

# A new mixing notion and functional central limit theorems for a sieve bootstrap in time series

PETER J. BICKEL<sup>1</sup> and PETER BÜHLMANN<sup>2</sup>

<sup>1</sup>*Department of Statistics, University of California, Berkeley, CA 94720-3860, USA.*

*e-mail: bickel@stat.berkeley.edu*

<sup>2</sup>*Seminar für Statistik, ETH-Zentrum, CH-8092 Zürich, Switzerland.*

*e-mail: buhlmann@stat.math.ethz.ch*

We study a bootstrap method for stationary real-valued time series, which is based on the sieve of autoregressive processes. Given a sample  $X_1, \dots, X_n$  from a linear process  $\{X_t\}_{t \in \mathbb{Z}}$ , we approximate the underlying process by an autoregressive model with order  $p = p(n)$ , where  $p(n) \rightarrow \infty$ ,  $p(n) = o(n)$  as the sample size  $n \rightarrow \infty$ . Based on such a model, a bootstrap process  $\{X_t^*\}_{t \in \mathbb{Z}}$  is constructed from which one can draw samples of any size.

We show that, with high probability, such a sieve bootstrap process  $\{X_t^*\}_{t \in \mathbb{Z}}$  satisfies a new type of mixing condition. This implies that many results for stationary mixing sequences carry over to the sieve bootstrap process. As an example we derive a functional central limit theorem under a bracketing condition.

*Keywords:* AR( $\infty$ ); ARMA; autoregressive approximation; bracketing; convex sets; linear process; MA( $\infty$ ); smooth bootstrap; stationary process; strong-mixing

## 1. Introduction

The bootstrap for AR( $\infty$ ) processes was proposed by Kreiss (1988; 1992). In Bühlmann (1997) it was developed further and called the ‘sieve bootstrap’. This is because the set of arbitrary but finite-order autoregressive processes can be viewed as an approximating family of finite-dimensional parametric models for an underlying true autoregressive process of order  $\infty$ . Following Bühlmann (1997), first define an MA( $\infty$ ) (or linear) process  $\{X_t\}_{t \in \mathbb{Z}}$  with expectation  $E[X_t] = \mu_X$  as follows:

$$X_t - \mu_X = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \quad (1.1)$$

where  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  is an i.i.d. sequence with common distribution  $F$ ,  $E[\varepsilon_t] = 0$ ,  $E|\varepsilon_t| < \infty$  and  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ . The set of all joint distributions on  $\mathbb{R}^{\mathbb{Z}}$  induced by such  $\{X_t\}_{t \in \mathbb{Z}}$  is a semi-parametric model indexed by  $\{F : \int x dF(x) = 0\} \times \{\{\psi_j\}_{j=0}^{\infty} \in \mathcal{L}_1 : \psi_0 = 1\}$ . An alternative definition leading to a slightly different set would be to require (1.1) for more restricted

distributions  $F$  with  $E|\varepsilon_t|^2 < \infty$ ,  $E[\varepsilon_t] = 0$  but now with  $\{\psi_j\}_{j=0}^\infty \in \ell_2$ . We are concerned here with a subset of the  $MA(\infty)$  processes which we call  $AR(\infty)$ , namely all processes representable as in (1.1) but also satisfying

$$\sum_{j=0}^{\infty} \phi_j (X_{t-j} - \mu_X) = \varepsilon_t, \quad \phi_0 = 1, \quad (1.2)$$

with  $\sum_{j=0}^{\infty} |\phi_j| < \infty$ . As is remarked in Bühlmann (1995), an  $MA(\infty)$  process is  $AR(\infty)$  if

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$$

has no zeros for  $|z| \leq 1$ ,  $z \in \mathbb{C}$ . Both the  $AR(\infty)$  and a fortiori  $MA(\infty)$  models are very rich. In particular, all stationary Gaussian processes can be approximated weakly by  $AR(\infty)$  models. The approximation we refer to is in the sense of weak convergence of finite-dimensional distributions of any order. In fact the sets of stationary process distributions obtainable as limits from (1.1) or (1.2) is quite large but far from exhaustive; see Bickel and Bühlmann (1996; 1997). Various authors – in particular, Tsay (1992) implicitly and Hjellvik and Tjøstheim (1995) explicitly – view ‘linear processes’ as being  $AR(\infty)$  (or approximable by  $AR(\infty)$ ).

Given this point of view, and given a sequence  $\{X_t\}_{t=1}^n$  from the process, it is reasonable to try to detect departures from this hypothesis of ‘linearity’ using various test statistics. This is the point of view of Hjellvik and Tjøstheim (1995) and Tsay (1992), save that Tsay considers parametric hypotheses such as Gaussian  $AR(p)$ . When dealing with the  $AR(\infty)$  hypothesis we face not only the choice of test statistics but also what critical value we should refer these statistics to. It is natural to try to estimate these critical values using a bootstrap appropriate to this hypothesis. Such a bootstrap for  $AR(\infty)$  processes was suggested and its properties were explored by Kreiss (1988; 1992) and developed further in Bühlmann (1997). Paparoditis and Streitberg (1992) use such a bootstrap for approximating the distribution of vector autocorrelations, Franke and Kreiss (1992) consider in more detail the bootstrap for real-valued finite-order ARMA processes. In Bühlmann (1997) it is established that the sieve bootstrap we discuss below gives correct approximations to the distributions of linear statistics such as  $\sum_{t=1}^n h(X_{t+1}, \dots, X_{t+m})$ , where  $h$  is smooth, or smooth functions thereof.

The statistics of Hjellvik and Tjøstheim (1995), however, involve estimates of the marginal densities of  $X_t$ , and statistics proposed by other authors (see Subba Rao and Gabr 1980), quite naturally force us to look at complicated functionals of the empirical distribution of the  $X_t$ s,  $(X_t, X_{t+1})$ s, and so on.

In this paper we introduce and study a variant of the sieve bootstrap for which we can show approximate validity of bootstrap critical values for such complicated nonlinear, non-regular statistics. In particular, we prove a functional central limit theorem under a bracketing condition for this sieve bootstrap. Such a result immediately implies that the sieve bootstrap works for estimators  $T_n$  which can be written as  $T_n = T(P_n)$ , where  $T$  is a (compactly) differentiable functional, in the sense of functional analysis, and  $P_n$  is an

empirical measure. In doing this we introduce some new notions of mixing which are of independent interest.

## 2. The smoothed sieve bootstrap

We briefly recall Kreiss’s bootstrap as given in Bühlmann (1997). Let  $\{X_t\}_{t \in \mathbb{Z}}$  be a real-valued, stationary linear process as given in (1.1) which also satisfies the infinite autoregressive representation as in (1.2). Given data  $X_1, \dots, X_n$  from such an  $\text{AR}(\infty)$  model as in (1.2), we use an autoregressive approximation as a sieve for the process  $\{X_t\}_{t \in \mathbb{Z}}$ . In a first step we fit an autoregressive process, with increasing order  $p(n)$  as the sample size  $n$  increases. We then estimate the coefficients  $\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n}$  corresponding to model (1.2), usually (but not necessarily) by the Yule–Walker estimates, which allows us to calculate centred residuals. Then we resample by the bootstrap as in Efron (1979) from these centred residuals yielding  $\varepsilon_t^*, t \in \mathbb{Z}$ . Finally, we construct a sieve bootstrap sample according to an  $\text{AR}(p(n))$  process with coefficients  $\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n}$ , that is,

$$\sum_{j=0}^{p(n)} \hat{\phi}_{j,n}(X_{t-j}^* - \bar{X}) = \varepsilon_t^*. \tag{2.1}$$

It is shown in Bühlmann (1997) that the sieve bootstrap process  $\{X_t^*\}_{t \in \mathbb{Z}}$  can be again inverted and represented as a linear process

$$X_t^* - \bar{X} = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \varepsilon_{t-j}^*, \quad \hat{\psi}_{0,n} = 1, \tag{2.2}$$

where the coefficients  $\{\hat{\psi}_{j,n}\}_{j=0}^{\infty}$  arise by inverting the estimated autoregressive transfer function  $\hat{\Phi}_n(z) = \sum_{j=0}^{p(n)} \hat{\phi}_{j,n} z^j, z \in \mathbb{C}, |z| \leq 1$ , that is,

$$\hat{\Psi}_n(z) = 1/\hat{\Phi}_n(z) = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} z^j, \quad z \in \mathbb{C}, |z| \leq 1.$$

Moreover, the behaviour of the coefficients  $\{\hat{\psi}_{j,n}\}_{j=0}^{\infty}$  is again controllable. Roughly speaking, if  $\sum_{j=0}^{\infty} j^r |\psi_j| < \infty$  ( $r \in \mathbb{N}_0$ ) then there exists a random variable  $n_0(\omega)$  such that  $\sup_{n \geq n_0} \sum_{j=0}^{\infty} j^r |\hat{\psi}_{j,n}| < \infty$  almost surely; see Bühlmann (1995). However, the bootstrap process as represented in (2.2) is not known to be mixing with mixing coefficients that can be bounded in some uniform sense over all realizations  $\omega$  of the underlying probability space. This is due to the fact that the distribution of the innovations  $\varepsilon_t^*$  is discrete and also changing with sample size  $n$ . All the literature for verifying some type of mixing property of a linear process assumes that the distribution of the innovations has a density or that the distribution is dominated by Lebesgue measure in some neighbourhood of the expectation of the innovation; see Gorodetskii (1977), Doukhan (1994). We leave it as an open question if the process in (2.1) or equivalently in (2.2) possesses some classical kind of mixing property which holds uniformly over all  $\omega$ s.

On the other hand, some type of mixing property of the sieve bootstrap process is needed

to describe and analyse the probabilistic behaviour of the bootstrap process as in (2.1) or in (2.2). We basically want to say that if the underlying process of (1.1) is linear and mixing, then the sieve bootstrap process is again linear and mixing. If so, many of the results for the underlying process  $\{X_t\}_{t \in \mathbb{Z}}$  would carry over to the sieve bootstrap process  $\{X_t^*\}_{t \in \mathbb{Z}}$ . For example, we would like a result in the spirit of Giné and Zinn (1990), which says in the i.i.d. set-up that the bootstrap for empirical processes works if and only if the corresponding empirical process for the original observations converges properly.

We propose a modified sieve bootstrap process which meets some of these goals. The idea is to resample residuals from a density estimate or, equivalently, to resample from a smooth empirical distribution of the residuals. The concept of constructing bootstrap schemes by resampling from a smooth empirical distribution is not new and has been studied in the i.i.d. set-up by Silverman and Young (1987), Hall *et al.* (1989), Falk and Reiss (1989a; 1989b) and others. The reason we use it, ensuring some new notion of mixing, is, however, entirely different than the second-order optimality invoked in the literature.

Here is our bootstrap scheme. Denote by  $X_1, \dots, X_n$  a sample from the model as in (1.2). We always assume that the distribution of the innovations  $\varepsilon_t$  has a density  $f_\varepsilon(\cdot)$  with respect to Lebesgue measure.

(1) Fit an autoregressive model of order  $p = p(n) \rightarrow \infty$ ,  $p(n) = o(n)$  ( $n \rightarrow \infty$ ) by estimating the parameters by the Yule–Walker method (cf. Brockwell and Davis 1987, Chapter 8.1). We denote the corresponding estimates by  $\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n}$  and the residuals by

$$\hat{\varepsilon}_{t,n} = \sum_{j=0}^{p(n)} \hat{\phi}_{j,n}(X_{t-j} - \bar{X}), \quad \hat{\phi}_{0,n} = 1 \quad t = p + 1, \dots, n.$$

(2) Compute a kernel density estimate for  $f_\varepsilon(\cdot)$ , based on the residuals,

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

where  $h = h(n)$  is a bandwidth with  $h = h(n) \rightarrow 0$ ,  $h(n)^{-1} = o(n)$  ( $n \rightarrow \infty$ ). Then resample

$$\varepsilon_t^* \text{ i.i.d. } \sim \hat{f}_{\hat{\varepsilon}}(x + \hat{\mu}_\varepsilon) dx, \quad t \in \mathbb{Z},$$

where  $\hat{\mu}_\varepsilon = \int_{-\infty}^{\infty} x \hat{f}_{\hat{\varepsilon}}(x) dx$ .

(3) Generate the smoothed sieve bootstrap process  $\{X_t^*\}_{t \in \mathbb{Z}}$  as in (2.1).

In the following we denote bootstrap quantities which correspond to this resampling scheme by an asterisk  $*$ . The smoothed sieve bootstrap now inherits the approximating order  $p = p(n)$  and the bandwidth  $h = h(n)$  which have to be chosen by the user.

Before giving all the technical details and derivations, we display a result which easily follows from the new mixing property of the smoothed sieve bootstrap process and our general results in empirical processes (see Sections 3 and 4). Consider a causal and invertible ARMA( $p_0, q_0$ ) process,

$$X_t = \sum_{j=1}^{p_0} \xi_j X_{t-j} + \sum_{j=1}^{q_0} \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad 0 \leq p_0, q_0 < \infty, t \in \mathbb{Z}, \tag{2.3}$$

with transfer functions  $\Xi(z) = 1 - \sum_{j=1}^{p_0} \xi_j z^j$ ,  $\Theta(z) = 1 + \sum_{j=1}^{q_0} \theta_j z^j$  ( $|z| \leq 1$ ). We typically assume

$$\Xi(z) \neq 0, \Theta(z) \neq 0 \text{ for } |z| \leq 1 + \kappa \ (\kappa > 0) \text{ and } \Xi(\cdot), \Theta(\cdot) \text{ have no common zeros.} \tag{2.4}$$

Let

$$Z_n(x) = n^{-1/2} \sum_{t=1}^n (1_{[X_t \leq x]} - F(x)), \quad x \in \mathbb{R},$$

$$Z_n^*(x) = n^{-1/2} \sum_{t=1}^n (1_{[X_t^* \leq x]} - F^*(x)), \quad x \in \mathbb{R}$$

where  $F(\cdot)$  and  $F^*(\cdot)$  denote the cdf of  $X_t$  and  $X_t^*$ , respectively. Denote by  $\Rightarrow$  weak convergence in the cadlag space  $\mathcal{D}(\mathbb{R})$  with respect to the supremum norm.

**Corollary 2.1.** *Assume that  $\{X_t\}_{t \in \mathbb{Z}}$  is an ARMA( $p_0, q_0$ ) process as defined in (2.3) with  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  an i.i.d sequence,  $\varepsilon_t \sim f_\varepsilon(x)dx$ ,  $\sup_{x \in \mathbb{R}} f_\varepsilon(x) < \infty$ ,  $E|\varepsilon_t|^4 < \infty$  and satisfying (2.4). For estimating  $f_\varepsilon(\cdot)$ , let  $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$  and  $h(n) = o(1)$ ,  $h(n)^{-1} = O(n^{1/2-\eta})$  ( $n \rightarrow \infty$ ) ( $\eta > 0$ ). Moreover, assume that the autoregressive sieve grows as  $p(n)/C_{\kappa'} \log(n) \rightarrow 1$  ( $n \rightarrow \infty$ ) for  $C_{\kappa'} = (2 \log 1 + \kappa')^{-1}$ ,  $0 < \kappa' \leq \kappa$ . Then*

$$Z_n \Rightarrow Z,$$

$$Z_n^* \Rightarrow Z \text{ in probability,}$$

where  $Z$  is the limiting Gaussian process given in Corollary 4.1.

The condition about  $p(n)$  is reasonable since the underlying ARMA process is geometrically strong-mixing and hence the approximating order  $p(n)$  should grow very slowly. The proof of Corollary 2.1 follows from much more general results; more precise arguments are given at the end of Section 4.

### 3. Mixing property of smoothed sieve bootstrap process

We will establish in this section a type of mixing property for the linear process  $\{X_t\}_{t \in \mathbb{Z}}$  in (1.1) or (1.2) and its smoothed sieve bootstrap counterpart  $\{X_t^*\}_{t \in \mathbb{Z}}$  in (2.2) or (2.1), respectively. Denote by  $\mathcal{M}_a^b = \sigma(\{X_j; a \leq j \leq b\})$  the  $\sigma$ -algebras with events that belong to the ‘time interval’  $[a, b]$ . Moreover, we denote the strong-mixing coefficients by

$$\alpha(k) = \sup_{A \in \mathcal{M}_{-\infty}^0, B \in \mathcal{M}_k^\infty} |\mathbb{P}[A \cap B] - \mathbb{P}[A]\mathbb{P}[B]|.$$

For the bootstrap we analogously define

$$\alpha^*(k) = \sup_{A \in \mathcal{M}_{-\infty}^*, B \in \mathcal{M}_k^\infty} |\mathbb{P}^*[A \cap B] - \mathbb{P}^*[A]\mathbb{P}^*[B]|,$$

where  $\mathcal{M}_a^b = \sigma(\{X_j^*; a \leq j \leq b\})$ .

We do not know whether the strong-mixing property for the smoothed sieve bootstrap holds. We will introduce a weaker type of mixing condition which is still powerful enough to establish quite general results and show that the smoothed sieve bootstrap satisfies this weaker condition.

