| \right).$$

For sequences of random $k \times l$-matrices W_n convergence in probability, almost surely and in r-th mean is defined as the corresponding convergence of the sequence $vec\,(W_n)$.

We finally note the following simple fact: Suppose $Z_1, Z_2, \ldots,$ and Z are non-random vectors, then $Z_n \xrightarrow{p} Z$, $Z_n \xrightarrow{a.s.} Z$, and $Z_n \xrightarrow{r-th} Z$ each hold if and only if $Z_n \to Z$ as $n \to \infty$. That is, in this case all of the concepts of convergence of random vectors introduced above coincide with the usual convergence concept for sequences of vectors in $\mathbf{R}^k$.

2.2 Convergence in Distribution

Let $\hat{\theta}_n$ be an estimator for a real-valued parameter θ and assume $\hat{\theta}_n \xrightarrow{p} \theta$. If G_n denotes the cumulative distribution function (c.d.f.) of $\hat{\theta}_n$, i.e., $G_n(z) = P\left(\hat{\theta}_n \le z\right)$, then as $n \to \infty$

$$G_n(z) \to \begin{cases} 0 & \text{for } z < \theta \\ 1 & \text{for } z > \theta. \end{cases} \tag{2.4}$$

To see this observe that $P(\hat{\theta}_n \le z) = P(\hat{\theta}_n - \theta \le z - \theta) \le P(\left|\hat{\theta}_n - \theta\right| \ge \theta - z)$ for $z < \theta$, and $P(\hat{\theta}_n \le z) = 1 - P(\hat{\theta}_n > z) = 1 - P(\hat{\theta}_n - \theta > z - \theta) \ge 1 - P(\left|\hat{\theta}_n - \theta\right| > z - \theta)$ for $z > \theta$. The result in (2.4) shows that the distribution of $\hat{\theta}_n$ "collapses" into the degenerate distribution at θ, i.e., into

$$G(z) = \begin{cases} 0 & \text{for } z < \theta \\ 1 & \text{for } z \ge \theta. \end{cases} \tag{2.5}$$

Consequently, knowing that $\hat{\theta}_n \xrightarrow{p} \theta$ does not provide information about the shape of G_n. As a point of observation note that $G_n(z) \to G(z)$ for $z \ne \theta$, but $G_n(z)$ may not converge to $G(z) = 1$ for $z = \theta$. For example, if $\hat{\theta}_n$ is distributed symmetrically around θ, then $G_n(\theta) = 1/2$ does not converge to $G(\theta) = 1$.

This raises the question how we can obtain information about G_n based on some limiting process. Consider, e.g., the case where $\hat{\theta}_n$ is the sample mean of i.i.d. random variables with mean θ and variance σ^2. Then $\hat{\theta}_n \xrightarrow{p} \theta$ in light of Corollary 2.11, since $E\hat{\theta}_n = \theta$ and $var(\hat{\theta}_n) = \sigma^2/n \to 0$. Consequently, as discussed above, the distribution of $\hat{\theta}_n$ "collapses" into the degenerate distribution at θ. Observe, however, that the re-scaled variable $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ has mean zero and variance σ^2. This indicates that the distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ will *not* collapse to a degenerate distribution, but its c.d.f. will "converge" to some non-degenerate limiting c.d.f. To formalize this idea we need to define an appropriate notion of convergence of c.d.f.'s.[2]

Definition 2.16 *(Convergence in distribution) Let F_1, F_2, $\ldots$, and F denote c.d.f.'s on* $\mathbf{R}$*. Then F_n converges to F in distribution (or in law) if*

$$\lim_{n\to\infty} F_n(z) = F(z)$$

for all $z \in \mathbf{R}$ that are continuity points of F. We then write $F_n \xrightarrow{d} F$ or $F_n \xrightarrow{L} F$.

Let Z_1, Z_2, $\ldots$, and Z denote random variables with corresponding c.d.f.'s F_1, F_2, $\ldots$, and F, respectively. We then say that Z_n converges in distribution (or in law) to Z, if F_n converges to F in distribution. We write $Z_n \xrightarrow{d} Z$ or $Z_n \xrightarrow{L} Z$.

Remark 2.17 **(a)** *The reason for only requiring in the above definition that $F_n(z) \to F(z)$ converges for all continuity points of F is to accommodate situations as, e.g., the one discussed at the beginning of this subsection. Of course, if F is continuous, then $F_n \xrightarrow{d} F \Leftrightarrow F_n(z) \to F(z)$ for all $x \in$* $\mathbf{R}$.

(b) *As is evident from the definition the concept of convergence in distribution is defined completely in terms of the convergence of distribution functions. It is for that reason that the concept of convergence in distribution remains well defined even if the random variables are not defined on a common probability space.*

(c) *To further illustrate what convergence in distribution does not mean consider the following example: Let Y be a random variable that takes the values $+1$ and -1 with probability $1/2$. Define $Z_n = Y$ for $n \geq 1$ and $Z = -Y$. Then clearly $Z_n \xrightarrow{d} Z$ since Z_n and Z have the same distribution, but $|Z_n - Z| = 2$ for all $n \geq 1$. That is, convergence in distribution does not necessarily mean that the difference between random variables vanishes in the limit. More generally, if $Z_n \xrightarrow{d} Z$ and one replaces the sequence Z_n by a sequence Z_n^* that has the same marginal distributions, then also $Z_n^* \xrightarrow{d} Z$.*

The next theorem provides several equivalent characterizations of convergence in distribution.

[2]We note that the rescaled quantities like $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ typically do not converge a.s. or i.p., and thus a new notion of convergence is needed for these quantities.

Theorem 2.18 [3] *Consider the cumulative distribution functions* F, F_1, F_2, ... *Let* Q, Q_1, Q_2, ... *denote the corresponding probability measures on* $\mathbf{R}$, *and let* ϕ, ϕ_1, ϕ_2, ... *denote the corresponding characteristic functions. Then the following statements are equivalent:*

(i) $F_n \xrightarrow{d} F$

(ii) $\lim_{n\to\infty} Q_n(A) = Q(A)$ *for all Borel sets* $A \subseteq \mathbf{R}$ *that are* Q*-continuous, i.e., for all Borel sets* A *whose boundary* ∂A *satisfies* $Q(\partial A) = 0$.

(iii) $\lim_{n\to\infty} \int f dF_n = \int f dF$ *for all bounded and continuous real valued functions* f *on* $\mathbf{R}$.

(iv) $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ *for all* $t \in \mathbf{R}$.

If, furthermore the cumulative distribution functions F, F_1, F_2, ... *have moment generating functions* M, M_1, M_2, ... *in some common interval* $[-t_*, t_*]$, $t_* > 0$, *then (i), (ii), (iii) or (iv) are, respectively, equivalent to*

(v) $\lim_{n\to\infty} M_n(t) = M(t)$ *for all* $t \in [-t_*, t_*]$.

Remark 2.19 *The equivalence of (i) and (ii) of the theorem can be reformulated as* $Z_n \xrightarrow{d} Z \Leftrightarrow P(Z_n \in A) \to P(Z \in A)$ *for all Borel sets* A *with* $P(Z \in \partial A) = 0$. *The equivalence of (i) and (iii) can be expressed equivalently as* $Z_n \xrightarrow{d} Z \Leftrightarrow Ef(Z_n) \to Ef(Z)$ *for all bounded and continuous real valued functions* f *on* $\mathbf{R}$.

The following theorem relates convergence in probability to convergence in distribution.

Theorem 2.20 $Z_n \xrightarrow{p} Z$ *implies* $Z_n \xrightarrow{d} Z$. *(Of course, the theorem also holds if* $Z_n \xrightarrow{p} Z$ *is replaced by* $Z_n \xrightarrow{a.s.} Z$ *or* $Z_n \xrightarrow{r-th} Z$, *since the latter imply the former.)*

Proof. Let $f(z)$ be any bounded and continuous real valued function, and let C denote the bound. Then $Z_n \xrightarrow{p} Z$ implies $f(Z_n) \xrightarrow{p} f(Z)$ by the results on convergence in probability of transformed sequences given in Theorem 2.28 in Section 2.3. Since $|f(Z_n(\omega)| \leq C$ for all n and $\omega \in \Omega$ it then follows from Theorems 2.14 and 2.9 that $Ef(Z_n) \to Ef(Z)$, and hence $Z_n \xrightarrow{d} Z$ by Theorem 2.18. ■

[3] See, e.g., Billingsley (1968), p. 12, Billingsley (1979), p. 345, and Serfling (1980), p. 16.

The converse of the above theorem does not hold in general, i.e., $Z_n \xrightarrow{d} Z$ does not imply $Z_n \xrightarrow{p} Z$. To see this consider the following example: Let $Z \sim N(0,1)$ and put $Z_n = (-1)^n Z$. Then Z_n does not converge almost surely or in probability. But since each $Z_n \sim N(0,1)$, evidently $Z_n \xrightarrow{d} Z$.

Convergence in distribution to a constant is, however, equivalent to convergence in probability to that constant.

Theorem 2.21 *Let $c \in \mathbf{R}$, then $Z_n \xrightarrow{d} c$ is equivalent to $Z_n \xrightarrow{p} c$.*

Proof. Because of Theorem 2.20 we only have to show that $Z_n \xrightarrow{d} c$ implies $Z_n \xrightarrow{p} c$. Observe that for any $\varepsilon > 0$

$$\begin{aligned} & P(|Z_n - c| > \varepsilon) \\ = \; & P(Z_n - c < -\varepsilon) + P(Z_n - c > \varepsilon) \\ \leq \; & P(Z_n \leq c - \varepsilon) - P(Z_n \leq c + \varepsilon) + 1 \\ = \; & F_n(c - \varepsilon) - F_n(c + \varepsilon) + 1 \end{aligned}$$

where F_n is the c.d.f. of Z_n. The c.d.f. of $Z = c$ is

$$F(z) = \begin{cases} 0 & z < c \\ 1 & z \geq c \end{cases}.$$

Hence, $c - \varepsilon$ and $c + \varepsilon$ are continuity points of F. Since $Z_n \xrightarrow{d} Z$ it follows that $F_n(c - \varepsilon) \to F(c - \varepsilon) = 0$ and $F_n(c + \varepsilon) \to F(c + \varepsilon) = 1$. Consequently,

$$0 \leq P(|Z_n - c| > \varepsilon) \leq F_n(c - \varepsilon) + 1 - F_n(c + \varepsilon) \to 0 + 1 - 1 = 0.$$

This shows $Z_n \xrightarrow{p} Z = c$. ■

In general convergence in distribution does not imply convergence of moments; in fact the moments may not even exist. However, we have the following result.

