Split invariance principles for stationary processes

István Berkes[1], Siegfried Hörmann and Johannes Schauer[2]

Graz University of Technology, Université Libre de Bruxelles
and Graz University of Technology

Abstract

The results of Komlós, Major and Tusnády give optimal Wiener approximation of partial sums of i.i.d. random variables and provide an extremely powerful tool in probability and statistical inference. Recently Wu [52] obtained Wiener approximation of a class of dependent stationary processes with finite p-th moments, $2 < p \leq 4$, with error term $o(n^{1/p}(\log n)^{\gamma})$, $\gamma > 0$, and Liu and Lin [27] removed the logarithmic factor, reaching the Komlós-Major-Tusnády bound $o(n^{1/p})$. No similar results exist for $p > 4$ and in fact, no existing method for dependent approximation yields an a.s. rate better than $o(n^{1/4})$. In this paper we show that allowing a second Wiener component in the approximation, we can get rates near to $o(n^{1/p})$ for arbitrary $p > 2$. This extends the scope of applications of the results essentially, as we illustrate it by proving new limit theorems for increments of stochastic processes and statistical tests for short term (epidemic) changes in stationary processes. Our method works under a general weak dependence condition similar to those in [27] and [52] covering, among others, wide classes of linear and nonlinear time series models and classical dynamical systems.

Keywords: Stationary processes, strong invariance principle, KMT approximation, dependence, increments of partial sums

MSC 2000: 60F17, 60G10, 60G17

[1]) Research supported by FWF grant S9603-N23 and OTKA grants K 67961 and K 81928

[2]) Research supported by FWF grant S9603-N23

1 Introduction

Let $X, X_1, X_2, \ldots$ be i.i.d. random variables with mean 0 and variance 1 and let $S_n = \sum_{k \le n} X_k$. Komlós, Major and Tusnády [24], [25] showed that if $E(e^{t|X|}) < \infty$ for some $t > 0$ then, after suitably enlarging the probability space, there exists a Wiener process $\{W(t),\, t \ge 0\}$ such that

$$S_n = W(n) + O(\log n) \qquad \text{a.s.} \tag{1}$$

Also, if $E|X|^p < \infty$ for some $p > 2$, they proved the approximation

$$S_n = W(n) + o(n^{1/p}) \qquad \text{a.s.} \tag{2}$$

The remainder terms in (1) and (2) are optimal. In the case when only $EX^2 < \infty$ is assumed, Strassen [46] obtained

$$S_n = W(n) + o\big((n \log\log n)^{1/2}\big) \qquad \text{a.s.} \tag{3}$$

Without additional moment assumptions the rate in (3) is also optimal, see Major [28]. Relation (3) is a useful invariance principle for the law of the iterated logarithm; on the other hand, it does not imply the CLT for $\{X_n\}$. This difficulty was removed by Major [29] who showed that under $EX^2 < \infty$ there exists a Wiener process W and a numerical sequence $\tau_n \sim n$ such that

$$S_n = W(\tau_n) + o(n^{1/2}) \qquad \text{a.s.} \tag{4}$$

Thus allowing a slight perturbation of the approximating Wiener process one can reach the remainder term $o(n^{1/p})$ also for $p = 2$, making the result applicable for a wide class of CLT type results. The case of strong approximation under the moment condition $EX^2 h(|X|) < \infty$ where $h(x) = o(x^\varepsilon)$, $x \to \infty$ for any $\varepsilon > 0$, has been cleared up completely by Einmahl [17].

The previous results, which settle the strong approximation problem for i.i.d. random variables with finite variances, provide powerful tools in probability and statistical inference, see e.g. the book of Shorack and Wellner [45]. Starting with Strassen [47], a wide literature has dealt with extensions of the above results for weakly dependent sequences, but the existing results are much weaker than in the i.i.d. case. Recently, however, Wu [52] showed that for a large class of weakly dependent stationary sequences $\{X_n\}$ satisfying $E|X_1|^p < \infty$, $2 < p \le 4$, we have the approximation

$$S_n = W(n) + o\big(n^{1/p}(\log n)^\gamma\big) \qquad \text{a.s.}$$

for some $\gamma > 0$ and Liu and Lin [27] removed the logarithmic factor in the error term, reaching the optimal Komlós-Major-Tusnády bound. The proofs do not work for $p > 4$ and in fact, no existing method for dependent approximation yields an a.s. rate better than $o(n^{1/4})$. On the other hand, many important limit theorems in probability and statistics involve norming sequences smaller than $n^{1/4}$, making such results inaccessible by invariance methods. The purpose of the present paper is to fill this gap and provide a new type of approximation theorem reaching nearly the Komlós-Major-Tusnády rate for any $p > 2$.

As noted above, reaching the error term $o(n^{1/2})$ for i.i.d. sequences with finite variance requires a perturbation of the approximating Wiener process W. In the case of dependent processes we will also need a similar perturbation and, more essentially, we will include a second Wiener process in the approximation, whose scaling factor is smaller than that of W and thus it will not affect the asymptotic behavior of the main term. Specifically, for a large class of weakly dependent stationary processes $\{Y_k\}$ with finite p-th moments, $2 < p < \infty$, we will prove the approximation

$$\sum_{k=1}^{n} Y_k = W_1\big(s_n^2\big) + W_2\big(t_n^2\big) + O\left(n^{\frac{1+\eta}{p}}\right) \quad \text{a.s.,} \tag{5}$$

where $\{W_1(t),\, t \ge 0\}$ and $\{W_2(t),\, t \ge 0\}$ are standard Wiener processes and s_n, t_n are numerical sequences with

$$s_n^2 \sim \sigma^2 n, \quad t_n^2 \sim \sigma^2 n^{\gamma} \quad \text{with } \sigma^2 > 0 \text{ and } 0 < \gamma < 1.$$

The new element in (5) is the term $W_2(t_n^2)$ which, by its smaller scaling, does not disturb the asymptotic properties of $W_1(s_n^2)$. Note that the processes W_1, W_2 are not independent, but this will not present any difficulties in applications. (See also Proposition 1 in the next section.) The number η depends on the weak dependence rate of $\{Y_k\}$ (introduced below), and can be made arbitrarily small under suitable rate conditions.

For $p \ge 1$ let $\|Y\|_p = (E|Y|^p)^{1/p}$ denote the L^p-norm of the random variable Y. If A and B are subsets of $\mathbb{Z}$, we let $d(A,B) = \inf\{|a-b| : a \in A, b \in B\}$.

Definition 1. *Let $\{Y_k, k \in \mathbb{Z}\}$ be a stochastic process, let $p \ge 1$ and let $\delta(m) \to 0$. We say that $\{Y_k, k \in \mathbb{Z}\}$ is* weakly $\mathcal{M}$-dependent in L^p with rate function $\delta(\cdot)$ *if:*

(A) *For any $k \in \mathbb{Z}$, $m \in \mathbb{N}$ one can find a random variable $Y_k^{(m)}$ with finite p-th moment such that*

$$\big\|Y_k - Y_k^{(m)}\big\|_p \le \delta(m).$$

(B) *For any disjoint intervals $I_1, \dots, I_r$ ($r \in \mathbb{N}$) of integers and any positive integers $m_1, \dots, m_r$, the vectors $\{Y_j^{(m_1)}, j \in I_1\}, \dots, \{Y_j^{(m_r)}, j \in I_r\}$ are independent provided $d(I_k, I_l) > \max\{m_k, m_l\}$ for $1 \le k < l \le r$.*

We remark that our dependence condition is naturally preserved under smooth transformations. For example, if $\{Y_k\}$ is weakly $\mathcal{M}$-dependent in L^p with rate $\delta(\cdot)$ and h is a Lipschitz continuous function with Lipschitz constant K, then $\{h(Y_k)\}$ is also weakly $\mathcal{M}$-dependent in L^p with rate function $K\delta(\cdot)$.

Note that (B) implies that for any fixed m the sequence $\{Y_k^{(m)}, k \in \mathbb{Z}\}$ is an m-dependent process. Hence, sequences satisfying conditions (A) and (B) are approximable, in the L^p sense, by m-dependent processes of any fixed order $m \ge 1$

with termwise approximation error $\delta(m)$. In other words, sequences in Definition 1 are close to m-dependent sequences, the value of m depending on the required closeness, explaining the terminology. Since $\|Y_k\|_p \le \|Y_k^{(m)}\|_p + \|Y_k - Y_k^{(m)}\|_p$, condition (A) implies that $E|Y_k|^p$ is finite. Using L^p–distance is convenient for our theorems, but, depending on the application, other distances can be used in part (A) of Definition 1. For example, defining (as usual) the L_0 norm of a random variable X by

$$\|X\|_0 = \inf\{\varepsilon > 0 : P(|X| \ge \varepsilon) < \varepsilon\},$$

(A) could be replaced by

$$\left\|Y_k - Y_k^{(m)}\right\|_0 \le \delta(m).$$

Such a definition requires no moment assumptions and turns out to provide a useful dependence measure for studying empirical processes (see [3]).

Trivially the previous definition covers m-dependent processes for any fixed m (see also Subsection 3.1), but, in contrast to the very restrictive condition of m-dependence, weak $\mathcal{M}$-dependence holds for a huge class of stationary sequences, including those studied in Wu [50], [52] and Liu and Lin [27]. In the case when $\{Y_k, k \in \mathbb{Z}\}$ allows a Wiener-Rosenblatt representation

$$Y_k = f(\varepsilon_k, \varepsilon_{k-1}, \ldots), \qquad k \in \mathbb{Z} \tag{6}$$

with an i.i.d. sequence $\{\varepsilon_k,\ k \in \mathbb{Z}\}$, weak $\mathcal{M}$-dependence is very close to Wu's physical dependence condition in [52], except that we allow a larger freedom in choosing the approximating random variables $Y_k^{(m)}$, compared with the choice in [50], [52] via coupling. (For sufficient criteria for the representation (6), see Rosenblatt [39], [40], [41].) Note that instead of (6) we may also assume a two-sided representation

$$Y_k = f(\ldots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \ldots), \qquad k \in \mathbb{Z} \tag{7}$$

of $\{Y_k\}$. In case when $\{Y_k, k \in \mathbb{Z}\}$ allows the representation (7) with mixing $\{\varepsilon_k\}$, Definition 1 is a modified version of NED (see Subsection 3.2), a weak dependence condition which appeared already in Ibragimov [21] and has been brought forward in Billingsley [5] (see also [30], [31]). Later NED has been successfully used in the econometrics literature to establish weak dependence of dynamic time series models (see e.g. [34]). In Section 3 we will discuss further the connection between weak $\mathcal{M}$-dependence with known weak dependence conditions. We stress that the definition of weak $\mathcal{M}$-dependence does not assume the representation (6) or (7), although it was motivated by this case. The reason for using our more general definition is to illuminate the essential structural condition on $\{Y_k\}$ required for our theorems.

The main results of our paper are formulated in Section 2. In Section 3 we give several examples. Applications of the theorems can be found in Section 4 and Section 5, while Section 6 contains the proofs of the main theorems.

2 Main Theorems

We write $a_n \ll b_n$ if $\overline{\lim}_{n\to\infty} |a_n/b_n| < \infty$.

Theorem 1. *Let $p > 2$, $\eta > 0$ and let $\{Y_k, k \in \mathbb{Z}\}$ be a centered stationary sequence, weakly $\mathcal{M}$-dependent in L^p with rate function*

$$\delta(m) \ll m^{-A}, \tag{8}$$

where

$$A > \frac{p-2}{2\eta}\left(1 - \frac{1+\eta}{p}\right) \vee 1, \quad (1+\eta)/p < 1/2. \tag{9}$$

Then the series

$$\sigma^2 = \sum_{k \in \mathbb{Z}} EY_0 Y_k \tag{10}$$

is absolutely convergent and $\{Y_k, k \in \mathbb{Z}\}$ can be redefined on a new probability space together with two standard Wiener processes $\{W_1(t),\, t \geq 0\}$ and $\{W_2(t),\, t \geq 0\}$ such that

$$\sum_{k=1}^{n} Y_k = W_1(s_n^2) + W_2(t_n^2) + O\left(n^{\frac{1+\eta}{p}}\right) \quad \text{a.s.}, \tag{11}$$

where $\{s_n\}$ and $\{t_n\}$ are nondecreasing numerical sequences with

$$s_n^2 = \sigma^2 n + O(n^{1-\epsilon}), \qquad t_n^2 \sim \sigma^2 n^\gamma \tag{12}$$

for some $0 < \gamma, \epsilon < 1$.

