

ON THE RANGE OF VALIDITY OF THE AUTOREGRESSIVE SIEVE BOOTSTRAP

JENS-PETER KREISS, EFSTATHIOS PAPARODITIS, AND DIMITRIS N. POLITIS

ABSTRACT. We explore the limits of the autoregressive (AR) sieve bootstrap, and show that its applicability extends well beyond the realm of linear time series as has been previously thought. In particular, for appropriate statistics, the AR-sieve bootstrap is valid for stationary processes possessing a general Wold-type autoregressive representation with respect to a white noise; in essence, this includes all stationary, purely non-deterministic processes, whose spectral density is everywhere positive. Our main theorem provides a simple and effective tool in assessing whether the AR-sieve bootstrap is asymptotically valid in any given situation. In effect, the large-sample distribution of the statistic in question must only depend on the first and second order moments of the process; prominent examples include the sample mean and the spectral density. As a counterexample, we show how the AR-sieve bootstrap is not always valid for the sample autocovariance even when the underlying process is linear.

1. INTRODUCTION

Due to the different possible dependence structures that may occur in time series analysis, several bootstrap procedures have been proposed to infer properties of a statistic of interest. Validity of the different bootstrap procedures depends on the probabilistic structure of the underlying stochastic process $\mathbf{X} = (X_t : t \in \mathbb{Z})$ and/or on the particular statistic considered. Bootstrap schemes for time series rank from those imposing more parametric type assumptions on the underlying stochastic process class to those accounting only for some kind of mixing or weak dependence assumptions. For an overview see Bühlmann (2002), Lahiri (2003), Politis (2003), and Paparoditis and Politis (2009).

Date: May 31, 2011.

2000 Mathematics Subject Classification. Primary 62M10, 62M15; secondary 62G09.

Key words and phrases. Autoregression, bootstrap, time series.

A common assumption is that \mathbf{X} is a *linear time series*, i.e., that

$$(1.1) \quad X_t = \sum_{j=-\infty}^{\infty} b_j e_{t-j}, \quad t \in \mathbb{Z},$$

with respect to independent, identically distributed (i.i.d.) random variables (e_t) —often assumed to have mean zero and finite fourth order moments—, and for absolutely summable coefficients (b_j) ; this is not to be confused with the Wold-representation with respect to white noise, i.e., uncorrelated, errors that all stationary, purely non-deterministic processes possess. If $b_j = 0$ for all $j < 0$, then the linear process is called *causal*.

Stationary autoregressive (AR) processes of order p are members of the linear class (1.1) provided the autoregression is defined on the basis of i.i.d. errors. Model-based bootstrapping in the AR(p) case was among the first bootstrap proposals for time series; see e.g. Freedman (1984). The extension to the AR(∞) case was inevitable; this refers to the situation where the strictly stationary process X_t has the following linear infinite order autoregressive representation

$$(1.2) \quad X_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + e_t, \quad t \in \mathbb{Z},$$

with respect to i.i.d. errors e_t having mean zero, variance $0 < E(e_t^2) = \sigma_e^2$ and $E(e_t^4) < \infty$; here the coefficients π_j are assumed absolutely summable and $\pi(z) = 1 - \sum_{j=1}^{\infty} \pi_j z^j \neq 0$ for $|z| = 1$. The two representations, (1.1) and (1.2) are related; in fact, the class (1.2) is a subset of the linear class (1.1). Furthermore, it can be shown that the linear AR(∞) process (1.2) is causal if and only if $\pi(z) = 1 - \sum_{j=1}^{\infty} \pi_j z^j \neq 0$ for $|z| \leq 1$.

There is already a large body of literature dealing with applications and properties of the AR-sieve bootstrap. Kreiss (1988, 1992) established validity of this bootstrap scheme for different statistics including autocovariances and autocorrelations. Papanoditis and Streitberg (1991) established asymptotic validity of the AR-sieve bootstrap to infer properties of high order autocorrelations, and Papanoditis (1996) established its validity in a multivariate time series context. The aforementioned results required an exponential decay of the AR coefficients π_j as $j \rightarrow \infty$; Bühlmann (1997) extended the class of AR(∞) processes for which the AR-sieve bootstrap works by allowing a polynomially decay of the π_j coefficients. Furthermore, Bickel and Bühlmann (1999) introduced a mixing concept appropriate for investigating properties of the AR-sieve bootstrap which is related to the weak dependence concept of Doukhan and Louhichi (1999), while Choi and Hall (2000) focused on properties of the AR-sieve bootstrap-based confidence intervals.

A basic assumption in the current literature of the AR-sieve bootstrap is that \mathbf{X} is a linear AR(∞) process, that is, X_t is generated by (1.2) with (e_t) being an

i.i.d. process. One exception is the case of the sample mean $\bar{X}_n = n^{-1} \sum_{t=1}^n X_t$, where Bühlmann (1997) proved validity of the AR-sieve bootstrap also for the case where the assumption of i.i.d. errors in (1.2) can be relaxed to that of martingale differences, i.e., $E(e_t | \mathcal{E}_{t-1}) = 0$ and $E(e_t^2 | \mathcal{E}_{t-1}) = \sigma_e^2$ with $\mathcal{E}_{t-1} = \sigma(\{e_s : s \leq t-1\})$ the σ -algebra generated by the random variables $\{e_{t-1}, e_{t-2}, \dots\}$. Notice that the process (1.2) with innovations forming a martingale difference sequence is in some sense not “very far” from the linear process (1.2) with i.i.d. errors. In fact, some authors call the set-up of model (1.1) with martingale difference errors “weak linearity”, and the same would hold regarding (1.2); see e.g. Kokoszka and Politis (2011).

To elaborate, for a causal linear process the general L_2 -optimal predictor of X_{t+k} based on its past X_t, X_{t-1}, \dots , namely the conditional expectation $E(X_{t+k} | X_s, s \leq t)$ of X_{t+k} , is identical to the best *linear* predictor $\mathcal{P}_{\mathcal{M}_t}(X_{t+k})$; here k is assumed positive, and \mathcal{P}_C denotes orthogonal projection onto the set C and $\mathcal{M}_s = \overline{\text{span}}\{X_j : j \leq s\}$, i.e., the closed linear span generated by the random variables $\{X_j : j \leq s\}$. The property of linearity of the optimal predictor is shared by causal processes that are only weakly linear. Recently, under the assumption of weak linearity with (1.2), Poskitt (2007) claimed validity of the AR-sieve bootstrap for a much wider class of statistics that are defined as smooth functions of means. However, this claim does not seem to be correct in general. In particular, our Example 3.2 of Section 3 contradicts Theorem 2 of Poskitt (2007); see Remark 3.2 in what follows.

The aim of the present paper is to explore the limits of the AR-sieve bootstrap, and to give a definitive answer to the question concerning for which classes of statistics, and for which dependence structures, is the AR-sieve bootstrap asymptotically valid. Moreover, we also address the question what the AR-sieve bootstrap really does when it is applied to data stemming from a stationary process not fulfilling strict regularity assumptions such as linearity or weak linearity. In order to do this, we examine in detail in Section 2 processes possessing a so-called general autoregressive representation with respect to white noise errors; these form a much wider class of processes than the linear $\text{AR}(\infty)$ class described by (1.2).

