Mixing Property and Functional Central Limit Theorems for a Sieve Bootstrap in Time Series

Peter J. Bickel * and Peter Bühlmann †
University of California
Berkeley

September 1995

Abstract

We study a bootstrap method for stationary real-valued time series, which is based on the method of sieves. We restrict ourselves to autoregressive sieve bootstraps. Given a sample $X_1, \ldots, X_n$ 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 $p(n) \to \infty$, $p(n) = o(n)$ as the sample size $n \to \infty$. 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 give a novel result which says 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.

Key words and phrases. AR(∞), ARMA, autoregressive approximation, bracketing, convex sets, linear process, MA(∞), smooth bootstrap, stationary process, strong-mixing.

Short title: Sieve Bootstrap in Time Series

*Research supported in part by grants NSA MDA 904-94-H-2020 and NSF DMS 95049555.
†Research supported in part by the Swiss National Science Foundation.

1 Introduction

In Bühlmann (1995b) one of us investigated what we like to call the AR(∞) sieve bootstrap for time series. Following Bühlmann (1995b) first define an MA(∞) (or linear) process $\{X_t\}_{t\in\mathbb{Z}}$ with expectation $\mathbb{E}[X_t] = \mu_X$ as follows:

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

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence with common distribution $F, \mathbb{E}[\varepsilon_t] = 0, \mathbb{E}|\varepsilon_t| < \infty$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$. The set of all joint distributions on $\mathbb{R}^\infty$ induced by such $\{X_t\}_{t\in\mathbb{Z}}$ is a semiparametric model indexed by $\{F : \int x dF(x) = 0\} \times \{\{\psi_j\}_{j=0}^{\infty} \in \ell_1 : \psi_0 = 1\}$. An alternative definition leading to a slightly different set would be to require (1.1) for more restricted distributions F with $\mathbb{E}|\varepsilon_t|^2 < \infty, \mathbb{E}[\varepsilon_t] = 0$ but now only ask that $\{\psi_j\}_{j=0}^{\infty} \in \ell_2$. We are concerned here with a subset of the MA(∞) processes which we call AR(∞), namely all processes representable as in (1.1) but also satisfying,

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

with $\sum_{j=0}^{\infty} |\phi_j| < \infty$. As is remarked in Bühlmann (1995a) an MA(∞) process is AR(∞) 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(∞) models, the approximation we refer to is in the sense of weak convergence of finite dimensional distributions of any order. In fact as we shall discuss elsewhere the sets of stationary process distributions obtainable as limits from (1.1) or (1.2) is quite large but far from exhaustive. Various authors, in particular Tsay (1992), implicitly, and Hjellvik and Tjøstheim (1995) explicitly view 'linear processes' as being AR(∞) (or approximable by AR(∞)).

Given this point of view it is reasonable given a sequence $\{X_t\}$, $1 \leq t \leq n$, from the process 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(k). When dealing with the AR(∞) hypothesis we face not only the choice of test statistics but also what critical value to which we should refer these statistics. It is natural to try to estimate these critical values using a bootstrap appropriate to this hypothesis. Such a bootstrap was suggested by Kreiss (1988) and its properties were explored in part by Bühlmann (1995b). The results of that paper establish that the sieve bootstrap we discuss below gives correct approximations to the distributions of linear statistics such as $\sum_{t=1}^{n} h(X_{t+1}, \ldots, X_{t+m})$, where h is smooth, or smooth functions thereof.

The statistics of Hjellvik and Tjøstheim however involve estimates of the marginal densities of X_t and statistics proposed by other authors, cf. 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, etc.

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, nonregular 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 may be of independent interest.

In separate work joint with John Rice we shall develop systematically and study empirically a number of test statistics and diagnostics to which the theory of this paper can be applied.

2 The smoothed sieve bootstrap

The autoregressive sieve bootstrap has been studied in Kreiss (1988) and in Bühlmann (1995b). We briefly recall the method as given in Bühlmann (1995b). Let $\{X_t\}_{t\in\mathbb{Z}}$ be a real-valued, stationary linear process as given in (1.1) which satisfies also the infinite autoregressive representation as in (1.2). Given data $X_1, \ldots, X_n$ from an AR(∞) 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 $\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n}$ corresponding to model (1.2), usually (but not necessarily) by the Yule-Walker estimates, which allows us to calculate centered residuals. Then we resample by the bootstrap as in Efron (1979) from these centered residuals yielding

$$\varepsilon_t^*,\ t \in \mathbb{Z}.$$

Finally we construct a sieve bootstrap sample according to an AR($p(n)$) process with coefficients $\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n}$, i.e.,

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

It is then also shown in Bühlmann (1995b) that the sieve bootstrap process $\{X_t^*\}_{t\in\mathbb{Z}}$ can be again inverted and represented as a linear process

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

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

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

Moreover the behavior 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 (1995a). 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 ω of the underlying probability space. This is due to the fact that the distribution of the innovations ε_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 the Lebesgue measure in some neighborhood of the expectation of the innovation, cf. Gorodetskii (1977), Doukhan (1994). We leave it as an open question if the process in (2.1) or equivalently in (2.2) possesses some kind of mixing property which holds uniformly over all ω's.

On the other hand, some type of mixing property of the sieve bootstrap process would be desirable to describe and analyze the probabilistic behavior of the bootstrap process as in (2.1) or in (2.2). It would basically say that if the underlying process as in (1.1) is linear and mixing, then the sieve bootstrap process would be again linear and mixing. In particular, many of the results for the underlying process $\{X_t\}_{t\in\mathbb{Z}}$ would carry over to the sieve bootstrap process $\{X_t^*\}_{t\in\mathbb{Z}}$. As an example, the sieve bootstrap for empirical processes would work under similar conditions as for the original process; this would be 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 will present a modification which achieves some mixing property for the sieve bootstrap process. 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, DiCiccio and Romano (1989), Falk and Reiss (1989a,b). Somewhat related ideas are the Bayesian bootstrap, introduced by Rubin (1981), the generally weighted bootstrap, cf. Haeusler, Mason and Newton (1991), Mason and Newton (1992), and the 'm out of n bootstrap', cf. Arcones and Giné (1989), Bickel, Götze and van Zwet (1994) and Politis and Romano (1994). For time series, the idea of weighted bootstrapping with weights that are now correlated is considered in Künsch (1989), formula (2.12), Bühlmann (1993) and Bühlmann and Künsch (1994). Summarizing, it is often desirable to use smooth over non-smooth bootstrap techniques. Therefore, a smoothed sieve bootstrap can be justified also from a statistical point of view of gaining performance.

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

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

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

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

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

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

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

where $\hat{\mu}_\varepsilon = \int_{-\infty}^{\infty} x\hat{f}_{\hat{\varepsilon}}(x)dx$. The centering forces that ε_t^* has conditional mean zero.
(III) 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 inherits now the approximating order $p = p(n)$ and the bandwidth $h = h(n)$ which have to be chosen by the statistician.

3 Mixing property of smoothed sieve bootstrap process

We will establish in this section some 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 $\mathcal{M}_a^b = \sigma(\{X_j; a \le j \le b\})$ the σ-algebras with events that belong to the 'time interval' $[a, b]$. Moreover we denote the strong-mixing coefficients by

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

For the bootstrap we analogously define

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

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

Showing the strong-mixing property for the smoothed sieve bootstrap seems to be a difficult task. 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 variation norm between the joint probability and the product of the marginal probabilities. This definition allows to bound covariances

$$|Cov(Z_1, Z_2)| \le 8\|Z_1\|_{q_1}\|Z_2\|_{q_2}\alpha^{1/q_3}(k),\ 1 \le q_1, q_2, q_3 \le \infty,\ q_1^{-1} + q_2^{-1} + q_3^{-1} = 1.$$

for any measurable variable $Z_1 \in \mathcal{M}_{-\infty}^0,\ Z_2 \in \mathcal{M}_k^\infty$, cf. Doukhan (1994, Th.3, Ch.1.2.2). However, we often only want to bound

$$|Cov(g_1(X_{-d_1+1}, \ldots, X_0), g_2(X_k, \ldots, X_{k+d_2-1}))|,$$

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 σ-algebras, i.e., 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, cf. Doukhan (1994, Ch.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, cf. Bhattacharya and 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 (cf. Lemma 5.4) which works under such more restrictive assumptions. The new idea is here that we do not aim to bound a variation norm (between the joint and the corresponding product of marginal probabilities) over any measurable events in a σ-subfield; this approach is explained in Doukhan (1994, Ch.1.1), where various notions for mixing are defined as measures of dependence between σ-subfields.