Note that for any fixed $p > 2$ and $\eta > 0$, condition (9) applies if A is large enough, providing a remainder term close to the optimal remainder term $o(n^{1/p})$ in the Komlós-Major-Tusnády approximation.

As the proof of Theorem 1 will show, the sequences $\{s_n\}$ and $\{t_n\}$ in (11) have a complementary character. More precisely, there is a partition $\mathbb{N} = G_1 \cup G_2$ (provided by the long and short blocks in a traditional blocking argument) and a representation

$$s_n^2 = \sum_{k=1}^{n} \sigma_k^2, \qquad t_n^2 = \sum_{k=1}^{n} \tau_k^2 \qquad (n = 1, 2, \ldots)$$

such that σ_k^2 converges to σ^2 on G_1 and equals 0 on G_2, and τ_k^2 converges to σ^2 on G_2 and equals 0 on G_1. In particular,

$$\overline{\lim}_{n\to\infty}\,(s_{n+1}^2 - s_n^2) = \overline{\lim}_{n\to\infty}\,(t_{n+1}^2 - t_n^2) = \sigma^2, \tag{13}$$

and both liminf's are equal to 0.

It is natural to ask if $W_1(s_n^2)$ in (11) can be replaced by $W_1(\sigma^2 n)$, a fact that would considerably simplify applications. In the case when the Y_k are uncorrelated with variance 1, we have $\sigma^2 = 1$ and our proof yields $s_n^2 = n - c_n n^\gamma$ for infinitely many n (namely at the end of the long blocks), where $c_n \to c$ for some positive constant c. For small η (which is the interesting case), γ is close to 1, see (14) below. Thus replacing $W_1(s_n^2)$ by $W_1(n)$ introduces an additional error term that ruins the error term $O(n^{(1+\eta)/p})$ in (11). Note again the irregular character of $s_{n+1}^2 - s_n^2$, having both σ^2 and 0 as limit points. The situation is similar to the Wiener approximation of partial sums of i.i.d. random variables under assuming only finite second moments when we have (4) with a numerical sequence $\tau_n \sim n$, but in general (4) does not hold with $\tau_n = n$. (See Major [28], [29].) Note, however, that in our case the large difference between s_n^2 and n is a consequence of the method and we do not claim that another construction cannot yield the approximation (11) with $s_n^2 = \sigma^2 n$. However, the presence of s_n^2 in (11) does not limit the applicability of our strong invariance principle: s_n^2 and t_n^2 are explicitly calculable nonrandom numbers and as we will see, applying limit theorems for $W_1(s_n^2)$ is as easy as for $W(n)$.

The numerical value of γ in (12) plays no role in the applications in this paper, but for later applications we note that if

$$A > \frac{p-2}{2\eta(1-\varepsilon_0)^2}\Big(1 - \frac{1+\eta}{p}\Big) \vee 1$$

for some $0 < \varepsilon_0 < 1$, then we can choose

$$\gamma = 1 - \varepsilon_0 \frac{2\eta(1-\varepsilon_0)}{p - 2(1+\eta\varepsilon_0)}. \tag{14}$$

As we already mentioned in the Introduction, the processes W_1 and W_2 are not independent. While for our applications this is not important, the following Proposition might be useful for possible further applications.

Proposition 1. *Under the assumptions of Theorem 1 we have*

$$\mathrm{Corr}(W_1(s_n), W_2(t_m)) \to 0, \quad \text{as } m, n \to \infty. \tag{15}$$

Our next theorem is the analogue of Theorem 1 in the case of an exponential decay in the dependence condition.

Theorem 2. *Let $p > 2$ and let $\{Y_k, k \in \mathbb{Z}\}$ be a centered stationary sequence, weakly $\mathcal{M}$-dependent in L^p with rate function*

$$\delta(m) \ll \exp(-\varrho m), \quad \varrho > 0. \tag{16}$$

Then the series (10) *is absolutely convergent and $\{Y_k, k \in \mathbb{Z}\}$ can be redefined on a new probability space together with two standard Wiener processes $\{W_1(t), t \geq 0\}$ and $\{W_2(t), t \geq 0\}$ such that*

$$\sum_{k=1}^{n} Y_k = W_1\big(s_n^2\big) + W_2\big(t_n^2\big) + O\big(n^{1/p}\log^2 n\big) \quad \text{a.s.}, \tag{17}$$

where $\{s_n\}$ and $\{t_n\}$ are nondecreasing numerical sequences such that $s_n^2 = \sigma^2 n + O(n/\log n)$, $t_n^2 \sim \sigma^2 n/\log n$ and (13) holds.

Using the law of the iterated logarithm for W_2, relation (11) implies

$$\sum_{k=1}^{n} Y_k = W_1\big(s_n^2\big) + O\big(n^{1/2-\lambda}\big) \quad \text{a.s.}, \tag{18}$$

for some $\lambda > 0$, which is the standard form of strong invariance principles. However, since γ in (12) is typically near to 1, the λ in (18) can be very small, and thus the effect of the very strong error term $O(n^{(1+\eta)/p})$ in (11) is lost.

The proof of the strong approximation theorems in Wu [52] depends on martingale approximation, while Liu and Lin [27] use approximation of the partial sums of $\{Y_k\}$ by partial sums of m-dependent r.v.'s. Our approach differs from both, using a direct approximation of separated block sums of $\{Y_k\}$ by independent r.v.'s, an idea used earlier in [2], [3], [4], [20]. In this approach, the second Wiener process W_2 is provided by the sum of short block sums. As we just pointed out, the method yields a satisfactory remainder term for all $p > 2$, but if we replace $W_2(t_n^2)$ by the upper bound provided by the law of the iterated logarithm, the resulting remainder term will be rather poor. The question if one can get a remainder term near $O(n^{1/p})$ in the simple (one-term) Wiener approximation for any $p > 2$ remains open.

3 Examples of weakly $\mathcal{M}$-dependent processes

The classical approach to weak dependence, developed in the seminal papers of Rosenblatt [38] and Ibragimov [21], uses the strong mixing property and its variants like β, ϱ, ϕ and ψ mixing, combined with a blocking technique to connect the partial sum behavior of $\{Y_k\}$ with that of independent random variables. This method yields very sharp results (for a complete account of the classical theory see Bradley [7]), but verifying mixing conditions of the above type is not easy and even when they apply (e.g. for Markov processes), they typically require strong smoothness conditions on the process. For example, for the AR(1) process

$$Y_k = \frac{1}{2} Y_{k-1} + \varepsilon_k$$

with Bernoulli innovations, strong mixing fails to hold (cf. Andrews [1]). Recognizing this fact, an important line of research in probability theory in past years has been to find weak dependence conditions which are strong enough to imply satisfactory asymptotic results, but which are sufficiently general to be satisfied in typical applications. Several conditions of this kind have been found, in particular by the French school, see [10], [11], [12], [15], [36], [37]. A different type of mixing conditions, the so-called physical and predictive dependence measures have been introduced by Wu [50] for stationary processes $\{Y_k\}$ admitting the representation (6) where $\{\varepsilon_k,\, k \in \mathbb{Z}\}$ is an i.i.d. sequence and $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ is a Borel-measurable function. These conditions are particularly easy to handle, since they are defined in terms of the algorithms which generate the process $\{Y_k\}$. Weak $\mathcal{M}$-dependence, although formally not requiring a representation of the form (6), is closely related to Wu's mixing conditions and works best for processes $\{Y_k\}$ having a representation (6) or its two-sided version (7). The examples below will clear up the exact connection of our weak $\mathcal{M}$-dependence condition with the mixing conditions in Wu [50], [52] and Liu and Lin [27].

3.1 m-dependent processes

Definition 1 implies that $\{Y_k,\, k \in \mathbb{Z}\}$ can be approximated, for every $m \geq 1$, by an m-dependent process with termwise L^p error $\delta(m)$. If $\{Y_k,\, k \in \mathbb{Z}\}$ itself is m-dependent for some fixed $m = m_0$ and $K := \sup_{k\in\mathbb{Z}} \|Y_k\|_p < \infty$, then Definition 1 is satisfied with

$$\delta(j) = \begin{cases} K & \text{if } j < m_0 \\ 0 & \text{if } j \geq m_0, \end{cases}$$

and $Y_k^{(n)} = 0$ if $n < m_0$ and $Y_k^{(n)} = Y_k$ if $n \geq m_0$. In other words, m-dependent sequences with uniformly bounded L^p norms are weakly $\mathcal{M}$-dependent with the above parameters. It is worth mentioning that m-dependent processes in general do not have representation (7) (see e.g. [8], [48]).

3.2 NED processes

Under (7) our condition can be directly compared to NED. We recall the

Definition 2 (NED)**.** *A sequence $\{Y_k,\, k \in \mathbb{Z}\}$ having representation* (7) *is called* NED over $\{\varepsilon_k\}$ under L^p-norm with rate function $\delta(\cdot)$ *if for any $k \in \mathbb{Z}$, $m \geq 1$,*

$$\left\|Y_k - E\left[Y_k|\mathcal{F}_{k-m}^{k+m}\right]\right\|_p \leq \delta(m),$$

where $\mathcal{F}_{k-m}^{k+m}$ is the σ-algebra generated by $\varepsilon_{k-m}, \ldots, \varepsilon_{k+m}$.

Clearly, if $\{\varepsilon_k\}$ is an independent sequence, then $Y_k^{(m)} = E\left[Y_k|\mathcal{F}_{k-m}^{k+m}\right]$ satisfies (B) of Definition 1. Hence if $\{Y_k\}$ is NED over $\{\varepsilon_k\}$ in L^p-norm with rate function $\delta(\cdot)$ where $\{\varepsilon_k\}$ is an independent sequence, then $\{Y_k\}$ is weakly $\mathcal{M}$-dependent with the same p, $\delta(\cdot)$.

As our examples below will show, for weakly $\mathcal{M}$-dependent sequences the construction for $Y_k^{(m)}$ is not restricted to $E\left[Y_k|\mathcal{F}_{k-m}^{k+m}\right]$, but is often more conveniently established by truncation or coupling methods.

3.3 Linear processes

Let $Y_k = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{k-j}$ with i.i.d. innovations $\{\varepsilon_j, j \in \mathbb{Z}\}$. If $a_j = 0$ for $j < 0$, then the sequence $\{Y_k, k \in \mathbb{Z}\}$ is causal. Liu and Lin [27] and Wang, Lin and Gulati [49] studied strong approximations of the partial sums with Gaussian processes (in the short- and long-memory cases).

We define $Y_k^{(m)}$ as

$$Y_k^{(m)} = \sum_{j=-\lfloor m/2\rfloor}^{\lfloor m/2\rfloor} a_j \varepsilon_{k-j}.$$

This directly ensures that condition (B) holds. To verify condition (8) we will assume that $E|\varepsilon_0|^p < \infty$ for some $p > 2$ as well as $|a_j| \ll |j|^{-(A+1)}$ $(j \to \infty)$. Then we get, using the Minkowski inequality,

$$\begin{aligned}
\|Y_k - Y_k^{(m)}\|_p &= \left\| \sum_{|j|>m/2} a_j \varepsilon_{k-j} \right\|_p \\
&\le \sum_{|j|>m/2} \|a_j \varepsilon_{k-j}\|_p \\
&= (E|\varepsilon_0|^p)^{1/p} \sum_{|j|>m/2} |a_j| \ll m^{-A}.
\end{aligned}$$

Thus if A is large enough, Theorem 1 applies. Obviously if $|a_j| \ll \rho^{|j|}$ with some $0 < \rho < 1$ then (16) holds and Theorem 2 applies.

