

WEAK DEPENDENCE: MODELS AND APPLICATIONS TO ECONOMETRICS

PATRICK ANGO NZE
Université Lille 3

PAUL DOUKHAN
Université de Cergy Pontoise

In this paper we discuss weak dependence and mixing properties of some popular models. We also develop some of their econometric applications. Autoregressive models, autoregressive conditional heteroskedasticity (ARCH) models, and bilinear models are widely used in econometrics. More generally, stationary Markov modeling is often used. Bernoulli shifts also generate many useful stationary sequences, such as autoregressive moving average (ARMA) or ARCH(∞) processes. For Volterra processes, mixing properties obtain given additional regularity assumptions on the distribution of the innovations.

We recall associated probability limit theorems and investigate the nonparametric estimation of those sequences.

1. INTRODUCTION

Mixing is now systematically used in time series where martingale assumptions and results cannot be directly employed. Mixing has proved particularly useful in cases where nonlinearities appear, such as autoregressive conditional heteroskedasticity (ARCH) modeling in econometrics. This success relies on powerful limit theorems proved under mixing conditions (see, among others, Doukhan, 1994; Rio, 2000; Doukhan, 2002). These limit results serve as basic tools for computation of the significance level and power of statistical tests. Mixing assumptions can be used in more general frameworks involving fading memory (asymptotic independence between functions of the past and the future of the process), such as near epoch dependence (NED) of a mixing process.

We first thank the editor for the huge amount of additional editorial work provided for this review paper. The efficiency of the numerous referees was especially useful. The error pointed out in Hall and Horowitz (1996) was the origin of the present paper, and we thank the referees for asking for a more detailed treatment of a correct proof for this paper in Section 2.3. Also we thank Marc Henry and Rafal Wojakowski for a very careful rereading of the paper. An anonymous referee has been particularly helpful in the process of revision of the paper. The authors thank him for his numerous suggestions of improvement, including important results on negatively associated sequences and a thorough update in standard English. Address correspondence to Patrick Ango Nze, Université Lille 3, UFR MSES, BP 149, 59653, Villeneuve d'Ascq Cedex, France; e-mail: angonze@univ-lille3.fr. Paul Doukhan, Université de Cergy Pontoise, J. E. 2070 CNRS-Mathématiques, Bât. Les Chênes, 33, Bd. du Port, 95011 Cergy-Pontoise Cedex, France, and CREST-LS, timbre J340, 3, Av. Pierre Larousse, 92245 Malakoff, France; e-mail: doukhan@ensae.fr.

We recall here the mixing properties of some models used in econometrics. Simultaneously, we present a different approach to limit theorems when mixing does not hold (which really may occur, as shown in Andrews, 1984, example (4.16)). For the sake of simplicity, our exposition mainly focuses on one-dimensional time series.

The paper is organized as follows. To provide deep econometric motivations, Section 2 exposes several situations where the various weak dependence conditions arise, and after some generic examples, we consider specific problems, including unit root problems, parametric problems, sieve bootstrap, and semiparametric estimation problems in Sections 2.1, 2.2, 2.4, and 2.5. Section 2.3 considers generalized method of moments (GMM) estimation in which the Doukhan and Louhichi (1999) weak dependence condition allows one to provide a complete proof of the results in Hall and Horowitz (1996). Indeed, the latter authors improperly claim a mixing property of their models to prove their consistency results. Finally, Section 2.6 considers nonparametric estimation problems.

Section 3 makes precise the mathematical framework of weak dependence needed in the previous section. It describes some classical concepts of fading memory (mixing conditions, the association property) and also the new weak dependence conditions introduced by Doukhan and Louhichi (1999). After this, Section 4 provides numerous classes of models commonly used in econometrics and finance and focuses on their weak dependence properties. Section 5 recalls some probabilistic limit theorems available in those cases. Extensions of Donsker's functional central limit theorem (FCLT) and the FCLT for the cumulative distribution function are discussed. Section 6 is devoted to functional estimation. Consistency and CLTs are discussed here. Proofs are given in Appendix A, and Appendix B recalls the main probabilistic tools.

Finally, we remark that the limit theorems of Section 3 and the asymptotic results for functional estimation in Section 6 are provided for very large classes of models (Sections 3 and 4). Hence, more general time series formulations such as those in Section 4 allow us to extend directly the classical results of Section 2.

2. ECONOMETRICS AND DEPENDENCE

Time series analysis is a major part of econometrics. Here we provide several examples of interest in which it is essential to consider dependent structures instead of simple independence. In some situations, classical tools of weak dependence such as mixing are useless. For instance, when bootstrap techniques are used, no mixing conditions can be expected. Consider the following example concerning bootstrap: let a stationary autoregressive sequence be generated by an independent and identically distributed (i.i.d.) sequence $(\xi_n)_{n \in \mathbb{Z}}$:

$$X_n = r(X_{n-1}) + \xi_n. \quad (2.1)$$

Standard nonparametric estimation techniques provide an estimate of the autoregression function r . Let \hat{r} be a convenient estimator of r . Given data (X_1, \dots, X_n) from the sequence (2.1), another autoregressive process is defined by

$$\widehat{X}_n = \hat{r}(\widehat{X}_{n-1}) + \widehat{\xi}_n. \quad (2.2)$$

The innovations $(\widehat{\xi}_n)$ are drawn according to the empirical measure of the estimated residuals, $\widehat{\xi}_i = X_i - \hat{r}(X_{i-1}), 1 \leq i \leq n$. No mixing assumption can be expected for the previous model (2.2); see (4.16) in Section 4. However, a new concept of fading memory can still be applied. Bickel and Bühlmann (1999) set up such a new weak dependence condition to build critical bootstrap values for a linearity test in linear models. Doukhan and Louhichi (1999) have extended it to fit models such as positively dependent sequences, Markov chains (with or without topological assumptions), and Bernoulli shifts (see Definition 4.1). The Bernoulli shifts are defined in Assumption 1 of Hall and Horowitz (1996) and are used throughout that paper. The previously mentioned weak dependence conditions yield standard results concerning convergence in distribution with an \sqrt{n} -normalization.

Another application of these results concerns linearity tests in time series analysis. Rios (1996) considers a stationary functional autoregressive model (2.1) where $r = L + C$ is the decomposition of the autoregression function into a sum of linear (L) and nonlinear (C) components. Local linearity of r is then tested via the null hypothesis

$$\mathbf{H}_0: \int (r''(x))^2 w(x) dx = 0,$$

where the weight function w has compact support.

Still another problem of interest is to test the independence of the innovations $(\xi_n)_{n \in \mathbb{Z}}$ in a regression model

$$X_n = aY_n + \xi_n.$$

This can be performed using the Durbin–Watson statistic. The latter can be written as a continuous functional of the Donsker line of the sequence $(\xi_n)_{n \in \mathbb{Z}}$.

2.1. Unit Root Tests

Consider a stationary autoregressive sequence $(X_n)_{n \in \mathbb{Z}}$ generated by an i.i.d. sequence $(\xi_n)_{n \in \mathbb{Z}}$,

$$X_n = aX_{n-1} + \xi_n.$$

A classical problem is to test whether there is a unit root (i.e., $a = 1$). In the specific context of aggregate time series, the assumption of white noise innovations seems to be rather strong. Phillips (1987) develops unit root tests for

mixing and heterogeneously distributed innovations. The ordinary least squares estimate \hat{a} is shown to be a continuous functional of the Donsker line of the sequence $(\xi_n)_{n \in \mathbb{Z}}$. As an application of the FCLT, Phillips shows that a unit root test can be based on the fact that under the null hypothesis $\mathbf{H}_0: a = 1$,

$$n(\hat{a} - 1) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \frac{1}{2 \int_0^1 W_t^2 dt} \left(W_1^2 - \frac{\sigma_{\xi}^2}{\sigma^2} \right),$$

where W denotes the standard Brownian motion and $\sigma^2 = \sum_{-\infty}^{\infty} \text{Cov}(\xi_0, \xi_k)$. The author works with stationary strong mixing sequences, and conditions under which the FCLT result holds true are reported in Section 5.1. This result can be obtained under a weak dependent context detailed in Section 4. The conditions for which Donsker’s theorems hold are described in Section 5.1. This example, as the author suggests, can be generalized to error sequences $(\xi_n)_{n \in \mathbb{Z}}$ that allow for heteroskedasticity. See also Mills (1999) for a discussion of the Dickey–Fuller unit root test in autoregressive models when errors fluctuate about a non-zero mean.

2.2. Parametric Problems

GMM estimation procedures involve an estimate $\hat{\theta}_n$, which is a solution of the arg-min problem $J_n(\hat{\theta}_n) = \min_{\theta \in \Theta} J_n(\theta)$, where

$$J_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^n g(X_i, \theta) \right)' \Omega \left(\frac{1}{n} \sum_{i=1}^n g(X_i, \theta) \right). \tag{2.3}$$

Here $\Theta \subset \mathbb{R}^d$ is a finite-dimensional parameter set, and $g(\cdot, \cdot)$ is a given function such that $\mathbb{E}_{\theta_0} g(X_1, \theta_0) = 0$, where θ_0 is the true parameter point. In the time series context, the positive semidefinite matrix Ω is often replaced (see Hall and Horowitz, 1996, equation (3.2)) by an asymptotically optimal weight matrix estimate

$$\Omega_n(\theta) = \frac{1}{n} \sum_{i=1}^n g(X_i, \theta)g(X_i, \theta)' + \sum_{j=1}^{\kappa} H(X_i, X_{i+j}, \theta),$$

$$H(x, y, \theta) = g(x, \theta)g(y, \theta)' + g(y, \theta)g(x, \theta)',$$

and κ is such that $\mathbb{E}g(X_i, \theta)g(X_j, \theta)' = 0$ if $|i - j| > \kappa$. The statistic to test $\mathbf{H}_0: \theta = \theta_0$ is $J_n(\theta) = K_n(\hat{\theta}_n)'K_n(\hat{\theta}_n)$, where $K_n(\theta) = (1/\sqrt{n})\Omega_n(\theta)^{1/2} \times \sum_{i=1}^n g(X_i, \theta)$ (the square root of a symmetric positive matrix is uniquely defined).

2.2.1. Block Bootstrap. A bootstrap procedure allows one to estimate the limit distribution of an estimate.

We describe a block-bootstrap procedure that is adapted to the times series $(X_i)_{i \in \mathbb{N}}$. Let $b = b(n)$ and $l = l(n)$ denote the number and the length of the blocks. Then $bl = n$ and the block j is $(\tilde{X}_{(j-1)l+1}, \dots, \tilde{X}_{jl})$ (see Künsch, 1989). In this construction, a suitable form of \tilde{X}_k is $\hat{H}(\hat{\xi}_k, \hat{\xi}_{k-1}, \dots, \hat{\xi}_{k-l+1}, 0, \dots)$ if the process $X_k = H(\xi_k, \xi_{k-1}, \dots)$ is a Bernoulli shift as defined in Section 4.3. Here \hat{H} and $\hat{\xi}_k$ are obtained through filtering and estimation procedures in the simple case of a linear process $(H(z_0, z_1, \dots) = \sum_k a_k z_k$; see Section 2.4); in the general setting, one needs to develop additional estimation procedures. To describe the asymptotic properties of such processes one needs to know the limiting asymptotic behavior of Bernoulli shifts.

2.2.2. Conditional Bootstrap. A simple local conditional bootstrap is investigated by Ango Nze, Bühlmann, and Doukhan (2002). In that paper, it is shown that asymptotic properties can be obtained using the same weak dependence techniques. The following central limit theorem (CLT) holds under standard mixing assumptions:

$$T_n(\theta) = \sqrt{n} \Sigma_n^{-1} (\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_d(0, I_d), \quad (2.4)$$

where the diagonal matrix Σ_n has d entries. GMM techniques naturally involve an unknown covariance matrix. To estimate such limiting distributions it is natural to use bootstrap techniques.

2.3. Bootstrapping Critical Values for GMM Estimators

Let $(\tilde{X}_i^*)_{1 \leq i \leq n}$ denote a block-bootstrap sample and let $g^*(x, \theta) = g(x, \theta) - \mathbb{E}^* g(x, \hat{\theta}_n)$. The expectation is taken with respect to the bootstrap distribution. The GMM estimate $\hat{\theta}_n^*$ solves the arg-min problem

$$J_n^*(\theta) = \left(\frac{1}{n} \sum_{i=1}^n g^*(X_i^*, \theta) \right)' \Omega \left(\frac{1}{n} \sum_{i=1}^n g^*(X_i^*, \theta) \right) \quad (2.5)$$

if the matrix Ω is known.

Hall and Horowitz (1996) make the erroneous statement that such Bernoulli shifts are strong mixing. However, the procedure used by Hall and Horowitz makes the bootstrap work. The weak dependence condition as defined in Doukhan and Louhichi (1999) allows us to rigorously prove the consistency of the Hall and Horowitz procedures. More precisely, if $X_n = h(\epsilon_n, \epsilon_{n-1}, \dots)$ for some i.i.d. sequence $(\epsilon_i)_{i \in \mathbb{Z}}$, their Assumption 1 is¹

$$\mathbb{E} \|h(\epsilon_n, \epsilon_{n-1}, \dots) - h(\epsilon_n, \epsilon_{n-1}, \dots, \epsilon_{n-m}, 0, 0, \dots)\| \leq \frac{e^{-dm}}{d}.$$

This condition holds for linear processes, and it is claimed to imply geometric strong mixing by the authors. Simple example (4.16) proves that this does not hold in general. This condition, however, does imply weak dependence in

Doukhan and Louhichi (1999). Consequently, a tail inequality for sums of functions of the sequence $\xi_n = f(X_n, \theta)$ can be derived. It is the main tool to prove the validity of the bootstrap in this dependent setting.

The preceding procedure can be used for testing the null hypothesis $\mathbf{H}_0: \theta = \theta_0$ against the bilateral alternative. The studentized statistic $T_n(\theta)$ described in (2.4) and the critical value Q_α then satisfy the relation that, under \mathbf{H}_0 ,

$$\mathbb{P}(|T_n(\theta)| > Q_\alpha) = \alpha + \mathcal{O}(1/n).$$

Hall and Horowitz show that the bootstrap studentized statistic $T_n^*(\theta) = \sqrt{n} \Sigma_n^{*-1}(\hat{\theta}_n^* - \hat{\theta}_n)$ and the previously mentioned statistic $T_n(\theta)$ have close laws, in the sense that

$$\mathbb{P}\left(\sup_{z \in \mathbb{R}} |\mathbb{P}^*(T_n^*(\theta) \leq z) - \mathbb{P}(T_n(\theta) \leq z)| > n^{-a}\right) = o(n^{-a}), \tag{2.6}$$

for a relevant integer $2a$, with $a \geq 1 + \xi$, and the range of $\xi \in [0, 1]$ is formulated according to the dependence assumptions prescribed. This relation comes from an Edgeworth development. It yields an improved acceptance rule for the test of \mathbf{H}_0 :

$$\mathbb{P}(|T_n^*(\theta)| > Q_\alpha^*) = \alpha + \mathcal{O}(n^{-1-\xi}).$$

Andrews (2002) points out that a direct computation of the bootstrap critical value Q_α^* is a hard problem and that the common estimating procedure, which is based upon B bootstrapped, independent copies (from the law of large numbers, it follows that $Q_\alpha^*(B) \xrightarrow[B \rightarrow \infty]{P} Q_\alpha^*$) is also difficult to implement. Indeed, the computations involve the minimization of B nonlinear functionals. A numerical improvement is brought to bear in Andrews’s paper. A bootstrap estimator $\hat{\theta}_{n,k}^*$ is computed by applying the Newton–Raphson algorithm. The initial value is $\hat{\theta}_{n,0}^* = \hat{\theta}_n$, and k iterations are made ($k \geq 3$). A bootstrap studentized statistic $T_{n,k}^*$ is now available, for which the computation of the critical bootstrap values $Q_\alpha^*(B)$ is much easier, because the problem is linear. The method is claimed to be as accurate as the one discussed by Hall and Horowitz. In fact, the author states a similar result to (2.6) with respect to $\mathbb{P}^*(T_{n,k}^*(\theta) \leq z)$ and $\mathbb{P}^*(T_n^*(\theta) \leq z)$. The assumptions are those of Hall and Horowitz. Therefore, the previously mentioned Assumption 1 must also be read in the context of the comments we have already made about Hall and Horowitz’s paper. For the sake of completeness we present a corrected version of Lemma 1 in Hall and Horowitz (1996), which is proved in Section B.4 of Appendix B.

LEMMA 2.1. *Let (ξ_n) be a stationary ψ_1 -weakly dependent (see Definition 3.4) sequence with $\mathbb{E}\xi_n = 0$ such that $\theta_r = \mathcal{O}(e^{-ar})$ for some $a > 0$, as $r \uparrow \infty$, and $\mathbb{P}(|\xi_1| \geq z) = \mathcal{O}(|z|^{-33})$, as $|z| \rightarrow \infty$. Then² $R_n = n^{-1} \sum_{i=1}^n \xi_i$ satisfies*

$$\lim_{n \rightarrow \infty} n\mathbb{P}(|R_n| > n^{-[(2+\epsilon)/5]}) = 0.$$

Following precisely the same steps as in Hall and Horowitz (1996), we thus prove, by only replacing their Lemma 1 by Lemma 2.1, that bootstrapping critical values for GMM estimators are valid.

Now Theorems 1–3 in Hall and Horowitz (1996) are rigorously proved. A paper on this topic by the authors is in preparation to provide sharper results; for instance, the exponent 33 in the previous lemma is unnatural, and it can be changed.

2.4. Sieve Bootstrap

Bickel and Bühlmann (1999) tackle the problem of the “sieve bootstrap” for a one-sided linear process

$$X_n - \mu = \xi_0 + \sum_{t=1}^{\infty} a_t \xi_{n-t}, \quad (2.7)$$

where (ξ_n) is a sequence of i.i.d. random variables (r.v.s) with $\mathbb{E}\xi_0 = 0$ and the density function f_{ξ} , and where $\sum_{t=1}^{\infty} |a_t| < \infty$ and $\mu = \mathbb{E}X_n$. Under the assumption that the function $\Psi(z) = 1 + \sum_{t=1}^{\infty} a_t z^t$ has no root in the closed unit circle, the process (2.7) admits an AR(∞) representation

$$(X_n - \mu) + \sum_{t=1}^{\infty} b_t (X_{n-t} - \mu) = \xi_n \quad (2.8)$$

with $\sum_{t=1}^{\infty} |b_t| < \infty$. Equation (2.8) is fitted with an autoregressive process of finite order $p(n)$ ($p(n)/n \rightarrow 0$, $p(n) \rightarrow \infty$). Using estimated residuals, the resampling (i.i.d.) innovation process $(\xi_n^*)_{n \in \mathbb{Z}}$ is constructed by smoothing the empirical process based on those residuals by a kernel density estimate of the density f_{ξ} . Finally, the smoothed sieve bootstrap sample $(X_n^*)_{n \in \mathbb{Z}}$ is defined by resampling the AR($p(n)$) process from innovations $(\xi_n^*)_{n \in \mathbb{Z}}$:

$$(X_n^* - \bar{X}) + \sum_{t=1}^{p(n)} b_t (X_{n-t}^* - \bar{X}) = \xi_n^*. \quad (2.9)$$

The purpose of the paper was to carry over a weak dependence property (here strong mixing) of the initial sequence $(X_n)_{n \in \mathbb{Z}}$ to the sieve processes $(X_n^*)_{n \in \mathbb{Z}}$ (a classic and a smoothed version were examined in the paper). The goal is unrealistic for the classic bootstrap sample because the distribution of the bootstrapped innovations is discrete. Proving a mixing property for the smoothed sieve bootstrap sample eludes the efforts of the authors. In the latter case, it nevertheless appears that limit theorems can be proved by another method. It consists of using the following property:

$$|\text{Cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))| \leq 4 \|g_1\|_{\infty} \|g_2\|_{\infty} \nu(k; \mathcal{C}^{d_1}, \mathcal{C}^{d_2}), \quad (2.10)$$

with $d_1, d_2 \in \mathbb{N}$ and for smooth functions g_1, g_2 belonging to the classes \mathcal{C}^{d_1} and \mathcal{C}^{d_2} (see equation (3.1) in Bickel and Bühlmann for the definition of the class \mathcal{C}^d and some examples). The new dependence coefficient ν is less than the strongly mixing coefficient. Bickel and Bühlmann (1999) cannot prove that the sieve sequence (X_n^*) is strongly mixing. A weak dependence condition is now defined by the ν coefficient. Bickel and Bühlmann prove that it is satisfied by both this sequence and a smooth version of the resampled innovations. For instance, Bickel and Bühlmann prove that if the sequence $(X_n)_{n \in \mathbb{Z}}$ satisfies some regularity conditions ensuring that $\alpha_k = \mathcal{O}(k^{-\gamma})$ (recall that $\nu_k \leq \alpha_k$), then the sieve bootstrap process $(X_n^*)_{n \in \mathbb{Z}}$ satisfies a ν -mixing condition with a polynomial rate $\nu^*(k; \mathcal{C}^{d_1}, \mathcal{D}^{d_2}) = \mathcal{O}(k^{-L\gamma})$ for relevant classes $\mathcal{C}^{d_1}, \mathcal{D}^{d_2}$ and a positive constant L . See Theorem 3.2 in Bickel and Bühlmann (1999, p. 422) for more details.