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary, real-valued process. We denote by $\omega_g(A) = \sup_{\mathbf{y},\mathbf{z}\in A} |g(\mathbf{y}) - g(\mathbf{z})|$, $g : \mathbb{R}^d \to \mathbb{R}$, $A \subseteq \mathbb{R}^d$, $B(\mathbf{x},\delta) = \{\mathbf{y}; \|\mathbf{x}-\mathbf{y}\| \le \delta\} \subseteq \mathbb{R}^d$, $\mathbf{x} \in \mathbb{R}^d$, $\delta \in \mathbb{R}^+$, $\|.\|$ 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 $g_{\mathbf{y}} : \mathbb{R}^d \to \mathbb{R}$, $g_{\mathbf{y}}(\mathbf{x}) = g(\mathbf{x}+\mathbf{y})$ $(\mathbf{x},\mathbf{y} \in \mathbb{R}^d)$ the translation of the function $g(.)$. We also denote in the sequel by $\|g\|_q = (\mathbb{E}|g(X_1,\ldots,X_d)|^q)^{1/q}$ $(1 \le q < \infty)$ and by $\|g\|_\infty = \sup_{\mathbf{x}} |g(x_1,\ldots x_d)|$.

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

$$\begin{aligned}
&\sup_{g\in\mathcal{C}^d} \|g\|_\infty < \infty,\\
&\sup_{g\in\mathcal{C}^d}\sup_{\mathbf{y}\in\mathbf{R}^d} \mathbb{E}[\omega_{g_{\mathbf{y}}}(B((X_1,\ldots,X_d),\delta))] \le const.\delta^\lambda, \text{ for all } 0<\delta<1, \text{ for some } \lambda>0,\\
&d \in \mathbb{N}.
\end{aligned} \tag{3.1}$$

We then say that $(\mathcal{C}^d, \lambda)$ satisfies (3.1). Of course, this depends also on the d-dimensional marginal distribution of the underlying process $\{X_t\}_{t\in\mathbb{Z}}$, but we usually do not mention it. If clear from the context, or if the value of λ is not of particular interest, we suppress the constant λ.

We now present our new mixing notion and define the so called ν-mixing coefficient for the stationary process $\{X_t\}_{t\in\mathbb{Z}}$ as

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

where $(\mathcal{C}^{d_1},\lambda_1)$, $(\mathcal{D}^{d_2},\lambda_2)$ satisfy (3.1) for possibly different $\lambda_1,\lambda_2>0$, $d_1,\ d_2\in\mathbb{N}$.

The expectation in condition (3.1) is meant with respect to the probability measure of the process $\{X_t\}_{t\in\mathbb{Z}}$ for which we define the ν-mixing coefficients. Whenever we write $\nu(.;\mathcal{C}^{d_1},\mathcal{D}^{d_2})$ we implicitly mean that $(\mathcal{C}^{d_1},\lambda_1)$, $(\mathcal{D}^{d_2},\lambda_2)$ satisfy (3.1) for some $\lambda_1,\ \lambda_2>0$. We say that the stationary process $\{X_t\}_{t\in\mathbb{Z}}$ is ν-mixing with respect to $(\mathcal{C}^{d_1},\ \mathcal{D}^{d_2})$ if $\nu(k;\mathcal{C}^{d_1},\ \mathcal{D}^{d_2}) \to 0$ for $k\to\infty$. The factor $1/(4\|g_1\|_\infty\|g_2\|_\infty)$ in the definition of ν-mixing is essential in order to get good bounds for estimating covariances as given in the following Lemma.

Lemma 3.1 *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary real-valued process and let $\mathcal{C}^{d_1}$, $\mathcal{D}^{d_2}$ $(d_1,\ d_2 \in \mathbb{N})$ be classes of measurable functions that satisfy the condition (3.1). Then*

(i) $\nu(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) \le \alpha(k)$,

(ii) for $g_1 \in \mathcal{C}^{d_1},\ g_2 \in \mathcal{D}^{d_2},\ 1 \le q_1,\ q_2,\ q_3 \le \infty$ *with* $q_1^{-1}+q_2^{-1}+q_3^{-1}=1$,
$|Cov(g_1(X_{-d_1+1},\ldots,X_0), g_2(X_k,\ldots X_{k+d_2-1}))| \le 8\|g_1\|_{q_1}\|g_2\|_{q_2}\nu^{1/q_3}(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2})$.

Proof: Assertion (i) follows immediately by definition. Assertion (ii) follows in the same way as in the case of α-mixing sequences. We note that a first step

$$|Cov(g_1(X_{-d_1+1},\ldots,X_0), g_2(X_k,\ldots X_{k+d_2-1}))| \le 4\|g_1\|_\infty\|g_2\|_\infty\nu(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2})$$

follows immediately by the definition of the ν-mixing coefficient. Now in a second step we consider the case $\|g_1\|_p<\infty$, $\|g_2\|_\infty<\infty$, $1<p<\infty$.
Define $g_1^{upp} = g_1(X_{-d_1+1},\ldots,X_0)1_{[|g_1(X_{-d_1+1},\ldots,X_0)|>M]}$ and
$g_1^{low} = g_1(X_{-d_1+1},\ldots,X_0)1_{[|g_1(X_{-d_1+1},\ldots,X_0)|\le M]}$. Thus

$$\begin{aligned} & |Cov(g_1(X_{-d_1+1},\ldots,X_0), g_2(X_k,\ldots X_{k+d_2-1}))| \\ = \ & |Cov(g_1^{low}+g_1^{upp}, g_2(X_k,\ldots X_{k+d_2-1}))| \\ \le \ & 4M\|g_2\|_\infty\nu(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) + 2\|g_2\|_\infty\mathbb{E}|g_1^{upp}|. \end{aligned}$$

Now by Hölder's inequality $\mathbb{E}|g_1^{upp}| \le \|g_1\|_p^p M^{-p+1}$. By choosing M such that $\|g_1\|_p^p M^{-p} = \nu(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2})$ we arrive at

$$\begin{aligned} & |Cov(g_1(X_{-d_1+1},\ldots,X_0), g_2(X_k,\ldots X_{k+d_2-1}))| \\ \le \ & 6\|g_1\|_p\|g_2\|_\infty\nu^{1-1/p}(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}). \end{aligned}$$

Now consider the situation $\|g_1\|_p<\infty,\ \|g_2\|_q<\infty$, $p^{-1}+q^{-1}<1$. Analogously as above we define g_2^{upp} and g_2^{low} with a truncation point M'. Then with the covariance inequality above and Hölder's inequality,

$$\begin{aligned} & |Cov(g_1(X_{-d_1+1},\ldots,X_0), g_2(X_k,\ldots X_{k+d_2-1}))| \\ \le \ & 6M'\|g_1\|_p\nu^{1-1/p}(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) + 2M'\|g_1\|_p\|g_2^{upp}\|_{p/(p-1)}. \end{aligned}$$

Again by Hölder's inequality we get $\|g_2^{upp}\|_{p/(p-1)} \le \|g_2\|_q^{q(p-1)/p}M'^{-q(p-1)/p+1}$. Hence by choosing M' such that $\|g_2\|_q^{q(p-1)/p}M'^{-q(p-1)/p} = \nu^{1-1/p}(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2})$ we complete the proof. See also Doukhan (1994, Th.3, Ch.1.2.2) □

Often one is interested in estimating covariances of products. Suppose that $(\mathcal{C}_1^{d_1},\lambda_1),\ldots,(\mathcal{C}_r^{d_r},\lambda_r)$ all satisfy (3.1) for some $\lambda_s>0,\ d_s\in\mathbb{N}$ $(s=1,\ldots,r)$. Then we define

$$\begin{aligned} \otimes_{i=1}^r \mathcal{C}_i^{d_i} = \{ \quad & g_{i_1}\cdot\ldots\cdot g_{i_m} : \mathbb{R}^{\prod_{j=1}^m d_{i_j}} \to \mathbb{R}; g_{i_j}\in\mathcal{C}_{i_j}^{d_{i_j}},\ i_j\in\{1,\ldots,r\}, \\ & j=1,\ldots,m,\ m\le r\}. \end{aligned} \tag{3.2}$$

Then every 'subproduct' $\otimes_{j=1}^m\mathcal{C}_{i_j}^{d_{i_j}} \subseteq \otimes_{i=1}^r\mathcal{C}_i^{d_i}$ $(j=1,\ldots,m\le r)$ and $(\otimes_{i=1}^r\mathcal{C}_i^{d_i},\lambda)$ satisfies again (3.1) with $\lambda=\min\{\lambda_s; 1\le s\le r\}$. This fact enables us to establish the same moment inequalities for centered sums as for α-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 $i=1,\ldots,r$.