3.4 Nonlinear time series

Let the time series $\{Y_k, k \in \mathbb{Z}\}$ be defined by the stochastic recurrence equation

$$Y_k = G(Y_{k-1}, \varepsilon_k), \tag{19}$$

where G is a measurable function and $\{\varepsilon_k, k \in \mathbb{Z}\}$ is an i.i.d. sequence. For example, ARCH(1) processes (see e.g. Engle [18]) which play an important role in econometrics literature, are included in this setting. Sufficient conditions for the existence of a stationary solution of (19) can be found in Diaconis and Freedman [13]. Note that iterating (19) yields $Y_k = f(\ldots, \varepsilon_{k-1}, \varepsilon_k)$ for some measurable function f. This suggests defining the approximating random variables $Y_k^{(m)}$ as $Y_k^{(m)} = f(\ldots, 0, 0, \varepsilon_{k-m}, \ldots, \varepsilon_k)$. Note, however, that this definition does not guarantee the convergence and thus the existence of $Y_k^{(m)}$. The coupling used by Wu [50], avoids this problem by defining

$$Y_k^{(m)} = f(\ldots, \varepsilon_{k-m-2}^{(k)}, \varepsilon_{k-m-1}^{(k)}, \varepsilon_{k-m}, \ldots, \varepsilon_k),$$

where $\{\varepsilon_k^{(l)}, k \in \mathbb{Z}\}$, $l = 1, 2, \ldots$, are i.i.d. sequences with the same distribution as $\{\varepsilon_k, k \in \mathbb{Z}\}$ that are mutually independent. These random variables satisfy condition (B). Results from Shao and Wu [44] show that under some simple technical assumption on G

$$\left\|Y_k - Y_k^{(m)}\right\|_p \ll \exp(-\rho m)$$

holds with some $p > 0$ and $\rho > 0$. Thus for $p > 2$, Theorem 2 applies.

3.5 Augmented GARCH sequences

Augmented GARCH sequences were introduced by Duan [16] and turned out to be very useful in applications in macroeconomics and finance. The model is quite general and many popular processes are included in its framework. Among others the well known GARCH ([6]), AGARCH ([14]) and EGARCH model ([33]) are covered.

We consider the special case of augmented GARCH(1,1) sequences, i.e. sequences $\{Y_k, k \in \mathbb{Z}\}$ defined by

$$Y_k = \sigma_k \varepsilon_k, \tag{20}$$

where the conditional variance σ_k^2 is given by

$$\Lambda(\sigma_k^2) = c(\varepsilon_{k-1})\Lambda(\sigma_{k-1}^2) + g(\varepsilon_{k-1}). \tag{21}$$

Here $\{\varepsilon_k, k \in \mathbb{Z}\}$ is a sequence of i.i.d. errors and $\Lambda(x)$, $c(x)$ and $g(x)$ are real-valued measurable functions. To solve (21) for σ_k^2 one usually assumes that $\Lambda^{-1}(x)$ exists. Necessary and sufficient conditions for the existence of a strictly stationary solution of (20) and (21) were given by Duan [16] and Aue et al. [2]. Under some technical conditions stated in Hörmann [20] (Lemma 1, Lemma 2 and Remark 2) one can show that augmented GARCH sequences are weakly $\mathcal{M}$-dependent in L^p-norm with exponential rate.

3.6 Linear processes with dependent innovations

Linear processes $Z_k = \sum_{j=-\infty}^{\infty} a_j Y_{k-j}$ with dependent innovations $\{Y_k\}$ have obtained considerable interest in the financial literature. A common example are autoregressive (AR) processes with augmented GARCH innovations, see e.g. [26].

Assume that $\{Y_k, k \in \mathbb{Z}\}$ is weakly $\mathcal{M}$-dependent in L^p with rate function $\delta(\cdot)$. In combination with the results of Section 3.3 one can easily obtain conditions on δ assuring that the linear process $\{Z_k, k \in \mathbb{Z}\}$ defined above is also weakly $\mathcal{M}$-dependent in L^p-norm with a rate function δ^* depending on (a_j) and δ.

Strong approximation results for linear processes with dependent errors were also obtained by Wu and Min [51].

3.7 Ergodic sums

Let f be a real measurable function with period 1 such that $\int_0^1 f(\omega)\, d\omega = 0$ and $\int_0^1 |f(\omega)|^p\, d\omega < \infty$ for some $p > 2$. Set

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

and $B_n^2 = \int_0^1 S_n^2(\omega)\, d\omega$. Then S_n defines a partial sum process on the probability space $([0,1), \mathcal{B}_{[0,1)}, \lambda_{[0,1)})$, where $\mathcal{B}_{[0,1)}$ and $\lambda_{[0,1)}$ are the Borel σ–algebra and Lebesgue measure on $[0, 1)$. The strong law of large numbers for $f(2^k\omega)$ is a consequence of the ergodic theorem, for central and functional central limit theorems see Kac [23], Ibragimov [22] and Billingsley [5].

Let $Y_k(\omega) = f\left(2^k \omega\right)$ and define the random variable $\varepsilon_k(\omega)$ to be 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 and we have $\varepsilon_k = \pm 1$, each with probability $1/2$. This gives the representation

$$Y_k = f\left(\sum_{j=1}^{\infty} \varepsilon_{k+j} 2^{-j}\right) = g(\varepsilon_{k+1}, \varepsilon_{k+2}, \ldots).$$

We can now make use of the coupling method described in Section 3.4 and approximations

$$Y_k^{(m)} = g(\varepsilon_{k+1}, \varepsilon_{k+2}, \ldots, \varepsilon_{k+m}, \varepsilon_{k+m+1}^{(k)}, \varepsilon_{k+m+2}^{(k)}, \ldots).$$

Changing for some $\omega \in [0,1)$ the digits $\varepsilon_k(\omega)$ for $k > m$ will give an ω' with $|\omega - \omega'| \le 2^{-m}$. If f is Lipschitz continuous of some order γ then we have

$$|Y_k - Y_k^{(m)}| = O\big(2^{-\gamma m}\big),$$

and thus for any $p \ge 1$ $\{Y_k\}$ is weakly $\mathcal{M}$-dependent in L^p-norm with an exponentially decaying rate function.

4 Increments of stochastic processes

For arbitrary $\lambda > 0$, relation (18) has many useful applications in probability and statistics, for example, it implies a large class of limit theorems on CLT and LIL behavior and for various other functionals of weakly dependent sequences. However, many refined limit theorems for partial sums require a remainder term better than $O(n^{1/4})$ and no existing method for dependent sequences provides such a remainder term. The purpose of the next two sections is to show how to deal with such limit theorems via our approximation results in Section 2.

Let $\{Y_k, k \in \mathbb{Z}\}$ be a stationary random sequence and let $0 < a_n \le n$ be a nondecreasing sequence of real numbers. In this section, we investigate the order of magnitude of

$$\max_{1\le k\le n-a_n} \max_{1\le \ell\le a_n} \left| \sum_{j=k+1}^{k+\ell} Y_j \right|.$$

Such results have been obtained by Csörgő and Révész [9] for i.i.d. sequences and the Wiener process. In particular, they obtained the following result ([9], Theorem 1.2.1).

Theorem 3. *Let $\{a_T, T \ge 0\}$ be a positive nondecreasing function satisfying*

(a) $0 < a_T \le T$*;*

(b) T/a_T *is nondecreasing.*

Set

$$\beta_T = \left(2a_T\left[\log\frac{T}{a_T} + \log\log T\right]\right)^{-1/2}. \tag{22}$$

Then

$$\overline{\lim}_{T\to\infty} \max_{0\le t\le T-a_T} \max_{0\le s\le a_T} \beta_T \big|W(t+s) - W(t)\big| = 1.$$

Using strong invariance, a similar result can be obtained for partial sums of i.i.d. random variables under suitable moment conditions, see [9], pp. 115–118. For slowly growing a_T, this requires a very good remainder term in the Wiener approximation of partial sums, using the full power of the Komlós-Major-Tusnády theorems. As an application of our main theorems in Section 2, we now extend Theorem 3 for dependent stationary processes. To simplify the formulation and to clarify the connection between the remainder term in our approximation theorems in Section 2 and the increment problem, we introduce the following

Assumption 1. *Let $\{Y_k\}$ be a random sequence which can be redefined on a new probability space together with two standard Wiener processes $\{W_1(t),\, t \geq 0\}$ and $\{W_2(t),\, t \geq 0\}$ such that*

$$\sum_{k=1}^{n} Y_k = W_1\big(s_n^2\big) + W_2\big(t_n^2\big) + O\big(E_n\big) \quad \text{a.s.}, \tag{23}$$

where $\{E_n\}$ is some given sequence and $\{s_n^2\}$ and $\{t_n^2\}$ are nondecreasing sequences satisfying

$$s_n^2 \sim \sigma^2 n, \quad t_n^2 = o(n), \quad \overline{\lim}_{k\to\infty}\,(s_{k+1}^2 - s_k^2) = \overline{\lim}_{k\to\infty}\,(t_{k+1}^2 - t_k^2) = \sigma^2. \tag{24}$$

We will prove the following result.

Theorem 4. *Let $\{Y_k\}$ be a sequence of random variables satisfying Assumption 1 and put $S_n = \sum_{k=1}^{n} Y_k$. Let a_T be a positive nondecreasing function such that*

(a) $0 < a_T \leq T$*;*

(b) T/a_T *is nondecreasing;*

(c) a_T *is regularly varying at ∞ with index $\varrho \in (0, 1]$.*

Let β_T be defined by (22). Then under the condition

$$\beta_T E_T = o(1) \tag{25}$$

we have

$$\overline{\lim}_{n\to\infty} \max_{1\leq k\leq n-a_n} \max_{1\leq \ell\leq a_n} \beta_n \big|S_{k+\ell} - S_k\big| = \sigma^2. \tag{26}$$

Given a function a_T and a weakly $\mathcal{M}$-dependent sequence $\{Y_k\}$ with parameters $p, \delta(\cdot)$, we can compute, using Theorem 4, a rate of decrease for $\delta(\cdot)$ and a value for $p > 2$ such that the fluctuation result (26) holds. For example, if $a_T = T^\alpha$, $0 < \alpha < 1$, then (26) holds if $p > 4/\alpha$ and $\delta(m) \ll m^{-p/2}$.

We note that for i.i.d. observations only assumptions (a) and (b) are required. It remains open whether a more general version of our Theorem 4 which does not require assumption (c) can be proved.

Recently Zholud [53] obtained a distributional version of Theorem 3 by showing that the functional

$$\max_{0\leq t\leq T-a_T} \max_{0\leq s\leq a_T} \big(W(t+s) - W(t)\big)$$

converges weakly, suitably centered and normalized, to the extremal distribution with distribution function $e^{-e^{-x}}$. Using this fact and our a.s. invariance principles, a distributional version of Theorem 4 can be obtained easily. Since the argument is similar to that for (26), we omit the details.

Let $|A|$ be the cardinality of a set A. For the proof of Theorem 4 we need the following simple

Lemma 1. *Assume that $\{d_k, k \geq 1\}$ is a non-increasing sequence of positive numbers such that $\sum_{k=1}^{\infty} d_k = \infty$. Let $A \subset \mathbb{N}$ have positive density, i.e.*

$$\liminf_{n \to \infty} |A \cap \{1, \ldots, n\}|/n > 0.$$

Then $\sum_{k=1}^{\infty} d_k I\{k \in A\} = \infty$.

