

# The notion of $\psi$ -weak dependence and its applications to bootstrapping time series

Paul Doukhan, Michael Neumann

► **To cite this version:**

Paul Doukhan, Michael Neumann. The notion of  $\psi$ -weak dependence and its applications to bootstrapping time series. *Probability Surveys*, Institute of Mathematical Statistics (IMS), 2008, 5, pp.146-168. 10.1214/06-PS086 . hal-00141589

**HAL Id: hal-00141589**

**<https://hal.archives-ouvertes.fr/hal-00141589>**

Submitted on 3 May 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THE NOTION OF WEAK DEPENDENCE AND ITS APPLICATIONS TO  
BOOTSTRAPPING TIME SERIES

Paul Doukhan

Laboratoire de Statistique du CREST, Timbre J340  
3, avenue Pierre Larousse, 92240 MALAKOFF, and  
SAMOS, Statistique Appliquée et Modélisation Stochastique,  
Université Paris 1, Centre Pierre Mendès France,  
90 rue de Tolbiac, F-75634 Paris Cedex 13, France  
E-mail: doukhan@ensae.fr

Michael H. Neumann

Friedrich-Schiller-Universität Jena  
Institut für Stochastik  
Ernst-Abbe-Platz 2, 07743 Jena, Germany  
E-mail: mi.neumann@tu-bs.de

**Abstract**

We give an introduction to the notion of weak dependence which is more general than mixing and allows to treat for example processes driven by discrete innovations as they appear with time series bootstrap. As a typical example, we analyze autoregressive processes and their bootstrap analogues in detail and show how weak dependence can be easily derived from a contraction property of the process. Furthermore, we provide an overview of classes of processes possessing the property of weak dependence and describe important probabilistic results under such an assumption.

December 1, 2006

---

*2000 Mathematics Subject Classification.* Primary 60E15; secondary 62E99.

*Keywords and Phrases.* Autoregressive processes, autoregressive bootstrap, mixing, weak dependence.

*Short title.* Weak dependence.

## 1. MIXING VS. WEAK DEPENDENCE

For a long time mixing conditions have been the dominating type of conditions for imposing a restriction on the dependence between time series data. They are considered to be useful since they are fulfilled for many classes of processes and since they allow to derive tools similar to those in the independent case. On the other hand, it turns out that certain classes of processes which are of interest in statistics are not mixing although a successive decline of the influence of past states takes place. The simplest example of such a process is an AR(1)-process,  $X_t = \theta X_{t-1} + \varepsilon_t$ , where the innovations are independent and identically distributed with  $P(\varepsilon_t = 1) = P(\varepsilon_t = -1) = 1/2$  and  $0 < \theta \leq 1/2$ ; see also Rosenblatt (1980). It is clear that this process has a stationary distribution supported on  $[-2, 2]$ , and for a process in the stationary regime, it can be seen from the equality  $X_t = \varepsilon_t + \theta\varepsilon_{t-1} + \dots + \theta^{t-s-1}\varepsilon_{s+1} + \theta^{t-s}X_s$  that a past state  $X_s$  can always be recovered from  $X_t$ . (Actually, since  $|\varepsilon_t| > \theta|\varepsilon_{t-1}| + \dots + \theta^{t-s-1}|\varepsilon_{s+1}| + \theta^{t-s}|X_s|$   $X_t$  has always the same sign as  $\varepsilon_t$  which means that we can recover  $\varepsilon_t$  and therefore  $X_{t-1}$  from  $X_t$ . Continuing in this way we can finally compute  $X_s$ .) This, however, excludes any of the commonly used mixing properties to hold. On the other hand,  $X_s$  loses its impact on  $X_t$  as  $t \rightarrow \infty$ .

Besides this somehow artificial example, there are many other processes of this type which are of great interest in statistics. For example, for bootstrapping a linear autoregressive process of finite order, it is most natural to estimate first the distribution of the innovations by the empirical distribution of the (possibly re-centered) residuals and to generate then a bootstrap process iteratively by drawing independent bootstrap innovations from this distribution. Now it turns out that commonly used techniques to prove mixing for autoregressive processes fail; because of the discreteness of the bootstrap innovations it is in general impossible to construct a coupling of two processes with different initial values.

Inspired by such problems, Doukhan and Louhichi (1999) and Bickel and Bühlmann (1999) introduced the alternative notions of weak dependence and  $\nu$ -mixing, respectively, which focus on covariances rather than the total variation norm between the joint distribution and the product of marginal distributions of random variables. A slightly simplified version of Doukhan and Louhichi's (1999) definition is given here:

**Definition 1.1.** A process  $(X_t)_{t \in \mathbb{Z}}$  is called *weakly dependent* if there exists a universal null sequence  $(\epsilon_r)_{r \in \mathbb{N}}$  such that, for any  $k$ -tuple  $(s_1, \dots, s_k)$  and any  $l$ -tuple  $(t_1, \dots, t_l)$  with  $s_1 \leq \dots \leq s_k < s_k + r = t_1 \leq \dots \leq t_l$  and arbitrary measurable functions  $g : \mathbb{R}^k \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^l \rightarrow \mathbb{R}$  with  $\|g\|_\infty \leq 1$  and  $\|h\|_\infty \leq 1$ , the following inequality is fulfilled:

$$|\text{cov}(g(X_{s_1}, \dots, X_{s_k}), h(X_{t_1}, \dots, X_{t_l}))| \leq \psi(k, l, \text{Lip } g, \text{Lip } h) \epsilon_r.$$

Here  $\text{Lip } h$  denotes the Lipschitz modulus of continuity of  $h$ , that is,

$$\text{Lip } h = \sup_{x \neq y} \frac{|h(x) - h(y)|}{\|x - y\|_{l_1}},$$

where  $\|z\|_{l_1} = \sum_i |z_i|$ , and  $\psi : \mathbb{N}^2 \times \mathbb{R}_+^2 \rightarrow [0, \infty)$  is an appropriate function.

*Remark 1.* In Bickel and Bühlmann (1999), another type of weak dependence, called  $\nu$ -mixing, was introduced. Similarly to Definition

*Remark 2. (Some classes of weak dependence)*

Specific functions  $\psi$  yield notions of weak dependence appropriate to describe various examples of models:

- $\kappa$ -weak dependence for which  $\psi(u, v, a, b) = uvab$ ; in this case we simply denote  $\epsilon_r$  as  $\kappa_r$ .
- $\kappa'$  (causal) weak dependence for which  $\psi(u, v, a, b) = vab$ ; in this case we simply denote  $\epsilon_r$  as  $\kappa'_r$ . This is the causal counterpart of  $\kappa$  coefficients which we recall only for completeness.
- $\eta$ -weak dependence,  $\psi(u, v, a, b) = ua + vb$ ; in this case we write  $\epsilon_r = \eta_r$  for short.
- $\theta$ -weak dependence is a causal dependence which refers to  $\psi(u, v, a, b) = vb$ ; we simply denote  $\epsilon_r = \theta_r$  (Dedecker and Doukhan (2003)) for this causal counterpart of  $\eta$  coefficients.
- $\lambda$ -weak dependence  $\psi(u, v, a, b) = uvab + ua + vb$ ; in this case we write  $\epsilon_r = \lambda_r$  (Doukhan and Wintenberger (2006)). Besides the fact that it includes  $\eta$ - and  $\kappa$ -weak dependence, this new notion of  $\lambda$ -weak dependence is convenient, for example, for Bernoulli shifts with associated inputs.