### 3.1. A new notion of mixing

The strong-mixing concept for a stationary process  $\{X_t\}_{t \in \mathbb{Z}}$  is based on the variational norm between the joint probability and the product of the marginal probabilities. This definition allows us to bound covariances

$$|\text{cov}(Z_1, Z_2)| \leq 8 \|Z_1\|_{q_1} \|Z_2\|_{q_2} \alpha^{1/q_3}(k), \quad 1 \leq q_1, q_2, q_3 \leq \infty, q_1^{-1} + q_2^{-1} + q_3^{-1} = 1,$$

for any measurable variable  $Z_1 \in \mathcal{M}_{-\infty}^0, Z_2 \in \mathcal{M}_k^\infty$ ; see Doukhan (1994, Theorem 3, Chapter 1.2.2). However, we often only need to bound

$$|\text{cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))|,$$

with  $d_1, d_2 \in \mathbb{N}, g_1, g_2$  measurable and ‘nice’ functions.

This suggests two generalizations. First, we only consider separation between finite-time generated  $\sigma$ -algebras, that is, we consider  $\mathcal{M}_{-d_1+1}^0$  and  $\mathcal{M}_k^{k+d_2-1}, d_1, d_2 \in \mathbb{N}$ ; this is not a new generalization (see Doukhan 1994, Chapters 1.1 and 1.3). Second, we restrict ourselves to bound covariances only for certain subclasses of bounded functions. Our restrictions on the function classes are in the same spirit as the sufficient and necessary conditions for uniformity classes in the theory of weak convergence (see Bhattacharya and Ranga Rao 1976). We restrict ourselves to such a subclass of functions so that we can estimate the difference between the bootstrap and the underlying true covariances. In doing so we make use of Berry’s smoothing lemma (see Lemma 5.4) which works under such more restrictive assumptions. The new idea here is that we do not aim to bound a variational norm which measures how close the expectations of any bounded functions are under the two measures, but a weaker norm which gives us such an estimate only for bounded functions which are not too rough.

Let  $\{X_t\}_{t \in \mathbb{Z}}$  be a stationary, real-valued process. Let  $\omega_g(A) = \sup_{\mathbf{y}, \mathbf{z} \in A} |g(\mathbf{y}) - g(\mathbf{z})|, g: \mathbb{R}^d \rightarrow \mathbb{R}, A \subseteq \mathbb{R}^d, B(\mathbf{x}, \delta) = \{\mathbf{y}; \|\mathbf{x} - \mathbf{y}\| \leq \delta\} \subseteq \mathbb{R}^d, \mathbf{x} \in \mathbb{R}^d, \delta \in \mathbb{R}^+, \|\cdot\|$  the Euclidean norm in  $\mathbb{R}^d, d \in \mathbb{N}$ . Below we will also consider an averaged translated modulus of oscillation. For this we denote by  $g_{\mathbf{y}}: \mathbb{R}^d \rightarrow \mathbb{R}, g_{\mathbf{y}}(\mathbf{x}) = g(\mathbf{x} + \mathbf{y}) (\mathbf{x}, \mathbf{y} \in \mathbb{R}^d)$  the translation of the function  $g(\cdot)$ . We also let  $\|g\|_\infty = \sup_{\mathbf{x}} |g(x_1, \dots, x_d)|$  and  $\|g\|_q = (E|g(X_1, \dots, X_d)|^q)^{1/q} (1 \leq q < \infty)$ , where the dependence on the probability distribution of  $X_1, \dots, X_d$  is usually suppressed.

Our definition of mixing comes along with a class  $\mathcal{E}^d$  of measurable functions from  $\mathbb{R}^d$  to  $\mathbb{R}$  which satisfies

$$\sup_{g \in \mathcal{E}^d} \|g\|_\infty < \infty,$$

$$\sup_{g \in \mathcal{E}^d, g \neq 0} \sup_{y \in \mathbb{R}^d} \int \frac{\omega_{g_y}(B(\mathbf{x}, \delta))}{\|g\|_\infty} dQ(\mathbf{x}) \leq \text{const. } \delta^\lambda, \tag{3.1}$$

for all  $0 < \delta < 1$ , for some  $\lambda > 0$ ,  $Q$  a probability measure in  $\mathbb{R}^d$ ,  $d \in \mathbb{N}$ .

We then say that  $(\mathcal{E}^d, \lambda, Q)$  satisfies (3.1). Often, the value  $\lambda$  and the probability measure  $Q$  are not of particular interest and we associate with  $\mathcal{E}^d$  such a function class satisfying (3.1) for some  $\lambda$  and  $Q$ .

**Example 3.1** *Indicator functions of intervals in  $\mathbb{R}^d$ .* The class of functions

$$\mathcal{E}^d = \{g: \mathbb{R}^d \rightarrow \mathbb{R}; g = 1_{[(\infty, b_1] \times \dots \times (-\infty, b_d)]}, (b_1, \dots, b_d) \in \mathbb{R}^d\}$$

satisfies (3.1) with  $\lambda = 1$  and distribution  $Q$  having a bounded density with respect to Lebesgue measure.

**Example 3.2** *Simple functions of convex sets in  $\mathbb{R}^d$ .* The class of functions

$$\mathcal{E}^d = \left\{ g: \mathbb{R}^d \rightarrow \mathbb{R}; g = \sum_{j=1}^m c_j 1_{[C_j]}; 0 < K_1 \leq |c_j| \leq K_2 < \infty, C_j \in \{\text{convex sets in } \mathbb{R}^d\} \forall j \right\},$$

$m \in \mathbb{N}$  fixed, satisfies (3.1) with  $\lambda = 1$  and distribution  $Q$  having a density  $f$ , such that  $f(\mathbf{x}) = \tilde{f}(\|\mathbf{x}\|)$ ,  $\|\cdot\|$  the Euclidean norm in  $\mathbb{R}^d$ , and  $\tilde{f}$  is differentiable with  $\int_0^\infty |\tilde{f}'(y)| dy < \infty$  and  $\lim_{y \rightarrow \infty} f(y) = 0$ ; see Bhattacharya and Ranga Rao (1976, Theorem 3.1).

**Example 3.3** *Lipschitz functions of order  $\lambda$ .* Denote by  $\|\cdot\|$  the Euclidean norm in  $\mathbb{R}^d$ . The class of functions

$$\mathcal{E}^d = \{g: \mathbb{R}^d \rightarrow \mathbb{R}; 0 < K_1 \leq \|g\|_\infty \leq K_2 < \infty, \sup_{\mathbf{x}, \mathbf{y}} \{|g(\mathbf{x}) - g(\mathbf{y})| / \|\mathbf{x} - \mathbf{y}\|^\lambda\} \leq C < \infty\},$$

$0 < \lambda \leq 1$ , satisfies (3.1) with the same  $\lambda$  and with any distribution  $Q$ .

We now present our new mixing notion and define what we call the  $\nu$ -mixing coefficient for the stationary process  $\{X_t\}_{t \in \mathbb{Z}}$  as

$$\nu(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) = \sup \left\{ \left| \frac{\text{cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))}{4\|g_1\|_\infty \|g_2\|_\infty} \right|; g_1 \in \mathcal{E}^{d_1}, g_2 \in \mathcal{D}^{d_2} \right\},$$

where  $(\mathcal{E}^{d_1}, \lambda_1, Q_1), (\mathcal{D}^{d_2}, \lambda_2, Q_2)$  satisfy (3.1),

with possibly different  $\lambda_1, \lambda_2 > 0, Q_1, Q_2$  and  $d_1, d_2 \in \mathbb{N}$ . If  $g_1$  and/or  $g_2$  in the definition of  $\nu(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2})$  is the zero function, we use the convention  $0/0 = 0$ . Whenever we write

$\nu(\cdot; \mathcal{E}^{d_1}, \mathcal{D}^{d_2})$  we implicitly mean that  $(\mathcal{E}^{d_1}, \lambda_1, Q_1), (\mathcal{D}^{d_2}, \lambda_2, Q_2)$  satisfy (3.1) for some  $\lambda_1, \lambda_2 > 0, Q_1, Q_2$ . We say that the stationary process  $\{X_t\}_{t \in \mathbb{Z}}$  is  $\nu$ -mixing with respect to  $(\mathcal{E}^{d_1}, \mathcal{D}^{d_2})$  if  $\nu(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \rightarrow 0$  for  $k \rightarrow \infty$ .

It is not hard to construct examples of stationary processes  $\{X_t\}_{t \in \mathbb{Z}}$  and function classes  $\mathcal{E}^{d_1}, \mathcal{D}^{d_2}$ , so that  $\{X_t\}_{t \in \mathbb{Z}}$  is  $\nu(\cdot; \mathcal{E}^{d_1}, \mathcal{D}^{d_2})$ -mixing but not strong-mixing, that is,  $\lim_{k \rightarrow \infty} \nu(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) = 0$  and  $\liminf_{k \rightarrow \infty} \alpha(k) > 0$ . On the other hand, for a stationary real-valued process and measurable functions  $g_1$  and  $g_2$ ,

$$|\text{cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))| \leq 4\|g_1\|_\infty\|g_2\|_\infty\alpha(k),$$

(see Doukhan 1994, Lemma 3, Chapter 1.2.2), so that by the definition of  $\nu(\cdot; \mathcal{E}^{d_1}, \mathcal{D}^{d_2})$ ,

$$\nu(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \alpha(k),$$

where  $\mathcal{E}^{d_1}, \mathcal{D}^{d_2}$  ( $d_1, d_2 \in \mathbb{N}$ ) are classes of measurable functions that satisfy the condition (3.1) for some  $\lambda_1 > 0, Q_1$  and  $\lambda_2 > 0, Q_2$ , respectively. A more refined covariance inequality in terms of  $\nu$ -mixing coefficients is given in Lemma 4.1.

For the smoothed sieve bootstrap process  $\{X_t^*\}_{t \in \mathbb{Z}}$  as described in Section 2 we define

$$\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) = \sup \left\{ \left| \frac{\text{cov}^*(g_1(X_{-d_1+1}^*, \dots, X_0^*), g_2(X_k^*, \dots, X_{k+d_2-1}^*))}{4\|g_1\|_\infty\|g_2\|_\infty} \right|; \right. \\ \left. g_1 \in \mathcal{E}^{d_1}, g_2 \in \mathcal{D}^{d_2} \right\}$$

with  $(\mathcal{E}^{d_1}, \lambda_1, Q_1), (\mathcal{D}^{d_2}, \lambda_2, Q_2)$  satisfying (3.1), where  $Q_i$  is the marginal distribution of  $(X_1, \dots, X_{d_i})$  ( $i = 1, 2$ ) from the true underlying process  $\{X_t\}_{t \in \mathbb{Z}}$ . We again use the convention  $0/0 = 0$  in the definition of  $\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2})$ .

### 3.2. Assumptions and main results

We now present the framework we are working with and make some general assumptions about the stationary, real-valued process  $\{X_t\}_{t \in \mathbb{Z}}$  from which we observe a sample  $X_1, \dots, X_n$ .

- (A1) Model (1.2) holds with  $\Phi(z) = \sum_{j=0}^\infty \phi_j z^j$  bounded away from zero for  $|z| \leq 1$  ( $z \in \mathbb{C}$ ) and the autoregressive coefficients decay as  $|\phi_j| = O(j^{-\beta})$  ( $j \rightarrow \infty$ ),  $\beta > 1$ .
- (A2) The innovations  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  are i.i.d., with  $E|\varepsilon_t|^s < \infty, s \geq 1$ , and have a distribution which admits a density  $f_\varepsilon(\cdot)$  with respect to Lebesgue measure. Moreover,  $\int_{-\infty}^\infty |f_\varepsilon(x) - f_\varepsilon(x+c)|dx \leq \text{const. } c, \forall c \in \mathbb{R}$ .

As an example, ARMA( $p_0, q_0$ ) models ( $p_0 < \infty, q_0 < \infty$ ) usually satisfy our assumption (A1) with an exponential decay of the coefficients  $\{\phi_j\}_{j=0}^\infty$ .

**Theorem 3.1.** *Assume that (A1) with  $\beta > 1$  and (A2) with  $s \geq 1$  hold. Then*

$$\alpha(k) \leq \text{const. } k^{-\gamma}, \quad \text{for all } \gamma < \frac{s\beta - s - 2}{s + 1} \quad (k \in \mathbb{N}).$$



**Proof.** By using existence of  $s$ th moments, this follows directly from Gorodetskii's (1977) result.  $\square$

For the mixing property of the smoothed sieve bootstrap we make, in addition to (A1) and (A2), the following general assumptions.

- (A3)  $p(n) \rightarrow \infty$ ,  $p(n) = o((n/\log(n))^{1/(2(\beta-1))})$  ( $n \rightarrow \infty$ ) for the same  $\beta$  as in (A1).
- (A4) The kernel  $K(\cdot)$  for estimating  $f_\varepsilon(\cdot)$  satisfies:  $K(\cdot)$  is a density of a probability measure with  $\int_{-\infty}^{\infty} xK(x)dx = 0$ ,  $\int_{-\infty}^{\infty} x^2K(x)dx \neq 0$ ,  $\int_{-\infty}^{\infty} |K(x) - K(x+c)|dx \leq \text{const.}c \forall c \in \mathbb{R}$ ,  $\int_{-\infty}^{\infty} |x|^s K(x)dx < \infty$  for the same  $s$  as in (A2). Moreover, the bandwidth satisfies

$$h(n) = o(1), h(n)^{-1} = o(n) \quad (n \rightarrow \infty)$$

$$h(n)^{-1} \max\{p(n)^{-\vartheta^2/(\vartheta+1)}, p(n)(\log(n)/n)^{1/2}\} = O(1) \quad (n \rightarrow \infty),$$

for some  $\vartheta \in \mathbb{N}$ ,  $\vartheta < \beta - 1$ ,  $\beta$  as in (A1).

- (A5) The triples  $(\mathcal{C}^{d_1}, \lambda_1, \mathcal{Q}_1)$ ,  $(\mathcal{D}^{d_2}, \lambda_2, \mathcal{Q}_2)$  ( $d_1, d_2 \in \mathbb{N}$ ) that come along with the definition of the  $\nu$ -mixing coefficients satisfy (3.1) for some  $\lambda_1, \lambda_2 > 0$  and  $\mathcal{Q}_i$ , the marginal distribution of  $(X_1, \dots, X_{d_i})$  ( $i = 1, 2$ ) from the process  $\{X_t\}_{t \in \mathbb{Z}}$  defined by (A1) and (A2).

**Remark 3.1.** Assumption (A3) is common in autoregressive approximation; see An *et al.* (1982) and Bühlmann (1995). If the approximating order is chosen by the data through the Akaike information criterion (AIC), then Shibata (1980) has shown that  $\hat{p}_{AIC} \sim \text{const.}n^{1/(2\beta)}$  (if  $|\phi_j| \sim \text{const.}j^{-\beta}$  as  $j \rightarrow \infty$ ), which satisfies (A3).

**Remark 3.2.** A slightly less general set of conditions for the parameters  $p(n)$ ,  $h(n)$  in (A3) and (A4) is given by

$$p(n)/(\text{const.}n^{c_p}) \rightarrow 1 \quad (n \rightarrow \infty), \quad 0 < c_p < 1/(2(\beta - 1)),$$

$$h(n)^{-1} = O(\max\{n^{c_p \vartheta^2/(\vartheta+1)}, n^{1/2-c_p}/(\log(n))^{1/2}\}), \quad h(n) = o(1),$$

or

$$p(n)/\text{const.} \log(n) \rightarrow 1 \quad (n \rightarrow \infty),$$

$$h(n)^{-1} = O((\log(n))^{\vartheta^2/(\vartheta+1)}), \quad h(n) = o(1).$$

**Remark 3.3.** The assumption about the interplay between the bandwidth  $h(n)$  and the approximating order  $p(n)$  in (A4) covers a wide range of situations. By taking  $p(n) = \text{const.}n^{1/(2\beta)}$  (this is the order of  $\hat{p}_{AIC}$ ) and  $h(n) = \text{const.}n^{-1/5}$  (this is the usual order for estimating  $f_\varepsilon(\cdot)$  with respect to the mean square error), (A4) holds for any  $\beta > 4$ .

The assumption about the interplay between  $h(n)$  and  $p(n)$  is in some sense counter-intuitive. When choosing a small bandwidth  $h$ , corresponding to small bias in density estimation, we are forced to take a large enough  $p$ , corresponding again to small bias in autoregressive approximation. The condition may be an artefact of our proof. It is needed for controlling the  $\nu^*$ -mixing coefficient for large separation lags: we obtain a bound which only depends on the smoothness of the bootstrap innovations, and not on some closeness between the bootstrap and the true underlying distribution, indicating that bias questions are irrelevant for mixing bounds.

Assumption (A5) will restrict our focus to function classes which inherit the probability law  $Q$  from the underlying original process. Our mixing statements for the smoothed sieve bootstrap exclusively correspond to such classes.