Proof. First note that by our assumption $\sum_{k=1}^{n} I\{k \in A\} \geq \mu n$ for some $\mu > 0$ as long as $n \geq n_0$. Using Abel summation we can write

$$\sum_{k=1}^{n} d_k = n d_n + \sum_{k=1}^{n-1} k(d_k - d_{k+1}).$$

Hence, by our assumptions

$$n d_n + \sum_{k=n_0}^{n-1} k(d_k - d_{k+1}) \to \infty, \quad (n \to \infty).$$

From $d_k - d_{k+1} \geq 0$ it follows (again using Abel summation) that for $n \geq n_0$

$$\begin{aligned}\sum_{k=1}^{n} d_k I\{k \in A\} &= d_n \sum_{k=1}^{n} I\{k \in A\} + \sum_{k=1}^{n-1} (d_k - d_{k+1}) \sum_{j=1}^{k} I\{j \in A\} \\ &\geq \mu \left(n d_n + \sum_{k=n_0}^{n-1} k(d_k - d_{k+1}) \right) \to \infty, \quad (n \to \infty).\end{aligned}$$

□

Proof of Theorem 4. For sake of simplicity we carry out the proof for $\sigma = 1$. From (23) and the triangular inequality we infer that

$$\begin{aligned}&\overline{\lim}_{n\to\infty} \max_{1 \leq k \leq n-a_n} \max_{1 \leq \ell \leq a_n} \beta_n \big| S_{k+\ell} - S_k \big| \\ &\quad \leq \overline{\lim}_{n\to\infty} \max_{1 \leq k \leq n-a_n} \max_{1 \leq \ell \leq a_n} \beta_n \big| W_1\big(s_{k+\ell}^2\big) - W_1\big(s_k^2\big) \big| \\ &\qquad + \overline{\lim}_{n\to\infty} \max_{1 \leq k \leq n-a_n} \max_{1 \leq \ell \leq a_n} \beta_n \big| W_2\big(t_{k+\ell}^2\big) - W_2\big(t_k^2\big) \big| \\ &\qquad + \overline{\lim}_{n\to\infty} \beta_n O\big(E_n\big) \\ &\quad = A_1 + A_2 + A_3.\end{aligned}$$

By (25) $A_3 = 0$. Since $a_n \to \infty$ (this is implicit in (c)), we conclude from (24) that for any $\varepsilon > 0$ some n_0 exists, such that for all $n \geq n_0$

$$\sup_{k \geq 1} \{s_{k+a_n}^2 - s_k^2\} \leq (1+\varepsilon) a_n \quad \text{and} \quad s_n^2 \leq (1+\varepsilon) n.$$

Set $T = (1+\varepsilon)n$ and define $a_{T,\varepsilon} = (1+\varepsilon)a_{T/(1+\varepsilon)}$. Then $a_{T,\varepsilon}$ satisfies (a) and (b) and for $n \geq n_0$ we have

$$\begin{aligned}
&\max_{1\leq k\leq n-a_n}\max_{1\leq \ell\leq a_n}\beta_n\big|W_1\big(s^2_{k+\ell}\big)-W_1\big(s^2_k\big)\big| \\
&\qquad\leq \sup_{0\leq t\leq s^2_{n-a_n}}\ \sup_{0\leq s\leq(1+\varepsilon)a_n}\beta_n\big|W_1(t+s)-W_1(t)\big| \\
&\qquad\leq \sup_{0\leq t\leq T-a_{T,\varepsilon}}\ \sup_{0\leq s\leq a_{T,\varepsilon}}\beta_{T/(1+\varepsilon)}\big|W_1\big(t+s\big)-W_1\big(t\big)\big|.
\end{aligned}$$

Let

$$\beta_{T,\varepsilon} = \left(2a_{T,\varepsilon}\left[\log\frac{T}{a_{T,\varepsilon}} + \log\log T\right]\right)^{-1/2}.$$

By application of Theorem 1.2.1 in Csörgő and Révész [9] (which requires (a) and (b)) we get

$$\overline{\lim}_{T\to\infty}\sup_{0\leq t\leq T-a_{T,\varepsilon}}\ \sup_{0\leq s\leq a_{T,\varepsilon}}\beta_{T,\varepsilon}\big|W_1\big(t+s\big)-W_1\big(t\big)\big| = 1 \quad \text{a.s.}$$

Since $\overline{\lim}_{T\to\infty}\beta_{T/(1+\varepsilon)}/\beta_{T,\varepsilon} = (1+\varepsilon)^{1/2}$, and ε can be chosen arbitrarily small we have shown that $A_1 \leq 1$ a.s.

It is not surprising that due to (24) similar arguments will lead to $A_2 = 0$ a.s.

The proof will be completed if we show that $A_1 \geq 1$. Let $\{n_k\}$ be a non-decreasing sequence of integers with $n_k \to \infty$. By (23), the triangular inequality and $A_2 = A_3 = 0$ we obtain

$$\begin{aligned}
&\overline{\lim}_{n\to\infty}\max_{1\leq k\leq n-a_n}\max_{1\leq \ell\leq a_n}\beta_n\big|S_{k+\ell}-S_k\big| \\
&\qquad\geq \overline{\lim}_{n\to\infty}\max_{1\leq k\leq n-a_n}\max_{1\leq \ell\leq a_n}\beta_n\big|W_1\big(s^2_{k+\ell}\big)-W_1\big(s^2_k\big)\big| \\
&\qquad\geq \overline{\lim}_{k\to\infty}\beta_{n_k}\big|W_1\big(s^2_{n_k}\big)-W_1\big(s^2_{n_k-a_{n_k}}\big)\big|.
\end{aligned}$$

We now proceed similarly as in Csörgő and Révész [9] for the proof of Step 2 of their Theorem 1.2.1. We will distinguish between the cases $\lim a_T/T = \rho$ with $\rho < 1$ and $\rho = 1$. Since both times we can use the same conceptual idea, we shall treat here only $\rho < 1$.

Set $n_1 = 1$. Given n_k define n_{k+1} such that $n_{k+1} - a_{n_{k+1}} = n_k$. This equation will in general have no integer solutions, but for the sake of simplicity we assume that (n_k) and (a_{n_k}) are $\mathbb{Z}$–valued. Since (s^2_n) is non-decreasing, we conclude that the increments $\Delta(k) = W_1\big(s^2_{n_k}\big) - W_1\big(s^2_{n_{k-1}}\big)$ are independent. By the second Borel-Cantelli lemma it suffices to show now that

$$\sum_{k=1}^{\infty} P(\beta_{n_k}|\Delta(k)| \geq 1-\varepsilon) = \infty \quad \text{for all } \varepsilon > 0. \tag{27}$$

For all large enough $k \in \mathbb{N}$ for which $s^2_{n_k} - s^2_{n_{k-1}} \geq (1-\varepsilon/2)a_{n_k}$, the estimates in [9] give

$$P(\beta_{n_k}|\Delta(k)| \geq 1-\varepsilon) \geq \left(\frac{a_{n_k}}{n_k\log n_k}\right)^{1-\varepsilon}.$$

It is also shown in [9] that $\sum_{k=1}^{\infty}\left(\frac{a_{n_k}}{n_k \log n_k}\right)^{1-\varepsilon} = \infty$. Thus, in view of Lemma 1 it remains to show that $A = \{k \geq 1 | s_{n_k}^2 - s_{n_{k-1}}^2 \geq (1-\varepsilon/2)a_{n_k}\}$ has a positive density. By (24) we have

$$\begin{aligned}
(s_{n_k}^2 - s_{n_1}^2)/n_k &= \sum_{j=2}^{k}(s_{n_j}^2 - s_{n_{j-1}}^2)/n_k \\
&\leq C_0 \sum_{\substack{2\leq j\leq k \\ j\in A}} (n_j - n_{j-1})/n_k + \sum_{\substack{2\leq j\leq k \\ j\in A^c}} (1-\varepsilon/2)(n_j - n_{j-1})/n_k \\
&\leq C_0 \sum_{\substack{2\leq j\leq k \\ j\in A}} (n_j - n_{j-1})/n_k + (1 - n_1/n_k)(1-\varepsilon/2),
\end{aligned}$$

for some $C_0 > 0$ which is independent of k. Now if A had density zero, the limsup of the right-hand side of the last relation would be $1 - \varepsilon/2$. This can be easily proved, using that $(n_j - n_{j-1})$ is regularly varying by assumption (c). The liminf of the left-hand side above is 1. Thus, A must have positive density and the proof is completed. □

5 Change-point tests with an epidemic alternative

In this section we apply our invariance principles to a change-point problem. Let $\{Y_k, k \in \mathbb{Z}\}$ be a zero mean process. Further let $X_k = Y_k + \mu_k$, where μ_k, $k \in \mathbb{Z}$, are unknown constants. We want to test the hypothesis

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_n = \mu$$

against the "epidemic alternative"

$$\begin{aligned}
H_A : &\text{ There exist } 1 \leq m_1 < m_2 \leq n \text{ such that } \mu_k = \mu \text{ for} \\
&k \in \{1, \ldots, n\} \setminus \{m_1+1, \ldots, m_2\} \text{ and } \mu_k = \mu + \Delta \text{ if } k \in \{m_1+1, \ldots, m_2\}.
\end{aligned}$$

It should be noted that the variables m_1, m_2 and Δ may depend on the sample size n. As it is common in the change-point literature, this dependence is suppressed in the notation.

Without loss of generality we assume that $\sigma = 1$. To detect a possible epidemic change it is natural to compare the increments of the process to a proportion of the total sum. More specifically, assume for the moment that X_k are independent and that we know when the epidemic starts and ends. Set $S_k = X_1 + \ldots + X_k$. Then by the law of large numbers $I(m_1, m_2) = |S_{m_2} - S_{m_1} - (m_2 - m_1)\, S_n/n| \gg m_2 - m_1$. If no change occurs, however, by the central limit theorem $I(m_1, m_2) = O_P\big(\sqrt{m_2 - m_1}\,\big)$. In general we do not know m_1 and m_2. Thus, a natural test statistic is

$$\max_{1\leq i<j\leq n} |S_j - S_i - (j-i)\, S_n/n|.$$

Clearly we are required to normalize the above test statistic appropriately. Following Račkauskas and Suquet in [35] we define

$$UI(n,\alpha) = n^{-1/2} \max_{1\le i<j\le n} \frac{|S_j - S_i - (j-i)\, S_n/n|}{[(j-i)/n\,(1-(j-i)/n)]^{\alpha}}$$

with $0 < \alpha < 1/2$. As we will see below, the parameter α plays an important role. The closer α is to 1/2 the "shorter" epidemics can be detected with this test. The price, however, is that in order to obtain the limiting law under H_0 with "large" α (close to 1/2) requires a.s. invariance principles with error n^{ϵ}, ϵ close to zero. Choosing $\alpha \ge 1/2$ would result in a degenerate limiting distribution under H_0.

Proposition 2 (Asymptotics under H_0)**.** *If the stationary sequence $\{Y_k, k\in\mathbb{Z}\}$ satisfies Assumption 1 with $E_n = o(n^{1/2-\alpha})$ and H_0 holds, then*

$$\sigma^{-1} UI(n,\alpha) \xrightarrow{\mathcal{D}} \sup_{0<s<t<1} \frac{|B(t)-B(s)|}{[(t-s)(1-(t-s))]^{\alpha}},$$

where $\{B(t), t\in[0,1]\}$ is a Brownian bridge.