It turns out that the notion of weak dependence is more general than mixing and allows to treat, for example, also Markovian processes driven

by discrete innovations as they appear with time series bootstrap. In the next section we consider as an instructive example linear autoregressive processes of finite order and a corresponding bootstrap version thereof. We will demonstrate that the desired property of weak dependence readily follows from a contraction property which is typical for such models under standard conditions on the parameters. The approach described there is also applicable to proving weak dependence for many other classes of processes. Section

## 2. AUTOREGRESSIVE PROCESSES AND THEIR BOOTSTRAP ANALOGUES

In this section we intend to give a gentle introduction to the basic ideas commonly used for verifying weak dependence. Most parts in this section are specialized to autoregressive processes of finite order and their bootstrap analogues.

We consider first a general real-valued stationary process  $(X_t)_{t \in \mathbb{Z}}$ . A simple and in many cases the most promising way of proving a property of weak dependence is via contraction arguments. For probability distributions  $P$  and  $Q$  on  $(\mathbb{R}^d, \mathcal{B}^d)$  with finite mean, we define the metric

$$d(P, Q) = \inf_{(X, Y): X \sim P, Y \sim Q} \mathbb{E} \|X - Y\|_{l_1}.$$

For  $d = 1$  and the  $L_2$  instead of the  $L_1$  distance, we obtain Mallows distance; see Mallows (1972). It is well known that such distances are suitable for metrizing weak convergence, that is,  $d(P_n, P) \xrightarrow{n \rightarrow \infty} 0$  implies  $P_n \Longrightarrow P$ ; see e.g. Bickel and Freedman (1981). Similar distances have also been used in the context of Markov processes to derive convergence of stationary distributions from convergence of the conditional distributions; see e.g. Dobrushin (1970) and Neumann and Paparoditis (2005). The following lemma shows that closeness of the conditional distributions in the above metric gives rise to estimates for covariances.

**Lemma 2.1.** *Suppose that  $(X_t)_{t \in \mathbb{Z}}$  is a real-valued stationary process. Furthermore, let  $s_1 \leq \dots \leq s_k \leq t_1 \leq \dots \leq t_l$  be arbitrary and let  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  and  $h : \mathbb{R}^l \rightarrow \mathbb{R}$  be measurable functions. Let  $(X'_t)_{t \in \mathbb{Z}}$  be another version of the process, where  $\mathbb{X}_{s_k} = (X_{s_k}, X_{s_k-1}, \dots)$  is independent of  $\mathbb{X}'_{s_k} =$*

$(X'_{s_k}, X'_{s_k-1}, \dots)$ . If  $g$  is bounded and  $\mathbb{E}|h(X_{t_1}, \dots, X_{t_l})| < \infty$ , then

$$\begin{aligned} & |\text{cov}(g(X_{s_1}, \dots, X_{s_k}), h(X_{t_1}, \dots, X_{t_l}))| \\ & \leq \|g\|_\infty \text{Lip } h \mathbb{E}d\left(P^{X_{t_1}, \dots, X_{t_l} | \mathbb{X}_{s_k}}, P^{X'_{t_1}, \dots, X'_{t_l} | \mathbb{X}'_{s_k}}\right). \end{aligned}$$

This lemma shows that a property of weak dependence follows from a convergence of the conditional distributions as the time gap to the lagged variables tends to infinity. The latter property can often be shown by appropriate coupling arguments. Note that there is a close connection to the notion of  $\tau$ -dependence introduced by Dedecker and Prieur (2004). According to their Lemma 5,  $\mathbb{E}d(P^{X_t | \mathbb{X}_s}, P^{X'_t | \mathbb{X}'_s})$  is actually equal to their coefficient  $\tau(t-s)$ . Dedecker and Prieur used such coupling arguments to derive exponential inequalities and other interesting results, with applications to density estimation.

In the rest of this section we restrict our attention to a real-valued autoregressive process  $(X_t)_{t \in \mathbb{Z}}$ , which obeys the equation

$$X_t = \theta_1 X_{t-1} + \dots + \theta_p X_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z}. \quad (2.1)$$

The innovations  $(\varepsilon_t)_{t \in \mathbb{Z}}$  are assumed to be independent and identically distributed with  $\mathbb{E}\varepsilon_t = 0$ . Furthermore, we make the standard assumption that the characteristic polynomial  $\theta(z) = 1 - \theta_1 z - \dots - \theta_p z^p$  has no zero in the unit circle. It is well known that there exists then a stationary solution to the model equation (

Convergence of  $\mathbb{E}d(P^{X_{t_1}, \dots, X_{t_l} | \mathbb{X}_{s_k}}, P^{X'_{t_1}, \dots, X'_{t_l} | \mathbb{X}'_{s_k}})$  as  $t_1 - s_k \rightarrow \infty$  can now be shown by a simple coupling argument. For this purpose, we consider a second (stationary) version of the autoregressive process,  $(X'_t)_{t \in \mathbb{Z}}$ , where  $\mathbb{X}'_{s_k}$  is independent of  $\mathbb{X}_{s_k}$ . Note that  $(X'_t)_{t \in \mathbb{Z}}$  can also be written as a linear process,

$$X'_t = \sum_{k=0}^{\infty} \alpha_k \varepsilon'_{t-k}.$$

Independence of  $\mathbb{X}'_{s_k}$  and  $\mathbb{X}_{s_k}$  is equivalent to the fact that  $\varepsilon_{s_k}, \varepsilon_{s_k-1}, \dots$  and  $\varepsilon'_{s_k}, \varepsilon'_{s_k-1}, \dots$  are independent. On the other hand, we have some freedom to couple the innovations after time  $s_k$ . Here we only have to take care that both sequences  $(\varepsilon_t)_{t \in \mathbb{Z}}$  and  $(\varepsilon'_t)_{t \in \mathbb{Z}}$  consist of independent random variables. A reasonably good coupling is obtained by feeding both processes after time  $s_k$  with one and the same sequence of innovations, that is,

$\varepsilon'_{s_k+1} = \varepsilon_{s_k+1}, \varepsilon'_{s_k+2} = \varepsilon_{s_k+2}, \dots$ . This gives that

$$X_t - X'_t = \sum_{l=0}^{\infty} \alpha_{t-s_k+l} (\varepsilon_{s_k-l} - \varepsilon'_{s_k-l}). \quad (2.2)$$

Since  $d(P^{X_{t_1}, \dots, X_{t_l} | \mathbb{X}_{s_k}}, P^{X'_{t_1}, \dots, X'_{t_l} | \mathbb{X}'_{s_k}}) \leq \mathbb{E} \left( \sum_{j=1}^l |X_{t_j} - X'_{t_j}| \middle| \mathbb{X}_{s_k}, \mathbb{X}'_{s_k} \right)$  we obtain in conjunction with (

**Lemma 2.2.** *Let  $(X_t)_{t \in \mathbb{Z}}$  and  $(X'_t)_{t \in \mathbb{Z}}$  be two versions of the autoregressive process as described above. If  $\mathbb{E}|\varepsilon_0| < \infty$ , then*

$$\mathbb{E}d \left( P^{X_{t_1}, \dots, X_{t_l} | \mathbb{X}_{s_k}}, P^{X'_{t_1}, \dots, X'_{t_l} | \mathbb{X}'_{s_k}} \right) \leq 2l K_\epsilon \frac{1}{1 - \rho_\epsilon} \rho_\epsilon^{t_1 - s_k} \mathbb{E}|\varepsilon_0|.$$

Lemma

**Corollary 2.1.** *Suppose that  $(X_t)_{t \in \mathbb{Z}}$  is a stationary process satisfying the above conditions. Furthermore, let  $s_1 \leq \dots \leq s_k \leq t_1 \leq \dots \leq t_l$  be arbitrary and let  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  and  $h : \mathbb{R}^l \rightarrow \mathbb{R}$  be measurable functions. If  $g$  is bounded and  $\mathbb{E}|h(X_{t_1}, \dots, X_{t_l})| < \infty$ , then*

$$|\text{cov}(g(X_{s_1}, \dots, X_{s_k}), h(X_{t_1}, \dots, X_{t_l}))| \leq \|g\|_\infty \text{Lip } h K_1 l \rho_\epsilon^{t_1 - s_k},$$

where  $K_1 = 2 \frac{1}{1 - \rho_\epsilon} K_\epsilon \mathbb{E}|\varepsilon_0|$ .

