Berry-Esseen bounds for econometric time series

Siegfried Hörmann

Département de Mathématique, Université Libre de Bruxelles, Campus Plaine, CP210, B-1050 Brussels - Belgium
E-mail address: `shormann@ulb.ac.be`

Abstract. We derive uniform and non–uniform error bounds in the normal approximation under a general dependence assumption. Our method is tailor made for dynamic time series models employed in the econometric literature but it is also applicable for many other dependent processes. Neither stationarity nor any smoothness conditions of the underlying distributions are required. If the introduced weak dependence coefficient decreases with a geometric rate then we obtain, up to a multiplicative logarithmic factor, the same convergence rate as in the central limit theorem for independent random variables.

1. Introduction

Let $X_1, X_2, \ldots$ be random variables with $EX_k = 0$ and $EX_k^2 < \infty$. Further let $S_n = X_1 + \cdots + X_n$ and $B_n^2 = ES_n^2$. One of the fundamental questions in probability and mathematical statistics is whether such a sequence satisfies the central limit theorem, i.e. whether

$$P(S_n \leq xB_n) \to \Phi(x) \qquad \text{as } n \to \infty, \tag{1.1}$$

where $\Phi(x)$ is the standard normal distribution function. Many basic statistical procedures require more information, such as large deviation probabilities or precise error term estimates (see, e.g., Dufour and Hallin (1992)). Thus it is desirable to determine the speed of convergence in (1.1). Let

$$\Delta_n(x) = |P(S_n \leq xB_n) - \Phi(x)|.$$

In this paper we give bounds for the normal approximation error $\Delta_n(x)$ if the underlying sample $\{X_k\}$ is dependent. Our concept is tailor made for time series studied in finance and macroeconomics. As an application we obtain fairly sharp bounds (with an explicit constant) for ARCH/GARCH processes, threshold autoregressive processes, near epoch dependent (NED) sequences and linear processes with dependent errors, etc. Furthermore, the method is not limited to special time series and might have some general interest.

A common way to measure dependence is to employ different mixing conditions and, as we will show below, several bounds for $\Delta_n(x)$ for mixing sequences already exist. Despite its prominent role and its various "ready to use theorems", mixing

---

Received by the editors , accepted .

2000 *Mathematics Subject Classification.* 60F05, 91B84.

Key words and phrases. Berry–Esseen, m–dependence, weak–dependence, dynamic nonlinear processes.

theory is oftentimes of limited use for the practitioner. For example, many of the processes studied in modern time series literature satisfy mixing conditions only under restrictive smoothness and regularity assumptions. In the past decade this led several authors to develop new dependence concepts which are more convenient in applications (see, e.g., Andrews (1987), Doukhan and Louhichi (1999), Pötscher and Prucha (1997), Wu (2005)). While these new methods have been used to obtain numerous central and empirical central limit theorems, only a few results exist for corresponding convergence rates.

The purpose of this note is to study the convergence rate in the central limit theorem adopting a dependence measure which is tailor made for dynamic non-linear models, i.e. models where the data generating process $\{X_k\}$ is of the form

$$X_k = f_k(\ldots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \ldots), \tag{1.2}$$

where $f_k : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is measurable, and $\{\varepsilon_k\}$ are independent random variables. (Usually ε_k are real valued but our theory applies if the ε_k take values in some more general space.) In context of econometrics $\{\varepsilon_k\}$ corresponds to a process of *exogenous variables* or *disturbances*. We would like to note that while in time series analysis the processes are often assumed to be *causal*, that is

$$X_k = f_k(\ldots, \varepsilon_{k-1}, \varepsilon_k),$$

non-causal processes appear in the theory of *rational expectations* models or in spatial statistics. In most practical examples causal $\{X_k\}$ are homogenous Markov processes obtained as stationary solutions of some stochastic recurrence equations. Oftentimes it is then possible to obtain geometric ergodicity and β–mixing for $\{X_k\}$ by using the theory of Markov models. However, the application of this apparatus requires restrictive smoothness and moment conditions for the innovations. For example, to verify β-mixing for GARCH sequences Carrasco and Chen (2002) require, among others, that the $\{\varepsilon_k\}$ have a continuous density which is positive on the whole line and that $E\varepsilon_k^2 < \infty$. More general assumptions were required by Boussama (1998). With our method we can circumvent mixing theory. In fact, our approach is much more general and has some further advantages over competitors: it does not require causality or stationarity, no smoothness assumptions and is easily verifiable for processes that are given as in (1.2). We demonstrate applicability in Section 4 below by means of various examples. The presented dependence measure is closely related to NED and L^p–approximability which play an important role in the econometrics and financial literature.

For the convenience of the reader we will recall now some well known results for the convergence rate in the CLT if $X_1, X_2, \ldots$ are independent. The Berry–Esseen theorem (Berry (1941), Esseen (1945); see, e.g., Petrov (1995)) states that if the X_k have absolute third moments, then

$$\sup_{-\infty<x<\infty} \Delta_n(x) \leq C B_n^{-3} \sum_{k=1}^{n} E|X_k|^3. \tag{1.3}$$

Here and in the sequel C will denote an absolute constant which may have different values at different places. In case of i.i.d. random variables (1.3) implies a rate $n^{-1/2}$. Employing a symmetric Bernoulli sequence $\{X_k\}$, i.e. $X_k = \pm 1$ with probability $1/2$, it can be easily seen (c.f. Petrov (1995, p. 150)) that this rate cannot be improved without additional conditions on the distribution of the random variables.

Under the more general assumption $E|X_k|^p < \infty$ for some $2 < p \leq 3$ the following version of (1.3) is valid (see Petrov (1965, 1995)):

$$\sup_{-\infty<x<\infty} \Delta_n(x) \leq C(p) B_n^{-p} \sum_{k=1}^{n} E|X_k|^p. \tag{1.4}$$

Here $C(p)$ depends solely on p. In contrast to the uniform estimates (1.3) and (1.4), there are also non–uniform estimates available, which take into account not only the sample size, but also the value of x. Under $E|X_1|^p < \infty$ for some $2 < p \leq 3$, Bikelis (1966) showed that

$$\Delta_n(x) \leq C(p) B_n^{-p} (1+|x|)^{-p} \sum_{k=1}^{n} E|X_k|^p \qquad \text{for all } x \in \mathbb{R}. \tag{1.5}$$

In what follows, we shall give a brief discussion on related results under dependence. For this purpose we recall some classical mixing concepts. Let $\{X_k\}$ be a random sequence and denote by $\mathcal{F}_a^b$ $(-\infty \leq a < b \leq \infty)$ the σ–algebra generated by $X_a, \ldots, X_b$. Then

$$\alpha(n) = \sup\{|P(A \cap B) - P(A) \cdot P(B)| : A \in \mathcal{F}_{-\infty}^k,\ B \in \mathcal{F}_{k+n}^{\infty}\}, \tag{1.6}$$

$$\beta(n) = \frac{1}{2} \sup \Big\{ \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)| : \tag{1.7}$$

$$(A_i)_{i=1}^I \text{ and } (B_j)_{j=1}^J \text{ are finite partitions of } \Omega \text{ with } A_i \in \mathcal{F}_{-\infty}^k,\ B_j \in \mathcal{F}_{k+n}^{\infty} \Big\},$$

$$\rho(n) = \sup\{E|\xi\eta| : \xi \in \mathcal{F}_{-\infty}^k,\ E\xi = 0, E\xi^2 \leq 1,\ \eta \in \mathcal{F}_{k+n}^{\infty},\ E\eta = 0,\ E\eta^2 \leq 1\}, \tag{1.8}$$

$$\varphi(n) = \sup\{|P(B|A) - P(B)| : A \in \mathcal{F}_{-\infty}^k, P(A) > 0, B \in \mathcal{F}_{k+n}^{\infty}\}. \tag{1.9}$$

For stationary $\{X_k\}$ these are independent of k, otherwise the supremum in (1.6)–(1.9) is also taken also over $k \in \mathbb{Z}$. If the corresponding coefficient goes to zero for $n \to \infty$, we say that the sequence is either α, β, ρ, or φ–mixing. We have $\alpha(n) \leq \rho(n) \leq 2\sqrt{\varphi(n)}$ and $2\alpha(n) \leq \beta(n) \leq \varphi(n)$ (for details see, e.g., Bradley (2007) or Doukhan (1994)).