Theorem 2.22 [4] *Suppose $Z_n \xrightarrow{d} Z$ and suppose that $\sup_n E|Z_n|^r < \infty$ for some $0 < r < \infty$. Then for all $0 < s < r$ we have $E|Z|^s < \infty$ and $\lim_{n\to\infty} E|Z_n|^s = E|Z|^s$. If, furthermore, Z^s and Z_n^s are well-defined for all n, then also $\lim_{n\to\infty} EZ_n^s = EZ^s$.*

Remark 2.23 *Since $Z_n \xrightarrow{p} Z$ and $Z_n \xrightarrow{a.s.} Z$ imply $Z_n \xrightarrow{d} Z$, Theorem 2.22 provides sufficient conditions under which $Z_n \xrightarrow{p} Z$ and $Z_n \xrightarrow{a.s.} Z$ imply convergence of moments. These conditions are an alternative to those of Theorem 2.14.*

[4] See, e.g., Serfling (1980), pp. 13-14.

The concept of convergence in distribution can be generalized to sequences of random vectors Z_n taking their values in $\mathbf{R}^k$. Contrary to the approach taken in generalizing the notions of convergence in probability, almost surely and in r-th mean to the vector case, the appropriate generalization is here *not* obtained by simply requiring that the component sequences $Z_n^{(i)}$ converge in distribution for $i = 1, \ldots, k$. Such an attempt at generalizing the notion of convergence in distribution would yield a nonsensical convergence concept as is illustrated by Example 2.26 below. The proper generalization is given in the following definition.

Definition 2.24 *Let F_1, F_2, ..., and F denote c.d.f.'s on $\mathbf{R}^k$. Then F_n converges to F in distribution (or in law) if*

$$\lim_{n\to\infty} F_n(z) = F(z)$$

for all $z \in \mathbf{R}^k$ that are continuity points of F. We then write $F_n \xrightarrow{d} F$ or $F_n \xrightarrow{L} F$.

Let Z_1, Z_2, ..., and Z denote random vectors raking their values in $\mathbf{R}^k$ with corresponding c.d.f.'s F_1, F_2, ..., and F, respectively. We then say that Z_n converges in distribution (or in law) to Z, if F_n converges to F in distribution. We write $Z_n \xrightarrow{d} Z$ or $Z_n \xrightarrow{L} Z$.

All the results presented in this subsection so far also hold for the multivariate case (with $\mathbf{R}^k$ replacing $\mathbf{R}$). Convergence in distribution of a sequence of random matrices W_n is defined as convergence in distribution of $vec(W_n)$.

The next theorem states that convergence in distribution of the joint distributions implies convergence in distribution of the marginal distributions.

Theorem 2.25 *$F_n \xrightarrow{d} F$ implies $F_n^i \xrightarrow{d} F^{(i)}$ and $Z_n \xrightarrow{d} Z$ implies $Z_n^{(i)} \xrightarrow{d} Z^{(i)}$, where $F_n^{(i)}$ and $F^{(i)}$ denote the i-th marginal distribution of F_n and F, and $Z_n^{(i)}$ and $Z^{(i)}$ denote the i-th component of Z_n and Z, respectively.*

Proof. The result follows from Theorem 2.28 below, since projections are continuous. ■

However, as alluded to in the above discussion, the converse of Theorem 2.25 is not true. That is convergence in distribution of the marginal distributions is not equivalent to convergence in distribution of the joint distribution, as is illustrated by the following counter example.

Example 2.26 *Let $Z \sim N(0,1)$ and let*

$$Z_n = \begin{pmatrix} Z \\ (-1)^n Z \end{pmatrix}.$$

Clearly, each component of Z_n converges in distribution to $N(0,1)$. However, for n even the distribution of Z_n is concentrated on the line $\{(z,z) : z \in \mathbf{R}\}$, whereas for n odd the distribution of Z_n is concentrated on the line $\{(z,-z) : z \in \mathbf{R}\}$. Consequently, the random vectors Z_n do not converge in distribution.

The following result is frequently useful in reducing questions about convergence in distribution of random vectors to corresponding questions about convergence in distribution of random variables.

Theorem 2.27 *(Cramér-Wold device) Let Z_1, Z_2, ..., and Z denote random vectors taking their values in $\mathbf{R}^k$. Then the following statements are equivalent:*

(i) $Z_n \xrightarrow{d} Z$

(ii) $\alpha' Z_n \xrightarrow{d} \alpha' Z$ *for all* $\alpha \in \mathbf{R}^k$.

(iii) $\alpha' Z_n \xrightarrow{d} \alpha' Z$ *for all* $\alpha \in \mathbf{R}^k$ *with* $\|\alpha\| = 1$.

Proof. The equivalence of (ii) and (iii) is obvious. We now prove the equivalence of (i) with (iii). Let $\phi_n(t)$ and $\phi(t)$, denote respectively the characteristic function of Z_n and Z. According to the multivariate version of Theorem 2.18 we have $Z_n \xrightarrow{d} Z$ if and only if $\phi_n(t) \to \phi(t)$ for all $t = (t_1, \ldots, t_k)' \in \mathbf{R}^k$. Let $\phi_n^\alpha(s)$ and $\phi^\alpha(s)$ denote the characteristic functions of $\alpha' Z_n$ and $\alpha' Z$, respectively. Again, $\alpha' Z_n \xrightarrow{d} \alpha' Z$ if and only if $\phi_n^\alpha(s) \to \phi^\alpha(s)$ for all $s \in \mathbf{R}$. Observe that for $t \neq 0$ we have

$$\phi_n(t) = E(\exp(it' Z_n)) = E(\exp(is\alpha' Z_n) = \phi_n^\alpha(s)$$

with $\alpha = t/\|t\|$ and $s = \|t\|$. Note that $\|\alpha\| = 1$. Similarly, $\phi(t) = \phi^\alpha(s)$. Consequently, $\phi_n(t) \to \phi(t)$ for all $t \neq 0$ if and only if $\phi_n^\alpha(s) \to \phi^\alpha(s)$ for all $s \neq 0$ and all α with $\|\alpha\| = 1$. Since $\phi_n(0) = \phi(0) = 1$ and $\phi_n^\alpha(0) = \phi^\alpha(0) = 1$, the proof is complete observing that $t = 0$ if and only if $s = 0$. ■

2.3 Convergence Properties and Transformations

We are often interested in the convergence properties of transformed random vectors or variables. In particular, suppose Z_n converges to Z in a certain mode, then given a function g we may ask the question whether or not $g(Z_n)$ converges to $g(Z)$ in the same mode. The following theorem answers the question in the confirmative, provided g is continuous (in the sense specified below). Part (a) of the theorem is commonly referred to as Slutzky's theorem.

Theorem 2.28 [5] *Let Z_1, Z_2, ..., and Z be random vectors in $\mathbf{R}^k$. Furthermore, let $g : \mathbf{R}^k \to \mathbf{R}^s$ be a Borel-measurable function and assume that g is*

[5] See, e.g., Serfling (1980), p. 24.

continuous with P_Z-probability one (where P_Z denotes the probability measure induced by Z on $\mathbf{R}^k$).[6] *Then*

(a) $Z_n \xrightarrow{p} Z$ *implies* $g(Z_n) \xrightarrow{p} g(Z)$,

(b) $Z_n \xrightarrow{a.s.} Z$ *implies* $g(Z_n) \xrightarrow{a.s.} g(Z)$,

(c) $Z_n \xrightarrow{d} Z$ *implies* $g(Z_n) \xrightarrow{d} g(Z)$.

In the special case where $Z = c$ is a constant or a vector of constants, the continuity condition on g in the above theorem only requires that the function g is continuous at c.

As special cases of Theorem 2.28 we have, e.g., the following corollaries.

Corollary 2.29 *Let W_n and V_n be sequences of k-dimensional random vectors. Suppose $W_n \to W$ and $V_n \to V$ i.p. [a.s.], then*

$$\begin{aligned} W_n \pm V_n &\to W \pm V \qquad \text{i.p. [a.s.]}, \\ W_n' V_n &\to W' V \qquad \text{i.p. [a.s.]}. \end{aligned}$$

In case $k = 1$,

$$W_n / V_n \to W / V \qquad \text{i.p. [a.s.]}$$

if $V \neq 0$ with probability one, and where W_n/V_n is set to an arbitrary value on the event $\{V_n = 0\}$.[7]

Proof. The assumed convergence of W_n and V_n implies that $Z_n = (W_n', V_n')'$ converges to $Z = (W', V')'$ i.p. [a.s.] in view of Theorem 2.15. The corollary then follows from Theorem 2.28(a),(b) since the maps $g_1(w,v) = w + v$, $g_2(w,v) = w - v$, $g_3(w,v) = w'v$ are continuous on all of $\mathbf{R}^{2k}$, and since the map $g_4(w,v) = w/v$ if $v \neq 0$ and $g_4(w,v) = c$ for $v = 0$ (with c arbitrary) is continuous on $A = \mathbf{R} \times (\mathbf{R} - \{0\})$, observing furthermore that $P_Z(A) = 1$ since $V \neq 0$ with probability 1. ■

The proof of the following corollary is completely analogous.

Corollary 2.30 *Let W_n and V_n be sequences of random matrices of fixed dimension. Suppose $W_n \to W$ and $V_n \to V$ i.p. [a.s.], then*

$$\begin{aligned} W_n \pm V_n &\to W \pm V \qquad \text{i.p. [a.s.]}, \\ W_n V_n &\to W V \qquad \text{i.p. [a.s.]}, \end{aligned}$$

[6] That is, let $A \subseteq \mathbf{R}^k$ denote the set of continuity points of g, then $P_Z(A) = P(Z \in A) = 1$. Of course, if g is continuous on $\mathbf{R}^k$, then $A = \mathbf{R}^k$ and the condition $P_Z(A) = 1$ is trivially satisfied.