2.5. A Semiparametric Estimation Problem

We follow the presentation in Robinson (1989). He considers an economic variable observable at time n that is an $R \times 1$ vector of r.v.s $(W_n)_{n \in \mathbb{Z}}$. We observe W_n at time $n = 1 - P, 2 - P, \dots, T$ where P is nonnegative and T large. Hypotheses of economic interest often involve a subset $X_n = B(W'_n, \dots, W'_{n-P})$ of the array $(W'_n, \dots, W'_{n-P})'$; for this B is a $J \times (PR)$ matrix formed from the PR -rowed identity matrix I_{PR} by omitting $PR - J$ rows (which means that in B , $PR - J$ elements of W_n, \dots, W_{n-P} are deleted). Thus, in B , elements of $W_n, W_{n-1}, \dots, W_{n-P}$ that are not in X_n are deleted, and X_n can have elements in common with $X_{n+P-1}, \dots, X_{n+1}, X_{n-1}, \dots, X_{n-P}$. Let $X_n = (Y'_n, Z'_n)'$, where Y_n and Z_n are $K \times 1$ and $L \times 1$ vectors ($K + L = J$). The problem of interest is to test the hypothesis $\mathbb{E}(Y_n|Z_n) = 0$ against the alternative $\mathbb{E}(Y_n|Z_n) \neq 0$. This null hypothesis is written in the form $\tau = \int_{\mathbb{R}^L} H(z, z) f^2(z) dz = 0$ for $M = 0$ and $\tau = \int_{\mathbb{R}^L} H(z, z) (f(z), f^{(1)}(z)', \dots, f^{(M)}(z)')' f(z) dz = 0$ for some $M > 0$ and some function $H(z, z)$ defined as $H(z_1, z_2) = \int_{\mathbb{R}^K \times \mathbb{R}^K} G(x_1, x_2) dF(y_1|z_1) dF(y_2|z_2)$ for some convenient function G and where $F(A|z) = \mathbb{P}(Y_n \in A|Z_n = z)$ for any Borel set A of \mathbb{R}^K and $z \in \mathbb{R}^L$ and $x_1 = (y'_1, z'_1)'$ and $x_2 = (y'_2, z'_2)'$. Here $f^{(j)}(z)$ denotes the vector of j -partial derivatives of f .

An example of this framework is given by $X_n = (Y'_n, Z'_n)'$, where $Y_n = (t_n, s'_n)'$ and $Z_n = v_n$. The regression model

$$t_n = \beta'(s_n - \mathbb{E}_n s_n) + \gamma' \mathbb{E}_n s_n + u_n \tag{2.11}$$

is of common use in econometrics. Here s_n, t_n, v_n are, respectively, scalars, $p \times 1$, and $q \times 1$; they are observable random sequences whereas the innovation process (u_n) is centered and unobservable, so that $\mathbb{E}(u_n|s_n, t_n) = 0$; we denote $\mathbb{E}_n(\cdot) = \mathbb{E}(\cdot|v_n)$. In the case of a weakly dependent and stationary innovation process, Robinson (1989) considers the hypothesis $\mathbf{H}_0: \beta = 0$.

In this case, the hypothesis can be written as before, and Robinson calculates $\beta = \tau$ where $K = p + 1$, $L = 1$, $M = 0$, and $G(x_1, x_2) = (t_1 - t_2)s_1\phi(v_1)$ for some function $\phi: \mathbb{R}^q \rightarrow \mathbb{R}$ (usually $\phi \equiv 1$). Robinson considers the statistics $\hat{\lambda} = n\hat{\tau}'\hat{\Omega}^{-1}\hat{\tau}$ constructed from the n -sample (X_1, \dots, X_n) . Here, $\hat{\tau} = (1/n^2h^L)\sum_{i,j=1}^n G(X_i, X_j)k(Z_i - Z_j/h)$ is a U -statistic and $\hat{\Omega}$ is the natural estimator of the covariance matrix of $\hat{\tau}$. One such estimate is $\hat{\Omega} = (1/n)\sum_{i=1}^n c_i c_i'$. Tapered versions might be preferred (see Robinson, 1989, formula (2.21)); here $c_i = \sum_{j=1}^n d_{i,j} + d_{i,i}$ with $d_{i,j} = G(X_i, X_j)\bar{k}(Z_i - Z_j)/h$, where $\bar{k}(z) = h^{-L}(k(z), h^{-1}k^{(1)}(z), \dots, h^{-M}k^{(M)}(z)')$. Under β -mixing assumptions, Robinson proves that these estimates are \sqrt{n} -consistent and satisfy a CLT. Under a natural β -mixing condition, Robinson proves in fact that the statistic $\hat{\lambda}$ has asymptotically a χ^2 -distribution if $\beta_j = \mathcal{O}(j^{-b})$ where $b > \mu/(\mu - 2)$ under the moment assumption $\sup_{i,j} \mathbb{E}|G(X_i, X_j)|^\mu < \infty$.

The β -mixing assumption allows us to compare the joint distribution of the initial sequence with respect to a sequence of r.v.s with *independent* blocks. This reconstruction is due to Berbee's coupling lemma, no matter how big the size of the blocks may be. Yoshihara (1976) derives a covariance inequality that fits to U -statistics. A way to get rid of β -mixing conditions is to consider an independent realization $\tilde{X}_1, \dots, \tilde{X}_n$ of the trajectory X_1, \dots, X_n . Now a simpler estimator of τ is given by

$$\tilde{\tau} = \frac{1}{n^2h^L} \sum_{i,j=1}^n G(X_i, \tilde{X}_j)k\left(\frac{Z_i - \tilde{Z}_j}{h}\right).$$

The asymptotic behavior of this expression is easy to derive under alternative weak dependence conditions by using our results because $\tilde{\tau} = (1/h^L)\sum_{i=1}^n W_{n,h}(X_i)$ is the numerator of a Nadaraya–Watson kernel for the regression estimation problem $\mathbb{E}(s_1(t_1 - t)|v_1 = z)$ in the special case of the previous example. In fact this trick avoids the corresponding coupling construction for U -statistics. For another application of semiparametric problems, see, for example, Chen and Fan (1999).

2.6. Nonparametric Problems

For a stationary process $(Z_t)_{t \in \mathbb{Z}}$ with $Z_t = (X_t, Y_t)$, an important quantity is the regression function $r(x) = \mathbb{E}(Y_0|X_0 = x)$. Various methods to fit such a function have been developed. Nadaraya–Watson kernel estimates are very popular; see, for instance, Rosenblatt (1991), Prakasa (1983), and Robinson (1983). Let K be some kernel function that integrates to 1, Lipschitzian and with compact support.

Among other problems, one may wish to estimate the volatility of financial times series, $v(x) = \text{Var}(X_t|X_{t-1} = x)$. The question enters the general framework because $v(x) = v_2(x) - v_1^2(x)$, where $v_j(x) = \mathbb{E}(X_t^j|X_0 = x)$.

Another important problem of econometric interest is to estimate the marginal density f of a stationary sample. Density kernel estimators built on K are available. Density and regression functions of derivatives can also be estimated by using analogous procedures.

Finally, conditional quantiles are linked to the conditional distribution $F(y|x) = \mathbb{P}(X_1 \leq y | X_0 = x)$. More precisely, we denote by $q(t|x) = \inf\{y; F(y|x) > t\}$ the generalized (right-continuous, with left limits) inverse of the monotone function $y \mapsto F(y|x)$. Consistent estimators of the conditional regression $\mathbb{E}(Y_t(y)|X_t = x)$, where $Y_t(y) = \mathbb{1}_{X_{t+1} \leq y}$, provide information on the previous conditional quantiles.

3. WEAK DEPENDENCE CONDITIONS

Various generalizations of independence have been introduced to answer the econometric questions discussed in Section 2. The martingale setting was the first extension of the independence framework (Hall and Heyde, 1980); weakening martingale conditions yields mixingales. Martingale conditions are written in terms of conditional expectations, and they seem to be quite restrictive in econometric practice. NED is a more flexible tool for modeling fading memory. The ergodic theorem is the first limit theorem proved for dependent sequences. Another point of view is given by the mixing properties of stationary sequences in the sense of ergodic theory: uniform versions of such properties are the forthcoming mixing properties. Those conditions are also based on independence properties of the underlying generated σ -algebras. They are also difficult to check (see Doukhan, 1994).

Our aim is to promote the weak dependence properties, which will be seen to be much easier to prove.

3.1. Mixing

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let \mathcal{U}, \mathcal{V} be two sub σ -algebras of \mathcal{A} . Various measures of dependence between \mathcal{U} and \mathcal{V} have been introduced; among them we recall

$$\alpha(\mathcal{U}, \mathcal{V}) = \sup\{|\mathbb{P}(U \cap V) - \mathbb{P}(U)\mathbb{P}(V)|; U \in \mathcal{U}, V \in \mathcal{V}\},$$

$$\beta(\mathcal{U}, \mathcal{V}) = \mathbb{E} \sup\{|\mathbb{P}(V|\mathcal{U}) - \mathbb{P}(V)|; V \in \mathcal{V}\},$$

$$\rho(\mathcal{U}, \mathcal{V}) = \sup\{|\text{Corr}(u, v)|; u \in \mathbb{L}^2(\mathcal{U}), v \in \mathbb{L}^2(\mathcal{V})\},$$

$$\phi(\mathcal{U}, \mathcal{V}) = \sup\{|\mathbb{P}(V|U) - \mathbb{P}(V)|; U \in \mathcal{U}, V \in \mathcal{V}, \mathbb{P}(U) > 0\}.$$

These coefficients are, respectively, the strong mixing coefficient $\alpha(\mathcal{U}, \mathcal{V})$ of Rosenblatt (1956), the absolute regularity coefficient $\beta(\mathcal{U}, \mathcal{V})$ of Wolkonski and Rozanov (1959, 1961), the maximal correlation coefficient $\rho(\mathcal{U}, \mathcal{V})$ of Kolmogorov and Rozanov (1960), and the uniform mixing coefficient $\phi(\mathcal{U}, \mathcal{V})$ of Ibragimov (1962).

Let $X = (X_n)_{n \in \mathbb{Z}}$ be a discrete time stationary process. We denote $X_A = \{X_t; t \in A\}$ the A -marginal of X with $A \subset \mathbb{Z}$. Finally, $\sigma(Z)$ will denote the σ -algebra generated by an r.v. Z .

For any coefficient previously defined, say, $c(\cdot, \cdot)$, we shall call the process \mathbf{X} a c -mixing process if $\lim_{k \rightarrow \infty} c_{\mathbf{X}, k} = 0$, where $c_{\mathbf{X}, k} = c(\sigma(X_{]-\infty, 0]}), \sigma(X_{[k, +\infty[}))$. The following relations hold:

ϕ -mixing $\Rightarrow \rho$ -mixing $\Rightarrow \alpha$ -mixing, and

ϕ -mixing $\Rightarrow \beta$ -mixing $\Rightarrow \alpha$ -mixing,

and no reverse implication holds in general. See Doukhan (1994) for more information.

3.2. Mixingales and Near Epoch Dependence

Let $X = (X_n)_{n \in \mathbb{Z}}$ be a real-valued process. We let $\|X_n\|_p = (\mathbb{E}(|X_n|^p))^{1/p}$ ($1 \leq p < \infty$).

DEFINITION 3.1 (McLeish, 1975; Andrews, 1988). *Let $p \geq 1$ and let $(\mathcal{F}_n)_{n \in \mathbb{Z}}$ be an increasing sequence of σ -algebras. The sequence $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}}$ is called an \mathbb{L}_p -mixingale if there exist nonnegative sequences $(c_n)_{n \in \mathbb{Z}}$ and $(\psi(n))_{n \in \mathbb{Z}}$ such that $\psi(n) \rightarrow 0$ as $n \rightarrow \infty$ and for all integers $n \in \mathbb{Z}$, $k \geq 0$,*

$$\|X_n - \mathbb{E}(X_n | \mathcal{F}_{n+k})\|_p \leq c_n \psi(k+1), \quad (3.12)$$

$$\|\mathbb{E}(X_n | \mathcal{F}_{n-k})\|_p \leq c_n \psi(k). \quad (3.13)$$

This property of fading memory is easier to handle than the martingale condition. A more general concept is the NED on a mixing process. Its definition can be found in the work by Billingsley (1968), who considered functions of ϕ -mixing processes.

DEFINITION 3.2 (Pötscher and Prucha, 1991a, 1991b). *Let $p \geq 1$. We consider a c -mixing process $(V_n)_{n \in \mathbb{Z}}$. For any integers $i \leq j$, set $\mathcal{F}_i^j = \sigma(V_i, \dots, V_j)$. The sequence $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}}$ is called an \mathbb{L}_p -NED process on the c -mixing process $(V_n)_{n \in \mathbb{Z}}$ if there exist nonnegative sequences $(c_n)_{n \in \mathbb{Z}}$ and $(\psi(n))_{n \in \mathbb{Z}}$ such that $\psi(n) \rightarrow 0$ as $n \rightarrow \infty$ and for all integers $n \in \mathbb{Z}$, $k \geq 0$,*

$$\|X_n - \mathbb{E}(X_n | \mathcal{F}_{n-k}^{n+k})\|_p \leq c_n \psi(k).$$

This approach is developed in detail in Pötscher and Prucha (1991). Functions of $\text{MA}(\infty)$ processes can be handled using the NED concept. For instance, limit theorems can be deduced for sums of such functions of $\text{MA}(\infty)$ processes. These previous definitions translate the fact that a k -period-ahead in the first case, both ahead and backward in the second definition—projection is convergent to the unconditional mean. They are known to be satisfied by a

wide class of models. For example, martingale differences can be described as \mathbb{L}_1 -mixingale sequences, and linear processes with martingale difference innovations also.

3.3. Association

The notion of association was introduced independently by Esary, Proschan, and Walkup (1967) and Fortuin, Kastelyn, and Ginibre (1971).

The motivations of those authors were radically different because the first group of authors was working in reliability theory and the others in mechanical statistics. The condition of the second group of authors is known as FKG inequality.

DEFINITION 3.3. *The sequence $(X_t)_{t \in \mathbb{Z}}$ is associated, if for all coordinate-wise increasing real-valued functions h and k ,*

$$\text{Cov}(h(X_t, t \in A), k(X_t, t \in B)) \geq 0$$

for all finite disjoint subsets A and B of \mathbb{Z} and $\mathbb{E}(h(X_t, t \in A)^2 + k(X_t, t \in B)^2) < \infty$.

This extends the positive correlation assumption to model the notion that two stochastic processes have a tendency to evolve in a similar way.

This definition is deeper than the simple positive correlatedness. Besides the evident fact that it does not assume that the variances exist, one can easily construct orthogonal (hence positively correlated) sequences that do not have the association property. An important difference between the preceding conditions is that its uncorrelatedness implies independence of an associated sequence (Newman, 1984). Let, for instance, (ξ_k, η_k) be independent and i.i.d. $\mathcal{N}(0, 1)$ sequences. Then the sequence $(X_n)_{n \in \mathbb{Z}}$ defined by $X_k = \xi_k(\eta_k - \eta_{k-1})$ is neither correlated nor independent, and hence it is not an associated sequence. Heredity of association only holds under monotonic transformations. This unpleasant restriction will disappear under the assumption of weak dependence.

The preceding property of associated sequences was a guideline for the forthcoming definition of weak dependence. It contains the idea that weakly correlated associated sequences are also “weakly dependent.” The very explicit inequality (B.2) proves that this idea makes sense.

On the opposite side, negatively associated sequences of r.v.s are defined by a similar relation as the aforementioned covariance inequality, except for the sign of this inequality. Shao (2000) provides a lucid summary of this type of association. Then he points out a crucial property of domination by comparable independent sequences. This property breaks the seemingly parallel definitions of positively and negatively associated sequences. We shall develop this idea further.

3.4. Weak Dependence

Here we shall make more explicit the asymptotic independence between “past” and “future.” Roughly speaking, for convenient functions h and k , we shall assume that

$$\text{Cov}(h(\text{“past”}), k(\text{“future”}))$$

is small when the distance between the “past” and the “future” is sufficiently large. Define by \mathbb{L}^∞ the union of the sets \mathbb{L}^u of numerical bounded measurable functions on some euclidean space \mathbb{R}^u ($u \in \mathbb{N}^*$) and $\|\cdot\|_\infty$ the corresponding uniform norm. We define the Lipschitz modulus of a function $h : \mathbb{R}^u \rightarrow \mathbb{R}$ by

$$\text{Lip}(h) = \sup_{x \neq y} \frac{|h(x) - h(y)|}{\|x - y\|_1}, \quad \text{where } \|(z_1, \dots, z_u)\|_1 = |z_1| + \dots + |z_u|,$$

if $x = (z_1, \dots, z_u)$. Consider the class

$$\mathcal{L} = \{h \in \mathbb{L}^\infty; \|h\|_\infty \leq 1, \text{Lip}(h) < \infty\}. \tag{3.14}$$

DEFINITION 3.4 (Doukhan and Louhichi, 1999). *A sequence $(X_n)_{n \in \mathbb{Z}}$ is called $(\theta, \mathcal{L}, \psi)$ -weak dependent if there exist a sequence $\theta = (\theta_r)_{r \in \mathbb{N}}$ decreasing to zero at infinity and a function ψ with arguments $(h, k, u, v) \in \mathcal{L}^2 \times \mathbb{N}^2$ such that for any u -tuple (i_1, \dots, i_u) and any v -tuple (j_1, \dots, j_v) with $i_1 \leq \dots \leq i_u < i_u + r \leq j_1 \leq \dots \leq j_v$, one has*

$$|\text{Cov}(h(X_{i_1}, \dots, X_{i_u}), k(X_{j_1}, \dots, X_{j_v}))| \leq \psi(h, k, u, v)\theta_r \tag{3.15}$$

if the functions h and k are defined, respectively, on \mathbb{R}^u and on \mathbb{R}^v .

Notice that the sequence θ depends both on the class \mathcal{L} and on the function ψ . The function ψ can in fact depend on all its arguments, contrary to the case of bounded mixing sequences. This definition is hereditary through images by convenient functions.

The examples of interest to follow involve the function $\psi_1(h, k, u, v) = u\text{Lip}(h) + v\text{Lip}(k)$, $\psi'_1(h, k, u, v) = v\text{Lip}(k)$, $\psi_2(h, k, u, v) = uv\text{Lip}(h)\text{Lip}(k)$, and $\psi'_2(h, k, u, v) = v\text{Lip}(h)\text{Lip}(k)$. For example, proving that an MA(∞) process $X_n = \sum_{k \geq 0} a_k \xi_{n-k}$ based on an i.i.d. sequence such that $\mathbb{E}|\xi_0| < \infty$ and $\sum_k |a_k| < \infty$ is ψ'_1 -weakly dependent with $\theta_r = \mathbb{E}|\xi_0| \sum_{k > r} |a_k|$ is based on the decomposition $X_n = \bar{X}_n + \underline{X}_n$ with $\bar{X}_n = \sum_{k < r} a_k \xi_{n-k}$. In this case, assuming for simplicity that $v = 1$ and $j_1 = n$, we have³

$$|\text{Cov}(h(\text{“past”}), k(X_n))| = |\text{Cov}(h(\text{“past”}), k(X_n) - k(\bar{X}_n))| \leq \text{Lip}(h)\mathbb{E}|\underline{X}_n|.$$

4. MODELS: DEPENDENCE PROPERTIES

4.1. Markovian Models

4.1.1. Models with a Markovian Representation. Let $(\xi_n)_{n \in \mathbb{Z}}$ be a real-valued i.i.d. sequence and let M be some function. We consider vector-valued models driven by the equation

$$X_n = M(X_{n-1}, \dots, X_{n-p}, \xi_n).$$

To justify the title of this section, note that the vector-valued sequence $(X_n^{(p)})_{n \in \mathbb{Z}}$, where $X_n^{(p)} = (X_{n-1}, \dots, X_{n-p})$, is Markovian. Using Proposition 7.6 of Kallenberg (1997) proves that any Markov process has such a representation.

Under “reasonable” assumptions (described subsequently) such models can be rewritten as ergodic Markov chains (see Meyn and Tweedie, 1993; Tjøstheim, 1990). Thus, the stationarity assumption is reachable. An interesting class is given by

$$X_n = R(X_{n-1}, \dots, X_{n-p}, \eta_n) + S(X_{n-1}, \dots, X_{n-p})\xi_n,$$

where $(\xi_n)_{n \in \mathbb{Z}}$ and $(\eta_n)_{n \in \mathbb{Z}}$ are two mutually independent i.i.d. sequences and the function S satisfies $S(x_1, \dots, x_p) \geq s > 0$ for some $s \in \mathbb{R}$, and any real numbers x_1, \dots, x_p and the functions R and S essentially satisfy contraction assumptions (for developments, see Doukhan, 1994; Ango Nze, 1995, 1998; Duflo, 1990).