Now we define the autoregressive bootstrap. We assume that observations  $X_{1-p}, \dots, X_n$  are available. Let  $\hat{\theta}_n = (\hat{\theta}_{n,1}, \dots, \hat{\theta}_{n,p})'$  be any consistent estimator of  $\theta = (\theta_1, \dots, \theta_p)'$ , that is,  $\hat{\theta}_n \xrightarrow{P} \theta$ , as  $n \rightarrow \infty$ . (The least squares and the Yule-Walker estimator are even  $\sqrt{n}$ -consistent.) Let  $\mathbf{X}_t = (X_{t-1}, \dots, X_{t-p})'$  be the vector of the  $p$  lagged observations at time  $t$ . We define residuals

$$\tilde{\varepsilon}_t = X_t - \mathbf{X}_t' \hat{\theta}_n$$

and re-center them as

$$\hat{\varepsilon}_t = \tilde{\varepsilon}_t - \tilde{\varepsilon}.,$$

where  $\tilde{\varepsilon} = (1/n) \sum_{t=1}^n \tilde{\varepsilon}_t$ . Now we draw independent bootstrap innovations  $\varepsilon_t^*$  from the empirical distribution  $P_n$  given by the  $\hat{\varepsilon}_t$ . A bootstrap version of the autoregressive process is now obtained as

$$X_t^* = \hat{\theta}_{n,1} X_{t-1}^* + \dots + \hat{\theta}_{n,p} X_{t-p}^* + \varepsilon_t^*. \quad (2.3)$$

For simplicity, we assume that  $(X_t^*)_{t \in \mathbb{Z}}$  is in its stationary regime. (This will be justified by (i) of the next lemma.) Before we state weak dependence

of the bootstrap process, we show that it inherits those properties from the initial process which were used for proving weak dependence.

**Lemma 2.3.** *Suppose that  $\mathbb{E}\varepsilon_0^2 < \infty$  and  $\widehat{\theta}_n \xrightarrow{P} \theta$  are fulfilled.*

- (i) *With a probability tending to 1,  $(X_t^*)_{t \in \mathbb{Z}}$  can be written as a stationary causal linear process,*

$$X_t^* = \sum_{k=0}^{\infty} \widehat{\alpha}_{n,k} \varepsilon_{t-k}^*,$$

where, for all  $\epsilon > 0$ , there exists a  $\widetilde{K}_\epsilon < \infty$  such that, with  $\rho_\epsilon = (1 + \epsilon)/\rho$ ,

$$\mathbb{P}\left(|\widehat{\alpha}_{n,k}| \leq \widetilde{K}_\epsilon \rho_\epsilon^k \quad \forall k \in \mathbb{N}\right) \xrightarrow{n \rightarrow \infty} 1.$$

- (ii)  $\mathbb{E}(\varepsilon_t^{*2} | X_{1-p}, \dots, X_n) \xrightarrow{P} \mathbb{E}\varepsilon_t^2$ .

Armed with the basic properties stated in Lemma

**Lemma 2.4.** *Suppose that the initial process  $(X_t)_{t \in \mathbb{Z}}$  satisfies the above conditions and that the bootstrap process  $(X_t^*)_{t \in \mathbb{Z}}$  is in its stationary regime. Let  $(X_t^{*'})_{t \in \mathbb{Z}}$  be another version of the bootstrap process, where  $\mathbb{X}_{s_k}^* = (X_{s_k}^*, X_{s_k-1}^*, \dots)$  is (conditionally on  $X_{1-p}, \dots, X_n$ ) independent of  $\mathbb{X}_{s_k}^{*'} = (X_{s_k}^{*'}, X_{s_k-1}^{*'}, \dots)$ . For any  $\epsilon > 0$ , let  $\rho_\epsilon = (1 + \epsilon)/\rho$  and  $\widetilde{K}_\epsilon < \infty$  be an appropriate constant. Then there exists a sequence of events  $\Omega_n$  such that  $\mathbb{P}(\Omega_n) \xrightarrow{n \rightarrow \infty} 1$  and if  $\Omega_n$  occurs, then*

$$\mathbb{E}d\left(P^{X_{t_1}^*, \dots, X_{t_l}^* | \mathbb{X}_{s_k}^*}, P^{X_{t_1}^{*'}, \dots, X_{t_l}^{*'} | \mathbb{X}_{s_k}^{*'}}\right) \leq 2l \widetilde{K}_\epsilon \rho_\epsilon^{t_1 - s_k} \sqrt{\mathbb{E}\varepsilon_0^2}.$$

From Lemma

**Corollary 2.2.** *Suppose that the conditions of Lemma*

*Let  $s_1 \leq \dots \leq s_k \leq t_1 \leq \dots \leq t_l$  be arbitrary and let  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  and  $h : \mathbb{R}^l \rightarrow \mathbb{R}$  be measurable functions. If  $g$  is bounded and  $\mathbb{E}|h(X_{t_1}^*, \dots, X_{t_l}^*)| < \infty$ , then*

$$\left| \text{cov}\left(g(X_{s_1}^*, \dots, X_{s_k}^*), h(X_{t_1}^*, \dots, X_{t_l}^*)\right) \right| \leq \|g\|_\infty \text{Lip } h K_2 l \rho_\epsilon^{t_1 - s_k},$$

where  $K_2 = 2 \widetilde{K}_\epsilon \sqrt{\mathbb{E}\varepsilon_0^2}$ .



Besides the useful property of weak dependence of the bootstrap process, the asymptotic validity of a bootstrap approximation requires that the (multivariate) stationary distributions of the bootstrap process converge to those of the initial process. Often, and in the case of the autoregressive bootstrap in particular, one has no direct access to these stationary distributions. However, according to Lemma 4.2 in Neumann and Paparoditis (2005), convergence of the stationary distributions can be derived from an appropriate convergence of conditional distributions. The latter, however, follows directly from  $\hat{\theta}_n \xrightarrow{P} \theta$  and  $\hat{\varepsilon}_t^* \xrightarrow{d} \varepsilon_t$ . Therefore, consistency of the autoregressive bootstrap can be shown by simple arguments which were already used for proving weak dependence of the bootstrap; for details see Section 4.2 in Neumann and Paparoditis (2005).

*Remark 3.* Motivated by the desire to have some sort of mixing for a smoothed sieve bootstrap for linear processes, Bickel and Bühlmann (1999) considered a condition called  $\nu$ -mixing which is similar to the notion of weak dependence in our Definition

In contrast, the approach described here is fundamentally different. We intend to prove weak dependence for processes driven by innovations with a possibly discrete distribution and achieve this goal by exploiting a contraction property of the initial and the bootstrap process.