Lemma 3.2 *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary real-valued process. Assume that $g:\mathbb{R}^d\to\mathbb{R}\in\mathcal{C}^d$ satisfying (3.1). Then the following holds true.*

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

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

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

$$\begin{aligned}&\mathbb{E}\left|n^{-1/2}\sum_{t=1}^{n}(g(X_{t+1},\ldots,X_{t+d})-\mathbb{E}[g(X_{t+1},\ldots,X_{t+d})])\right|^{2r}\\&\le const.\left((n\tau^2)+\ldots+(n\tau^2)^r\right),\ r\in\mathbb{N}.\end{aligned}$$

Proof: By using Lemma 3.1 (ii) the statements follow as in Yokoyama (1980) and Andrews and Pollard (1994), respectively. □.

We remark that the bounds in Lemma 3.2 will often be applied to a class $\tilde{\mathcal{C}}^d=\{g_1-g_2;g_1,\ g_2\in\mathcal{C}^d\}$. But $(\tilde{\mathcal{C}}^d,\lambda)$ satisfies (3.1) whenever $(\mathcal{C}^d,\lambda)$ does with the same $\lambda>0$. This property comes into play when proving stochastic equicontinuity for ν-mixing sequences, see section 4.

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

$$\mathcal{C}^d=\left\{g:\mathbb{R}^d\to\mathbb{R};g=1_{[(\infty,b_1]\times\ldots\times(-\infty,b_d]]},\ (b_1,\ldots,b_d)\in\mathbb{R}^d\right\}$$

satisfies (3.1) with $\lambda=1$ if the d-dimensional marginal distribution of $\{X_t\}_{t\in\mathbb{Z}}$ has a bounded density.

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

$$\mathcal{C}^d=\left\{g:\mathbb{R}^d\to\mathbb{R};g=\sum_{j=1}^{m}c_j1_{[C_j]};\ c_j\in\mathbb{R},\ C_j\in\{\text{convex sets in }\mathbb{R}^d\}\right\},\ m\in\mathbb{N},$$

satisfies (3.1) with $\lambda=1$ if the d-dimensional marginal distribution of $\{X_t\}_{t\in\mathbb{Z}}$ has a density f, such that $f(\mathbf{x})=\tilde{f}(\|\mathbf{x}\|)$ and $\tilde{f}$ is differentiable with $\int_0^\infty|\tilde{f}'(y)|dy<\infty$ and $lim_{y\to\infty}\tilde{f}(y)=0$, cf. Bhattacharya and Rao (1976, Th.3.1).

Example 3.3. (Lipschitz functions of order λ). Denote by $\|.\|$ the Euclidean norm on $\mathbb{R}^d$. The class of functions

$$\mathcal{C}^d=\left\{g:\mathbb{R}^d\to\mathbb{R};\sup_{\mathbf{x}}|g(\mathbf{x})|<\infty,\ \sup_{\mathbf{x},\mathbf{y}}\{|g(\mathbf{x})-g(\mathbf{y})|/\|\mathbf{x}-\mathbf{y}\|^\lambda\}\le C<\infty\right\},\ 0<\lambda\le1,$$

satisfies (3.1) with the same λ.

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

$$\nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) = \sup\{ \quad \left|\frac{Cov^*(g_1(X^*_{-d_1+1},\ldots,X^*_0), g_2(X^*_k,\ldots X^*_{k+d_2-1}))}{4\|g_1\|_\infty\|g_2\|_\infty}\right|; \\ g_1\in\mathcal{C}^{d_1},\ g_2\in\mathcal{D}^{d_2}\},$$

with $(\mathcal{C}^{d_1},\lambda_1)$, $(\mathcal{D}^{d_2},\lambda_2)$ satisfying (3.1), where the expectations in (3.1) is taken with respect to the true underlying process $\{X_t\}_{t\in\mathbb{Z}}$.
Fortunately, it suffices to take expectations in (3.1) with respect to the true underlying probability measure so that this condition is verifiable.

3.2 Assumptions and main results

We present now 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 $X_1,\ldots,X_n$.

(A1) Model (1.2) holds with $\Phi(z)=\sum_{j=0}^\infty \phi_j z^j$ bounded away from zero for $|z|\le 1$ ($z\in\mathbb{C}$) and the autoregressive coefficients decay like $|\phi_j| = O(j^{-\beta})$, $\beta>3$ ($j\to\infty$).

(A2) The innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are i.i.d., with $\mathbb{E}|\varepsilon_t|^s<\infty$ and have a distribution which admits a density $f_\varepsilon(.)$ with respect to the Lebesgue measure. Moreover, $\int_{-\infty}^\infty |f_\varepsilon(x)-f_\varepsilon(x+c)|dx \le const.c$, $\forall\ c\in\mathbb{R}$.

As an example, ARMA(p,q) models ($p<\infty$, $q<\infty$) usually satisfy our assumptions (A1) and (A2) with an exponential decay of the coefficients $\{\phi_j\}_{j=0}^\infty$.

Theorem 3.1 *Assume that (A1) and (A2) with s=2 hold. Then*

$$\alpha(k)\le const.k^{-\gamma},\ \gamma<\beta-3/2.$$

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

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

(A3) The kernel K(.) for estimating $f_\varepsilon(.)$ satisfies: $K(.)$ is a density of a probability measure with $\int_{-\infty}^\infty xK(x)dx=0$, $\int_{-\infty}^\infty x^2K(x)dx\ne 0$, $\int_{-\infty}^\infty |K(x)-K(x+c)|dx\le const.c$ $\forall\ c\in\mathbb{R}$, $\int_{-\infty}^\infty |x|^sK(x)dx<\infty$ for the same s as in (A2). Moreover, the bandwidth satisfies

$$h(n)^{-1}=o(n)\ (n\to\infty)$$
$$h(n)^{-1}max\{(log(n)p(n)/n)^{\vartheta/(2(\vartheta+1))}, p(n)^{-\vartheta^2/(\vartheta+1)}\} = O(1)\ (n\to\infty),$$
$$\vartheta<\beta-1,\ \vartheta\in\mathbb{N},$$

for the same β as in (A1).

(A4) $p(n)=o((n/log(n))^{1/(2(\beta-1))})$ for the same β as in (A1).

(A5) The pairs $(\mathcal{C}^{d_1}, \lambda_1)$, $(\mathcal{D}^{d_2}, \lambda_2)$ $(d_1,\ d_2 \in \mathbb{N})$ that come along with the definition of the ν-mixing coefficients satisfy (3.1) for some $\lambda_1, \lambda_2 > 0$, where the expectation in (3.1) is taken with respect to the probability measure of the process defined by (A1) and (A2). Moreover we assume that $\inf_{g_1 \in \mathcal{C}^{d_1}} \|g_1\|_\infty > 0$, $\inf_{g_2 \in \mathcal{D}^{d_2}} \|g_2\|_\infty > 0$.

Assumption (A4) is a usual assumption in autoregressive approximation, cf. An et al. (1982) and Bühlmann (1995a). If the approximating order is chosen by the data through AIC, then Shibata (1980) has shown that $\hat{p}_{AIC} \sim const.n^{1/(2\beta)}$, which satisfies (A4). Assumption (A3) describes the interplay between the bandwidth $h(n)$ and the approximating order $p(n)$, both depending on the sample size n. By taking $p(n) = const.n^{1/(2\beta)}$ (this is of the order of $\hat{p}_{AIC}$) and $h(n) = const.n^{-1/5}$ (this is of the optimal order for estimating $f_\varepsilon(.)$ with respect to the mean square error), (A3) holds for any $\beta > 3$. Our assumption (A3) restricts the bandwidth to be not too small; if the underlying density f_ε is very smooth, we are allowed to take a large bandwidth $h(n)$. Finally, we restrict ourselves to second-order kernels K in (A3) so that we are able to resample from positive densities $\hat{f}_{\hat{\varepsilon}}$.

Theorem 3.2 *Assume that (A1)-(A5) hold with $\beta > 3$ in (A1), $s \geq 4$ in (A2) and $\lambda > 0$ in (A5). Then*

$$\nu^*(k; \mathcal{C}^{d_1}, \mathcal{D}^{d_2}) \leq const.k^{-\gamma^* \frac{s\lambda}{s(1+2\lambda+d)+d}} \text{ in probability,}$$

where $d = d_1 + d_2$, $\gamma^* = ([\beta] - 3)/2$ if $\beta \notin \mathbb{N}$, $\gamma^* = (\beta - 4)/2$ if $\beta \in \mathbb{N}$.

Proof: 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 $\alpha(k) \leq const.k^{-\gamma}$ (see Theorem 3.1), we can always write

$$\nu^*(k; \mathcal{C}^{d_1}, \mathcal{D}^{d_2}) \leq const.k^{-\gamma L},\ L < \frac{(\beta - 4)s\lambda}{(2\beta - 3)(s(1 + 2\lambda + d) + d)}.$$

If (A1) holds for all $s \in \mathbb{N}$, then $L < \lambda(\beta - 4)/((2\beta - 3)(1 + 2\lambda + d))$. Note that often the case $d_1 = d_2 = 1$ $(d = 2)$ and $\lambda = 1$ applies. We further note that the decay of $\nu^*(.; \mathcal{C}^{d_1}, \mathcal{D}_{d_2})$ is still polynomial.