[7] The event $\{V_n = 0\}$ has probability approaching zero, and hence it is irrelevant which value is assigned to W_n/V_n on this event.

Furthermore

$$W_n V_n^{-1} \to W V^{-1} \quad \text{and} \quad V_n^{-1} W_n \to V^{-1} W \qquad \text{i.p. [a.s.]}$$

if V is nonsingular with probability one, and where $W_n V_n^{-1}$ and $V_n^{-1} W_n$ are set to an arbitrary matrix of appropriate dimension on the event $\{V_n$ singular$\}$. (The matrices are assumed to be of conformable dimensions.)

The following example shows that convergence in probability or almost surely in Corollaries 2.29 and 2.30 *cannot* be replaced by convergence in distribution.

Example 2.31 *Let $U \sim N(0,1)$ and define $W_n = U$ and $V_n = (-1)^n U$. Then*

$$W_n + V_n = \begin{cases} 2U \sim N(0,4) & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$$

Clearly, $W_n + V_n$ does not converge in distribution, although $W_n \xrightarrow{d} U$ and $V_n \xrightarrow{d} U$.

The reason behind this negative result is again the fact that convergence in distribution of the components of a random vector does in general not imply convergence in distribution of the entire random vector. Of course, if the entire random vector $Z_n = (W_n', V_n')'$ converges in distribution to $Z = (W', V')'$ then $W_n \pm V_n \xrightarrow{d} W \pm V$, $W_n' V_n \xrightarrow{d} W'V$ as a consequence of Theorem 2.28; also, if $k = 1$ and $V \neq 0$ with probability 1, then $W_n / V_n \xrightarrow{d} W/V$.

However, there is an important special case in which we can conclude that $Z_n = (W_n', V_n')' \xrightarrow{d} Z = (W', V')'$ from knowing that $W_n \xrightarrow{d} W$ and $V_n \xrightarrow{d} V$: This is the case where $V = c$ and c is a constant vector.

Theorem 2.32 *Let W_n and V_n be sequences of $k \times 1$ and $l \times 1$ random vectors, respectively. Let W be a $k \times 1$ random vector and let $V = c$ be a constant vector in $\mathbf{R}^l$. Suppose $W_n \xrightarrow{d} W$ and $V_n \xrightarrow{d} c$ (or equivalently $V_n \xrightarrow{p} c$ in light of Theorem 2.21). Then $Z_n = (W_n', V_n')' \xrightarrow{d} Z = (W', V')' = (W', c')'$.*

Proof. Let $\phi_n(t)$ and $\phi(t)$ denote, respectively, the characteristic function of Z_n and Z. To show that $Z_n \xrightarrow{d} Z$ it suffices to show that $\phi_n(t) \to \phi(t)$ for all $t \in \mathbf{R}^{k+l}$ in light of the multivariate version of Theorem 2.18. Let $t = (s', u')'$ with $s \in \mathbf{R}^k$ and $u \in \mathbf{R}^l$ arbitrary. Observing that $|\exp(is'W_n)| = 1 = |\exp(iu'c)|$, we have

$$\begin{aligned} & \left|\phi_n(t) - \phi(t)\right| \qquad (2.6) \\ = & \left| E\left(e^{is'W_n} e^{iu'V_n} - e^{is'W} e^{iu'c} \right) \right| \\ \leq & E\left[\left| e^{is'W_n} \right| \left| e^{iu'V_n} - e^{iu'c} \right| \right] + \left| e^{iu'c} \right| \left| E\left(e^{is'W_n} - e^{is'W} \right) \right| \\ \leq & E\left| e^{iu'V_n} - e^{iu'c} \right| + \left| E\left(e^{is'W_n} - e^{is'W} \right) \right| \\ = & E\left| e^{iu'V_n} - e^{iu'c} \right| + \left| \phi_n^W(s) - \phi^W(s) \right|, \end{aligned}$$

where $\phi_n^W(s)$ and $\phi^W(s)$ denote, respectively, the characteristic function of W_n and W. Since $V_n \xrightarrow{p} c$ it follows from Theorem 2.28 that $\exp(iu'V_n) - \exp(iu'c) \xrightarrow{p} 0$. Observing that $|\exp(iu'V_n) - \exp(iu'c)| \leq 2$ it follows furthermore from Theorem 2.14 that $E|\exp(iu'V_n) - \exp(iu'c)| \to 0$. By assumption $W_n \xrightarrow{d} W$. It then follows again from the multivariate version of Theorem 2.18 that $\phi_n^W(s) \to \phi^W(s)$. Thus both terms in the last line of (2.6) converge to zero, and hence $\phi_n(t) \to \phi(t)$. ■

Given Theorem 2.32 the following result follows immediately from Theorem 2.28.

Corollary 2.33 *Let W_n and V_n be sequences of $k \times 1$ and $l \times 1$ random vectors, respectively. Let W be a $k \times 1$ random vector and c a constant vector in $\mathbf{R}^l$. Suppose $W_n \xrightarrow{d} W$ and $V_n \xrightarrow{d} c$ (or equivalently $V_n \xrightarrow{p} c$). Let $g : \mathbf{R}^k \times \mathbf{R}^l \to \mathbf{R}^s$ be a Borel measurable function and assume that g is continuous in every point of $A \times \{c\}$ where $A \subseteq \mathbf{R}^k$ satisfies $P(W \in A) = 1$. Then $g(W_n, V_n) \xrightarrow{d} g(W, c)$.*

As a further corollary we have the following useful results.

Corollary 2.34 *Let W_n and V_n be sequences of $k \times 1$ and $l \times 1$ random vectors, let A_n and B_n be a sequences of $l \times k$ and $k \times k$ random matrices, respectively. Furthermore, let W be a $k \times 1$ random vector, let c be a $l \times 1$ non-stochastic vector, and let A and B be some non-stochastic $l \times k$ and $k \times k$ matrices.*

(a) *For $k = l$*

$$W_n \xrightarrow{d} W, V_n \xrightarrow{p} c \quad \textit{implies} \quad \begin{array}{l} W_n \pm V_n \xrightarrow{d} W \pm c \\ W_n' V_n \xrightarrow{d} W'c. \end{array}$$

(If $c = 0$, then $W_n' V_n \xrightarrow{d} 0$ and hence also $W_n' V_n \xrightarrow{p} 0$).

(b) *For $k = l = 1$*

$$W_n \xrightarrow{d} W, V_n \xrightarrow{p} c \quad \textit{implies} \quad \begin{array}{ll} W_n/V_n \xrightarrow{d} W/c & \textit{if } c \neq 0, \\ V_n/W_n \xrightarrow{d} c/W & \textit{if } P(W = 0) = 0. \end{array}$$

(c)

$$W_n \xrightarrow{d} W,\ V_n \xrightarrow{p} c,\ A_n \xrightarrow{p} A \quad \textit{implies} \quad A_n W_n + V_n \xrightarrow{d} AW + c,$$

(d)

$$W_n \xrightarrow{d} W,\ B_n \xrightarrow{p} B \quad \textit{implies} \quad W_n' B_n W_n \xrightarrow{d} W'BW.$$

Of course, if in the above corollary $W \sim N(\mu, \Sigma)$, then $AW + c \sim N(A\mu + c, A\Sigma A')$. If $W \sim N(0, I_k)$ and B is idempotent of rank p, then $W'BW \sim \chi(p)$.

2.4 Orders of Magnitude

In determining the limiting behavior of sequences of random variables it is often helpful to employ notions of orders of relative magnitudes. We start with a review of the concepts of order of magnitudes for sequences of real numbers.

Definition 2.35 *(Order of magnitude of a sequence of real numbers) Let a_n be a sequence of real numbers and let c_n be a sequence of positive real numbers. We then say a_n is at most of order c_n, and write $a_n = O(c_n)$, if there exists a constant $M < \infty$ such that $c_n^{-1} |a_n| \le M$ for all $n \in \mathbf{N}$. We say a_n is of smaller order than c_n, and write $a_n = o(c_n)$, if $c_n^{-1} |a_n| \to 0$ as $n \to \infty$. (The definition extends to vectors and matrices by applying the definition to each element or, equivalently, to the norm.)*

The following results concerning the algebra of order in magnitude operations are often useful.

Theorem 2.36 *Let a_n and b_n be sequences of real numbers, and let c_n and d_n be sequences of positive real numbers.*

(a) *If $a_n = o(c_n)$ and $b_n = o(d_n)$, then $a_n b_n = o(c_n d_n)$, $|a_n|^s = o(c_n^s)$ for $s > 0$, $a_n + b_n = o(\max\{c_n, d_n\}) = o(c_n + d_n)$.*

(b) *If $a_n = O(c_n)$ and $b_n = O(d_n)$, then $a_n b_n = O(c_n d_n)$, $|a_n|^s = O(c_n^s)$ for $s > 0$, $a_n + b_n = O(\max\{c_n, d_n\}) = O(c_n + d_n)$.*

(c) *If $a_n = o(c_n)$ and $b_n = O(d_n)$, then $a_n b_n = o(c_n d_n)$.*

We now generalize the concept of order of magnitude from sequences of real numbers to sequences of random variables.

Definition 2.37 *(Order in probability of a sequence of random variables) Let Z_n be a sequence of random variables, and let c_n be a sequence of positive real numbers. We then say Z_n is at most of order c_n in probability, and write $Z_n = O_p(c_n)$, if for every $\varepsilon > 0$ there exists a constant $M_\varepsilon < \infty$ such that $P(c_n^{-1} |Z_n| \ge M_\varepsilon) \le \varepsilon$. We say Z_n is of smaller order in probability than c_n, and write $Z_n = o_p(c_n)$, if $c_n^{-1} |Z_n| \xrightarrow{p} 0$ as $n \to \infty$. (The definition extends to vectors and matrices by applying the definition to each element or, equivalently, to the norm.)*

The algebra of order in probability operations O_p and o_p is identical to that of order in magnitude operations O and o presented in the theorem above; see, e.g., Fuller (1976), p. 184.