*Remark 4.* Arguing in the same way as above we could also establish the property of weak dependence for nonlinear autoregressive processes,

$$X_t = m(X_t) + \varepsilon_t, \quad t \in \mathbb{Z},$$

where  $(\varepsilon_t)_{t \in \mathbb{Z}}$  is a sequence of independent and identically distributed innovations. If  $\text{Lip } m < 1$ , then we have obviously a contraction property being fulfilled which immediately yields weak dependence.

It is interesting to note that such a contraction property can still be proved if  $\text{Lip } m < 1$  is not fulfilled. To this end, define the *local* Lipschitz modulus of continuity

$$\Delta(x) = \sup_{y \neq x} \frac{|m(y) - m(x)|}{|y - x|}$$

and assume that

$$\rho = \sup_x \mathbb{E} [\Delta(m(x) + \varepsilon_0)] < 1.$$

Then

$$d\left(\mathbb{P}^{X_{t+k}|X_t=x}, \mathbb{P}^{X_{t+k}|X_t=y}\right) \leq \rho^{k-1} \cdot \Delta(x) \cdot |x - y|, \quad (2.4)$$

which implies weak dependence by Lemma

### 3. SOME EXAMPLES OF WEAKLY DEPENDENT SEQUENCES

Note first that sums of independent weakly dependent processes admit the common weak dependence property where dependence coefficients are the sums of the initial ones. We now provide a non-exhaustive list of weakly dependent sequences with their weak dependence properties. Further examples may be found in Doukhan and Louhichi (1999). Let  $X = (X_t)_{t \in \mathbb{Z}}$  be a stationary process.

- (1) If this process is either a Gaussian process or an associated process and  $\lim_{t \rightarrow \infty} |\text{cov}(X_0, X_t)| = 0$ , then it is a  $\kappa$ -weakly dependent process such that  $\kappa_r = O\left(\sup_{t \geq r} |\text{cov}(X_0, X_t)|\right)$ . It is also  $\kappa'$ -weakly dependent with  $\kappa'_r = O\left(\sum_{t \geq r} |\text{cov}(X_0, X_t)|\right)$ .
- (2)  $ARMA(p, q)$  processes and more generally causal or non-causal linear processes:  $X = (X_t)_{t \in \mathbb{Z}}$  is a such that

$$X_t = \sum_{k=-\infty}^{\infty} a_k \xi_{t-k} \quad \text{for } t \in \mathbb{Z},$$

where  $(a_k)_{k \in \mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$  and  $(\xi_t)_{t \in \mathbb{Z}}$  is a sequence of independent and identically distributed random variables with  $\mathbb{E}\xi_t = 0$ . If  $a_k = O(|k|^{-\mu})$  with  $\mu > 1/2$ , then  $X$  is a  $\eta$ -weakly dependent process with  $\eta_r = O\left(\frac{1}{r^{\mu-1/2}}\right)$ . In the general case of dependent innovations, properties of weak dependence are proved in Doukhan and Wintenberger (2006).

- (3)  $GARCH(p, q)$  processes and more generally  $ARCH(\infty)$  processes:  $X = (X_t)_{t \in \mathbb{Z}}$  is a such that

$$X_t = \rho_t \cdot \xi_t \quad \text{with} \quad \rho_t^2 = b_0 + \sum_{k=1}^{\infty} b_k X_{t-k}^2 \quad \text{for } k \in \mathbb{Z},$$

with a sequence  $(b_k)_k$  depending on the initial parameters in the case of a  $GARCH(p, q)$  process and a sequence  $(\xi_t)_{t \in \mathbb{Z}}$  of independent and

identically distributed innovations. Then, if  $\mathbb{E}(|\xi_0|^m) < \infty$ , with the condition of stationarity,  $\|\xi_0\|_m^2 \cdot \sum_{j=1}^{\infty} |b_j| < 1$ , and if:

- there exists  $C > 0$  and  $\mu \in ]0, 1[$  such that  $\forall j \in \mathbb{N}$ ,  $0 \leq b_j \leq C \cdot \mu^{-j}$ , then  $X$  is a  $\theta$ -weakly dependent process with  $\theta_r = O(e^{-c\sqrt{r}})$  and  $c > 0$  (this is the case of *GARCH*( $p, q$ ) processes).
- there exists  $C > 0$  and  $\nu > 1$  such that  $\forall k \in \mathbb{N}$ ,  $0 \leq b_k \leq C \cdot k^{-\nu}$ , then  $X$  is a  $\eta$ -weakly dependent process with  $\theta_r = O(r^{-\nu+1})$  (Doukhan, Teyssière and Winant (2006) introduce vector valued LARCH( $\infty$ ) models including the previous ones).

- (4) Causal bilinear processes introduced by Giraitis and Surgailis (2002) are such that

$$X_t = \xi_t \left( a_0 + \sum_{k=1}^{\infty} a_k X_{t-k} \right) + c_0 + \sum_{k=1}^{\infty} c_k X_{t-k}, \quad \text{for } k \in \mathbb{Z}.$$

Assume that there exists  $m \geq 1$  with  $\|\xi_0\|_m < \infty$  and  $\|\xi_0\|_m \cdot \left( \sum_{k=1}^{\infty} |a_k| + \sum_{k=1}^{\infty} |c_k| \right) < 1$ . Then, if:

- $\left\{ \begin{array}{l} \exists K \in \mathbb{N} \text{ such that } \forall k > K, a_k = c_k = 0, \text{ or,} \\ \exists \mu \in ]0, 1[ \text{ such that } \sum_k |c_k| \mu^{-k} \leq 1 \text{ and } \forall k \in \mathbb{N}, 0 \leq a_k \leq \mu^k \end{array} \right.$ , then  $X$  is a  $\theta$ -weakly dependent process with  $\theta_r = O(e^{-c\sqrt{r}})$ , for some  $c > 0$ ;
- $\forall k \in \mathbb{N}$ ,  $c_k \geq 0$ , and  $\exists \nu_1 > 2$  and  $\exists \nu_2 > 0$  such that  $a_k = O(k^{-\nu_1})$  and  $\sum_k c_k k^{1+\nu_2} < \infty$ , then  $X$  is a  $\theta$ -weakly dependent process with  $\eta_r = O\left(\left(\frac{r}{\log r}\right)^d\right)$ ,  $d = \max\left\{-\nu_1, -\frac{\nu_2 \cdot \delta}{\delta + \nu_2 \cdot \log 2}\right\}$  (see Doukhan, Teyssière and Winant (2006)).

- (5) Non-causal LARCH( $\infty$ ) processes  $X = (X_t)_{t \in \mathbb{Z}}$  satisfying

$$X_t = \xi_t \cdot \left( a_0 + \sum_{k \in \mathbb{Z} \setminus \{0\}} a_k X_{t-k} \right), \quad t \in \mathbb{Z},$$

where  $\|\xi_0\|_{\infty} < \infty$  (bounded random variables) and  $(a_k)_{k \in \mathbb{Z}}$  is a sequence of real numbers such that  $\lambda = \|\xi_0\|_{\infty} \cdot \sum_{j \neq 0} |a_j| < 1$  (stationarity condition). Assume that the sequence  $(a_k)_{k \in \mathbb{Z}}$  satisfies  $a_k = O(|k|^{-\mu})$  with  $\mu > 1$ , then  $X$  is a  $\eta$ -weakly dependent process with  $\eta_r = O\left(\frac{1}{r^{\mu-1}}\right)$  (see Doukhan, Teyssière and Winant (2006)).