There is also some interest in the case where the autoregressive coefficients ϕ_j in model (1.2) decay exponentially. As examples we mention ARMA(p, q) models $(p < \infty,\ q < \infty)$. 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 ν-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') The same assumptions for the kernel $K(.)$ as in (A3) but the bandwidth satisfies

$$h(n)^{-1} = O(\max\{n^{Clog(1+\Delta)}, n^{1/2-\eta}\}), \text{ for some } \eta > 0,$$
$$0 < \Delta < \min\{\kappa, exp(1/(2C)) - 1\},$$

with the same κ as in (A1') and the same C as in (A4') below.

(A4') $p(n)/(Clog(n)) \to 1$ $(n \to \infty)$, $C \in \mathbb{R}^+$.

Assumption (A1') is almost the same as in Kreiss (1988). Assumption (A4') reflects the behavior of AIC, because $\hat{p}_{AIC} \sim const.log(n)$, 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 (A4') which then would simplify (A3'). The error for estimating $\Phi(z)$ in $|z| \leq 1$ is given by

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

If the behavior of the true coefficients $\{\phi_j\}_{j=0}^{\infty}$ were known, a typical approach would be to choose $p(n)$ such that $\sum_{j=p(n)+1}^{\infty} |\phi_j| \sim const.n^{-1/2}$. Assuming that $|\phi_j| \sim const.(1+\kappa)^{-j}$ $(j \to \infty)$, we then would choose $p(n) = p_\kappa(n) = C_\kappa log(n)$ with $C_\kappa = (2log(1+\kappa))^{-1}$. Then, for the condition on the bandwidth $h(n)$ in (A3'), $\Delta < \kappa$, $n^{C_\kappa log(1+\Delta)} = n^{log(1+\Delta)/(2log(1+\kappa))}$ and hence the only remaining condition on the bandwidth would be

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

Even with less knowledge we can simplify. Suppose we only know κ (but not necessarily the largest κ in (A1')), we then can set $C = C_\kappa = (2log(1+\kappa))^{-1}$ and the only condition on the bandwidth would be as above.

Theorem 3.3 *Assume that (A1') with* $\kappa > 0$ *and (A2) with* $s = 2$ *hold. Then*

$$\alpha(k) \leq const.\rho^k, \ (1+\kappa)^{-1} < \rho < 1,$$

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'), (A4') with* $C \in \mathbb{R}^+$ *and (A5) with* $\lambda > 0$ *hold. Then*

$$\nu^*(k; \mathcal{C}^{d_1}, \mathcal{D}^{d_2}) \leq const.(\rho^*)^k, \ \textit{in probability},$$
$$(1+\tilde{\kappa})^{-\frac{s\lambda}{s(1+2\lambda+d)+d}} < \rho^* < 1, \ d = d_1 + d_2,$$

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 $C = C_\kappa = (2log(1+\kappa))^{-1}$ *in (A4') we have*

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

Proof: 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), a faster decay of the autoregressive coefficients in (A1) and then make use of the Borel-Cantelli Lemma, i.e., one would show complete convergence.

4 Smoothed sieve bootstrap and empirical processes

As an application of the results in section 3.2 we show that for the smoothed sieve bootstrap, as some kind of conditionally stationary and mixing process, some general functional central limit theorems hold. We closely follow the approach in Andrews and Pollard (1994), which considers empirical processes for strong-mixing, stationary processes.

4.1 General empirical process

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 $\{X_t\}_{t\in\mathbb{Z}}$. To show that the bootstrap works for estimating such (smooth) functionals, one therefore has to consider the empirical process based on the vectorized samples $\{\mathbf{X}_t = (X_t, \ldots, X_{t+q-1})\}_{t=1}^{n-q+1}$ and $\{\mathbf{X}_t^* = (X_t^*, \ldots, X_{t+q-1}^*)\}_{t=1}^{n-q+1}$, respectively, cf. Bühlmann (1994). We need then the ν-mixing property with respect to classes of functions: $\mathbb{R}^q \to \mathbb{R}$.

Let $\mathcal{F}^q$ be a class of measurable functions from $\mathbb{R}^q \to \mathbb{R}$. We introduce now 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 σ-algebra of $\mathbb{R}^d$, and by $Pf = \int f(x)dP(x)$ 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, cf. Giné and Zinn (1990)) for the metric induced by $\|.\|_{\mathcal{F}^q}$; here $\|h\|_{\mathcal{F}^q} = \sup_{f\in\mathcal{F}^q} |h(f)|$, where $h : \mathcal{F}^q \to \mathbb{R}$. We restrict ourselves to uniformly bounded classes $\mathcal{F}^q$ which satisfy a bracketing condition in the following sense. Let $\rho(f) = \|f\|_2 = (\mathbb{E}|f(\mathbf{X}_1)|^2)^{1/2}$ be a pseudo-norm on $\mathcal{F}^q$ and denote by $N(.) = N(.; \mathcal{F}^q, \rho)$ the bracketing number, which is defined as

$$
\begin{aligned}
N(\delta) = \min_M \{ \quad & \exists f_1, \ldots, f_M \text{ and } b_1, \ldots b_M \text{ with } \rho(b_i) \le \delta \ \forall i \text{ such that :} \\
& \forall f \in \mathcal{F}^q \ \exists i \text{ for which } |f - f_i| \le b_i \quad \}.
\end{aligned}
$$

A bracketing condition assumes now a certain decay of $N(\delta)$ as a function of δ.

Example 4.1. (Parametric family of Lipschitz functions). Consider the class of functions

$$
\mathcal{F}^q = \{f : \mathbb{R}^q \to \mathbb{R}; f = f(.;\theta),\ \theta \in \Theta\},
$$

with Θ a bounded subset of $\mathbb{R}^k$ such that

$$
\begin{aligned}
& \sup_\theta \|f(.,\theta)\|_\infty < \infty, \\
& \sup_{\mathbf{x},\mathbf{y}} \sup_\theta |\{(f(\mathbf{x},\theta) - f(\mathbf{y},\theta))/\|\mathbf{x}-\mathbf{y}\|^\gamma\} \le C < \infty,\ 0 < \lambda \le 1,\ \|.\| \text{ the Euclidean norm on } \mathbb{R}^q, \\
& |f(\mathbf{x},\theta_1) - f(\mathbf{x},\theta_2)| \le L(\mathbf{x})\|\theta_1 - \theta_2\|^\tau,\ \text{and}\|L\|_2 < \infty,\ \tau > 0,\ \|.\| \text{ the Euclidean norm on } \mathbb{R}^k.
\end{aligned}
$$

Then $\mathcal{F}^q$ satisfies (3.1) with the same λ and the bracketing number satisfies $N(\delta; \mathcal{F}^q, \rho) \le const.\delta^{-k/\tau}$. This example is a straightforward extension of the example in Andrews and Pollard (1994, Sec.2).

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

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

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

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

Lemma 4.1 *(Stochastic Equicontinuity) Assume that (A1)-(A4) hold with $s \geq 4$ in (A2) and that the function class $\mathcal{F}^q$ satisfies (A5) for some $\lambda > 0$. Moreover assume that every $f \in \mathcal{F}^q$ has at most countably many discontinuities. In addition we assume that the parameters s in (A2), λ in (A5) and the dimension q are such that*

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

where $r \in \mathbb{N}$, $C > 0$ and γ^* as in Theorem 3.2. Moreover assume that

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

for the same r and C. Then $\forall\eta > 0\ \exists\delta > 0$ such that

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

Proof: We use the ν-mixing property of $\{X_t^*\}_{t\in\mathbb{Z}}$ with respect to the pair $(\otimes_{i=1}^{2r-1}\mathcal{F}^q, \otimes_{i=1}^{2r-1}\mathcal{F}^q)$ (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 3.2 (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 (A5), $\mathcal{F}^q$ is a uniformity class, cf. Bhattacharya and Rao (1976, Th.2.4). $\square$

Under the conditions of Lemma 4.1 the empirical process $Z_n(.)$ converges weakly to some Gaussian process $Z(.)$, indexed by $\mathcal{F}^q$, with ρ-continuous sample paths and with $\mathbb{E}[Z(f)] = 0$, $f \in \mathcal{F}^q$ and

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

see Andrews and Pollard (1994, Cor.2.3).

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

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

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

$$|\mathbb{P}^*[R_n^*(f_1), \ldots R_n^*(f_h) \le x] - \mathbb{P}[R(f_1), \ldots, R(f_h) \le x]| \le \eta \text{ on the set } A_n,\ \forall n \ge n_0,$$
$$\mathbb{P}[A_n] \to 1\ (n \to \infty),$$

where for each $n \in \mathbb{N}$, the set A_n is universal $\forall\ f_1, \ldots, f_h\ \in \mathcal{F}^q,\ h \in \mathbb{N}$.