The following results apply to strictly stationary sequences. If $\alpha(n)$ or $\rho(n)$ is $\leq Ke^{-\beta n}$ $(K, \beta > 0)$ Tikhomirov (1980) proved that $\sup_{-\infty<x<\infty} \Delta_n(x) \leq A(\log n)^{p-1} n^{1-p/2}$, where A depends solely on K, β and p. Here and in the sequel the value of $p \in (2,3]$ is related to the moment assumption $E|X_i|^p < \infty$. Under $E|X_1|^3 < \infty$, Bentkus et. al (1997) obtain a bound of order $O\big((\log n)^{\delta} n^{-1/2}\big)$ for a general class of asymptotically normal statistics which are functions of n observations of an absolute regular sequence. Uniform bounds for ρ–mixing sequences with polynomial rate are also given in Zuparov (1992). For φ–mixing sequences Grin obtained a rate of order $O\big((\log n)^{1/3} n^{-1/2}\big)$. Convergence rates of order $O(n^{-1/2})$ so far have only been obtained under more restrictive regularity conditions. See for example Rio (1996) or Bolthausen (1982b).

There exist much less results for non–stationary sequences. The following relaxation of stationarity is due to Sunklodas (1984). Assume that $\{X_k\}$ are α–mixing with geometric rate, and that $B_n^2 \geq cn$ for some $c > 0$. Then

$$\sup_{-\infty<x<\infty} \Delta_n(x) = O\Big(\max_{1\leq k\leq n} E|X_k|^p (\log B_n)^{p-1} B_n^{2-p} \Big).$$

One of the most seminal contributions to the theory of normal approximation is due to Stein (1972). He provides a method going without the previous Fourier analytic approaches, which are difficult to apply under dependence. Chen and Shao (2004) used Stein's method to obtain very sharp results under local dependence. In case of m–dependent sequences they show for example

$$\sup_{-\infty<x<\infty} \Delta_n(x) \leq 75(10m+1)^{p-1} B_n^{-p} \sum_{k=1}^{n} E|X_k|^p,$$

(see Lemmas 5.2–5.3). Their results improve upon previous work for m–dependent sequences, as e.g. those of Tikhomirov (1980), Heinrich (1984) or Sunklodas (1999).

Rates of convergence in the central limit theorem have also been obtained for dependent and not necessarily mixing processes. See e.g. Hall and Heyde (1981), Bolthausen (1982a), Rinott and Rotar (1999), Ouchti (2005) or El Machkouri and Ouchti (2007) for martingales and Birkel (1988) or Dewan and Pakasa Rao (2005) for associated sequences.

The rest of the paper is organized as follows: in Section 2 below we introduce our dependence concept. The main theorems and several applications are stated in Section 3–4. The proofs are given in Section 5.

2. Dependence condition

Despite their prominent role in probability theory, a major disadvantage of diverse mixing concepts is that their verification is difficult in practice. Hence, frequently additional and more restrictive assumptions than actually necessary are imposed on the underlying random sequence $\{X_k\}$, in order to verify a certain mixing condition. Furthermore, in order to apply at all, mixing typically requires strong smoothness conditions on the process. For example, for the AR(1) process

$$X_n = \frac{1}{2} X_{n-1} + \varepsilon_n$$

with independent Bernoulli innovations $\{\varepsilon_n\}$ even the weakest mixing assumption, namely α–mixing, fails to hold (cf. Andrews (1984)). We shall introduce now a dependence concept, which is on the one hand general enough to contain a fairly large class of important processes and which, on the other hand, is easy to verify in practice. The principal idea behind mixing and related weak dependence concepts is the assumption of a fading memory of the process $\{X_k\}$. If the separation m between the two sets of random variables $\{X_k,\, k \leq n\}$ and $\{X_k,\, k > n+m\}$ is large, then the mutual dependence of these sets should be small in some sense. Our idea to formalize this heuristics is given below. We define for $p \geq 0$ $\|X\|_p = (E|X|^p)^{1/p}$. Hence for $p \geq 1$ this is the usual L^p norm. We recall that a sequence $\{Z_k\}$ is called m–dependent, if for each n the two sets of random variables $\{Z_k,\, k \leq n\}$ and $\{Z_k,\, k > n+m\}$ are independent.

Definition 2.1. Let $p > 0$ and let $\{m_n\}$ be a sequence of non–decreasing natural numbers. A random process $\{X_k,\, k \in \mathbb{Z}\}$ is called $\{m_n\}$–approximable in L^p of size $\{a_n\}$, if there exist m–dependent sequences $\{X_{km},\, k \in \mathbb{Z}\}$ $(m = 1, 2, \ldots)$ such that

$$\sum_{k=1}^{n} \|X_k - X_{km_n}\|_p = o(a_n). \tag{2.1}$$

We will write shortly $\{X_k\} \in \mathbf{W}(L^p, \{m_n\}, \{a_n\})$.

Remark 2.2. By Lyapounov's inequality it follows for $0 < q \leq p$ that $\|X_k - X_{km_n}\|_q \leq \|X_k - X_{km_n}\|_p$. Hence, if $a_n = O(a'_n)$ and $m_n \leq m'_n$, then it follows that $\mathbf{W}(L^p, \{m_n\}, \{a_n\}) \subset \mathbf{W}(L^q, \{m'_n\}, \{a'_n\})$.

Remark 2.3. In this paper the most important case is when $a_n = B_n$. Then we shall solely write $\{X_k\} \in \mathbf{W}(L^p, \{m_n\})$.

This alternative method to describe dependence is in the spirit of similar concepts as those in Ibragimov (1962), Billingsley (1968) or McLeish (1975a, 1975b). The crucial idea behind these methods is to approximate the original process with auxiliary processes whose asymptotic is known. In our case m–dependent processes (where $m = m_n$) are used. Martingale approximations have for example been used by Gordin (1969) or Wu (2006). If the approximation error is small enough, then the properties of the auxiliary processes carry over. For example Barbour et al. (2000) showed how iterates of expanding maps can be closely tied to an m–dependent sequence. Their construction also allows to obtain error bounds in the normal approximation for these specific processes.

In order to apply our method, we need a simple way to construct m–dependent approximations for the original sequence. As it has been outlined in the introduction many important processes in the literature have a representation of the form

$$X_k = f_k(\dots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \dots), \tag{2.2}$$

where $\{\varepsilon_k,\, k \in \mathbb{Z}\}$ is a sequence of independent random variables and where $f_k : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ are Borel–measurable. (See also Section 4 for several examples.) Provided $E|X_k| < \infty$, a natural definition for the approximations is

$$X_{km} = E[X_k | \mathcal{F}_{k-m}^{k+m}], \tag{2.3}$$

where $\mathcal{F}_a^b = \sigma(\varepsilon_a, \dots, \varepsilon_b)$. The so defined X_{km} can be represented as

$$X_{km} = f_{km}(\varepsilon_{k-m}, \dots, \varepsilon_k, \dots \varepsilon_{k+m}),$$

where $f_{km} : \mathbb{R}^{2m+1} \to \mathbb{R}$ is measurable and consequently, by the independence of $\{\varepsilon_k\}$ the sequences $\{X_{km},\, k \in \mathbb{Z}\}$ are $2m$–dependent. (Note that if X_{km} are $2m$–dependent, then $X'_{km} = X_{km'}$ with $m' = \lfloor m/2 \rfloor$ are m–dependent. Thus, Definition 2.1 formally applies.) In fact, for $p \geq 1$ the conditional mean $E[X_k|\mathcal{F}_{k-m}^{k+m}]$ is (up to a constant multiplicative factor) the best possible approximation in L^p norm of all $\mathcal{F}_{k-m}^{k+m}$ measurable random variables. To see this, we note that if a random variable $X \in L^p(\Omega, \mathcal{A}, P)$ then for every $\mathcal{M} \subset \mathcal{A}$ we have by Jensen's inequality $E|X|^p = E\big[E[|X|^p|\mathcal{M}]\big] \geq E\big[|E[X|\mathcal{M}]|^p\big]$. Thus by the $\mathcal{F}_{k-m}^{k+m}$ measurability of X_{km} it follows that

$$\|X_{km} - E[X_k|\mathcal{F}_{k-m}^{k+m}]\|_p = \|E[X_{km} - X_k|\mathcal{F}_{k-m}^{k+m}]\|_p \leq \|X_{km} - X_k\|_p.$$

Therefore by the triangular inequality $\|X_k - E[X_k|\mathcal{F}_{k-m}^{k+m}]\|_p \leq 2\|X_k - X_{km}\|_p$. For non–linear functionals the computation of the conditional mean in (2.3) might be difficult, and it is therefore convenient to allow for a more general definition of X_{km}. A simple alternative construction is

$$X_{km} = f_k(\dots, 0, 0, \varepsilon_{k-m}, \dots, \varepsilon_k, \dots \varepsilon_{k+m}, 0, 0, \dots), \tag{2.4}$$

provided this functional is still well defined. (Of course any other constants could be chosen instead of 0 in the above construction.)