**Theorem 3.2.** *Assume that (A1)–(A5) hold, with  $\beta \geq 2$  in (A1),  $s \geq 4$  in (A2) and  $\lambda = \min\{\lambda_1, \lambda_2\} > 0$  in (A5). Then*

$$\mathbb{P}[\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \text{const.} \cdot k^{-\gamma^* \frac{s\lambda}{s(1+2\lambda+d)+d}} \forall k \in \mathbb{N}] \rightarrow 1 \quad (n \rightarrow \infty),$$

where  $d = d_1 + d_2$ ,  $\gamma^* = (s[\beta] - s - 2)/(s + 1)$  if  $\beta \notin \mathbb{N}$ ,  $\gamma^* = s\beta - s - 3/s + 1$  if  $\beta \in \mathbb{N}$ .

The proof is given in Section 5.

Theorem 3.2 describes the ‘loss’ for the decaying speed of the bootstrap compared to the original mixing coefficients. By setting  $\alpha(k) \leq \text{const.} \cdot k^{-\gamma}$  (see Theorem 3.1), we can always write

$$\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \text{const.} \cdot k^{-\gamma L}, \quad L < \frac{(s\beta - s - 3)s\lambda}{(s\beta - s - 2)(s(1 + 2\lambda + d) + d)}.$$

If (A2) holds for all  $s \in \mathbb{N}$ , then  $L < \lambda/(1 + 2\lambda + d)$ . Note that often the case  $d_1 = d_2 = 1$  ( $d = 2$ ) and  $\lambda = 1$  applies which then yields  $L < 1/5$ . We further note that the decay of  $\nu^*(\cdot; \mathcal{E}^{d_1}, \mathcal{D}^{d_2})$  is still polynomial.

There is also some interest in the case where the autoregressive coefficients  $\phi_j$  in model (1.2) decay exponentially. As examples we mention ARMA( $p_0, q_0$ ) models ( $p_0 < \infty, q_0 < \infty$ ). Then the mixing coefficients decay also at an exponential rate. Under more restrictive assumptions than before, the smoothed sieve bootstrap process  $\{X_t^*\}_{t \in \mathbb{Z}}$  is again  $\nu$ -mixing with exponentially decaying coefficients. We strengthen the assumptions as follows.

(A1') Model (1.2) holds with  $\Phi(z) = \sum_{j=0}^{\infty} \phi_j z^j$  bounded away from zero for  $|z| \leq 1 + \kappa$  and  $\sum_{j=0}^{\infty} |\phi_j| (1 + \kappa)^j < \infty$  for some  $\kappa > 0$ .

(A3')  $p(n)/(C \log n) \rightarrow 1 \quad (n \rightarrow \infty)$ ,  $C \in \mathbb{R}^+$ .

(A4') The same assumptions for the kernel  $K(\cdot)$  as in (A4), but the bandwidth satisfies

$$h(n)^{-1} = O(\max\{n^{C \log(1+\Delta)}, n^{1/2-\eta}\}),$$

$$\text{for some } \eta > 0, \text{ for some } 0 < \Delta < \min\{\kappa, \exp(1/(2C)) - 1\},$$

with the same  $\kappa$  as in (A1') and the same  $C$  as in (A3').

Assumption (A1') is almost the same as in Kreiss (1988; 1992). Assumption (A3') reflects the behaviour of the AIC under the condition (A1') since then  $\hat{p}_{AIC} \sim \text{const.} \log(n)$ ; cf. Shibata (1980). However, we allow a general constant  $C \in \mathbb{R}^+$ . We now briefly discuss a specific choice of the constant  $C$  in (A3') which then would simplify (A4'). The error for estimating  $\Phi(z)$  in  $|z| \leq 1$  is given by

$$\sup_{|z| \leq 1} |\hat{\Phi}_n(z) - \Phi(z)| = O((\log(n)/n)^{1/2}) + O\left(\sum_{j=p(n)+1}^{\infty} |\phi_j|\right) \text{ almost surely.}$$

A typical approach would be to choose  $p(n)$  such that  $\sum_{j=p(n)+1}^{\infty} |\phi_j| \sim \text{const.} n^{-1/2}$ . Assuming that  $|\phi_j| \sim \text{const.} (1 + \kappa)^{-j}$  ( $j \rightarrow \infty$ ), we then would choose  $p(n) = p_\kappa(n) = C_\kappa \log(n)$  with  $C_\kappa = (2 \log(1 + \kappa))^{-1}$ . Then, for the condition on the bandwidth  $h(n)$  in (A4'),  $\Delta < \kappa$ ,  $n^{C_\kappa \log(1+\Delta)} = n^{\log(1+\Delta)/(2 \log(1+\kappa))}$  and hence the only remaining condition on the bandwidth would be

$$h(n)^{-1} = O(n^{1/2-\eta}) \text{ for some } \eta > 0, \quad h(n) = o(1).$$

**Theorem 3.3.** *Assume that (A1') with  $\kappa > 0$  and (A2) with  $s \geq 1$  hold. Then*

$$\alpha(k) \leq \text{const.} \rho^k, \quad \text{for all } (1 + \kappa)^{-s/(s+1)} < \rho < 1 \text{ (} k \in \mathbb{N}\text{)}.$$

**Proof.** This follows directly from Gorodetskii's (1977) result. □

For the smoothed sieve bootstrap we can show:

**Theorem 3.4.** *Assume that (A1') with  $\kappa > 0$ , (A2) with  $s \geq 4$ , (A3') with  $C \in \mathbb{R}^+$ , (A4') and (A5) with  $\lambda = \min\{\lambda_1, \lambda_2\} > 0$  hold. Then*

$$\mathbb{P}[\nu^*(k; \mathcal{E}^{d_1}, \mathcal{G}^{d_2}) \leq \text{const.} (\rho^*)^k, \forall k \in \mathbb{N}] \rightarrow 1 \text{ (} n \rightarrow \infty\text{)},$$

$$\text{for all deterministic } (1 + \tilde{\kappa})^{-\frac{\tilde{\kappa}^2 \lambda}{(s+1)(s(1+2\tilde{\lambda}+d)+d)}} < \rho^* < 1, \quad d = d_1 + d_2,$$

where  $0 < \tilde{\kappa} < \min\{\kappa, \exp(1/(2C)) - 1\}$  and  $\tilde{\kappa}$  is restricted to be appropriately close to  $\min\{\kappa, \exp(1/(2C)) - 1\}$ . In particular, by choosing  $C = C_\kappa = (2 \log(1 + \kappa))^{-1}$  in (A3') we have

$$(1 + \kappa)^{-\frac{\tilde{\kappa}^2 \lambda}{(s+1)(s(1+2\tilde{\lambda}+d)+d)}} < \rho^* < 1, \quad d = d_1 + d_2.$$

The proof is outlined in Section 5.

Our results are stated in probability. One way to extend them to hold almost surely is to assume higher moments in (A2) and a faster decay of the autoregressive coefficients in (A1), and then to make use of the Borel–Cantelli lemma to show complete convergence.

### 4. Smoothed sieve bootstrap and central limit theorems

We first establish more refined covariance inequalities in terms of the  $\nu$ -mixing coefficients. To obtain them, we enlarge the function classes  $\mathcal{E}^{d_1}, \mathcal{D}^{d_2}$  by some truncated functions. For  $0 < M < \infty$ , let

$$t_M(g)(\mathbf{x}) = \begin{cases} g(\mathbf{x}), & \text{if } |g(\mathbf{x})| \leq M, \\ \text{sign}(g(\mathbf{x}))M, & \text{if } |g(\mathbf{x})| > M. \end{cases}$$

Define by

$$\mathcal{E}_{\text{ext},C}^d = \mathcal{E}^d \cup \{g; g = t_M(g), g \in \mathcal{E}^d, C < M < \infty\}, \quad C \geq 0,$$

the extension of the class  $\mathcal{E}^d$ , closed under all truncations  $t_M(\cdot)$ ,  $C < M < \infty$ .

**Example 4.1** *Extensions of single functions.* Consider  $\mathcal{E}^d = \{g\}$  ( $g \neq 0$ ) which satisfies (3.1) for some  $\lambda > 0$  and some  $Q$ . Assume  $2^{1/q} \|g\|_q > C > 0$  for some  $1 \leq q \leq \infty$ . The extension  $\mathcal{E}_{\text{ext},C}^d$  plays an important role; see Lemma 4.1(i). For such  $C$ , the triple  $(\mathcal{E}_{\text{ext},C}^d, \lambda, Q)$  satisfies (3.1) again.

**Example 4.2** *Extensions of indicator and simple functions.* For  $(\mathcal{E}^d, \lambda, Q)$  as in Examples 3.1 or 3.2, the extension  $\mathcal{E}_{\text{ext},0}^d$  plays an important role; see Lemma 4.1(ii). Since truncated indicator functions are either the function itself or the zero function, and similarly for simple functions, the extension  $(\mathcal{E}_{\text{ext},0}^d, \lambda, Q)$  satisfies (3.1) again.

**Lemma 4.1.** *Let  $\{X_t\}_{t \in \mathbb{Z}}$  be a stationary real-valued process and let  $\mathcal{E}^{d_1}, \mathcal{D}^{d_2}$  ( $d_1, d_2 \in \mathbb{N}$ ) be classes of measurable functions that satisfy the condition (3.1) for some  $\lambda_1 > 0, Q_1$  and  $\lambda_2 > 0, Q_2$ , respectively.*

(i) *If  $\mathcal{E}^{d_1} = \{g_1\}$  and  $\mathcal{D}^{d_2} = \{g_2\}$  and if  $\|g_1\|_{q_1} > 0, \|g_2\|_{q_2} > 0$  for some  $1 \leq q_1, q_2 \leq \infty$ , then for  $q_1^{-1} + q_2^{-1} + q_3^{-1} = 1$ ,*

$$|\text{cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))| \leq 8 \|g_1\|_{q_1} \|g_2\|_{q_2} \nu^{1/q_3}(k; \mathcal{E}_{\text{ext},C_1}^{d_1}, \mathcal{D}_{\text{ext},C_2}^{d_2}),$$

where  $0 < C_1 < \|g_1\|_{q_1} 2^{1/q_1}$  and  $0 < C_2 < \|g_2\|_{q_2} 2^{1/q_2}$ .

(ii) *If  $g_1 \in \mathcal{E}^{d_1}, g_2 \in \mathcal{D}^{d_2}$ , then for  $q_1^{-1} + q_2^{-1} + q_3^{-1} = 1$  ( $1 \leq q_1, q_2, q_3 \leq \infty$ ),*

$$\sup \left\{ \frac{|\text{cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))|}{\|g_1\|_{q_1} \|g_2\|_{q_2}}; g_1 \in \mathcal{E}_{\text{ext},0}^{d_1}, \mathcal{D}_{\text{ext},0}^{d_2}, g_1, g_2 \neq 0 \right\} \leq 8 \nu^{1/q_3}(k; \mathcal{E}_{\text{ext},0}^{d_1}, \mathcal{D}_{\text{ext},0}^{d_2}).$$

**Remark 4.1.** Theorem 3.2 and 3.3 give bounds for  $\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2})$ . If the extensions  $\mathcal{E}_{\text{ext},C_1}^{d_1}, \mathcal{D}_{\text{ext},C_2}^{d_2}$  ( $C_1, C_2 > 0$ ) or the extensions  $\mathcal{E}_{\text{ext},0}^{d_1}, \mathcal{D}_{\text{ext},0}^{d_2}$  again satisfy assumption (A5) in Section 3.2, then the same bounds (up to a constant with respect to  $k$  and  $n$ ) apply for  $\nu^*(k; \mathcal{E}_{\text{ext},C_1}^{d_1}, \mathcal{D}_{\text{ext},C_2}^{d_2})$  ( $C_1, C_2 > 0$ ) or  $\nu^*(k; \mathcal{E}_{\text{ext},0}^{d_1}, \mathcal{D}_{\text{ext},0}^{d_2})$ , respectively.

**Proof.** Assertions (i) and (ii) can be proved simultaneously. We note that a first step

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

follows immediately by the definition of the  $\nu$ -mixing coefficient. We now consider the case where  $0 < q_1^{-1} + q_2^{-1} < 1$ ; the situation  $q_1^{-1} + q_2^{-1} = 1$  follows by Hölder's inequality. Consider first the case  $1 < q_1 < \infty, q_2 = \infty$ . Define

$$g_1^{\text{low},M}(\mathbf{x}) = t_M(g_1)(\mathbf{x}), \quad g_1^{\text{upp},M}(\mathbf{x}) = g(\mathbf{x}) - g_1^{\text{low},M}(\mathbf{x}).$$

Note that  $g_1^{\text{low},M}$  is in  $\mathcal{E}_{\text{ext},C_1}^{d_1}$  ( $C_1 < M$ ), satisfying (3.1). By writing  $g_1^{\text{low}}$  for  $g_1^{\text{low},M}(\mathbf{x})$  and  $g_1^{\text{upp}}$  for  $g_1^{\text{upp},M}(\mathbf{x})$ , we thus have

$$\begin{aligned} |\text{cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))| &= |\text{cov}(g_1^{\text{low}} + g_1^{\text{upp}}, g_2(X_k, \dots, X_{k+d_2-1}))| \\ &\leq 4M \|g_2\|_\infty \nu(k; \mathcal{E}_{\text{ext},C_1}^{d_1}, \mathcal{D}^{d_2}) + 2 \|g_2\|_\infty E |g_1^{\text{upp}}|. \end{aligned}$$

Since  $|g_1^{\text{upp}}| \leq |g_1| 1_{\{|g_1| > M\}}$ , we have by Hölder's inequality  $E |g_1^{\text{upp}}| \leq \|g_1\|_{q_1}^{q_1} M^{-q_1+1}$ . By choosing  $M$  such that  $\|g_1\|_{q_1}^{q_1} M^{-q_1} = \nu(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2})$ , we arrive at

$$|\text{cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))| \leq 6 \|g_1\|_{q_1} \|g_2\|_\infty \nu^{1-1/q_1}(k; \mathcal{E}_{\text{ext},C_1}^{d_1}, \mathcal{D}^{d_2}).$$

Note that we have chosen the truncation  $M = \|g_1\|_{q_1} / \nu^{1/q_1}(k; \mathcal{E}_{\text{ext},C_1}^{d_1}, \mathcal{D}^{d_2}) \geq \|g_1\|_{q_1} 2^{1/q_1}$ , since  $\nu(k; \mathcal{E}_{\text{ext},C_1}^{d_1}, \mathcal{D}^{d_2}) \leq 2^{-1}$  for all  $k$ . We need to include the functions  $t_M(g), g \in \mathcal{E}^{d_1}$ , in the extended class  $\mathcal{E}_{\text{ext},C_1}^{d_1}$  so that  $C_1 < M$ , explaining the restriction about  $C_1$  in assertion (i).

Now consider the situation  $1 < q_1, q_2 < \infty$ . Analogously as above, we define  $g_2^{\text{upp}}$  and  $g_2^{\text{low}}$  with a truncation point  $M'$ . Note that  $g_2^{\text{low},M'}$  is in  $\mathcal{D}_{\text{ext},C_2}^{d_2}$  ( $C_2 < M'$ ), satisfying (3.1). Then with the covariance inequality above and Hölder's inequality,

$$\begin{aligned} |\text{cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))| \\ \leq 6M' \|g_1\|_{q_1} \nu^{1-1/q_1}(k; \mathcal{E}_{\text{ext},C_1}^{d_1}, \mathcal{D}_{\text{ext},C_2}^{d_2}) + 2M' \|g_1\|_{q_1} \|g_2^{\text{upp}}\|_{q_1/(q_1-1)}. \end{aligned}$$

Again by Hölder's inequality we get  $\|g_2^{\text{upp}}\|_{q_1/(q_1-1)} \leq \|g_2\|_{q_2}^{q_2(q_1-1)/q_1} M'^{-q_2(q_1-1)/q_1+1}$ . Now choose  $M'$  such that  $\|g_2\|_{q_2}^{q_2(q_1-1)/q_1} M'^{-q_2(q_1-1)/q_1} = \nu^{1-1/q_1}(k; \mathcal{E}_{\text{ext},C_1}^{d_1}, \mathcal{D}_{\text{ext},C_2}^{d_2})$ , which yields the desired covariance inequality. Note that  $M' = \|g_2\|_{q_2} \nu^{1/q_2}(k; \mathcal{E}_{\text{ext},C_1}^{d_1}, \mathcal{D}_{\text{ext},C_2}^{d_2}) \geq \|g_2\|_{q_2} 2^{1/q_2}$ . We need  $C_2 < M'$ , which explains the restriction about  $C_2$  in assertion (i).  $\square$

Often one is interested in estimating covariances of products. Suppose that  $(\mathcal{E}_1^{d_1}, \lambda_1, Q_1), \dots, (\mathcal{E}_r^{d_r}, \lambda_r, Q_r)$  all satisfy (3.1) for some  $\lambda_i > 0, Q_i, d_i \in \mathbb{N}$  ( $i = 1, \dots, r$ ). Then we define

$$\otimes_{i=1}^r \mathcal{E}_i^{d_i} = \{g_1 \cdot \dots \cdot g_r : \mathbb{R}^{\sum_{i=1}^r d_i} \rightarrow \mathbb{R}; g_i \in \mathcal{E}_i^{d_i}, i = 1, \dots, r\}. \tag{4.1}$$

Without loss of generality, we assume that the constant function  $g \equiv 1$  is an element of every  $\mathcal{E}_i^{d_i}$  so that subproducts like  $g_{i_1} \cdot \dots \cdot g_{i_m}, i_j \in \{1, \dots, r\}, i_j \neq i_{j'} (j \neq j'), j = 1, \dots, m, m \leq r$ , are elements of  $\otimes_{i=1}^r \mathcal{E}_i^{d_i}$ .