Theorem 4.1 *Assume the conditions of Lemma 4.1. Moreover assume fidi-convergence*

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

Then

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

Proof: The result follows directly from fidi-convergence and Lemma 4.1. □

Fidi-convergence of Z_n^* is usually not directly available because $\{X_t^*\}_{t \in \mathbb{Z}}$ satisfies by Theorem 3.2 only a ν-mixing property. This does not allow to use one of the usual blocking techniques.

Theorem 4.2 *Assume the conditions of Lemma 4.1. Then*

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

Proof: It remains to show fidi-convergence

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

We remark here that every $f \in \mathcal{F}^q$ is Q-continuous, Q being the probability measure of $(X_1, \ldots, X_q)$, which admits a density with respect to the Lebesgue measure, i.e., 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 with $h = 1$ and $q = 1$, the general case for $h \in \mathbb{N}$ follows by the Cramér-Wold device, and for $q \in \mathbb{N}$ in a straightforward, but notationally more awkward way. We follow the same strategy as in Bühlmann (1995b, proof of Th. 3.3) 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 $Z_{n,M}^*(.)$ by means of the variables $\{X_{t,M}^*\}_{t=1}^N$. By exploiting the M-dependence we get in a straightforward way as in Bühlmann (1995b),

$$Z_{n,M}^*(f) \xrightarrow{d^*} Z_M(f) \text{ in probability universal over } \mathcal{F}^q. \tag{4.1}$$

Here $Z_M(f)$ 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 $Z_n^*(f)$ by $Z_{n,M}^*(f)$ and $Z(f)$ by $Z_M(f)$ becomes negligible for large M. We first show that

$$\mathbb{P}[Z_M(f) \le c] \to \mathbb{P}[Z(f) \le c] \ (M \to \infty), \ c \in \mathbb{R}. \tag{4.2}$$

Formula (4.2) follows by showing

$$\sum_{k=-M}^{M} Cov(f(X_{0,M}), f(X_{k,M})) \to \sum_{k=-\infty}^{\infty} Cov(f(X_0), f(X_k)) \ (M \to \infty).$$

But this holds true by using the mixing property of $\{X_t\}_{t\in\mathbb{Z}}$, the boundedness and Q-continuity of f and $(X_{0,M}, X_{k,M}) \xrightarrow{d} (X_0, X_k)$ $(M \to \infty)$, c.f. Bhattacharya and Rao (1976, Th 1.3).
Finally we show that $\forall \eta > 0\ \exists M_0(\eta)\ \exists n_0(\eta)$ such that

$$Var^*(Z_{n,M}^*(f) - Z_n^*(f)) \le \eta \text{ on a set } A_{n,1}, \forall n \ge n_0\ \forall M \ge M_0 \tag{4.3}$$

where $A_{n,1}$ is universal $\forall\ f \in \mathcal{F}^q$ and $\mathbb{P}[A_{n,1}] \to 1$ $(n \to \infty)$.
We have by the mixing property of $\{X_{t,M}^*\}_{t\in\mathbb{Z}}$ (see Theorem 3.2, the bounds for the mixing coefficients translate directly to the truncated process $\{X_{t,M}^*\}_{t\in\mathbb{Z}}$),

$$\begin{aligned} & Var^*(Z_{n,M}^*(f) - Z_n^*(f)) \\ \le\ & const.(\mathbb{E}^*|f(X_0^*) - f(X_{0,M}^*) - \mathbb{E}^*[f(X_0^*] + \mathbb{E}^*[f(X_{0,M}^*)]|^{2+\delta})^{2/(2+\delta)} \\ \le\ & const.(\mathbb{E}^*|f(X_0^*) - f(X_{0,M}^*)|^2)^{2/(2+\delta)} \text{ on a set } A_{n,2},\ \delta > 0 \end{aligned}$$

where $A_{n,2}$ is universal $\forall\ f \in \mathcal{F}^q$ and $\mathbb{P}[A_{n,2}] \to 1$ $(n \to \infty)$; we have used the boundedness of $f \in \mathcal{F}^q$ and the covariance inequality in Lemma 3.1 (ii).
Now by using the convergence of the bootstrap probabilities to the original probabilities (see also Lemma 5.5), these convergences holding on some set $A_{n,3}$, universal $\forall f \in \mathcal{F}^q$, with $\mathbb{P}[A_{n,3}] \to 1$ $(n \to \infty)$, we arrive at, cf. Bhattacharya and Rao (1976, Th 1.3),

$$\mathbb{E}^*|f(X_0^*) - f(X_{0,M}^*)|^2 - \mathbb{E}|f(X_0) - f(X_{0,M})|^2 = o(1) \text{ on the set } A_{n,3}. \tag{4.4}$$

But $\mathbb{E}|f(X_0) - f(X_{0,M})|^2 \to 0$ $(M \to \infty)$, hence by setting $A_{n,1} = A_{n,2} \cap A_{n,3}$ we have shown (4.3).
By (4.1)-(4.3) we have shown $Z_n^*(f) \xrightarrow{d^*} Z(f)$ in probability universal over $\mathcal{F}^q$. □

We just remark that by replacing (A1), (A3) and (A4) by (A1'), (A3') and (A4') respectively, we get better bounds on the ν-mixing coefficients and hence need less conditions on the bracketing numbers.

4.2 Empirical process on $\mathbb{R}^q$

We specialize now our results from section 4.1 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 $(-\infty, \mathbf{x}] = \times_{i=1}^{q}(-\infty, x_i]$. By Example 3.1 we know that (A5) holds for $\mathcal{F}^q$ with $\lambda = 1$, if the q-dimensional marginal distribution of the process $\{X_t\}_{t\in\mathbb{Z}}$ has a bounded density. Denote the c.d.f. of $\mathbf{X}_t$ and $\mathbf{X}_t^*$ by $F^{(q)}(.)$ and $F^{(q)*}(.)$, respectively. Defining '$\leq$' componentwise, the empirical process and its bootstrap counterpart can then be written as

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

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

$$\sum_{k=1}^{\infty} k^{2r-2}k^{-\gamma^*\frac{sC}{(s(3+4rq-2q)+4rq-2q)(2r+C)}} < \infty,\ \textit{for some } C>0,\ r\in\mathbb{N}$$
$$\textit{with } r > \frac{2q}{1-C/(2+C)},\ \gamma^* \textit{ as in Theorem 3.2.}$$

Then

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

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

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

Proof: The result is basically a consequence of Theorem 4.2. Note that the assumption $\sup_{x\in\mathbb{R}} f_\varepsilon(x)<\infty$ implies that the q-dimensional marginal distribution of $(X_t,\ldots,X_{t+q})$ has a bounded density and hence $\lambda=1$. We remark that it is sufficient to work in the cadlag-space $\mathcal{D}([0,1]^q)$. This claim follows by applying the Continuous Mapping Theorem to the continuous map

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

cf. Bühlmann (1994, Remark on p.998).
We work with $\lambda=1$ (see above) for the condition in Lemma 4.1, the bracketing condition then holds since the index space $[0,1]^q$ is compact and hence $N(\delta)\leq const.\delta^{-2q}$. $\square$.

5 Proofs

In the sequel we denote by $\mathcal{B}(S)$ the Borel σ-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 ε_t^* $i.i.d.\sim\hat{f}_\varepsilon(x+\hat{\mu}_\varepsilon)dx$. We will show in Lemma 5.3 that $\alpha^*(k)\leq const.h(n)^{-1}k^{-\gamma^*}$ in probability, yielding for $k\geq h(n)^{-1/\zeta}$, $\zeta\in\mathbb{R}^+$,

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

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

$$\begin{aligned} & \nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) \le \alpha(k) \\ + & \sup\{\frac{|Cov^*(g_1(X^*_{-d_1+1},\ldots,X^*_0),g_2(X^*_k,\ldots,X^*_{k+d_2-1}))}{4\|g_1\|_\infty\|g_2\|_\infty} \\ - & \frac{Cov(g_1(X_{-d_1+1},\ldots,X_0),g_2(X_k,\ldots,X_{k+d_2-1}))|}{4\|g_1\|_\infty\|g_2\|_\infty}\}, \end{aligned}$$

where the supremum is over all $g_1 \in \mathcal{C}^{d_1},\ g_2 \in \mathcal{D}^{d_2}$.
The denominator $4\|g_1\|_\infty\|g_2\|_\infty$ can be bounded by a constant, uniformly over $\mathcal{C}^{d_1},\ \mathcal{D}^{d_2}$ by assumption (A5). For bounding the difference of the covariances we introduce now a moment (pseudo-) norm

$$\|Q_1 - Q_2\|_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}} = \sup\{|\int g(\mathbf{x})(dQ_1 - dQ_2)(x)|; g \in \mathcal{C}^{d_1} \otimes \mathcal{D}^{d_2}\}, \tag{5.2}$$