Another very useful method to obtain X_{km} in (2.1) is the following coupling method. For each $\ell \in \mathbb{Z}$ we define an independent sequence $\{\varepsilon_k^{(\ell)}, k \in \mathbb{Z}\}$ with $\varepsilon_k^{(\ell)} \overset{\mathcal{L}}{=} \varepsilon_k$ such that the sequences (ε_k), $(\varepsilon_k^{(\ell)})$, $\ell \in \mathbb{Z}$, are mutually independent. This is always possible by enlarging the original probability space. Now set

$$X_{km} = f_k(\ldots, \varepsilon_{k-m-1}^{(k)}, \varepsilon_{k-m}, \ldots, \varepsilon_k, \ldots, \varepsilon_{k+m}, \varepsilon_{k+m+1}^{(k)}, \ldots). \tag{2.5}$$

Obviously the X_{km} are again $2m$–dependent. However, they are no longer $\mathcal{F}_{k-m}^{k+m}$ measurable and thus formally approximation concepts like L^p–approximability or NED (see Section 4.2) are not applicable. One advantage of the coupling method is that the random variables X_{km} have the same marginal distributions as the X_k's. This will be useful here, since some of our results require conditions on the moments $E|X_{km}|^p$. Furthermore, in contrast to (2.4), it is clear that the variables X_{km} are well defined.

The most important case of (2.2) is when $f_k = f$ and $\{\varepsilon_k\}$ are i.i.d. Then $\{X_k\}$ is a stationary and ergodic sequence. In fact, most stationary and ergodic processes in practice can be represented as a shift process of i.i.d. random variables. See Rosenblatt (1959, 1961, 1971) for general sufficient criteria for the representation (2.2). Especially it is well known that (2.2) holds for many popular time series models (cf. Priestley (1988), Stine (1997), Tong (1990)).

We emphasize that in an abstract sense (2.2) is not required for our method, but it is (2.2) that gives the possibility for an easy construction of X_{km} in (2.1).

3. Results

For the rest of the paper we agree on the following notation: $S_n = X_1 + \cdots + X_n$ and $B_n^2 = \mathrm{Var}(S_n)$. The "approximation–depth sequence" (m_n) is assumed to be a sequence of positive and non–decreasing integers. Our results depend crucially on the order of magnitude of the constants

$$e_{p,n,m} = B_n^{-1} \sum_{k=1}^{n} \|X_k - X_{km}\|_p. \tag{3.1}$$

We will assume throughout that $\{X_k\} \in \mathbf{W}(L^p, \{m_n\})$, which implies that $e_{p,n,m_n} \to 0$ as $n \to \infty$. We are now ready to formulate our main results.

Theorem 3.1. *Let $\{X_k\} \in \mathbf{W}(L^p, \{m_n\})$ for some $p \in (2, 3]$. Then if n is sufficiently large we have*

$$\sup_{-\infty<x<\infty} \Delta_n(x) \le 2^p \cdot 76\,(10m_n + 1)^{p-1} \left(B_n^{-p} \sum_{k=1}^{n} E|X_k|^p + e_{p,n,m_n}^p \right) + 2\, e_{p,n,m_n}^{p/(p+1)}. \tag{3.2}$$

We note that the constants in Theorem 3.1 and also in sequel results could be slightly improved. For example $2\, e_{p,n,m_n}^{p/(p+1)}$ in (3.2) can be replaced with $(1 + (2\pi)^{-1/2} + \delta) e_{p,n,m_n}^{p/(p+1)}$, where $\delta > 0$ can be chosen arbitrary small. For the sake of simplicity and ease of notation we work with slightly coarser estimates.

Of course, the order of magnitude of e_{p,n,m_n} depends on the growth speed of m_n. The faster m_n grows the smaller is e_{p,n,m_n}. To get an optimal bound in (3.2) will thus require to correctly balance the speed of growth of e_{p,n,m_n} and m_n. Our applications in Section 4 show that in many important special cases choosing

$m_n = \lfloor H \log n \rfloor$ is optimal in the sense that for sufficiently large H the constants e_{p,n,m_n} satisfy

$$e_{p,n,m_n}^{p/(p+1)} \vee m_n^{p-1} e_{p,n,m_n}^{p} = o\Big(m_n^{p-1} B_n^{-p} \sum_{k=1}^{n} E|X_k|^p \Big).$$

(As usual $\lfloor x \rfloor$ denotes the integer part of the real number x.) We obtain then up to a multiplicative factor $(\log n)^{p-1}$ the same order in the normal approximation as for independent random variables.

Our next result is a non-uniform version of Theorem 3.1. Here the (absolute) constant in the estimate remains undetermined.

Theorem 3.2. *Let $\{X_k\} \in \mathbf{W}(L^p, \{m_n\})$ for some $p \in (2,3]$. Then if n is sufficiently large we have*

$$\Delta_n(x) \le C(1+|x|)^{-p} \times \Bigg\{ m_n^{p-1} \Big(B_n^{-p} \sum_{k=1}^{n} E|X_k|^p + e_{p,n,m_n}^{p} \Big) + (-\log e_{p,n,m_n})^{(p+1)/2} e_{p,n,m_n}^{p/(p+1)} \Bigg\}, \tag{3.3}$$

where C is an absolute constant.

As a matter of fact the computation of e_{p,n,m_n} might be difficult if p is not an integer. In order to get a simpler condition, we shall give now a version of Theorem 3.1 and Theorem 3.2 under the weaker assumption $\{X_k\} \in \mathbf{W}(L^2, \{m_n\})$.

Theorem 3.3. *Let $\{X_k\} \in \mathbf{W}(L^2, \{m_n\})$ and assume that $E|X_k|^p < \infty$ for some $p \in (2,3]$. Assume further that there is a constant D such that*

$$\sum_{k=1}^{n} E|X_{km_n}|^p \le D \sum_{k=1}^{n} E|X_k|^p \quad \textit{ultimately}. \tag{3.4}$$

Then for sufficiently large n

$$\sup_{-\infty<x<\infty} \Delta_n(x) \le 76D\,(10m_n+1)^{p-1} B_n^{-p} \sum_{k=1}^{n} E|X_k|^p + 2\, e_{2,n,m_n}^{2/3}$$

and

$$\Delta_n(x) \le C(1+|x|)^{-2} \Bigg\{ D\, m_n^{p-1} B_n^{-p} \sum_{k=1}^{n} E|X_k|^p + \big(-\log e_{2,n,m_n} \big)^{3/2} e_{2,n,m_n}^{2/3} \Bigg\},$$

where C is an absolute constant.

In return for the weaker assumption $\{X_k\} \in \mathbf{W}(L^2, \{m_n\})$ we have to require the additional condition (3.4) in Theorem 3.3. For example, if the X_{km} are constructed via the coupling method (2.5) then (3.4) is trivially satisfied with $D = 1$. We also notice that if (3.4) holds in Theorem 3.1 or Theorem 3.2, then the factor

$$B_n^{-p} \sum_{k=1}^{n} E|X_k|^p + e_{p,n,m_n}^{p}$$

in (3.2) and (3.3) can be replaced with

$$D\, 2^{-p} \Big(B_n^{-p} \sum_{k=1}^{n} E|X_k|^p \Big).$$

In the following Theorem we consider the important special case where the sequences $\{(X_k, X_{km})\}$ are stationary, for each $m \geq 1$. Here condition (3.4) reduces to

$$E|X_{1m_n}|^p \leq D\, E|X_1|^p \quad \text{ultimately.} \tag{3.5}$$

We will further assume a logarithmic growth rate for m_n, since this is the right approximation–depth for most of our examples in the next section.

Corollary 3.4. *Let* $\{X_k\} \in \mathbf{W}(L^2, \{\lfloor H \log n \rfloor\}, \{n^h\})$*, where* $H > 0$ *and* $h < 2 - 3p/4$*. Assume that for each* $m \geq 1$ *the sequence* $\{(X_k, X_{km})\}$ *is stationary. Let* $E|X_1|^p < \infty$*,* $2 < p \leq 3$*, and assume that* (3.5) *holds. Then*

$$\sigma^2 = EX_1^2 + 2 \sum_{k \geq 2} E(X_1 X_k) \tag{3.6}$$

converges absolutely and if $\sigma^2 > 0$*, then for sufficiently large* n

$$\sup_{-\infty < x < \infty} \Delta_n(x) \leq 76\, E|X_1/\sigma|^p\, D\, \big(10 H \log n\big)^{p-1} n^{1-p/2},$$

and

$$\Delta_n(x) \leq C\, (1 + |x|)^{-2}\ E|X_1/\sigma|^p\, D\, \big(H \log n\big)^{p-1} n^{1-p/2},$$

where C *is an absolute constant.*

4. Applications

4.1. *Iterated random functions.* We briefly outline the construction of Markov chains via iterated random functions as it can be found in the paper of Diaconis and Freedman (1999). We also refer to Diaconis and Freedman (1999) for several interesting applications, ranging from fractal images to queuing theory. The theory is also applicable for many non–linear time series models, like threshold autoregressive models (Tong (1990)), bilinear autoregressive models (Haggan and Ozaki (1981)) or ARCH models (Engle (1982)).