Our theoretical results in Section 3 provide an effective and simple tool for gauging consistency of the AR-sieve bootstrap. They imply that for certain classes of statistics the range of the validity of the AR-sieve bootstrap goes far beyond that of the linear class (1.1). On the other hand, for other classes of statistics, like for instance autocorrelations, validity of the AR-sieve bootstrap is restricted to the linear process class (1.1), while for statistics like autocovariances, the AR-sieve bootstrap is only valid for the linear $\text{AR}(\infty)$ class (1.2). But even in the case of the linear autoregression (1.2) with infinite order, the theory developed in this paper provides a further generalization of existing results since it establishes validity of this bootstrap procedure under the simple assumption of absolute summability of the coefficients

π_j , thus relaxing previous assumptions referring to exponential or polynomial decay of these coefficients.

The remaining of the paper is organized as follows. Section 2 develops the background concerning the Wold-type infinite order AR representation that is required to study the AR-sieve bootstrap, and states the necessary assumptions to be imposed on the underlying process class and on the parameters of the bootstrap procedure. Section 3 presents our main result and discusses its implications by means of several examples. Proofs and technical details are deferred to the final Section 5.

2. THE AR-SIEVE BOOTSTRAP AND GENERAL AUTOREGRESSIVE REPRESENTATION

Here, and throughout the paper, we assume that we have observations X_1, \dots, X_n stemming from a strictly stationary process \mathbf{X} . Let $T_n = T_n(X_1, \dots, X_n)$ be an estimator of some unknown parameter θ of the underlying stochastic process \mathbf{X} . Suppose that for some appropriately increasing sequence of real numbers $\{c_n : n \in \mathbb{N}\}$ the distribution $\mathcal{L}_n = \mathcal{L}(c_n(T_n - \theta))$ has a non-degenerated limit. The AR-sieve bootstrap proposal to estimate the distribution \mathcal{L}_n goes as follows.

Step 1: Select an order $p = p(n) \in \mathbb{N}$, $p \ll n$ and fit a p -th order autoregressive model to X_1, X_2, \dots, X_n . Denote by $\hat{a}(p) = (\hat{a}_j(p), j = 1, 2, \dots, p)$, the Yule-Walker autoregressive parameter estimators, that is $\hat{a}(p) = \hat{\Gamma}(p)^{-1} \hat{\gamma}_p$ where for $0 \leq h \leq p$,

$$\hat{\gamma}_X(h) = n^{-1} \sum_{t=1}^{n-|h|} (X_t - \bar{X}_n)(X_{t+|h|} - \bar{X}_n),$$

$$\bar{X}_n = n^{-1} \sum_{t=1}^n X_t, \hat{\Gamma}(p) = (\hat{\gamma}_X(r-s))_{r,s=1,2,\dots,p} \text{ and } \hat{\gamma}_p = (\hat{\gamma}_X(1), \hat{\gamma}_X(2), \dots, \hat{\gamma}_X(p))'.$$

Step 2: Let $\tilde{\varepsilon}_t(p) = X_t - \sum_{j=1}^p \hat{a}_j(p) X_{t-j}$ $t = p+1, p+2, \dots, n$, be the residuals of the autoregressive fit and denote by \hat{F}_n the empirical distribution function of the centered residuals $\hat{\varepsilon}_t(p) = \tilde{\varepsilon}_t(p) - \bar{\varepsilon}$, where $\bar{\varepsilon} = (n-p)^{-1} \sum_{t=p+1}^n \tilde{\varepsilon}_t(p)$. Let $(X_1^*, X_2^*, \dots, X_n^*)$ be a set of observations from the time series $\mathbf{X}^* = \{X_t^* : t \in \mathbb{Z}\}$ where $X_t^* = \sum_{j=1}^p \hat{a}_j(p) X_{t-j}^* + e_t^*$ and the e_t^* 's are independent random variables having identical distribution \hat{F}_n .

Step 3: Let $T_n^* = T_n(X_1^*, X_2^*, \dots, X_n^*)$ be the same estimator as T_n based on the pseudo-time series $X_1^*, X_2^*, \dots, X_n^*$ and θ^* the analogue of θ associated with the bootstrap process \mathbf{X}^* . The AR-sieve bootstrap approximation of \mathcal{L}_n is then given by $\mathcal{L}_n^* = \mathcal{L}^*(c_n(T_n^* - \theta^*))$.

In the above (and in what follows), $\mathcal{L}^*, E^*, \dots$ will denote probability law, expectation, etc. in the bootstrap world (conditional on the data X_1, \dots, X_n).

Note that the use of Yule-Walker estimators in Step 1 is essential and guarantees—among other things—that the complex polynomial $\hat{A}_p(z) = 1 - \sum_{j=1}^p \hat{a}_j(p) z^j$ has

no roots on or within the unit disc $\{z \in \mathbb{C} : |z| \leq 1\}$, see the discussion before equation (2.22), i.e., the bootstrap process \mathbf{X}^* always is a stationary and causal autoregressive process.

The question considered in this paper is when can the bootstrap distribution \mathcal{L}_n^* correctly approximate the distribution \mathcal{L}_n of interest, and moreover what the AR-sieve bootstrap does if the latter is not the case. To this end, let us first discuss a general autoregressive representation of stationary processes.

Recall that by the well-known Wold representation, every purely nondeterministic, stationary and zero-mean stochastic process $\mathbf{X} = \{X_t : t \in \mathbb{Z}\}$ can be expressed as

$$(2.1) \quad X_t = \sum_{j=1}^{\infty} b_j u_{t-j} + u_t,$$

where $\sum_{j=1}^{\infty} b_j^2 < \infty$ and $u_t = X_t - \mathcal{P}_{\mathcal{M}_{t-1}}(X_t)$ is a zero mean, white noise “innovation” process with finite variance $0 < \sigma_u^2 = E(u_t^2) < \infty$; recall that $\mathcal{M}_s = \overline{\text{span}}\{X_j : j \leq s\}$.

Less known is that for all purely nondeterministic, stationary and zero-mean time series unique autoregressive coefficients $(a_k : k \in \mathbb{N})$ exist that only depend on the autocovariance function of the time series (X_t) , such that for any $n \in \mathbb{N}$,

$$(2.2) \quad \mathcal{P}_{\mathcal{M}_{t-1}}(X_t) = \sum_{k=1}^n a_k X_{t-k} + e_{t,n}, \quad t \in \mathbb{Z},$$

where $(e_{t,n} : t \in \mathbb{Z})$ is stationary and $e_{t,n} \in \overline{\text{span}}\{X_s : s \leq t - n - 1\}$.

Under the additional assumption that the coefficients $(a_k, k \in \mathbb{N})$ are absolute summable, i.e., $\sum_{k=1}^{\infty} |a_k| < \infty$, one then obtains an autoregressive, Wold-type representation of the underlying process given by

$$(2.3) \quad X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \varepsilon_t, \quad t \in \mathbb{Z}.$$

Here again $(\varepsilon_t : t \in \mathbb{Z})$ denotes a white noise, i.e., uncorrelated, process with finite variance $\sigma_\varepsilon^2 = E\varepsilon_t^2$ which fullfills

$$(2.4) \quad \sigma_\varepsilon^2 = \gamma_X(0) - \sum_{k=1}^{\infty} a_k \gamma_X(k),$$

where $\gamma_X(\cdot)$ denotes the autocovariance function of \mathbf{X} .