The triple  $(\otimes_{i=1}^r \mathcal{E}_i^{d_i}, \lambda, Q)$  satisfies (3.1), where  $\lambda = \min\{\lambda_i; 1 \leq i \leq r\}$  and  $Q$  is any

distribution on  $\mathbb{R}^{\sum_{i=1}^r d_i}$  which has as marginal distributions  $Q_1, \dots, Q_r$ ; for example,  $Q = Q_1 \times \dots \times Q_r$ . This fact enables us to establish the same moment inequalities for centred sums as for  $\alpha$ -mixing sequences. We abbreviate by  $\otimes_{i=1}^r \mathcal{C}^d = \otimes_{i=1}^r \mathcal{C}_i^{d_i}$  with  $\mathcal{C}_i^{d_i} = \mathcal{C}^d$  for  $i = 1, \dots, r$ .

**Lemma 4.2.** *Let  $\{X_t\}_{t \in \mathbb{Z}}$  be a stationary real-valued process. Assume that  $g : \mathbb{R}^d \rightarrow \mathbb{R} \in \mathcal{C}^d$  satisfying (3.1) for some  $\lambda > 0, Q$ . Then the following holds true.*

(i) (Yokoyama's inequality.) *If  $\sum_{k=0}^\infty (k+1)^{r-1} \nu^{\delta/(2r+\delta)}(k; (\otimes_{i=1}^{2r-1} \mathcal{C}^d)_{\text{ext},0}) < \infty, \delta > 0$ , then*

$$\mathbb{E} \left| n^{-1/2} \sum_{t=1}^n (g(X_{t+1}, \dots, X_{t+d}) - \mathbb{E}[g(X_{t+1}, \dots, X_{t+d})]) \right|^{2r} \leq \text{const.} \|g\|_{2r+\delta}^{2r}, \quad r \in \mathbb{N},$$

where const. is universal for all  $g \in \mathcal{C}^d$ .

(ii) (Doukhan and Portal's inequality.) *Denote by  $Z_t = g(X_{t+1}, \dots, X_{t+d}) - \mathbb{E}[g(X_1, \dots, X_d)]$ ,  $t \in \mathbb{Z}$ . Assume  $|Z_t| \leq 1 \forall t, \mathbb{E}|Z_t|^2 \leq \tau^{2+\delta}, \delta > 0$  and  $\sum_{k=0}^\infty (k+1)^{2r-2} \nu^{\delta/(2r+\delta)}(k; (\otimes_{i=1}^{2r-1} \mathcal{C}^d)_{\text{ext},0}, (\otimes_{i=1}^{2r-1} \mathcal{C}^d)_{\text{ext},0}) < \infty$ . Then*

$$\mathbb{E} \left| n^{-1/2} \sum_{t=1}^n (g(X_{t+1}, \dots, X_{t+d}) - \mathbb{E}[g(X_{t+1}, \dots, X_{t+d})]) \right|^{2r} \leq \text{const.} ((n\tau^2)^r + \dots + (n\tau^2)^r), \quad r \in \mathbb{N},$$

where const. is universal for all  $g \in \mathcal{C}^d$ .

**Remark 4.2.** For a single function  $g$  with  $\|g\|_1 > 0$ , Lemma 4.2 also holds with  $\nu(\cdot; (\otimes_{i=1}^{2r-1} \mathcal{C}^d)_{\text{ext},C}, (\otimes_{i=1}^{2r-1} \mathcal{C}^d)_{\text{ext},C})$  for some  $C > 0$ ; see also Lemma 4.1(i).

**Proof.** By using Lemma 4.1(ii) the statements follow as in Yokoyama (1980) and Doukhan and Portal (1987) or Andrews and Pollard (1994), respectively.  $\square$

### 4.1. Central limit theorems

In the context of time series one often estimates a functional which depends on the  $q$ -dimensional marginal distribution of the underlying real-valued process  $\{X_t\}_{t \in \mathbb{Z}}$ . We study here consistency of the smoothed sieve bootstrap for sums

$$Z_n = (n - q + 1)^{-1/2} \sum_{t=1}^{n-q+1} (f(X_t, \dots, X_{t+q-1}) - \mathbb{E}[f(X_t, \dots, X_{t+q-1})]) \quad (q \in \mathbb{N}).$$

The smoothed sieve bootstrap version is

$$Z_n^* = (n - q + 1)^{-1/2} \sum_{t=1}^{n-q+1} (f(X_t^*, \dots, X_{t+q-1}^*) - \mathbb{E}^*[f(X_t^*, \dots, X_{t+q-1}^*)]).$$

**Theorem 4.1.** Assume that (A1)–(A4) hold with  $\beta > 1$  in (A1),  $s \geq 4$  in (A2),  $f$  satisfies (3.1) for some  $\lambda > 0$  and for  $Q$ , the marginal distribution of  $(X_1, \dots, X_q)$  from the process  $\{X_t\}_{t \in \mathbb{Z}}$  defined by (A1) and (A2); also,  $f$  has at most countably many discontinuities. In addition, we assume that the parameters  $\beta, s, \lambda$  and the dimension  $q$  are such that

$$\gamma^* > \frac{s(1 + 2\lambda + 2q) + 2q}{s\lambda},$$

where  $\gamma^*$  is as in Theorem 3.2, depending on  $\beta$ . Then,

$$Z_n \xrightarrow{d} \mathcal{N}(0, \sigma^2), \sigma^2 = \sum_{k=-\infty}^{\infty} \text{cov}(f(X_0, \dots, X_{q-1}), F(X_k, \dots, X_{k+q-1})),$$

$$Z_n^* \xrightarrow{d^*} \mathcal{N}(0, \sigma^2) \text{ in probability.}$$

**Remark 4.3.** Theorem 4.1 is an extension of Theorem 3.3 in Bühlmann (1997) to possibly non-differentiable functions  $f$ . Under a geometric decay of  $|\phi_j|$  as  $j \rightarrow \infty$  in (A1), Theorem 4.1 holds for all  $q \in \mathbb{N}$ .

**Proof.** We remark here that  $f$  is  $Q$ -continuous,  $Q$  being the probability measure of  $(X_1, \dots, X_q)$ , which admits a density with respect to Lebesgue measure, that is,  $f$  is continuous except on a set with  $Q$ -probability zero. This is a requirement we will need.

For simplicity, we sketch here the case  $q = 1$ . The general case for  $q \in \mathbb{N}$  is then straightforward, but in a notationally more awkward way. We follow the same strategy as in the proof of Theorem 3.3 in Bühlmann (1997, pp. 144–146) by applying a truncation technique to the moving-average representation of  $X_t^*$ ; see (2.2). We write  $X_{t,M}^* = \sum_{j=0}^M \hat{\psi}_{j,n} \varepsilon_{t-j}^*$  and define  $Z_{n,M}^*(\cdot)$  by means of the variables  $\{X_{t,M}^*\}_{t=1}^n$ . By exploiting the  $M$ -dependence we obtain straightforwardly, as in Bühlmann (1997),

$$Z_{n,M}^* \xrightarrow{d^*} Z_M \text{ in probability.} \tag{4.2}$$

Here  $Z_M$  is the limit based on the truncated  $X_{t,M}$ s,  $X_{t,M} = \sum_{j=0}^M \psi_j \varepsilon_{t-j}$ .

Then we show that the effect of replacing  $Z_n^*$  by  $Z_{n,M}^*$  and  $Z$  by  $Z_M$  becomes negligible for large  $M$ . We first show that

$$\mathbb{P}[Z_M \leq c] \rightarrow \mathbb{P}[Z \leq c] \quad (M \rightarrow \infty), \quad c \in \mathbb{R}. \tag{4.3}$$

Formula (4.3) follows by showing

$$\sum_{k=-M}^M \text{cov}(f(X_{0,M}), f(X_{k,M})) \rightarrow \sum_{k=-\infty}^{\infty} \text{cov}(f(X_0), f(X_k)) \quad (M \rightarrow \infty).$$

But this holds true by using the strong-mixing property of  $\{X_t\}_{t \in \mathbb{Z}}$ , the boundedness and  $Q$ -continuity of  $f$  and  $(X_{0,M}, X_{k,M}) \xrightarrow{d} (X_0, X_k) \quad (M \rightarrow \infty)$ ; cf. Bhattacharya and Ranga Rao (1976, Theorem 1.3).

Finally, we show that for all  $\eta > 0$  there exist an  $M_0(\eta)$  and an  $n_0(\eta)$  such that

$$\text{var}^*(Z_{n,M}^* - Z_n^*) \leq \eta \text{ on a set } A_n, \quad \forall n \geq n_0, \forall M \geq M_0 \tag{4.4}$$

where  $\mathbb{P}[A_n] \rightarrow 1$  ( $n \rightarrow \infty$ ).

We use a similar strategy to that for proving Theorem 3.2 (compare with formula (5.3) and the covariance inequality for strong-mixing processes at the beginning of Section 3.1) but now for the vectorized process  $\{Y_t^*\}_{t \in \mathbb{Z}} = \{X_t^*, X_{t,M}^*\}_{t \in \mathbb{Z}}$ , and bound

$$\begin{aligned} |\text{cov}^*(f(X_{0,M}^*) - f(X_0^*), f(X_{k,M}^*) - f(X_k^*))| &\leq |\text{cov}(f(X_{0,M}) - f(X_0), f(X_{k,M}) - f(X_k))| \\ &+ |\text{cov}^*(f(X_{0,M}^*) - f(X_0^*), f(X_{k,M}^*) - f(X_k^*)) - \text{cov}(f(X_{0,M}) - f(X_0), f(X_{k,M}) - f(X_k))| \\ &\leq 32 \|f(X_{0,M}) - f(X_0) - \mathbb{E}[f(X_{0,M})] + \mathbb{E}[f(X_0)]\|_{2+\delta}^2 \alpha^{\delta/(2+\delta)}(k) + B_n, \quad \delta > 0, \end{aligned} \tag{4.5}$$

where  $B_n$  is the bound from Berry’s smoothing lemma (Lemma 5.2), which is bounded by  $O_P(h(n)^{\{s\lambda/[s(1+\lambda+2q)+2q]\}})$ ; see Lemma 5.5. This inequality will be used for  $k$  not too large. On the other hand, by Lemma 5.3,

$$\begin{aligned} |\text{cov}^*(f(X_{0,M}^*) - f(X_0^*), f(X_{k,M}^*) - f(X_k^*))| \\ \leq 16 \|f\|_\infty^2 \alpha^*(k) \leq \text{const.} h(n)^{-1} k^{-\gamma^*} \text{ in probability.} \end{aligned} \tag{4.6}$$

This inequality will be used for  $k$  large.

We then bound by (4.5) and (4.6), for  $0 < \kappa < 1$ ,  $\delta > 0$ ,

$$\begin{aligned} \text{var}^*(Z_{n,M}^* - Z_n^*) &\leq \sum_{k=-n+1}^{n-1} |\text{cov}^*(f(X_{0,M}^*) - f(X_0^*), f(X_{k,M}^*) - f(X_k^*))| \\ &\leq \sum_{k=-B_n^{-(1-\kappa)}}^{B_n^{-(1-\kappa)}} (32 \|f(X_{0,M}) - f(X_0) - \mathbb{E}[f(X_{0,M})] + \mathbb{E}[f(X_0)]\|_{2+\delta}^2 \alpha^{\delta/(2+\delta)}(k) + B_n) \\ &+ \text{const.} \sum_{|k| > B_n^{-(1-\kappa)}} h(n)^{-1} k^{-\gamma^*} \\ &\leq \text{const.} \|f(X_{0,M}) - f(X_0) - \mathbb{E}[f(X_{0,M})] + \mathbb{E}[f(X_0)]\|_{2+\delta}^2 + o(1) \text{ on a set } A_n, \end{aligned}$$

with  $\mathbb{P}[A_n] \rightarrow 1$  ( $n \rightarrow \infty$ ). We have used here the restrictions on the parameters  $\gamma^*$ ,  $s$ ,  $\lambda$ ,  $q$  in the assumptions (choose  $\kappa > 0$  close to zero), which imply  $\sum_{|k| > B_n^{-(1-\kappa)}} h(n)^{-1} k^{-\gamma^*} = o_P(1)$  and also the summability of the mixing coefficients  $\alpha^{\delta/(2+\delta)}(\cdot)$  (choose  $\delta > 0$  large enough). Note also that  $\mathbb{E}|f(X_{0,M}) - f(X_0) - \mathbb{E}[f(X_{0,M})] + \mathbb{E}[f(X_0)]|^{2+\delta} \leq \text{const.} \mathbb{E}|f(X_{0,M}) - f(X_0)|^2$ , since  $\|f\|_\infty < \infty$ . But  $\mathbb{E}|f(X_0) - f(X_{0,M})|^2 \rightarrow 0$  ( $M \rightarrow \infty$ ), hence we have shown (4.4).

By (4.2)–(4.4) we have shown  $Z_n^* \xrightarrow{d^*} Z$  in probability. □



### 4.2. General empirical process

For (smooth) functionals, more general than in Section 4.1, depending on the  $q$ -dimensional marginal distribution of the underlying real-valued process  $\{X_t\}_{t \in \mathbb{Z}}$ , one would like to have consistency of the smoothed sieve bootstrap for empirical processes. We consider empirical processes based on the vectorized samples  $\{\mathbf{X}_t = (X_t, \dots, X_{t+q-1})\}_{t=1}^{n-q+1}$  and  $\{\mathbf{X}_t^* = (X_t^*, \dots, X_{t+q-1}^*)\}_{t=1}^{n-q+1}$ , respectively; see Bühlmann (1994). We then need the  $\nu$ -mixing property with respect to classes of functions  $\mathbb{R}^q \rightarrow \mathbb{R}$  and follow closely the approach in Andrews and Pollard (1994), who consider empirical processes for strong-mixing, stationary processes.

Let  $\mathcal{F}^q$  be a class of measurable functions from  $\mathbb{R}^q \rightarrow \mathbb{R}$ . We now introduce some notation and terminology. Denote by  $P$  a probability measure on  $(\mathbb{R}^q, \mathcal{B}(\mathbb{R}^q))$ ,  $\mathcal{B}(\mathbb{R}^q)$  the Borel  $\sigma$ -algebra of  $\mathbb{R}^d$ , and let  $Pf = \int f(x)dP(x)$  for  $f \in \mathcal{F}^q$ . Furthermore, we denote by  $\Rightarrow$  weak convergence in the function space  $l^\infty(\mathcal{F}^q)$  (in the Hoffmann–Jørgensen sense; see Giné and Zinn 1990) for the metric induced by  $\|\cdot\|_{\mathcal{F}^q}$ ; here  $\|h\|_{\mathcal{F}^q} = \sup_{f \in \mathcal{F}^q} |h(f)|$ , where  $h: \mathcal{F}^q \rightarrow \mathbb{R}$ . We restrict ourselves to uniformly bounded classes  $\mathcal{F}^q$  which satisfy a bracketing condition in the following sense. Let  $\rho(f) = \|f\|_2 = (E|f(\mathbf{X}_1)|^2)^{1/2}$  be a pseudo-norm in  $\mathcal{F}^q$  and denote by  $N(\cdot) = N(\cdot; \mathcal{F}^q, \rho)$  the bracketing number, which is defined as

$$N(\delta) = \min_M \{ \exists f_1(\delta), \dots, f_M(\delta) \in \mathcal{F}^q \text{ and } b_1(\delta), \dots, b_M(\delta) \text{ with } \rho(b_i(\delta)) \leq \delta \forall i \text{ such that:}$$

$$\forall f \in \mathcal{F}^q \exists i \text{ for which } |f - f_i(\delta)| \leq b_i(\delta) \}.$$

Note that the bracketing functions  $b_i(\delta)$  need not belong to  $\mathcal{F}^q$ . A bracketing condition now assumes a certain decay of  $N(\delta)$  as a function of  $\delta$ .