Let S be a complete separable metric space equipped with the metric ρ and let $(\Theta, \mathcal{A}, \mu)$ be a probability space. Further let $\{f_\theta,\, \theta \in \Theta\}$ be a parametric family of measurable functions from S onto itself. We consider now a Markov chain moving around in S according to the following rule: after starting in some x_0 we pick a $\theta_1 \in \Theta$ at random from μ and set $Y_1(x_0) = f_{\theta_1}(x_0)$. Repeating this experiment independently with the new starting point $Y_1(x_0)$ we obtain $Y_2(x_0)$, etc. Hence the process at time n is

$$Y_n(x_0) = f_{\theta_n} \circ f_{\theta_{n-1}} \circ \cdots \circ f_{\theta_1}(x_0).$$

Assuming that the functions f_θ are Lipschitz continuous, i.e.

$$\rho(f_\theta(x), f_\theta(y)) \leq K_\theta \rho(x, y) \quad \text{for all } x, y \in S,$$

Theorem 1.1 in Diaconis and Freedman (1999) gives conditions that imply that the induced Markov chain has a stationary distribution π. Alternative conditions were derived by Wu and Shao (2004). Lemma 4.1 below is immediate from Theorem 2 in Wu and Shao (2004). We let $\{\theta_n\}$ be an i.i.d. sequence with marginals $\theta_n \sim \mu$. Further we let $\{\theta'_n\}$ be an independent copy of $\{\theta_n\}$.

Lemma 4.1. *Assume that there exist* $y_0 \in S$ *and* $\alpha > 0$ *such that* $E\rho(y_0, f_\theta(y_0))^\alpha < \infty$*. Assume further that there is an* $x_0 \in S$*, an* $\alpha > 0$*, and constants* $r = r(\alpha) \in (0,1)$*,* $C = C(\alpha) > 0$ *such that* $E\rho(Y_n(x), Y_n(x_0))^\alpha \le Cr^n \rho(x, x_0)^\alpha$ *for all* $x \in S$ *and* $n \ge 1$*. Then for all* $x \in S$ *the limit*

$$\lim_{k\to\infty} f_{\theta_n} \circ f_{\theta_{n-1}} \circ \cdots \circ f_{\theta_{n-k}}(x)$$

exists almost surely and does not dependent on x*. If* Y_n *denotes the limit, then*

$$Y_n = M(\ldots, \theta_{n-1}, \theta_n)$$

with a measurable function $M : \Theta^{\mathbb{N}} \to S$*. The sequence is strictly stationary and ergodic and satisfies*

$$\begin{aligned} &Y_n = f_{\theta_n}(Y_{n-1});\\ &Y_n \sim \pi;\\ &E\rho(Y_n, M(\ldots, \theta'_{n-m-1}, \theta'_{n-m}, \theta_{n-m+1}, \ldots, \theta_n))^\alpha \le Cr^m. \end{aligned} \tag{4.1}$$

Property (4.1) is called the geometric moment contraction property. In order to get a real valued sequence which inherits the moment contraction property we define $X_n = T(Y_n)$, where $T : S \to \mathbb{R}$ is Lipschitz continuous, i.e. there is a constant L such that $|T(x) - T(y)| \le L\,\rho(x, y)$ for all $x, y \in S$. Proposition 4.2 below is an immediate consequence of Corollary 3.4.

Proposition 4.2. *Assume that the conditions of Lemma 4.1 hold with* $\alpha = 2$*. For some Lipschitz continuous function* $T : S \to \mathbb{R}$ *we define* $X_n = T(Y_n)$ *and we assume that* $E|X_0|^p < \infty$*,* $p \in (2,3]$*. Let* $H > (1 - 3p/4)/\log r$*. Then the series in* (3.6) *converges absolutely and for sufficiently large* n

$$\sup_{-\infty<x<\infty} \Delta_n(x) \le 76\ E|X_1/\sigma|^p\ (10H\log n)^{p-1} n^{1-p/2},$$

and

$$\Delta_n(x) \le C\,(1 + |x|)^{-2}\ E|X_1/\sigma|^p\ (H\log n)^{p-1} n^{1-p/2},$$

where C *is an absolute constant.*

4.2. *NED sequences. Near epoch dependence* (NED) has been successfully used in the econometrics literature to establish weak dependence of many important dynamic time series models (see e.g. Pötscher and Prucha (1997)).

Definition 4.3 (Andrews (1987))**.** Let $\{X_k\}$ and $\{\varepsilon_k\}$ be two random sequences defined on the same probability space. Then the process $\{X_k\}$ is called near epoch dependent (NED) on the basis process $\{\varepsilon_k\}$ if

$$\nu_m = \sup_{k\in\mathbb{Z}} \|X_k - E[X_k|\varepsilon_{k-m}, \ldots, \varepsilon_{k+m}]\|_2 \tag{4.2}$$

tends to zero for $m \to \infty$.

Under NED or the more general concept of L^p–approximability (see Pötscher and Prucha (1997)) (functional) central limit theorems and laws of large numbers have been obtained. We refer to Pötscher and Prucha (1997) for detailed results and further references. Lemma 4.4 below shows when our results apply to NED–sequences.

Lemma 4.4. *Let* ν_m *be given as in* (4.2)*. Assume that* $\{X_k\}$ *is NED on an independent basis sequence* $\{\varepsilon_k\}$*. For any positive sequence* $\{\kappa_n\}$ *with* $\kappa_n \nearrow \infty$ *we have* $\{X_k\} \in \mathbf{W}(L^2, \{m_n\}, \{n\nu_{m_n}\kappa_n\})$*. If* $\nu_{m_n} = o(B_n/n)$ *then* $\{X_k\} \in \mathbf{W}(L^2, \{m_n\})$*.*

Example 1. Let $\{y_k\}$ be a GARCH(1,1) sequence. I.e.

$$y_k = \varepsilon_k \sigma_k \tag{4.3}$$

where

$$\sigma_k^2 = \delta + \alpha\sigma_{k-1}^2 + \beta y_{k-1}^2, \tag{4.4}$$

with i.i.d. $\{\varepsilon_k\}$, $\delta > 0$ and $\alpha, \beta \geq 0$. Nelson (1990) showed that a strictly stationary solution of (4.3) and (4.4) with $E|y_0|^p < \infty$ exists if and only if $\varrho_p := E(\alpha + \beta\varepsilon_0^2)^{p/2} < 1$. The unique solution is given by

$$y_k = \varepsilon_k\sqrt{\delta}\Big(1 + \sum_{\ell=1}^{\infty}\prod_{i=1}^{\ell}\big(\alpha + \beta\varepsilon_{k-i}^2\big)\Big)^{1/2}.$$

Assume that $\varrho_p < 1$ for some $p \in (2,3]$. We notice that then also $\varrho_q < 1$ for all $0 < q \leq p$. Using the independence of the ε_k we obtain with some routine arguments

$$\begin{aligned}
\nu_m &= \|y_0 - E[y_0|\varepsilon_{-m}, \cdots, \varepsilon_0]\|_2 \\
&\leq \Big\|y_0 - \varepsilon_0\sqrt{\delta}\Big(\sum_{\ell=1}^{m}\prod_{i=1}^{\ell}\big(\alpha + \beta\varepsilon_{-i}^2\big)\Big)^{1/2}\Big\|_2 \\
&\leq \sqrt{\delta}\,\|\varepsilon_0\|_2\Big(E\sum_{\ell=m+1}^{\infty}\prod_{i=1}^{\ell}\big(\alpha + \beta\varepsilon_{-i}^2\big)\Big)^{1/2} \\
&= O\big(\varrho_2^{m/2}\big).
\end{aligned}$$

Now let $m_n = \lfloor H\,\log n\rfloor$ where $H > 2(1-3p/4)/\log\varrho_2$. Then Corollary 3.4 applies.

The example shows, that our method can significantly improve upon standard methods (using β–mixing) when applied to GARCH(1,1) sequences. Besides the fact that we do not need any of the smoothness assumptions for the density of the error sequence mentioned in the Introduction, we also get non–unform bounds, and bounds when only $p < 3$ moments exist. As for the uniform bounds we can explicitly determine the absolute constant in the approximation error.