Under the absolute summability assumption on the autoregressive coefficients (a_k) —conditions for which will be given in Lemma 2.1 in the sequel—we have that $X_t - \mathcal{P}_{\mathcal{M}_{t-1}}(X_t) = X_t - \sum_{k=1}^{\infty} a_k X_{t-k}$; this implies that the white noise process (u_t) appearing in (2.1) coincides with the white noise process (ε_t) in (2.3). Notice that this does not mean that if we have an arbitrary one sided moving average representation of a time series (X_t) , even with summable coefficients, that this moving

average representation is the Wold-representation of the process; see Remark 2.1 for an example. Furthermore, let f_X be the spectral density of \mathbf{X} , i.e.,

$$f_X(\lambda) = (2\pi)^{-1} \sum_{h \in \mathbb{Z}} \gamma(h) \exp\{-i\lambda h\}, \quad \lambda \in [-\pi, \pi].$$

Then, from (2.3) one immediately obtains that

$$(2.5) \quad \left| 1 - \sum_{k=1}^{\infty} a_k e^{-ik\lambda} \right|^2 \cdot f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi}, \quad \lambda \in [-\pi, \pi],$$

which implies that for strictly positive spectral densities f_X the power series $A(z) := 1 - \sum_{k=1}^{\infty} a_k z^k$ has no zeroes with $|z| = 1$. For more details of the autoregressive Wold representation (2.3) see Pourahmadi (2001), Lemma 6.4(b), (6.10) and (6.12). It is worth mentioning that in the historical evolution of Wold decompositions the autoregressive variant preceded the moving average one.

Remark 2.1. If we consider a purely nondeterministic and stationary time series possessing a standard one-sided moving average representation, and if we additionally assume that the spectral density is bounded away from zero and that the moving average coefficients b_j are absolutely summable, then this would imply that the polynomial $B(z) = 1 + \sum_{j=1}^{\infty} b_j z^j$ has no zeroes with magnitude equal to one. There may of course exist zeroes within the unit disk. But since the closed unit disk is compact and $B(z)$ represents a holomorphic function there could exist only finitely many zeroes with magnitude less than one. Following the technique described in Kreiss and Neuhaus (2006), Section 7.13, one may switch to another moving average model for which the polynomial has no zeroes within the unit disk. This procedure definitely changes the white noise process; for example, if the white noise process in the assumed moving average representation consists of independent random variables, this desirable feature typically is lost when switching to the moving average model with all zeroes within the unit disk removed. In fact, only the property of uncorrelatedness is preserved. The modified moving average process allows then for an autoregressive representation of infinite order and this process, because of the uniqueness of the autoregressive representation, coincides with the one in (2.3).

The following simple example, taken from Brockwell and Davis (1991, Example 3.5.2), illustrates these points. Based on i.i.d. random variables (e_t) with mean zero and finite and non-vanishing variance σ_e^2 , construct the simple MA(1)-process

$$(2.6) \quad X_t = e_t - 2e_{t-1}, \quad t \in \mathbb{Z}.$$

This MA(1)-model is not invertible to an autoregressive process. However, a general autoregressive representation as described above exists. In order to obtain this representation denote by L the usual lag-operator and consider $B(L) := 1 - 2L$ as

well as $\tilde{B}(L) := 1 - 0.5L$. Of course

$$(2.7) \quad X_t = \tilde{B}(L) \frac{B(L)}{\tilde{B}(L)} e_t.$$

Since $|B(e^{-i\lambda})|^2/|\tilde{B}(e^{-i\lambda})|^2 = 4$ we obtain that

$$\varepsilon_t := \frac{B(L)}{\tilde{B}(L)} e_t = e_t - \frac{3}{2} \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^{j-1} e_{t-j}.$$

Again (ε_t) is a (uncorrelated) white noise process with variance $\sigma_\varepsilon^2 = 4\sigma_e^2$. Moreover we have

$$(2.8) \quad X_t = \varepsilon_t - 0.5\varepsilon_{t-1} = - \sum_{j=1}^{\infty} 0.5^j X_{t-j} + \varepsilon_t.$$

Obviously $\varepsilon_t = X_t - \mathcal{P}_{\mathcal{M}_{t-1}}(X_t)$ which means that (2.8) and not (2.6) is the Wold-representation of the time series (X_t) . This also means that the modified moving average (or Wold-) representation of the process (X_t) possesses only uncorrelated innovations (ε_t) instead of independent innovations (e_t) . But the representation (2.8) with uncorrelated innovations has the advantage that it indeed possesses an autoregressive representation of infinite order. Of course, via the described modification, we do not change any property of the process (X_t) . But, and this is essential, the modification leading to the general AR(∞)-representation typically destroys a existing independence property of the white noise in a former moving average representation.

To elaborate, the problem of understanding the stochastic properties of the innovation process in linear time series has been thoroughly investigated in the literature. Breidt and Davis (1991) showed that time reversibility of a linear process is equivalent to the fact that the i.i.d. innovations e_t are Gaussian and used this result to derive for a class of linear processes uniqueness of moving average representations with i.i.d. non-Gaussian innovations and to discuss the stochastic properties of the innovation process appearing in alternative moving average representations for the same process class. Breidt et al. (1995) used such results to initialize autoregressive processes in Monte Carlo generation of conditional sample paths running autoregressive processes backward in time and Andrews et al. (2007) for estimation problems for all-pass time series models. Properties of the innovation process in non Gaussian, non-invertible time series have been also discussed in Lii and Rosenblatt (1982) and Lii and Rosenblatt (1996).

As we have seen the variances of e_t and ε_t do not coincide and the same is true for the fourth order cumulant $E(e_t^4)/\sigma_e^4 - 3$ which will be of some importance later. Using the fact that ε_t is defined via a linear transformation on the i.i.d. sequence

(e_t) we obtain by straightforward computation

$$(2.9) \quad \frac{E(\varepsilon_t^4)}{(E(\varepsilon_t^2))^2} - 3 = \frac{2}{5} \frac{E(e_1^4)}{\sigma_e^4} - \frac{6}{5},$$

which only equals $E(e_1^4)/\sigma_e^4 - 3$ in case the latter quantity is equal to 0, e.g., when the e_t are normally distributed. The normally distributed case always leads to the fact that uncorrelatedness and independence are equivalent, thus implying that the white noise process in the general autoregressive representation always consists of independent and normally distributed random variables which leads for the autoregressive sieve bootstrap in some cases to a considerable simplification as we will see later. \square

In order to get conditions which ensure the absolute summability of the autoregressive coefficients ($a_k, k \in \mathbb{N}$) one can go back to an important paper by Baxter (1962). Informally speaking it is the smoothness of the spectral density f_X which ensures summability of these coefficients. To be more precise we have the following result.

Lemma 2.1.

(i) *If f_X is strictly positive and continuous and if $\sum_{h=0}^{\infty} h^r |\gamma_X(h)| < \infty$ for some $r \geq 0$ then*

$$(2.10) \quad \sum_{h=0}^{\infty} h^r |a_h| < \infty.$$

(ii) *If f_X is strictly positive and possesses $k \geq 2$ derivatives then*

$$(2.11) \quad \sum_{h=0}^{\infty} h^r |\gamma_X(h)| < \infty \quad \forall r < k - 1.$$

PROOF Cf. Baxter (1962), p. 140 and 142. \square

The uniquely determined autoregressive coefficients (a_k) are closely related to the coefficients of an optimal (in the mean square sense) autoregressive fit of order p , or equivalently, to prediction coefficients based on the finite past. To be precise denote the minimizers of

$$(2.12) \quad E \left(X_t - \sum_{r=1}^p c_r X_{t-r} \right)^2$$

by $a_1(p), \dots, a_p(p)$, which of course are solutions of the following Yule-Walker linear equations

$$(2.13) \quad \begin{pmatrix} \gamma_X(0) & \cdots & \gamma_X(p-1) \\ \vdots & \ddots & \vdots \\ \gamma_X(p-1) & \cdots & \gamma_X(0) \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_p \end{pmatrix} = \begin{pmatrix} \gamma_X(1) \\ \vdots \\ \gamma_X(p) \end{pmatrix}.$$

Recall from Brockwell and Davis (1991), Prop. 5.1.1, that the covariance matrix $\Gamma(p)$ on the left hand side is for all p invertible provided $\gamma_X(0) > 0$ and $\gamma_X(h) \rightarrow 0$ as $h \rightarrow \infty$.