Proof. Using (23) and assuming for simplicity that $\sigma = 1$, we obtain

$$\begin{aligned} UI(n,\alpha) &\le n^{-1/2} \max_{1\le i<j\le n} \frac{|W_1(s_j^2) - W_1(s_i^2) - (j-i)\, W_1(s_n^2)/n|}{[(j-i)/n\,(1-(j-i)/n)]^{\alpha}} \\ &\quad + n^{-1/2} \max_{1\le i<j\le n} \frac{|W_2(t_j^2) - W_2(t_i^2) - (j-i)\, W_2(t_n^2)/n|}{[(j-i)/n\,(1-(j-i)/n)]^{\alpha}} \\ &\quad + O\big(n^{-1/2+\alpha} E_n\big) \\ &= n^{-1/2} \max_{1\le i<j\le n} T_{i,j}^{(1)} + n^{-1/2} \max_{1\le i<j\le n} T_{i,j}^{(2)} + o(1). \end{aligned}$$

It is easy to see that $n^{-1/2}\max_{1\le i<j\le n} T_{i,j}^{(2)}$ tends to zero. Since we can get a similar lower bound for $UI(n,\alpha)$ we have

$$UI(n,\alpha) = n^{-1/2} \max_{(i,j)\in\mathcal{M}_n} T_{i,j}^{(1)} + o_P(1),$$

where $\mathcal{M}_n = \{(i,j) | 1\le i<j\le n\}$. Let us partition $\mathcal{M}_n$ into $\mathcal{M}_{1,n} = \{(i,j)|1\le i<j\le n;\, n\gamma_n < j-i < n(1-\gamma_n)\}$, $\mathcal{M}_{2,n} = \{(i,j)|1\le i<j\le n; n\gamma_n \ge j-i\}$ and $\mathcal{M}_{3,n} = \{(i,j)|1\le i<j\le n; j-i\ge n(1-\gamma_n)\}$, where $\gamma_n \to 0$ will be defined later. By our assumptions on the sequence $\{s_j^2\}$ there exists a $\tau > 0$ such that $s_j^2 - s_i^2 \le \tau(j-i)$ for all $1\le i\le j$ and that $s_n^2 \le (2-\tau\gamma_n)n$ if $n\ge n_0$. We have for large enough n

$$\begin{aligned} &n^{-1/2} \max_{(i,j)\in\mathcal{M}_{2,n}} T_{i,j}^{(1)} \\ &\quad\le 2n^{\alpha-1/2} \max_{(i,j)\in\mathcal{M}_{2,n}} \left\{ \frac{|W_1(s_j^2) - W_1(s_i^2)|}{(j-i)^{\alpha}} \right\} + 2n^{-1/2}\gamma_n^{1-\alpha} |W_1(s_n^2)| \\ &\quad\le 2n^{\alpha-1/2} \max_{1\le h\le n\gamma_n} \sup_{0\le t\le(2-\tau\gamma_n)n} \sup_{0\le s\le\tau h} \left\{ \frac{|W_1(t+s) - W_1(t)|}{h^{\alpha}} \right\} + o_P(1). \end{aligned}$$

For arbitrary $\epsilon > 0$ we get by Lemma 1.2.1 in Csörgő and Révész [9] that there is a constant C which is independent of n and ϵ such that

$$\begin{aligned} &P\left(\max_{1 \le h \le n\gamma_n} \sup_{0 \le t \le (2-\tau\gamma_n)n} \sup_{0 \le s \le \tau h} \left\{ \frac{|W_1(t+s) - W_1(t)|}{h^\alpha} \right\} > \epsilon\, n^{1/2-\alpha} \right) \\ &\quad \le \sum_{h=1}^{n\delta_n} P\left(\sup_{0 \le t \le 2n - \tau h} \sup_{0 \le s \le \tau h} |W_1(t+s) - W_1(t)| > \epsilon\, h^{1/2} (n/h)^{1/2-\alpha} \right) \\ &\quad \le \sum_{h=1}^{n\delta_n} \frac{Cn}{h}\, e^{-\frac{\epsilon^2}{3}(n/h)^{1-2\alpha}} \to 0 \quad (n \to \infty). \end{aligned}$$

Hence $n^{-1/2} \max_{(i,j)\in\mathcal{M}_{2,n}} T_{i,j}^{(1)} = o_P(1)$. In the same fashion one can show that $n^{-1/2} \max_{(i,j)\in\mathcal{M}_{3,n}} T_{i,j}^{(1)} = o_P(1)$. Therefore $UI(n,\alpha) = n^{-1/2} \max_{(i,j)\in\mathcal{M}_{1,n}} T_{i,j}^{(1)} + o_P(1)$. Some further basic estimates give

$$\begin{aligned} &n^{-1/2} \max_{(i,j)\in\mathcal{M}_{1,n}} T_{i,j}^{(1)} \\ &\quad = n^{-1/2} \max_{(i,j)\in\mathcal{M}_{1,n}} \frac{|W_1(j) - W_1(i) - (j-i)W_1(n)/n|}{\big[(j-i)/n\,(1-(j-i)/n)\big]^\alpha} \\ &\qquad + O\left(\frac{n^{-1/2}}{\gamma_n^\alpha} \max_{1\le i\le n} |W_1(i) - W_1(s_i^2)| \right). \end{aligned}$$

Since $s_n^2 \sim n$ it follows that there is a null sequence $\{\epsilon_n\}$ such that $\max_{1\le i\le n} |i - s_i^2| \le \epsilon_n n$. Hence

$$\max_{1\le i\le n} |W_1(i) - W_1(s_i^2)| \le \sup_{0\le t\le n} \sup_{0\le s\le 2\epsilon_n n} |W_1(t+s) - W_1(t)|.$$

Setting $\gamma_n = \epsilon_n$ and applying again Lemma 1.2.1 in [9] it can be seen that

$$\sup_{0\le t\le n} \sup_{0\le s\le 2\epsilon_n n} |W_1(t+s) - W_1(t)| = o_P\big(n^{1/2}\gamma_n^\alpha\big).$$

Consequently

$$UI(n,\alpha) = n^{-1/2} \max_{(i,j)\in\mathcal{M}_{1,n}} \frac{|W_1(j) - W_1(i) - (j-i)W_1(n)/n|}{\big[(j-i)/n\,(1-(j-i)/n)\big]^\alpha} + o_P(1). \tag{28}$$

Since the line of argumentation is very similar to what we have shown before, we note now without proof that $\mathcal{M}_{1,n}$ in the right hand side of (28) can be replaced by $\mathcal{M}_n$. The rest of the proof of Proposition 2 is standard. □

The next proposition shows that this test is consistent. Let $\ell = m_2 - m_1$ denote the length of the epidemic.

Proposition 3 (Asymptotics under H_A)**.** *Let $\{Y_k, k \in \mathbb{Z}\}$ be a mean zero process, weakly $\mathcal{M}$-dependent in L^p with $p \geq 2$ and $\delta(\cdot)$ satisfying*

$$\sum_{m \geq 1} \delta(m) < \infty.$$

Let $X_k = Y_k + \mu_k$, $k \in \mathbb{Z}$. Assume that H_A holds and that

$$\lim_{n\to\infty} \frac{(\ell(n-\ell))^{1-\alpha}}{n^{3/2-2\alpha}} |\Delta| = \infty. \tag{29}$$

Then $UI(n,\alpha) \xrightarrow{\mathcal{P}} \infty$.

Proof. Under the alternative hypothesis H_A we have $X_k = Y_k + \mu$ for $k \in \{1, \ldots, n\} \setminus \{m_1 + 1, \ldots, m_2\}$ and $X_k = Y_k + \mu + \Delta$ for $k \in \{m_1 + 1, \ldots, m_2\}$. To find a lower bound for $UI(n,\alpha)$ we study the numerator of the test statistic corresponding to the true epidemic. Thus we look at

$$\begin{aligned}
S_{m_2} - S_{m_1} - S_n(m_2/n - m_1/n) = \\
&= (1-\ell/n)(S_{m_2} - S_{m_1}) - (\ell/n)(S_n - (S_{m_2} - S_{m_1})) \\
&= \frac{\ell(n-\ell)}{n}\Delta + (1-\ell/n)\sum_{j=m_1+1}^{m_2} Y_j - (\ell/n)\left(\sum_{j=1}^{m_1} Y_j + \sum_{j=m_2+1}^{n} Y_j\right) \\
&= \frac{\ell(n-\ell)}{n}\Delta + R_n.
\end{aligned}$$

With the help of the moment inequality stated in Proposition 6.1 below we get

$$\begin{aligned}
\mathrm{Var}(n^{-1/2}R_n) &= O((1-\ell/n)^2(\ell/n) + (\ell/n)^2(1-\ell/n) + 2(\ell/n)^{3/2}(1-\ell/n)^{3/2}) \\
&= O((\ell/n)(1-\ell/n)),
\end{aligned}$$

and thus $n^{-1/2}R_n = O_P((\ell/n)^{1/2}(1-\ell/n)^{1/2})$. Thus we have shown that

$$\begin{aligned}
UI(n,\alpha) &\geq n^{1/2}\left((\ell/n)(1-\ell/n)\right)^{1-\alpha} |\Delta| - O_P\left(((\ell/n)(1-\ell/n))^{1/2-\alpha}\right) \\
&= \frac{(\ell(n-\ell))^{1-\alpha}}{n^{3/2-2\alpha}} |\Delta| - O_P\left(((\ell/n)(1-\ell/n))^{1/2-\alpha}\right).
\end{aligned} \tag{30}$$

To conclude the proof we note that $\lim_{n\to\infty} ((\ell/n)(1-\ell/n))^{1/2-\alpha} = 0$ if $\ell = o(n)$ (or $n - \ell = o(n)$ respectively) and $((\ell/n)(1-\ell/n))^{1/2-\alpha} \leq 1$ in general. Consequently condition (29) together with relation (30) finishes the proof. $\square$

For example, if Δ is independent of n then condition (29) will hold for $\ell \sim cn$, $c \in (0,1)$. In case that $n^\nu \ll \ell \ll n - n^\nu$, $\nu > 0$, condition (29) holds if $(1-2\alpha)/(1-\alpha) < 2\nu$. That is, choosing α close to $1/2$ allows to detect relatively "short" ("long") epidemics.

6 Proof of the main theorems

6.1 A moment inequality

In the proofs of our theorems we will use the following moment inequality which may be of separate interest.

Proposition 4. *Let $\{Y_k,\, k \in \mathbb{Z}\}$ be a centered stationary sequence, weakly $\mathcal{M}$-dependent in L^p with $p \geq 2$ and a rate function $\delta(\cdot)$ satisfying*

$$D_p := \sum_{m=0}^{\infty} \delta(m) < \infty.$$

Then for any $n \in \mathbb{N}$, $b \in \mathbb{Z}$ we have

$$E\left|\sum_{k=b+1}^{b+n} Y_k\right|^p \leq C_p\, n^{p/2}, \tag{31}$$

where C_p is a constant depending on p and the sequence $\{Y_k\}$.

Proof. By stationarity, we can assume $b = 0$. Let first $p = 2$. We use below that $\sup_{m\geq 0} \|Y_k^{(m)}\|_p \leq \|Y_1\|_p + D_p$. Without loss of generality we assume that $EY_k^{(m)} = 0$ for all $k \in \mathbb{Z}$ and $m \in \mathbb{N}$. Since

$$Y_k Y_{k+j} = \Big(Y_k - Y_k^{(j-1)}\Big) Y_{k+j} + Y_k^{(j-1)} \Big(Y_{k+j} - Y_{k+j}^{(j-1)}\Big) + Y_k^{(j-1)} Y_{k+j}^{(j-1)}$$

we get by Assumption (B) that for $j \geq 1$

$$\begin{aligned}
\big|EY_k Y_{k+j}\big| &\leq \Big|E\big[\big(Y_k - Y_k^{(j-1)}\big)Y_{k+j}\big]\Big| + \Big|E\big[Y_k^{(j-1)}(Y_{k+j} - Y_{k+j}^{(j-1)})\big]\Big| \\
&\leq \big\|Y_{k+j}\big\|_2 \big\|Y_k - Y_k^{(j-1)}\big\|_2 + \big\|Y_k^{(j-1)}\big\|_2 \big\|Y_{k+j} - Y_{k+j}^{(j-1)}\big\|_2 \\
&\leq \Big(\big\|Y_{k+j}\big\|_2 + \big\|Y_k^{(j-1)}\big\|_2\Big)\delta(j-1) \\
&\leq \big(2\|Y_1\|_2 + D_2\big)\delta(j-1).
\end{aligned} \tag{32}$$

From relation (32) we infer, letting $S_n = \sum_{k=1}^n Y_k$,

$$\begin{aligned}
ES_n^2 &= \sum_{k=1}^{n} EY_k^2 + 2 \sum_{1\leq k<l\leq n} EY_k Y_l \\
&\leq n\|Y_1\|_2^2 + 2\left[\sum_{1\leq k\leq n-1} |EY_k Y_{k+1}| + \ldots + \sum_{1\leq k\leq 2} |EY_k Y_{k+n-2}| + E|Y_1 Y_n|\right] \\
&\leq n\|Y_1\|_2^2 + 2\big(2\|Y_1\|_2 + D_2\big)\left[(n-1)\delta(0) + \ldots + 2\delta(n-3) + \delta(n-2)\right] \\
&\leq n\Big(\|Y_1\|_2^2 + 2D_2\big(2\|Y_1\|_2 + D_2\big)\Big) =: C_2\, n.
\end{aligned}$$

This shows (31) for $p=2$.