We study here the smoothed sieve bootstrapped empirical process. The empirical process  $\{Z_n(f)\}_{f \in \mathcal{F}^q}$  is defined by

$$Z_n(f) = (n - q + 1)^{1/2}(P_n(f) - Pf), \quad P_n(f) = (n - q + 1)^{-1} \sum_{t=1}^{n-q+1} \delta_{\mathbf{X}_t}(f),$$

where  $\delta_{\mathbf{x}}$  denotes the point mass at  $\mathbf{x} \in \mathbb{R}^q$ . Its smoothed sieve bootstrapped counterpart  $\{Z_n^*(f)\}_{f \in \mathcal{F}^q}$  is defined by

$$Z_n^*(f) = (n - q + 1)^{1/2}(P_n^*(f) - E^*[P_n^*(f)]), \quad P_n^*(f) = (n - q + 1)^{-1} \sum_{t=1}^{n-q+1} \delta_{\mathbf{X}_t^*}(f).$$

For the purpose of applying the results on bounding  $\nu^*(k; \cdot, \cdot, \cdot)$ -mixing coefficients, we assume:

- (A6) Let  $\mathcal{H}^q = \mathcal{F}^q \cup \tilde{\mathcal{F}}^q \cup \mathcal{B}^q$ , where  $\tilde{\mathcal{F}}^q = \{f_1 - f_2; f_1, f_2 \in \mathcal{F}^q\}$  and  $\mathcal{B}^q = \{b_1(\delta), \dots, b_M(\delta); 0 < \delta < 1\}$  is the class of bracketing functions. Assume that the extension  $\mathcal{H}_{\text{ext},0}$  satisfies (3.1) for some  $\lambda > 0$  and for  $Q$ , the marginal distribution of  $(X_1, \dots, X_q)$  from the process  $\{X_t\}_{t \in \mathbb{Z}}$  defined by (A1) and (A2). Moreover, assume that every  $f \in \mathcal{F}^q$  has at most countably many discontinuities.

**Lemma 4.3 (Stochastic equicontinuity).** Assume that (A1)–(A4) and (A6) hold with  $\beta > 1$  in (A1) and  $s \geq 4$  in (A2). In addition, assume that the parameters  $\beta$  in (A1),  $s$  in (A2),  $\lambda$  in (A6) and the dimension  $q$  are such that

$$2r - \gamma^* \frac{S\lambda C}{(s(1 + 2\lambda + 4rq - 2q) + 4rq - 2q)(2r + C)} < 1,$$

where  $r \in \mathbb{N}$ ,  $C > 0$  and  $\gamma^*$  is as in Theorem 3.2, depending on  $\beta$ . Moreover, assume that

$$\int_0^1 x^{-C/(2+C)} N(x; \mathcal{F}^q, \rho^*)^{1/(2r)} dx < \infty$$

for the same  $r$  and  $C$  and with  $\rho^*(f) = (\mathbb{E}^*|f(\mathbf{X}_1^*)|^2)^{1/2}$ . Then for all  $\eta > 0$  there exists a  $\delta > 0$  such that

$$\limsup_{n \rightarrow \infty} (\mathbb{E}^* \mid \sup_{\rho(f-g) < \delta} |Z_n^*(f) - Z_n^*(g)|^{2r})^{1/(2r)} < \eta \text{ in probability.}$$

**Remark 4.4.** Lemma 4.3 covers only ‘parametric’ function classes having bracketing numbers  $N(x; \mathcal{F}^q, \rho^*)$  which increase at most polynomially in  $1/x$  as  $x \rightarrow 0$ . Note also that the result does not hold uniformly in  $r$ .

**Remark 4.5.** The metric entropy condition in Lemma 4.3 is with respect to the bootstrap pseudo-norm  $\rho^*(\cdot)$ . We now give an important example which relates it directly to the metric entropy  $N(\cdot; \mathcal{F}^q, \rho)$ .

**Example 4.3** Indicator functions of intervals in  $\mathbb{R}^q$ . Consider the class  $\mathcal{F}^q = \{1_{(-\infty, \mathbf{x}]}; \mathbf{x} \in \mathbb{R}^q\}$  for  $q \geq 1$ . Notation is as in Section 4.3. Denote by  $F^{(1)*}(x) = \mathbb{P}^*[X_1^* \leq x]$  and let  $U_t^* = F^{(1)*}(X_t^*)$  ( $t \in \mathbb{Z}$ ) and  $\mathbf{U}_t^* = (U_t^*, \dots, U_{t+q-1}^*)$ . Then  $U_t^* \sim \text{Unif}([0, 1])$ . Denote for  $\mathbf{u} \in [0, 1]^q$ ,

$$\tilde{Z}_n^*(\mathbf{u}) = n^{-1/2} \sum_{t=1}^{n-q+1} (1_{[\mathbf{U}_t^* \leq \mathbf{u}]} - \mathbb{P}^*[\mathbf{U}_1^* \leq \mathbf{u}]).$$

Note that  $\mathbb{P}^*$ -a.s.,

$$\tilde{Z}_n^*(\mathbf{u}) = Z_n^*(\mathbf{x}) = n^{-1/2} \sum_{t=1}^{n-q+1} (1_{[\mathbf{X}_t^* \leq \mathbf{x}]} - \mathbb{P}^*[\mathbf{X}_1^* \leq \mathbf{x}]), \quad \mathbf{x}_i = (F^{(1)*})^{-1}(\mathbf{u}_i), \quad i \in \{1, \dots, q\}.$$

As also pointed out in the proof of Corollary 4.1 in Section 4.3, it is sufficient to work with the empirical process  $\tilde{Z}_n^*(\cdot)$  and the function class  $\tilde{\mathcal{F}}^q = \{1_{(-\infty, \mathbf{u}]}; \mathbf{u} \in [0, 1]^q\}$ . For the pseudo-norm,  $\tilde{\rho}^*(\mathbf{u}) = (\mathbb{E}^*|1_{[\mathbf{U}_1^* \leq \mathbf{u}]}|^2)^{1/2} = (\mathbb{P}^*[\mathbf{U}_1^* \leq \mathbf{u}])^{1/2} \leq \min_{1 \leq i \leq q} (\mathbf{u}_i)^{1/2}$  since  $U_t^* \sim \text{Unif}([0, 1])$  for all  $t$ . This yields the metric entropy bound  $N(x; \tilde{\mathcal{F}}^q, \tilde{\rho}^*) \leq \text{const.}x^{-2q}$  which is essentially the same as for  $N(x; \tilde{\mathcal{F}}^q, \tilde{\rho})$  with  $\tilde{\rho}(\mathbf{u}) = (\mathbb{E}|1_{[\mathbf{U}_1 \leq \mathbf{u}]}|^2)^{1/2}$ .

**Proof of Lemma 4.3.** We use the  $\nu$ -mixing property of  $\{X_t^*\}_{t \in \mathbb{Z}}$  with respect to the pair  $(\otimes_{i=1}^{2r-1} \mathcal{F}^q, \otimes_{i=1}^{2r-1} \tilde{\mathcal{F}}^q)$  (see Theorem 3.2) and follow the proof of Theorem 2.2 in Andrews

and Pollard (1994). In particular, we make use of our Lemma 4.2(ii). First, we work with  $\rho^*(f) = (\mathbb{E}^*|f(\mathbf{X}_1^*)|^2)^{1/2}$  and then use the fact that

$$\sup_{f \in \mathcal{F}^q} |\rho^*(f)^2 - \rho(f)^2| = o_P(1).$$

This inequality holds since  $\mathbf{X}_1^* \xrightarrow{d^*} \mathbf{X}_1$  in probability (cf. Lemma 5.5 below) and, by (A6),  $\mathcal{F}^q$  is a uniformity class; see Theorem 2.4 of Bhattacharya and Ranga Rao (1976).  $\square$

Under the conditions of Lemma 4.3 but with the metric entropy condition with respect to  $\rho$ , the empirical process  $Z_n(\cdot)$  converges weakly to some Gaussian process  $Z(\cdot)$ , indexed by  $\mathcal{F}^q$ , with  $\rho$ -continuous sample paths and with  $\mathbb{E}[Z(f)] = 0$ ,  $f \in \mathcal{F}^q$  and

$$\text{cov}(Z(f), Z(g)) = \sum_{k=-\infty}^{\infty} \text{cov}(f(X_0), g(X_k));$$

see Corollary 2.3 of Andrews and Pollard (1994). We assume below that weak convergence of  $Z_n(\cdot)$  to  $Z(\cdot)$  holds.

In the following we sometimes make statements about weak convergence, holding in probability in a universal sense over all  $f \in \mathcal{F}^q$ . Let  $R_n^*(f)$  be a random variable with respect to the bootstrap measure  $\mathbb{P}^*$  and  $R(f)$  a random variable of the underlying original probability space. We say that

$$R_n^*(f_1), \dots, R_n^*(f_h) \xrightarrow{d^*} R(f_1), \dots, R(f_h) \text{ in probability universal over } \mathcal{F}^q$$

if the following holds. For every continuity point  $\mathbf{x} \in \mathbb{R}^h$  of the distribution of  $(R(f_1), \dots, R(f_h))$ ,  $\forall \eta > 0$  there exist an  $n_0 = n_0(\eta)$  and a sequence  $\{A_n\}_{n \in \mathbb{N}}$  of (universal) sets such that

$$|\mathbb{P}^*[(R_n^*(f_1), \dots, R_n^*(f_h)) \leq \mathbf{x}] - \mathbb{P}[(R(f_1), \dots, R(f_h)) \leq \mathbf{x}]| \leq \eta \text{ on the set } A_n, \forall n \geq n_0,$$

$$\mathbb{P}[A_n] \rightarrow 1 \text{ (} n \rightarrow \infty \text{),}$$

where for each  $n \in \mathbb{N}$ , the set  $A_n$  is universal  $\forall f_1, \dots, f_h \in \mathcal{F}^q$ ,  $h \in \mathbb{N}$  and  $\mathbf{x} \leq \mathbf{y}$  is defined componentwise as  $x_i \leq y_i$  ( $i = 1, \dots, h$ ).

**Theorem 4.2.** *Assume the conditions of Lemma 4.3. Moreover, assume finite-dimensional convergence*

$$(Z_n^*(f_1), \dots, Z_n^*(f_h)) \xrightarrow{d^*} (Z(f_1), \dots, Z(f_h)) \text{ in probability universal over } \mathcal{F}^q.$$

Then

$$Z_n^* \Rightarrow Z \text{ in probability.}$$

**Proof.** The result follows directly from finite-dimensional convergence and Lemma 4.3.  $\square$

Finite-dimensional convergence of  $Z_n^*$  is usually not directly available because  $\{X_t^*\}_{t \in \mathbb{Z}}$

satisfies by Theorem 3.2 only a  $\nu$ -mixing property. This does not allow us to use one of the usual blocking techniques.

**Theorem 4.3.** *Assume the conditions of Lemma 4.3. Then*

$$Z_n^* \Rightarrow Z \text{ in probability.}$$

**Proof.** It remains to show finite-dimensional convergence

$$(Z_n^*(f_1), \dots, Z_n^*(f_h)) \xrightarrow{d^*} (Z(f_1), \dots, Z(f_h)) \text{ in probability universal over } \mathcal{F}^q.$$

But this follows from a multivariate version of Theorem 4.1 which can be established by the Cramér–Wold device. The conditions of Theorem 4.1 are implied by the current assumptions.  $\square$

We just remark that by replacing (A1), (A3) and (A4) by (A1’), (A3’) and (A4’) respectively, we obtain better bounds on the  $\nu$ -mixing coefficients and hence need fewer conditions on the bracketing numbers.

### 4.3. Empirical process on $\mathbb{R}^q$

We now specialize our results from Section 4.2 to the classical empirical process on  $\mathbb{R}^q$ ,  $q \in \mathbb{N}$ , based on the vectorized observations  $\{\mathbf{X}_t\}_{t=1}^{n-q+1}$  and  $\{\mathbf{X}_t^*\}_{t=1}^{n-q+1}$ , respectively. That is,  $\mathcal{F}^q = \{1_{(-\infty, \mathbf{x}]}; \mathbf{x} \in \mathbb{R}^q\}$ , where  $(-\infty, \mathbf{x}] = \times_{i=1}^q (-\infty, \mathbf{x}_i]$ . By Examples 3.1 and 4.2 we know that (A6) holds for  $\mathcal{H}^q$  with  $\lambda = 1$ , if the  $q$ -dimensional marginal distribution of the process  $\{X_t\}_{t \in \mathbb{Z}}$  has a bounded density. Note that the bracketing class  $\mathcal{B}^q$  can be taken again as a class of indicator functions. Denote the cdf of  $\mathbf{X}_t$  and  $\mathbf{X}_t^*$  by  $F^{(q)}(\cdot)$  and  $F^{(q)*}(\cdot)$ , respectively. Defining ‘ $\leq$ ’ componentwise, the empirical process and its bootstrap counterpart can then be written as

$$Z_n(\mathbf{x}) = (n - q + 1)^{-1/2} \sum_{t=1}^{n-q+1} (1_{[X_t \leq \mathbf{x}]} - F^{(q)}(\mathbf{x})), \quad \mathbf{x} \in \mathbb{R}^q,$$

$$Z_n^*(\mathbf{x}) = (n - q + 1)^{-1/2} \sum_{t=1}^{n-q+1} (1_{[X_t^* \leq \mathbf{x}]} - F^{(q)*}(\mathbf{x})), \quad \mathbf{x} \in \mathbb{R}^q.$$

**Corollary 4.1.** *Assume that (A1)–(A4) hold with  $\beta > 1$  in (A1),  $s \geq 4$  in (A2) and  $\sup_{\mathbf{x} \in \mathbb{R}^q} f_\varepsilon(\mathbf{x}) < \infty$ . In addition, assume that the parameters  $s$  in (A2) and the dimension  $q$  are such that*

$$2r - \gamma^* \frac{s(\frac{r}{q} - 1)}{(s(3 + 4rq - 2q) + 4rq - 2q)(r + \frac{r}{q} - 1)} < 1 \text{ for some } r \geq q + 1, r \in \mathbb{N},$$

with  $\gamma^*$  as in Theorem 3.2, depending on  $\beta$ .

Then

$$Z_n^* \Rightarrow Z \text{ in probability,}$$

where  $Z$  is the limiting Gaussian process of  $Z_n$  with mean zero and

$$\text{cov}(Z(\mathbf{x}), Z(\mathbf{y})) = \sum_{k=-\infty}^{\infty} \text{cov}(1_{[\mathbf{x}_0 \leq \mathbf{x}]}, 1_{[\mathbf{x}_k \leq \mathbf{y}]})$$

**Remark 4.6.** Under a geometric decay for  $|\phi_j|$  as  $j \rightarrow \infty$  in (A1), Corollary 4.1 holds for all  $q \in \mathbb{N}$ . The choice  $r = q + 1$  is sufficient, since  $\gamma^*$  can be chosen arbitrary large.

**Remark 4.7.** The region for the parameters  $q, s, \beta$  satisfying the conditions in Corollary 4.1 can be identified. The condition on  $q, s, \gamma^*$  is equivalent to  $v(r; q, s, \gamma^*) < 0$  for some  $r \geq q + 1, r \in \mathbb{N}$ , where  $v(r; q, s, \gamma^*)$  is a cubic polynomial in  $r$ , namely

$$v(r; q, s, \gamma^*) = r^3 \{16 + 16q + 16s + 16qs\} + r^2 \left\{ -16 - 32q - 4s - 32qs + \frac{12s}{q} \right\} + r \left\{ 4 + 20q - 14s + 20qs - \frac{6s}{q} - \frac{2\gamma^*}{q} \right\} - 4q + 6s + 2\gamma^* - 4qs.$$

The roots  $v_1 \leq v_2 \leq v_3$  of the equation  $v(r; q, s, \gamma^*) = 0$  can be computed in closed form. Since  $\gamma^* = \gamma^*(s, \beta)$ , these roots are of the form  $v_i = v_i(q, s, \beta), i = 1, 2, 3$ . The condition in Corollary 4.1 on  $q, s, \gamma^*$ , or equivalently on  $q, s, \beta$  is then implied by  $v_1(q, s, \beta) > q + 1$ .

**Proof of Corollary 4.1.** The result is basically a consequence of Theorem 4.3. Consider first  $\tilde{Z}_n^*(\cdot)$  as defined in Example 4.3. Since for  $\mathbf{u}, \mathbf{v} \in [0, 1]^q$ ,

$$\text{cov}^*(1_{[U_0^* \leq \mathbf{u}]}, 1_{[U_k^* \leq \mathbf{v}]} ) = \text{cov}^*(1_{[X_0^* \leq \{(F^{(1)*})^{-1}(\mathbf{u}_i)\}_{i=1}^q]} , 1_{[X_k^* \leq \{(F^{(1)*})^{-1}(\mathbf{v}_i)\}_{i=1}^q]} ),$$

the process  $(U_t^*)_{t \in \mathbb{Z}}$  is  $\nu$ -mixing with respect to  $\tilde{\mathcal{F}}^q$  with coefficients bounded by the mixing coefficients of  $(X_t^*)_{t \in \mathbb{Z}}$  with respect to  $\mathcal{F}^q$  (we use the same notation as in Example 4.3). Note that the assumption  $\sup_{x \in \mathbb{R}} f_\varepsilon(x) < \infty$  implies that the  $q$ -dimensional marginal distribution of  $(X_t, \dots, X_{t+q})$  has a bounded density and hence  $\lambda = 1$  in assumption (A6) which is needed to control the  $\nu^*$ -mixing behaviour of  $(X_t^*)_{t \in \mathbb{Z}}$ . Moreover, as noted in Example 4.3,  $N(x, \tilde{\mathcal{F}}^q, \tilde{\rho}^*) \leq \text{const} \cdot x^{-2q}$ . Thus, by using  $r > q(2 + C)/2$  ( $r \in \mathbb{N}$ ), the bracketing condition in Lemma 4.3 holds. (We then use  $\lambda = 1$  (see above) and the optimal  $C = 2r/q - 2 - \eta$  ( $\eta > 0$  arbitrarily small) in Lemma 4.3.) Therefore

$$\tilde{Z}_n^* \Rightarrow \tilde{Z} \text{ in probability,} \tag{4.7}$$

where  $\tilde{Z}(\cdot)$  is a Gaussian process with  $E[\tilde{Z}(\mathbf{u})] = 0$  and  $\text{cov}(\tilde{Z}(\mathbf{u}), \tilde{Z}(\mathbf{v})) = \sum_{k=-\infty}^{\infty} \text{cov}(1_{[\{F^{(1)}(X_i)\}_{i=0}^{q-1} \leq \mathbf{u}]}, 1_{[\{F^{(1)}(X_i)\}_{i=k}^{k+q-1} \leq \mathbf{v}]}) (\mathbf{u}, \mathbf{v} \in [0, 1]^q)$ . Here we have used the fact that

$$\sup_{x \in \mathbb{R}} |F^{(1)*}(x) - F^{(1)}(x)| = o_P(1), \tag{4.8}$$

by continuity of  $F^{(1)}(\cdot)$ .