4.3. *Linear processes with dependent innovations.* Let $\{a_i\}$ be a real–valued and absolute summable sequence and define the linear process

$$X_k = \sum_{i=-\infty}^{\infty} a_i Y_{k-i}. \tag{4.5}$$

We are interested in the case where $\{Y_k\}$ is a dependent sequence. Invariance principles for the partial sums of linear processes with dependent innovations have been studied by Wu and Min (2005).

Proposition 4.5. *Let $\{Y_k\}$ be a zero–mean sequence in $\mathbf{W}(L^p, \{m_n\}, \{a_n\})$ for which $M = \sup_{k\in Z}\|Y_k\|_p < \infty$. Let $\{X_k\}$ be defined as in (4.5). Further let $\phi_m = \sup_{k\in\mathbb{Z}}\|Y_k - Y_{km}\|_p$ and*

$$\nu_m = m\phi_m + \sum_{|i|>m}|a_i|.$$

If $\nu_{m_n} = o(B_n/n)$, then $\{X_n\} \in \mathbf{W}(L^p, \{m_n\})$.

Proof. Setting

$$X_{km} = \sum_{i=-m}^{m} a_i Y_{k-i,m},$$

we get $4m$–dependent approximations for X_k. Repeated application of the Minkowski's inequality gives

$$\begin{aligned}\|X_k - X_{km}\|_p &\leq \Big\| \sum_{|i|>m} a_i Y_{k-i} \Big\|_p + \Big\| \sum_{i=-m}^{m} a_i (Y_{k-i} - Y_{k-i,m}) \Big\|_p \\ &\leq M \sum_{|i|>m} |a_i| + (2m+1)\phi_m = O(\nu_m).\end{aligned}$$

□

Example 2. In the econometric literature an important class of linear processes with dependent errors is defined by ARMA(p, q) processes with GARCH–type innovations. I.e. the process $\{X_k,\, k \in \mathbb{Z}\}$ is given by the relation

$$X_k - \phi_1 X_{k-1} - \cdots - \phi_p X_{k-p} = Y_k + \theta_1 Y_{k-1} + \cdots + \theta_q Y_{k-q}, \tag{4.6}$$

with some real coefficients ϕ_i, $i \in \{1, \ldots, p\}$ and θ_j, $j \in \{1, \ldots, q\}$ and it is assumed that $\{Y_k,\, k \in \mathbb{Z}\}$ is some GARCH type process which should not be specified here. If a stationary and causal solution of (4.6) exists, then it can be represented as a linear process $\sum_{i\geq 0} \psi_i Y_{k-i}$ with exponentially decreasing coefficients ψ_i (see Brockwell and Davis (1991)). Similar as we just showed for the GARCH(1,1), we proved in Hörmann (2008) that for a large class of GARCH models (including EGARCH, AGARCH, threshold models etc.) m–dependent approximations $\{Y_{km}\}$ to the original sequence $\{Y_k\}$ can be obtained, such that

$$\|Y_k - Y_{km}\|_2 \leq \text{const} \cdot \varrho^m \qquad (\varrho < 1).$$

Hence, for ARMA processes with a causal representation and errors specified by the GARCH processes given in Hörmann (2008) we get $\nu_m \leq \text{const}\cdot e^{-\delta m}$ for some $\delta > 0$. If $p \in (2, 3]$ moments exist, application of Corollary 3.4 gives $\sup_{-\infty<x<\infty} \Delta_n(x) = O\big((\log n)^{p-1} n^{1-p/2}\big)$.

4.4. *Sums of the form* $\sum f_k(2^k\omega)$. This example serves to demonstrate the applicability of our method outside the time series framework. Let $(f_k)_{k\geq 1}$ be a sequence of measurable functions defined on the unit interval, such that $\int_0^1 f_k(\omega)\, d\omega = 0$ and $\int_0^1 |f_k(\omega)|^p < \infty$ for some $p \in (2, 3]$. In addition we let $\hat{f}_k(\omega)$ be the 1–periodic extension to the positive real line, i.e. $\hat{f}_k(x) = f_k(x - \lfloor x \rfloor)$. Further we set

$$S_n(\omega) = \sum_{k=1}^{n} \hat{f}_k(2^k\omega), \quad \omega \in [0, 1),$$

and $B_n^2 = \int_0^1 S_n^2(\omega)\, d\omega$. Notice that if λ denotes the Lebesgue measure then we have here

$$\Delta_n(x) = |\lambda\{\omega \in [0, 1) : S_n(\omega) \leq x B_n\} - \Phi(x)|.$$

Under the present setup McLeish (1975a) obtained a Kolmogorov type law of large numbers. For $f_k = f$ central and functional central limit theorems have been obtained by Ibragimov (1967) and Billingsley (1968).

We define the modulus of continuity w_f of a function f on the unit interval

$$w_f(\delta) = \sup_{\substack{0 \le s,t < 1 \\ |s-t| < \delta}} |f(s) - f(t)|, \quad 0 < \delta < 1.$$

Proposition 4.6. *Assume that for a sequence $(m_n)_{n \ge 1}$ of positive integers we have*

$$\rho_n = \frac{1}{B_n} \sum_{k=1}^{n} w_{f_k}\big(2^{-m_n}\big) = o(1).$$

Then for sufficiently large n

$$\sup_{-\infty < x < \infty} \Delta_n(x) \le 76\,(10m_n + 1)^{p-1} B_n^{-p} \sum_{k=1}^{n} \int_0^1 |f_k(\omega)|^p \, d\omega + 2\,\rho_n^{p/(p+1)}$$

and

$$\Delta_n(x) \le C(1+|x|)^{-p} \left\{ m_n^{p-1} B_n^{-p} \sum_{k=1}^{n} \int_0^1 |f_k(\omega)|^p \, d\omega + \big(-\log \rho_n \big)^{(p+1)/2} \rho_n^{p/(p+1)} \right\},$$

where C is an absolute constant.

Proof. Let $X_k(\omega) = \hat{f}_k(2^k \omega)$. Define the random variable $\varepsilon_k(\omega)$ equal to the k–th digit in the binary expansion of ω. Ambiguity can be avoided by the convention to take terminating expansions whenever possible. Then $\{\varepsilon_k\}$ is an i.i.d. sequence where ε_k takes values 0 and 1 with probability 1/2. Obviously we have the representation

$$X_k = f_k\Big(\sum_{j=1}^{\infty} \varepsilon_{k+j} 2^{-j} \Big) = M_k(\varepsilon_{k+1}, \varepsilon_{k+2}, \ldots).$$

Using a one–sided version of the coupling construction method, we define m–dependent approximations

$$X_{km} = M_k(\varepsilon_{k+1}, \varepsilon_{k+2}, \ldots, \varepsilon_{k+m}, \varepsilon^{(k)}_{k+m+1}, \varepsilon^{(k)}_{k+m+2}, \ldots).$$

Changing for some $\omega \in [0,1)$ the digits $\varepsilon_k(\omega)$ for $k > m$ will give an ω' with $|\omega - \omega'| \le 2^{-m}$. And therefore

$$|X_k - X_{km}| \le w_{f_k}\big(2^{-m}\big).$$

Thus $e_{p,n,m_n} \le \rho_n$. Applying Theorems 3.1–3.2 directly would yield a little bit weaker result as in Proposition 4.6. Since the X_{km} are constructed via the coupling method we have $E|X_{km}|^p = E|X_k|^p$ for all $k, m \ge 1$. Hence we get sharper estimates in (5.7) and (5.8). □

Example 3. If $f_k = f$, $\int_0^1 |f(\omega)|^p \, d\omega$ is finite and $w_f(h) \le \text{const} \cdot |h|^\beta$, $\beta > 0$, our method yields

$$\Delta_n(x) \le C_1 (1 + |x|)^{-p} (\log n)^{p-1} n^{1-p/2}$$

and

$$\sup_{-\infty < x < \infty} \Delta_n(x) \le C_2 (\log n)^{p-1} n^{1-p/2}.$$

We notice that under these assumptions Ibragimov (1967) obtained the slightly better uniform bound $C_1 (\log n)^{p/2-1} n^{1-p/2}$. See also Ladohin and Moskvin (1971).

5. Proof

In the sequel we let $S_{nm} = X_{1m} + \cdots + X_{nm}$. Without loss of generality we can assume that $EX_k = EX_{km} = 0$ for all $k, m \geq 1$. We set $B_{nm} = E(S_{nm}^2)$ and recall the definition of e_{p,n,m_n} in (3.1). Note that if $\{X_k\}$ is in $\mathbf{W}(L^p, \{m_n\})$ then it follows that $e_{q,n,m_n} \to 0$ for all $0 < q \leq p$. To show our main results we need some preliminary lemmas.