For instance, ARMA(p, q) processes

$$Y_n = \sum_{i=1}^p a_i Y_{n-i} + \xi_n + \sum_{j=1}^q b_j \xi_{n-j}$$

have such a Markov representation in the case when the roots of the polynomial $\sum_{j=1}^q b_j x^j$ lie outside the unit disk. Indeed, $X_n = (Y_n, Y_{n+1}, \dots, Y_{n+\ell-1})$, where $\ell = \max\{p, q + 1\}$ and $Y_{n,j} = \mathbb{E}[Y_j | Y_i : i \leq n]$ is a Markov process. See Mokkadem (1990).

A further example is that of bilinear models,⁴ which are popular in econometrics

$$X_n = (a + b\xi_{n-2})X_{n-2} + (c + d\xi_{n-1})X_{n-1} + \xi_n.$$

Examples of such models are also doubly stochastic autoregressive processes:

$$X_n = \eta_n X_{n-1} + \xi_n.$$

Econometricians have introduced generalized ARCH-GARCH processes:

$$X_n = r(X_{n-1}, \dots, X_{n-p}) + s(X_{n-1}, \dots, X_{n-p})\xi_n$$

to model conditional variances (interpreted as, e.g., an asset volatility in finance theory) that change over time (for further references, see Bollerslev, 1986). These models are known to satisfy the NED property of Definition (3.2).

Note that functional autoregressive models correspond to constant functions s (see Bollerslev, 1986).

Moreover, threshold models are those for which r is linear on a partition of the space into polygonal regions. For example, Petrucci and Woolford (1984) study threshold autoregressive models such as

$$X_n = aX_{n-1}^+ + bX_{n-1}^- + \xi_n,$$

where $x^+ = \max(x, 0)$ and $x^- = \min(x, 0)$. This model is ergodic if $a < 1$, $b < 1$, and $ab < 1$ and has geometric rates of convergence in total variation to the stationary limit if the centered sequence (ξ_n) has finite exponential moment and its distribution has a density with respect to Lebesgue measure. If, for instance, $(a, b) = (\frac{1}{2}, -2)$, the function $r(x) = ax^+ + bx^-$ is relevant, but it is not a contraction.

ARCH or GARCH models are those with nonconstant functions s , such as square roots of nonnegative polynomials with degree 2, namely,

$$X_n = aX_{n-1} + (b^2 + c^2X_{n-1}^2)^{1/2}\xi_n,$$

with $|a| + |c| < 1$. Vector-valued versions of such models can also be described. They include GARCH models. He and Teräsvirta's paper (1999) looks at the existence of marginal moments and conditions for stationarity of GARCH models. The following example of a Markovian nonmixing sequence is given in Andrews (1984) and Rosenblatt (1985). This is the (Markov) AR(1)-process with binomial innovations ($\mathbb{P}(\xi_0 = 0) = \mathbb{P}(\xi_0 = 1) = \frac{1}{2}$):

$$X_n = \frac{1}{2}(X_{n-1} + \xi_n). \quad (4.16)$$

This is also the Bernoulli shift $X_n = H(\xi_n, \xi_{n-1}, \dots)$ with $H(x) = \sum_{k=0}^{\infty} 2^{-(k+1)}x_k$. Full definitions of Bernoulli shifts will be given in Section 4.3. This model has stationary uniform distribution on the interval $[0, 1]$, but it satisfies no mixing condition. Indeed, the innovations ξ_j ($j \leq n$) are the digits of the dyadic expansion of X_n ; hence, X_k is a deterministic function of X_n for $k \leq n$. An extension of this model to innovations taking p different values is immediate; for this, one can use the numeration in basis p . The process $X_n = 0.\xi_n\xi_{n-1}\dots$ is the solution of the recurrence equation $X_n = (1/p)(X_{n-1} + \xi_n)$ if the innovations are uniform on $\{0, 1, \dots, p-1\}$.

4.1.2. Weak Dependence Properties. Lipschitzian models (see Duflo, 1990) are multivariate Markov models, defined recursively through $X_n = M(X_{n-1}, \xi_n)$ and the assumptions that

$$\mathbb{E}\|M(x, \xi_n) - M(y, \xi_n)\|^S \leq a\|x - y\|^S \quad \text{and} \quad \mathbb{E}\|M(0, \xi_n)\|^S < \infty,$$

for some $0 \leq a < 1$ and $S \geq 1$. Here, the \mathbb{R}^p -space where the process lives is equipped with some—not necessarily euclidean—norm $\|\cdot\|$. Duflo (1990) introduces the concept of stability of such Markov chains. She proves their geometric stability. That is, denoting $\Phi_n(X_0)$ to be precisely the initial state X_0 , there exists some $c \in [0, 1[$ such that for any $x, y \in \mathbb{R}^p$

$$\|\Phi_n(x) - \Phi_n(y)\| \leq c^n \|x - y\|.$$

In the particular case where X_0 's distribution is the stationary probability measure,

$$\|\Phi_n(x) - \Phi_n(X_0)\| \leq c^n \|x - X_0\|.$$

Using those results, Doukhan and Louhichi (1999) deduce $(\theta, \mathcal{L}, \psi'_1)$ -weak dependence. In fact, under the assumptions that follow, one has $\theta_r = \mathcal{O}(a^r)$ and $\psi'_1(h, k, u, v) = v \text{Lip}k$. Here neither stationarity nor any further regularity assumption on the sequence of innovations is required. Such contraction properties are also used by Pötscher and Prucha (1991). More general AR(p) non-linear models, $X_n = M(X_{n-1}, \dots, X_{n-p}; \xi_n)$, have the same properties, if, for example, $\mathbb{E}|M(0; \xi_n)| < \infty$ and, for some constants $a_j \geq 0, 1 \leq j \leq p$ with $\sum_{j=1}^p a_j < 1$,

$$\mathbb{E}|M(x_1, \dots, x_p; \xi_n) - M(y_1, \dots, y_p; \xi_n)| \leq \sum_{j=1}^p a_j |x_j - y_j|.$$

The more recent papers by Diaconis and Friedmann (1999) and Jarner and Tweedie (2001) provide a wide range of examples in this spirit. Alsmeyer and Fuh (2001) give conditions for arithmetic decay of the weak dependence coefficient sequence. Both papers study iterated random sequences $M_n = F(\varepsilon_n, M_{n-1})$ for independent sequences (ε_n) and some F , measurable and Lipschitz in the second variable. The process (M_n) takes values in a complete separable metric space (E, d) and forms a Markov chain. Under the assumption of existence of the unique invariant distribution π , both papers prove, using different methods, that $\tilde{d}(\mathbb{P}(M_n \in \cdot | M_0 = x), \pi(\cdot)) \leq A_x(n + 1)^{-p}$ if for some $x_0 \in E$ and $p > 0$

$$\mathbb{E} \log^{p+1}(1 + \text{Lip}(F(\varepsilon_1, \cdot))) < \infty \quad \text{and} \quad \mathbb{E} \log^{p+1}(1 + d(F(\varepsilon_1, x_0), x_0)) < \infty. \tag{4.17}$$

The distance \tilde{d} is the Prohorov metric associated with d .

4.1.3. Mixing Properties. Mixing properties of the models with a Markovian representation, $X_n = M(X_{n-1}, \dots, X_{n-p}, \xi_n)$, are described in Meyn and Tweedie (1993). The preceding models are ergodic under suitable assumptions on ξ_0 's distribution.

Assume that $\mathbb{E}|\xi_0| < \infty$ and assume the existence of an almost surely non-vanishing density f for ξ_0 's law. Then, under contraction assumptions on the function M , one can prove that, under the invariant initial distribution, $\beta_n =$

$\mathcal{O}(n^{-b})$ or $\beta_n = \mathcal{O}(e^{-bn})$. If $M(X_{n-1}, \dots, X_{n-p}, \xi_n) = R(X_{n-1}, \dots, X_{n-p}) + \xi_n$, then the second relation holds if, for example,

$$|R(x_1, \dots, x_p)| \leq A + a_1|x_1| + \dots + a_p|x_p| \quad \text{with } a_1 + \dots + a_p < 1 \quad (4.18)$$

and R is continuous. (It is enough that R be continuous out of a null set; e.g., a piecewise continuous function R is relevant, as in the previously mentioned threshold model by Petruccielli and Woolford, 1984. For more details, see Doukhan, 1994). In fact, Davydov (1973) has proved that Harris recurrent Markov chains are ergodic and β -mixing when stationary; moreover, denoting by μ the stationary distribution of the Markov chain X_n , by P the transition probability kernel, and by $\|\cdot\|_{TV}$ the norm in total variation, one has that

$$\beta_n = \int \|P^n(x, \cdot) - \mu\|_{TV} \mu(dx).$$

Returning to the more specific models introduced before, Ango Nze (1995) proves that (4.18) implies the β -mixing property under the assumption that ξ_0 's distribution has a density with respect to Lebesgue measure. The mixing coefficients decrease at a geometric rate. If, moreover, $p = 1$ in the preceding (functional AR(1)) model, he proves (see also Doukhan, 1994) that, under the previous assumptions on the white noise (ξ_n),

$$-\lim_{n \rightarrow \infty} \beta_n = 0 \quad \text{if } |R(x)| \leq |x| - \mathbb{E}|\xi_0| - \epsilon \quad \text{for some } \epsilon > 0,$$

x is large enough, and R is locally bounded.

The expression $-\beta_n = \mathcal{O}(n^{-b})$ if $|R(x)| \leq |x| - c|x|^{b/(1+b)} + A$ for some constants $c, A > 0$, and any real number x . The function R is continuous. A more general result is obtained in Ango Nze (1998) if $\mathbb{E}|\xi_0|^S < \infty$. Veretennikov (1999) improves on the previous hyperbolic mixing decay assumptions. Under a local Doëblin condition (implied by the preceding absolute continuity assumptions on ξ_0 's distribution), he proves that $\beta_n = \mathcal{O}(n^{-b})$ if $b < S/2 - 1$ where $S > 4$ satisfies $\mathbb{E}|\xi_0|^S < \infty$. The existence of the stationary distribution is proved under the relaxed condition $S \geq 2$. Improved results are provided in Fort and Moulines (2002); they are clarified in the work by Jarner and Tweedie (2002), where constants are explicitly given.

The expression $-\beta_n = \mathcal{O}(e^{-bn})$ if $|R(x)| \leq B|x| + A$ for some constants $B < e^{-b}$ and $A > 0$ and any real number x . The function R is continuous.

If the innovations have a finite exponential moment, $\mathbb{E}e^{c|\xi_0|} < \infty$, Mokkadem (1990) proves that the assumptions $|R(x)| \leq |x| - \epsilon$ for some $\epsilon > 0$ and $|x|$ big enough to ensure an analogous result: the mixing sequence (β_n) decays at an exponential rate.

The expression $-\phi_n = \mathcal{O}(e^{-bn})$ if $R(x)$ is a bounded function, continuous outside a null set, and ξ_0 's law is not orthogonal to Lebesgue measure; moreover, the stationarity is no longer required. Unfortunately, this drastic boundedness condition excludes, for example, the linear autoregressive processes.

The preceding results provide (upper) bounds for the mixing coefficients. It is a much harder problem to derive both upper and lower bounds of the mixing sequences from the assumptions about the model; for results about some types of Markov sequences, see Davydov (1973). Meyn and Tweedie (1993) also provide necessary and sufficient conditions for geometric ergodicity of threshold autoregressive linear processes (the function R is piecewise linear; see also Cline and Pu, 1998). Doubly stochastic AR models are geometrically ergodic if ξ 's distribution has an absolutely continuous component and $\mathbb{E}|\xi_0| < \infty$ and $\mathbb{E}|\eta_0| < 1$ (see Tjøstheim, 1986).

For the other models, we refer to Pham (1986), Doukhan (1994), and Ango Nze (1998).

4.2. Associated Sequences

Associated sequences with finite variances are $(\theta, \mathcal{L}, \psi_2)$ -weak dependent with $\theta_r = \sup_{i \geq r} \text{Cov}(X_0, X_i)$ (see Doukhan and Louhichi, 1999). Note that broad classes of examples of associated processes result from the fact that any independent sequence is associated and that monotonicity preserves association (for this, see Newman, 1984).

The case of Gaussian sequences is analogous. One may also consider combinations of sums of Gaussian and associated sequences, or Bernoulli shifts driven by stationary, associated, instead of i.i.d. sequences.

Linear processes with nonnegative coefficients are associated, and so are functional autoregressive processes with nondecreasing regression functions. Note that for associated or Gaussian sequences, the function ψ_2' replaces ψ_2 if $\theta_r = \sup_{i \geq r} |\text{Cov}(X_0, X_i)|$ is replaced by $\theta_r = \sum_{i \geq r} |\text{Cov}(X_0, X_i)|$.

Giraitis, Kokoszka, and Leipus (2000) consider ARCH(∞)-models (4.19) with nonnegative coefficients and nonnegative inputs. In that case the models are also associated.

4.3. Bernoulli Shifts

DEFINITION 4.1. *Let $(\xi_i)_{i \in \mathbb{Z}}$ be a sequence of i.i.d. real-valued r.v.s and the function $H: \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}$ be measurable. The sequence $(X_n)_{n \in \mathbb{Z}}$ is called a Bernoulli shift if it is defined by $X_n = H(\xi_{n-j}, j \in \mathbb{Z})$.*

We refer the reader to Ornstein and Weiss (1990), where such models are motivated through deep ergodic theoretic arguments.

One-sided shifts are defined as $X_n = H(\xi_n, \xi_{n-1}, \xi_{n-2}, \dots, \xi_0, \xi_{-1}, \xi_{-2}, \dots)$, that is, $H: \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$. The model described in equation (4.16) is an example of such a shift: $H(x_0, x_1, \dots) = \sum_{j=0}^{\infty} 2^{-1-j} x_j$. The previous model is a simple example of a weakly dependent but possibly nonmixing sequence.

4.3.1. Markov Sequences. A general situation where sequences are one sided is the following Markov stationary setting. Consider a Markov process driven

by the updating equation $X_t = M(X_{t-1}, \xi_t)$, for some i.i.d. sequence $(\xi_t)_{t \in \mathbb{Z}}$; then the function H if it exists is defined implicitly by the relation $H(x) = M(H(x'), x_0)$, where $x = (x_0, x_1, x_2, \dots)$, $x' = (x_1, x_2, x_3, \dots)$. To consider more general Markov sequences one may also refer to the previous section devoted to Markov processes. To prove such Bernoulli shift representations, Mokkadem (1990) and Meyn and Tweedie (1993) use the tools of control theory.

4.3.2. *Chaotic Representations.* We now specialize the analysis to chaotic expansions associated with the discrete chaos generated by the sequence $(\xi_i)_{i \in \mathbb{Z}}$. Let $x = (x_i)_{i \in \mathbb{Z}}$; we write in a condensed formulation $H(x) = \sum_{k=0}^{\infty} H^{(k)}(x)$, where $H^{(k)}(x)$ denotes the k th-order chaos contribution, $H^{(0)}(x) = a_0^{(0)}$, is only a centering constant and for $k > 0$,

$$H^{(k)}(x) = \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} \dots \sum_{j_k=-\infty}^{\infty} a_{j_1, \dots, j_k}^{(k)} x_{j_1} x_{j_2} \dots x_{j_k},$$

or in short, in vector notation, $H^{(k)}(x) = \sum_{j \in \mathbb{Z}^k} a_j^{(k)} x_j$.

Processes associated with a finite number of chaotic terms (i.e., $H^{(k)} = 0$ if $k > k_0$) are also called *Volterra processes*. The first example of such a Volterra process is clearly the class of linear processes that includes autoregressive moving average (ARMA) processes: it corresponds to the consideration of just a term in the first chaos (i.e., $k = 1$ in the previous representation); it is widely used in the field of statistics (see, e.g., Rosenblatt, 1985). A simple and general condition for \mathbb{L}^1 -convergence of such series is, still in a condensed notation, $\sum_{k=0}^{\infty} \{ \sum_{j \in \mathbb{Z}^k} |a_j^{(k)}| \} \mathbb{E} |\xi_0|^k < \infty$.

The simple bilinear process, $X_t = (a + b\xi_{t-1})X_{t-1} + \xi_t$, is stationary if $c = \mathbb{E} |a + b\xi_0| < 1$ (see, e.g., Tong, 1981). It is a Bernoulli shift with $H(x) = x_0 + \sum_{j=1}^{\infty} x_j \prod_{s=1}^j (a + bx_s)$, for $x = (x_i)_{i \in \mathbb{N}}$.

More general affine models are considered in Mokkadem (1990).

ARCH(∞)-models (see Giraitis et al., 2000) are given by a sequence $(b_j)_{j \geq 1}$ and an i.i.d. sequence of r.v.s $(\xi_j)_{j \geq 0}$ through the recursive relation

$$X_t = \left(a + \sum_{j=1}^{\infty} b_j X_{t-j} \right) \xi_t. \tag{4.19}$$

Such models have a stationary representation with the chaotic expansion

$$X_t = a \sum_{\ell=1}^{\infty} \sum_{j_1=1}^{\infty} \dots \sum_{j_{\ell}=1}^{\infty} b_{j_1} \dots b_{j_{\ell}} \xi_{t-j_1} \dots \xi_{t-(j_1+\dots+j_{\ell})}$$

under the simple assumption $\mathbb{E} |\xi_0| \sum_{j=1}^{\infty} |b_j| < 1$.

4.3.3. *Mixing Properties.* Finite moving averages $X_n = H(\xi_n, \xi_{n-1}, \dots, \xi_{n-m})$ are trivially m -dependent.

The Bernoulli shift $X_n = H(\xi_n, \xi_{n-1}, \dots)$ (with $H(x) = \sum_{k=0}^\infty 2^{-(k+1)}x_k$) is not mixing; this is again example (4.16) of a Markovian, nonmixing,⁵ sequence. The difference between two such independent processes of this type provides an example of a nonassociated and nonmixing process.

Hence one cannot expect sufficient condition for mixing in such weak shifts.

4.3.4. $(\theta, \mathcal{L}, \psi)$ -Weak Dependence. Contrary to mixing conditions, it can be proved that even two-sided sequences can be $(\theta, \mathcal{L}, \psi)$ -weak dependent. For instance, for infinite moving averages $X_t = \sum_{i=-\infty}^\infty a_i \xi_{t-i}$, $\theta_r = 2\mathbb{E}|\xi_0| \times \sum_{|i|>r/2} |a_i|$ and $\psi(h, k, u, v) = (u\text{Lip}(h) + v\text{Lip}(k))$. Note also, for completeness, that (NED) conditions can also be deduced for such two-sided models. More generally, we can state the following definition.

DEFINITION 4.2. For any integer $k > 0$, we denote by δ_r any number such that

$$\sup_{i \in \mathbb{Z}} \mathbb{E}|H(\xi_{i-j}, j \in \mathbb{Z}) - H(\xi_{i-j} \mathbb{1}_{|j|<r}, j \in \mathbb{Z})| \leq \delta_r.$$

Such sequences $(\delta_r)_{r \in \mathbb{Z}^+}$ are related to the modulus of uniform continuity of H ; that is, if for positive constants $(a_i)_{i \in \mathbb{Z}}, 0 < b \leq 1$, the inequality $|H(u_i, i \in \mathbb{Z}) - H(v_i, i \in \mathbb{Z})| \leq \sum_{i \in \mathbb{Z}} a_i |u_i - v_i|^b$ holds for any sequences $(u_i), (v_i) \in \mathbb{R}^\mathbb{Z}$, and if the sequence $(\xi_i)_{i \in \mathbb{Z}}$ has a finite moment of order b , then one can choose $\delta_r = \sum_{|i| \geq r} a_i \mathbb{E}|\xi_i|^b$.

PROPOSITION 4.1 (Doukhan and Louhichi, 1999). *Bernoulli shifts are $(\theta, \mathcal{L}, \psi)$ -weak dependent with $\theta_r = 2\delta_{r/2}$ and $\psi(h, k, u, v) = 4(u\|k\|_\infty \text{Lip}(h) + v\|h\|_\infty \text{Lip}(k))$. If, moreover, the Bernoulli shift is one sided, then it is $(\theta, \mathcal{L}, \psi)$ -weak dependent with $\theta_r = \delta_r$ and $\psi(h, k, u, v) = 2v\text{Lip}k\|h\|_\infty$.*

We turn back to Volterra expansions. A suitable bound for δ_r corresponds here to the stationarity condition

$$\delta_r = \sum_{k=0}^\infty \left\{ \sum_{j \in \mathbb{Z}^k; \|j\|_\infty > r} |a_j^{(k)}| \right\} \mathbb{E}|\xi_0|^k < \infty.$$

The one-sided example of a simple bilinear process, $X_t = (a + b\xi_{t-1})X_{t-1} + \xi_t$, with convergent chaotic representation for $c = \mathbb{E}|a + b\xi_0| < 1$ satisfies $\delta_r = \theta_r = c^r(r + 1)/(1 - c)$; it has a geometric rate of decay under a stationarity condition set out by Tong (1981). The stochastic volatility model

$$X_n = \mu + \xi_n \exp\left(\frac{h_n}{2}\right), \quad h_n = \sum_{j=1}^\infty \beta_j \eta_{n-j}$$

is another example yielding a one-sided chaotic decomposition. The sequence (η_j) is assumed to consist of independent r.v.s and to be independent of the cen-

tered reduced sequence (ξ_n) . The chaotic representation converges if $\sum |\beta_j| < \infty$ and $\prod \mathbb{E} \exp(|\beta_j \eta_0|) < \infty$.