It remains to show the convergence  $Z_n^*(\cdot) \Rightarrow Z(\cdot)$ . This follows by applying the continuous mapping theorem to the continuous maps,

$$H^* : \mathcal{D}([0, 1]^q) \rightarrow \mathcal{D}(\mathbb{R}^q), z \mapsto z \circ (F^{(1)*}, \dots, F^{(1)*}),$$

$$H : \mathcal{D}([0, 1]^q) \rightarrow \mathcal{D}(\mathbb{R}^q), z \mapsto z \circ (F^{(1)}, \dots, F^{(1)}),$$

where  $\mathcal{D}$  denotes the cadlag space.

By (4.8) we get  $\|H^*(z) - H(z)\|_{\mathcal{D}^q} = o_P(1)$  for any continuous  $z \in \mathcal{C}([0, 1]^q)$ . Using this and (4.7) we can easily show  $H^*(\tilde{Z}_n^*) \Rightarrow H(\tilde{Z})$  in probability. The proof is completed by noting that  $H(\tilde{Z})$  has the same distribution as  $Z(\cdot)$ . □

**Proof of Corollary 2.1.** We invoke the much stronger result in Theorem 3.4 about the geometric  $\nu^*$  mixing property of the sieve bootstrapped process. Conditions (A1'), (A2) with  $s \geq 4$ , (A3'), (A4') (see the discussion about the choice of  $p(n)$  and  $h(n)$  following assumptions (A1'), (A2') and (A4') in Section 3.2) and (A5) with  $\lambda = 1$  hold. Now the proof follows as for Corollary 4.1; we can replace  $\gamma^*$  by an arbitrary large value. The result follows for any parameter value of  $s \geq 4$ . □

## 5. Proofs

In what follows we denote by  $\mathcal{B}(S)$  the Borel  $\sigma$ -algebra of a metric space  $S$ . We first outline the idea for proving Theorem 3.2. The same idea is used for proving Theorem 3.4. The strategy is to split the problem into two cases with small and large separation lags  $k$ .

If  $k$  is large (or arbitrary), we use Gorodetskii's (1977) result by exploiting the linear representation (2.2) and the fact that  $\varepsilon_i^*$  i.i.d.  $\sim \hat{f}_\varepsilon(x + \hat{\mu}_\varepsilon)dx$ . We will show in Lemma 5.3 that  $\alpha^*(k) \leq \text{const.}h(n)^{-1}k^{-\gamma^*}$  in probability, yielding for  $k \geq h(n)^{-1/\zeta}$ ,  $\zeta \in \mathbb{R}^+$ ,

$$\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \alpha^*(k) \leq \text{const.}k^{-(\gamma^* - \zeta)} \text{ in probability.} \tag{5.1}$$

On the other hand, we first use the general fact that

$$\begin{aligned} \nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) &\leq \alpha(k) \\ &+ \sup \left\{ \frac{|\text{cov}^*(g_1(X_{-d_1+1}^*, \dots, X_0^*), g_2(X_k^*, \dots, X_{k+d_2-1}^*))|}{4\|g_1\|_\infty\|g_2\|_\infty} \right. \\ &\quad \left. - \frac{\text{cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))}{4\|g_1\|_\infty\|g_2\|_\infty} \right\}, \end{aligned}$$

where the supremum is over all  $g_1 \in \mathcal{E}^{d_1}$ ,  $g_2 \in \mathcal{D}^{d_2}$ .

For bounding the difference of the covariances we now introduce a moment (pseudo-) norm

$$\|Q_1 - Q_2\|_{\mathcal{E}^{d_1}, \mathcal{D}^{d_2}} = \sup \left\{ \left| \frac{\int_{\mathbb{R}^{d_1+d_2}} g(\mathbf{x})(dQ_1 - dQ_2)(\mathbf{x})}{\|g\|_\infty} \right|; g \in \mathcal{E}^{d_1} \otimes \mathcal{D}^{d_2} \right\}, \tag{5.2}$$

where  $Q_1, Q_2$  are probability measures on  $(\mathbb{R}^{d_1+d_2}, \mathcal{B}(\mathbb{R}^{d_1+d_2}))$  (for the definition of  $\mathcal{E} \otimes \mathcal{D}$  see (4.1)).

Using the fact that for  $g = g_1 \cdot g_2$ ,  $\|g\|_\infty = \|g_1\|_\infty \|g_2\|_\infty$ , the difference of covariances can now be bounded as

$$\begin{aligned} & \frac{1}{\|g_1\|_\infty \|g_2\|_\infty} |\text{cov}^*(g_1(X_{-d_1+1}^*, \dots, X_0^*), g_2(X_k^*, \dots, X_{k+d_2-1}^*)) \\ & \quad - \text{cov}(g_1(X_{-d_1+1}, \dots, X_0), g_2(X_k, \dots, X_{k+d_2-1}))| \\ & = \frac{1}{\|g_1\|_\infty \|g_2\|_\infty} \left| \int_{\mathbb{R}^{d_1+d_2}} g_1 \cdot g_2(\mathbf{x})(d\mathbb{P}^* - d\mathbb{P})(\mathbf{x}) \right. \\ & \quad \left. - \int_{\mathbb{R}^{d_1}} g_1(\mathbf{x})d\mathbb{P}^*(\mathbf{x}) \int_{\mathbb{R}^{d_2}} g_2(\mathbf{x})d\mathbb{P}^*(\mathbf{x}) + \int_{\mathbb{R}^{d_1}} g_1(\mathbf{x})d\mathbb{P}(\mathbf{x}) \int_{\mathbb{R}^{d_2}} g_2(\mathbf{x})d\mathbb{P}(\mathbf{x}) \right| \\ & \leq \|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{E}^{d_1}, \mathcal{D}^{d_2}} + 2\|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{E}^{d_1}, \mathcal{D}^{d_2}}. \end{aligned}$$

This means that we bound

$$v^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \alpha(k) + 3\|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{E}^{d_1}, \mathcal{D}^{d_2}}. \tag{5.3}$$

In Lemma 5.5 we will give the bound  $\|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{E}^{d_1}, \mathcal{D}^{d_2}} = O_P(b(n))$ , where  $b(n)$  is a function of the tuning parameters  $p(n)$  and  $h(n)$  and of the sample size  $n$ . In particular, under the assumptions about the bandwidth  $h(n)$  in (A4) we obtain  $\|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{E}^{d_1}, \mathcal{D}^{d_2}} = O_P(h(n)^c)$  for some  $c \in \mathbb{R}^+$ , yielding then for  $k \leq h(n)^{-1/\zeta}$ ,  $\|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{E}^{d_1}, \mathcal{D}^{d_2}} \leq \text{const} \cdot k^{-\zeta c}$  in probability and hence for  $k \leq h(n)^{-1/\zeta}$ ,

$$v^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \text{const} \cdot k^{-\zeta c} \text{ in probability.} \tag{5.4}$$

Putting (5.1) and (5.4) together, we minimize over  $\zeta$ .

We now give some preliminary results. The first has to do with moving-average representations of autoregressive approximations. We recall the definition for the coefficients  $\{\hat{\psi}_{j,n}\}_{j=0}^\infty$ , which arise by inverting the estimated autoregressive transfer function, (compare with (2.1) and (2.2)).

**Lemma 5.1.** *Assume that model (1.2) holds with  $\varepsilon_t$  iid.,  $E[\varepsilon_t] = 0$ ,  $E|\varepsilon_t|^4 < \infty$ . Suppose that  $\Phi(z)$  is bounded away from zero for  $|z| \leq 1$  ( $z \in \mathbb{C}$ ) (see (A1)),  $\sum_{j=0}^\infty j^r |\phi_j| < \infty$  and  $p(n) = o((n/\log(n))^{1/(2r+2)})$ ,  $r \in \mathbb{N}$ . Then:*

(i) *there exists a random variable  $n_0(\omega)$  such that*

$$\sup_{n \geq n_0(\omega)} \sum_{j=0}^\infty j^r |\hat{\psi}_{j,n}| < \infty \text{ almost surely;}$$

(ii) for  $a(n) \rightarrow \infty, a(n) = o(n) (n \rightarrow \infty)$ ,

$$\sup_{s \in \mathbb{N}_0} \sum_{j=s+1}^{s+a(n)} |\hat{\psi}_{j,n} - \psi_j| = O(a(n)p(n)^{-r}) + O(a(n)(\log(n)/n)^{1/2}) \text{ almost surely.}$$

**Proof.** Assertion (i) is Theorem 3.1 in Bühlmann (1995). Assertion (ii) follows from Theorem 3.2 in Bühlmann (1995). □

**Lemma 5.2.** *Assume the conditions of Lemma 5.1, and in addition that  $E|\varepsilon_t|^s < \infty, s \geq 4$ . Suppose that the kernel  $K(\cdot)$  for estimating  $f_\varepsilon(\cdot)$  is a probability density and satisfies  $\int_{-\infty}^\infty xK(x)dx = 0, \int_{-\infty}^\infty x^2K(x)dx \neq 0, \int_{-\infty}^\infty |x|^sK(x)dx < \infty$  for the same  $s$ , and the bandwidth satisfies  $h(n) \rightarrow 0, h(n)^{-1} = o(n) (n \rightarrow \infty)$ . Then:*

- (i)  $E^*[(\varepsilon_t^*)^w] - E[(\varepsilon_t)^w] = O_P(h(n)^2) + O_P(p(n)(\log(n)/n)^{1/2}) + o_P(p(n)^{-r}), w \leq s;$
- (ii)  $E^*|\varepsilon_t^*|^s = O_P(1).$

**Proof.** We have

$$\begin{aligned} E^*[(\varepsilon_t^*)^w] &= \int_{-\infty}^\infty x^w \hat{f}_\varepsilon(x + \hat{\mu}_\varepsilon) dx = (n - p)^{-1} \sum_{t=p+1}^n \int_{-\infty}^\infty (hu + \hat{\varepsilon}_{t,n} - \hat{\mu}_\varepsilon)^w K(u) du \\ &= (n - p)^{-1} \sum_{t=p+1}^n (\hat{\varepsilon}_{t,n})^w + O_P(\hat{\mu}_\varepsilon + h(n)^2). \end{aligned} \tag{5.5}$$

We write

$$\hat{\varepsilon}_{t,n} = \varepsilon_t + Q_{t,n} + R_{t,n} - (\bar{X} - \mu_X) \sum_{j=0}^\infty \phi_j, \tag{5.6}$$

where  $Q_{t,n} = \sum_{j=0}^p (\hat{\phi}_{j,n} - \phi_{j,n})(X_{t-j} - \bar{X}), R_{t,n} = \sum_{j=0}^\infty (\phi_{j,n} - \phi_j)(X_{t-j} - \bar{X})$ . Here  $\phi_p = (\phi_{1,n}, \dots, \phi_{p,n})'$  are the solutions of the theoretical Yule–Walker equations  $\Gamma_p \phi_p = -\gamma_p$ ; see Brockwell and Davis (1987, Chapter 8.1). Now in a similar manner to the proof of Lemma 5.3 in Bühlmann (1997, pp. 137–138),

$$\begin{aligned} |Q_{t,n}| &\leq \max_{0 \leq j \leq p(n)} |\hat{\phi}_{j,n} - \phi_{j,n}| \sum_{j=0}^{p(n)} |X_{t-j} - \bar{X}| \\ &= O((\log(n)/n)^{1/2}) \sum_{j=0}^p |X_{t-j} - \bar{X}|, \text{ the } O\text{-term being a.s.,} \end{aligned} \tag{5.7}$$

(see Theorem 2.1 of Hannan and Kavalieris 1986), and

$$E|R_{t,n}|^w \leq \text{const.} \left( \sum_{j=p+1}^\infty |\phi_j| \right)^w = o(p(n)^{-wr}), \tag{5.8}$$



where we have used Baxter’s inequality (see the proof of (3.1) in Bühlmann 1995). Since  $\hat{\mu}_\varepsilon = (n - p)^{-1} \sum_{t=p+1}^n \hat{\varepsilon}_{t,n}$  we complete the proof by using (5.5)–(5.8) and applying a binomial expansion for  $(\hat{\varepsilon}_{t,n})^w$ .

Assertion (ii) follows immediately by using the representation as in (5.5). □

### 5.1. Mixing property for large separation lags

**Lemma 5.3.** *Assume that (A1)–(A4) hold with  $\beta > 1$  in (A1) and  $s \geq 4$  in (A2). Then*

$$\mathbb{P}[\alpha^*(k) \leq \text{const.} h(n)^{-1} k^{-\gamma^*} \forall k \in \mathbb{N}] \rightarrow 1 \quad (n \rightarrow \infty),$$

where  $\gamma^*$  is defined as in Theorem 3.2.

**Proof.** We use representation (2.2) and use the result under existence of  $s$ th moments in Gorodetskii (1977). His condition (i) follows immediately by (A4), (ii) follows by Lemma 5.2 and (iii) by (A1) and (A3); see Lemma 2.2 and Theorem 3.1 in Bühlmann (1995). The constant  $\gamma^*$  shows up by using Lemma 5.1(i) (note that this lemma handles only  $r \in \mathbb{N}$ ). □

We remark here that Lemma 5.3 holds true if we weaken the assumptions on the bandwidth  $h(n)$  in (A4) to the sole condition  $h(n) = o(1)$ ,  $h(n)^{-1} = o(n)$  ( $n \rightarrow \infty$ ). Bounding bootstrap covariances for non-smooth functions of shortly lagged variables cannot be done effectively with Lemma 5.3. An analysis such as that in the following Section 5.2 completes the establishment of useful bounds for short lags.

### 5.2. Moment norm between bootstrap and true measure

Denote by  $\mathbb{P}_{k;d_1,d_2}[C] = \mathbb{P}[(X_{-d_1+1}, \dots, X_0, X_k, \dots, X_{k+d_2-1}) \in C]$ ,  $C \in \mathcal{B}(\mathbb{R}^d)$ ,  $d = d_1 + d_2$ ,  $k \in \mathbb{N}$ . We define  $\mathbb{P}_{k;d_1,d_2}^*$  analogously for the bootstrap. By the definition of the  $\nu$ -mixing coefficients and the boundedness of  $g_1 \in \mathcal{C}^{d_1}$ ,  $g_2 \in \mathcal{D}^{d_2}$  from above and below we have, cf. (5.3),

$$\nu^*(k; \mathcal{C}^{d_1}, \mathcal{D}^{d_2}) \leq \alpha(k) + 3 \|\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{C}^{d_1}, \mathcal{D}^{d_2}}, \quad \|\cdot\|_{\mathcal{C}^{d_1}, \mathcal{D}^{d_2}} \text{ as in (5.2).}$$

Our next aim is to bound

$$\sup_{k \in \mathbb{N}} \|\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{C}^{d_1}, \mathcal{D}^{d_2}}.$$

To do so we will compare this quantity with the variational norm of a ‘smoothed difference’  $\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}$ . The variational norm for a probability measure  $Q$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  is defined as

$$\|Q\|_{V;d} = 2 \sup_{C \in \mathcal{B}(\mathbb{R}^d)} |Q[C]|.$$

In the following we denote by  $Q_1 \star Q_2$  the convolution of some signed measures  $Q_1$  and  $Q_2$ .

**Lemma 5.4 (Berry’s smoothing lemma)** . Let  $\delta(n) = o(1)$  ( $n \rightarrow \infty$ ) and  $\{K_{\delta(n)}\}_{n \in \mathbb{N}}$  be a sequence of probability measures on  $\mathbb{R}^d$  with  $\sup_{n \in \mathbb{N}} K_{\delta(n)}(\{\|\mathbf{x}\| \leq \delta(n)\}) > \frac{1}{2} \forall n \in \mathbb{N}$ ,  $\|\cdot\|$  the Euclidean norm in  $\mathbb{R}^d$ . Assume that (A5) holds. Then  $\forall n \in \mathbb{N}$ ,

$$\sup_{k \in \mathbb{N}} \|\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{G}^{d_1}, \mathcal{G}^{d_2}} \leq \text{const.} \sup_{k \in \mathbb{N}} \|(\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}) \star K_{\delta(n)}\|_{V;d} + \text{const.} \delta(n)^\lambda,$$

where  $\lambda = \min\{\lambda_1, \lambda_2\}$ ,  $d = d_1 + d_2$ .