Now by slight modifications of Baxter (1962), Theorem 2.2 (cf. also Pourahmadi (2001), Theorem 7.22) we obtain the following helpful result relating the coefficients $a_k(p)$ of the p th order autoregressive fit to the (a_k) of the general autoregressive representation.

Lemma 2.2. *Assume that f_X is strictly positive and continuous and that $\sum_{h=0}^{\infty} (1+h)^r |\gamma_X(h)| < \infty$ for some $r \geq 0$. Then there exists $p_o \in \mathbb{N}$ and $C > 0$ (both depending on f_X only) such that for all $p \geq p_o$:*

$$(2.14) \quad \sum_{k=0}^p (1+k)^r |a_k(p) - a_k| \leq C \cdot \sum_{k=p+1}^{\infty} (1+k)^r |a_k|,$$

as well as

$$(2.15) \quad \sum_{k=1}^{\infty} (1+k)^r |a_k| < \infty.$$

This means that we typically can achieve a polynomial rate of convergence of $a_k(p)$ towards a_k .

As already mentioned, $\gamma_X(0) > 0$ and $\gamma_X(h) \rightarrow 0$ as $h \rightarrow \infty$ ensure non-singularity of all autocovariance matrices appearing in the left hand side of (2.13). Since these matrices are positive semidefinite this means that under these conditions $\Gamma(p)$ actually is positive definite. This in turn with Kreiss and Neuhaus (2006), Section 8.7, implies that the polynomial $A_p(z) = 1 - \sum_{k=1}^p a_k(p)z^k$ has no zeroes in the closed unit disk. We can even prove a slightly stronger result.

Lemma 2.3. *Assume that f_X is strictly positive and continuous, that $\sum_{h=0}^{\infty} |\gamma_X(h)| < \infty$ and $\gamma_X(0) > 0$. Then there exists $\delta > 0$ and $p_o \in \mathbb{N}$ such that for all $p \geq p_o$:*

$$(2.16) \quad \inf_{|z| \leq 1 + \frac{1}{p}} \left| 1 - \sum_{k=1}^p a_k(p)z^k \right| \geq \delta > 0.$$

The uniform convergence of $A_p(z)$ towards $A(z)$ on the closed unit disk immediately implies the following corollary to Lemma 2.3.

Corollary 2.1. *Under the assumption of Lemma 2.3 we have*

$$(2.17) \quad A(z) = 1 - \sum_{j=1}^{\infty} a_j z^j \neq 0, \quad \forall |z| \leq 1.$$

Lemma 2.3 and Corollary 2.1 now enable us to invert the power series $A(z)$ as well as the polynomial $A_p(z)$. Let us denote

$$(2.18) \quad \left(1 - \sum_{j=1}^{\infty} a_j z^j\right)^{-1} = 1 + \sum_{j=1}^{\infty} \alpha_j z^j, \quad \forall |z| \leq 1$$

and for all p large enough (because of Lemma 2.3)

$$(2.19) \quad \left(1 - \sum_{j=1}^p a_j(p) z^j\right)^{-1} = 1 + \sum_{j=1}^{\infty} \alpha_j(p) z^j, \quad \forall |z| \leq 1 + \frac{1}{p}.$$

From (2.19) one immediately obtains that

$$(2.20) \quad |\alpha_j(p)| \leq C \cdot \left(1 + \frac{1}{p}\right)^{-j} \quad \forall j \in \mathbb{N}.$$

A further auxiliary result contains the transfer of the approximation property of $a_j(p)$ for a_k to the respective coefficients $\alpha_k(p)$ and α_k of the inverted series. For such a result we make use of a weighted version of Wiener's lemma, cf. Gröchenig (2007).

Lemma 2.4. *Under the assumptions of Lemma 2.3 and additionally $\sum_{h=0}^{\infty} (1+h)^r |\gamma_X(h)| < \infty$ for some $r \geq 0$ there exists a constant $C > 0$ such that for all p large enough*

$$(2.21) \quad \sum_{j=1}^{\infty} (1+j)^r |\alpha_j(p) - \alpha_j| \leq C \cdot \sum_{j=p+1}^{\infty} (1+j)^r |a_j| \rightarrow_{p \rightarrow \infty} 0.$$

In a final step of this section we now move on to estimators of the coefficients $a_k(p)$. The easiest one might think of is to replace in (2.13) the theoretical autocovariance function by its sample version $\hat{\gamma}_X(h)$. Denote the resulting Yule-Walker estimators of $a_k(p)$ by $\hat{a}_k(p)$, $k = 1, \dots, p$. These Yule-Walker estimators are under the typically satisfied assumption that $\hat{\gamma}_X(0) > 0$ uniquely determined and moreover fulfill (by the same arguments already used) the desired property that

$$(2.22) \quad \hat{A}_p(z) = 1 - \sum_{k=1}^p \hat{a}_k(p) z^k \neq 0, \quad \forall |z| \leq 1.$$

Thus we can also invert the polynomial $\hat{A}_p(z)$ and we denote

$$(2.23) \quad \left(1 - \sum_{k=1}^p \hat{a}_k(p) z^k\right)^{-1} = 1 + \sum_{k=1}^{\infty} \hat{\alpha}_k(p) z^k, \quad |z| \leq 1.$$

We require that the estimators $(\widehat{a}_k(p) : k = 1, \dots, p)$ converge—even at a slow rate—to their theoretical counterparts, namely

(A1) $p(n)^2 \cdot \sum_{1 \leq k \leq p(n)} |\widehat{a}_k(p(n)) - a_k(p(n))| = \mathcal{O}_P(1)$, where $p(n)$ denotes a sequence of integers converging to infinity at a rate to be specified.

Assumption **(A1)** for example is met if a sufficient fast rate of convergence for the empirical autocovariances towards their theoretical counterparts can be guaranteed. The convergence property of $\widehat{a}_k(p)$ carries over to the corresponding coefficients $\widehat{\alpha}_k(p)$ of the inverted polynomials (cf. (2.23)) as is specified in the following Lemma.

Lemma 2.5. *Under the assumptions of Lemma 2.3 and **(A1)** we have uniformly in $k \in \mathbb{N}$*

$$(2.24) \quad |\widehat{\alpha}_k(p(n)) - \alpha_k(p(n))| \leq \left(1 + \frac{1}{p(n)}\right)^{-k} \frac{1}{p(n)^2} \mathcal{O}_P(1).$$

3. VALIDITY OF THE AR-SIEVE BOOTSTRAP

3.1. Functions of Generalized Means. Consider a general class of estimators

$$(3.1) \quad T_n = f \left(\frac{1}{n-m+1} \sum_{t=1}^{n-m+1} g(X_t, \dots, X_{t+m-1}) \right),$$

discussed in Künsch (1989), cf. Example 2.2; here $g : \mathbb{R}^m \rightarrow \mathbb{R}^d$ and $f : \mathbb{R}^d \rightarrow \mathbb{R}$. For this class of statistics, Bühlmann (1997) proved validity of the AR-sieve bootstrap under the main assumption of an invertible linear process with i.i.d. innovations for the underlying process (X_t) ; this means a process which admits an autoregressive representation (1.2). The class of statistics given in (3.1) is quite rich and contains, for example, versions of sample autocovariances, autocorrelations, partial autocorrelations, Yule-Walker estimators and the standard sample mean as well.