4.3.5. Linear Processes. Using suitable assumptions on the law of ξ_0 , the one-sided linear processes $X_n = \sum_{t=0}^{\infty} a_t \xi_{n-t}$ satisfy β -mixing conditions. This requires the absolute continuity of ξ_0 's density. If $\mathbb{E}\xi_0 = 0$, and if, moreover, for some $\delta > 0$, $\mathbb{E}|\xi_0|^{1+\delta} < \infty$ and $\sum_{t=0}^{\infty} |a_t| < \infty$, then $\beta_n \leq C \sum_{l=n}^{\infty} (\sum_{k=l}^{\infty} |a_k|)^{\delta/1+\delta}$, for some $C > 0$ (Pham and Tran, 1985). See Doukhan⁶ (1994) for a bibliography; nevertheless, the study of two-sided linear sequences herein is not complete.

5. LIMIT THEOREMS

The aim of this section is to present the state of the art of the limit theorems for stationary sequences.

5.1. The Donsker Line

Consider a stationary sequence $(X_n)_{n \in \mathbb{Z}}$. We assume that this sequence is integrable and centered at expectation

$$\mathbb{E}X_0 = 0.$$

Denote by $[x]$ the integer part of a real number x ($[x] \leq x < [x] + 1$). The Donsker line $(D_n(t))_{t \in [0,1]}$ is defined for any sample with positive size n as the following continuous time process:

$$D_n(t) = \sum_{k=1}^{[nt]} X_k + (nt - [nt])X_{[nt]+1}.$$

We consider the following convergence result in the space $\mathcal{C}([0,1])$ of continuous functions on the unit interval when the sample size n grows to infinity.

THEOREM 5.1. *The following functional convergence holds in the space $\mathcal{C}([0,1])$ under any of the weak dependence conditions formulated subsequently:*

$$\frac{1}{\sqrt{n}} D_n(t) \xrightarrow[n \rightarrow \infty]{\mathcal{C}([0,1])} \sigma W_t.$$

Here, $\sigma^2 = \sum_{-\infty}^{\infty} \text{Cov}(X_0, X_k)$ (the series is assumed to be convergent).

Recall that here $(W_t)_{t \in [0,1]}$ is the standard Brownian motion; that is, W denotes the centered Gaussian real-valued process with covariance function

$$\mathbb{E}W_s W_t = \min\{s, t\}.$$

To avoid triviality we shall also assume that $\sigma \neq 0$.

The preceding FCLT is known to hold in the cases that follow.

5.1.1. *Strong Mixing Case (α -Mixing).* Recall that the quantile function Q of the distribution of X_0 is also the càdlàg (left continuous with right limits) inverse of the tail function $s \rightarrow \mathbb{P}(|X_0| > s)$ and that α^{-1} is the càdlàg inverse of the monotone function $t \rightarrow \alpha_{[t]}$. The condition

$$\int_0^1 \alpha^{-1}(u)Q^2(u) du < \infty \tag{DMR}$$

implies the FCLT in Theorem 5.1 and also the convergence of the series in the definition of σ^2 (for more details, see Rio, 2000). Let $\delta > 0$. Assume that moment of order $(2 + \delta)$ of X_0 is finite. Condition (DMR) is equivalent to

$$\sum_{n=0}^\infty n^{2/\delta} \alpha_n < \infty.$$

The previous FCLT result was obtained by Davydov (1973) under the slightly stronger assumption $\sum_{n=0}^\infty \alpha_n^{\delta/(2+\delta)} < \infty$. In other words, both series converge for the same hyperbolic mixing decays $\alpha_n \sim n^{-a}$ for $a > 1 + 2/\delta$. Note that no gain seems to be obtained here when one considers β -mixing sequences.

5.1.2. *ρ -Mixing Case.* The condition

$$\sum_{n=0}^\infty \rho(2^n) < \infty$$

with $\mathbb{E}X_0^2 < \infty$ implies the FCLT, as proven by Shao (1988). It is well known that the preceding conditions ensure nice behavior of second-order moments.

5.1.3. *Associated Case.* The condition

$$\sum_{n=0}^\infty \text{Cov}(X_0, X_n) < \infty$$

implies the functional convergence (see Newman and Wright, 1981). Clearly, this condition is also necessary to ensure that Theorem 5.1 holds.

Notice that the property “orthogonality implies independence” makes this condition credible for this very special case of an associated sequence.

The FCLT is phrased for a strictly stationary negatively associated sequence in identical terms. In fact, the tightness of D_n/\sqrt{n} is shown using an exponential inequality that results from a comparison theorem on moment inequalities proved by Shao (2000).

5.1.4. *Nonlinear Functions of Linear Processes.* We consider two-sided linear sequences. If the coefficients in the linear process $X_k = \sum_{t=-\infty}^\infty a_t \xi_{k-t}$ satisfy $a_t = \mathcal{O}(t^{-(D+1)/2})$, then $\text{Cov}(X_0, X_n) = r(n) = \mathcal{O}(n^{-D})$ is the decorrelation rate of the sequence. The FCLT holds under the condition $D > 1$ (see Giraitis and Surgailis, 1986).

Giraitis and Surgailis (1986) prove the result for the partial sums of $\Phi(X_k)$ if Φ is some polynomial function with Appell rank m (see definition that follows). Recall that Appell polynomials are defined through the relation

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} A_k(x) \mathbb{E} \exp(zX_0) = \exp(zx).$$

Like Hermite polynomials, they satisfy a recursive relation: for all $k \in \mathbb{Z}_+$,

$$A'_k = kA_{k-1}.$$

Setting $\mu_k = \mathbb{E}X_0^k$, one obtains, for instance,

$$A_0(x) = 1, \quad A_1(x) = x - \mu_1, \quad A_2(x) = x^2 - 2\mu_1x + 2\mu_1^2 - \mu_2, \dots,$$

$$A_k(x) = x^k + \dots.$$

Let F denote the distribution function of X_0 . An analytic function Φ such that $\mathbb{E}\Phi^2(X_0) < \infty$ has an uniquely defined Appell expansion:

$$\Phi(x) = \sum_{k=0}^{\infty} \frac{c_k}{k!} A_k(x), \quad \text{with } c_k = \int_{-\infty}^{\infty} \Phi^{(k)}(x)F(dx) = \mathbb{E}\Phi^{(k)}(X_0).$$

Now, if the distribution function F is infinitely differentiable, then, setting $f = F'$ for the density function, one obtains $c_k = (-1)^k \int_{-\infty}^{\infty} \Phi(x)f^{(k)}(x) dx$. A straightforward integration by parts yields

$$\mathbb{E}A_k(X_0)Q_l(X_0) = k! \quad \text{if } k = l \quad (\text{and } = 0 \text{ otherwise}), \quad \text{with } Q_k(x) = \frac{f^{(k)}(x)}{f(x)}.$$

This means that the system of functions $(A_k, Q_k)_{k \geq 0}$ is biorthogonal.

It is suitable to define the Appell rank of Φ as the smallest integer m such that $c_m \neq 0$. Appell rank is thus uniquely defined at least for polynomials. The system of Appell polynomials is not orthogonal (except for the special case of Hermite polynomials, which are associated with Gaussian distributions). Hence existence and uniqueness of such expansions follow from additional conditions such as analyticity. We refer to Giraitis and Surgailis (1986) for details. Giraitis and Surgailis (1986) assume the existence of moments of any order and $\sum_{n=0}^{\infty} |\text{Cov}(X_0, X_n)|^m < \infty$. The functional convergence is ensured by the Chentsov tightness criterion, given in Appendix B. The reason is that the method of moments is used to prove the CLT.

Concerning one-sided sequences, Ho and Hsing (1997) obtain an analogous CLT for more general nonlinear functionals of a one-sided linear sequence. The idea is to approximate such nonlinear functions of a one-sided linear sequence by m -dependent moving averages that are easily shown to satisfy an \sqrt{n} -CLT. The main assumptions are $\mathbb{E}\xi_0^4 < \infty$ and the following regularity condition. The regularity is twofold. For any subset J of the set \mathbb{N} of nonnegative integers, define $X^{(J)} = \sum_{k \in J} a_k \xi_k$ ($X^{(\mathbb{N})} = X_0$).

A C^1 -condition on the functions

$$\Phi_J(x) = \mathbb{E}\Phi(x + X^J)$$

is required if $\text{Card}(J)$ is large enough. Hence, even a very weak regularity condition on the marginal innovations such as $|\mathbb{E} \exp(iu\xi_0)| \leq C/(1 + |u|^\delta)$ for a small $\delta > 0$ implies the regularity of the distribution of Y^J , hence the regularity needed on Φ_J 's. Now Burkholder's inequality for martingales with $\mathbb{E}|\xi_0|^{4+\delta} < \infty$ still yields the Chentsov tightness criterion.

Such linear sequences are also $(\theta, \mathcal{L}, \psi)$ -weakly dependent sequences. Therefore, one may refer to Section 4.3.

5.1.5. Gaussian Processes. The same finite-dimensional CLT as for linear processes holds for instantaneous functions of Gaussian stationary sequences when one replaces m by the Hermite rank of an arbitrary function Φ such that $\mathbb{E}\Phi^2(Y_k) < \infty$ (see Breuer and Major, 1983).

However, Donsker's Theorem 5.1 requires an additional tightness condition. Chambers and Slud (1989) introduce such a condition in terms of the coefficients of Φ 's Hermite expansion. Assume that $\Phi = \sum_{k=m}^\infty a_k H_k$, where $\sum_{k=m}^\infty 3^{k/2} \sqrt{k!} |a_k| < \infty$. This means that an exponential decay of the coefficients is needed to obtain Donsker's theorem.

Chambers and Slud provide a result for general stationary processes that are built on a Gaussian chaos. Such functionals may fail to be instantaneous functions $\Phi(Y_n)$. They can be written as general Bernoulli shifts of $(Y_n)_{n \in \mathbb{Z}}$:

$$X_n = H(Y_n, Y_{n-1}, Y_{n-2}, \dots).$$

The authors also consider instantaneous functionals of Gaussian sequences satisfying CLT but not the Donsker theorem.

Recall, however, that a smooth lower bound assumption on the spectral density of the process yields ρ -mixing. Hence Theorem 5.1 still holds under slight additional conditions on the Gaussian process. The assumption concerning the function Φ is unchanged: $\mathbb{E}\Phi^2(\mathcal{N}(0,1)) < \infty$.

5.1.6. $(\theta, \mathcal{L}, \psi)$ -Weak Dependence. Assume a $(\theta, \mathcal{L}, \psi)$ -weak dependence condition with $\theta_r = \mathcal{O}(r^{-a})$, for the stationary sequence $(X_n)_{n \in \mathbb{Z}}$. Suppose also that for some $\delta > 0$, $\mathbb{E}|X_0|^{4+\delta} < \infty$.

Then if the function ψ associated with weak dependence is ψ_1 (respectively, ψ_2), Doukhan and Louhichi (1999) prove the FCLT if

$$a \geq 2 + 4/\delta \quad (\text{resp. } a > 2).$$

So, without any regularity condition on innovations, Theorem 5.1 holds for a bounded Lipschitz function of a linear process if $a_k = \mathcal{O}(k^{-D})$ when $D > 3$. The latter doesn't need to be one sided, whereas Ho and Hsing (1997) need this assumption. Moreover, the functions Φ available are more general than those considered by Giraitis and Surgailis (1986).

Finally, a faster hyperbolic decay of coefficients in place of the boundedness assumption for Φ together with the finiteness of the fourth-order moment for the innovations yields the FCLT.

Define the class $\mathcal{F}_{exp} = \{h_{s,u}, k_{s,u}; (u, s) \in \mathbb{N} \times \mathbb{R}\}$ of real-valued functions

$$h_{s,u}(x_1, \dots, x_u) = \sin(s(x_1 + \dots + x_u))$$

$$k_{s,u}(x_1, \dots, x_u) = \cos(s(x_1 + \dots + x_u)).$$

Assuming \mathcal{F}_{exp} -weak dependence, we can even improve on the preceding condition by the following uniform bound on the set of integers $i \leq j \leq k \leq l$ and $j + r \leq k$

$$\text{Cov}(X_i X_j, X_k X_l) = \mathcal{O}(r^{-2}) \quad \text{and} \quad \text{Cov}(X_j, X_k) = \mathcal{O}(r^{-2}).$$

5.1.7. Martingales and Generalizations. In this section, we consider conditions in terms of conditional expectations with respect to an adapted filtration. We first recall that Theorem 5.1 holds for martingales with stationary square integrable increments such that $\mathbb{E}X_0^2 < \infty$ (see Billingsley, 1968). More generally, let $(X_n)_{n \in \mathbb{Z}}$ be a process adapted to the filtration $(\mathcal{M}_n)_{n \in \mathbb{Z}}$: X_n is \mathcal{M}_n -measurable for any $n \in \mathbb{Z}$. The following result is proved by Dedecker and Rio (2000). All ergodicity assumptions are gone. Let $T: \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}^{\mathbb{Z}}$ denote the right shift operator (i.e., $(T(x))_n = x_{n+1}, n \in \mathbb{Z}$). Denote by \mathcal{I} the tail σ -algebra of T -invariant Borel sets of $\mathbb{R}^{\mathbb{Z}}$.

THEOREM 5.2. *Assume that $\mathbb{E}X_n^2 < \infty, \mathbb{E}X_n = 0$ for any $n \in \mathbb{Z}$ and*

$$\sum_{n=0}^{\infty} X_0 \mathbb{E}(X_n | \mathcal{M}_0)$$

is a convergent series in \mathbb{L}^1 . Denote by $S_n = \sum_{i=1}^n X_i$ the partial sums. Then the sequence $\mathbb{E}(X_0^2 + 2X_0 S_n | \mathcal{I})$ converges in \mathbb{L}^1 to some r.v. η and, conditionally on the tail σ -algebra \mathcal{I} , the process $D_n(t)/\sqrt{n}$ converges to the Brownian motion ηW_t .

Remark. This result provides a FCLT with a limit process that is not Gaussian in general. If the sequence is ergodic, a standard Donsker theorem holds. Indeed, the ergodicity assumption implies that the r.v. η is almost surely constant. Hall and Heyde (1980) give this theorem under a more restrictive \mathbb{L}^2 assumption: both series

$$\sum_{n=0}^{\infty} \mathbb{E}(X_n | \mathcal{M}_0) \quad \text{and} \quad \sum_{n=0}^{\infty} (X_n - \mathbb{E}(X_n | \mathcal{M}_0))$$

converge in \mathbb{L}^2 .

Theorem 5.1 under the condition (DMR) stated in Section 5.1.1 can also be derived as a corollary of Theorem 5.2.

The following corollary can be deduced from Theorem 5.2. Consider a stationary Markov sequence $(\xi_n)_{n \in \mathbb{Z}}$ with stationary distribution μ and transition operator P . Let $X_n = g(\xi_n)$ be a centered at expectation, nonlinear functional of $(\xi_n)_{n \in \mathbb{Z}}$. Then the FCLT holds under the condition of convergence of the series $\sum_{n=0}^\infty gP^n g$ in $\mathbb{L}^1(\mu)$.

5.2. The Empirical Cumulative Distribution

Let us consider a stationary sequence $(X_n)_{n \in \mathbb{Z}}$. We assume without loss of generality that the marginal distribution of this sequence is uniform on $[0, 1]$. The cumulative distribution of the empirical process, E_n , of the sequence $(X_n)_{n \in \mathbb{Z}}$ at time n is defined as $(1/n)E_n(x)$, where

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

We consider the following convergence result in the Skohorod space $\mathcal{D}(\mathbb{R})$ when the sample size n converges to infinity:

$$\frac{1}{\sqrt{n}} E_n(x) \xrightarrow[n \rightarrow \infty]{\mathcal{D}(\mathbb{R})} \bar{B}(x).$$

Here $(\bar{B}(x))_{x \in \mathbb{R}}$ is the dependent analogue of a Brownian bridge; that is, \bar{B} denotes the centered Gaussian process with covariance function given by

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

Note that for independent sequences with a marginal cumulative distribution function F , this just means that $\bar{B}(x) = B(F(x))$ for some standard Brownian bridge B ; this justifies the name generalized Brownian bridge.

THEOREM 5.3. *The following functional convergence holds in the Skohorod space of real-valued càdlàg functions on the real line, $\mathcal{D}(\mathbb{R})$, under the weak dependence conditions detailed in the next sections:*

$$\frac{1}{\sqrt{n}} E_n(x) \xrightarrow[n \rightarrow \infty]{\mathcal{D}(\mathbb{R})} \bar{B}(x).$$

The preceding covariance function can be rewritten as

$$\Gamma(x, y) = \text{Cov}(\bar{B}(x), \bar{B}(y)) = \Gamma_{iid}(x, y) + \sum_{k \neq 0} T_k(x, y).$$

For the i.i.d. case, it is equivalent to $\Gamma(x, y) = \Gamma_{iid}(x, y) = F(x) \wedge F(y) - F(x)F(y)$; as the supremum of two regular functions, this term is intrinsically singular on the diagonal $x = y$. This is no longer the case for the other terms $T_k(x, y) = \text{Cov}(\mathbb{1}_{\{X_0 \leq x\}}, \mathbb{1}_{\{X_k \leq y\}})$. If, for instance, the second-order marginals

(X_0, X_k) have a continuous joint density, then $T_k(x, y)$ is a C^2 -function. It is well known that the regularity properties of a Gaussian process are determined by those of its covariance function. Hence, the distortion of the Brownian bridge due to this series is not very important. Either finite-dimensional or empirical functional convergence (EFCLT) is known to hold in the following cases.

5.2.1. Strong Mixing Case (α -Mixing). The condition $\sum_{n=0}^{\infty} \alpha_n < \infty$ implies finite-dimensional convergence. The EFCLT holds if, for some $a > 1$,

$$\alpha_n = \mathcal{O}(n^{-a}).$$

This result, proved by Rio (2000), improves on the previous conditions $a > 1 + \sqrt{2}$ formerly given by Shao and Yu (1996) and on the condition $a > 3$ from Yoshihara (1973). This condition is close to the previous summability, necessary to ensure finite-dimensional convergence.

5.2.2. Absolute Regularity Condition (β -Mixing). The condition $\sum_{n=0}^{\infty} \beta_n < \infty$ implies finite-dimensional convergence. Doukhan, Massart, and Rio (1995) obtain the EFCLT Theorem 5.3 when $\beta_n = \mathcal{O}(n^{-1}(\log n)^{-a})$, for some $a > 2$. Here tightness obtains with an additional loss term of order $(\log^2 n)$. Finally, Rio (2000) obtains the simple (and optimal) sufficient condition for EFCLT

$$\sum_{n=0}^{\infty} \beta_n < \infty.$$

In a previous paper Arcones and Yu (1994) have proved CLTs for empirical processes indexed by so-called V.C. subgraph classes of functions, not necessarily bounded (for more details, see van der Vaart and Wellner, 1996, p. 141). This context contrasts with the common conditions in terms of bracketing numbers. The EFCLT obtains for uniformly bounded classes in the p th mean under the β -mixing condition that

$$n^{p/(p-2)} (\log(n))^{2(p-1)/(p-2)} \beta_n \xrightarrow{n \rightarrow \infty} 0 \quad (p \in]2, \infty[).$$

5.2.3. ρ -Mixing Case. The condition

$$\sum_{n=0}^{\infty} \rho(2^n) < \infty$$

implies finite-dimensional convergence (see Peligrad, 1987). Shao and Yu (1996) obtain the EFCLT under the same condition.

5.2.4. Gaussian Subordinated Case. Let $(Y_n)_{n \in \mathbb{Z}}$ be a standard Gaussian stationary process: $\mathbb{E}Y_0 = 0$, $\mathbb{E}Y_0^2 = 1$. Consider a function Φ with Hermite rank

m such that $\mathbb{E}\Phi^2(Y_0) < \infty$. Csörgő and Mielniczuk (1996) prove the EFCLT for the subordinated field $X_n = \Phi(Y_n)$ under the natural condition

$$\sum_{n=0}^{\infty} |\text{Cov}(Y_0, Y_n)|^m < \infty.$$

Hence even if the indicator functions do not satisfy the additional tightness assumption in Chambers and Slud (1989), the tightness of the empirical process follows from the diagram formula. In this case the Kolmogorov–Smirnov statistics are convergent even if the Donsker theorem does not hold for finite-dimensional distributions of the empirical cumulative process.

5.2.5. Associated Case. We assume that the marginal distribution of X_0 is uniform on an interval $[0,1]$. Then the condition $\sum_{n=0}^{\infty} \text{Cov}^{1/3}(X_0, X_n) < \infty$ implies finite-dimensional convergence. Louhichi (2000) obtains the EFCLT Theorem 5.3 under the condition that, for some $a > 4$,

$$\text{Cov}(X_0, X_n) = \mathcal{O}(n^{-a}).$$

Her result improves on the condition of Shao and Yu (1996): $a > (3 + \sqrt{33})/2$.