**Proof.** We use formula (11.26) in Bhattacharya and Ranga Rao (1976). To bound the moment norm we need a bound for some type of modulus of oscillation; but our assumption (A5) is exactly tailored to this problem. The quantity to be bounded is

$$R = \sup_{k \in \mathbb{N}} \sup_{g \in \mathcal{G}^{d_1} \otimes \mathcal{G}^{d_2}} \sup_{\mathbf{y} \in \mathbb{R}^d} \int_{\mathbb{R}^d} \frac{\omega_{g_y}(B(\mathbf{x}, \delta))}{\|g\|_\infty} d\mathbb{P}_{k;d_1,d_2}(\mathbf{x}) \quad (d = d_1 + d_2).$$

Denote by  $g = g_1 \cdot g_2 \in \mathcal{G}^{d_1} \otimes \mathcal{G}^{d_2}$ ,  $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)' \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$  and  $Q_i$  the distributions of  $(X_1, \dots, X_{d_i})$  ( $i = 1, 2$ ) as defined in (A5). Then, since  $\|g\|_\infty = \|g_1\|_\infty \|g_2\|_\infty$ ,

$$\begin{aligned} & \int_{\mathbb{R}^d} \frac{\omega_{g_y}(B(\mathbf{x}, \delta))}{\|g\|_\infty} d\mathbb{P}_{k;d_1,d_2}(\mathbf{x}) \\ & \leq \frac{\|g_2\|_\infty}{\|g_1\|_\infty \|g_2\|_\infty} \int_{\mathbb{R}^{d_1}} \omega_{(g_1)_{y_1}}(B(\mathbf{x}_1, \delta)) dQ_1(\mathbf{x}_1) + \frac{\|g_1\|_\infty}{\|g_1\|_\infty \|g_2\|_\infty} \int_{\mathbb{R}^{d_2}} \omega_{(g_2)_{y_2}}(B(\mathbf{x}_2, \delta)) dQ_2(\mathbf{x}_2). \end{aligned}$$

Hence by (A5) we bound  $R$ , as defined above, by  $\text{const.} \delta^\lambda$ , which is the second term on the right-hand side in the assertion. □

We now make use of the smoothing idea: choose  $K_{\delta(n)}$  smooth such that its Fourier transform vanishes for large arguments. Together with Berry’s lemma (Lemma 5.4) we will show:

**Lemma 5.5.** Assume that (A1)–(A5) hold with  $\beta > 1$  in (A1),  $s \geq 4$  in (A2) and  $\lambda > 0$  in (A5). Then for all  $d_1, d_2 \in \mathbb{N}$ ,

$$\sup_{k \in \mathbb{N}} \|\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{G}^{d_1}, \mathcal{G}^{d_2}} = O_P(\xi(n)^{\frac{s\lambda}{s(1+\lambda+d)+d}}), \quad d = d_1 + d_2,$$

where  $\xi(n) = \max\{h(n), p(n) (\log(n)/n)^{1/2}, p(n)^{-\vartheta^2/(\vartheta+1)}\}$ ,  $\vartheta < \beta - 1$ ,  $\vartheta \in \mathbb{N}$ . Moreover, the assumptions about the bandwidth  $h(n)$  in (A4) yield

$$\sup_{k \in \mathbb{N}} \|\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{G}^{d_1}, \mathcal{G}^{d_2}} = O_P(h(n)^{\frac{s\lambda}{s(1+\lambda+d)+d}}), \quad d = d_1 + d_2,$$

**Proof.** To simplify notation we always denote by  $\vartheta$  an integer less than  $\beta - 1$ . By Lemma 5.4 we want to bound

$$\sup_{k \in \mathbb{N}} \|(\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}) \star K_{\delta(n)}\|_{V;d} = 2 \sup_{k \in \mathbb{N}} \sup_{C \in \mathcal{B}(\mathbb{R}^d)} |(\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}) \star K_{\delta(n)}[C]|.$$

We make a similar choice of  $K_{\delta(n)}$  as in (13.8)–(13.11) in Bhattacharya and Ranga Rao (1976), that is,  $K_{\delta(n)}$  has a density

$$\prod_{i=1}^n g_{\delta(n),2s}(x_i), \quad g_{a,2m}(x) = \text{const.} \left( \frac{\sin(ax)}{ax} \right)^2 m \quad \text{a density on } \mathbb{R}.$$

Then  $\sup_{n \in \mathbb{N}} K_{\delta(n)}[\|\mathbf{x}\| \leq \delta(n)] > \frac{1}{2}$  for  $n$  large enough (this is a condition in Lemma 5.4) and for the Fourier transform of  $K_{\delta(n)}$  we have

$$\int_{\mathbb{R}^d} \exp(i\mathbf{y} \cdot \mathbf{x}) K_{\delta(n)}(d\mathbf{x}) = 0 \text{ if } \mathbf{y} \notin [-2s\delta(n)^{-1}, 2s\delta(n)^{-1}]^d, \tag{5.9}$$

where  $\mathbf{y} \cdot \mathbf{x} = \sum_{i=1}^d y_i x_i$ ; see (10.9) in Bhattacharya and Ranga Rao (1976).

In the following we let  $J(n; d) = [-2s\delta(n)^{-1}, 2s\delta(n)^{-1}]^d$ . Let  $C \in \mathcal{B}(\mathbb{R}^d)$ . Then by Fourier inversion

$$\sup_{\substack{k \in \mathbb{N} \\ k \in \mathbb{N}}} |(\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}) \star K_{\delta(n)}[C]| \leq \text{const.} \int_C \int_{J(n;d)} \sup_{k \in \mathbb{N}} |\varphi_{k;d_1,d_2}^*(\mathbf{x}) - \varphi_{k;d_1,d_2}(\mathbf{x})| d\mathbf{x} d\mathbf{y}, \tag{5.10}$$

where  $\varphi_{k;d_1,d_2}(\mathbf{x}) = E[\exp(i\mathbf{x} \cdot \mathbf{X})]$ ,  $\mathbf{X} = (X_{-d_1+1}, \dots, X_0, X_k, \dots, X_{k+d_2-1})'$ , and analogously for  $\varphi_{k;d_1,d_2}^*$ .

To bound (5.10) much of the work boils down to estimating  $|\varphi_{k;d_1,d_2}^*(\mathbf{x}) - \varphi_{k;d_1,d_2}(\mathbf{x})|$ . We use the linear representations (1.1) and (2.2) and write

$$\begin{aligned} \varphi_{k;d_1,d_2}(\mathbf{x}) &= \exp(i\mathbf{x} \cdot \mu_X \mathbf{1}) \prod_{j=0}^{\infty} \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x}) \prod_{j=1}^{k+d_2-1} \varphi_{\varepsilon}(\mathbf{f}_{j,k} \cdot \tilde{\mathbf{x}}), \\ \varphi_{k;d_1,d_2}^*(\mathbf{x}) &= \exp(i\mathbf{x} \cdot \bar{X} \mathbf{1}) \prod_{j=0}^{\infty} \varphi_{\varepsilon^*}(\hat{\mathbf{h}}_{j,k} \cdot \mathbf{x}) \prod_{j=1}^{k+d_2-1} \varphi_{\varepsilon^*}(\hat{\mathbf{f}}_{j,k} \cdot \tilde{\mathbf{x}}), \end{aligned}$$

where  $\mathbf{h}_{j,k} = (\psi_{j-d_1+1}, \dots, \psi_j, \psi_{j+k}, \dots, \psi_{j+k+d_2-1})'$ ,  $\mathbf{f}_{j,k} = (\psi_{k-j}, \dots, \psi_{k+d_2-1-j})'$ ,  $\hat{\mathbf{h}}_{j,k}$  and  $\hat{\mathbf{f}}_{j,k}$  analogously with  $\hat{\psi}_{j,n}$  instead of  $\psi_j$ ,  $\mathbf{x} = (x_1, \dots, x_{d_1}, x_{d_1+1}, \dots, x_d)'$ ,  $\tilde{\mathbf{x}} = (x_{d_1+1}, \dots, x_d)'$ ,  $\varphi_{\varepsilon}(x) = E[\exp(ix\varepsilon_0)]$ ,  $\varphi_{\varepsilon^*}(x) = E^*[\exp(ix\varepsilon_0^*)]$ ,  $x \in \mathbb{R}$ . Here we have followed the convention that  $\psi_j = \hat{\psi}_{j,n} = 0$  for  $j < 0$ . We then obtain

$$\begin{aligned}
 & \sup_{k \in \mathbb{N}} |\varphi_{k; d_1, d_2}^*(\mathbf{x}) - \varphi_{k; d_1, d_2}(\mathbf{x})| \\
 & \leq |\exp(\mathbf{i}x \cdot \bar{X}\mathbf{1}) - \exp(\mathbf{i}x \cdot \mu_X \mathbf{1})| + \sup_{k \in \mathbb{N}} \left| \prod_{j=0}^{q(n)} \varphi_{\varepsilon^*}(\hat{\mathbf{h}}_{j,k} \cdot \mathbf{x}) - \prod_{j=0}^{q(n)} \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x}) \right| \\
 & \quad + \sup_{k \in \mathbb{N}} \left| \prod_{j=q(n)+1}^{\infty} \varphi_{\varepsilon^*}(\hat{\mathbf{h}}_{j,k} \cdot \mathbf{x}) - \prod_{j=q(n)+1}^{\infty} \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x}) \right| \\
 & \quad + \sup_{k \in \mathbb{N}} \left| \prod_{j=1}^{k+d_2-1} \varphi_{\varepsilon^*}(\hat{\mathbf{f}}_{j,k} \cdot \tilde{\mathbf{x}}) - \prod_{j=1}^{k+d_2-1} \varphi_{\varepsilon}(\mathbf{f}_{j,k} \cdot \tilde{\mathbf{x}}) \right| \\
 & = I(\mathbf{x}) + \sup_{k \in \mathbb{N}} II_k(\mathbf{x}) + \sup_{k \in \mathbb{N}} III_k(\mathbf{x}) + \sup_{k \in \mathbb{N}} IV_k(\mathbf{x}), \tag{5.11}
 \end{aligned}$$

where  $q(n) \rightarrow \infty$ ,  $q(n) = o(n)$  ( $n \rightarrow \infty$ ).

By a Taylor expansion we obtain

$$\sup_{\mathbf{x} \in J(n; d)} I(\mathbf{x}) \leq \delta(n)^{-1} O_P(n^{-1/2}). \tag{5.12}$$

Again by using a Taylor expansion, we obtain

$$\begin{aligned}
 \sup_{k \in \mathbb{N}} \sup_{\mathbf{x} \in J(n; d)} III_k(\mathbf{x}) & \leq \sup_{k \in \mathbb{N}} \sup_{\mathbf{x} \in J(n; d)} \sum_{j=q(n)+1}^{\infty} |\varphi_{\varepsilon^*}(\hat{\mathbf{h}}_{j,k} \cdot \mathbf{x}) - \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x})| \\
 & \leq (E|\varepsilon_t| + E^*|\varepsilon_t^*|) \sup_{k \in \mathbb{N}} \sup_{\mathbf{x} \in J(n; d)} \sum_{j=q(n)+1}^{\infty} (|\hat{\mathbf{h}}_{j,k} \cdot \mathbf{x}| + |\mathbf{h}_{j,k} \cdot \mathbf{x}|) \\
 & \leq \text{const.} \delta(n)^{-1} q(n)^{-\vartheta} \text{ in probability,} \tag{5.13}
 \end{aligned}$$

where the last inequality follows from Lemma 5.2(ii), Lemma 5.1(i) and  $|\phi_j| = O(j^{-\beta})$  ( $j \rightarrow \infty$ ), which implies  $\sum_{j=q(n)+1}^{\infty} |\mathbf{h}_{j,k}| = o(q(n)^{-\vartheta})$ .

Most work is needed for bounding  $II_k(\mathbf{x})$  (and similarly  $IV_k(\mathbf{x})$ ). We have

$$\begin{aligned}
 II_k(\mathbf{x}) & \leq \sum_{j=0}^{q(n)} |\varphi_{\varepsilon^*}(\hat{\mathbf{h}}_{j,k} \cdot \mathbf{x}) - \varphi_{\varepsilon^*}(\mathbf{h}_{j,k} \cdot \mathbf{x})| + \sum_{j=0}^{q(n)} |\varphi_{\varepsilon^*}(\mathbf{h}_{j,k} \cdot \mathbf{x}) - \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x})| \\
 & = II.1_k(\mathbf{x}) + II.2_k(\mathbf{x}).
 \end{aligned}$$

By a Taylor expansion we obtain

$$\begin{aligned}
 \sup_{k \in \mathbb{N}} \sup_{\mathbf{x} \in J(n; d)} II.1_k(\mathbf{x}) & \leq \text{const.} \delta(n)^{-1} \left( \sup_{s \in \mathbb{N}_0} \sum_{j=s}^{s+q(n)} |\hat{\psi}_{j,n} - \psi_{j,n}| + \sup_{k \in \mathbb{N}} \sum_{j=k}^{k+q(n)+d_2-1} |\hat{\psi}_{j,n} - \psi_{j,n}| \right) \\
 & = \delta(n)^{-1} (O_P(q(n)p(n)^{-\vartheta}) + O_P(q(n)(\log(n)/n)^{1/2})), \tag{5.14}
 \end{aligned}$$

where we have used Lemma 5.1(ii) for the last inequality.

For bounding  $II.2_k(\mathbf{x})$  we consider

$$\begin{aligned} & \varphi_{\varepsilon^*}(\mathbf{h}_{j,k} \cdot \mathbf{x}) - \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x}) \\ &= (n-p)^{-1} \sum_{t=p+1}^n \exp(i\mathbf{h}_{j,k} \cdot \mathbf{x}(\hat{\varepsilon}_{t,n} - \hat{\mu}_{\varepsilon})) \int_{-\infty}^{\infty} \exp(i\mathbf{h}_{j,k} \cdot \mathbf{x}uh(n))K(u)du - \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x}) \\ &= (n-p)^{-1} \sum_{t=p+1}^n \exp(i\mathbf{h}_{j,k} \cdot \mathbf{x}(\hat{\varepsilon}_{t,n} - \hat{\mu}_{\varepsilon}))(1 + E(h(n); j, \mathbf{x})) - \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x}), \end{aligned} \quad (5.15)$$

where

$$|E(h(n); j, \mathbf{x})| \leq |\mathbf{h}_{j,k} \cdot \mathbf{x}|h(n) \int_{-\infty}^{\infty} |u|K(u)du \leq \text{const.}h(n)|\mathbf{h}_{j,k} \cdot \mathbf{x}|.$$

On the other hand,

$$\begin{aligned} & (n-p)^{-1} \sum_{t=p+1}^n \exp(i\mathbf{h}_{j,k} \cdot \mathbf{x}(\hat{\varepsilon}_{t,n} - \hat{\mu}_{\varepsilon})) - \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x}) \\ &= (n-p)^{-1} \sum_{t=p+1}^n \exp(i\mathbf{h}_{j,k} \cdot \mathbf{x}\varepsilon_t)(1 + D(h(n), t; j, \mathbf{x})) - \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x}), \end{aligned} \quad (5.16)$$

where

$$|D(h(n), t; j, \mathbf{x})| \leq |\mathbf{h}_{j,k} \cdot \mathbf{x}||\hat{\varepsilon}_{t,n} - \hat{\mu}_{\varepsilon} - \varepsilon_t|,$$

and hence (see (5.7)–(5.8))

$$(n-p)^{-1} \sum_{t=p+1}^n |D(h(n), t; j, \mathbf{x})| = |\mathbf{h}_{j,k} \cdot \mathbf{x}|(O_P(p(n)(\log(n)/n)^{1/2}) + o_P(p(n)^{-9})). \quad (5.17)$$

(Here the  $O_P$  terms are uniformly bounded in  $j, k$  and  $\mathbf{x}$ .)