- (6) Causal and non-causal Volterra processes write as  $X_t = \sum_{p=1}^{\infty} Y_t^{(p)}$  with

$$Y_t^{(p)} = \sum_{\substack{j_1 < j_2 < \dots < j_p \\ j_1, \dots, j_p \in \mathbb{Z}}} a_{j_1, \dots, j_p} \xi_{t-j_1} \cdots \xi_{t-j_p}, \quad \text{for } t \in \mathbb{Z}.$$

Assume  $\sum_{p=0}^{\infty} \sum_{\substack{j_1 < j_2 < \dots < j_p \\ j_1, \dots, j_p \in \mathbb{Z}}} |a_{j_1, \dots, j_p}|^m \|\xi_0\|_m^p < \infty$ , with  $m > 0$ , and

that there exists  $p_0 \in \mathbb{N} \setminus \{0\}$  such that  $a_{j_1, \dots, j_p} = 0$  for  $p > p_0$ . If  $a_{j_1, \dots, j_p} = O\left(\max_{1 \leq i \leq p} \{|j_i|^{-\mu}\}\right)$  with  $\mu > 0$ , then  $X$  is a  $\eta$ -weakly dependent process with  $\eta_r = O\left(\frac{1}{r^{\mu+1}}\right)$  (see Doukhan (2002)).

Finite order Volterra processes with dependent inputs are also considered in Doukhan and Wintenberger (2006): again,  $\eta$ -weakly dependent innovations yields  $\eta$ -weak dependence and  $\lambda$ -weakly dependent innovations yields  $\lambda$ -weak dependence of the process.

- (7) Very general models are the causal or non-causal infinite memory processes  $X = (X_t)_{t \in \mathbb{Z}}$  such that

$$X_t = F(X_{t-1}, X_{t-2}, \dots; \xi_t), \quad \text{and} \quad X_t = F(X_s, s \neq t; \xi_t),$$

where the functions  $F$  defined either on  $\mathbb{R}^{\mathbb{N} \setminus \{0\}} \times \mathbb{R}$  or  $\mathbb{R}^{\mathbb{Z} \setminus \{0\}} \times \mathbb{R}$  satisfy

$$\begin{aligned} \|F(0; \xi_0)\|_m &< \infty, \\ \|F((x_j)_j; \xi_0) - F((y_j)_j; \xi_0)\|_m &\leq \sum_{j \neq 0} a_j |x_j - y_j|, \end{aligned}$$

with  $a = \sum_{j \neq 0} a_j < 1$ . Then, works in progress by Doukhan and Wintenberger as well as Doukhan and Truquet, respectively, prove that a solution of the previous equations is stationary in  $L^m$  and either  $\theta$ -weakly dependent or  $\eta$ -weakly dependent with the following decay rate for the coefficients:

$$\inf_{p \geq 1} \left\{ a^{r/p} + \sum_{|j| > p} a_j \right\}.$$

This provides the same rates as those already mentioned for the cases of  $ARCH(\infty)$  or  $LARCH(\infty)$  models.

## 4. SOME PROBABILISTIC RESULTS

In this section, we present results derived under weak dependence which are of interest in probability and statistics. This collection clearly shows that the notion of weak dependence, although being more general than mixing, allows one to prove results very similar to those in the mixing case.

**4.1. Donsker invariance principle.** We consider a stationary, zero mean, and real valued sequence  $(X_t)_{t \in \mathbb{Z}}$  such that

$$\mu = \mathbb{E}|X_0|^m < \infty, \quad \text{for a real number } m > 2. \quad (4.1)$$

We also set

$$\sigma^2 = \sum_{t \in \mathbb{Z}} \text{cov}(X_0, X_t) = \sum_{t \in \mathbb{Z}} \mathbb{E}X_0X_t, \quad (4.2)$$

$W$  denotes the standard Brownian motion and

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} X_i, \quad \text{for } t \in [0, 1], \quad n \geq 1. \quad (4.3)$$

We now present versions of the Donsker weak invariance principle under weak dependence assumptions.

**Theorem 4.1** (Donsker type results). *Assume that the zero mean stationary process  $(X_t)_{t \in \mathbb{Z}}$  satisfies (*

**$\lambda$ -dependence.** *The process is  $\lambda$ -weakly dependent and satisfies  $\lambda_r = O(r^{-\lambda})$  (as  $r \uparrow \infty$ ) for  $\lambda > 4 + 2/(m - 2)$ .*

**$\theta$ -dependence.** *The process is  $\theta$ -weakly dependent and satisfies  $\theta_r = O(r^{-\theta})$  (as  $r \uparrow \infty$ ) for  $\theta > 1 + 1/(m - 2)$ .*

*Remark 5.* The result for  $\kappa'$ -weak dependence is obtained in Bulinski and Shashkin (2005). Results under  $\kappa$ - and  $\lambda$ -weak dependences are proved in Doukhan and Wintenberger (2006); note that  $\eta$ -weak dependence implies  $\lambda$ -weak dependence and the Donsker principle then holds under the same decay rate for the coefficients. The result for  $\theta$ -weak dependence is due to Dedecker and Doukhan (2003). A few comments on these results are now in order:

- The difference of the above conditions under  $\kappa$  and  $\kappa'$  assumptions is natural. The observed loss under  $\kappa$ -dependence is explained by the fact that  $\kappa'$ -weakly

dependent sequences satisfy  $\kappa_r' \geq \sum_{s \geq r} \kappa_s$ . This simple bound directly follows from the definitions.

- Actually, it is enough to assume the  $\theta$ -weak dependence inequality for any positive integer  $u$  and only for  $v = 1$ . Hence, for any 1-bounded function  $g$  from  $\mathbb{R}^u$  to  $\mathbb{R}$  and any 1-bounded Lipschitz function  $h$  from  $\mathbb{R}$  to  $\mathbb{R}$  with Lipschitz coefficient  $\text{Lip}(h)$ , it is enough to assume  $|\text{cov}(g(X_{i_1}, \dots, X_{i_u}), h(X_{i_u+i}))| \leq \theta_i \text{Lip}(h)$ , for any  $u$ -tuple  $i_1 \leq i_2 \leq \dots \leq i_u$ .

**4.2. Empirical process.** Let  $(X_t)_{t \in \mathbb{Z}}$  a real-valued stationary process. We use a quantile transform to obtain that the marginal distribution of this sequence is the uniform law on  $[0, 1]$ . The empirical process of the sequence  $(X_t)_{t \in \mathbb{Z}}$  at time  $n$  is defined as  $\frac{1}{\sqrt{n}} E_n(x)$  where

$$E_n(x) = \sum_{k=1}^n \left( \mathbb{1}_{(X_k \leq x)} - \mathbb{P}(X_k \leq x) \right).$$

Note that  $E_n = n(F_n - F)$  if  $F_n$  and  $F$  denote the empirical distribution function and the marginal distribution function, respectively. We consider the following convergence result in the Skohorod space  $D([0, 1])$  when the sample size  $n$  tends to infinity:

$$\frac{1}{\sqrt{n}} E_n(x) \xrightarrow{d} \bar{B}(x).$$

Here  $(\bar{B}(x))_{x \in [0, 1]}$  is the dependent analogue of a Brownian bridge, that is  $\bar{B}$  denotes a centered Gaussian process with covariance given by

$$\mathbb{E} \bar{B}(x) \bar{B}(y) = \sum_{k=-\infty}^{\infty} (\mathbb{P}(X_0 \leq x, X_k \leq y) - \mathbb{P}(X_0 \leq x) \mathbb{P}(X_k \leq y)). \quad (4.4)$$