The necessary smoothness assumptions on the functions f and g are described below; these are identical to the ones imposed by Bühlmann (1997).

(A2) $f(y)$ has continuous partial derivatives for all y in a neighborhood of $\theta = E g(X_t, \dots, X_{t+m-1})$ and the differentials $\sum_{i=1}^m \partial f(y) / \partial x_i |_{x=\theta}$ do not vanish. The function g has continuous partial derivatives of order h ($h \geq 1$) that satisfy a Lipschitz condition.

We intend to investigate in this section what the autoregressive sieve bootstrap really mimics if it is applied to statistics of the form (3.1) and the observations X_1, \dots, X_n

do *not* stem from a linear $\text{AR}(\infty)$ process (1.2). To be precise, we only assume that we observe X_1, \dots, X_n from a process satisfying the following assumption **(A3)**.

(A3) $(X_t : t \in \mathbb{Z})$ is a zero mean, strictly stationary and purely nondeterministic stochastic process. The autocovariance function γ_X satisfies $\sum_{h=0}^{\infty} h^r |\gamma_X(h)| < \infty$ for some $r \in \mathbb{N}$ specified in the respective results and the spectral density f_X is bounded and strictly positive. Furthermore $E(X_t^4) < \infty$.

Notice that the processes class described by **(A3)** is large enough and includes several of the commonly used linear and nonlinear time series models including stationary and invertible autoregressive moving-average (ARMA) processes, ARCH processes, GARCH processes and so on. Summability of the autocovariance function implies that the spectral density f_X exists, is bounded and continuous. We also added in **(A3)** the assumption of finite fourth order moments of the time series. This assumption seems to be unavoidable due to the autoregressive parameter estimation involved in Step 1 of the AR-sieve bootstrap procedure and in regard of assumption **(A1)**.

From Section 2 we know that if (X_t) satisfies assumption **(A3)** then it possesses an autoregressive representation with an uncorrelated white noise process (ε_t) : cf. equation (2.3). Because of the strict stationarity of (X_t) we have that the time series (ε_t) is strictly stationary as well and thus that the marginal distribution $\mathcal{L}(\varepsilon_t)$ of ε_t does not depend on t .

Theorem 3.1 is the main result of this section. To state it, we define the *companion* autoregressive process $\tilde{\mathbf{X}} = (\tilde{X}_t : t \in \mathbb{Z})$ where \tilde{X}_t is generated as follows

$$(3.2) \quad \tilde{X}_t = \sum_{j=1}^{\infty} a_j \tilde{X}_{t-j} + \tilde{\varepsilon}_t \quad t \in \mathbb{Z};$$

here $(\tilde{\varepsilon}_t)$ consists of i.i.d. random variables whose marginal distribution of $\tilde{\varepsilon}_t$ is identical to that of ε_t from 2.3, i.e., $\mathcal{L}(\tilde{\varepsilon}_t) = \mathcal{L}(\varepsilon_t)$. It is worth mentioning that all second order properties of (\tilde{X}_t) and (X_t) , like autocovariance function and spectral density, coincide while the probabilistic characteristics beyond second order quantities of both stationary processes are not necessarily the same. Now, let \tilde{T}_n be the same statistic as T_n defined in (3.1) but with X_t replaced by \tilde{X}_t , i.e.,

$$(3.3) \quad \tilde{T}_n = f \left(\frac{1}{n-m+1} \sum_{t=1}^{n-m+1} g \left(\tilde{X}_t, \dots, \tilde{X}_{t+m-1} \right) \right),$$

The main message of Theorem 3.1 is that the AR-sieve bootstrap applied to data X_1, \dots, X_n in order to approximate the distribution of the statistic (3.1) will generally lead to an asymptotically consistent estimation of the distribution of the statistic

\tilde{T}_n . This implies that for the class of statistics (3.1), the AR-sieve bootstrap will work *if and only if* the limiting distributions of T_n and of \tilde{T}_n are identical.

Theorem 3.1. *Assume assumptions (A1), (A2), (A3) for $r = 1$ and the moment condition $E \varepsilon_t^{2(h+2)}$ (cf. (A2) for the definition of h and (2.3) for the definition of ε_t), the condition $p(n) = o((n/\log n)^{1/4})$ on the order of the approximating autoregression to the data and the following two further assumptions*

- (A4) *The empirical distribution function F_n of the random variables $\varepsilon_1, \dots, \varepsilon_n$ converges weakly to the distribution function F of $\mathcal{L}(\varepsilon_1)$.*
- (A5) *The empirical moments $1/n \sum_{t=1}^n \varepsilon_t^r$ converge in probability to $E \varepsilon_1^r$ for all $r \leq 2(h+2)$.*

Then,

$$(3.4) \quad d_K \left(\mathcal{L}^* \left(\sqrt{n}(T_n^* - f(\theta^*)) \right), \mathcal{L} \left(\sqrt{n}(\tilde{T}_n - f(\tilde{\theta})) \right) \right) = o_P(1)$$

as $n \rightarrow \infty$. Here $\theta^* = E g(X_t^*, \dots, X_{t+m-1}^*)$, $\tilde{\theta} = E g(\tilde{X}_t, \dots, \tilde{X}_{t+m-1})$ and d_K denotes the Kolmogorov distance.

Some remarks are in order.

Remark 3.1.

- (i) Assumption (A4) does imply that we need some conditions on the dependence structure of the random variables ε_t . For instance, a standard mixing condition on (ε_t) suffices to ensure (A4): cf. Politis, Romano and Wolf (1999), Theorem 2.1.
- (ii) Assumption (A5) on the empirical moments is fulfilled if we ensure that sufficiently high empirical moments of the underlying strictly stationary time series X_t itself would converge in probability to their theoretical counterparts.
- (iii) As we already pointed out, Theorem 3.1 states that the AR-sieve bootstrap mimics the behavior of the companion autoregressive process (\tilde{X}_t) as far as statistics of the form (3.1) are considered. Of course if (X_t) is a linear process with i.i.d. innovations and if the corresponding moving average representation is invertible leading to an infinite order autoregression (1.2) with i.i.d. innovations, then the AR-sieve bootstrap works asymptotically as is already known. Moreover, for general process satisfying assumption (A3), if we are in the advantageous situation that the existing dependence structure of the innovations (ε_t) appearing in the general AR(∞) representation (2.3) does not show up in the limiting distribution of T_n , then Theorem 3.1 implies that the AR sieve bootstrap works. We will illustrate this point by several examples later on.

Now we discuss relevant specializations of Theorem 3.1. Notice that the advantage of this theorem is that in order to check validity of the AR-sieve bootstrap, one only needs to check whether the asymptotic distributions of $T_n = T_n(X_1, \dots, X_n)$ based on the observed time series is identical to the distribution of the statistic $\tilde{T}_n = T_n(\tilde{X}_1, \dots, \tilde{X}_n)$ based on fictitious observations $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ from the companion process $\tilde{\mathbf{X}}$. If and only if this is the case, then the AR-sieve bootstrap works asymptotically.