Moreover, by the i.i.d. structure of  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  and the boundedness of  $\exp(ix), x \in \mathbb{R}$ , we obtain by some well-known exponential inequalities, for instance Bernstein’s inequality,

$$\sup_{|x| \leq n^r} |(n-p)^{-1} \sum_{t=p+1}^n \exp(ix\varepsilon_t) - \varphi_{\varepsilon}(x)| = O_P(n^{-1/2+\eta}), \text{ for any } 0 < \eta < \frac{1}{2},$$

where  $r$  is an arbitrary exponent in  $\mathbb{R}^+$ . This is a stronger version of formula (2.4) in Singh (1981). But this implies

$$\sup_{k \in \mathbb{N}} \sup_{j \in \mathbb{N}_0} \sup_{\mathbf{x} \in J(n,d)} \left| (n-p)^{-1} \sum_{t=p+1}^n \exp(i\mathbf{h}_{j,k} \cdot \mathbf{x}\varepsilon_t) - \varphi_{\varepsilon}(\mathbf{h}_{j,k} \cdot \mathbf{x}) \right| = O_P(n^{1/2-\eta}), \text{ for any } 0 < \eta < \frac{1}{2}. \quad (5.18)$$

Therefore by (5.15)–(5.18) we obtain

$$\begin{aligned} \sup_{k \in \mathbb{N}} \sup_{\mathbf{x} \in J(n;d)} II_{2k}(\mathbf{x}) &\leq \delta(n)^{-1}(O_P(h(n)) + O_P(p(n)(\log(n)/n)^{1/2}) + o_P(p(n)^{-\vartheta})) \\ &\quad + O_P(q(n)n^{-1/2+\eta}), \quad \eta > 0. \end{aligned} \tag{5.19}$$

Hence by (5.14) and (5.19),

$$\begin{aligned} \sup_{k \in \mathbb{N}} \sup_{\mathbf{x} \in J(n;d)} II_k(\mathbf{x}) &= \delta(n)^{-1}(O_P(h(n)) + O_P((q(n) + p(n))(\log(n)/n)^{1/2}) + O_P(q(n)p(n)^{-\vartheta})) \\ &\quad + O_P(q(n)n^{-1/2+\eta}), \quad \eta > 0. \end{aligned} \tag{5.20}$$

Similarly, we obtain

$$\sup_{k \in \mathbb{N}} \sup_{\mathbf{x} \in J(n;d)} IV_k(\mathbf{x}) = O(\sup_{k \in \mathbb{N}} \sup_{\mathbf{x} \in J(n;d)} II_k(\mathbf{x})) + O(\sup_{k \in \mathbb{N}} \sup_{\mathbf{x} \in J(n;d)} III_k(\mathbf{x})). \tag{5.21}$$

Then we have by (5.11)–(5.13), (5.20)–(5.21),

$$\begin{aligned} \sup_{k \in \mathbb{N}} \sup_{\mathbf{x} \in J(n;d)} |\varphi_{k;d_1,d_2}^*(\mathbf{x}) - \varphi_{k;d_1,d_2}(\mathbf{x})| &\leq \delta(n)^{-1}(O_P(q(n)^{-\vartheta}) + O_P(h(n)) + O_P((q(n) + p(n))(\log(n)/n)^{1/2}) + O_P(q(n)p(n)^{-\vartheta})) \\ &\quad + O_P(q(n)n^{-1/2+\eta}) \\ &= \delta(n)^{-1}(O_P(q(n)^{-\vartheta}) + O_P(h(n)) + O_P((q(n) + p(n))(\log(n)/n)^{1/2}) + O_P(q(n)p(n)^{-\vartheta})), \end{aligned} \tag{5.22}$$

where the last bound follows since  $\eta > 0$  is arbitrary.

What remains is to integrate the error term in (5.22); see (5.10). Let  $r(n) \rightarrow \infty$ ,  $r(n) = o(n)$  ( $n \rightarrow \infty$ ). Denote by  $C_1 = C \cap [-r(n), r(n)]^d$ ,  $C_2 = C \setminus C_1$ . By Markov’s inequality we obtain

$$\sup_{k \in \mathbb{N}} |\mathbb{P}_{k;d_1,d_2} \star K_{\delta(n)}[C_2]| = O(r(n)^{-s}), \quad \sup_{k \in \mathbb{N}} |\mathbb{P}_{k;d_1,d_2}^* \star K_{\delta(n)}[C_2]| = O_P(r(n)^{-s}). \tag{5.23}$$

Hence by (5.10), (5.22) and (5.23)

$$\begin{aligned} \sup_{k \in \mathbb{N}} |(\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}) \star K_{\delta(n)}[C]| &\leq r(n)^d \delta(n)^{-d-1} (O_P(q(n)^{-\vartheta}) + O_P(h(n)) + O_P((q(n) + p(n))(\log(n)/n)^{1/2}) \\ &\quad + O_P(q(n)p(n)^{-\vartheta})) + O_P(r(n)^{-s}), \end{aligned}$$

and therefore by Lemma 5.4

$$\sup_{k \in \mathbb{N}} \|\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{L}^{d_1, \mathcal{G}^{d_2}}} = O_P(r(n)^d \delta(n)^{-d-1} \xi(n)) + O_P(r(n)^{-s}) + O(\delta(n)^{\lambda}), \tag{5.24}$$

where  $\xi(n) = \max\{q(n)^{-\vartheta}, h(n), (q(n) + p(n))(\log(n)/n)^{1/2}, q(n)p(n)^{-\vartheta}\}$ . By choosing

$q(n) = p(n)^{\vartheta/(\vartheta+1)}$  we obtain  $\xi(n) = \max\{h(n), p(n)(\log(n)/n)^{1/2}, p(n)^{-\vartheta^2/(\vartheta+1)}\}$ . By choosing the optimal orders for  $r(n)$  and  $\delta(n)$  the right-hand side in (5.24) is of the order

$$O_p(\xi(n)^{\frac{\lambda}{s(1+\lambda+d)+d}}).$$

This completes the proof. □

### 5.3. Proofs of Theorem 3.2 and Theorem 3.4

We first give the proof of Theorem 3.2. To do this, we combine the results in Section 5.1 and 5.2. By Lemma 5.3 we know

$$\mathbb{P}[\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \alpha^*(k) \leq \text{const.} h(n)^{-1} k^{-\gamma^*} \forall k \in \mathbb{N}] \rightarrow 1 \quad (n \rightarrow \infty),$$

and hence,

$$\mathbb{P}[\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \text{const.} k^{-(\gamma^* - \zeta)} \forall k \geq (h(n)^{-1})^{1/\zeta}] \rightarrow 1 \quad (n \rightarrow \infty), \zeta \in \mathbb{R}^+. \quad (5.25)$$

On the other hand, we have by (5.3), Theorem 3.1 and Lemma 5.5 (using the second statement) for  $k < (h(n)^{-1})^{1/\zeta}$ ,

$$\begin{aligned} \nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) &\leq \alpha(k) + \text{const.} \|\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{E}^{d_1}, \mathcal{D}^{d_2}} \\ &\leq \text{const.} k^{-\gamma} + O_p(h(n)^{\frac{\lambda}{s(1+\lambda+d)+d}}) \end{aligned}$$

and since the  $O_p(\cdot)$  term is uniform in  $k$  (see Lemma 5.5), we obtain

$$\begin{aligned} \mathbb{P}[\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \text{const.} k^{-\gamma} + \text{const.} k^{-\frac{\zeta \lambda}{s(1+\lambda+d)+d}} \forall k < (h(n)^{-1})^{1/\zeta}] \\ \rightarrow 1 \quad (n \rightarrow \infty), \zeta \in \mathbb{R}^+. \end{aligned} \quad (5.26)$$

By choosing  $\zeta$  yielding the best rate for  $\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2})$ , that is,  $\zeta = \gamma^*(s(1 + \lambda + d) + d)/(s(1 + 2\lambda + d) + d)$ , we obtain by (5.25) and (5.26) the result of Theorem 3.2.

We now sketch the arguments for proving Theorem 3.4 which are very similar to the proof of Theorem 3.2. We first show the following: there exists a random variable  $n_0(\omega)$  such that

$$\sup_{n \geq n_0(\omega)} \sum_{j=0}^{p(n)} |\hat{\phi}_{j,n}| (1 + \tilde{\kappa})^j < \infty \text{ almost surely,} \quad 0 < \tilde{\kappa} < \min\{\kappa, \exp(1/(2C)) - 1\}. \quad (5.27)$$

We have, for any  $0 < c < \frac{1}{2}$

$$\begin{aligned} \sup_{|z| \leq 1 + \tilde{\kappa}} |\hat{\Phi}_n(z) - \Phi(z)| &\leq \max_{1 \leq j \leq p(n)} |\hat{\phi}_{j,n} - \phi_j| (1 + \tilde{\kappa})^{p(n)} + \sum_{j=p(n)+1}^{\infty} |\phi_j| (1 + \tilde{\kappa})^j \\ &= O((\log(n)/n)^{1/2}) O(n^{1/2-c}) + o(1) = o(1) \text{ almost surely.} \end{aligned}$$

(Use the result of Theorem 2.1 of Hannan and Kavalieris (1986) and Baxter's inequality;

compare with (5.7) and (5.8)). Formula (5.27), together with (A1'), implies that, for  $n$  sufficiently large, we can invert  $\hat{\Phi}_n(z)$  in  $|z| \leq 1 + \tilde{\kappa}$ ; we then obtain instead of Lemma 5.1,

$$\sup_{n \geq n_0(\omega)} |\hat{\psi}_{j,n}| \leq \text{const.}(1 + \tilde{\kappa})^{-j} \text{ almost surely,}$$

$$\sup_{s \in \mathbb{N}_0} \sum_{j=s+1}^{s+a(n)} |\hat{\psi}_{j,n} - \psi_j| = O(a(n)(1 + \tilde{\kappa})^{-p(n)}) + O(a(n)(\log(n)/n)^{1/2}) \text{ almost surely.}$$

Lemma 5.2 remains the same with  $p(n) = O(\log(n))$ . Lemma 5.3 becomes

$$\mathbb{P}[\alpha^*(k) \leq \text{const.}h(n)^{-1}\tilde{\rho}^k \ \forall k \in \mathbb{N}] \rightarrow 1 \quad (n \rightarrow \infty), \quad (1 + \tilde{\kappa})^{-s/(s+1)} < \tilde{\rho} < 1. \quad (5.28)$$

(Compare with Theorem 3.3.)

Lemma 5.4 remains exactly the same. It is plausible that we obtain the same bound as in Lemma 5.5, since the assumptions (A1)–(A5) are generally weaker than the assumptions of Theorem 3.4, that is,

$$\sup_{k \in \mathbb{N}} \|\mathbb{P}_{k;d_1,d_2}^* - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{E}^{d_1}, \mathcal{D}^{d_2}} = O_P(h(n)^{\frac{s_2}{s(1+\tilde{\kappa}+d)+d}}), \quad d = d_1 + d_2. \quad (5.29)$$

However, we have to re-examine the interplay of the tuning parameters  $h(n)$  and  $p(n)$ . Some quantities now change; we choose  $q(n) = \text{const.}\log(n)$  such that the (old) expression  $q(n)^{-\theta}$  becomes something of order  $n^{-1/2}$ . By (A3'),  $p(n) \sim C \log(n)$  and instead of the (old) expression  $p(n)^{-\theta}$  we have  $(1 + \tilde{\kappa})^{-p(n)}$ . Then  $\xi(n)$  in (5.24) equals  $\max\{h(n), n^{-1/2+\eta}, \log(n)(1 + \tilde{\kappa})^{-p(n)}\}$ ; note that for deriving this the  $O_P(q(n)n^{-1/2+\eta})$  term in (5.20) dominates in the derivation of (5.22). By choosing  $\tilde{\kappa}$  appropriately close to  $\min\{\kappa, \exp(1/(2C)) - 1\}$ , we know that by (A4')  $\max\{h(n), \log(n)(1 + \tilde{\kappa})^{-p(n)}\} = O(h(n))$ . This then explains why (5.29) holds.

Now by (5.28), for  $\tilde{\rho} < \tau < 1$ ,

$$\mathbb{P}[\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \text{const.}(\tilde{\rho}/\tau)^k \ \forall k > -\log(h(n)^{-1})/\log(\tau)] \rightarrow 1 \quad (n \rightarrow \infty),$$

and by (5.29), for  $\tilde{\rho} < \tau < 1$ ,

$$\mathbb{P}[\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \text{const.}\tau^{k\frac{s_2}{s(1+\tilde{\kappa}+d)+d}} \ \forall k \leq -\log(h(n)^{-1})/\log(\tau)] \rightarrow 1 \quad (n \rightarrow \infty).$$

By choosing  $\tau = \tilde{\rho}^{\frac{s(1+\tilde{\kappa}+d)+d}{s(1+2\tilde{\kappa}+d)+d}}$  we arrive at

$$\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \text{const.}\tilde{\rho}^{k\frac{s_2}{s(1+2\tilde{\kappa}+d)+d}} = \text{const.}(\rho^*)^k,$$

where  $(1 + \tilde{\kappa})^{-\frac{s_2}{(s+1)(s(1+2\tilde{\kappa}+d)+d)}} < \rho^* < 1$ , that is,

$$\mathbb{P}[\nu^*(k; \mathcal{E}^{d_1}, \mathcal{D}^{d_2}) \leq \text{const.}(\rho^*)^k \ \forall k \in \mathbb{N}] \rightarrow 1 \quad (n \rightarrow \infty).$$

### Acknowledgements

The research of the first author was supported in part by grants NSA MDA 904-94-H-2020 and NSF DMS 95049555. The research of the second author was supported in part by the



Swiss National Science Foundation. Comments by all three referees helped to correct and clarify the statements and expositions of our results.

## References

- An, H.-Z., Chen, Z.-G. and Hannan, E.J. (1982) Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.*, **10**, 926–936. (Correction **11**, 1018.)
- Andrews, D.W.K. and Pollard, D. (1994) An introduction to functional central limit theorems for dependent stochastic processes. *Internat. Statist. Rev.*, **62**, 119–132.
- Bhattacharya, R.N. and Ranga Rao, R. (1976) *Normal Approximation and Asymptotic Expansions*. New York: Wiley.
- Bickel, P.J. and Bühlmann, P. (1996) What is a linear process? *Proc. Nat. Acad. Sci. U.S.A.*, **93**, 12 128–12 131.
- Bickel, P.J. and Bühlmann, P. (1997) Closure of linear processes. *J. Theoret. Probab.*, **10**, 445–479.
- Brockwell, P.J. and Davis, R.A. (1987) *Time Series: Theory and Methods*. New York: Springer-Verlag.
- Bühlmann, P. (1994) Blockwise bootstrapped empirical processes for stationary sequences. *Ann. Statist.*, **22**, 995–1012.
- Bühlmann, P. (1995) Moving-average representation for autoregressive approximations. *Stochastic Process. Appl.*, **60**, 331–342.
- Bühlmann, P. (1997) Sieve bootstrap for time series. *Bernoulli*, **3**, 123–148.
- Doukhan, P. (1994) *Mixing. Properties and Examples*. Lecture Notes in Statist. 85, New York: Springer-Verlag.
- Doukhan, P. and Portal, F. (1987) Weak invariance principles for the empirical distribution function in a multidimensional and mixing environment [in French]. *Probab. Math. Statist.*, **8**, 117–132.
- Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7**, 1–26.
- Falk, M. and Reiss, R.-D. (1989a) Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. *Ann. Probab.*, **17**, 362–371.
- Falk, M. and Reiss, R.-D. (1989b) Bootstrapping the distance between smooth bootstrap and sample quantile distribution. *Probab. Theory Related Fields*, **82**, 177–186.
- Franke, J. and Kreiss, J.-P. (1992) Bootstrapping stationary autoregressive moving-average models. *J. Time Ser. Anal.*, **13**, 297–317.
- Giné, E. and Zinn, J. (1990) Bootstrapping general empirical measures. *Ann. Probab.*, **18**, 851–869.
- Gorodetskii, V.V. (1977) On the strong mixing property for linear sequences. *Theory Probab. Appl.*, **22**, 411–413.
- Hall, P., DiCiccio, T.J. and Romano, J.P. (1989) On smoothing and the bootstrap. *Ann. Statist.*, **17**, 692–704.
- Hannan, E.J. and Kavalieris, L. (1986) Regression, autoregression models. *J. Time Ser. Anal.*, **7**, 27–49.
- Hjellvik, V. and Tjøstheim, D. (1995) Nonparametric tests of linearity for time series. *Biometrika*, **82**, 351–368.
- Kreiss, J.-P. (1988) Asymptotic statistical inference for a class of stochastic processes. *Habilitationsschrift*, Universität Hamburg.
- Kreiss, J.-P. (1992) Bootstrap procedures for  $AR(\infty)$ -processes. In K.H. Jöckel, G. Rothe and W. Sendler (eds), *Bootstrapping and Related Techniques*, pp. 107–113. Lect. Notes in Econom. and Math. Systems 376. Heidelberg: Springer-Verlag.
- Paparoditis, E. and Streitberg, B. (1992) Order identification statistics in stationary autoregressive

- moving-average models: vector autocorrelations and the bootstrap. *J. Time Ser. Anal.*, **13**, 415–434.
- Shibata, R. (1980) Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.*, **8**, 147–164.
- Silverman, B.W. and Young, G.A. (1987) The bootstrap: To smooth or not to smooth? *Biometrika*, **74**, 469–479.
- Singh, K. (1981) On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.*, **9**, 1187–1195.
- Subba Rao, T. and Gabr, M.M. (1980) A test for linearity of stationary time series. *J. Time Ser. Anal.*, **1**, 145–158.
- Tsay, R.S. (1992) Model checking via parametric bootstraps in time series analysis. *Appl. Statist.*, **41**, 1–15.
- Yokoyama, R. (1980) Moment bounds for stationary mixing sequences. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, **52**, 45–57.

Received September 1995 and revised March 1998.