5.2.6. One-Sided Linear Processes. The EFCLT Theorem 5.3 holds if, for some $0 < \gamma \leq 1$ and $S, C, \Delta > 0$, with $S\Delta > 2\gamma$, we have

$$\mathbb{E}|\xi_0|^S < \infty, \quad \sum_{t=0}^{\infty} |a_t|^\gamma < \infty \quad \text{and} \quad |\mathbb{E} \exp(iu\xi_0)| \leq \frac{C}{1 + |u|^\Delta}.$$

If the innovations have moments of any order, the existence of some $\Delta > 0$ such that the preceding inequality holds implies the result.

On the other hand, a lower order moment assumption for the innovation allows higher regularity properties. If, for example, $\gamma = 1$ and $S = 2$, the inversion formula shows the existence of a \mathbb{L}^1 -integrable density. If now $\gamma = 1$ and $S = 4$, the density must be \mathbb{L}^2 -integrable.

For $\gamma = 1$, this result recovers the proof of the EFCLT in Doukhan and Surgailis (1998) when the covariance series is absolutely convergent. The latter paper considers the case $S = 4$; however, Burkholder’s inequality for martingales yields the general result in a straightforward way. Giraitis and Surgailis (1994) gave a hint of the available results in a long memory dependence framework.

5.2.7. $(\theta, \mathcal{L}, \psi)$ -Weak Dependence. The sequence $(X_n)_{n \in \mathbb{Z}}$ is assumed to satisfy a weak dependence condition we now present:

$$\sup_{f \in \mathcal{F}} \left| \text{Cov} \left(\prod_{i=1}^2 f(X_{t_i}), \prod_{i=3}^4 f(X_{t_i}) \right) \right| \leq \theta_r, \tag{5.20}$$

where $\mathcal{F} = \{x \rightarrow \mathbb{1}_{s < x \leq t}, \text{ for } s, t \in [0, 1]\}$, $0 \leq t_1 \leq t_2 \leq t_3 \leq t_4$, 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 5.1. *Let $(X_n)_{n \in \mathbb{Z}}$ be a stationary sequence such that (5.20) holds with*

$$\theta_r = \mathcal{O}(r^{-5/2-\nu}), \tag{5.21}$$

for some $\nu > 0$. Then the sequence of processes $((1/\sqrt{n})\{E_n(t); t \in [0, 1]\})_{n > 0}$ is tight in the Skorohod space $\mathcal{D}([0, 1])$.

In the same way, stationary mixing sequences satisfy the conditions of Proposition 5.1 if $\alpha_r = \mathcal{O}(r^{-5/2-\nu})$. This condition is slightly sharper than Yoshihara’s condition, $\alpha_r = \mathcal{O}(r^{-3-\nu})$ for some $\nu > 0$. Yet, it is slightly sharper than the corresponding result in Shao and Yu (see Theorem 2.2 therein), and the result of Rio (2000) improves on both of them.

THEOREM 5.4. *Suppose that $(X_n)_{n \in \mathbb{Z}}$ is $(\theta, \mathcal{L}_1, \psi_j)$ -weak dependent. If either $j = 1$ and $\theta_r = \mathcal{O}(r^{-15/2-\nu})$ or $j = 2$ and $\theta_r = \mathcal{O}(r^{-5-\nu})$, then the empirical functional convergence holds.*

Remark. The use of the space \mathcal{L} allows one to work with each of the classes of models in the previous section (association and Gaussian sequences enter the first case, whereas the second one corresponds to Bernoulli shifts). This yields new results for Bernoulli shifts and apparently for Markov sequences. Note that Rio’s condition for α -mixing sequences improves this result. Moreover, Yu (1993) proves the same result for associated sequences with an exponent loss term 1.5.

6. FUNCTIONAL ESTIMATION

We consider a stationary process $(Z_t)_{t \in \mathbb{Z}}$ with $Z_t = (X_t, Y_t)$ where $X_t, Y_t \in \mathbb{R}$. The quantity of interest is the regression function $r(x) = \mathbb{E}(Y_0 | X_0 = x)$. Let K be some kernel function integrated to 1, Lipschitzian with compact support. The kernel estimators are defined by

$$\hat{f}(x) = \hat{f}_{n,h}(x) = \frac{1}{nh} \sum_{t=1}^n K\left(\frac{x - X_t}{h}\right),$$

$$\hat{g}(x) = \hat{g}_{n,h}(x) = \frac{1}{nh} \sum_{t=1}^n Y_t K\left(\frac{x - X_t}{h}\right),$$

$$\hat{r}(x) = \hat{r}_{n,h}(x) = \frac{\hat{g}_{n,h}(x)}{\hat{f}_{n,h}(x)} \quad \text{if } \hat{f}_{n,h}(x) \neq 0; \quad \hat{r}(x) = 0 \text{ otherwise.}$$

Here $h = (h_n)_{n \in \mathbb{N}}$ is a sequence of positive real numbers. We always assume that $h_n \rightarrow 0, nh_n \rightarrow \infty$ as $n \rightarrow \infty$.

DEFINITION 6.1. Let $\rho = a + b$ with $(a, b) \in \mathbb{N} \times]0; 1]$. Define the set of ρ -regular functions \mathcal{C}_ρ by

$$\mathcal{C}_\rho = \{u : \mathbb{R} \rightarrow \mathbb{R}; u \in \mathcal{C}_a \text{ and } \exists A \geq 0 |u^{(a)}(x) - u^{(a)}(y)| \leq A|x - y|^b \text{ for all } x, y \text{ in any compact subset}\}.$$

Here, \mathcal{C}_a is the set of a -times continuously differentiable functions.

Assuming $g \in \mathcal{C}_\rho$, one can choose a kernel function K of order ρ (not necessarily nonnegative) such that the bias b_h satisfies

$$b_h(x) = \mathbb{E}(\hat{g}(x)) - g(x) = \mathcal{O}(h^\rho)$$

uniformly on any compact subset of \mathbb{R} (see Rosenblatt, 1991). If, moreover, ρ is an integer with $b = 1, \rho = a - 1$, then with an appropriately chosen kernel K of order $\rho, b_h(x) \sim (g^{(\rho)}(x)/\rho!)h^\rho \int s^\rho K(s) ds$, uniformly on any compact interval.

In view of the asymptotic analysis we assume that the marginal density $f(\cdot)$ of X_0 exists and is continuous. Moreover, $f(x) > 0$ for any point x of interest, and the regression function $r(\cdot) = \mathbb{E}(Y_0 | X_0 = \cdot)$ exists and is continuous. Finally, for some $p \geq 1, x \rightarrow g_p(x) = f(x)\mathbb{E}(|Y_0|^p | X_0 = x)$ exists and is continuous. We set $g = fr$ with obvious shorthand notation. Moreover, we impose one of the following moment conditions. Either

$$\mathbb{E}|Y_0|^S < \infty, \text{ for some } S \geq p \tag{6.22}$$

or

$$\mathbb{E} \exp(|Y_0|) < \infty. \tag{6.23}$$

6.1. Second-Order Properties

We consider first the properties of $\hat{g}(x)$. We also consider the following conditionally centered equivalent of g_2 appearing in the asymptotic variance of the estimator \hat{r} ,

$$G_2(x) = f(x) \text{Var}(Y_0 | X_0 = x) = g_2(x) - f(x)r^2(x).$$

Assume that the densities of the pairs $(X_0, X_k), k \in \mathbb{Z}^+$, exist and are uniformly bounded: $\sup_{k>0} \|f_{(k)}\|_\infty < \infty$. Moreover, uniformly over all $k \in \mathbb{Z}^+$, the functions

$$r_{(k)}(x, x') = \mathbb{E}(|Y_0 Y_k| | X_0 = x, X_k = x') \tag{6.24}$$

are continuous. Under these assumptions, the functions $g_{(k)} = f_{(k)} r_{(k)}$ are locally bounded.

THEOREM 6.1. *Suppose that the stationary sequence $(Z_t)_{t \in \mathbb{Z}}$ satisfies the conditions (6.23) and (6.24) with $p = 2$. Suppose that $n^\delta h \rightarrow \infty$ for some $\delta \in]0, 1[$. Then*

$$\sqrt{nh}(\hat{g}(x) - \mathbb{E}\hat{g}(x)) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(0, g_2(x) \int K^2(u) du\right)$$

and

$$\sqrt{nh}([\hat{g}(x) - r(x)\hat{f}(x)] - \mathbb{E}[\hat{g}(x) - r(x)\hat{f}(x)]) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(0, G_2(x) \int K^2(u) du\right)$$

under any of the weak dependence condition formulated subsequently.

To consider asymptotics for the ratio estimator \hat{r} we use a method, already used by Collomb (1984), that consists of studying higher order asymptotics. It is the topic of the next section.

THEOREM 6.2. *Suppose that the stationary sequence $(Z_t)_{t \in \mathbb{Z}}$ satisfies the conditions (6.23) and (6.24) with $p = 2$. Consider a positive kernel K . Let $f, g \in C_\rho$ for some $\rho \in]0, 2]$, and $nh^{1+2\rho} \rightarrow 0$. Then, for all x belonging to any compact subset of \mathbb{R} ,*

$$\sqrt{nh}(\hat{r}(x) - r(x)) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{G_2(x)}{f^2(x)} \int K^2(u) du\right)$$

under any of the weak dependence conditions formulated subsequently.

6.1.1. Strong Mixing Case (α -Mixing). Theorems 6.1 and 6.2 hold if $h_n \rightarrow 0$, $nh_n/\log(n) \rightarrow \infty$, and if $\alpha_n = \mathcal{O}(n^{-a})$ for some $a > \max\{6\delta, 2 + 2\delta\}$. The proof is based on a Bernstein grouping argument. Besides, Robinson (1983) proves the CLT result in Theorem 5.2 under condition (5.5) for $a > 2S/(S - 2)$ without the assumption of positivity of the kernel K .

6.1.2. ρ -Mixing Case. The estimators in Theorem 6.1 obey the CLT under the mixing assumption $\sum_{i=1}^{\infty} \rho(2^i) < \infty$ and if the bandwidth condition $h_n \rightarrow 0$, $nh_n/\log^2(n) \rightarrow \infty$ is fulfilled (see Peligrad, 1995).

The latter CLT Theorem 6.2 holds if, for some $\gamma > 0$, $\sum_{i=1}^{\infty} (\rho(2^i))^{2/(2+\gamma)} < \infty$ and if the bandwidth condition $h_n \rightarrow 0$, $nh_n/\log^2(n) \rightarrow \infty$, holds. The proof is based on a triangular CLT in Peligrad (1996) combined with moment inequalities B.1 from Shao (1995).

6.1.3. $(\theta, \mathcal{L}, \psi)$ -Weak Dependence. Assuming that the sequence $(Z_t)_{t \in \mathbb{Z}}$ is $(\theta, \mathcal{L}, \psi'_j)$ -weak dependent with $\theta_r = \mathcal{O}(r^{-a})$ and $a > 2 + j$, for $j = 1$ or $j = 2$, Ango Nze et al. (2000) prove that, uniformly in x belonging to any compact subset of \mathbb{R} ,

$$\text{Var}(\hat{g}(x)) = \frac{1}{nh} g_2(x) \int K^2(u) du + o\left(\frac{1}{nh}\right)$$

and

$$\text{Var}(\hat{g}(x) - r(x)\hat{f}(x)) = \frac{1}{nh} G_2(x) \int K^2(u) du + o\left(\frac{1}{nh}\right).$$

The exponential moment assumption can be relaxed. Suppose that the stationary sequence $(Z_t)_{t \in \mathbb{Z}}$ satisfies conditions (6.22) and (6.24) with $p = 2, S > 2$. The former results then hold if the sequence $(Z_t)_{t \in \mathbb{Z}}$ is $(\theta, \mathcal{L}, \psi_j)$ -weak dependent with $\theta_r = \mathcal{O}(r^{-a})$ and $a > [(2 + j)S - 4]/(S - 2) + [2/\delta(S - 2)]$, for $j = 1$ or $j = 2$.

The CLT convergence Theorem 6.1 holds, under the conditions (6.23) and (6.24) with $p = 2$ if the stationary sequence $(Z_t)_{t \in \mathbb{Z}}$ is $(\theta, \mathcal{L}, \psi_j)$ -weak dependent with $\theta_r = \mathcal{O}(r^{-a})$ and

$$a > \alpha_j(\delta) = \min\left(\max(2 + j, 3(2 + j)\delta), \max\left(2 + j + \frac{1}{\delta}, \frac{2 + 2(2 + j)\delta}{1 + \delta}\right)\right),$$

for $j = 1$ or $j = 2$. These results extend the results of Doukhan and Louhichi (1999), valid for the case of the density function \hat{f} , to the estimate \hat{g} under weak dependence with either ψ_1 or ψ_2 . Indeed, the first right-hand-side term is obtained by Bernstein’s blocking technique described in Appendix B. The second right-hand-side term results from the application of the Lindeberg method (see Rio, 2000).

The CLT convergence Theorem 6.2 relies on the expansion

$$u^{-1} = \sum_{i=0}^p (-1)^i \frac{(u - u_0)^i}{u_0^{i+1}} + (-1)^{p+1} \frac{(u - u_0)^{p+1}}{uu_0^{p+1}}, \tag{6.25}$$

where $p = 2, u = b_n, u_0 = \mathbb{E}b_n = 1$, and $\hat{r}(x) = a_n/b_n$ (if $b_n \neq 0$) with

$$a_n = \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right) / \left(n \mathbb{E}K\left(\frac{x - X_0}{h_n}\right)\right) \quad \text{and}$$

$$b_n = \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) / \left(n \mathbb{E}K\left(\frac{x - X_0}{h_n}\right)\right).$$

Using the Rosenthal inequalities described in Appendix B and the aforementioned CLT, we obtain the CLT convergence Theorem 6.2 for the regression function, under conditions (6.23) and (6.24) with $p = 2$, if the stationary sequence $(Z_t)_{t \in \mathbb{Z}}$ is $(\theta, \mathcal{L}, \psi_j)$ -weak dependent with $\theta_r = \mathcal{O}(r^{-a})$,

$$a > \alpha_j(\delta) \quad \text{and} \quad a > \max\left(3, \frac{9(2 + j)}{7 - 4\delta}\right),$$

for $j = 1$ or $j = 2$.

The results stated in Theorem 6.1 and Theorem 6.2 also hold for finite-dimensional convergence. The components are asymptotically jointly independent, much in the same way as for i.i.d. sequences.

Moreover, Rios (1996) proves that the local linearity test can be handled in the strong mixing case. The function r is assumed to be ρ continuous. Then the plug-in estimator $\hat{T} = \int \hat{r}^2(x)w(x) dx$ converges to $T = \int r^2(x)w(x) dx$ if $\alpha_n = \mathcal{O}(n^{-a})$ and $a > 2 + 3/\rho$ and the bandwidth condition $h_n \in [n^{-1/10}, n^{-1/(2\rho-4)}]$.

6.2. Almost Sure Convergence Properties

THEOREM 6.3. *Let $(Z_t)_{t \in \mathbb{Z}}$ be a stationary sequence satisfying the conditions (6.23) and (6.24) with $p = 2$. Then under the conditions formulated in the sections that follow,*

- (i) *There exists a sequence $(\epsilon_n)_{n \in \mathbb{N}}$ with $nh/(\epsilon_n \log(n)) \rightarrow \infty$ as $n \rightarrow \infty$ such that for any $M > 0$,*

$$\sup_{|x| \leq M} |\hat{g}(x) - \mathbb{E}\hat{g}(x)| = \mathcal{O}\left(\sqrt{\frac{\epsilon_n \log(n)}{nh}}\right) \text{ almost surely.}$$

- (ii) *Assume now that $\inf_{|x| \leq M} f(x) > 0$. If $f, g \in \mathcal{C}_\rho$ for some $\rho \in]0, \infty[$, $h \sim (\epsilon_n \log(n)/n)^{1/(1+2\rho)}$, then*

$$\sup_{|x| \leq M} |\hat{f}(x) - r(x)| = \mathcal{O}\left\{\left(\frac{\epsilon_n \log(n)}{n}\right)^{\rho/(1+2\rho)}\right\} \text{ almost surely.}$$

Remark. Under condition (ii) of Theorem 6.3, but assuming only the weaker condition about the bandwidth sequence

$$n^\delta h \rightarrow \infty \text{ for some } \delta \in]0, 1[,$$

we obtain

$$\sup_{|x| \leq M} |\hat{f}(x) - r(x)| = o(1) \text{ almost surely.}$$

6.2.1. Strong Mixing Case (α -Mixing). Liebscher (1996) proves the uniform almost sure convergence at the optimal rate ($\epsilon_n = 1$) if $\alpha_r = \mathcal{O}(r^{-a})$, with $a > 4 + 3/\rho$.

6.2.2. ρ -Mixing Case. Peligrad (1991) states a uniform almost sure convergence result with $\epsilon_n = \log(n)$ if $\rho_k = \mathcal{O}(k^{-r})$, with $r > (\rho + 1)/2\rho$.

6.2.3. $(\theta, \mathcal{L}, \psi_1)$ -Weak Dependence Case. For the sake of simplicity, we only consider the geometrically dependent case.

THEOREM 6.4. *Let $(Z_t)_{t \in \mathbb{Z}}$ be a stationary sequence satisfying the conditions (6.23) and (6.24) with $p = 2$ and either $(\theta, \mathcal{L}, \psi_1)$ - or $(\theta, \mathcal{L}, \psi_2)$ -weak dependent with $\theta_r \leq a^r$ for some $0 < a < 1$.*

(i) If $nh/\log^4(n) \rightarrow \infty$, then for any $M > 0$, almost surely,

$$\sup_{|x| \leq M} |\hat{g}(x) - \mathbb{E}\hat{g}(x)| = \mathcal{O}\left(\frac{\log^2(n)}{\sqrt{nh}}\right).$$

(ii) For any $M > 0$, if $f, g \in \mathcal{C}_\rho$ for some $\rho \in]0, \infty[$, $h \sim (\log^4(n)/n)^{1/(1+2\rho)}$ and $\inf_{|x| \leq M} f(x) > 0$, then, almost surely,

$$\sup_{|x| \leq M} |\hat{r}(x) - r(x)| = \mathcal{O}\left\{\left(\frac{\log^4(n)}{n}\right)^{\rho/(1+2\rho)}\right\}.$$

7. CONCLUSION

This paper presented the new weak dependence condition of Doukhan and Louhichi (1999). It is related to some of the most popular conditions used by econometricians to transcribe the notion of fading memory. The new dependence condition has the advantage that it allows consideration of a broader class of models. This natural weak dependence condition also fits well with the more general (stationary) models used in econometrics. As we have illustrated, most applications of interest can be set out under this weaker dependence condition. Moreover, our framework is a very natural one for bootstrapping techniques. We have also provided several useful limit theorems.

NOTES

1. The function h takes its values in $L_x \times 1$ space equipped with a norm $\|\cdot\|$. The assumption is missing the symbol of mathematical expectation \mathbb{E} , as in Andrews (2002).

2. Here θ_r is a dependence coefficient, and it is not related to a statistical parameter denoted θ and estimated by $\hat{\theta}_n$.

3. Thanks to an anonymous referee, we prove that NED implies our weak dependence through the following inequalities. For simplicity, write $h = h(X_{i_1}, \dots, X_{i_u})$, $k = k(X_{j_1}, \dots, X_{j_v})$, then the Cauchy-Schwarz inequality gives

$$|\text{Cov}(h(X_{i_1}, \dots, X_{i_u}), k(X_{j_1}, \dots, X_{j_v}))| = |\mathbb{E}(h(k - \mathbb{E}(k|\mathcal{F}_{i_u})))| \leq \sqrt{\mathbb{E}h^2 \mathbb{E}(k - \mathbb{E}(k|\mathcal{F}_{i_u}))^2}$$

and the last expression can be bounded using the \mathbb{L}_2 -mixingale property of \mathbb{L}_2 -NED sequences. Clearly, this implication is not an equivalence between both notions. It is an open question whether or not these notions are equivalent.

4. Further technical details on this topic are provided by Granger and Andersen (1978), among other references.

5. Its stationary representation writes $X_n = \sum_{k=0}^\infty 2^{-k-1} \xi_{n-k}$. Here ξ_{n-k} is the k th digit in the binary expansion of the uniformly chosen number $X_n = 0.\xi_n \xi_{n-1} \dots \in [0, 1]$. This proves that X_n is a deterministic function of X_0 , which is the main argument to derive that such models are not mixing. The same arguments apply to the model described before of an autoregressive process with innovations taking p distinct values.

6. If $a_j = \mathcal{O}(j^{-a})$, then, under the preceding regularity and moment conditions, we have $\beta_n \sim n^{-b}$, where $b = ((a - 2)\delta - 1)/(1 + \delta)$ and $a > 2 + 1/\delta$. Therefore, $\sum_{n=0}^\infty \beta_n < \infty$ holds if

$a > 3 + 2/\delta$. If, for instance, $\delta = 1$, this becomes $a > 5$; on the other hand, when $\delta = \infty$, this becomes $a > 3$.