Example 3.1. (Sample Mean) Consider the case of the sample mean $T_n = \bar{X}_n \equiv n^{-1} \sum_{t=1}^n X_t$ and recall that under standard and mild regularity conditions (e.g. mixing or weak dependence) we typically obtain that the sample mean of stationary time series satisfies $\sqrt{n}T_n \Rightarrow N(0, \sum_{h=-\infty}^{\infty} \gamma_X(h))$ as $n \rightarrow \infty$ where \Rightarrow stands for weak convergence. Thus the asymptotic distribution of the sample mean depends only on the second order properties of the underlying process \mathbf{X} and since the companion process $\tilde{\mathbf{X}}$ has the same second order properties as \mathbf{X} we immediately get by Theorem 3.1 that the AR-sieve bootstrap asymptotically works in the case of the mean for general stationary time series for which the spectral density is strictly positive. Even the strict stationarity is not necessary in this case. This is a novel and significant extension of the results of Bühlmann (1997).

Example 3.2. (Sample Autocovariances) For $0 \leq h < n$, let $T_n = \hat{\gamma}(h) \equiv n^{-1} \sum_{t=1}^{n-h} (X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n)$ be the sample autocovariance at lag h . Let us assume that $\sum_{h_1, h_2, h_3=-\infty}^{\infty} |\text{cum}(X_t, X_{t+h_1}, X_{t+h_2}, X_{t+h_3})| < \infty$ holds and denote by

$$f_4(\lambda_1, \lambda_2, \lambda_3) = \sum_{h_1, h_2, h_3=-\infty}^{\infty} \text{cum}(X_t, X_{t+h_1}, X_{t+h_2}, X_{t+h_3}) e^{i \sum_{r=1}^3 \lambda_r h_r}$$

the fourth order cumulant spectrum of \mathbf{X} . Under standard and mild regularity conditions, see for instance Dahlhaus (1985), Theorem 2.1 and Taniguchi and Kakizwa (2000), Chapter 6.1, it is known that $\sqrt{n}(T_n - \gamma(h)) \Rightarrow N(0, \tau^2)$ where

$$\tau^2 = 2\pi \int_{-\pi}^{\pi} 4 \cos(\lambda h)^2 f^2(\lambda) d\lambda + \int \int 4 \cos(\lambda_1 h) \cos(\lambda_2 h) f_4(-\lambda_1, \lambda_2, -\lambda_2) d\lambda_1 d\lambda_2.$$

Notice that in contrast to the case of the sample mean, the limiting distribution of the sample autocovariance depends also on the fourth order moment structure of the underlying process \mathbf{X} . Now, to check validity of the AR-sieve bootstrap we have to derive the asymptotic distribution of the sample autocovariances for the companion autoregressive process (\tilde{X}_t) . This can be easily done, since the autoregressive polynomial in the general autoregressive representation (2.3) is always invertible (cf. Corollary 2.1) from which we immediately get a one-sided moving average representation with i.i.d. innovations $(\tilde{\varepsilon}_t)$ for the companion process (\tilde{X}_t) . Furthermore, the fourth order cumulant spectrum of (\tilde{X}_t) is given by

$$\tilde{f}_4(\lambda_1, \lambda_2, \lambda_3) = (2\pi)^{-3} \left(\frac{E \varepsilon_1^4}{(E \varepsilon_1^2)^2} - 3 \right) \alpha(\lambda_1) \alpha(\lambda_2) \alpha(\lambda_3) \alpha(-\lambda_1 - \lambda_2 - \lambda_3),$$

where $\alpha(\lambda) = \sum_{j=0}^{\infty} \alpha_j \exp\{-ij\lambda\}$ and the coefficients (α_k) are those appearing in (2.18); see Section 2. Thus for the sample autocovariance \tilde{T}_n we get from Brockwell and Davis (1991), Prop. 7.3.1, that $\sqrt{n}(\tilde{T}_n - \gamma(h)) \Rightarrow N(0, \tilde{\tau}^2)$ where

$$(3.5) \quad \begin{aligned} \tilde{\tau}^2 &= 2\pi \int_{-\pi}^{\pi} 4 \cos(\lambda h)^2 f^2(\lambda) d\lambda + \int \int 4 \cos(\lambda_1 h) \cos(\lambda_2 h) \tilde{f}_4(-\lambda_1, \lambda_2, -\lambda_2) d\lambda_1 d\lambda_2 \\ &= \left(\frac{E \varepsilon_1^4}{(E \varepsilon_1^2)^2} - 3 \right) (\gamma(h))^2 + \sum_{k=-\infty}^{\infty} (\gamma(k)^2 + \gamma(k+h) \gamma(k-h)). \end{aligned}$$

Since the variances τ^2 and $\tilde{\tau}^2$ of the asymptotic distributions of T_n and \tilde{T}_n do not coincide in general, we conclude by Theorem 3.1 that the AR-sieve bootstrap fails for sample autocovariances. Notice that this failure is due to the fact that in general the limiting distribution of sample autocovariances depends additionally on the fourth order moment structure f_4 of the underlying process \mathbf{X} , and this structure may substantially differ from that of the companion process $\tilde{\mathbf{X}}$.

Interestingly enough, even if the underlying process \mathbf{X} is a linear time series, i.e., satisfies (1.1), the AR-sieve bootstrap may fail for the sample autocovariances. To see why, note that from the aforementioned proposition of Brockwell and Davis (1991), the asymptotic distribution of T_n satisfies $\sqrt{n}(T_n - \gamma(h)) \Rightarrow N(0, \tau_L^2)$ where τ_L^2 is given by

$$(3.6) \quad \tau_L^2 = \left(\frac{E e_1^4}{(E e_1^2)^2} - 3 \right) \left(\gamma(h)^2 + \sum_{k=-\infty}^{\infty} (\gamma(k)^2 + \gamma(k+h) \gamma(k-h)) \right).$$

Special attention is now due to the factor of the first summand of (3.6), which is the fourth order cumulant of the i.i.d. process (e_t) . Recall the asymptotic distribution of the sample autocovariances for the companion autoregressive process (\tilde{X}_t) and especially its variance given in (3.5). The two asymptotic variances given in (3.6) and (3.5) are in general not the same since the fourth order cumulant of the two innovation processes (e_t) and (ε_t) are not necessarily the same. We refer to Remark 2.1 for an example. Of course, all appearing autocovariances are identical since we do not change the second order properties of the process when switching from (X_t) to (\tilde{X}_t) . The consequence is that for sample autocovariances, the AR-sieve bootstrap generally does not work even for linear processes of the type (1.1) and this is true even if the process is causal; see Remark 2.1 and example (2.6).

Remark 3.2. If the innovations $\{e_t\}$ in (1.1) are not necessarily i.i.d. but form a martingale difference sequence, then we are in the above described situation more general than in the case of (1.1) with i.i.d. innovations and thus the limiting distribution of sample autocovariances and also of statistics of the type (3.1), is not correctly mimicked by the AR-sieve bootstrap. This contradicts Theorem 2 of Poskitt (2007).

We conclude this example by mentioning that in the case where the i.i.d. innovations e_t in (1.1) are normally distributed, it follows that the random variables ε_t are Gaussian as well, and that in both expressions (3.6) and (3.5) the fourth order cumulants appearing as factors in the first summands vanish. This means that for the special case of Gaussian time series fulfilling assumption (1.1) the AR-sieve bootstrap works.