Once (31) is established for p, it holds for all $0<q\le p$. Indeed, by Lyapunov's inequality, relation (31) implies

$$E\left|\sum_{k=b+1}^{b+n} Y_k\right|^q \le C_p^{q/p} n^{q/2} \tag{33}$$

for any $0<q\le p$. In particular, (31) holds with $p=1$.

Next we prove (31) for all integers $p>2$. Clearly, if $C_p \ge \|Y_1\|_p^p$, then the inequality

$$E|S_n|^p \le C_p\, n^{p/2} \tag{34}$$

holds for $n=1$. Using a double induction argument, we show now that for some constant C_p, relation (34) holds for all $n\in\mathbb{N}$. More precisely, we show that if (34) holds for $p-1$ and all $n\in\mathbb{N}$ and also for p and $n\le n_0$, then it will also hold for p and $n\le 2n_0$.

For $k\le n$ put $S_k^n = Y_k+Y_{k+1}+\ldots+Y_n$. We have

$$\begin{aligned}
E|S_{2n}|^p &= E\left|S_n + S_{n+1}^{2n}\right|^p \\
&= E\left|\sum_{k=1}^n \left(Y_k - Y_k^{(n-k)}\right) + \sum_{k=1}^n \left(Y_{n+k} - Y_{n+k}^{(k-1)}\right) + \sum_{k=1}^n Y_k^{(n-k)} + \sum_{k=1}^n Y_{n+k}^{(k-1)}\right|^p \\
&\le \Bigg(\sum_{k=1}^n \left\|Y_k - Y_k^{(n-k)}\right\|_p + \sum_{k=1}^n \left\|Y_{n+k} - Y_{n+k}^{(k-1)}\right\|_p \qquad (35)\\
&\qquad\qquad + \left\|\sum_{k=1}^n Y_k^{(n-k)} + \sum_{k=1}^n Y_{n+k}^{(k-1)}\right\|_p\Bigg)^p \\
&\le \left(2D_p + \left\|\sum_{k=1}^n Y_k^{(n-k)} + \sum_{k=1}^n Y_{n+k}^{(k-1)}\right\|_p\right)^p \\
&=: \left(2D_p + \|Z_n+W_n\|_p\right)^p. \qquad (36)
\end{aligned}$$

For some positive constants ψ_p that will be specified later, we choose C_p so that $C_p^{1/p} > D_p/\psi_p$. Then if $n\le n_0$

$$\begin{aligned}
E|Z_n|^p &\le \left(\|S_n\|_p + \|S_n - Z_n\|_p\right)^p \\
&\le \left(\|S_n\|_p + D_p\right)^p \\
&\le (1+\psi_p)^p C_p n^{p/2}.
\end{aligned}$$

By the induction assumption, this relation holds with arbitrary n for all integer moments of order $\le p-1$. The same estimate applies for $E|W_n|^p$. Due to Assumption (B) in Definition 1, the random variables Z_n and W_n are independent.

Thus

$$\begin{aligned}
E|Z_n+W_n|^p &\le E|Z_n|^p + E|W_n|^p + \sum_{m=1}^{p-1}\binom{p}{m} E|Z_n|^m E|W_n|^{p-m} \\
&\le n^{p/2}\left[2(1+\psi_p)^p C_p + \sum_{m=1}^{p-1}\binom{p}{m}(1+\psi_m)^m(1+\psi_{p-m})^{p-m} C_m C_{p-m}\right] \\
&=: n^{p/2}\Big[2(1+\psi_p)^p C_p + R_p\Big]. \qquad (37)
\end{aligned}$$

Hence (36) and (37) and our assumptions on C_p imply that

$$\begin{aligned}
E|S_{2n}|^p &\le \left(2\psi_p C_p^{1/p} + n^{1/2}\Big[2(1+\psi_p)^p C_p + R_p\Big]^{1/p}\right)^p \\
&\le C_p n^{p/2}\left(2\psi_p + \Big[2(1+\psi_p)^p + R_p/C_p\Big]^{1/p}\right)^p. \qquad (38)
\end{aligned}$$

Choosing ψ_p small enough and then choosing C_p large enough, we can always achieve that the term in brackets of (38) is $\le \sqrt{2}$, provided that $p > 2$, and that the inequality $C_p^{1/p} > D_p/\psi_p$ mentioned before is satisfied. Hence we have for every $n \le n_0$ that $E|S_{2n}|^p \le C_p(2n)^{p/2}$, proving (34) for all even numbers $n \le 2n_0$. The case of odd n is similar. The proof of Proposition is finished for integer p.

For general $p > 2$ we have by the result shown before that (31) holds for $\lfloor p \rfloor$. (As usual, $\lfloor p \rfloor$ denotes the integer part of the real number p.) To finish the proof we need the following inequality which will be proven below:

$$|a+b|^p \le |a|^p + |b|^p + \sum_{k=1}^{\lfloor p \rfloor}\binom{p}{k}\Big(|a|^k|b|^{p-k} + |b|^k|a|^{p-k}\Big), \quad p \in [1,\infty). \qquad (39)$$

Using (39) we get a similar estimate for $E|Z_n+W_n|^p$ as in (37) and the proof can be finished along the same lines as for integer p.

Verification of (39)*:* Let $x \in [0,1]$. We recall that $(1+x)^p$ can be expanded in the binomial series

$$(1+x)^p = \sum_{k\ge 0}\binom{p}{k}x^k,$$

with

$$\binom{p}{k} = \frac{p(p-1)\cdots(p-k+1)}{k!}. \qquad (40)$$

From (40) it is clear that for $k \ge \lfloor p \rfloor + 2$ we have sign$\left\{\binom{p}{k}\right\} = (-1)^{k-\lfloor p \rfloor+1}$. This immediately yields for $k = \lfloor p \rfloor + 2\ell$ with $\ell \ge 1$,

$$\binom{p}{k}x^k + \binom{p}{k+1}x^{k+1} \le \binom{p}{k}x^k + \binom{p}{k+1}x^k = \binom{p+1}{k+1}x^k < 0.$$

Consequently

$$\sum_{k\ge\lfloor p \rfloor+2}\binom{p}{k}x^k < 0,$$

and

$$(1+x)^p \le \sum_{k=0}^{\lfloor p\rfloor+1} \binom{p}{k} x^k. \tag{41}$$

Now consider $|a+b|^p$. If $|a| \ge |b|$ then we infer from (41) that

$$\begin{aligned}
|a+b|^p &\le |a|^p \left(1 + \left|\frac{b}{a}\right|\right)^p \\
&\le |a|^p \sum_{k=0}^{\lfloor p\rfloor+1} \binom{p}{k} \left|\frac{b}{a}\right|^k \\
&= |a|^p + \sum_{k=1}^{\lfloor p\rfloor} \binom{p}{k} |b|^k |a|^{p-k} + \binom{p}{\lfloor p\rfloor+1} |b|^p \left|\frac{b}{a}\right|^{\lfloor p\rfloor+1-p}.
\end{aligned}$$

Thus (39) follows from $\binom{p}{\lfloor p\rfloor+1}\left|\frac{b}{a}\right|^{\lfloor p\rfloor+1-p} \le 1$. Interchanging the roles of a and b completes the proof. □

Using Móricz [32, Theorem 1]) we get

Corollary 1. *Under the assumptions of Proposition 6.1 with $p > 2$, we have for any $2 < q \le p$ and any $n \in \mathbb{N}$, $b \in \mathbb{Z}$*

$$E \max_{1\le k\le n} \left| \sum_{j=b+1}^{b+k} Y_j \right|^q \le C'_{p,q}\, n^{q/2},$$

where the constants $C'_{p,q}$ only depend on p, q and the sequence $\{Y_k\}$.

A slightly weaker result can also be derived from Proposition 6.1 for the case of $0 < q \le 2$.

6.2 Proofs of Theorems 1 and 2

We give the proof of Theorem 1. Note first of all that $\delta(m) = \left\|Y_k - Y_k^{(m)}\right\|_p \ge \left\|Y_k - Y_k^{(m)}\right\|_2$ and consequently (32) holds when the L^2–norm is replaced by the L^p–norm. Since $A > 1$ in (9), we infer that the series in (10) is absolutely convergent.

Let us specify some constants that will be used for the proof. By our assumption on A it is possible to find a constant $0 < \varepsilon_0 < 1/2$ such that

$$A > \frac{p-2}{2\eta(1-\varepsilon_0)^2}\left(1 - \frac{1+\eta}{p}\right).$$

Then we set

$$\delta = \frac{\beta}{1+\alpha} \quad \text{with} \quad \alpha = \frac{2\eta(1-\varepsilon_0)}{p-2(1+\eta)}, \quad \beta = (1-\varepsilon_0)\alpha. \tag{42}$$

For some $\varepsilon_1 > 0$ (which will be specified later) we now define $m_k = \lfloor \varepsilon_1 k^\delta \rfloor$. The first step in the proof of (11) is to show that it is sufficient to provide the strong approximation for the perturbed sequence $Y_k' = Y_k^{(m_k)}$. We notice that our main assumption (8) yields $\|Y_k - Y_k'\|_p \ll k^{-A\delta}$. If $A\delta < 1$, then

$$\begin{aligned}
&P\left(\max_{2^n \le k \le 2^{n+1}} \left|\sum_{j=1}^{k} (Y_j - Y_j')\right| > \frac{1}{n} 2^{\frac{n}{p}(1+\eta)}\right) \\
&\le P\left(\sum_{j=1}^{2^{n+1}} \left|Y_j - Y_j'\right| > \frac{1}{n} 2^{\frac{n}{p}(1+\eta)}\right) \\
&\le 2^{-n(1+\eta)} n^p \left(\sum_{j=1}^{2^{n+1}} \|Y_j - Y_j'\|_p\right)^p \\
&\ll 2^{-c_1 n} n^p,
\end{aligned}$$

where $c_1 = (1+\eta) - (1 - A\delta)p > 0$. Thus by the Borel-Cantelli lemma we have almost surely

$$\sum_{j=1}^{k} Y_j = \sum_{j=1}^{k} Y_j' + o\big(k^{(1+\eta)/p}\big) \quad \text{a.s..}$$

If $A\delta \ge 1$ we get an (even better) error term of order $o\big(k^{1/p}\big)$.

The main part of the proof of Theorem 1 is based on a blocking argument. We partition $\mathbb{N}$ into disjoint blocks:

$$\mathbb{N} = J_1 \cup I_1 \cup J_2 \cup I_2 \cup \ldots,$$

where $|I_k| = \lfloor k^\alpha \rfloor$ and $|J_k| = \lfloor k^\beta \rfloor$ with α, β as in (42). Let us further set

$$I_k = \{\underline{i}_k, \ldots, \bar{i}_k\} \quad \text{and} \quad J_k = \{\underline{j}_k, \ldots, \bar{j}_k\}$$

and

$$\xi_k = \sum_{j \in I_k} Y_j' \quad \text{and} \quad \eta_k = \sum_{j \in J_k} Y_j'.$$

Note that $\bar{i}_k = O\big(k^{1+\alpha}\big)$. Provided that ε_1 in the definition of m_k is chosen small enough, this will imply that

$$|J_k| = \lfloor k^\beta \rfloor > \lfloor \varepsilon_1 \underline{i}_k^\delta \rfloor = m_{\underline{i}_k}$$

and hence by Assumption (B) it follows that $\{\xi_k\}$ and $\{\eta_k\}$ each define a sequence of independent random variables.

The following lemma by Sakhanenko [42] (cf. also Shao [43]) is our crucial ingredient for the construction of the approximating processes.