REFERENCES

- Alsmeyer, G. & C.D. Fuh (2001) Limit theorems for iterated random functions by regenerative methods. *Stochastic Processes and Their Applications* 96, 123–142.
- Andrews, D.W.K. (1984) Non strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–934.
- Andrews, D.W.K. (1988) Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* 4, 458–467.
- Andrews, D.W.K. (2002) Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica* 70, 119–162.
- Andrews, D.W.K. & D. Pollard (1994) An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review* 62, 119–132.
- Ango Nze, P. (1995) *Critères d'ergodicité de modèles markoviens: Estimation non paramétrique sous des hypothèses de dépendance*. Ph.D. Dissertation, Université Dauphine, Paris.
- Ango Nze, P. (1998) Critères d'ergodicité géométrique ou arithmétique de modèles linéaires perturbés à représentation markovienne. *Comptes Rendus de l'Académie des Sciences, Serie I, Paris t.* 326, 371–376.
- Ango Nze, P., P. Bühlmann, & P. Doukhan (2002) Weak dependence beyond mixing and asymptotics for nonparametric regression. *Annals of Statistics* 30, 397–430.
- Arcones, M.A. & B. Yu (1994) Central limit theorems for empirical and u-processes of stationary sequences. *Journal of Theoretical Probability* 7, 47–71.
- Berbee, H.C.P. (1979) *Random Walks with Stationary Increments and Renewal Theory*. Mathematical centre tracts, Amsterdam.
- Bickel, P. & P. Bühlmann (1999) A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* 5, 413–446.
- Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bradley, R.C. & W. Bryc (1985) Multilinear forms and measures of dependence between random variables. *Journal of Multivariate Analysis* 16, 335–367.
- Breuer, P. & P. Major (1983) Central limit theorems for non-linear functionals of Gaussian random fields. *Journal of Multivariate Analysis* 13, 425–441.
- Chambers, D. & D. Slud (1989) Central limit theorems for non-linear functionals of stationary Gaussian processes. *Probability Theory and Related Fields* 80, 323–346.
- Chen, X. & Y. Fan (1999) Consistent hypotheses testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* 91, 373–401.
- Cline, D. & H. Pu (1998) Verifying irreducibility and continuity of a nonlinear time series. *Statistics and Probability Letters* 40(2), 139–148.
- Collomb, G. (1984) Propriétés de convergence presque complète du prédicteur à noyau. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 66, 441–460.
- Coulon-Prieur, C. & P. Doukhan (2000) A triangular CLT for weakly dependent sequences. *Statistics and Probability Letters* 47, 61–68.
- Csörgő, M. & J. Mielniczuk (1996) The empirical process of a short-range dependent stationary sequence under Gaussian subordination. *Probability Theory and Related Fields* 104, 15–25.
- Davydov, Y. (1973) Mixing conditions for Markov chains. *Theory of Probability and Its Applications* 18, 313–328.
- Dedecker, J. & E. Rio (2000) On the functional central limit theorem for stationary processes. *Annales de l'institut Henri Poincaré, série B* 36, 1–34.
- Diaconis, P. & D. Freedman (1999) Iterated random functions. *SIAM Review* 41, 45–76
- Doukhan, P. (1994) *Mixing: Properties and Examples*. Lecture Notes in Statistics 85. Springer Verlag.

- Doukhan, P. (2002) Models, inequalities and limit theorems for stationary sequences. In P. Doukhan, G. Oppenheim, & M. Taqqu (eds.), *Long Range Dependence, Theory and Applications*. Birkhäuser.
- Doukhan, P., P. Massart, & E. Rio (1995) Invariance principles for absolutely regular empirical processes. *Annales de l'institut Henri Poincaré, série B* 31, 393–427.
- Doukhan, P. & S. Louhichi (1999) A new weak dependence condition and application to moment inequalities. *Stochastic Processes and Their Applications* 84, 313–343.
- Doukhan, P. & D. Surgailis (1998) Functional central limit theorem for the empirical process of short memory linear processes. *Comptes Rendus de l'Académie des Sciences, Paris* 326, 87–92.
- Dufflo, M. (1990) *Méthodes récursives aléatoires*. Masson. (English edition, Springer, 1996).
- Esary, J., F. Proschan, & D. Walkup (1967) Association of random variables with applications. *Annals of Mathematical Statistics* 38, 1466–1476.
- Fort, G. & E. Moulines (2002) Computable bounds for polynomial ergodicity. *Stochastic Processes and Their Applications*. Forthcoming.
- Fortuin, C., P. Kastelyn, & J. Ginibre (1971) Correlation inequalities on some ordered sets. *Communications in Mathematical Physics* 22, 89–103.
- Giraitis, L., P. Kokoszka, & R. Leipus (2000) Stationary ARCH models: Dependence structure and CLT. *Econometric Theory* 16, 3–22.
- Giraitis, L. & D. Surgailis (1986) Multivariate Appell polynomials and the central limit theorem. In E. Eberlein & M. Taqqu (eds.), *Dependence in Probability and Statistics*, pp. 27–51. Birkhäuser.
- Giraitis, L. & D. Surgailis (1994) A Central Limit Theorem for the Empirical Process of a Long Memory Linear Sequence. Preprint.
- Granger, C.W.J. & A.P. Andersen (1978) *An Introduction to Bilinear Time Series Models*. Vandenhoeck and Ruprecht.
- Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Applications*. Academic Press.
- Hall, P. & J. Horowitz (1996) Bootstrap critical values for tests based on generalized method of moment estimators. *Econometrica* 64, 891–916.
- He, C. & T. Teräsvirta (1999) Properties of moments of a family of GARCH models. *Journal of Econometrics* 92, 173–192.
- Ho, H.-C. & T. Hsing (1997) Limit theorems for functionals of moving averages. *Annals of Probability* 25, 1636–1669.
- Ibragimov, I.A. (1962) Some limit theorems for stationary processes. *Theory of Probability and Its Applications* 7, 349–382.
- Ibragimov, I.A. & Y.V. Linnik (1971) Independent and stationary sequences of random variables. Wolters-Noordhoff.
- Ibragimov, R. & S. Sharakhmentov (2001) The exact constant in the Rosenthal inequality for random variables with mean zero. *Theory of Probability and Its Applications* 46, 127–132.
- Jarner, S.F. & G.O. Roberts (2002) Polynomial convergence rates of Markov chains. *Annals of Applied Probability* 12, 224–247.
- Jarner, S.F. & R.L. Tweedie (2001) Locally contacting iterated functions and stability of Markov chains. *Journal of Applied Probability* 38, 494–507.
- Kallenberg, O. (1997) *Foundations of Modern Probability*. Springer-Verlag.
- Kolmogorov, A.N. & Y.A. Rozanov (1960) On the strong mixing conditions for stationary Gaussian sequences. *Theory of Probability and Its Applications* 5, 204–207.
- Künsch, H.R. (1989) The jackknife and bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
- Liebscher, E. (1996) Strong convergence of sums of α -mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications* 65, 69–98.
- Lin, Z.Y. & C.R. Lu (1996) *Limit Theory for Mixing Dependent Random Variables*. Kluwer Academic Publishers.
- Louhichi, S. (2000) Weak convergence for empirical processes of associated sequences. *Annales de l'Institut Henri Poincaré, série B* 36, 547–567.
- McLeish, D.L. (1975) A maximal inequality and dependent strong laws. *Annals of Probability* 3, 829–839.

- Meyn, S.P. & R.L. Tweedie (1993) Markov chains and stochastic stability. *Communications in Control and Engineering*.
- Mills, T.C. (1999) *The Econometric Modelling of Financial Time Series*. Cambridge University Press.
- Mokkadem, A. (1990) Propriétés de mélange des modèles autorégressifs polynômes. *Annales de l'Institut Henri Poincaré, série B* 26, 219–260.
- Moricz, F.A., R.J. Serfling, & W.F. Stout (1982) Moment and probability bounds with quasi-superadditive structure for the maximum partial sum. *Annals of Probability* 10, 1032–1040.
- Newman, C.M. (1984) Asymptotic independence and limit theorems for positively and negatively dependent random variables. In Y.L. Tong (ed.), *Inequalities in Statistics and Probability, IMS Hayward, CA Lecture Notes-Monograph Series* 5, pp. 127–140.
- Newman, C.M. & A.L. Wright (1981) An invariance principle for certain dependent sequences. *Annals of Probability* 9, 671–675.
- Ornstein, D.S. & G. Weiss (1990) Statistical properties of chaotic systems. *Bulletin of the American Mathematical Society* 24, 11–116.
- Peligrad, M. (1987) The convergence of moments in ρ -mixing sequences of random variables. *Proceedings of the American Mathematical Society* 101, 142–148.
- Peligrad, M. (1991) Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence assumptions. *Stochastics and Stochastic Reports* 40, 147–168.
- Peligrad, M. (1996) On the asymptotic normality of weak dependent random variables. *Journal of Theoretical Probabilities* 9, 703–715.
- de la Peña, V.H., R. Ibragimov, & S. Sharakhmetov (2003) On extremal distributions and sharp L_p -bounds for sums of multilinear forms. *Annals of Probability* 31, 630–675.
- Petrucci, J. & S.W. Woolford (1984) A threshold AR(1) model. *Journal of Applied Probability* 21, 270–286.
- Pham, T.D. (1986) The mixing property of bilinear and generalized random coefficient autoregressive models. *Stochastic Processes and Their Applications* 23, 291–300.
- Pham, T.D. & L.T. Tran (1985) Some mixing properties of time series models. *Stochastic Processes and Their Applications* 19, 297–303.
- Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.
- Pollard, D. (1981) Limit theorems for empirical processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 57, 181–195.
- Pötscher, B.M. & I.R. Prucha (1991a) Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part I: Consistency and approximation concepts. *Econometric Reviews* 10, 125–216.
- Pötscher, B.M. & I.R. Prucha (1991b) Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part II: Asymptotic normality. *Econometric Reviews* 10, 253–326.
- Prakasa, R. (1983) *Nonparametric Functional Estimation*. Academic Press.
- Rio, E. (1994) Inégalités de moments pour les suites stationnaires et fortement mélangées. *Comptes Rendus de l'Académie des Sciences, Paris* 318, 355–360.
- Rio, E. (2000) Théorie asymptotique des processus aléatoires faiblement dépendants. *Mathématiques et applications* 31.
- Rios, R. (1996) *Utilisation des techniques non paramétriques et semi paramétriques en statistique des données dépendantes*. Ph.D. Dissertation, Université Paris sud, Paris.
- Robinson, P.M. (1983) Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.
- Robinson, P.M. (1989) Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies* 56, 511–534.
- Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition. *Proceedings of National Academy of Sciences, U.S.A.* 42, 43–47.
- Rosenblatt, M. (1985) *Stationary Sequences and Random Fields*. Birkhäuser.
- Rosenblatt, M. (1991) Stochastic curve estimation. *NSF-CBMS Regional Conference Series in Probability and Statistics* 3.

- Shao, Q.M. (1988) Remark on the invariance principle for ρ -mixing sequences of random variables. *Chinese Annals of Mathematics* 9A, 409–412.
- Shao, Q.M. (1995) Maximal inequalities for partial sums of ρ -mixing sequences. *Annals of Probability* 23, 948–965.
- Shao, Q.M. (2000) A comparison theorem on moment inequalities between negatively associated and independent random variables. *Journal of Theoretical Probability* 13, 343–356.
- Shao, Q.M. & H. Yu (1996) Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability* 24, 2052–2078.
- Tjøstheim, D. (1986) Some doubly stochastic time series models. *Journal of Time Series Analysis* 7, 51–72.
- Tjøstheim, D. (1990) Non-linear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.
- Tong, H. (1981) A note on a Markov bilinear stochastic process in discrete time. *Journal of Time Series Analysis* 2 (4), 279–284.
- van der Vaart, A.W. & J.A. Wellner (1996) Weak convergence and empirical processes. Springer-Verlag.
- Veretennikov, A.Y. (1999) On polynomial mixing and convergence rate for stochastic difference and differential equation. *Theory of Probability and Its Applications* 44, 361–374.
- Withers, C.S. (1981) Central limit theorem for dependent variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 57, 509–534.
- Wolkonski, V.A. & Y.A. Rozanov (1959) Some limit theorems for random functions, Part I. *Theory of Probability and Its Applications* 4, 178–197.
- Wolkonski, V.A. & Y.A. Rozanov (1961) Some limit theorems for random functions, Part II. *Theory of Probability and Its Applications* 6, 186–198.
- Yoshihara, K. (1973) Billingsley's theorems on empirical processes of strong mixing sequences. *Yokohama Mathematical Journal* 23, 1–7.
- Yoshihara, K. (1976) Limiting behaviour of U-statistics for stationary, absolutely regular processes. *Zeitung Wahrscheinlichkeitstheorie und verwandte Gebiete* 35, 237–252.
- Yu, H. (1993) A Glivenko-Cantelli lemma and weak convergence for empirical processes of associated sequences. *Probability Theory and Related Fields* 95, 357–370.
- Zhao, L. & Z. Fang (1985) Strong convergence of kernel estimates of nonparametric regression functions. *Chinese Annals of Mathematics* 6B, 147–155.

APPENDIX A: PROOFS

A.1. Proofs for Section 4: Models.

Proof of Proposition 4.1. Let $A = i_1, \dots, i_u$ and $B = j_1, \dots, j_v$ be two finite subsets of \mathbb{Z} with $i_1 \leq \dots \leq i_u < i_u + r \leq j_1 \leq \dots \leq j_v$. Set $X_n = H(\xi_{n-j}, j \in \mathbb{Z})$ and, for a given integer k , $X_n^{(k)} = H(\xi_{n-j} \mathbb{1}_{|j| < k}, j \in \mathbb{Z})$. Then for any functions $h, k \in \mathcal{L}$ with obvious notations

$$\begin{aligned}
 & \text{Cov}(h(X_n, n \in A), k(X_n, n \in B)) \\
 & \equiv \text{Cov}(h_A(X_n, n \in \mathbb{Z}), k_B(X_n, n \in \mathbb{Z})) \\
 & = \text{Cov}(h_A(X_n, n \in \mathbb{Z}) - h_A(X_n^{(k)}, n \in \mathbb{Z}), k_B(X_n, n \in \mathbb{Z})) \\
 & \quad + \text{Cov}(h_A(X_n^{(k)}, n \in \mathbb{Z}), k_B(X_n, n \in \mathbb{Z}) - k_B(X_n^{(k)}, n \in \mathbb{Z})) \\
 & \quad + \text{Cov}(h_A(X_n^{(k)}, n \in \mathbb{Z}), k_B(X_n^{(k)}, n \in \mathbb{Z})).
 \end{aligned}$$

If $p \leq r/2$, the r.v.s $h_A(X_n^{(k)}, n \in \mathbb{Z})$ and $k_B(X_n^{(k)}, n \in \mathbb{Z})$ are measurable with respect to the independent σ -fields $\sigma(\xi_i: i_1 - p < i < i_u + p)$ and $\sigma(\xi_j: j_1 - p < j < j_v + p)$. Therefore, the last covariance term is null. Besides, set

$$C = |\text{Cov}(h_A(X_n, n \in \mathbb{Z}) - h_A(X_n^{(k)}, n \in \mathbb{Z}), k_B(X_n, n \in \mathbb{Z}))|.$$

Then,

$$C \leq 2\|k\|_\infty \mathbb{E}|h_A(X_n, n \in \mathbb{Z}) - h_A(X_n^{(k)}, n \in \mathbb{Z})| \leq 4u\|k\|_\infty \delta_k \text{Lip}(h)$$

in the one-sided case. The two-sided case is handled similarly. ■

A.2. Proofs for Section 5: FCLT.

Sketch of the Proof of Theorem 5.1 in the α -Mixing Case. With the notations of Rosenthal inequality in Lemma B.10 in Appendix B, S_n/\sqrt{n} converges in distribution to the normal distribution $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 = \lim n^{-1} \text{Var}(S_n)$ is assumed to be positive. The tightness of the process $(D_n(t))$ is derived according to Lemma B.1. Because the sequence (S_n^2/n) is uniformly integrable, there exists a convex function G increasing faster than x at infinity, such that $\mathbb{E}G(S_n^2/n) \leq 1$. By a blocking argument, it follows that (see details in Rio, 2000, Proposition 2)

$$\mathbb{P}(S_n^* \geq (3 + 2\lambda)\sqrt{n}) \leq \frac{p}{G(p)} + p\alpha_{q-1} + \frac{2}{G(\lambda^2)}. \tag{A.1}$$

Here $p = \lceil \sqrt{n} \rceil$ and $q = \lfloor n/p \rfloor$. The tightness condition obtains as soon as $n\alpha_n \rightarrow 0$. ■

Sketch of the Proof of Theorem 5.1 in the ρ -Mixing Case. The proof follows the same lines as in the α -mixing case. Details are developed in the book by Lin and Lu (1996) (see Section 4.1 therein). ■

Sketch of the Proof of Theorem 5.1 in the Weak Dependent Case. Lemma B.12 and a maximal inequality by Moricz, Serfling, and Stout (1982) yield

$$\mathbb{E}|S_n|^{2+\delta} = \mathcal{O}(n^{1+\delta/2}) \tag{A.2}$$

as soon as for any increasing sequence of integers $0 \leq i < j < k \leq l$

$$\sum_{m=0}^\infty m \mathbb{E}|X_0 X_m| < \infty \quad \text{and} \quad \text{Cov}(X_i X_j, X_k X_l) = \mathcal{O}((k - j)^{-2}). \tag{A.3}$$

Moreover, this entails that $\sigma^2 = \lim n^{-1} \text{Var}(S_n) > 0$, so that the finite-dimensional convergence is obtained. The tightness of the process is a consequence of (A.3). The first part of (A.3) follows from the covariance bound $|\text{Cov}(X_0, X_r)| = \mathcal{O}(\theta_r^{(2+\delta)/(4+\delta)})$. The latter bound follows from $\text{Cov}(X_i X_j, X_k X_l) = \mathcal{O}(\theta_{k-j}^{\delta/(4+\delta)})$. ■

A.3. Proofs for Section 5: Empirical Process.

Proof of Proposition 5.1. Using the Rosenthal inequality in Lemma B.12, we get

$$\begin{aligned} \|E_n(t) - E_n(s)\|_4 &\leq C\sqrt{n} \sum_{r=0}^{n-1} \min\{r^{-5/2-\nu}, |t - s|\} + \left(n \sum_{r=0}^{n-1} (r + 1)^2 \theta_r \right)^{1/4}, \\ &\leq \sqrt{n} (|t - s|^{(a-1)/2a} + n^{(2-a)/4}). \end{aligned}$$

The conclusion follows from Lemma 9.4 in Shao and Yu (1996). ■

Proof of Theorem 5.4. Lemma B.13.

A.4. Proofs for Section 6: Functional Estimation.

A.4.1. Second-order Properties.

Proof of Theorem 6.1. We proceed as in Rio (2000) and more specifically as in Coulon-Prieur and Doukhan (2000) for density estimation.

Consider a sequence $(W_n)_{n \in \mathbb{N}}$ of i.i.d. r.v.s with standard normal distribution independent of $((X_t, Y_t)_{t \in \mathbb{Z}}$ and set $M(n) = \log(n)$, $nh\sigma_n^2 = \text{Var}(\sum_{t=1}^n T_t(x))$. Here

$$T_t(x) = (Y_t \mathbb{1}_{\{|Y_t| \leq M(n)\}} + M(n) \mathbb{1}_{\{Y_t > M(n)\}} - M(n) \mathbb{1}_{\{Y_t \leq -M(n)\}}) K\left(\frac{x - X_t}{h}\right) \quad (0 \leq t \leq n) \tag{A.4}$$

denotes the estimator truncated at level $M(n)$ by a Lipschitz continuous function. Define

$$\xi_t = \frac{1}{\sigma_n \sqrt{nh}} (T_t(x) - \mathbb{E}T_t(x)),$$

$$S_k = \sum_{t=1}^k \xi_t, \quad 1 \leq k \leq n, \quad \text{with } S_0 = 0,$$

$$\tau_k = \sum_{t=k}^n V_t, \quad 1 \leq k \leq n, \quad \text{with } \tau_{n+1} = 0,$$

where $v_k/\sigma_n^2 = |\text{Var}(S_k) - \text{Var}(S_{k-1})|$ and $V_k = (\sqrt{v_k}/\sigma_n)W_k$. Let φ denote a three times differentiable function with bounded derivatives up to order 3 and consider the following r.v.s:

$$U_t = S_{t-1} + \tau_{t+1}, \quad R_t(x) = \varphi(U_t + x) - \varphi(U_t) - \frac{v_t}{2\sigma_n^2} \varphi''(U_t) \quad (1 \leq t \leq n).$$

We are interested in establishing that

$$\sigma_n \rightarrow \left(g_2(x) \int K^2(u) du \right)^{1/2}, \tag{A.5}$$

$$\sqrt{nh}(\hat{g}(x) - \mathbb{E}\hat{g}(x)) - \sigma_n S_n \xrightarrow{L^2} 0, \tag{A.6}$$

$$S_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0,1). \tag{A.7}$$

We consider either a $(\theta, \mathcal{L}, \psi_1)$ - or a $(\theta, \mathcal{L}, \psi'_1)$ -weakly dependent sequence $(Z_t)_{t \in \mathbb{Z}}$. We shall follow the practical abuse of notation, in which the same letter C is used for different constants.