Lemma 5.1. *For every $\delta > 0$, every $m, n \geq 1$ and every $x \in \mathbb{R}$ the following estimate holds:*

$$\begin{aligned} |P(S_n \leq xB_n) - \Phi(x)| &\leq A_0(x,\delta) + A_1(m,n,\delta) \\ &\quad + \max\{A_2(m,n,x,\delta) + A_3(m,n,x,\delta), A_4(m,n,x,\delta) + A_5(m,n,x,\delta)\}, \end{aligned}$$

where

$$\begin{aligned} &A_0(x,\delta) = |\Phi(x) - \Phi\big(x+\delta\big)|; \\ &A_1(m,n,\delta) = P(|S_n - S_{nm}| > \delta B_n); \\ &A_2(m,n,x,\delta) = |P(S_{nm} \leq (x+\delta)B_n) - \Phi\big((x+\delta)B_n/B_{nm}\big)|; \\ &A_3(m,n,x,\delta) = |\Phi\big((x+\delta)B_n/B_{nm}) - \Phi\big((x+\delta)\big)|; \\ &A_4(m,n,x,\delta) = A_2(m,n,x,-\delta) \quad \text{and} \quad A_5(m,n,x,\delta) = A_3(m,n,x,-\delta). \end{aligned}$$

Proof. Since $\{S_n \leq xB_n\} \subset \{S_{nm} \leq (x+\delta)B_n\} \cup \{S_{nm} - S_n > \delta B_n\}$ we obtain

$$P(S_n \leq xB_n) \leq P\big(S_{nm} \leq (x+\delta)B_n\big) + P(|S_n - S_{nm}| > \delta B_n). \tag{5.1}$$

Similarly it follows that

$$P(S_n \leq xB_n) \geq P\big(S_{nm} \leq (x-\delta)B_n\big) - P(|S_n - S_{nm}| > \delta B_n).$$

By (5.1) and the triangular inequality we get

$$\begin{aligned} P(S_n \leq xB_n) - \Phi(x) \leq |P\big(S_{nm} \leq (x+\delta)B_n\big) - \Phi\big((x+\delta)\big)| \\ + A_0(x,\delta) + A_1(m,n,\delta). \end{aligned}$$

Using again the triangular inequality we can split up the first term on the right above in $A_2(m,n,x,\delta) + A_3(m,n,x,\delta)$. With the same argument we obtain a lower bound. Then A_2 has to be replaced with A_4 and A_3 with A_5. □

The next two Lemmas are special cases of Theorem 2.6 in Chen and Shao (2004) and give uniform and nonuniform Berry–Esseen bounds for m–dependent random variables.

Lemma 5.2. *Let $Z_1, Z_2, \ldots, Z_n$ be m–dependent random variables with zero mean and finite $E|Z_i|^p$ for $2 < p \leq 3$. Then*

$$\sup_{-\infty < x < \infty} \Delta_n(x) \leq 75(10m+1)^{p-1} B_n^{-p} \sum_{i=1}^n E|Z_i|^p.$$

Lemma 5.3. *Let $Z_1, Z_2, \ldots, Z_n$ be m–dependent random variables with zero mean and finite $E|Z_i|^p$ for $2 < p \leq 3$. Then there is an absolute constant c_0, such that*

$$\Delta_n(x) \leq c_0(1+|x|)^{-p} m^{p-1} B_n^{-p} \sum_{i=1}^n E|Z_i|^p.$$

Lemma 5.4. *Let* $\{X_k\} \in \mathbf{W}(L^2, \{m_n\})$. *Then*

$$\limsup_{n\to\infty} \frac{1}{e_{2,n,m_n}} \left(\left| \frac{B_n}{B_{nm_n}} - 1 \right| \vee \left| \frac{B_{nm_n}}{B_n} - 1 \right| \right) \leq 2.$$

Especially we have $B_n \sim B_{nm_n}$ *for* $n \to \infty$.

Proof. Note that $|B_n^2 - B_{nm_n}^2| = E|S_n + S_{nm_n}||S_n - S_{nm_n}|$. Hence by some basic inequalities we infer

$$\begin{aligned} |B_n^2 - B_{nm_n}^2| &\leq \|S_n + S_{nm_n}\|_2 \|S_n - S_{nm_n}\|_2 \\ &\leq \big(\|S_n\|_2 + \|S_{nm_n}\|_2\big)\|S_n - S_{nm_n}\|_2 \\ &\leq \big(2\|S_n\|_2 + \|S_n - S_{nm_n}\|_2\big)\|S_n - S_{nm_n}\|_2, \end{aligned}$$

where we used $\|S_{nm_n}\|_2 \leq \|S_n\|_2 + \|S_n - S_{nm_n}\|_2$. From the definition of e_{p,n,m_n} and the Minkowski inequality it follows that

$$\|S_n - S_{nm_n}\|_p \leq e_{p,n,m_n} B_n. \tag{5.2}$$

Hence

$$|B_n^2 - B_{nm_n}^2| \leq (2B_n + e_{2,n,m_n} B_n) e_{2,n,m_n} B_n.$$

The latter relation implies that

$$\begin{aligned} \left| \frac{B_n - B_{nm_n}}{B_n} \right| &\leq \left| \frac{B_n - B_{nm_n}}{B_n} \right| \left| \frac{B_n + B_{nm_n}}{B_n} \right| \\ &\leq (2 + e_{2,n,m_n}) e_{2,n,m_n}. \end{aligned}$$

□

Proof of Theorem 3.2. We use Lemma 5.1 to estimate $\Delta_n(x)$. Since the bound given there is uniform in its parameters we can use $m = m_n$ and

$$\delta = \delta_n(x) = e_{p,n,m_n}^{p/(p+1)}(1 + |x|).$$

With these values we estimate the terms $A_i \quad (i = 1, 2, \ldots, 5)$ of Lemma 5.1.

In order to bound

$$A_0(x, \delta_n(x)) = \big|\Phi(x) - \Phi\big(x + \delta_n(x)\big)\big|,$$

we distinguish two cases. First we assume that $1 + |x| < (-2\log e_{p,n,m_n})^{1/2}$. By the mean value theorem it follows that

$$\begin{aligned} \big|\Phi(x) - \Phi\big(x + \delta_n(x)\big)\big| &\leq (2\pi)^{-1/2}\delta_n(x) \\ &\leq (2\pi)^{-1/2}(1 + |x|)^{-p}(-2\log e_{p,n,m_n})^{(p+1)/2} e_{p,n,m_n}^{p/(p+1)}. \end{aligned}$$

Now we assume that $1 + |x| \geq (-2\log e_{p,n,m_n})^{1/2}$. If $\operatorname{sign}(x) = -1$ then

$$x + \delta_n(x) = -(1 + |x|)(1 - e_{p,n,m_n}^{p/(p+1)}) + 1.$$

Hence we can choose an n_0 which is independent of the eligible x such that $x + \delta_n(x) < 0$ if $n \geq n_0$. Thus

$$\begin{aligned} \big|\Phi(x) - \Phi\big(x + \delta_n(x)\big)\big| &\leq 2\Phi\big(x + \delta_n(x)\big) \\ &\leq 2\Big(1 - \Phi\Big((1 + |x|)(1 - e_{p,n,m_n}^{p/(p+1)}) - 1\Big)\Big) = b_n(x). \end{aligned}$$

If $\mathrm{sign}(x) = 1$ we have $\big|\Phi(x) - \Phi\big(x + \delta_n(x)\big)\big| \le 2\big(1 - \Phi(x)\big) \le b_n(x)$. We recall the well know inequality

$$1 - \Phi(T) \le (2\pi)^{-1/2} \frac{1}{T}\ e^{-T^2/2} \quad \text{for all } T \ge 1.$$

Our assumptions on x and $e_{p,n,m_n} \to 0$ imply that for each $\gamma > 0$ we can choose an $n_1 = n_1(\gamma)$, such that for all $n \ge n_1$

$$(1 + |x|)\big(1 - e_{p,n,m_n}^{p/(p+1)}\big) - 1 > (1 + |x|)(1 - \gamma)^{1/2}.$$

If γ is chosen small enough we obtain

$$\begin{aligned}\big|\Phi(x) - \Phi\big(x + \delta_n(x)\big)\big| &\le (2\pi)^{-1/2} \frac{2}{(1 + |x|)(1 - \gamma)^{1/2}}\ \exp\Big(- (1 + |x|)^2 (1 - \gamma)/2\Big) \\ &\le (1 + |x|)^{-p} (1 + |x|)^{p-1} \exp\Big(- (1 + |x|)^2 (1 - \gamma)/2\Big) \\ &\le (1 + |x|)^{-p} (-2 \log e_{p,n,m_n})^{(p-1)/2} e_{p,n,m_n}^{1-\gamma}.\end{aligned}$$