A sequence of random variables Z_n that is $O_p(1)$ is also said to be "stochastically bounded" or "bounded in probability". The next theorem gives sufficient conditions for a sequence to be stochastically bounded.

Theorem 2.38 *(a) Suppose* $E|Z_n|^r = O(1)$ *for some* $r > 0$*, then* $Z_n = O_p(1)$.

(b) Suppose $Z_n \xrightarrow{d} Z$*, then* $Z_n = O_p(1)$.

Proof. Part (a) follows readily from Markov's inequality. To prove part (b) fix $\varepsilon > 0$. Now choose M_ε^* such that F is continuous at $-M_\varepsilon^*$ and M_ε^*, and $F(-M_\varepsilon^*) \leq \varepsilon/4$ and $F(M_\varepsilon^*) \geq 1 - \varepsilon/4$. Since every c.d.f. has at most a countable number of discontinuity points, such a choice is possible. By assumption $F_n(z) \to F(z)$ for all continuity points of F. Let n_ε be such that for all $n \geq n_\varepsilon$

$$|F_n(-M_\varepsilon^*) - F(-M_\varepsilon^*)| \leq \varepsilon/4$$

and

$$|F_n(M_\varepsilon^*) - F(M_\varepsilon^*)| \leq \varepsilon/4.$$

Then for $n \geq n_\varepsilon$

$$\begin{aligned} P(|Z_n| \geq M_\varepsilon^*) &\leq F_n(-M_\varepsilon^*) - F_n(M_\varepsilon^*) + 1 \\ &\leq F(-M_\varepsilon^*) - F(M_\varepsilon^*) + 1 + \varepsilon/2 \leq \varepsilon. \end{aligned}$$

Since $\lim_{M\to\infty} P(|Z_i| \geq M) = 0$ for each $i \in \mathbf{N}$ we can find an M_ε^{**} such that $P(|Z_i| \geq M_\varepsilon^{**}) \leq \varepsilon$ for $i = 1, \ldots, n_\varepsilon - 1$. Now let $M_\varepsilon = \max\{M_\varepsilon^*, M_\varepsilon^{**}\}$. Then $P(|Z_n| \geq M_\varepsilon) \leq \varepsilon$ for all $n \in \mathbf{N}$. ■

3 Laws of Large Numbers

Let Z_t, $t \in \mathbf{N}$, be a sequence of random variables with $EZ_t = \mu_t$. Furthermore let $\bar{Z}_n = n^{-1}\sum_{t=1}^n Z_t$ denote the sample mean, and let $\bar{\mu}_n = E\bar{Z}_n = n^{-1}\sum_{t=1}^n \mu_t$. A law of large numbers (LLN) then specifies conditions under which

$$\bar{Z}_n - E\bar{Z}_n = n^{-1}\sum_{t=1}^n (Z_t - \mu_t)$$

converges to zero either in probability or almost surely. If the convergence is in probability we speak of a weak LLN, if the convergence is almost surely we speak of a strong LLN. We note that in applications the random variables Z_t may themselves be functions of other random variables.

The usefulness of LLNs stems from the fact that many estimators can be expressed as (continuous) functions of sample averages of random variables, or differ from such a function only by a term that can be shown to converge to zero i.p. or a.s. Thus to establish the probability or almost sure limit of such an estimator we may try to establish in a first step the limits for the respective averages by means of LLNs. In a second step we may then use Theorem 2.28 to derive the actual limit for the estimator.

Example 3.1 *As an illustration consider the linear regression model* $y_t = x_t\theta + \varepsilon_t$*,* $t = 1, \ldots, n$*, where* y_t*,* x_t *and* ε_t *are all scalar and denote the dependent variable, the independent variable and the disturbance term in period* t*. The ordinary least squares estimator for the parameter* θ *is then given by*

$$\widehat{\theta}_n = \frac{\sum\limits_{t=1}^{n} x_t y_t}{\sum\limits_{t=1}^{n} x_t^2} = \theta + \frac{n^{-1}\sum\limits_{t=1}^{n} x_t \varepsilon_t}{n^{-1}\sum\limits_{t=1}^{n} x_t^2}$$

and thus $\widehat{\theta}_n$ *is seen to be a function of the sample averages of* $x_t\varepsilon_t$ *and* x_t^2*.*

3.1 Independent Processes

In this subsection we discuss LLNs for independent processes.

Theorem 3.2 [8] *(Kolmogorov's strong LLN for i.i.d. random variables) Let* Z_t *be a sequence of identically and independently distributed (i.i.d.) random variables with* $E|Z_1| < \infty$ *and* $EZ_1 = \mu$*. Then* $\bar{Z}_n \overset{a.s.}{\to} \mu$ *(and hence* $\bar{Z}_n \overset{i.p.}{\to} \mu$*) as* $n \to \infty$*.*

We have the following trivial but useful corollary.

Corollary 3.3 *Let* Z_t *be a sequence of i.i.d. random variables, and let* f *be a Borel measurable real function satisfying* $E|f(Z_1)| < \infty$*, then* $n^{-1}\sum_{t=1}^{n} f(Z_t) \overset{a.s.}{\to} Ef(Z_1)$ *as* $n \to \infty$*.*

The corollary, can, in particular be used to establish convergence of sample moments of, say, order p to the corresponding population moment by choosing $f(Z_t) = Z_t^p$.

We now derive the probability limit of the ordinary least squares estimator considered in Example 3.1 as an illustration.

Example 3.4 *Assume the setup of Example 3.1. Assume furthermore that the processes* x_t *and* ε_t *are i.i.d. with* $Ex_t^2 = Q_x$*,* $0 < Q_x < \infty$*,* $E|\varepsilon_t| < \infty$*, and* $E\varepsilon_t = 0$*, and that the two processes are independent of each other. Then* $x_t\varepsilon_t$ *is i.i.d., has finite expectation and satisfies* $Ex_t\varepsilon_t = Ex_t E\varepsilon_t = 0$*. Hence it follows from Theorem 3.2 that* $n^{-1}\sum_{t=1}^{n} x_t\varepsilon_t \overset{a.s.}{\to} 0$*. Corollary 3.3 implies* $n^{-1}\sum_{t=1}^{n} x_t^2 \overset{a.s.}{\to} Q_x$*. Applying Theorem 2.28 then yields* $\widehat{\theta}_n \overset{a.s.}{\to} \theta + 0/Q_x = \theta$*.*

The assumption in Theorem 3.2 that the random variables are identically distributed can be relaxed at the expense of maintaining additional assumptions on the second moments.

[8] See, e.g., Shiryayev (1984), pp. 366.

Theorem 3.5 [9] *(Kolmogorov's strong LLN for i.d. random variables) Let* Z_t *be a sequence of independently distributed (i.d.) random variables with* $EZ_t = \mu_t$ *and* $var(Z_t) = \sigma_t^2 < \infty$. *Suppose* $\sum_{t=1}^{\infty} \sigma_t^2/t^2 < \infty$. *Then* $\bar{Z}_n - \bar{\mu}_n \stackrel{a.s.}{\to} 0$ *as* $n \to \infty$.

The condition $\sum_{t=1}^{\infty} \sigma_t^2/t^2 < \infty$ puts a restriction on the permissible variation in the σ_t^2. For example, it is satisfied if the sequence σ_t^2 is bounded.

3.2 Dependent Processes

The following weak LLN follows immediately from Corollary 2.11. In contrast to the above LLNs this theorem does not require the variables to be independently distributed, but only requires uncorrelatednes.

Theorem 3.6 *(Chebychev's weak LLN for uncorrelated random variables) Let* Z_t *be a sequence of uncorrelated random variables with* $EZ_t = \mu_t$ *and* $var(Z_t) = \sigma_t^2 < \infty$. *Suppose* $var(\bar{Z}_n) = n^{-2} \sum_{t=1}^{n} \sigma_t^2 \to 0$ *as* $n \to \infty$, *then* $\bar{Z}_n - \bar{\mu}_n \stackrel{p}{\to} 0$.

The condition on the variance in Theorem 3.6 is weaker than the corresponding condition in Theorem 3.5 in view of Kronecker's lemma; see, e.g., Shiryayev (1984), p. 365. The condition is clearly satisfied if the sequence σ_t^2 is bounded.

A class of dependent processes that is important in econometrics and statistics is the class of martingale difference sequences. For example, the score of the maximum likelihood estimator evaluated at the true parameter value represents (under mild regularity conditions) a martingale difference sequence.

Definition 3.7 *(Martingale difference sequence) Let* $\mathfrak{F}_t$, $t \geq 0$, *be a sequence of σ-fields such that* $\mathfrak{F}_0 \subseteq \mathfrak{F}_1 \subseteq \ldots \subseteq \mathfrak{F}$. *Let* Z_t, $t \geq 1$, *be a sequence of random variables, then* Z_t *is said to be a martingale difference sequence (w.r.t. the sequence* $\mathfrak{F}_t$*), if* Z_t *is* $\mathfrak{F}_t$*-measurable,* $E|Z_t| < \infty$ *and*

$$E(Z_t \mid \mathfrak{F}_{t-1}) = 0$$

for all $t \geq 1$.

We note that if Z_t is a martingale difference sequence then $E(Z_t) = E(E(Z_t \mid \mathfrak{F}_{t-1})) = 0$ by the law of iterated expectations. Furthermore, since $E(Z_t Z_{t+k}) = E(Z_t E(Z_{t+k} \mid \mathfrak{F}_{t+k-1})) = 0$ for $k \geq 1$, we see that every martingale difference sequence is uncorrelated. We also note, if Z_t is a martingale difference sequence w.r.t. the σ-fields $\mathfrak{F}_t$, then it is also a martingale difference sequence w.r.t. the σ-fields $\mathfrak{G}_t$ where $\mathfrak{G}_t$ is generated by $\{Z_t, Z_{t-1}, \ldots, Z_1\}$ and $\mathfrak{G}_0 = \{\emptyset, \Omega\}$.

We now present a strong LLN for martingale difference sequences.

[9] See, e.g., Shiryayev (1984), pp. 364.