Formula (A.6) is easily proved using the exponential moment assumption (6.23).

To prove formula (A.7), we apply the so-called Lindeberg–Rio method (see Rio, 2000). Clearly,

$$|\mathbb{E}\varphi(S_n) - \mathbb{E}\varphi(W_0)| \leq \sum_{t=1}^n |\mathbb{E}R_t(\xi_t)| + \sum_{t=1}^n |\mathbb{E}R_t(V_t)|. \tag{A.8}$$

Because

$$\begin{aligned} |\mathbb{E}R_t(V_t)| &= \left| \mathbb{E} \left(\varphi(U_t + V_t) - \varphi(U_t) - V_t \varphi'(U_t) - \frac{V_t^2}{2} \varphi''(U_t) \right) \right| \\ &\leq \frac{1}{6} \mathbb{E}|V_t^3 \varphi^{(3)}(U_t + \vartheta V_t)|, \quad \text{with } 0 < \vartheta < 1, \\ &\leq \frac{\|\varphi^{(3)}\|_\infty}{6} \left(\frac{v_t}{\sigma_n^2} \right)^{3/2} \mathbb{E}|W_0|^3 \end{aligned}$$

and

$$\begin{aligned} \frac{v_t}{\sigma_n^2} &= |\text{Var}(S_t) - \text{Var}(S_{t-1})| \\ &\leq C \left(\frac{1}{n} + \sum_{j=1}^{t-1} \frac{1}{nh} \min \left(\frac{M^2(n)\theta_{t-j}}{h}, h^2 \right) \right), \end{aligned}$$

we obtain

$$\sum_{t=1}^n |\mathbb{E}R_t(V_t)| \leq \frac{C}{\sigma_n^3} \sum_{t=1}^n v_t^{3/2} \leq \frac{C}{n^{3/2}\sigma_n^3} \sum_{t=1}^n \left(1 + \sum_{j=1}^{t-1} \min \left(\frac{M^2(n)\theta_j}{h^2}, h \right) \right)^{3/2}. \tag{A.9}$$

Moreover,

$$\begin{aligned} R_t(\xi_t) &= \varphi(U_t + \xi_t) - \varphi(U_t) - \frac{v_t}{2\sigma_n^2} \varphi''(U_t) \\ &= \xi_t \varphi'(U_t) + \frac{1}{2} \left(\xi_t^2 - \frac{v_t}{2\sigma_n^2} \right) \varphi''(U_t) + \frac{1}{6} \xi_t^3 \varphi^{(3)}(U_t + \vartheta_t \xi_t), \quad \text{with } 0 < \vartheta_t < 1. \end{aligned}$$

It then follows that

$$\begin{aligned} \sum_{t=1}^n |\mathbb{E}R_t(\xi_t)| &\leq \left| \mathbb{E} \left(\sum_{t=1}^{n-1} \text{Cov}(\xi_0, \xi_t) \sum_{k=t+1}^n \varphi''(U_k) \right) \right| + \frac{1}{2} \left| \sum_{t=1}^n \mathbb{E}(\varphi''(U_t)(\xi_t^2 - \mathbb{E}\xi_t^2)) \right| \\ &\quad + \left| \sum_{t=1}^n \sum_{j=1}^{t-1} \text{Cov}(\varphi''(S_{t-1-j} + \tau_{j+1})\xi_{t-j}, \xi_t) \right| + \left| \sum_{t=1}^n \text{Cov}(\xi_t, \varphi'(\tau_{t+1})) \right| \\ &\quad + \frac{1}{2} \left| \sum_{t=1}^n \sum_{j=1}^{t-1} \text{Cov}(\varphi^{(3)}(S_{t-1-j} + \tau_{j+1} + \vartheta_{t-j}\xi_{t-j})\xi_{t-j}^2, \xi_t) \right| \\ &\quad + \frac{1}{6} \left| \sum_{t=1}^n \mathbb{E}(\varphi^{(3)}(U_t + \iota_t \xi_t)\xi_t^3) \right| = E_1 + E_2 + E_3 + 0 + E_4 + E_5. \end{aligned} \tag{A.10}$$

We can now bound the preceding five terms.

$$E_1 \leq \sum_{t=1}^n |\text{Cov}(\xi_0, \xi_t)| \left| \sum_{k=t+1}^n \mathbb{E}\varphi''(U_t) \right| \leq \frac{C}{nh\sigma_n^2} \sum_{t=1}^n (n-t) \min\left(\frac{\theta_t}{h}, h^2\right). \tag{A.11}$$

In the $(\theta, \mathcal{L}, \psi'_1)$ case,

$$\begin{aligned} E_2 &\leq \frac{1}{2} \sum_{t=2}^n \sum_{j=1}^{t-1} \left| \text{Cov}\left(\varphi^{(3)}(S_{j-1} + \iota_j \xi_j + \tau_{k+1}) \xi_j, \frac{\xi_t^2}{2}\right) \right| + \sum_{t=1}^n \left| \text{Cov}\left(\varphi''(\tau_{k+1}), \frac{\xi_t^2}{2}\right) \right| \\ &\leq \frac{CM(n)}{(\sigma_n \sqrt{nh})^3} \sum_{t=2}^n \sum_{j=1}^{t-1} \min\left(\frac{(t-j)M^2(n)\theta_j}{h}, h^2\right) + \frac{Ch}{\sigma_n^2}, \end{aligned} \tag{A.12}$$

$$E_3 \leq \sum_{t=1}^n \sum_{j=1}^t \frac{C}{nh\sigma_n^2} \min\left(\frac{M^2(n)(t-j)\theta_j}{h}, h^2\right), \tag{A.13}$$

$$E_4 \leq \sum_{t=1}^n \sum_{j=1}^t \frac{C}{(\sigma_n \sqrt{nh})^3} \min\left(\frac{M^3(n)(t-j)\theta_j}{h}, h^2 M(n)\right), \tag{A.14}$$

$$E_5 \leq \frac{C}{(\sigma_n \sqrt{nh})^3} \sum_{t=1}^n h \leq \frac{C}{\sigma_n^3 \sqrt{nh}}. \tag{A.15}$$

For a $(\theta, \mathcal{L}, \psi_1)$ -weak dependent sequence, again using (A.9)–(A.15), we need $\theta_r = O(r^{-a})$ with

$$a > 3 + \frac{1}{\delta}.$$

If the sequence $(Z_t)_{t \in \mathbb{Z}}$ is $(\theta, \mathcal{L}, \psi'_1)$ -weak dependent with $\theta_r = O(r^{-a})$ for some $a > 3$, then by (A.9)–(A.15) the right-hand side term of (A.8) tends to zero as n^{-1} .

The CLT for \hat{g} is now proved. The second assertion is a consequence of the first one, where Y_t is replaced by $Y_t - r(x)$. ■

A.4.2. Almost Sure Convergence.

Proof of Theorem 6.3. We keep usual notations and denote by C a universal constant (whose value can change from one place to another). Assume that $\mathbb{E}(\exp(|Y_0|)) < \infty$. Then

$$\mathbb{P}\left(\sup_{|x| \leq M} |\hat{g} - \tilde{g}|(x) > 0\right) \leq n\mathbb{P}(|Y_0| \geq M_0 \log(n)) \leq Cn^{1-M_0},$$

and, by the Cauchy–Schwarz inequality,

$$\sup_{|x| \leq M} \mathbb{E}(|\hat{g} - \tilde{g}|(x)) \leq \frac{1}{h} \mathbb{E}\left[|Y_0| \mathbb{1}_{\{|Y_0| \geq M_0 \log(n)\}} \left| K\left(\frac{x - X_0}{h}\right) \right|\right] \leq \frac{1}{h} h^{1/3} n^{-M_0}.$$

We can now reduce computations to the case of a density estimator, as in Doukhan and Louhichi (1999). Assume that the interval $[-M, M]$ is covered by L_ν intervals with diameter $1/\nu$ (here $\nu = \nu(n)$ depends on n , and we denote by I_j the j th interval and by x_j the

center of the interval). Assume that the relation $h\nu \rightarrow \infty$ holds (for $n \rightarrow \infty$). Assume that the compactly supported kernel K vanishes if $t > R_0$. Liebscher (1996) exhibits another kernel-type density estimate \tilde{g}' based on an even, continuous kernel, decreasing on $[0, \infty[$, constant on $[0, 2R_0]$, taking the value 0 at $t = 3R_0$. Then, he proves that

$$\sup_{x \in I_j} |\tilde{g} - \mathbb{E}\tilde{g}|(x) \leq |\tilde{g}(x_j) - \mathbb{E}\tilde{g}(x_j)| + \frac{C}{h\nu} (|\tilde{g}' - \mathbb{E}\tilde{g}'|(x_j) + 2|\mathbb{E}\tilde{g}'|(x_j)).$$

Therefore, for any $\lambda > 0$,

$$\begin{aligned} \mathbb{P}\left(\sup_{x \in [-M, M]} |\hat{g}(x) - \mathbb{E}\hat{g}(x)| \geq \frac{2\lambda}{\sqrt{nh}} + \frac{1}{h} h^{1/3} n^{-M_0} + C \frac{\log(n)}{h\nu}\right) \\ \leq n^{1-M_0} + L_\nu \mathbb{P}\left(|\tilde{g} - \mathbb{E}\tilde{g}|(x_1) \geq \frac{\lambda}{\sqrt{nh}}\right) + L_\nu \mathbb{P}\left(|\tilde{g}' - \mathbb{E}\tilde{g}'|(x_1) \geq \frac{\lambda}{\sqrt{nh}}\right). \end{aligned}$$

Exponential inequality (DPL), which can be found in Section B.4.1 of Appendix B, completes the proof of assertion (i). ■

Remark. Zhao and Fang (1985) prove almost sure convergence, uniform on compact sets, of the kernel regression estimator for strongly mixing stationary process under the same condition as in Theorem 6.3. Let us consider a strongly mixing processes that satisfies conditions (6.23) and (6.24) with $p = 2$. Assume that $f, g \in C_\rho$ for some $\rho \in]0, \infty[$ and $n^\delta h \rightarrow \infty$ for some $\delta \in]0, 1[$. If the moment condition (6.22) holds with $S > 4 + 2/\rho$, and if the mixing rate is $\alpha_r \leq r^{-a}$ for some $a > 4$ such that

$$a > \max\left(\frac{2S}{S-2}, \frac{S(S+4)(1+\rho)}{(S-2)(S\rho-4\rho+4)}, \frac{(S+1)(3+4\rho)}{\rho S-2-4\rho}\right),$$

then, almost surely,

$$\sup_{|x| \leq M} |\hat{f}(x) - r(x)| = O((n^{-1} \log(n))^{\rho/(1+2\rho)}).$$

APPENDIX B: TECHNICAL LEMMAS

B.1. Sufficient Conditions for Tightness.

B.1.1. Kolmogorov–Chentsov Criterion. To obtain functional convergence in distribution it is usual to make use of some chaining argument to prove tightness of the sequence $(Y_n(t))_{n \geq 1}$, where $Y_n(t) = D_n(t)/\sqrt{n}$, in the space $\mathcal{C}([0, 1])$. Chaining techniques can be found in Pollard (1981). For the sake of completeness, we recall the following tightness result deduced from the Arzela–Ascoli theorem.

LEMMA B.1 (Billingsley, 1968). *The sequence of processes $(Y_n(t))_{t \in [0, 1]}$ is a tight sequence in the space $\mathcal{C}([0, 1])$ for $n = 1, 2, \dots$ if for each $\epsilon, \eta > 0$, there exists a $\delta > 0$ and an integer n_0 such that for all $t \in [0, 1]$*

$$\frac{1}{\delta} \mathbb{P}\left(\sup_{s \in [t, t+\delta]} |Y_n(s) - Y_n(t)| \geq \epsilon\right) \leq \eta \quad \text{for } n \geq n_0.$$

To conclude the section we recall the standard Kolmogorov–Chentsov tightness criterion, adapted to moment inequalities.

LEMMA B.2 (Billingsley, 1968). *A sequence of continuous processes $(Z_n(t))_{t \in [0,1]}$ is tight in the space $\mathcal{C}([0,1])$ for $n = 1, 2, \dots$ if there exist constants $C > 0$, $p > 0$, $q > 1$, such that for any $s, t \in [0, 1]$:*

$$\mathbb{E}|Z_n(s) - Z_n(t)|^p \leq C|t - s|^q.$$

B.1.2. Tightness of the Empirical Process. We now recall a tightness criterion in the Skohorod space $D([0,1])$, of left continuous with right limits functions on the interval. Note that we can restrict ourselves to the case of marginal uniform distributions.

LEMMA B.3 (Billingsley, 1968). *A sequence of càdlàg (right continuous with left limits) processes $(Z_n(x))_{x \in [0,1]}$ is tight in $D([0,1])$ if*

- For any $\eta > 0$ there exists an $a > 0$ such that

$$\mathbb{P}(\sup_{x \in [0,1]} |Z_n(x)| > a) < \eta \quad \text{for any } n \geq 1.$$

- For any $\epsilon, \eta > 0$ there exists a sequence $0 = x_0 < x_1 < \dots < x_p = 1$ and an integer n_0 such that

$$\mathbb{P}(\max_{0 \leq i < p} \sup_{x \in [x_i, x_{i+1}[} |Z_n(x) - Z_n(x_i)| > \delta) < \epsilon \quad \text{for } n \geq n_0.$$

To conclude the section, we recall a chaining lemma, a proof of which can be found in Shao and Yu (1996); this result can be used to prove tightness of the empirical process when the fourth-order moment involved ($p = 4$) is bounded. The absolutely regular case is an exception for which a more general technique is used by Rio (2000).

LEMMA B.4 (Shao and Yu, 1996). *Let X_n be a stationary sequence with marginal cumulative function F . Then the empirical Brownian bridge $(E_n(x)/\sqrt{n})_{x \in \mathbb{R}}$ is a tight sequence in the Skohorod space $D(\mathbb{R})$ if there exist constants $C > 0$, $p > 2$, $q > 1$, and $u > 0$, $0 \leq v \leq 1$, with $u + v > 1$ such that for any $x, y \in \mathbb{R}$:*

$$\mathbb{E} \left| \frac{1}{\sqrt{n}} (E_n(x) - E_n(y)) \right|^p \leq C \left(|F(x) - F(y)|^q + \frac{|F(x) - F(y)|^v}{n^{u/2}} \right).$$

B.2. Tools under Mixing.

B.2.1. Covariance Inequalities. A fundamental covariance inequality due to Rio (2000) extends on the previously known ones. First, recall the following definition.

DEFINITION 9.1. *The quantile function Q_X of the real valued r.v. X is the càdlàg inverse of the tail function of $|X|$,*

$$Q_X(s) = \inf\{t > 0; \mathbb{P}(|X| > t) \leq s\}.$$

We are now in a position to state the fundamental covariance inequality.

LEMMA B.5 (Rio, 2000). *Let u, v be two real-valued r.v.s with finite variance. Then, setting $\alpha = \alpha(\sigma(u), \sigma(v))$, we have*

$$|\text{Cov}(u, v)| \leq 2 \int_0^{2\alpha} Q_u(t) Q_v(t) dt.$$

Because $|X|^p = \int_0^\infty \mathbb{1}_{\{t < |X|^p\}} dt$, we find by taking expectations that

$$\mathbb{E}|X|^p = \int_0^\infty \mathbb{P}(t < |X|^p) dt = \int_0^\infty \left[\int_0^1 \mathbb{1}_{\{s < \mathbb{P}(t < |X|^p)\}} ds \right] dt.$$

Hence a simple use of the Fubini theorem yields, for $p > 0$, the identity $\mathbb{E}|X|^p = \int_0^1 Q_X^p(s) ds$. Thus, using Hölder's inequality, we deduce

$$|\text{Cov}(u, v)| \leq 2^{1+(1/r)} \alpha^{1/r} \|u\|_p \|v\|_q, \quad \text{if } \frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1 \quad \text{and } p, q, r > 0.$$

Similar covariance inequalities are available for other mixing type sequences. Namely, we can state the following lemmas.

LEMMA B.6 (Bradley and Bryc, 1985). *Let u, v be two real-valued r.v.s. Set $\rho = \rho(\sigma(u), \sigma(v))$. Let $p > 0$ and $q > 0$ be such that $1/p + 1/q = 1$. Then*

$$|\text{Cov}(u, v)| \leq 3,000\rho(1 - \log \rho)^{\min\{(1/p), (1/q)\}} \|u\|_p \|v\|_q.$$

LEMMA B.7 (Ibragimov, 1962). *Let u, v be two real-valued r.v.s. Set $\phi = \phi(\sigma(u), \sigma(v))$. Let $p > 0$ and $q > 0$ be such that $1/p + 1/q = 1$. Then*

$$|\text{Cov}(u, v)| \leq 2\phi^{1/p} \|u\|_p \|v\|_q.$$

B.2.2. Reconstruction Lemmas. Such results exist in the α -mixing and β -mixing cases. The first one is due to Berbee (1979), and the second one is due to Rio (1995). The reader is referred to Rio (2000) for further details.

LEMMA B.8 (Berbee, 1979). *Let u, v be two r.v.s defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking their values in Polish spaces \mathcal{U}, \mathcal{V} . Then, enlarging the probability space if it is necessary, it is possible to define an r.v. v^* with the same distribution as v and such that u and v^* are independent r.v.s and*

$$\mathbb{P}(v \neq v^*) = \beta(\sigma(u), \sigma(v)).$$

We shall use a similar device for real-valued strong mixing sequences.

LEMMA B.9 (Rio, 2000). *Let \mathcal{A} be a σ -field of $(\Omega, \mathcal{F}, \mathbb{P})$ and let v be a real-valued r.v. with values (a.s.) in $[a, b]$. Suppose, furthermore, that there exists an r.v. U with uniform distribution on $[0, 1]$ independent of \mathcal{A} and $\sigma(v)$. Then there exists an r.v. v^* independent of \mathcal{A} with the same distribution as v , $\sigma(\mathcal{A}, v, U)$ -measurable and such that*

$$\mathbb{E}|v - v^*| \leq 2(b - a)\alpha(\mathcal{A}, \sigma(v)).$$

B.2.3. Rosenthal Inequalities. First applications of Rosenthal-type inequalities can be found in Billingsley (1968). They concern the Kolmogorov–Smirnov functional CLT for the empirical cumulative distribution of a ϕ -mixing sequence. A nice and general

presentation of applications of Rosenthal inequalities to functional CLT is provided in Andrews and Pollard (1994). Maximal Rosenthal inequalities are available either in the α -mixing or in the ρ -mixing case.

For simplicity we consider stationary sequences $(X_n)_{n \in \mathbb{Z}}$, such that for some real number $r \geq 2$,

$$\mathbb{E}X_0 = 0 \quad \text{and} \quad \mathbb{E}|X_0|^r < \infty,$$

and we set

$$S_n = X_1 + \dots + X_n \quad \text{and} \quad S_n^* = \max\{|S_1|, \dots, |S_n|\}.$$

Similarly, α^{-1} denotes the inverse of the decreasing function $s \rightarrow \alpha_{[s]}$ (where $[s] \in \mathbb{Z}$ denotes the integral part of $s \in \mathbb{R}$) defined by

$$\alpha^{-1}(t) = \sum_{k=0}^{\infty} \mathbb{1}_{\{\alpha_k > t\}} \quad \forall t \in \mathbb{R}.$$

LEMMA B.10 (Rio, 1994). *Let (X_n) be a strong mixing sequence and for $r \geq 2$ set*

$$M_{r,\alpha,n} = \int_{\alpha_n}^{\alpha_0} [\alpha^{-1}(t) Q_{X_0}(t)]^r \frac{dt}{\alpha^{-1}(t)} \quad \text{and} \quad M_{r,\alpha} = M_{r,\alpha,\infty}.$$

Then there exists a constant C_r only depending on r such that

$$\mathbb{E}(S_n^*)^r \leq C_r \{ \sqrt{(nM_{2,\alpha,n})^r} + nM_{r,\alpha,n} \}.$$

For example, an explicit computation gives

$$M_{r,\alpha} = M_{r,\alpha,\infty} < \infty \quad \text{if} \quad \sum_{k=0}^{\infty} (k+1)^{r-2} \int_0^{\alpha_k} Q_{X_0}^r(t) dt < \infty.$$

This implies the maximal Rosenthal inequality $\mathbb{E}(S_n^*)^r \leq C_r \{ \sqrt{(nM_{2,\alpha})^r} + nM_{r,\alpha} \}$.