We chose now $\gamma < 1/(p+1)$ and collect our estimates for $A_0(x, \delta_n(x))$. We conclude that for every $C > (2\pi)^{-1/2}$,

$$A_0(x, \delta_n(x)) \le C(1 + |x|)^{-p} (-2 \log e_{p,n,m_n})^{(p+1)/2} e_{p,n,m_n}^{p/(p+1)} \quad \text{ultimately.} \tag{5.3}$$

By the Markov–inequality and (5.2)

$$\begin{aligned}A_1(m_n, n, \delta_n(x)) &= P(|S_n - S_{nm_n}| > \delta_n(x) B_n) \\ &\le E|S_n - S_{nm_n}|^p (\delta_n(x) B_n)^{-p} \\ &\le e_{p,n,m_n}^p \big(\delta_n(x)\big)^{-p} \\ &= (1 + |x|)^{-p} e_{p,n,m_n}^{p/(p+1)}.\end{aligned} \tag{5.4}$$

According to Lemma 5.3 we have

$$A_2(m_n, n, x, \delta_n(x)) \le c_0 (1 + |H_n(x)|)^{-p} m_n^{p-1} B_{nm_n}^{-p} \sum_{k=1}^{n} E|X_{km_n}|^p, \tag{5.5}$$

where

$$\begin{aligned}1 + |H_n(x)| &= 1 + \left| \frac{(x + \delta_n(x)) B_n}{B_{nm_n}} \right| \\ &\ge 1 + \Big[(1 + |x|)\big(1 + e_{p,n,m_n}^{p/(p+1)}\big) - 1\Big] \frac{B_n}{B_{nm_n}} \\ &\ge 1 - \frac{B_n}{B_{nm_n}} + (1 + |x|) \frac{B_n}{B_{nm_n}}\ .\end{aligned}$$

By Lemma 5.4 it follows that

$$\limsup_{n \to \infty}\ \sup_{x \in \mathbb{R}} \frac{1 + |x|}{1 + |H_n(x)|} \le 1. \tag{5.6}$$

Note that

$$\begin{aligned}E|X_{km_n}|^p &\le (\|X_k\|_p + \|X_{km_n} - X_k\|_p)^p \\ &\le 2^p (\|X_k\|_p^p + \|X_k - X_{km_n}\|_p^p).\end{aligned}$$

Since for any real sequence (a_k) and for any $p \geq 1$ the relation $\sum_{k=1}^n |a_k|^p \leq \big(\sum_{k=1}^n |a_k|\big)^p$ holds, we conclude from the definition of e_{p,n,m_n}

$$\begin{aligned}\sum_{k=1}^n E|X_{km_n}|^p &\leq 2^p \bigg(\sum_{k=1}^n E|X_k|^p + \Big(\sum_{k=1}^n \|X_k - X_{km_n}\|_p\Big)^p\bigg) \\ &= 2^p \bigg(\sum_{k=1}^n E|X_k|^p + e_{p,n,m_n}^p B_n^p\bigg). \end{aligned} \tag{5.7}$$

Hence combining (5.5)–(5.7) and Lemma 5.4 we get for every $C > 2^p c_0$

$$A_2(m_n, n, x, \delta_n(x)) \leq C\big(1+|x|\big)^{-p} m_n^{p-1}\bigg(B_n^{-p}\sum_{k=1}^n E|X_k|^p + e_{p,n,m_n}^p\bigg) \quad \text{ultimately.} \tag{5.8}$$

Next we estimate A_3. By definition we have

$$A_3(m_n, n, x, \delta_n(x)) = \big|\Phi\big((x+\delta_n(x))B_n/B_{nm_n}\big) - \Phi\big(x+\delta_n(x)\big)\big|. \tag{5.9}$$

The mean value theorem and Lemma 5.4 give for $1+|x| \leq (-2\log e_{p,n,m_n})^{1/2}$

$$\begin{aligned}&\big|\Phi\big((x+\delta_n(x))B_n/B_{nm_n}\big) - \Phi\big(x+\delta_n(x)\big)\big| \\ &\quad\leq (2\pi)^{-1/2}\,\big|x+\delta_n(x)\big|\,\bigg|\frac{B_n}{B_{nm_n}} - 1\bigg| \\ &\quad\leq \text{const}\cdot(1+|x|)^{-p}\big(-\log e_{p,n,m_n}\big)^{(p+1)/2} e_{2,n,m_n} \\ &\quad= (1+|x|)^{-p}\ o\big(e_{p,n,m_n}^{p/(p+1)}\big).\end{aligned}$$

Essentially the same arguments we used to estimate A_0 in case of $1+|x| \geq (-2\log e_{p,n,m_n})^{1/2}$ can be used here, to show that the (5.9) is bounded by $(1+|x|)^{-p}\ o\big(e_{p,n,m_n}^{p/(p+1)}\big)$, for $1+|x| \geq (-2\log e_{p,n,m_n})^{1/2}$. Thus we have

$$A_3(m_n, n, x, \delta_n(x)) = (1+|x|)^{-p}\ o\big(e_{p,n,m_n}^{p/(p+1)}\big). \tag{5.10}$$

It is obvious that the terms A_4 and A_5 in Lemma 5.1 can be estimated in exactly the same way as A_2 and A_3. Using the estimates (5.3), (5.4), (5.8) and (5.10), the proof of Theorem 3.2 follows at once from Lemma 5.1. □

Proof of Theorem 3.1. We can use similar (in fact easier) arguments as in the proof of Theorem 3.2. Again we will employ Lemma 5.1, but now we choose $\delta = \delta_n = e_{p,n,m_n}^{p/(p+1)}$. Application of the mean value theorem and the Markov inequality yield

$$\sup_{-\infty<x<\infty} A_0(x, \delta_n) \leq (2\pi)^{-1/2} e_{p,n,m_n}^{p/(p+1)}; \tag{5.11}$$

$$A_1(m_n, n, \delta_n) \leq e_{p,n,m_n}^{p/(p+1)}; \tag{5.12}$$

By Lemma 5.2 it follows that

$$\sup_{-\infty<x<\infty} A_2(m_n, n, x, \delta_n) \leq 75(10m_n+1)^{p-1} B_{nm_n}^{-p}\sum_{k=1}^n E|X_{km_n}|^p;$$

Thus by Lemma 5.4 and (5.7) it is clear that

$$\sup_{-\infty<x<\infty} A_2(m_n, n, x, \delta_n) \leq$$

$$2^p \cdot 76\,(10m_n+1)^{p-1}\left(B_n^{-p}\sum_{k=1}^{n} E|X_k|^p + e_{p,n,m_n}^p\right) \quad \text{ultimately.} \tag{5.13}$$

By the elementary proposition (cf. Petrov (1995, Lemma 5.2))

$$\sup_{-\infty<x<\infty} |\Phi(px)-\Phi(x)| \leq (2\pi e)^{-1/2}\max\{p-1, p^{-1}-1\}$$

we conclude that

$$\sup_{-\infty<x<\infty} A_3(m_n, n, x, \delta_n) \leq (2\pi e)^{-1/2}\left(\left|\frac{B_n}{B_{nm_n}}-1\right| \vee \left|\frac{B_{nm_n}}{B_n}-1\right|\right).$$

Thus by Lemma 5.4 we infer that

$$\sup_{-\infty<x<\infty} A_3(m_n, n, x, \delta_n) = o\big(e_{p,n,m_n}^{p/(p+1)}\big). \tag{5.14}$$

The estimates for A_4 and A_5 are the same as for A_2 and A_3. Collecting our estimates (5.11)–(5.14) and plugging them into Lemma 5.1 finishes the proof. □

Proof of Theorem 3.3. The proof only requires some simple modifications of the proofs of Theorem 3.1–3.2 and will thus be omitted. □

Proof of Corollary 3.4. We first show that

$$\sum_{k=0}^{\infty} |E(X_1X_{k+1})| < \infty. \tag{5.15}$$

Without loss of generality we can assume that $E(X_{km}) = 0$ for all $k \in \mathbb{Z}$, $m \geq 1$. Now write

$$X_1X_{k+1} = (X_1 - X_{1m})X_{k+1} + X_{1m}(X_{k+1} - X_{k+1,m}) + X_{1m}X_{k+1,m}.$$

Using stationarity, (3.5) and the fact that X_{1m} and $X_{k+1,m}$ are independent if $m < k$, we get for $m = k-1$ that

$$\begin{aligned}|E(X_1X_{k+1})| &\leq E|(X_1 - X_{1,k-1})X_{k+1}| + E|X_{1,k-1}(X_{k+1} - X_{k+1,k-1})| \\ &\leq (\|X_1\|_2 + \|X_{1,k-1}\|_2)\|X_1 - X_{1,k-1}\|_2 \\ &\leq \big(1 + D^{1/p}\big)\|X_1\|_p\|X_1 - X_{1,k-1}\|_2 \quad \text{for large enough } k.\end{aligned}$$

Next note that $X_k \in \mathbf{W}(L^2, \{\lfloor H\log n\rfloor\}, \{n^h\})$ and stationarity imply that $\|X_1 - X_{1m_n}\|_2 = o(n^{h-1})$. Let t_n be such that $m_{t_n} = \lfloor H\log t_n\rfloor = n$. Since t_n obviously grows exponentially fast, we conclude that $\|X_1 - X_{1n}\|_2$ converges to zero at an exponential rate, and (5.15) follows.