Note that for independent sequences with a marginal distribution function  $F$ , this turns into  $\bar{B}(x) = B(x)$  for some standard Brownian bridge  $B$ ; this justifies the name of generalized Brownian bridge. We have:

**Theorem 4.2.** *Suppose that the stationary sequence  $(X_t)_{t \in \mathbb{Z}}$  has uniform marginal distribution and is either  $\eta$ -weakly dependent with  $\eta_r = O(r^{-15/2-\nu})$ , or  $\kappa$ -weakly dependent  $\kappa_r = O(r^{-5-\nu})$ , for some  $\nu > 0$ . Then the following empirical functional convergence holds true in the Skohorod space of real-valued càdlàg functions on the unit interval,  $D([0, 1])$ :*

$$\frac{1}{\sqrt{n}} E_n(x) \xrightarrow{d} \bar{B}(x).$$

*Remark 6.* Under strong mixing, the condition  $\sum_{n=0}^{\infty} \alpha_n < \infty$  implies fdi convergence. The empirical functional convergence holds if, in addition, for some  $a > 1$ ,  $\alpha_n = O(n^{-a})$  (see Rio (2000)). In an absolutely regular framework, Doukhan, Massart and Rio (1995) obtain the empirical functional convergence when, for some  $a > 2$ ,  $\beta_n = O(n^{-1}(\log n)^{-a})$ . Shao and Yu (1996) and Shao (1995) obtain the empirical functional convergence theorem when the maximal correlation coefficients satisfy the condition  $\sum_{n=0}^{\infty} \rho(2^n) < \infty$ .

To prove the result, we introduce the following dependence condition for a stationary sequence  $(X_t)_{t \in \mathbb{Z}}$ :

$$\sup_{f \in \mathcal{F}} |\text{cov}(f(X_{t_1})f(X_{t_2}), f(X_{t_3})f(X_{t_4}))| \leq \epsilon_r, \quad (4.5)$$

where  $\mathcal{F} = \{x \mapsto \mathbb{1}_{s < x \leq t}, \text{ for } s, t \in [0, 1]\}$ ,  $0 \leq t_1 \leq t_2 \leq t_3 \leq t_4$  and  $r = t_3 - t_2$  (in this case a weak dependence condition holds for a class of functions  $\mathbb{R}^u \rightarrow \mathbb{R}$  working only with the values  $u = 1$  or  $2$ ).

**Proposition 4.1.** *Let  $(X_n)$  be a stationary sequence such that (*

**4.3. Central limit theorems.** *First central limit theorems for weakly dependent sequences were given by Corollary A in Doukhan and Louhichi (1999) and Theorem 1 in Coulon-Prieur and Doukhan (2000). While the former result is for sequences of stationary random variables, the latter one is tailor-made for triangular arrays of asymptotically sparse random variables as they appear with kernel density estimators. Using their notion of  $\nu$ -mixing Bickel and Bühlmann (1999) proved a CLT for linear processes of infinite order and their (smoothed) bootstrap counterparts. Below we state a central limit theorem for general triangular schemes of weakly dependent random variables. Note that the applicability of a central limit theorem to bootstrap processes requires some robustness in the parameters of the underlying process since these parameters have to be estimated when it comes to the bootstrap. A result for a triangular scheme is therefore appropriate since the involved random variables have changing properties as  $n$  varies, but also for a fixed  $n$ . An interesting aspect of the following result is that no moment condition beyond Lindeberg's is required. Its proof uses the variant of Rio of the classical Lindeberg method.*

**Theorem 4.3.** *(Theorem 6.1 in Neumann and Paparoditis (2005))*

Suppose that  $(X_{n,k})_{k \in \mathbb{Z}}$ ,  $n \in \mathbb{N}$ , is a triangular scheme of (row-wise) stationary random variables with  $\mathbb{E}X_{n,k} = 0$  and  $\mathbb{E}X_{n,k}^2 \leq C < \infty$ . Furthermore, we assume that

$$\frac{1}{n} \sum_{k=1}^n \mathbb{E}X_{n,k}^2 I(|X_{n,k}|/\sqrt{n} > \epsilon) \xrightarrow{n \rightarrow \infty} 0 \quad (4.6)$$

holds for all  $\epsilon > 0$  and that

$$\sigma_n^2 = \sum_{k \in \mathbb{Z}} \mathbb{E}X_{n,0}X_{n,k} \xrightarrow{n \rightarrow \infty} \sigma^2 \in [0, \infty). \quad (4.7)$$

For  $n \geq n_0$ , there exists a monotonously nonincreasing and summable sequence  $(\theta_r)_{r \in \mathbb{N}}$  such that, for all indices  $s_1 < s_2 < \dots < s_u < s_u + r = t_1 \leq t_2$ , the following upper bounds for covariances hold true: for all measurable and quadratic integrable functions  $g : \mathbb{R}^u \rightarrow \mathbb{R}$ ,

$$|\text{cov}(g(X_{n,s_1}, \dots, X_{n,s_u}), X_{n,t_1})| \leq \sqrt{\mathbb{E}g^2(X_{n,s_1}, \dots, X_{n,s_u})} \theta_r, \quad (4.8)$$

for all measurable and bounded functions  $g : \mathbb{R}^u \rightarrow \mathbb{R}$ ,

$$|\text{cov}(g(X_{n,s_1}, \dots, X_{n,s_u}), X_{n,t_1}X_{n,t_2})| \leq \|g\|_\infty \theta_r, \quad (4.9)$$

where  $\|g\|_\infty = \sup_{x \in \mathbb{R}^u} |g(x)|$ . Then

$$\frac{1}{\sqrt{n}} (X_{n,1} + \dots + X_{n,n}) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

*Remark 7.* Conditions (

The following multivariate central limit theorem, simply applicable to triangular schemes of weakly dependent random vectors, was derived in Bardet, Doukhan, Lang and Ragache (2006). It fits better results where dependence does not play any role in the limit.

**Theorem 4.4.** (Theorem 1 in Bardet, Doukhan, Lang and Ragache (2006))

Suppose that  $(X_{n,k})_{k \in \mathbb{N}}$ ,  $n \in \mathbb{N}$ , is a triangular scheme of zero mean random vectors with values in  $\mathbb{R}^d$ . Assume that there exists a positive definite matrix  $\Sigma$  such that

$$\sum_{k=1}^n \text{Cov}(X_{n,k}) \xrightarrow{n \rightarrow \infty} \Sigma$$

and that for each  $\epsilon > 0$

$$\sum_{k=1}^n \mathbb{E}(\|X_{n,k}\|^2 \mathbb{I}_{\{\|X_{n,k}\| > \epsilon\}}) \xrightarrow{n \rightarrow \infty} 0,$$



where  $\|\cdot\|$  denotes the Euclidean norm. Furthermore, we assume the following condition is satisfied:

$$\sum_{k=2}^n \left| \text{cov}(e^{it'(X_{n,1}+\dots+X_{n,k-1})}, e^{it'X_{n,k}}) \right| \xrightarrow{n \rightarrow \infty} 0.$$

Then, as  $n \rightarrow \infty$ ,

$$\sum_{k=1}^n X_{n,k} \xrightarrow{d} \mathcal{N}_d(0_d, \Sigma).$$

**Remark.** One common point of those two results is the use of the classical Lindeberg assumption. Note that this assumption is more often checked by using a higher order moment condition. A main distinction between those two results is that while the first one yields direct application to partial sums, the second one is more adapted to triangular arrays where limit does not write as a sum. In this setting Doukhan and Wintenberger (2006) use Bernstein blocks to prove a CLT for partial sums.