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 $\mathcal{C}\otimes\mathcal{D}$ see (3.2)).
The difference of covariances can now be bounded like

$$\begin{aligned} & |Cov^*(g_1(X^*_{-d_1+1},\ldots,X^*_0),g_2(X^*_k,\ldots,X^*_{k+d_2-1})) \\ - & Cov(g_1(X_{-d_1+1},\ldots,X_0),g_2(X_k,\ldots,X_{k+d_2-1}))| \\ = & \Big|\int_{\mathbb{R}^{d_1 d_2}} g_1 g_2(\mathbf{x})(d\mathbb{P}^* - d\mathbb{P})(\mathbf{x}) \\ - & \int_{\mathbb{R}^{d_1}} g_1(\mathbf{x})d\mathbb{P}^*(\mathbf{x})\int_{\mathbb{R}^{d_2}} g_2(\mathbf{x})d\mathbb{P}^*(\mathbf{x}) + \int_{\mathbb{R}^{d_1}} g_1(\mathbf{x})d\mathbb{P}(\mathbf{x})\int_{\mathbb{R}^{d_2}} g_2(\mathbf{x})d\mathbb{P}(\mathbf{x})\Big| \\ \le & \|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}} + \|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}}(\|g_1\|_\infty + \|g_2\|_\infty). \end{aligned}$$

This means that we bound

$$\nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) \le \alpha(k) + const.\|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}}. \tag{5.3}$$

In Lemma 5.5 we will give the bound $\|\mathbb{P}^*-\mathbb{P}\|_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}} = O_P(b(n))$, 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 (A3) we get $\|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}} = O_P(h(n)^c)$ for some $c \in \mathbb{R}^+$, yielding then for $k \le h(n)^{-1/\zeta}$, $\|\mathbb{P}^* - \mathbb{P}\|_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}} \le const.k^{-\zeta c}$ in probability and hence for $k \le h(n)^{-1/\zeta}$,

$$\nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) \le const.k^{-\zeta c} \text{ in probability.} \tag{5.4}$$

Putting (5.1) and (5.4) together, we minimize over ζ.

We now give some preliminary results. The first one is dealing 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* ε_t *i.i.d.,* $\mathbf{E}[\varepsilon_t] = 0$, $\mathbf{E}|\varepsilon_t|^4 < \infty$. *Suppose that* $\Phi(z)$ *is bounded away from zero for* $|z| \le 1$ $(z \in \mathbb{C})$ *(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 (1995a). Assertion (ii) follows from Theorem 3.2 in Bühlmann (1995a). □

Lemma 5.2 *Assume the conditions of Lemma 5.1, but more generally we assume that $\mathbf{E}|\varepsilon_t|^s < \infty,\ s\geq 4$. Suppose that the kernel $K(.)$ for estimating $f_\varepsilon(.)$ is a probability density and it 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) $\mathbf{E}^*[(\varepsilon_t^*)^w] - \mathbf{E}[(\varepsilon_t)^w] = O_P(h(n)^2) + O_P(p(n)(log(n)/n)^{1/2}) + o_P(p(n)^{-r}),\ w\leq s$.

(ii) $\mathbf{E}^*|\varepsilon_t^*|^s = O_P(1)$.

Proof: We have

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

We write

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

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 $\phi_p = (\phi_{1,n},\ldots,\phi_{p,n})'$ are the solutions of the theoretical Yule-Walker equations , ${}_p\phi_p = -\gamma_p$, cf. Brockwell and Davis (1987, Ch.8.1). Now similarly as in Bühlmann (1995b, proof of Lem. 5.3),

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

cf. Hannan and Kavalieris (1986, Th.2.1).,

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

here we have used Baxter's inequality, cf. Bühlmann (1995a, proof of (3.1)).
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 $(\hat{\varepsilon}_{t,n})^w$.

The 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 $s = 4$ in (A2). Then*

$$\alpha^*(k) \le const.h(n)^{-1}k^{-\gamma^*} \text{ in probability,}$$

where γ^* is defined as in Theorem 3.2.

Proof: We use representation (2.2) and check the conditions in Gorodetskii (1977). His condition (i) follows immediately by (A3), (ii) follows by Lemma 5.2 and (iii) by (A1) and (A4), cf. Bühlmann (1995a, Lem.2.2 and Th.3.1). The constant γ^* 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 (A3) to the only condition $h(n) = o(1),\ h(n)^{-1} = o(n)\ (n \to \infty)$.

5.2 Moment norm between bootstrap and true measure

Denote by $\mathbb{P}_{k;d_1,d_2}[C] = \mathbb{P}[(X_{-d_1+1},\ldots,X_0,X_k,\ldots,X_{k+d_2-1}) \in C]$, $C \in \mathcal{B}(\mathbb{R}^d)$, $d = d_1 + d_2$, $k \in \mathbb{N}$. Analogously we define $\mathbb{P}^*_{k;d_1,d_2}[.]$ for the bootstrap. By the definition of the ν-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),

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

Our next aim is to bound

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

To do so we will compare this quantity with the variation norm of a 'smoothed difference' $\mathbb{P}^*_{k;d_1,d_2} - \mathbb{P}_{k;d_1,d_2}$. The variation 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 sequel 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 $\delta(n) = o(1)$ $(n \to \infty)$ and $\{K_{\delta(n)}\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathbb{R}^d$ with $\sup_{n\in\mathbb{N}} K_{\delta(n)}(\{\|x\| \le \delta(n)\}) > 1/2$ $\forall n \in \mathbb{N}$, $\|.\|$ the Euclidean norm on $\mathbb{R}^d$. Assume that (A1) and (A2) hold and $(\mathcal{C}^{d_1},\lambda_1)$, $(\mathcal{D}^{d_2},\lambda_2)$ satisfy (3.1), with expectations taken with respect to the probability measure of the true underlying process as defined by (A1) and (A2). Then $\forall n \in \mathbb{N}$, $\forall k \in \mathbb{N}$

$$\|\mathbb{P}^*_{k;d_1,d_2} - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}} \le const.\|(\mathbb{P}^*_{k;d_1,d_2} - \mathbb{P}_{k;d_1,d_2}) \star K_{\delta(n)}\|_{V;d} + const.\delta(n)^\lambda,$$

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

Proof: We use formula (11.26) in Bhattacharya and Rao (1976). To bound the covariance norm we need a bound for some type of modulus of oscillation; but our assumption (A5) is exactly tailored to this problem so that we can bound this modulus of oscillation uniformly over the classes $\mathcal{C}^{d_1}$, $\mathcal{D}^{d_2}$ by $const.\delta(n)^{\lambda}$. □

Now we 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 $s \geq 4$ in (A2) and $\lambda > 0$ in (A5). Then $\forall d_1, d_2 \in \mathbb{N}$, $\forall k \in \mathbb{N}$*

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

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 (A3) yield

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

Proof: To simplify notation we always denote by ϑ an integer $< \beta - 1$. By Lemma 5.4 we want to bound

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

We choose $K_{\delta(n)}$ similar as in Bhattacharya and Rao (1976, (13.8)-(13.11)), i.e., $K_{\delta(n)}$ has a density

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

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

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

where $\mathbf{y}\cdot\mathbf{x} = \sum_{i=1}^{d} y_i x_i$, cf. Bhattacharya and Rao (1976, (10.9)).
In the sequel we denote by $J(n;d) = [-2s\delta(n)^{-1}, 2s\delta(n)^{-1}]^d$. Let $C \in \mathcal{B}(\mathbb{R}^d)$. Then by Fourier inversion

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

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

To bound (5.10) much of the work boils down in 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

$$\varphi_{k;d_1,d_2}(\mathbf{x}) = exp(i\mathbf{x}\cdot\mu_X\mathbf{1}) \prod_{j=0}^{\infty} \varphi_\varepsilon(\mathbf{h}_j\cdot\mathbf{x}) \prod_{j=1}^{k+d_2-1} \varphi_\varepsilon(\mathbf{f}_j\cdot\tilde{\mathbf{x}}),$$