Theorem 3.8 [10] *Let Z_t be a martingale difference sequence with $var(Z_t) = \sigma_t^2 < \infty$. Suppose $\sum_{t=1}^{\infty} \sigma_t^2/t^2 < \infty$. Then $\bar{Z}_n \stackrel{a.s.}{\to} 0$ as $n \to \infty$.*

The above LLN contains Kolmogorov's strong LLN for independent random variables as a special case (with Z_t replaced by $Z_t - \mu_t$).

Many processes of interest in econometrics and statistics are correlated, and hence not covered by the above LLNs. In the following we present a strong LLN for strictly stationary processes, which allow for a wide range of correlation structures.

Definition 3.9 *(Strict stationarity) The sequence of random variables Z_t, $t \geq 1$, is said to be strictly stationary if $(Z_1, Z_2, \ldots, Z_n)$ has the same distribution as $(Z_{1+k}, Z_{2+k}, \ldots, Z_{n+k})$ for all $k \geq 1$ and $n \geq 1$.*

Definition 3.10 [11] *(Invariance and ergodicity of strictly stationary sequences) Let Z_t, $t \geq 1$, be a strictly stationary sequence.*

(a) *Consider the event*

$$A = \{\omega \in \Omega : (Z_1(\omega), Z_2(\omega), \ldots) \in B\}$$

with $B \in \mathfrak{B}^\infty$, where $\mathfrak{B}^\infty$ are the Borel sets of $\mathbf{R}^\infty$. Then A is said to be invariant if

$$A = \{\omega \in \Omega : (Z_{1+k}(\omega), Z_{2+k}(\omega), \ldots) \in B\}$$

for all $k \geq 1$.

(b) *The sequence Z_t is ergodic if every invariant event has probability one or zero.*

We note that every i.i.d. sequence of random variables is strictly stationary and ergodic. Furthermore, if Z_t is strictly stationary and ergodic, and $g : \mathbf{R}^\infty \to \mathbf{R}$ is measurable, then the sequence Y_t with $Y_t = g(Z_t, Z_{t+1}, \ldots)$ is again strictly stationary and ergodic.

We can now give the following strong LLN, which is often referred to as the Ergodic Theorem. This theorem contains Kolmogorov's strong LLN for i.i.d. random variables as a special case.

Theorem 3.11 [12] *Let Z_t be a strictly stationary and ergodic sequence with $E|Z_1| < \infty$ and $EZ_1 = \mu$. Then $\bar{Z}_n \stackrel{a.s.}{\to} \mu$ as $n \to \infty$.*

[10] See, e.g., Shiryayev (1984), pp. 487, or Davidson (1994), p. 314.
[11] See, e.g., Stout (1974), p. 180.
[12] See, e.g., Stout (1974), p. 181.

There is a large literature on LLNs for dependent processes beside the LLNs presented above. LLNs for weakly stationary processes, including linear processes and ARMA processes, can be found in Hannan (1970, Ch. IV.3); see also Phillips and Solo (1992). Important classes of dependent processes considered in econometrics and statistics are α-mixing, ϕ-mixing, near epoch dependent and L_p-approximable processes. LLNs for such processes are discussed in some detail in, e.g., Davidson (1974, Part IV) and Pötscher and Prucha (1997, Ch. 6), and in the references given therein; see also Davidson and de Jong (1997) for recent extensions.

3.3 Uniform Laws of Large Numbers

It is planned to include here a brief discussion of uniform laws of large numbers. Given space constraints, we would appreciate feedback from the editor on whether we should complete or drop this section.

4 Central Limit Theorems

Let Z_t, $t \in \mathbf{N}$, be a sequence of i.i.d. random variables with $EZ_t = \mu$ and $var(Z_t) = \sigma^2$, $0 < \sigma^2 < \infty$. Let $\bar{Z}_n = n^{-1} \sum_{t=1}^{n} Z_t$ denote the sample mean. By Kolmogorov's strong LLN for i.i.d. random variables (Theorem 3.2) it then follows that $\bar{Z}_n - E\bar{Z}_n$ converges to zero a.s. and hence i.p. This implies that the limiting distribution of $\bar{Z}_n - E\bar{Z}_n$ is degenerate at zero, and thus no insight is gained from this limiting distribution regarding the shape of the distribution of the sample mean for finite n; cp. the discussion at the beginning of Section 2.2. Suppose we consider the rescaled quantity

$$\sqrt{n}\left(\bar{Z}_n - E\bar{Z}_n\right) = n^{-1/2} \sum_{t=1}^{n} \left(Z_t - \mu\right). \tag{4.1}$$

Then the variance of the re-scaled expression is $\sigma^2 > 0$ for all n, indicating that its limiting distribution will not be degenerate. Theorems that provide results concerning the limiting distribution of expressions like (4.1) are called central limit theorems (CLTs). Rather than to center the respective random variables, as is done in (4.1), we assume in the following without loss of generality that the respective random variables have mean zero.

4.1 Independent Processes

4.1.1 Some Classical CLTs

In this subsection we will present several classical CLTs, starting with the Lindeberg-Lévy CLT.

Theorem 4.1 [13] *(Lindeberg-Lévy CLT) Let Z_t be a sequence of i.i.d. random variables with $EZ_t = 0$ and $var(Z_t) = \sigma^2 < \infty$. Then $n^{-1/2}\sum_{t=1}^{n} Z_t \xrightarrow{d} N(0, \sigma^2)$. (In case $\sigma^2 = 0$ the limit $N(0, 0)$ should be interpreted as the degenerate distribution having all its probability mass concentrated at zero.)*

Of course, if $\sigma^2 > 0$ the conclusion of the theorem can be written equivalently as $n^{-1/2}\sum_{t=1}^{n} Z_t/\sigma \xrightarrow{d} N(0, 1)$. Extensions of Theorem 4.1 and of any of the following central limit theorems to the vector case are readily obtained using the Cramér-Wold device (Theorem 2.27). To illustrate this we exemplarily extend Theorem 4.1 to the vector case.

Example 4.2 *Let Z_t be a sequence of i.i.d. k-dimensional random vectors with zero mean and finite variance covariance matrix Σ. Let $\xi_n = n^{-1/2}\sum_{t=1}^{n} Z_t$, let $\xi \sim N(0, \Sigma)$ (where $N(0, \Sigma)$ denotes a singular normal distribution if Σ is singular), and let α be some element of $\mathbf{R}^k$. Now consider the scalar random variables $\alpha'\xi_n = n^{-1/2}\sum_{t=1}^{n} \alpha' Z_t$. Clearly the summands $\alpha' Z_t$ are i.i.d. with mean zero and variance $\alpha'\Sigma\alpha$. It hence follows from Theorem 4.1 that $\alpha'\xi_n$ converges in distribution to $N(0, \alpha'\Sigma\alpha)$. Of course $\alpha'\xi \sim N(0, \alpha'\Sigma\alpha)$, and hence $\alpha'\xi_n \xrightarrow{d} \alpha'\xi$. Since α was arbitrary it follows from Theorem 2.27 that $\xi_n \xrightarrow{d} \xi$, which shows that the random vector $n^{-1/2}\sum_{t=1}^{n} Z_t$ converges in distribution to $N(0, \Sigma)$.*

Theorem 4.1 postulates that the random variables Z_t are i.i.d. The following theorems relax this assumption to independence. It proves helpful to define

$$\sigma_{(n)}^2 = \sum_{t=1}^{n} \sigma_t^2 \tag{4.2}$$

where $\sigma_t^2 = var(Z_t)$. For independent Z_t's clearly $\sigma_{(n)}^2 = n^2 var\left(\bar{Z}_n\right)$, and in case the Z_t's are i.i.d. with variance σ^2 we have $\sigma_{(n)}^2 = n\sigma^2$. To connect Theorem 4.1 with the subsequent CLTs observe that within the context of Theorem 4.1 we have $n^{-1/2}\sum_{t=1}^{n} Z_t/\sigma = \sum_{t=1}^{n} Z_t/\sigma_{(n)}$ (given $\sigma^2 > 0$).

Theorem 4.3 [14] *(Lindeberg-Feller CLT) Let Z_t be a sequence of independent random variables with $EZ_t = 0$ and $var(Z_t) = \sigma_t^2 < \infty$. Suppose that $\sigma_{(n)}^2 > 0$, except for finitely many n. If for every $\varepsilon > 0$*

$$\lim_{n\to\infty} \frac{1}{\sigma_{(n)}^2} \sum_{t=1}^{n} E\left[|Z_t|^2 \mathbf{1}\left(|Z_t| \geq \varepsilon\sigma_{(n)}\right)\right] = 0, \tag{L}$$

then $\sum_{t=1}^{n} Z_t/\sigma_{(n)} \xrightarrow{d} N(0, 1)$.

Condition (L) is called the Lindeberg condition. The next theorems employs in place of the Lindeberg condition a condition that is stronger but easier to verify.

Theorem 4.4 [15] *(Lyapounov CLT) Let Z_t be a sequence of independent random*

[13] See, e.g, Billingsley (1979), p. 308.
[14] See, e.g, Billingsley (1979), p. 310.
[15] See, e.g, Billingsley (1979), p. 312.

variables with $EZ_t = 0$ *and* $var(Z_t) = \sigma_t^2 < \infty$. *Suppose that* $\sigma_{(n)}^2 > 0$, *except for finitely many* n. *If for some* $\delta > 0$

$$\lim_{n\to\infty} \sum_{t=1}^{n} E\left|Z_t/\sigma_{(n)}\right|^{2+\delta} = 0, \qquad (P)$$

then $\sum_{t=1}^{n} Z_t/\sigma_{(n)} \xrightarrow{d} N(0,1)$.