It is also clear now that $B_n \sim \sigma n^{1/2}$. Further we notice that by our assumptions there is an $\alpha > 0$ such that $e_{2,n,m_n}^{2/3} = o\big(n^{1-p/2-\alpha}\big)$. Hence the proof follows by an application of Theorem 3.3. □

References

D. W. K. Andrews. Nonstrong mixing autoregressive processes. *J. Appl. Probab.* **21**, 930–934 (1984).

D. W. K. Andrews. Laws of large numbers for dependent nonidentically distributed random variables. *Econometrica* **55**, 1465–1471 (1987).

A. D. Barbour, R. M. Gerrard and G. Reinert. Iterates of expanding maps. *Probab. Theory Relat. Fields* **116**, 151–180 (2000).

A. C. Berry. The accuracy of Gaussian approximation to the sum of independent variables. *Trans. Amer. Math. Soc.* **49**, 122–136 (1941).

V. Bentkus, F. Götze, and A. Tikhomirov. Berry–Esseen bounds for statistics of weakly dependent samples. *Bernoulli* **3**, 329–349 (1997).

A. Bikelis. Estimates of the reminder in the central limit theorem. *Litovsk. Mat. Sb.* **6**, 323–346 (1966).

A. Birkel. On the convergence rate in the central limit theorem for associated processes. *Ann. Probab.* **16**, 1685–1698 (1988).

P. Billingsley. *Convergence of Probability Measures.* Wiley (1968).

E. Bolthausen. Exact convergence rates in some martingale central limit theorems. *Ann. Probab.* **10**, 672–688 (1982).

E. Bolthausen. The Berry–Esseén theorem for stronly mixing Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete* **60**, 283–289 (1982).

F. Boussama. *Ergodicité, mélange et estimation dans les modèles GARCH.* Doctoral thesis, Universit Paris 7 (1998).

R. C. Bradley. *Introduction to Strong Mixing Conditions, Vol. I–III.* Kendrick Press (2007).

P. J. Brockwell and R. A. Davis *Time Series: Theory and Methods (2nd ed.)* Springer–Verlag (1991).

M. Carrasco and X. Chen Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18**, 17–39 (2002).

L. Chen and Q. M. Shao. Normal approximation under local dependence. *Ann. Probab.* **32**, 1985–2028 (2004).

I. Dewan and B. L. S. Pakasa Rao. Non–uniform and uniform Berry–Esseen type bounds for stationary associated sequences. *J. Nonparametr. Stat.* **17**, 217–235 (2005).

P. Diaconis and D. Freedman. Iterated random functions. *SIAM Review* **41**, 45–75 (1999).

P. Doukhan. *Mixing: Properties and Examples.* Springer Lecture Notes (1994).

P. Doukhan and S. Louhichi. A new weak dependence condition and applications to moment inequalities. *Stoch. Proc. Appl.* **84**, 313–342 (1999).

J-M. Dufour and M. Hallin. Improved Berry-Esseen-Chebyshev bounds with statistical applications. *Econometric Theory* **8**, 223–240 (1992).

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007 (1982).

M. El Machkouri and L. Ouchti. Exact convergence rates in the central limit theorem for a class of martingales. *Bernoulli* **13**, 981–999 (2007).

C. G. Esséen. Fourier analysis of distribution functions: A mathematical study of the Laplpace–Gaussian law. *Acta Math.* **77**, 1–125 (1945).

M. I. Gordin. The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR* **188**, 739–741 (1969).

A. G. Grin. Refinements of the central limit theorem for sums of dependent random variables. In *Mathematical structures and modeling. No. 12*, 10–17, Omsk. Gos. Univ. (2003).

V. Haggan and T. Ozaki. Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* **68**, 189–196 (1981).

P. Hall and C. C. Heyde. Rates of convergence in the martingale central limit theorem. *Ann. Probab.* **9**, 395–404 (1981).

L. Heinrich. Nonuniform estimates and asymptotic expansions of the remainder in the central limit theorem for m–dependent random variables. *Math. Nachr.* **115**, 7–20 (1984).

S. Hörmann. Augmented GARCH sequences: Dependence structure and asymptotics. *Bernoulli* **14**, 543–561 (2008).

I. A. Ibragimov. Some limit theorems for stationary processes. *Theory Probab. Appl.* **7**, 349–382 (1962).

I. A. Ibragimov. The central limit theorem for sums of functions of independent random variables and sums of the form $\sum f(2^k t)$. *Theory Probab. Appl.* **12**, 596–607 (1967).

V. I. Ladohin and D. A. Moskvin. An estimate for the remainder term in the central limit theorem for sums of functions of independent random variables and sums of the form $\sum f(t2^k)$. *Theory Probab. Appl.* **16**, 116–125 (1971).

D. L. McLeish. A maximal inequality and dependent strong laws. *Ann. Probab.* **3**, 829–839 (1975).

D. L. McLeish. Invariance principles for dependent random variables. *Z. Wahrsch. Verw. Gebiete* **32**, 165–178 (1975).

D. B. Nelson. Stationarity and Persistence in the GARCH(1,1) Model. *Econometric Theory* **6**, 318–334 (1990).

L. Ouchti. On the rate of convergence in the central limit theorem for martingale difference sequences. *Ann. Inst. H. Poincaré Probab. Statist.* **41**, 35–43 (2005).

V. V. Petrov. An estimate of the deviation of the distribution function of a sum of independent random variables from the normal law. *Soviet Math. Doklady* **6**, 242–244 (1965).

V. V. Petrov. *Limit Theorems of Probability Theory.* Oxford Science Publications (1995).

B. M. Pötscher and I. R. Prucha. *Dynamic Nonlinear Econometric Models.* Springer (1997).

M. Priestley. *Nonlinear and Nonstationary Time Series Analysis.* Academic Press (1988).

I. Rinott and V. I. Rotar. Some estimates for the rate of convergence in the CLT for martingales. I. *Theory Probab. Appl.* **43**, 604–619 (1999).

E. Rio. Sur le théorème de Berry–Esseen pour les suites faiblement dépendantes. *Probab. Theory Relat. Fields* **104**, 255–282 (1996).

M. Rosenblatt. Stationary random processes as shifts of functions of independent random variables. *J. Math. Mech.* **8**, 665–681 (1959).

M. Rosenblatt. Independence and dependence. *Proc. 4th Berkeley Sympos. Math. Stat. Prob., Vol. II*, 431–443 (1961).

M. Rosenblatt. *Markov Processes. Structure and Asymptotic Behavior.* Springer (1971).

M. Rosenblatt. *Stationary Sequences and Random Fields.* Birkhäuser (1985).

C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2**, 583–602 (1972).

R. Stine. Nonlinear time series. In S. Kotz, C. Rea, and D. Banks, editors, *Encyclopedia of Statistical Sciences.* Wiley (1997).

Ĭ. Sunklodas. The rate of convergence in the central limit theorem for strongly mixing random variables. *Litovsk. Mat. Sb.* **24**, 174–185 (1984).

Ĭ. Sunklodas. A lower bound for the rate of convergence in the central limit theorem for m–dependent random fields. *Theory. Probab. Appl.* **43**, 162–169 (1999).

A. Tikhomirov. On the convergence rate in the central limit theorem for weakly dependent random variables. *Theory Probab. Appl.* **25**, 790–809 (1980).

H. Tong. *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press (1990).

W. B. Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences, USA* **102**, 14150–14154 (2005).

W. B. Wu. Strong invariance principles for dependent random variables. *Ann. Probab.* **35**, 2294–2320 (2006).

W. B. Wu and W. Min. On linear processes with dependent innovations. *Stoch. Process. Appl.* **115**, 939–958 (2005).

W. B. Wu and X. Shao. Limit theorems for iterated random functions. *J. Appl. Probab.* **41**, 425–436 (2004).

T. M. Zuparov. The rate of convergence in the central limit theorem for weakly dependent variables. *Theory Probab. Appl.* **36**, 783–792 (1992).