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

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

$$\begin{aligned}
& |\varphi^*_{k;d_1,d_2}(\mathbf{x}) - \varphi_{k;d_1,d_2}(\mathbf{x})| \\
\leq\ & |exp(i\mathbf{x}\cdot\bar{X}\mathbf{1}) - exp(i\mathbf{x}\cdot\mu_X\mathbf{1})| + |\prod_{j=0}^{q(n)} \varphi_{\varepsilon^*}(\hat{\mathbf{h}}_j\cdot\mathbf{x}) - \prod_{j=0}^{q(n)} \varphi_\varepsilon(\mathbf{h}_j\cdot\mathbf{x})| \\
+\ & |\prod_{j=q(n)+1}^{\infty} \varphi_{\varepsilon^*}(\hat{\mathbf{h}}_j\cdot\mathbf{x}) - \prod_{j=q(n)+1}^{\infty} \varphi_\varepsilon(\mathbf{h}_j\cdot\mathbf{x})| + |\prod_{j=1}^{k+d_2-1} \varphi_{\varepsilon^*}(\hat{\mathbf{f}}_j\cdot\tilde{\mathbf{x}}) - \prod_{j=1}^{k+d_2-1} \varphi_\varepsilon(\mathbf{f}_j\cdot\tilde{\mathbf{x}})| \\
=\ & I(\mathbf{x}) + II(\mathbf{x}) + III(\mathbf{x}) + IV(\mathbf{x}), \qquad (5.11)
\end{aligned}$$

where $q(n) \to \infty,\ q(n) = o(n)\ (n \to \infty)$.
By a Taylor expansion we get

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

Again by using a Taylor expansion we get

$$\begin{aligned}
\sup_{\mathbf{x}\in J(n;d)} III(\mathbf{x}) &\leq \sup_{\mathbf{x}\in J(n;d)} \sum_{j=q(n)+1}^{\infty} |\varphi_{\varepsilon^*}(\hat{\mathbf{h}}_j\cdot\mathbf{x}) - \varphi_\varepsilon(\mathbf{h}_j\cdot\mathbf{x})| \\
&\leq \sup_{\mathbf{x}\in J(n;d)} \sum_{j=q(n)+1}^{\infty} (|\hat{\mathbf{h}}_j\cdot\mathbf{x}| + |\mathbf{h}_j\cdot\mathbf{x}|) \leq const.\delta(n)^{-1} q(n)^{-\vartheta} \text{ almost surely}, \qquad (5.13)
\end{aligned}$$

where the last inequality follows from Lemma 5.1 (i) and $|\phi_j| = O(j^{-\beta})$ $(j \to \infty)$, which implies $\sum_{j=q(n)+1}^{\infty} |\mathbf{h}_j| = o(q(n)^{-\vartheta})$.
Most work is needed for bounding $II(\mathbf{x})$ (and similarly $IV(\mathbf{x})$). We have

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

By a Taylor expansion we get

$$\begin{aligned}
& \sup_{\mathbf{x}\in J(n;d)} II.1(\mathbf{x}) \leq const.\delta(n)^{-1} \left(\sum_{j=0}^{q(n)} |\hat{\psi}_{j,n} - \psi_{j,n}| + \sum_{j=k}^{k+q(n)+d_2-1} |\hat{\psi}_{j,n} - \psi_{j,n}| \right) \\
=\ & \delta(n)^{-1} \left(O(q(n)p(n)^{-\vartheta}) + O(q(n)(log(n)/n)^{1/2}) \right) \text{ almost surely}, \qquad (5.14)
\end{aligned}$$

where we used Lemma 5.1(ii) for the last inequality.
For bounding $II.2(\mathbf{x})$ we consider

$$\varphi_{\varepsilon^*}(\mathbf{h}_j\cdot\mathbf{x}) - \varphi_\varepsilon(\mathbf{h}_j\cdot\mathbf{x})$$

$$= (n-p)^{-1} \sum_{t=p+1}^{n} exp(i\mathbf{h}_j \cdot \mathbf{x}(\hat{\varepsilon}_{t,n} - \hat{\mu}_\varepsilon)) \int_{-\infty}^{\infty} exp(i\mathbf{h}_j \cdot \mathbf{x} u h(n)) K(u) du - \varphi_\varepsilon(\mathbf{h}_j \cdot \mathbf{x})$$

$$= (n-p)^{-1} \sum_{t=p+1}^{n} exp(i\mathbf{h}_j \cdot \mathbf{x}(\hat{\varepsilon}_{t,n} - \hat{\mu}_\varepsilon))(1 + E(h(n); j, \mathbf{x})) - \varphi_\varepsilon(\mathbf{h}_j \cdot \mathbf{x}), \tag{5.15}$$

where

$$\begin{aligned} |E(h(n); j, \mathbf{x})| &\leq |\mathbf{h}_j \cdot \mathbf{x}| h(n) \int_{-\infty}^{\infty} |u| K(u) du \\ &\leq const. h(n) |\mathbf{h}_j \cdot \mathbf{x}|. \end{aligned}$$

On the other hand,

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

where

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

and hence, see (5.7)-(5.8),

$$(n-p)^{-1} \sum_{t=p+1}^{n} |D(h(n), t; j, \mathbf{x})| = |\mathbf{h}_j \cdot \mathbf{x}| \left(O_P(p(n)(log(n)/n)^{1/2}) + o_P(p(n)^{-\vartheta}) \right). \tag{5.17}$$

(Here the O_P-terms are uniformly bounded in j 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 get by some well known exponential inequalities, e.g. Bernstein's inequality,

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

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

$$\begin{aligned} &\sup_{\mathbf{x} \in J(n;d)} |(n-p)^{-1} \sum_{t=p+1}^{n} exp(i\mathbf{h}_j \cdot \mathbf{x}\varepsilon_t) - \varphi_\varepsilon(\mathbf{h}_j \cdot \mathbf{x})| \\ = \; &O_P(n^{1/2-\eta}), \text{ for any } O < \eta < 1/2, \end{aligned} \tag{5.18}$$

where the O_P-term is uniformly bounded in j.
Therefore by (5.15)-(5.18) we get

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

Hence by (5.14) and (5.19)

$$\begin{aligned}
&\sup_{\mathbf{x}\in J(n;d)} II(\mathbf{x}) && (5.20)\\
=\;& \delta(n)^{-1}\left(O_P(h(n)) + O_P((q(n)+p(n))(log(n)/n)^{1/2}) + O_P(q(n)p(n)^{-\vartheta})\right) \\
+\;& O_P(q(n)n^{-1/2+\eta}),\ \ \eta > 0. && (5.21)
\end{aligned}$$

Similarly we get

$$\sup_{\mathbf{x}\in J(n;d)} IV(\mathbf{x}) = O(\sup_{\mathbf{x}\in J(n;d)} II(\mathbf{x})). \tag{5.22}$$

Then we have by (5.11)-(5.13), (5.20)-(5.22),

$$\begin{aligned}
&\sup_{\mathbf{x}\in J(n;d)} |\varphi^*_{k;d_1,d_2}(\mathbf{x}) - \varphi_{k;d_1,d_2}(\mathbf{x})| \\
\le\;& \delta(n)^{-1}\left(O_P(q(n)^{-\vartheta}) + O_P(h(n)) + O_P((q(n)+p(n))(log(n)/n)^{1/2}) + O_P(q(n)p(n)^{-\vartheta})\right) \\
+\;& O_P(q(n)n^{-1/2+\eta}) \\
=\;& \delta(n)^{-1}\left(O_P(q(n)^{-\vartheta}) + O_P(h(n)) + O_P((q(n)+p(n))(log(n)/n)^{1/2}) + O_P(q(n)p(n)^{-\vartheta})\right),
\end{aligned} \tag{5.23}$$

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

What remains is to integrate the error term in (5.23), see (5.10). Let $r(n) \to \infty$, $r(n) = o(n)$ $(n \to \infty)$. Denote by $C_1 = C \cap [-r(n), r(n)]^d$, $C_2 = C \setminus C_1$. By Markov's inequality we get

$$|\mathbb{P}_{k;d_1,d_2} \star K_{\delta(n)}[C_2]| = O(r(n)^{-s}),\ \ |\mathbb{P}^*_{k;d_1,d_2} \star K_{\delta(n)}[C_2]| = O_P(r(n)^{-s}). \tag{5.24}$$

Hence by (5.10), (5.23) and (5.24)

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

and therefore by Lemma 5.4

$$||\mathbb{P}^*_{k;d_1,d_2} - \mathbb{P}_{k;d_1,d_2}||_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}} = O_P(r(n)^d\delta(n)^{-d-1}\xi(n)) + O_P(r(n)^{-s}) + O(\delta(n)^{\lambda}), \tag{5.25}$$

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 get $\xi(n) = \max\{h(n), p(n)(log(n)/n)^{1/2}, p(n)^{-\vartheta^2/(\vartheta+1)}\}$.
By choosing the optimal orders for $r(n)$ and $\delta(n)$ the right hand side in (5.25) is of the order

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

This completes the proof. $\square$

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

$$\nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) \le \alpha^*(k) \le const.h(n)^{-1}k^{-\gamma^*} \text{ in probability,}$$

and hence,

$$\text{if } k \ge (h(n)^{-1})^{1/\zeta} \text{ then } \nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) \le const.k^{-(\gamma^*-\zeta)},\ \zeta \in \mathbb{R}^+. \tag{5.26}$$