Example 3.3. (Sample Autocorrelations) Consider the estimator $T_n = \widehat{\rho}(h) \equiv \widehat{\gamma}(h)/\widehat{\gamma}(0)$ of the autocorrelation $\rho_X(h) = \gamma_X(h)/\gamma_X(0)$. Due to the fact that for general processes satisfying assumption (A3) the limiting distribution of T_n depends also on the fourth order moment structure of the underlying process \mathbf{X} , cf. Theorem 3.1 of Romano and Thombs (1996), which is not mimicked correctly by the companion process $\widetilde{\mathbf{X}}$, the AR-sieve bootstrap fails. Fortunately the situation for autocorrelations is much better if we switch to linear processes of type (1.1). From Theorem 7.2.1 of Brockwell and Davis (1991), we obtain that $\sqrt{n}(T_n - \rho_X(h)) \Rightarrow N(0, v^2)$ where the asymptotic variance is given by Bartlett's formula:

$$v^2 = \sum_{k \in \mathbb{Z}} \left\{ (1 + 2\rho_X^2(h))\rho_X^2(k) + \rho_X(k-h)\rho_X(k+h) - 4\rho_X(h)\rho_X(k)\rho_X(k+h) \right\}.$$

As it is seen, in this case the asymptotic variance depends only on the autocorrelation function $\rho_X(\cdot)$ (or equivalently on the standardized spectral density $\gamma_X^{-1}(0)f_X(\cdot)$) of the underlying process \mathbf{X} . This means that the first summand in (3.6) which refers to the fourth order cumulant of the i.i.d. innovation process e_t in (1.1) does not show up in the limiting distribution of sample autocorrelations and this in turn leads to the fact that the asymptotic distribution of sample autocorrelations based on observations stemming from the process \mathbf{X} is identical to that of the sample autocorrelation based on observations stemming from the companion autoregressive process $\widetilde{\mathbf{X}}$. This is true since the companion autoregressive process $\widetilde{\mathbf{X}}$ shares all second order properties of the process \mathbf{X} . Hence, the AR-sieve bootstrap works for the autocorrelations given data from a linear process (1.1). We stress here the fact that this result is true regardless whether the representation (1.1) allows for an autoregressive inversion or not and holds even though the probabilistic properties of the underlying process (X_t) and of the autoregressive companion process (\widetilde{X}_t) beyond second order properties are not the same.

Remark 3.3. Similarly to the autocorrelation case, the validity of the AR-sieve bootstrap in the linear class (1.1) is shared by many different statistics whose large-sample distribution depends only on the (first and) second order moment structure of the underlying process. Examples include the partial autocorrelations or Yule-Walker estimators of autoregressive coefficients.

Remark 3.4. Consider statistics of the type (3.1) in the easiest case where $d = 1$, i.e.,

$$(3.7) \quad \frac{1}{n-m+1} \sum_{t=1}^{n-m+1} g(X_t, \dots, X_{t+m-1})$$

for a function $g : \mathbb{R}^m \rightarrow \mathbb{R}$. Here, the practitioner may approach this as the sample mean of observations Y_1, \dots, Y_{n-m+1} where $(Y_t = g(X_t, \dots, X_{t+m-1}) : t \in \mathbb{Z})$. Notice that strict stationarity as well as mixing properties easily carries over from (X_t) to (Y_t) . Thus, we may apply the AR-sieve bootstrap to the sample mean of Y_t , which works under quite general assumptions; cf. Remark 3.1 and Theorem 3.1. The only crucial assumption for establishing asymptotic consistency of the AR-sieve bootstrap is that we need the property that the spectral density f_Y of the transformed time series (Y_t) is strictly positive and continuous. Although this is not a very restrictive condition, it may be difficult to check since there seems to be no general result describing the behavior of spectral densities under nonlinear transformations. \square

3.2. Integrated Periodograms. The considerations of Example 3.2 and 3.3 can be transferred to integrated periodogram estimators for these quantities which lead us to the second large class of statistics that we will discuss. Denote, based on observations X_1, \dots, X_n , the periodogram $I_n(\lambda)$ defined by

$$(3.8) \quad I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t e^{-i\lambda t} \right|^2, \quad \lambda \in [0, \pi].$$

and consider a general class of integrated periodogram estimators defined by

$$(3.9) \quad M(I_n, \varphi) = \int_0^\pi \varphi(\lambda) I_n(\lambda) d\lambda,$$

where φ denotes an appropriately defined function on $[0, \pi]$. Under the main assumption that the underlying process \mathbf{X} has the representation (1.1), Dahlhaus (1985) investigated asymptotic properties of (3.9) and obtained the asymptotic distribution of $\sqrt{n}(M(I_n, \varphi) - M(f_X, \varphi))$. In particular, it has been shown that $\sqrt{n}(M(I_n, \varphi) - M(f_X, \varphi))$ converges, as $n \rightarrow \infty$, to a Gaussian distribution with zero mean and variance given by

$$(3.10) \quad (E e_1^4 / (E e_1^2)^2 - 3) \left(\int_0^\pi \varphi(\lambda) f_X(\lambda) d\lambda \right)^2 + 2\pi \int_0^\pi \varphi^2(\lambda) f_X^2(\lambda) d\lambda.$$

Notice that substituting $\varphi(\lambda)$ by $2 \cos(\lambda h)$, $h \in \mathbb{N}_0$, implies that $M(I_n, \varphi)$ equals the sample autocovariance of the observations X_1, \dots, X_n at lag h and (3.10) would then exactly turn to be the asymptotic variance given in (3.6).

As we will see in Theorem 3.2 the situation for integrated periodograms (3.9) is rather similar to that of empirical autocovariances which are of course special cases of integrated periodograms. Thus we only discuss briefly this rather relevant class

of statistics. As Theorem 3.2 shows, we obtain for this class that the AR-sieve bootstrap asymptotically mimics the behavior of $\sqrt{n}(M(\tilde{I}_n, \varphi) - M(f_X, \varphi))$, where \tilde{I}_n is defined as I_n with X_t replaced by the companion autoregressive time series \tilde{X}_t , that is

$$(3.11) \quad \tilde{I}_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n \tilde{X}_t e^{-i\lambda t} \right|^2, \quad \lambda \in [0, \pi].$$

Theorem 3.2. *Assume (A1) and (A3) with $r = 1$ and assume that for all $M \in \mathbb{N}$*

$$(3.12) \quad \mathcal{L}^* \left(\sqrt{n} (\hat{\gamma}_{X^*}(h) - E^* \hat{\gamma}_{X^*}(h)) : h = 0, 1, \dots, M \right) \Rightarrow N(0, V_M)$$

(in probability), where $\hat{\gamma}_{X^*}(h)$ denote the empirical autocovariances of the bootstrap observations $X_1^*, X_2^*, \dots, X_n^*$ and where

$$(3.13) \quad V_M = \left[\left(\frac{E \varepsilon_1^4}{(E \varepsilon_1^2)^2} - 3 \right) \gamma(i)\gamma(j) + \sum_{k=-\infty}^{\infty} \gamma(k)\gamma(k-i+j) + \gamma(k+j)\gamma(k-i) \right]_{i,j=0}^M.$$

Then we obtain for all φ bounded and with bounded variation that (in probability)

$$(3.14) \quad d_K \left(\mathcal{L}^* \left(\sqrt{n}(M(I_n^*, \varphi) - M(\hat{f}_{AR}, \varphi)) \right), \mathcal{L} \left(\sqrt{n}(M(\tilde{I}_n, \varphi) - M(f_X, \varphi)) \right) \right) \rightarrow 0.$$

Here $\hat{f}_{AR}(\lambda) = \frac{\hat{\sigma}^2(p(n))}{2\pi} \left| 1 - \sum_{j=1}^{p(n)} \hat{a}_j(p(n)) e^{-ij\lambda} \right|^{-2}$, $\lambda \in [0, \pi]$, and $\hat{\sigma}^2(p(n))^2 = E \hat{\varepsilon}_1(p(n))^2$, cf. Step 2 in the definition of the AR-sieve bootstrap procedure. Moreover the limiting Gaussian distribution of $\sqrt{n}(M(I_n^*, \varphi) - M(\hat{f}_{AR}, \varphi))$ possesses the following variance

$$(3.15) \quad (E \varepsilon_1^4 / (E \varepsilon_1^2)^2 - 3) \left(\int_0^\pi \varphi(\lambda) f_X(\lambda) d\lambda \right)^2 + 2\pi \int_0^\pi \varphi^2(\lambda) f_X^2(\lambda) d\lambda.$$

Remark 3.5.