**4.4. Probability and moment inequalities.** In this section we state inequalities of Bernstein and Rosenthal type. In the case of mixing, such inequalities can be easily derived by the well-known technique of replacing dependent blocks of random variables (separated an appropriate gap) by independent ones and then using the classical inequalities from the independent case; see for example Doukhan (1994) and Rio (2000). The concept of weak dependence is particularly suitable for deriving upper estimates for the cumulants of sums of random variables which give rise to rather sharp inequalities of Bernstein and Rosenthal type which are analogous to those in the independent case.

Based on a Rosenthal-type inequality, a first inequality of Bernstein-type was obtained by Doukhan and Louhichi (1999), however, with  $\sqrt{t}$  instead of  $t^2$  in the exponent. Dedecker and Prieur (2004) proved a Bennett inequality which can possibly be used to derive also a Bernstein inequality. A first Bernstein inequality with  $\text{var}(X_1 + \dots + X_n)$  in the asymptotically leading term of the denominator of the exponent has been derived in Kallabis and Neumann (2006), under a weak dependence condition tailor-made for causal processes with an exponential decay of the coefficients of weak dependence. The following result is a generalization which is also applicable to possibly

non-causal processes with a not necessarily exponential decay of the coefficients of weak dependence.

**Theorem 4.5.** (Theorem 1 in Doukhan and Neumann (2005))

Suppose that  $X_1, \dots, X_n$  are real-valued random variables defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  with  $\mathbb{E}X_i = 0$  and  $\mathbb{P}(|X_i| \leq M) = 1$ , for all  $i = 1, \dots, n$  and some  $M < \infty$ . Let  $\Psi : \mathbb{N}^2 \rightarrow \mathbb{N}$  be one of the following functions:

- (a)  $\Psi(u, v) = 2v$ ,
- (b)  $\Psi(u, v) = u + v$ ,
- (c)  $\Psi(u, v) = uv$ ,
- (d)  $\Psi(u, v) = \alpha(u + v) + (1 - \alpha)uv$ , for some  $\alpha \in (0, 1)$ .

We assume that there exist constants  $K, L_1, L_2 < \infty$ ,  $\mu \geq 0$ , and a non-increasing sequence of real coefficients  $(\rho(n))_{n \geq 0}$  such that, for all  $u$ -tuples  $(s_1, \dots, s_u)$  and all  $v$ -tuples  $(t_1, \dots, t_v)$  with  $1 \leq s_1 \leq \dots \leq s_u \leq t_1 \leq \dots \leq t_v \leq n$  the following inequality is fulfilled:

$$|\text{cov}(X_{s_1} \cdots X_{s_u}, X_{t_1} \cdots X_{t_v})| \leq K^2 M^{u+v-2} \Psi(u, v) \rho(t_1 - s_u), \quad (4.10)$$

where

$$\sum_{s=0}^{\infty} (s+1)^k \rho(s) \leq L_1 L_2^k (k!)^\mu \quad \forall k \geq 0. \quad (4.11)$$

Then

$$\mathbb{P}\left(\sum_{k=1}^n X_k \geq t\right) \leq \exp\left(-\frac{t^2/2}{A_n + B_n^{1/(\mu+2)} t^{(2\mu+3)/(\mu+2)}}\right), \quad (4.12)$$

where  $A_n$  can be chosen as any number greater than or equal to  $\sigma_n^2$  and

$$B_n = 2(K \vee M) L_2 \left( \left( \frac{2^{4+\mu} n K^2 L_1}{A_n} \right) \vee 1 \right).$$

A first Rosenthal-type inequality for weakly dependent random variables was derived by Doukhan and Louhichi (1999) via direct expansions of the moments of even order. Unfortunately, the variance of the sum did not explicitly show up in their bound. Using cumulant bounds in conjunction with Leonov and Shiryayev's formula the following tighter moment inequality was obtained in Doukhan and Neumann (2005).

**Theorem 4.6.** (Theorem 3 in Doukhan and Neumann (2005))

Suppose that  $X_1, \dots, X_n$  are real-valued random variables on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  with zero mean and let  $p$  be a positive integer. We assume that there exist constants  $K, M < \infty$ , and a non-increasing sequence of real coefficients  $(\rho(n))_{n \geq 0}$  such that, for all  $u$ -tuples  $(s_1, \dots, s_u)$  and all  $v$ -tuples  $(t_1, \dots, t_v)$  with  $1 \leq s_1 \leq \dots \leq s_u \leq t_1 \leq \dots \leq t_v \leq n$  and  $u + v \leq p$ , condition (

## 5. PROOFS

*Proof of Lemma* Let, for simplicity of notation,  $s_k = 0$ . Then

$$\begin{aligned} & \text{cov}(g(X_{s_1}, \dots, X_{s_k}), h(X_{t_1}, \dots, X_{t_l})) \\ &= \mathbb{E}[g(X_{s_1}, \dots, X_{s_k}) (\mathbb{E}(h(X_{t_1}, \dots, X_{t_l}) \mid X_{s_1}, \dots, X_{s_k}) - \mathbb{E}h(X_{t_1}, \dots, X_{t_l}))]. \end{aligned}$$

Now we obtain by Jensen's inequality for conditional expectations that

$$\begin{aligned} & |\text{cov}(g(X_{s_1}, \dots, X_{s_k}), h(X_{t_1}, \dots, X_{t_l}))| \\ & \leq \mathbb{E}[|g(X_{s_1}, \dots, X_{s_k})| \cdot |\mathbb{E}(h(X_{t_1}, \dots, X_{t_l}) \mid \mathbb{X}_0) - \mathbb{E}h(X_{t_1}, \dots, X_{t_l})|] \\ & \leq \mathbb{E}\left[|g(X_{s_1}, \dots, X_{s_k})| \cdot \left|\mathbb{E}(h(X_{t_1}, \dots, X_{t_l}) \mid \mathbb{X}_0) - \mathbb{E}(h(X'_{t_1}, \dots, X'_{t_l}) \mid \mathbb{X}'_0)\right|\right] \\ & \leq \text{Lip } h \mathbb{E}\left[|g(X_{s_1}, \dots, X_{s_k})| \cdot \mathbb{E}\left(\sum_{j=1}^l |X_{t_j} - X'_{t_j}| \mid \mathbb{X}_0, \mathbb{X}'_0\right)\right]. \end{aligned}$$

The assertion follows now immediately.  $\square$

*Proof of Lemma* The assertion follows immediately from (

*Proof of Lemma*(a) Let  $\widehat{\xi}_{n,1}, \dots, \widehat{\xi}_{n,p}$  be the roots of the characteristic polynomial  $\widehat{\theta}(z) = 1 - \widehat{\theta}_{n,1}z - \dots - \widehat{\theta}_{n,p}z^p$  of the bootstrap process. Since  $\widehat{\theta} \xrightarrow{P} \theta$  we obtain by Theorem 1.4 in Marden (1949) that

$$\min\{|\widehat{\xi}_{n,1}|, \dots, |\widehat{\xi}_{n,p}|\} \xrightarrow{P} \rho = \min\{|\xi_1|, \dots, |\xi_p|\}.$$