Condition (P) is called the Lyapounov condition. Condition (P) implies condition (L). It is readily seen that a sufficient condition for (P) is that

$$n^{-1}\sigma_{(n)}^2 = n^{-1}\sum_{t=1}^{n}\sigma_t^2 \geq const > 0$$

for n sufficiently large and that

$$\lim_{n\to\infty} \sum_{t=1}^{n} E\left|Z_t/\sqrt{n}\right|^{2+\delta} = 0.$$

In turn, sufficient conditions for those two conditions are, respectively,

$$\lim_{n\to\infty} n^{-1}\sigma_{(n)}^2 = \psi, \qquad 0 < \psi < \infty, \tag{4.3}$$

and

$$\sup_n n^{-1}\sum_{t=1}^{n} E\left|Z_t\right|^{2+\delta} < \infty. \tag{4.4}$$

We note that the conclusions of Theorems 4.3 and 4.4 can be stated equivalently as $n^{-1/2}\sum_{t=1}^{n} Z_t \xrightarrow{d} N(0,\psi)$, whenever additionally also (4.3) holds. In this context we also make the trivial observation, that for a sequence of independent random variables Z_t with zero mean and finite variances $\sigma_t^2 \geq 0$ the condition $n^{-1}\sigma_{(n)}^2 \to \psi = 0$ implies $n^{-1/2}\sum_{t=1}^{n} Z_t \xrightarrow{p} 0$ (Corollary 2.11), which can also be rewritten as $n^{-1/2}\sum_{t=1}^{n} Z_t \xrightarrow{d} N(0,\psi)$, $\psi = 0$.

The above CLTs were given for sequences of random variables $(Z_t, t \geq 1)$. They can be readily generalized to cover triangular arrays of random variables $(Z_{tn},\ 1 \leq t \leq n,\ n \geq 1)$. In fact Theorems 4.3 and 4.4 hold with Z_t replaced by Z_{tn} and σ_t^2 replaced by σ_{tn}^2; see, e.g., Billingsley (1979), pp. 310-312.

The need for CLTs for triangular arrays arises frequently in econometrics. One example is the derivation of the limiting distribution of the least squares estimator when different regressors grow at different rates. In this case one can still obtain a limiting normal distribution for the least squares estimator if the usual $\sqrt{n}$-norming is replaced with a normalization by an appropriate diagonal matrix. In essence, this entails renormalizing the i-th regressor by the square root of $\sum_{t=1}^{n} x_{ti}^2$, whose obvious dependence on n leads to the consideration of a CLT for quantities of the form $\sum_{t=1}^{n} c_{tn}u_t$ with u_t i.i.d.; see Theorem 4.6 below.

4.1.2 CLTs for Regression Analysis

In this subsection we present some CLTs that are geared towards regression analysis. As discussed above, within this context we will often need CLTs for a sequence of i.i.d. random variables multiplied by some time-varying scale factors, that may also depend on the sample size. We first give a general CLT that covers such situations as a corollary to the Lindeberg-Feller CLT.

Theorem 4.5 [16] *Let $\underline{Z}_t$ be a sequence of i.i.d. random variables with $E\underline{Z}_t = 0$ and $var(\underline{Z}_t) = 1$. Furthermore, let $(\sigma_{tn},\ 1 \leq t \leq n,\ n \geq 1)$ be a triangular array of real numbers, and define the triangular array Z_{tn} by $Z_{tn} = \sigma_{tn}\underline{Z}_t$. Suppose that $\sigma^2_{(n)} = \sum_{t=1}^n \sigma^2_{tn} > 0$, except for finitely many n. If*

$$\lim_{n\to\infty} \frac{\max_{1\leq t\leq n} \sigma^2_{tn}}{\sum_{t=1}^n \sigma^2_{tn}} = 0, \qquad (M)$$

then $\sum_{t=1}^n Z_{tn}/\sigma_{(n)} \xrightarrow{d} N(0,1)$.

A proof of the above theorem and proofs for the subsequent CLTs are given in the appendix. All of the subsequent CLTs in this section are based on Theorem 4.5.

Theorem 4.6 *Let u_t, $t \geq 1$, be a sequence of i.i.d. random variables with $Eu_t = 0$ and $Eu_t^2 = \sigma^2 < \infty$. Let X_n, $n \geq 1$, with $X_n = (x_{ti})$ be a sequence of real nonstochastic $n \times k$ matrices with*

$$\lim_{n\to\infty} \frac{\max_{1\leq t\leq n} x^2_{ti}}{\sum_{t=1}^n x^2_{ti}} = 0 \text{ for } i = 1, \ldots, k, \qquad (4.5)$$

where it is assumed that $\sum_{t=1}^n x^2_{ti} > 0$ for all but finitely many n. Define $W_n = X_n S_n^{-1}$ where S_n is a $k \times k$ diagonal matrix with the i-th diagonal element equal to $[\sum_{t=1}^n x^2_{ti}]^{1/2}$, and assume that $\lim_{n\to\infty} W_n' W_n = \Phi$ is finite. Let $\mathbf{u}_n = [u_1, \ldots, u_n]'$, then $W_n'\mathbf{u}_n \xrightarrow{d} N(0, \sigma^2\Phi)$.

The above theorem is given in Amemiya (1985), p. 97, for the case of nonsingular $\sigma^2\Phi$.[17] The theorem allows for trending (non-stochastic) regressors. For example, (4.5) holds for $x_{ti} = t^p$, $p > 0$. We note that in case of a single regressor $W_n'W_n = \Phi = 1$.

Theorem 4.7 *Let u_t, $t \geq 1$, be a sequence of i.i.d. random variables with $Eu_t = 0$ and $Eu_t^2 = \sigma^2 < \infty$. Let X_n, $n \geq 1$, with $X_n = (x_{ti})$ be a sequence of real nonstochastic $n \times k$ matrices with $\lim_{n\to\infty} n^{-1}X_n'X_n = Q$ finite. Let $\mathbf{u}_n = [u_1, \ldots, u_n]'$, then $n^{-1/2}X_n'\mathbf{u}_n \xrightarrow{d} N(0, \sigma^2 Q)$.*

[16] The theorem is given as Problem 27.6 in Billingsley (1979), p. 319.

[17] The proof given in Amemiya seems not entirely rigorous in that it does not take into account that the elements of S_n and hence those of W_n depend on the sample size n.

The theorem is, e.g., given in Theil (1971), pp. 380, for the case of nonsingular $\sigma^2 Q$. The theorem does not require that the elements of X_n are bounded in absolute value, as is often assumed in the literature.

We now use Theorems 4.6 and 4.7 to exemplarily give two asymptotic normality results for the least-squares estimator.

Example 4.8 *(Asymptotic normality of the least squares estimator) Consider the linear regression model*

$$y_t = \sum_{i=1}^{k} x_{ti}\beta_i + u_t, \qquad t \geq 1.$$

Suppose u_t and $X_n = (x_{ti})$ satisfy the assumption of Theorem 4.6. Assume furthermore that the matrix Φ in Theorem 4.6 is nonsingular. Then rank $(X_n) = k$ for large n and the least squares estimator for $\beta = (\beta_1, \ldots, \beta_k)'$ is then given by $\hat{\beta}_n = (X_n' X_n)^{-1} X_n' \mathbf{y}_n$ with $\mathbf{y}_n = (y_1, \ldots, y_n)'$. Since $\hat{\beta}_n - \beta = (X_n' X_n)^{-1} X_n' \mathbf{u}_n$ we have

$$S_n \left(\hat{\beta}_n - \beta\right) = S_n \left(X_n' X_n\right)^{-1} S_n' S_n'^{-1} X_n' \mathbf{u}_n = \left(W_n' W_n\right)^{-1} W_n' \mathbf{u}_n.$$

Since $\lim_{n\to\infty} W_n' W_n = \Phi$ and Φ is assumed to be nonsingular, we obtain

$$S_n \left(\hat{\beta}_n - \beta\right) \xrightarrow{d} N\left(0, \sigma^2 \Phi^{-1}\right)$$

as a consequence of Theorem 4.6. Note that this asymptotic normality result allows for trending regressors.

Now suppose that u_t and $X_n = (x_{ti})$ satisfy the assumptions of Theorem 4.7 and that furthermore Q is nonsingular. Then we obtain by similar argumentation

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) = \left(n^{-1} X_n' X_n\right)^{-1} \left(n^{-\frac{1}{2}} X_n' \mathbf{u}_n\right) \xrightarrow{d} N\left(0, \sigma^2 Q^{-1}\right).$$

We note that Theorem 4.7 does not hold in general if the regressors are allowed to be triangular arrays, i.e., the elements are allowed to depend on n. For example, suppose $k = 1$ and $X_n = [x_{11,n}, \ldots, x_{n1,n}]'$ where

$$x_{t1,n} = \begin{cases} 0 & t < n \\ \sqrt{n} & t = n \end{cases},$$

then $n^{-1} X_n' X_n = 1$ and $n^{-1/2} X_n' \mathbf{u}_n = u_n$. The limiting distribution of this expression is just the distribution of the u_t's, and hence not necessarily normal, violating the conclusion of Theorem 4.7.

We now give a CLT where the elements of X_n are allowed to be triangular arrays, but where we assume additionally that the elements of the X_n matrices are bounded in absolute value.

Theorem 4.9 *Let u_t, $t \geq 1$, be a sequence of i.i.d. random variables with $Eu_t = 0$ and $Eu_t^2 = \sigma^2 < \infty$. Let $(x_{ti,n}, 1 \leq t \leq n, n \geq 1)$, $i = 1, \ldots, k$, be triangular arrays of real numbers that are bounded in absolute value, i.e., $\sup_n \sup_{1 \leq t \leq n, 1 \leq i \leq k} |x_{ti,n}| < \infty$. Let $X_n = (x_{ti,n})$ denote corresponding sequences of $n \times k$ real matrices and let $\lim_{n \to \infty} n^{-1} X_n' X_n = Q$ be finite. Furthermore, let $\mathbf{u}_n = [u_1, \ldots, u_n]'$, then $n^{-1/2} X_n' \mathbf{u}_n \xrightarrow{d} N(0, \sigma^2 Q)$.*

Inspection of the proof of Theorem 4.9 shows that the uniform boundedness condition is stronger than is necessary and that it can be replaced by the condition $\max_{1 \leq t \leq n} |x_{ti,n}| = o\left(n^{1/2}\right)$ for $i = 1, \ldots, k$.

4.2 Dependent Processes

There is a large literature on CLTs for dependent processes. Due to space limitation we will only present here – analogously as in our discussion of LLNs – two CLTs for dependent processes. Both CLTs are given for martingale difference sequences. As discussed, martingale difference sequences represent an important class of stochastic processes in statistics. The first of the subsequent two theorems assumes that the process is strictly stationary.