Lemma 2. *Let $\{\xi_k\}$ be a sequence of centered independent random variables with finite p–th moments, $p > 2$. Then we can redefine $\{\xi_k\}$ on a suitable probability* *space, together with a sequence* $\{\xi_k^*\}$ *of independent normal random variables with* $E\xi_k^* = 0$, $E(\xi_k^*)^2 = E\xi_k^2$ *such that for any* $x > 0$, $m \geq 1$

$$P\left(\max_{1\leq k\leq m}\left|\sum_{j=1}^{k}\xi_j - \sum_{j=1}^{k}\xi_j^*\right| > x\right) \leq C\frac{1}{x^p}\sum_{j=1}^{m}E|\xi_j|^p,$$

where C *is an absolute constant.*

We shall now apply Lemma 2 to the sequences $\{\xi_k\}$ and $\{\eta_k\}$. For this purpose we need estimates of the moments $E|\xi_k|^p$, $E|\eta_k|^p$. By Minkowski's inequality and Proposition 6.1 we get

$$\begin{aligned} E|\xi_k|^p &\leq \left(\left\|\sum_{j\in I_k}Y_k\right\|_p + \sum_{j\in I_k}\|Y_j - Y_j'\|_p\right)^p \\ &= O\left(\left(|I_k|^{1/2} + |I_k|\cdot \underline{i}_k^{-A\delta}\right)^p\right). \end{aligned}$$

Some easy algebra shows that the restrictions on the parameters A, δ, α and ε_0 imply

$$|I_k|\cdot \underline{i}_k^{-A\delta} \ll k^{\alpha}\cdot k^{-A\delta(1+\alpha)} \ll k^{\alpha/2} \ll |I_k|^{1/2}.$$

A similar estimate holds for $E|\eta_k|^p$. Hence we can find constants F_p such that

$$E|\xi_k|^p \leq F_p|I_k|^{p/2} \quad \text{and} \quad E|\eta_k|^p \leq F_p|J_k|^{p/2},$$

where F_p does not depend on k.

Let $L_n = \sum_{k=1}^{n}|I_k|$. Then $L_n = O\big(n^{(1+\alpha)}\big)$. By our previous estimates and by Lemma 2 we infer that, after enlarging the probability space, we have

$$\begin{aligned} P\left(\max_{2^n\leq k\leq 2^{n+1}}\left|\sum_{j=1}^{k}\xi_j - \sum_{j=1}^{k}\xi_j^*\right| > L_{2^n}^{\frac{1+\eta}{p}}\right) &\leq L_{2^n}^{-(1+\eta)}\sum_{k=1}^{2^{n+1}}E|\xi_k|^p \\ &= O\Big(2^{[-(1+\alpha)(1+\eta)+\frac{\alpha p}{2}+1]n}\Big), \end{aligned} \tag{43}$$

where ξ_k^* is a sequence of independent and centered normal random variables with $E(\xi_k^*)^2 = E\xi_k^2$. In order that the exponent in (43) be negative we need $(1+\alpha)(1+\eta) > \frac{\alpha p}{2}+1$. This is equivalent to $\alpha < \frac{2\eta}{p-2(1+\eta)}$, which follows by (42). Thus, by the Borel-Cantelli lemma we obtain

$$\sum_{j=1}^{k}\xi_j = \sum_{j=1}^{k}\xi_j^* + O\left(L_k^{\frac{1+\eta}{p}}\right) \quad \text{a.s.}$$

By further enlarging the probability space we can write

$$\sum_{j=1}^{k}\xi_j = W_1\left(\sum_{j=1}^{k}\mathrm{Var}(\xi_j)\right) + O\left(L_k^{\frac{1+\eta}{p}}\right) \quad \text{a.s.,}$$

where $\{W_1(t),\, t \geq 0\}$ is a standard Wiener process. The same arguments show that

$$\sum_{j=1}^{k} \eta_j = W_2\left(\sum_{j=1}^{k} \mathrm{Var}(\eta_j)\right) + O\left(M_k^{\frac{1+\eta}{p}}\right) \quad \text{a.s.,}$$

where $\{W_2(t),\, t \geq 0\}$ is another standard Wiener process on the same probability space and $M_n = \sum_{k=1}^{n} |J_k|$.

It is a basic result that (32) and the stationarity of $\{Y_j\}$ imply

$$b_k^2 = \mathrm{Var}\left(\sum_{j \in I_k} Y_j\right) / |I_k| \to \sigma^2 \quad \text{and} \quad h_k^2 = \mathrm{Var}\left(\sum_{j \in J_k} Y_j\right) / |J_k| \to \sigma^2, \tag{44}$$

as $k \to \infty$. It can be easily shown that (44) remains true if the Y_j are replaced with Y_j'. Indeed, by the Minkowski inequality we infer that

$$\begin{aligned}\mathrm{Var}^{1/2}\left(\sum_{j \in I_k} Y_j'\right) &\leq \mathrm{Var}^{1/2}\left(\sum_{j \in I_k} Y_j\right) + \mathrm{Var}^{1/2}\left(\sum_{j \in I_k} (Y_j - Y_j')\right) \\ &\leq \mathrm{Var}^{1/2}\left(\sum_{j \in I_k} Y_j\right) + |I_k| \|Y_{\underline{i}_k} - Y_{\underline{i}_k}'\|_2.\end{aligned}$$

Furthermore, using the definitions of the introduced constants we obtain

$$\begin{aligned}\|Y_{\underline{i}_k} - Y_{\underline{i}_k}'\|_2 &\leq \|Y_{\underline{i}_k} - Y_{\underline{i}_k}'\|_p \\ &\ll \underline{i}_k^{-A\delta} \\ &\ll k^{-(\alpha+1)A\delta} \\ &\leq k^{-\beta} = k^{-(1-\epsilon_0)\alpha}, \quad \text{with } \varepsilon_0 < 1/2.\end{aligned}$$

Since by definition $|I_k| \ll k^\alpha$, we conclude that

$$\mathrm{Var}^{1/2}\left(\sum_{j \in I_k} Y_j'\right) / |I_k|^{1/2} \leq \mathrm{Var}^{1/2}\left(\sum_{j \in I_k} Y_j\right) / |I_k|^{1/2} + o(1) \quad \text{as } k \to \infty.$$

In the same manner a lower bound for $\mathrm{Var}^{1/2}\left(\sum_{j \in I_k} Y_j'\right) / |I_k|^{1/2}$ can be obtained. The second part of (44) is analogous.

For $\ell \in I_k$ we set $\sigma_\ell^2 = b_k^2$ and for $\ell \in J_k$ we set $\sigma_\ell^2 = 0$. Similarly define $\tau_\ell^2 = h_k^2$ if $\ell \in J_k$ and $\tau_\ell^2 = 0$ if $\ell \in I_k$. Summarizing our results so far we can write

$$\sum_{k=1}^{\bar{i}_n} Y_k = W_1\left(\sum_{k=1}^{\bar{i}_n} \sigma_k^2\right) + W_2\left(\sum_{k=1}^{\bar{i}_n} \tau_k^2\right) + O\left(\bar{i}_n^{\,(1+\eta)/p}\right) \quad \text{a.s.}$$

In other words we have shown (11) along the subsequence $\{\bar{i}_n\}$ with values of s_n that satisfy $s_n^2 \sim \sigma^2 n$ and (13), and values of t_n^2 that satisfy (12) and (13). To prove

$|s_n^2 - \sigma^2 n| = O(n^{1-\epsilon})$ we remark that (8) together with a routine estimate implies $|\sigma^2 - b_k^2| = O(|I_k|^{-A+1})$. Thus the first part of (11) follows from

$$|s_n^2 - \sigma^2 n| \le \sigma^2 \sum_{\substack{k\le n\\ k\in J}} 1 + \sum_{\substack{k\le n\\ k\in I}} |\sigma^2 - b_k^2|,$$

where I and J are the intergers within the long and short blocks, respectively.

To finish the proof we have to show that the fluctuations of the partial sums and the Wiener processes W_1 and W_2 within the blocks I_k are small enough. Since fluctuation properties of Wiener processes are easy to handle using standard deviation inequalities (see e.g. [9]), we only investigate the partial sums. By Corollary 1 we have

$$\begin{aligned} P\left(\sup_{\underline{i}_k \le \ell \le \bar{i}_k} \left|\sum_{j=\underline{i}_k}^{\ell} Y_j\right| > \underline{i}_k^{\frac{1+\eta}{p}}\right) &\le \underline{i}_k^{-(1+\eta)} E\left(\sup_{\underline{i}_k \le \ell \le \bar{i}_k} \left|\sum_{j=\underline{i}_k}^{\ell} Y_j\right|^p\right) \\ &\ll \underline{i}_k^{-(1+\eta)} |I_k|^{p/2} \\ &\ll k^{-(1+\eta)(1+\alpha)+\frac{\alpha p}{2}} \\ &= O\left(k^{-(1+\varepsilon_2)}\right), \end{aligned}$$

if $\varepsilon_2 > 0$ is chosen sufficiently small. The Borel-Cantelli lemma shows that we can also control the fluctuation within the blocks. Thus (11) is proven.

The proof of Theorem 2 is similar to the proof of Theorem 1 and will be therefore omitted. We only remark that under the exponential mixing rate logarithmic block sizes are required in the blocking argument.

6.3 Proof of Proposition 1

We use the notation introduced in the proof of Theorem 1. Further we let $I = I_1 \cup I_2 \cup \cdots$ and $J = J_1 \cup J_2 \cup \cdots$ and $M_n = \{1, \ldots, n\}$. By looking at the proof of Theorem 1, it readily follows that

$$\begin{aligned} \frac{1}{s_n} \sum_{i\in I\cap M_n} Y_i &= W_1(s_n^2)/s_n - X_n, \\ \frac{1}{t_m} \sum_{j\in J\cap M_n} Y_j &= W_2(t_m^2)/t_m - Z_m, \end{aligned}$$

where

$$X_n = o\Big((s_n^2)^{\frac{1+\eta}{p}-\frac{1}{2}}\Big) = o(1) \quad \text{a.s.} \quad \text{and} \quad Z_m = o\Big((t_m^2)^{\frac{1+\eta}{p}-\frac{1}{2}}\Big) = o(1) \quad \text{a.s.} \qquad (45)$$

Hence

$$\begin{aligned} &\mathrm{Corr}\big(W_1(s_n^2), W_2(t_m^2)\big) \\ &\quad = \mathrm{Corr}\left(\frac{1}{s_n} W_1(s_n^2), \frac{1}{t_m} W_2(t_m^2)\right) \\ &\quad = \mathrm{Corr}\left(\frac{1}{s_n} \sum_{i\in I\cap M_n} Y_j + X_n, \frac{1}{t_m} \sum_{j\in J\cap M_n} Y_j + Z_m\right). \end{aligned}$$

In order to calculate this correlation we need a couple of estimates.