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

$$\begin{aligned}
\nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) &\le \alpha(k) + const.\|\mathbb{P}^*_{k;d_1,d_2} - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}}\\
&\le const.k^{-\gamma} + O_P(h(n)^{\frac{s\lambda}{s(1+\lambda+d)+d}})\\
&\le const.k^{-\gamma} + const.k^{-\frac{\zeta s}{s(2+d)+d}} \text{ in probability, } \zeta \in \mathbb{R}^+. \qquad (5.27)
\end{aligned}$$

By choosing ζ yielding the best rate for $\nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2})$, i.e., $\zeta = \gamma^*(s(1+\lambda+d)+d)/(s(1+2\lambda+d)+d)$, we obtain by (5.26) and (5.27) 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

$$\sup_{n\ge n_0(\omega)} \sum_{j=0}^{p(n)} |\hat{\phi}_{j,n}|(1+\tilde{\kappa})^j < \infty \text{ almost surely, } 0 < \tilde{\kappa} < \min\{\kappa, exp(1/(2C))-1\}. \tag{5.28}$$

We have for any $0 < c < 1/2$,

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

(Use the result of Hannan and Kavalieris (1986, Th.2.1) and Baxter's inequality, compare with (5.7) and (5.8)).
Formula (5.28), together with (A1'), implies that, for n sufficiently large, we can invert $\hat{\Phi}_n(z)$ in $|z| \le 1+\tilde{\kappa}$, we then get instead of Lemma 5.1,

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

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

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

$$\alpha^*(k) \le const.h(n)^{-1}\tilde{\rho}^k, \text{ almost surely, } (1+\tilde{\kappa})^{-1} < \tilde{\rho} < 1. \tag{5.29}$$

(Compare with Theorem 3.3).
Lemma 5.4 remains exactly the same. It is plausible that we get the same bound as in Lemma 5.5, since the assumptions (A1)-(A5) are generally weaker than the assumptions of Theorem 3.4, i.e.,

$$\|\mathbb{P}^*_{k;d_1,d_2} - \mathbb{P}_{k;d_1,d_2}\|_{\mathcal{C}^{d_1},\mathcal{D}^{d_2}} = O_P(h(n)^{\frac{s\lambda}{s(1+\lambda+d)+d}}),\ d = d_1 + d_2. \tag{5.30}$$

However, we have to re-examine the interplay of the tuning parameters $h(n)$ and $p(n)$. Some quantities change now, we choose $q(n) = const.log(n)$ such that the (old) expression $q(n)^{-\vartheta}$ becomes something of the order $n^{-1/2}$. By (A4'), $p(n) \sim Clog(n)$ and instead of the (old) expression $p(n)^{-\vartheta}$ we have $(1+\tilde{\kappa})^{-p(n)}$. Then $\xi(n)$ in (5.25) 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 (5.23).
By choosing $\tilde{\kappa}$ appropriately close to $\min\{\kappa, exp(1/(2C)) - 1\}$, we know that by (A3') $\max\{h(n), log(n)(1+\tilde{\kappa})^{-p(n)}\} = O(h(n))$. This then explains that (5.30) holds.

Now by (5.29), for $k > -log(h(n)^{-1})/log(\tau)$, $\tilde{\rho} < \tau < 1$,

$$\nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) \le const.(\tilde{\rho}/\tau)^k \text{ in probability},$$

and by (5.30), for $k \le -log(h(n)^{-1})/log(\tau)$, $\tilde{\rho} < \tau < 1$

$$\nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) \le const.\tau^{k\frac{s\lambda}{s(1+\lambda+d)+d}} \text{ in probability.}$$

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

$$\begin{aligned}\nu^*(k;\mathcal{C}^{d_1},\mathcal{D}^{d_2}) \le const.\tilde{\rho}^{k\frac{s\lambda}{s(1+2\lambda+d)+d}} &= const.(\rho^*)^k \text{ in probability},\\ &\quad (1+\tilde{\kappa})^{-\frac{s\lambda}{s(1+2\lambda+d)+d}} < \rho^* < 1.\end{aligned}$$

References

[1] An,H.-Z., Chen,Z.-G. and Hannan,E.J. (1982). Autocorrelation, autoregression and autoregressive approximation. Ann. Statist. **10** 926-936 (Corr: **11** p1018).

[2] 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.

[3] Arcones,M. and Giné,E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. Ann. Inst. H. Poincare, Sect. B, Calcul Probab. Statist. **25**, 457-481. (Corr: **27**,583-595).

[4] Bhattacharya,R.N. and Ranga Rao,R. (1976). Normal Approximation and Asymptotic Expansions. Wiley, New York.

[5] Bickel,P.J., Götze,F. and van Zwet,W.R. (1994). Resampling fewer than n observations: gains, losses, and remedies for losses. Tech. Rep. 419, Dept. of Stat., U.C. Berkeley.

[6] Bühlmann,P. (1993). The Blockwise Bootstrap in Time Series and Empirical Processes. PhD thesis, ETH Zürich.

[7] Bühlmann,P. (1994). Blockwise bootstrapped empirical processes for stationary sequences. Ann. Statist. **22** 995-1012.

[8] Bühlmann,P. (1995a). Moving-average representation for autoregressive approximations. Tech. Rep. 423, Dept. of Stat., U.C. Berkeley.

[9] Bühlmann,P. (1995b). Sieve bootstrap for time series. Tech. Rep. 431, Dept. of Stat., U.C. Berkeley.

[10] Bühlmann,P. and Künsch,H.R. (1994). Block length selection in the bootstrap for time series. Res.Rep. 72. Seminar für Statistik, ETH Zürich.

[11] Brockwell,P.J. and Davis,R.A. (1987). Time Series: Theory and Methods. Springer, New York.

[12] Doukhan,P. (1994). Mixing. Properties and Examples. Lect. Notes in Stat. **85**. Springer, New York.

[13] Efron,B. (1979). Bootstrap methods: another look at the jackknife. Ann. Statist. **7** 1-26.

[14] Falk,M. and Reiss,R.-D. (1989a). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. Ann. Probab. **17**, 362- 371.

[15] Falk,M. and Reiss,R.-D. (1989b). Bootstrapping the distance between smooth bootstrap and sample quantile distribution. Prob. Th. Rel. Fields **82**, 177-186.

[16] Giné,E. and Zinn,J. (1990). Bootstrapping general empirical measures. Ann. Probab. **18** 851-869.

[17] Gorodetskii, V.V. (1977). On the strong mixing property for linear sequences. Theory Probab. Appl. **22**, 411-413.

[18] Haeusler,E., Mason,D.M. and Newton,M.A. (1991). Weighted bootstrapping of means. 91CWI Quart. **4**,213-228.

[19] Hall,P., DiCiccio,T.J. and Romano,J.P. (1989). On smoothing and the bootstrap. Ann. Statist. **17**, 692-704.

[20] Hannan,E.J. and Kavalieris,L. (1986). Regression, autoregression models. J. Time Series Anal. **7**, 27-49.

[21] Hjellvik,V. and Tjøstheim,D. (1995). Nonparametric tests of linearity for time series. To appear in Biometrika.

[22] Kreiss,J.-P. (1988). Asymptotic Statistical Inference for a Class of Stochastic Processes. Habilitationsschrift, Universität Hamburg.

[23] Künsch,H.R. (1989). The jackknife and the bootstrap for general stationary observations. Ann. Statist. **17** 1217-1241.

[24] Mason,D.M. and Newton,M.A. (1992). A rank statistics approach to the consistency of a general bootstrap. Ann. Statist. **20**,1611-1624.

[25] Politis,D.N. and Romano,J.P. (1994). Large sample confidence regions based on subsamples under minimal conditions. Ann. Statist. **22** 2031-2050.

[26] Subba Rao,T. and Gabr,M.M. (1980). A test for linearity of stationary time series. J. Time Series Anal. **1**, 145-158.

[27] Rubin,D.B. (1981). The Bayesian bootstrap. Ann. Statist. **9**, 130-134.

[28] Singh,K. (1981). On the asymptotic accuracy of Efron's bootstrap. Ann. Statist. **9** 1187-1195.

[29] Shibata,R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. Ann. Statist. **8** 147-164.

[30] Silverman,B.W. and Young,G.A. (1987). The bootstrap: To smooth or not to smooth? Biometrika **74** 469-479.

[31] Tsay,R.S. (1992). Model checking via parametric bootstraps in time series analysis. Appl. Statist. **41** 1-15.

[32] Yokoyama,R. (1980). Moment bounds for stationary mixing sequences. Z. Wahrsch. verw. Gebiete **52** 45-57.

Department of Statistics
Evans Hall
University of California
Berkeley, CA 94720
USA