Theorem 4.10 [18] *Let Z_t be a strictly stationary and ergodic martingale difference sequence with $var(Z_t) = \sigma^2 < \infty$. Then $n^{-1/2} \sum_{t=1}^{n} Z_t \xrightarrow{d} N(0, \sigma^2)$. (In case $\sigma^2 = 0$, the limit $N(0,0)$ should be interpreted as the degenerate distribution having all its probability mass concentrated at zero.)*

The above theorem contains the Lindeberg-Lévy CLT for i.i.d. random variables as a special case. The usefulness of Theorem 4.10 is illustrated by the following example.

Example 4.11 *Suppose y_t is a stationary autoregressive process of order one satisfying*

$$y_t = a y_{t-1} + \varepsilon_t,$$

where $|a| < 1$ and the ε_t's are i.i.d. with mean zero and variance σ^2, $0 < \sigma^2 < \infty$. Then $y_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j}$ is strictly stationary and ergodic. The least squares estimator calculated from a sample $y_0, y_1, \ldots, y_n$ is given by $\hat{a}_n = \sum_{t=1}^{n} y_t y_{t-1} / \sum_{t=1}^{n} y_{t-1}^2$ (with the convention that we set $\hat{a} = 0$ on the event $\left\{\sum_{t=0}^{n} y_{t-1}^2 = 0\right\}$). Thus

$$n^{1/2}(\hat{a}_n - a) = \left(n^{-1/2} \sum_{t=1}^{n} \varepsilon_t y_{t-1}\right) / \left(n^{-1} \sum_{t=1}^{n} y_{t-1}^2\right).$$

[18] See, e.g., Gänssler and Stute (1977), p. 372.

The denominator converges a.s. to $E\left(y_{t-1}^2\right) = \sigma^2 / \left(1-a^2\right) > 0$ by the Ergodic Theorem (Theorem 3.11). Observe that $Z_t = \varepsilon_t y_{t-1}$ satisfies $E\left(Z_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\right) = y_{t-1} E\left(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\right) = y_{t-1} E\left(\varepsilon_t\right) = 0$ since y_{t-1} is a (linear) function of $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ and since ε_t is independent of $\left\{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\right\}$. Hence, Z_t is a martingale difference sequence w.r.t. $\mathfrak{F}_t = \sigma\left(\varepsilon_t, \varepsilon_{t-1}, \ldots\right)$. As a function of ε_t and y_{t-1} the sequence Z_t is clearly strictly stationary and ergodic. Furthermore, $var\left(Z_t\right) = E\left(\varepsilon_t^2 y_{t-1}^2\right) = E\left(\varepsilon_t^2\right) E\left(y_{t-1}^2\right) = \sigma^4 / \left(1-a^2\right) < \infty$. Theorem 4.10 then implies

$$n^{-1/2} \sum_{t=1}^{n} \varepsilon_t y_{t-1} \xrightarrow{d} N\left(0, \sigma^4/(1-a^2)\right).$$

Combining this with the already established convergence of the denominator implies the asymptotic normality result

$$n^{1/2}\left(\hat{a}_n - a\right) \xrightarrow{d} N\left(0, 1-a^2\right).$$

Theorem 4.12 [19] *Let Z_t be a martingale difference sequence (w.r.t. $\mathfrak{F}_t$) with conditional variances $E(Z_t^2 \mid \mathfrak{F}_{t-1}) = \sigma_t^2$. Let $\sigma_{(n)}^2 = \sum_{t=1}^{n} \sigma_t^2$. Suppose*

$$n^{-1}\sigma_{(n)}^2 \xrightarrow{p} \psi, \qquad 0 < \psi < \infty, \tag{4.6}$$

and

$$\sum_{t=1}^{n} E\left(\left|Z_t/\sqrt{n}\right|^{2+\delta} \mid \mathfrak{F}_{t-1}\right) \xrightarrow{p} 0 \tag{4.7}$$

as $n \to \infty$ for some $\delta > 0$, then $n^{-1/2} \sum_{t=1}^{n} Z_t \xrightarrow{d} N(0, \psi)$.

Condition (4.7) is a conditional Lyapounov condition. A sufficient condition for (4.7) is

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left(\left|Z_t\right|^{2+\delta} |\mathfrak{F}_{t-1}\right) = O_p(1) \tag{4.8}$$

As mentioned above, there is an enormous body of literature on CLTs for dependent processes. For further CLTs for martingale difference sequences and related results see Hall and Heyde (1980). Central limit theorems for m-dependent and linear processes can, e.g., be found in Hannan (1970, Ch. IV.4), Anderson (1971, Ch. 7.7), or Phillips and Solo (1992). Classical references to central limit theorems for mixingales (including α-mixing, ϕ-mixing and near epoch dependent processes) are McLeish (1974, 1975). For additional discussions of CLTs see, e.g., Davidson (1974, Part V) and Pötscher and Prucha (1997, Ch.10) and the references given therein.

[19] See, e.g., Gänssler and Stute (1977), p. 365 and 370.

5 Further Readings

There is a large number of books available that provide a further in depth discussions of the material (or parts of the material) presented in this article. The list of such books includes texts by Billingsley (1968, 1979), Davidson (1994), Serfling (1980) and Shiryayev (1984), to mention a few. Hall and Heyde (1980) give a thorough discussion of martingale limit theory.

Recent books on asymptotic theory for least mean distance estimators (including maximum likelihood estimators) and generalized method of moments estimators for general classes of nonlinear models include texts by Bierens (1994), Gallant (1987), Gallant and White (1988), Pötscher and Prucha (1997), and White (1994). For recent surveys articles see, e.g., Newey and McFadden (1994), and Wooldridge (1994).

A Appendix

Proof of Theorem 4.5: Clearly, under the assumptions of the theorem the random variables $Z_{1n}, \ldots, Z_{nn}$ are independent for each n, with $EZ_{tn} = 0$ and $var(Z_{tn}) = \sigma^2_{tn} < \infty$. To prove the theorem it hence suffices to show that condition (M) implies condition (L) of Theorem 4.3, which also holds for triangular arrays as mentioned in Section 4.1.1. For brevity write a_n for $\max_{1\leq t\leq n} \sigma_{tn}/\sigma_{(n)}$. Then for $\varepsilon > 0$ we have

$$\begin{aligned} 0 &\leq \frac{1}{\sigma^2_{(n)}} \sum_{t=1}^{n} E\left[Z^2_{tn}\mathbf{1}\left(|Z_{tn}| \geq \varepsilon\sigma_{(n)}\right)\right] \leq \frac{1}{\sigma^2_{(n)}} \sum_{t=1}^{n} \sigma^2_{tn} E\left[\underline{Z}^2_t \mathbf{1}\left(a_n \left|\underline{Z}_t\right| \geq \varepsilon\right)\right] \\ &= E\left[\underline{Z}^2_1 \mathbf{1}\left(a_n \left|\underline{Z}_1\right| \geq \varepsilon\right)\right], \end{aligned}$$

since the $\underline{Z}_t$'s are i.i.d. and $\sigma^2_{(n)} = \sum_{t=1}^{n} \sigma^2_{tn}$. Since $a_n \to 0$ by assumption (M) and $\varepsilon > 0$, the event $\{a_n |\underline{Z}_1| \geq \varepsilon\} \downarrow \emptyset$. Consequently, $\underline{Z}^2_1 \mathbf{1}(a_n |\underline{Z}_1| \geq \varepsilon) \overset{a.s.}{\to} 0$. As this sequence of random vectors is dominated by $\underline{Z}^2_1$ and $E\left(\underline{Z}^2_1\right) < \infty$ holds, the Dominated Convergence Theorem implies $E\left[\underline{Z}^2_1 \mathbf{1}(a_n |\underline{Z}_1| \geq \varepsilon)\right] \to 0$ as $n \to \infty$, which shows that condition (L) holds indeed. ■

Proof of Theorem 4.6: Using the Cramér-Wold device (Theorem 2.27) it suffices to show that $\alpha' W_n' \mathbf{u}_n \overset{d}{\to} N\left(0, \sigma^2 \alpha' \Phi \alpha\right)$ for every $\alpha \in \mathbf{R}^k$. Consider first the case when $\alpha'\Phi\alpha = 0$ or $\sigma^2 = 0$. Then $var(\alpha' W_n' \mathbf{u}_n) = \sigma^2 \alpha' W_n' W_n \alpha \to \sigma^2 \alpha' \Phi \alpha = 0$. Since $E(\alpha' W_n' \mathbf{u}_n) = 0$, Corollary 2.11 implies that $\alpha' W_n' \mathbf{u}_n \overset{p}{\to} 0$ and hence in distribution. By our notational convention this can be written as $\alpha' W_n' \mathbf{u}_n \overset{d}{\to} N(0,0) = N\left(0, \sigma^2\alpha'\Phi\alpha\right)$. Next consider the case where $\sigma^2\alpha'\Phi\alpha > 0$. Then it suffices to establish that $\alpha' W_n' \mathbf{u}_n / \left[\sigma^2 \alpha' \Phi \alpha\right]^{\frac{1}{2}} \overset{d}{\to} N(0,1)$. Define $\underline{Z}_t = u_t/\sigma, Z_{tn} = \sigma_{tn}\underline{Z}_t$, and $\sigma_{tn} = [w_{t.,n}\alpha] / [\alpha'\Phi\alpha]^{1/2}$ with $w_{t.,n} = [w_{t1,n}, \ldots, w_{tk,n}]$ and where $w_{ti,n}$ denotes the (t,i)-th element of the matrix W_n. Then $\alpha' W_n' \mathbf{u}_n / \left[\sigma^2\alpha'\Phi\alpha\right]^{1/2} = \sum_{t=1}^{n} Z_{tn}$. Clearly the $\underline{Z}_t$'s are i.i.d. with zero mean and variance one. Next observe that

$$\sigma^2_{(n)} = \sum_{t=1}^{n} \sigma^2_{tn} = \sum_{t=1}^{n} [\alpha' w'_{t.,n} w_{t.,n} \alpha]/[\alpha'\Phi\alpha] = \alpha' W_n' W_n \alpha / [\alpha'\Phi\alpha] \to 1.$$