First we note that by definition of s_n^2 and t_n^2

$$s_n^2 \sim \sigma^2 |I \cap M_n| \quad \text{and} \quad t_m^2 \sim \sigma^2 |J \cap M_m|. \tag{46}$$

It readily follows from Proposition 6.1 that

$$\left\| \frac{1}{s_n} \sum_{i \in I \cap M_n} Y_i \right\|_p \leq C_p, \tag{47}$$

where C_p does not depend on n. Thus

$$\begin{aligned} \sup_{n \geq 1} \|X_n\|_p &= \sup_{n \geq 1} \left\| \frac{1}{s_n} \sum_{i \in I \cap M_n} Y_i - W_1(s_n^2)/s_n \right\|_p \\ &\leq \sup_{n \geq 1} \left\| \frac{1}{s_n} \sum_{i \in I \cap M_n} Y_i \right\|_p + \left\| W_1(1) \right\|_p < \infty, \end{aligned}$$

and hence $\{X_n^2\}$ is uniformly integrable. This and (45) show that $\operatorname{Var}(X_n) \to 0$; by the same arguments $\operatorname{Var}(Z_m) \to 0$. By (44)

$$\left\| \frac{1}{s_n} \sum_{i \in I \cap M_n} Y_i \right\|_2 \sim \sigma^2. \tag{48}$$

Thus by (47) and (48)

$$\begin{aligned} c_1(m,n) &:= \operatorname{Cov}\left(Z_m, \frac{1}{s_n} \sum_{i \in I \cap M_n} Y_i \right) \\ &\leq \operatorname{Var}^{1/2}(Z_m) \operatorname{Var}^{1/2}\left(\frac{1}{s_n} \sum_{i \in I \cap M_m} Y_i \right) \\ &= o(1) \quad \text{for } m, n \to \infty, \end{aligned}$$

and similarly

$$c_2(m,n) := \operatorname{Cov}\left(X_n, \frac{1}{t_m} \sum_{j \in J \cap M_m} Y_j \right) = o(1) \quad \text{for } m, n \to \infty.$$

Furthermore we have

$$\begin{aligned} B_1(n) &:= \operatorname{Var}^{1/2}\left(\frac{1}{s_n} \sum_{i \in I \cap M_n} Y_i + X_n \right) \\ &\geq \operatorname{Var}^{1/2}\left(\frac{1}{s_n} \sum_{i \in I \cap M_n} Y_i \right) - \operatorname{Var}^{1/2}(X_n) \\ &= \sigma + o(1) \quad \text{for } n \to \infty, \end{aligned}$$

and

$$B_2(m) := \mathrm{Var}^{1/2}\left(\frac{1}{t_m}\sum_{j\in J\cap M_m} Y_j + Z_m\right)$$
$$\geq \sigma + o(1) \quad \text{for } m\to\infty.$$

Finally we introduce the term

$$c_0(m,n) = \frac{1}{s_n t_m}\sum_{i\in I\cap M_n}\sum_{j\in J\cap M_m} \mathrm{Cov}\big(Y_i, Y_j\big).$$

We choose $r \geq 0$ such that $n \in I_{r+1}\cup J_{r+1}$ and we choose $v \geq 0$ such that $m \in I_{v+1}\cup J_{v+1}$ and recall that by Theorem 1 $\sum_{i\in\mathbb{Z}} |\operatorname{Cov}(Y_0, Y_i)| < \infty$. Hence if $v \leq 2r$ we have

$$\begin{aligned}
c_0(m,n) &\leq s_{\bar{i}_r}^{-1} t_{\bar{i}_v}^{-1} \sum_{i\in I_1\cup\cdots\cup I_{r+1}}\ \sum_{j\in J_1\cup\cdots\cup J_{v+1}} |\operatorname{Cov}\big(Y_i, Y_j\big)| \\
&\leq s_{\bar{i}_r}^{-1} t_{\bar{i}_v}^{-1} \sum_{j\in J_1\cup\cdots\cup J_{v+1}}\ \sum_{i\in\mathbb{Z}} |\operatorname{Cov}\big(Y_i, Y_j\big)| \\
&= s_{\bar{i}_r}^{-1} t_{\bar{i}_v}^{-1} \sum_{j\in J_1\cup\cdots\cup J_{v+1}}\ \sum_{i\in\mathbb{Z}} |\operatorname{Cov}\big(Y_i, Y_0\big)| \\
&\ll s_{\bar{i}_r}^{-1} t_{\bar{i}_v}^{-1}\big(|J_1| + \cdots + |J_{v+1}|\big) \\
&\ll s_{\bar{i}_r}^{-1} t_{\bar{i}_v}^{-1} t_{\bar{i}_{v+1}}^{2} = o(1) \quad \text{as } m,n\to\infty.
\end{aligned}$$

If $v > 2r$ we have to additionally show that

$$s_{\bar{i}_r}^{-1} t_{\bar{i}_v}^{-1} \sum_{i\in I_1\cup\cdots\cup I_{r+1}}\ \sum_{j\in J_{2r+1}\cup\cdots\cup J_{v+1}} |\operatorname{Cov}\big(Y_i, Y_j\big)| \to 0.$$

Now we have by (32) and assumptions (8), (9) that

$$\begin{aligned}
&s_{\bar{i}_r}^{-1} t_{\bar{i}_v}^{-1} \sum_{i\in I_1\cup\cdots\cup I_{r+1}}\ \sum_{j\in J_{2r+1}\cup\cdots\cup J_{v+1}} |\operatorname{Cov}\big(Y_i, Y_j\big)| \\
&\leq s_{\bar{i}_r}^{-1} t_{\bar{i}_v}^{-1} \sum_{\pi\geq 2r+1} |J_\pi| \sum_{\ell=1}^{r+1} |I_\ell| \big(d(I_\ell, J_\pi)\big)^{-1} \\
&\ll s_{\bar{i}_r}^{-1} t_{\bar{i}_v}^{-1} \sum_{\pi\geq 2r+1} \pi^\beta \sum_{\ell=1}^{r+1} \ell^\alpha \big(d(I_\ell, J_\pi)\big)^{-1}.
\end{aligned}$$

For $\ell \in \{1,\ldots,r+1\}$ and $\pi \geq 2r+1$ we have constants k_0 and k_1 independent of r and π such that

$$d(I_\ell, J_\pi) \geq k_0\big(\pi^{\alpha+1} - r^{\alpha+1}\big) \geq k_1 \pi^{\alpha+1}$$

and thus

$$\begin{aligned}
&s_{\bar{i}_r}^{-1}t_{\bar{i}_v}^{-1}\sum_{\pi\geq 2r+1}\pi^{\beta}\sum_{\ell=1}^{r+1}\ell^{\alpha}\big(d(I_\ell,J_\pi)\big)^{-1}\\
&\quad\ll s_{\bar{i}_r}^{-1}t_{\bar{i}_v}^{-1}r^{\alpha+1}\sum_{\pi\geq 2r+1}\pi^{\beta-\alpha-1}\\
&\quad\ll s_{\bar{i}_r}^{-1}t_{\bar{i}_v}^{-1}r^{1+\beta}\ll r^{-(\alpha-\beta)/2}=o(1)\quad\text{as } r\to\infty.
\end{aligned}$$

Using the definitions of c_0, c_1, c_2 and B_1 and B_2 we see that

$$\mathrm{Corr}\big(W_1(s_n^2),W_2(t_m^2)\big)=\frac{c_0(m,n)+c_1(m,n)+c_2(m,n)+\mathrm{cov}(X_n,Z_m)}{B_1(n)B_2(m)}. \tag{49}$$

We have shown that $c_0(m,n)+c_1(m,n)+c_2(m,n)+\mathrm{cov}(X_n,Z_m)\to 0$ as $m,n\to\infty$ while the denominator in (49) is bounded away from zero. This finishes the proof of Proposition 1.

Acknowledgement. The authors are indebted to Wei Biao Wu for raising the question leading to the present research and for valuable comments. We also thank two anonymous referees for several inspiring questions and remarks which led to a considerable improvement of the presentation.

References

[1] D.W.K. Andrews. Non strong mixing autoregressive processes. *J. Appl. Probab.* **21**, 930–934 (1984).

[2] A. Aue, I. Berkes, and L. Horváth. Strong approximation for the sums of squares of augmented GARCH sequences. *Bernoulli* **12**, 583–608 (2006).

[3] I. Berkes, S. Hörmann, and J. Schauer. Asymptotic results for the empirical process of stationary sequences. *Stoch. Proc. Appl.* **119**, 1298–1324 (2009).

[4] I. Berkes and L. Horváth. Strong approximation of the empirical process of GARCH sequences. *Ann. Appl. Probab.* **11** 789–809 (2001).

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

[6] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327 (1986).

[7] R.C. Bradley. *Introduction to Strong Mixing Conditions, Vol. I-III.* Kendrick Press, 2007.

[8] R.M. Burton, M. Goulet, and R. Meester. On 1-dependent processes and k-block factors. *Ann. Probab.* **21**, 2157–2168 (1993).

[9] M. Csörgő and P. Révész. *Strong Approximations in Probability and Statistics.* Academic Press, 1981.

[10] J. Dedecker, P. Doukhan, G. Lang, J. R. León, S. Louhichi, C. Prieur. *Weak dependence with examples and applications.* Lecture Notes in Statistics 190. Springer, 2007.

[11] J. Dedecker and C. Prieur. Coupling for τ-dependent sequences and applications. *J. Theor. Probab.* **17**, 861–885 (2004).

[12] J. Dedecker and C. Prieur. New dependence coefficients. Examples and applications to statistics. *Prob. Theory Related Fields* **132**, 203–236 (2005).

[13] P. Diaconis and D. Freedman. Iterated random functions. *SIAM Review* **41**, 45–76 (1999).

[14] Z. Ding, C. W. J. Granger and R. Engle. A long memory property of stock market returns and a new model. *J. Emp. Finance* **1**, 83–106 (1993).

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

[16] J.C. Duan. Augmented GARCH(p,q) process and its diffusion limit. *J. Econometrics* **79**, 97–127 (1997).

[17] U. Einmahl. Strong invariance principles for partial sums of independent random vectors. *Ann. Probab.* **15**, 1419–1440 (1987).

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

[19] L. Giraitis, P. M. Robinson and D. Surgailis. Long memory conditional heteroscedasticity. *Ann. Appl. Probab.* **10**, 1002–1024 (2000).

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

[21] I. A. Ibragimov. Some limit theorems for stationary processes. *Teor. Verojatnost. i Primenen.* **7**, 361–392 (1962).

[22] 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).

[23] M. Kac. On the distribution of values of sums of the type $\sum f(2^k t)$. *Ann. of Math.* **47**, 33–49 (1946).

[24] J. Komlós, P. Major and G. Tusnády. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32**, 111–131 (1975).

[25] J. Komlós, P. Major and G. Tusnády. An approximation of partial sums of independent RV's and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34**, 33–58 (1976).

[26] D. Lien and Y. Tse. Forecasting the Nikkei spot index with fractional cointegration. *J. Forecasting* **18**, 259–273 (1999).

[27] W. Liu and Z. Lin. Strong approximation for a class of stationary processes. *Stoch. Proc. Appl.* **119**, 249–280 (2009).

[28] P. Major. An approximation of partial sums of independent RV's. *Z. Wahrsch. Verw. Gebiete* **35**, 213–220 (1976).

[29] P. Major. An improvement of Strassen's invariance principle. *Ann. Probab.* **1**, 55–61 (1979).

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

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

[32] F. Móricz. Moment inequalities and the strong laws of large numbers. *Z. Wahrsch. Verw. Gebiete* **35**, 299–314 (1976).

[33] D. B. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**, 347–370 (1991).

[34] B. M. Pötscher and I. R. Prucha. *Dynamic Nonlinear Econometric Models*. Springer, 1997.

[35] A. Račkauskas and C. Suquet. Hölder norm test statistics for epidemic change. *J. Statist. Plann. Inference* **126**, 495–520 (2004).

[36] E. Rio. Processus empriques absoluments réguliers et entropie universelle. *Prob. Theory Related Fields* **111**, 585-608 (1998).

[37] E. Rio. Théorie asymptotique des processus aléatoires faiblement dépendants. *Collection Mathématiques & Applications* **31**. Springer 2000.

[38] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. USA* **42**, 43–47 (1956).

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

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

[41] M. Rosenblatt. *Markov Processes. Structure and Asymptotic Behavior*. Springer, 1971.

[42] A. I. Sakhanenko. Estimates in the invariance principle. In: *Proceedings of the Institute of Mathematics, vol. 5,* 27–44, (1985).

[43] Q. M. Shao. Strong approximation theorems for independent random variables and their applications. *J. Multivariate Anal.* **52**, 107–130 (1995).

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

[45] G. Shorack and J. Wellner. *Empirical Processes with Applications to Statistics.* Wiley, 1986.

[46] V. Strassen. An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **3** , 211–226 (1964).

[47] V. Strassen. Almost sure behavior of sums of independent random variables and martingales. *Proc. Fifth Berkeley Sympos. Math. Statist. Probab.,* Vol. II, Part 1, pp. 315–343. Univ. California Press 1967.

[48] V. de Valk. Hilbert space representations of m-dependent processes. *Ann. Probab.* **21**, 1550–1570 (1993).

[49] Q. Wang, Y.-X. Lin, and C. M. Gulati. Strong approximation for long memory processes with applications. *J. Theoret. Probab.* **16**, 377–389 (2003).

[50] W. B. Wu. Nonlinear system theory: Another look at dependence. *Proc. Nat. Acad. Sci. USA* **102**, 14150–14154 (2005).

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

[52] W. B. Wu. Strong invariance principles for dependent random variables. *Ann. Probab.* **35**, 2294–2320 (2007).

[53] D. Zholud. Extremes of Shepp statistics for the Wiener process. *Extremes* **11**, 339–351, 2008.