Assume now that $\mathbb{E}|X_0|^{r+\delta} < \infty$ for some positive δ . Then the integral $M_{r,\alpha} = \int_0^1 [\alpha^{-1}(t)]^{r-1} [Q_{X_0}(t)]^r dt$ can be bounded using Hölder's inequality and the relation $\int_0^1 [Q_{X_0}(t)]^p dt = \mathbb{E}|X_0|^p$:

$$M_{r,\alpha} \leq \left(\int_0^1 [\alpha^{-1}(t)]^{(r+\delta)(r-1)/\delta} dt \right)^{\delta/(r+\delta)} (\mathbb{E}|X_0|^{r+\delta})^{r/(r+\delta)}.$$

The integral on the right-hand side of the previous inequality can be written as a sum, yielding, after the use of the Abel transformation and the simple inequality $(n+1)^p - n^p \leq p(n+1)^{p-1}$ that follows for $p > 1$ from the use of the mean value theorem,

$$\begin{aligned} \int_0^1 [\alpha^{-1}(t)]^{(r+\delta)(r-1)/\delta} dt &= \sum_{n=0}^{\infty} n^{(r-1)(1+r/\delta)} (\alpha_n - \alpha_{n-1}) \\ &\leq \frac{(r+\delta)(r-1)}{\delta} \sum_{n=0}^{\infty} (n+1)^{r-2+r(r-1)/\delta} \alpha_n. \end{aligned}$$

If $\mathbb{E}|X_0|^{r+\delta} < \infty$ and $\sum_{n=0}^{\infty} (n+1)^{r-2+r(r-1)/\delta} \alpha_n < \infty$, the moment $M_{r,\alpha}$ is finite.

LEMMA B.11 (Shao, 1995). *Let (X_n) be a ρ -mixing sequence and set for $r \geq 2$ and some constant $K_{r,\rho}$ depending only on r and ρ :*

$$K_{r,\rho,n} = \exp\left(K_{r,\rho} \sum_{k=0}^{\lceil \log_2 n \rceil} \rho^{2^k/r}\right). \tag{B.1}$$

Then there exists a constant $C_{r,\rho}$ depending only on r and ρ such that

$$\mathbb{E}(S_n^*)^r \leq C_{r,\rho} \{K_{2,\rho,n}(n\|X\|_2^2)^{r/2} + K_{r,\rho,n}(n\|X\|_r^r)\}.$$

This follows from Bradley and Bryc’s lemma and from several important lemmas in Peligrad (1987). The work of Peligrad (1987) yields maximal bounds for the moments of sums of n -samples of order 2 and 4 that only involve the sums $\sum_{k=0}^{\lceil \log_2(n) \rceil} \rho(2^k)$.

B.3. Tools under Association. The following inequality is essential for studying associated r.v.s.

THEOREM B.1 (Newman, 1984). *For a pair of measurable numeric functions (f, g) defined on $A \subset \mathbb{R}^k$, we write $f \ll g$ if both functions $g + f$ and $g - f$ are nondecreasing with respect to each argument. Now let X be any associated random vector with range in A . Then*

$$(f_i \ll g_i, \text{ for } i = 1, 2) \Rightarrow (|\text{Cov}(f_1(X), f_2(X))| \leq \text{Cov}(g_1(X), g_2(X))).$$

This theorem follows simply from several applications of the definition to the coordinatewise nondecreasing functions $g_i - f_i$ and $g_i + f_i$. By an easy application of the preceding inequalities one can check that

$$|\text{Cov}(f(X), g(Y))| \leq \sum_{i=1}^k \sum_{j=1}^l \left\| \frac{\partial f}{\partial x_i} \right\|_{\infty} \left\| \frac{\partial g}{\partial y_j} \right\|_{\infty} \text{Cov}(X_i, Y_j), \tag{B.2}$$

for \mathbb{R}^k - or \mathbb{R}^l -valued associated random vectors X and Y and \mathcal{C}^1 functions f and g with bounded partial derivatives. For this, it suffices to note that $f \ll f_1$ if one defines

$$f_1(x_1, \dots, x_p) = \sum_{i=1}^p \left\| \frac{\partial f}{\partial x_i} \right\|_{\infty} |x_i|$$

and uses Theorem B.1.

Denote by $\mathcal{R}(z)$ the real part of the complex number z . Theorem B.1 can be extended to complex-valued functions, up to a factor 2 on the left-hand side of inequality (B.2). Indeed, we can now set $f \ll g$ if for any real number ω the mapping $t = (t_1, \dots, t_k) \rightarrow \mathcal{R}(g(t) + e^{i\omega(t_1 + \dots + t_k)} f(t))$ is nondecreasing with respect to each argument. Also, for any real numbers t_1, \dots, t_k ,

$$|\mathbb{E}e^{i(t_1 X_1 + \dots + t_k X_k)} - \mathbb{E}e^{it_1 X_1} \dots \mathbb{E}e^{it_k X_k}| \leq 2 \sum_{i=1}^k \sum_{j=1}^k |t_i| |t_j| \text{Cov}(X_i, X_j).$$

If now the r.v.s X_i have a density, bounded uniformly with respect to the index i , then

$$\gamma_r := \sup_{\substack{x,y \\ |i-j| \geq r}} \text{Cov}(\mathbb{1}_{\{X_i > x\}}, \mathbb{1}_{\{X_j > y\}}) \leq c \sup_{|i-j| \geq r} \text{Cov}^{1/3}(X_i, X_j). \tag{B.3}$$

This relation provides a connection between \mathcal{L} - and \mathcal{I} -weak dependences (the class \mathcal{I} will be defined in (B.11) of Section B.4) in the context of associated sequences.

The negatively associated r.v.s are much simpler to handle. The key result is a comparison theorem established by Shao (2000). For any convex function f on \mathbb{R} ,

$$\mathbb{E}f\left(\sum_{i=1}^n X_i\right) \leq \mathbb{E}f\left(\sum_{i=1}^n X_i^*\right) \tag{B.4}$$

for a given sequence $(X_i)_{1 \leq i \leq n}$ of negatively associated r.v.s and for any sequence $(X_i^*)_{1 \leq i \leq n}$ of independent r.v.s such that $X_i \stackrel{D}{=} X_i^*$ for each $i = 1, \dots, n$. To avoid triviality, we assume that the preceding right-hand-side term exists.

B.4. Tools under Weak Dependence.

B.4.1. Moment and Exponential Inequalities. Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of r.v.s with $\mathbb{E}X_n = 0$ and let $S_n = \sum_{i=1}^n X_i$. In this section, we give moment bounds for $|\mathbb{E}S_n^q|$, when $q \in \mathbb{N}$ and $q \geq 2$. For positive integers r , define coefficients of weak dependence as nondecreasing sequences $(C_{r,q})_{q \geq 2}$ such that

$$|\text{Cov}(X_{t_1} \times \dots \times X_{t_m}, X_{t_{m+1}} \times \dots \times X_{t_q})| \leq C_{r,q},$$

for $1 \leq t_1 \leq \dots \leq t_q \leq n$ and for integers $1 \leq m < q$, $t_{m+1} - t_m = r$. Doukhan and Louhichi (1999) provide explicit bounds $C_{r,q}$ to construct inequalities for partial sums S_n . Two kinds of bounds are considered, either

$$|\text{Cov}(X_{t_1} \times \dots \times X_{t_m}, X_{t_{m+1}} \times \dots \times X_{t_q})| \leq cq^\gamma M^{q-2} \theta_r \tag{B.5}$$

or

$$|\text{Cov}(X_{t_1} \times \dots \times X_{t_m}, X_{t_{m+1}} \times \dots \times X_{t_q})| \leq c \int_0^{\theta_r} Q_{X_{t_1}}(x) \times \dots \times Q_{X_{t_q}}(x) dx, \tag{B.6}$$

where Q_X still denotes X 's quantile function and $c, \gamma \geq 0$ denote real numbers. In the examples, bound (B.5) holds for bounded sequences such that $\|X_n\|_\infty \leq M$. For instance, $(\theta, \mathcal{L}, \psi)$ -weak dependence yields the bounds

$$C_{r,q} = \max_{1 \leq m < q} \psi(j^{\otimes m}, j^{\otimes(q-m)}, m, q - m) M^q \theta_r,$$

where $j(x) = x\mathbb{1}_{\{|x| \leq 1\}} + \mathbb{1}_{\{x > 1\}} - \mathbb{1}_{\{x < -1\}}$. As in Lemma 1 of Doukhan and Louhichi (1999), we see that under $(\theta, \mathcal{L}, \psi)$ -weak dependence with $\psi(h, k, u, v) = c(u, v) \times \text{Lip}(h)\text{Lip}(k)$, a bound is

$$C_{r,q} = \max_{1 \leq m < q} c(m, q - m) M^{q-2} \theta_r.$$

Bound (B.6) holds for more general r.v.s, as can be shown using moment or tail assumptions. A first consequence of the previous definitions is the following Marcinkiewicz–Zygmund inequality.

Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of r.v.s with $\mathbb{E}X_n = 0$ satisfying the condition

$$C_{r,q} = \mathcal{O}(r^{-q/2}). \tag{B.7}$$

Then, there exists a constant $B > 0$ independent of n for which

$$|\mathbb{E}S_n^q| \leq Bn^{q/2}. \tag{B.8}$$

The following lemma gives moment inequalities with $q \in \{2, 4\}$. It was essentially proved in Billingsley (1968, Lemmas 3 and 4, p. 172).

LEMMA B.12 (Doukhan and Louhichi, 1999). *If $(X_n)_{n \in \mathbb{N}}$ is a sequence of r.v.s with $\mathbb{E}X_n = 0$, then*

$$\mathbb{E}S_n^2 \leq 2n \sum_{r=0}^{n-1} C_{r,2}, \quad \mathbb{E}S_n^4 \leq 4! \left\{ \left(n \sum_{r=0}^{n-1} C_{r,2} \right)^2 + n \sum_{r=0}^{n-1} (r+1)^2 C_{r,4} \right\}. \tag{B.9}$$

The following theorems deal with higher order moments.

THEOREM B.2 (Doukhan and Louhichi, 1999). *Let $q \geq 2$ be some integer. Assume that dependence coefficients $C_{r,p}$ of the sequence (X_n) satisfy*

$$C_{r,p} = Ce^{\gamma p} M^{p-2} \theta_r, \tag{H_1}$$

for all integers $0 < p \leq q$ and for some positive constants M, γ, C . Then, for any integer $n \geq 2$,

$$|\mathbb{E}S_n^q| \leq \frac{(2q-2)!}{(q-1)!} e^{q\gamma} \left\{ \left(Cn \sum_{r=0}^{n-1} \theta_r \right)^{q/2} \vee \left(CM^{q-2} n \sum_{r=0}^{n-1} (r+1)^{q-2} \theta_r \right) \right\}. \tag{B.10}$$

Theorem B.2 is adapted to work with bounded sequences. Define the class \mathcal{I} by

$$\mathcal{I} = \{ \otimes_{i=1}^u g_{x_i}; x_i \in \mathbb{R}_+^*, u \in \mathbb{N}^* \}, \quad \text{where } g_x(y) = \mathbb{1}_{x \leq y} - \mathbb{1}_{y \leq -x}, \quad \forall x \in \mathbb{R}_+^*. \tag{B.11}$$

To consider the unbounded case, we shall consider $(\theta, \mathcal{I}, \psi)$ -weak dependence where $\psi(h, k, u, v) = c(u, v)$.

THEOREM B.3 (Doukhan and Louhichi, 1999). *Let $(X_n)_{n \in \mathbb{N}}$ be a $(\theta, \mathcal{I}, \psi)$ -weak dependent sequence with $\mathbb{E}X_n = 0$ and set $C_q = (\max_{u+v \leq q} c(u, v)) \vee 2$. Then*

$$|\mathbb{E}S_n^q| \leq \frac{(2q-2)!}{(q-1)!} \left(C_q \sum_{i=1}^n \int_0^1 (\theta^{-1}(u) \wedge n)^{q-1} Q_i^q(u) du \vee \left(C_2 \sum_{i=1}^n \int_0^1 (\theta^{-1}(u) \wedge n) Q_i^2(u) du \right)^{q/2} \right).$$

In the special case of strongly mixing and stationary sequences, this is Theorem 1 in Rio (1994). The explicit form of the constants compensates for the fact that we are restricted to even integers.

Remark. Exponential inequalities can be obtained using Theorem B.2. Define

$$M_{q,n} = n \sum_{r=0}^{n-1} (r+1)^{q-2} C_{r,q}.$$

We first suppose that for all integers $q \geq 2$ and n

$$C_{r,p} = Ce^{\gamma p} M^{p-2} \theta_r, \quad M_{q,n} \leq A_n \frac{q!}{\beta^q},$$

where β is some constant and A_n is a sequence independent of q . Then the following Doukhan–Portal type exponential inequality is available:

$$\mathbb{P}(|S_n| \geq x\sqrt{A_n}) \leq A \exp(-B\sqrt{\beta x}). \tag{DPL}$$

One can take here the constants $A = e^{4+1/12} \sqrt{8\pi}$ and $B = e^{5/2-\gamma/2}$. Let (X_n) be a sequence of r.v.s with $\mathbb{E}X_n = 0$. Inequality (DPL) holds when $C_{r,q} = C\sigma^2 M^{q-2} e^{\gamma q} e^{-br}$ for suitable constants $C, \sigma, \gamma, b > 0$, if $\|X_n\|_\infty \leq M$ and $\|X_n\|_2 \leq \sigma$, for any integer $n \geq 0$. In this case, $A_n = n\sigma^2$. For example, inequality (DPL) holds under $(\theta, \mathcal{L}, \psi)$ -weak dependence if $\theta_r = \mathcal{O}(e^{-br})$ and $\psi(h, k, u, v) \leq e^{\delta(u+v)} \text{Lip}(h)\text{Lip}(k)$ for some $\delta \geq 0$.

The use of combinatorics in the preceding inequalities makes them relatively weak. For instance, Bernstein’s inequality, valid for independent sequences, allows one to replace the term \sqrt{x} in the preceding inequality by x^2 under the same assumption $n\sigma^2 \geq 1$; in the mixing cases similar inequalities are also obtained using coupling arguments that are not available here.

Shao (2000) utilizes his main comparison result (expression (B.4)) to show that most of the inequalities for independent r.v.s remain true for negatively associated r.v.s, even with respect to the constants. For instance, a Rosenthal inequality is proved. Recalling that $S_n^* = \max\{|S_1|, \dots, |S_n|\}$, we have that for any real number $p > 2$,

$$\mathbb{E}S_n^{*p} \leq 2 \left(\frac{15p}{\ln(p)} \right)^p \left\{ \sqrt{(n \mathbb{E}X_0^2)^p} + n \mathbb{E}|X_0|^p \right\}. \tag{B.12}$$

An exponential inequality is also derived for negatively associated r.v.s. It is as sharp as the Bernstein’s inequality for independent r.v.s.

The results of Shao (2000) can also be combined with those in the papers by Ibragimov and Sharakhmentov (2001) and de la Peña, Ibragimov, and Sharakhmentov (2003) to obtain sharp versions of Rosenthal inequalities. These inequalities concern general moments of partial sums of negatively associated sequences of nonnegative and symmetric r.v.s and also mean zero r.v.s in the case of even power p . A symmetrization argument gives improvements of formula (B.12) in the context of mean zero r.v.s and arbitrary power p .

Proof of Lemma 2.1. As an application of the preceding results, let us turn to the proof of a correct version of Lemma 1 in Hall and Horowitz (1996). Set $\xi_{i,n} = \xi_i \mathbb{1}_{\{|\xi_i| \leq n^{1/16}\}} - \mathbb{E}\xi_i \mathbb{1}_{\{|\xi_i| \leq n^{1/16}\}}$ and let $\tilde{R}_n = (1/n) \sum_{i=1}^n \xi_{i,n}$. Then

$$\begin{aligned} \mathbb{P}(|R_n| > 2n^{-[(2+\epsilon)/5]}) &\leq \mathbb{P}(|\tilde{R}_n| > n^{-[(2+\epsilon)/5]}) + 2n^{1+[(2+\epsilon)/5]} \mathbb{E}|\xi_i \mathbb{1}_{\{|\xi_i| > n^{1/16}\}}| \\ &\leq n^{-[32(2+\epsilon)/5]} \mathbb{E}|R_n|^{32} + 2n^{1+[(2+\epsilon)/5]} \|\xi_1\|_4 \mathbb{P}^{3/4}(|\xi_1| > n^{1/16}) \\ &= n^{-1} \mathcal{O}(n^{-(11+32\epsilon)/5} + n^{(\epsilon/5)-(47/320)}) = o(n^{-1}). \end{aligned}$$

B.4.2. Central Limit Theorem. Following Withers (1981) and Newman (1984), the CLT holds if $\sigma^2 = \lim_{n \rightarrow \infty} (1/n) \text{Var}(X_1 + \dots + X_n) > 0$ (this limit is assumed to exist), and the sequence S_n^2/n is uniformly integrable.

Define by \mathcal{F}_{exp} the class of complex exponential functions of complex exponential functions f such that $f(x_1, \dots, x_p) = \exp(it_f(x_1 + \dots + x_p))$ for some $t_f \in \mathbb{R}$ and $p \in \mathbb{N}$. The function ψ is defined for this class by $\psi_{exp}(f, g, p, q) = pq(1 + t_f t_g)$. Let $(\theta_r^{\mathcal{F}_{exp}})$ be the dependence coefficient associated with the preceding function class.

The CLT holds for a $(\theta, \mathcal{L}, \psi)$ -weak dependent sequence with $\theta_r^{\mathcal{L}} = \mathcal{O}(r^{-2})$ and $\lim_{r \rightarrow \infty} \theta_r^{\mathcal{F}_{exp}} = 0$. The inclusion $\mathcal{F}_{exp} \subset \mathcal{L}$ implies that the CLT holds under a $(\theta, \mathcal{L}, \psi)$ -weak dependence condition, with $\psi(h, k, u, v) = \min(u, v)\text{Lip}(h)\text{Lip}(k) + 1$. Notice that the preceding condition $\theta_r^{\mathcal{L}} = \mathcal{O}(r^{-2})$ holds for associated r.v.s where $\sum_n \text{Cov}(X_0, X_n) < \infty$.

PROPOSITION B.1 (Doukhan and Louhichi, 1999). *The CLT holds under $(\theta, \mathcal{L}, \psi)$ -weak dependence if $\psi(h, k, u, v) = \min\{u, v\}\mu(\text{Lip}(h), \text{Lip}(v))$ for some locally bounded real function on \mathbb{R}_+^2 . It also holds if, for some $d > 0$, $\psi(h, k, u, v) = (u + v)^d \times \mu(\text{Lip}(h), \text{Lip}(v))$ and, for some $D > d$, $\theta_r = \mathcal{O}(r^{-D})$.*

The proof is based on a more general lemma (see, e.g., Ibragimov, 1962; Ibragimov and Linnik, 1971; and Withers, 1981) setting $Y_k = \phi(X_k)$. Let $(Y_n)_{n \in \mathbb{N}}$ be a stationary sequence. The idea is to split $S_n = \sum_{k=1}^n Y_k$ into Bernstein's blocks.

LEMMA B.13 (Ibragimov, 1962). *Let (Y_k) be a stationary sequence of centered r.v.s. Let $p = p(n)$, $q = q(n)$, $k = \lfloor n/(p + q) \rfloor$ be integer-valued functions satisfying*

$$p \rightarrow \infty, \quad q \rightarrow \infty, \quad q = o(p), \quad p = o(n), \quad \text{as } n \rightarrow \infty. \tag{B.13}$$

We define Bernstein blocks as

$$\begin{aligned} \epsilon_i &= \sum_{(i-1)(p+q)+1}^{ip+(i-1)q} Y_j, & \nu_i &= \sum_{ip+(i-1)q+1}^{i(p+q)} Y_j, & \text{for } 1 \leq i \leq k, & \text{ and} \\ \nu_{k+1} &= \sum_{k(p+q)+1}^n Y_j. \end{aligned}$$

Then we set

$$S_n = Z_k + Z'_{k+1}, \quad \text{with } Z_k = \sum_{i=1}^k \epsilon_i, \quad Z'_{k+1} = \sum_{i=1}^{k+1} \nu_i. \tag{B.14}$$

Denote $\sigma_n^2 = \text{Var } S_n$ and let g, h be either $x \rightarrow \cos x$ or $x \rightarrow \sin x$. If

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{\sigma_n^2} \mathbb{E} Z_{k+1}'^2 &= 0, & \lim_{n \rightarrow \infty} \frac{1}{\sigma_n^2} \sum_{i=1}^k \mathbb{E} |\epsilon_i|^2 &= 1, \\ \sum_{j=2}^k \left| \text{Cov} \left(g \left(\frac{t}{\sigma_n} \sum_{i=1}^{j-1} \epsilon_i \right), h \left(\frac{t}{\sigma_n} \epsilon_j \right) \right) \right| &\rightarrow 0, & \text{for all } t \in \mathbb{R}, \\ \lim_{n \rightarrow \infty} \frac{1}{\sigma_n^2} \sum_{i=1}^k \mathbb{E} |\epsilon_i|^2 \mathbb{1}_{|\epsilon_i| \geq \epsilon \sigma_n} &= 0, & \text{for all } \epsilon > 0, \end{aligned}$$

then the sequence S_n/σ_n converges in distribution to a Gaussian $\mathcal{N}(0,1)$ distribution.