Consequently condition (M) holds if $\lim\limits_{n\to\infty} \max_{1\leq t\leq n} \sigma^2_{tn} = 0$. Now

$$\begin{aligned} \sigma^2_{tn} &= [\sum_{i=1}^{k} w_{ti,n}\alpha_i]^2/[\alpha'\Phi\alpha] \\ &\leq [\sum_{i=1}^{k} w^2_{ti,n}][\sum_{i=1}^{k} \alpha^2_i]/[\alpha'\Phi\alpha] \end{aligned}$$

$$= \left[\sum_{i=1}^{k} x_{ti}^2/[\sum_{t=1}^{n} x_{ti}^2]\right]\left[[\sum_{i=1}^{k}\alpha_i^2]/[\alpha'\Phi\alpha]\right]$$

and hence

$$\max_{1\le t\le n}\sigma_{tn}^2 \le \left[\sum_{i=1}^{k}\max_{1\le t\le n} x_{ti}^2/[\sum_{t=1}^{n} x_{ti}^2]\right]\left[[\sum_{i=1}^{k}\alpha_i^2]/[\alpha'\Phi\alpha]\right].$$

Observing that by assumption $\max_{1\le t\le n} x_{ti}^2/[\sum_{t=1}^n x_{ti}^2] \to 0$ it follows that $\max_{1\le t\le n}\sigma_{tn}^2 \to 0$. Given condition (M) holds and $\sigma_{(n)}^2 \to 1$ the result now follows as a special case of Theorem 4.5. ■

Proof of Theorem 4.7: Using the Cramér-Wold device it suffices to show that $n^{-1/2}\alpha' X_n'\mathbf{u}_n \xrightarrow{d} N(0,\sigma^2\alpha' Q\alpha)$ for every $\alpha \in \mathbf{R}^k$. The case where $\alpha' Q\alpha = 0$ or $\sigma^2 = 0$ is treated similarly as in the proof of Theorem 4.7. Next consider the case where $\sigma^2\alpha' Q\alpha > 0$. Then it suffices to establish that $n^{-1/2}\alpha' X_n'\mathbf{u}_n/\left[\sigma^2\alpha' Q\alpha\right]^{1/2} \xrightarrow{d} N(0,1)$. Define $\underline{Z}_t = u_t/\sigma$, $Z_{tn} = \sigma_{tn}\underline{Z}_t$, and $\sigma_{tn} = n^{-1/2}a_t$ with $a_t = [x_{t.}\alpha]/[\alpha' Q\alpha]^{1/2}$ and where $x_{t.} = [x_{t1},\ldots,x_{tk}]$. Then $n^{-1/2}\alpha' X_n'\mathbf{u}_n/[\sigma^2\alpha' Q\alpha]^{1/2} = \sum_{t=1}^n Z_{tn}$. Clearly the $\underline{Z}_t$'s are i.i.d. with zero mean and variance one. Next observe that

$$\begin{aligned}\sigma_{(n)}^2 &= \sum_{t=1}^{n}\sigma_{tn}^2 = n^{-1}\sum_{t=1}^{n}a_t^2 = n^{-1}\sum_{t=1}^{n}[\alpha' x_{t.}' x_{t.}\alpha]/[\alpha' Q\alpha]\\ &= n^{-1}\alpha' X_n' X_n\alpha/[\alpha' Q\alpha] \to 1\end{aligned}$$

and that consequently by Theorem 3.5.5 in Amemiya (1985), p. 98,

$$\lim_{n\to\infty}\frac{\max_{1\le t\le n}\sigma_{tn}^2}{\sum_{t=1}^n\sigma_{tn}^2} = \lim_{n\to\infty}\frac{\max_{1\le t\le n}a_t^2}{\sum_{t=1}^n a_t^2} = 0.$$

That is, condition (M) holds. Since $\sigma_{(n)}^2 \to 1$ the result now follows as a special case of Theorem 4.5. ■

Proof of Theorem 4.9: The proof is similar to the proof of Theorem 4.7. In fact, the case where $\alpha' Q\alpha = 0$ or $\sigma^2 = 0$ is treated identically. Considering now the case where $\sigma^2\alpha' Q\alpha > 0$ it suffices to show that $n^{-1/2}\alpha' X_n'\mathbf{u}_n/[\sigma^2\alpha' Q\alpha]^{1/2} \xrightarrow{d} N(0,1)$. Define $\underline{Z}_t = u_t/\sigma$, $Z_{tn} = \sigma_{tn}\underline{Z}_t$, and $\sigma_{tn} = n^{-1/2}x_{t.,n}\alpha/[\alpha' Q\alpha]^{1/2}$ where $x_{t.,n} = [x_{t1,n},\ldots,x_{tk,n}]$. Then $n^{-1/2}\alpha' X_n'\mathbf{u}_n/[\sigma^2\alpha' Q\alpha]^{1/2} = \sum_{t=1}^n Z_{tn}$. Clearly the random variables $\underline{Z}_t$ are i.i.d. with zero mean and variance one. Next observe that

$$\sigma_{(n)}^2 = \sum_{t=1}^{n}\sigma_{tn}^2 = n^{-1}\sum_{t=1}^{n}\alpha' x_{t.,n}' x_{t.,n}\alpha/[\alpha' Q\alpha] = n^{-1}\alpha' X_n' X_n\alpha/[\alpha' Q\alpha] \to 1.$$

Consequently

$$\lim_{n\to\infty} \frac{\max_{1\le t\le n} \sigma_{tn}^2}{\sum_{t=1}^n \sigma_{tn}^2} = 0$$

will follow, if we can show that $\lim\limits_{n\to\infty} \max_{1\le t\le n} \sigma_{tn}^2 = 0$. Recall that by assumption there exists a finite constant c such that $|x_{ti,n}| \le c$ for all $1 \le t \le n, n \ge 1$, and $1 \le i \le k$. Now

$$\begin{aligned} \max_{1\le t\le n} \sigma_{tn}^2 &= \max_{1\le t\le n} n^{-1}\alpha' x'_{t.,n} x_{t.,n}\alpha/\alpha' Q\alpha = \max_{1\le t\le n} n^{-1}[\sum_{i=1}^k \alpha_i x_{ti,n}]^2/\alpha' Q\alpha \\ &\le \max_{1\le t\le n} n^{-1}\left[\sum_{i=1}^k x_{ti,n}^2\right]\left[\sum_{i=1}^k \alpha_i^2\right] \le n^{-1}kc^2\left[\sum_{i=1}^k \alpha_i^2\right] \to 0. \end{aligned}$$

Since $\sigma_{(n)}^2 \to 1$ the result now follows as a special case of Theorem 4.5. ∎

References

[1] Amemiya, T. *Advanced Econometrics*, Cambridge: Harvard University Press, 1985.

[2] Anderson, T.W. *The Statistical Analysis of Time Series*, New York: Wiley, 1971.

[3] Andrews, D.W.K. Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers. *Econometrica* 55, 1987, 1465-1471.

[4] Bierens, H.J. *Topics in Advanced Econometrics*, Cambridge: Cambridge University Press, 1994.

[5] Billingsley, P. *Convergence of Probability Measures*, New York: Wiley, 1968.

[6] Billingsley, P. *Probability and Measure*, New York: Wiley, 1979.

[7] Davidson, J. *Stochastic Limit Theory*, Oxford: Oxford University Press, 1994.

[8] Davidson, J., and R. de Jong. Strong Laws of Large Numbers for Dependent Heterogeneous Processes: A Synthesis of Recent New Results. *Econometric Reviews* 16, 1997, 251-279.

[9] Fuller, W.A. *Introduction to Statistical Time Series*, New York: Wiley, 1976.

[10] Gallant, A.R. *Nonlinear Statistical Models*, New York: Wiley, 1987.

[11] Gallant, A.R. and H. White. *A Unified Theory of Estimation and Inference in Nonlinear Dynamic Models*, New York: Basil Blackwell, 1988.

[12] Gänssler, P., and W. Stute. *Wahrscheinlichkeitstheorie*, New York: Springer Verlag, 1977.

[13] Hall, P., and C.C. Heyde. *Martingale Limit Theory and Its Application*, New York: Academic Press, 1980.

[14] Hannan, E.J. *Multiple Time Series*, New York: Wiley, 1970.

[15] McLeish, D.L. Dependent Central Limit Theorems and Invariance Principles. *Annals of Probability* 2, 1974, 620-628.

[16] McLeish, D.L. Invariance Principles for Dependent Variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 32, 1975, 165-178.

[17] Newey, W.K. Uniform Convergence in Probability and Stochastic Equicontinuity. *Econometrica* 59, 1991, 1161-1167.

[18] Newey, W.K., and D.L. McFadden. Large Sample Estimation and Hypothesis Testing. In R.F. Engle and D.L. Mc Fadden (eds.), *Handbook of Econometrics*, Vol. IV, pp. 2113-2245, New York: Elsevier Science B.V., 1994.

[19] Phillips, P.C.B., and V. Solo. Asymptotics for Linear Processes. *Annals of Statistics* 20, 1992, 971-1001.

[20] Pötscher, B.M., and I.R. Prucha. A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes. *Econometrica* 57, 1989, 675-683.

[21] Pötscher, B.M., and I.R. Prucha. Generic Uniform Convergence and Equicontinuity Concepts for Random Functions: An Exploration of the Basic Structure. *Journal of Econometrics* 60, 1994, 23-63.

[22] Pötscher, B.M., and I.R. Prucha. *Dynamic Nonlinear Econometric Models, Asymptotic Theory*, New York: Springer Verlag, 1997.

[23] Serfling, R.J. *Approximation Theorems of Mathematical Statistics*, New York: Wiley, 1980.

[24] Shiryayev, A.N. *Probability*, New York: Springer Verlag, 1984.

[25] Stout, W.F. *Almost Sure Convergence*, New York: Academic Press, 1974.

[26] Theil, H. *Principles of Econometrics*, New York, Wiley, 1971.

[27] White, H. *Estimation, Inference and Specification Analysis*, Cambridge: Cambridge University Press, 1994.

[28] Wooldridge, J.M. Estimation and Inference for Dependent Processes. In R.F. Engle and D.L. Mc Fadden (eds.), *Handbook of Econometrics*, Vol. IV, pp. 2641-2738, New York: Elsevier Science B.V., 1994.