Therefore, we have, for any  $\epsilon > 0$ , that

$$\mathbb{P}\left(\min\{|\widehat{\xi}_{n,1}|, \dots, |\widehat{\xi}_{n,p}|\} \geq \rho/(1 + \epsilon/2)\right) \xrightarrow{n \rightarrow \infty} 1. \quad (5.1)$$

Thus there exists a stationary solution to equation (

*Proof of inequality* ( We prove this result by a simple coupling argument. Let  $(X_t)_{t \in \mathbb{Z}}$  and  $(X'_t)_{t \in \mathbb{Z}}$  be two versions of the autoregressive process with  $X_t = x$  and  $X'_t = y$ . We construct a coupling simply by feeding both processes after time  $t$  with the same sequence of innovations  $\varepsilon_{t+1}, \varepsilon_{t+2}, \dots$ , that is, we

have  $X_{t+l+1} = m(X_{t+l}) + \varepsilon_{t+l+1}$  and  $X'_{t+l+1} = m(X'_{t+l}) + \varepsilon_{t+l+1}$  ( $l \geq 0$ ). We obtain by this construction that

$$\begin{aligned} |X_{t+k} - X'_{t+k}| &\leq \Delta(X_{t+k-1})|X_{t+k-1} - X'_{t+k-1}| \\ &\leq \cdots \leq \Delta(X_{t+k-1}) \cdots \Delta(X_{t+1})\Delta(x)|x - y|. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} d\left(\mathbb{P}^{X_{t+k}|X_t=x}, \mathbb{P}^{X_{t+k}|X_t=y}\right) &\leq \mathbb{E}|X_{t+k} - X'_{t+k}| \\ &\leq \mathbb{E}(\Delta(X_{t+k-1}) \cdots \Delta(X_{t+1}) | X_t = x) \cdot \Delta(x) \cdot |x - y|. \end{aligned}$$

Using the Markov property we see that

$$\begin{aligned} &\mathbb{E}(\Delta(X_{t+k-1}) \cdots \Delta(X_{t+1}) | X_t = x) \\ &= \mathbb{E}[\mathbb{E}(\Delta(X_{t+k-1}) | X_{t+k-2}) \cdots \mathbb{E}(\Delta(X_{t+1}) | X_t = x)] \\ &\leq \rho^{k-1}, \end{aligned}$$

which yields the assertion.  $\square$

## REFERENCES

- ANDREWS, D. (1984). Non strong mixing autoregressive processes. *J. Appl. Probab.* **21**, 930–934.
- BARDET, J. M., DOUKHAN, P., LANG G. AND RAGACHE, N. (2006). A Lindeberg central limit theorem for dependent processes and its statistical applications. Manuscript.
- BICKEL, P. J. AND BÜHLMANN, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* **5**, 413–446.
- BICKEL, P. J. AND FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196–1217.
- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. New-York: Wiley.
- BIRKHOFF, G. D. (1931). Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. USA* **17**, 656–660.
- BROCKWELL, P. J. AND DAVIS, R. A. (1991). *Time Series: Theory and Methods, 2nd edition*. New York: Springer.
- BULINSKI, A. AND SHASHKIN, A. (2005). Strong invariance principle for dependent random fields, *Preprint*.
- DEDECKER, J. AND DOUKHAN, P. (2003). A new covariance inequality and applications. *Stoch. Proc. Appl.* **106**, **1**, 63–80.
- COULON-PRIEUR, C. AND DOUKHAN, P. (2000). A triangular central limit theorem under a new weak dependence condition. *Statist. Probab. Lett.* **47**, 61–68.
- DEDECKER, J. AND PRIEUR, C. (2004). Coupling for  $\tau$ -dependent sequences and applications. *J. Theor. Probab.* **17**, 861–885.

- DEDECKER, J. AND RIO, E. (2000). On the functional central limit theorem for stationary processes. *Annales de l'institut Henri Poincaré série B* **36**, 1–34.
- DOBRUSHIN, R. L. (1970). Describing a system of random variables by conditional distributions. *Teor. Veroyatn. Primen.* **15**, 469–497 (in Russian). [English translation: *Theory Probab. Appl.* **15**, 458–486.]
- DOUKHAN, P. (1994). *Mixing: Properties and Examples*. Lecture Notes in Statistics 85, Springer Verlag.
- DOUKHAN, P. (2002). Limit theorems for stationary sequences. in *Long Range Dependence, theory and Applications*, editors P. Doukhan, G. Oppenheim and M. Taqqu, Birkhäuser.
- DOUKHAN, P. AND LOUHICHI, S. (1999). A new weak dependence condition and application to moment inequalities. *Stoch. Proc. Appl.* **84**, 313–342.
- DOUKHAN, P., MADRE, H. AND ROSENBAUM, M. (2005). ARCH type bilinear weakly dependent models. Manuscript.
- DOUKHAN, P., MASSART, P. AND RIO, E. (1995). Invariance principle for the empirical measure of a weakly dependent process. *Ann. I.H.P.* **31-2**, 393–427.
- DOUKHAN, P. AND NEUMANN, M. H. (2005). Probability and moment inequalities for sums of weakly dependent random variables, with applications. Manuscript.
- DOUKHAN, P., TEYSSIERE, G. AND WINANT, M. H. (2005). Vector valued ARCH infinity processes, in *Statistics for times series*, Lecture Notes in Statistics, Bertail, Doukhan, & Soulier editors, to appear.
- DOUKHAN, P. AND WINTENBERGER, O. (2006). An invariance principle for new weakly dependent stationary models using sharp moment assumptions. Manuscript.
- GIRAITIS, L. AND SURGAILIS, D. (2002). ARCH-type bilinear models with double long memory, *Stoch. Proc. Appl.* **100**, 275–300.
- GORODETSKII, V. V. (1977). On the strong mixing property for linear sequences. *Teor. Veroyatn. Primen.* **22** 421–423, (in Russian). [English translation: *Theory Probab. Appl.* **22**, 411–413.]
- KALLABIS, R. S. AND NEUMANN, M. H. (2006). An exponential inequality under weak dependence. *Bernoulli* **12**, 333–350.
- MALLOWS, C. L. (1972). A note on asymptotic joint normality. *Ann. Math. Statist.* **43**, 508–515.
- MARDEN, M. (1949). *The Geometry of the Zeros of a Polynomial in a Complex Variable*. Mathematical Surveys Number III, Amer. Math. Soc., New York.
- NEUMANN, M. H. AND PAPARODITIS, E. (2005). Goodness-of-fit tests for Markovian time series models. Technical Report No. 16/2005, Department of Mathematics and Statistics, University of Cyprus.
- PITT, L. (1982). Positively correlated normal variables are associated. *Ann. Probab.* **10**, 496–499.
- RIO, E. (1995). About the Lindeberg method for strongly mixing sequences. *ESAIM, Probab. Statist.* **1**, 35–61.

- RIO, E. (2000). *Théorie asymptotique pour des processus aléatoires faiblement dépendants*. SMAI, Mathématiques et Applications 31, Springer.
- ROSENBLATT, M. (1980). Linear processes and bispectra. *J. Appl. Probab.* **17**, 265–270.
- SHAO, Q. M. (1995). Maximal inequalities for partial sums of  $\rho$ -mixing sequences. *Annals of Probability* **23**, 948–965.
- SHAO, Q. M. AND YU, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability* **24**, 2052–2078.
- STOUT, W. (1974). *Almost sure convergence*. Academic Press, New York.