- (i) Assumptions under which (3.12) is fulfilled are given in Theorem 3.1 since sample autocovariances belong to the class (3.1).
- (ii) Theorem 3.2 implies that if $\int_0^\pi \varphi(\lambda) f(\lambda) d\lambda = 0$ then the AR-sieve bootstrap asymptotically works for the integrated periodogram statistics $M(I_n, \varphi)$ for time series fulfilling (1.1). This follows immediately by a comparison of the asymptotic variances (3.10) and (3.15). Clearly, the same result holds true if the underlying time series is normally distributed since in this case both innovation processes, (ε_t) and (e_t) , are Gaussian and therefore the fourth order cumulants vanish. In all other cases the AR-sieve bootstrap does not work in general, since the fourth order cumulant $E \varepsilon_1^4 / (E \varepsilon_1^2)^2 - 3$ does not necessarily coincide with $E e_1^4 / (E e_1^2)^2 - 3$; see Example 3.2.

Relevant statistics for which we can take advantage of the condition $\int_0^\pi \varphi(\lambda) f(\lambda) d\lambda = 0$ are the so-called ratio statistics which are defined by

$$(3.16) \quad R(I_n, \varphi) = \frac{M(I_n, \varphi)}{\int_0^\pi I_n(\lambda) d\lambda}.$$

For this class of statistics, Dahlhaus and Janas (1996) showed that under the same assumptions of a linear process of type (1.1) one obtains that $\sqrt{n}(R(I_n, \varphi) - R(f_X, \varphi))$ has a Gaussian limiting distribution with mean zero and variance given by

$$(3.17) \quad \frac{2\pi \int_0^\pi \psi^2(\lambda) f_X^2(\lambda) d\lambda}{\left(\int f_X(\lambda) d\lambda\right)^4}, \quad \text{where } \psi(\lambda) = \varphi(\lambda) \int f_X(\lambda) d\lambda - \int \varphi(\lambda) f_X(\lambda) d\lambda.$$

Thus, exactly as in the case of sample autocorrelations the fourth order cumulant term (cf. (3.10)) of the i.i.d. innovation process disappears and therefore again the following corollary to Theorem 3.2 is true. This corollary states that the AR-sieve bootstrap works for ratio statistics under the quite general assumption that the underlying process is a linear time series (1.1) with i.i.d. innovations and a strictly positive spectral density which under model (1.1) is always continuous.

Corollary 3.1. *Under the assumptions of Theorem 3.2 we have that (in probability)*

$$(3.18) \quad d_K \left(\mathcal{L}^* \left(\sqrt{n}(R(I_n^*, \varphi) - R(\hat{f}_{\text{AR}}, \varphi)) \right), \mathcal{L} \left(\sqrt{n}(R(I_n, \varphi) - R(f, \varphi)) \right) \right) \rightarrow 0.$$

Moreover, the limiting Gaussian distribution of $\sqrt{n}(R(I_n^, \varphi) - R(\hat{f}_{\text{AR}}, \varphi))$ possesses the variance given in (3.17).*

Another class of integrated periodogram estimators is that of nonparametric estimators of the spectral density f_X which are obtained from (3.9) if we allow for the function φ to depend on n . In particular let $\varphi_n(\lambda) = K_h(\omega - \lambda)$ for some $\omega \in [0, \pi]$ where $h = h(n)$ is a sequence of positive numbers (bandwidths) approaching zero as $n \rightarrow \infty$, $K_h(\cdot) = h^{-1}K(\cdot/h)$ and K is a kernel function satisfying the following assumption

(A6) K is a nonnegative kernel function with compact support $[-\pi, \pi]$. The Fourier transform k of K is assumed to be a symmetric, continuous and bounded function satisfying $k(0) = 2\pi$ and $\int_{-\infty}^\infty k^2(u) du < \infty$.

Denote by $f_{n,X}$ be the resulting integrated periodogram estimator, that is,

$$(3.19) \quad f_{n,X}(\omega) = \int_{-\pi}^\pi K_h(\omega - \lambda) I_n(\lambda) d\lambda.$$

Notice that the asymptotic properties of the estimator (3.19) of the spectral density are identical to those of its discretized version $\hat{f}_{n,X}(\omega) = (nh)^{-1} \sum_j K_h(\omega - \lambda_j) I_n(\lambda_j)$, where $\lambda_j = 2\pi j/n$ are the Fourier frequencies, as well as of so-called lag-window estimators; cf. Priestley (1981).

Now, let $f_{n,X}^*$ be the same estimator as (3.19) based on the AR-sieve bootstrap periodogram $I_n^*(\lambda) = (2\pi n)^{-1} |\sum_{t=1}^n X_t^* \exp\{i\lambda t\}|^2$. We then obtain the following Theorem.

Theorem 3.3. *Under the assumptions of Theorem 3.2 with $r = 2$ and assumption (A6), we have that (in probability)*

$$(3.20) \quad d_K \left(\mathcal{L}^* \left(\sqrt{nh} (f_{n,X}^*(\lambda) - \widehat{f}_{\text{AR}}(\lambda)) \right), \mathcal{L} \left(\sqrt{nh} (f_{n,X}(\lambda) - f_X(\lambda)) \right) \right) \rightarrow 0.$$

Moreover, conditionally on X_1, X_2, \dots, X_n ,

$$(3.21) \quad \sqrt{nh} E^* (f_{n,X}^*(\lambda) - \widehat{f}_{\text{AR}}(\lambda)) \rightarrow \begin{cases} 0 & , \text{ if } n^{-1/5}h \rightarrow 0 \\ \frac{1}{4\pi} f_X''(\lambda) \int u^2 K(u) du & , \text{ if } n^{-1/5}h \rightarrow 1, \end{cases}$$

where f_X'' denotes the second derivative of f_X and

$$(3.22) \quad nh \text{Var} (f_{n,X}^*(\lambda)) \rightarrow (1 + \delta_{0,\pi}) f_X^2(\lambda) (2\pi)^{-1} \int K^2(u) du,$$

where $\delta_{0,\pi} = 1$ if $\lambda = 0$ or π and $\delta_{0,\pi} = 0$ otherwise.

Recall that under Assumption **(A3)** it has been shown under different regularity conditions (see Shao and Wu (2007)) that $\sqrt{nh}(f_{n,X}(\lambda) - f_X(\lambda))$ converges to a Gaussian distribution with mean and variance given by the expression on the right hand side of (3.21) and (3.22) respectively. Thus the above theorem implies that for spectral density estimators like (3.19), the AR-sieve bootstrap asymptotically is valid for a very broad class of stationary time series that goes far beyond the linear processes class (1.1).

Corollary 3.1 and Theorem 3.3 highlight an interesting relation between frequency domain bootstrap procedures, like for instance those proposed by Franke and Härdle (1992), and Dahlhaus and Janas (1996) and the AR-sieve bootstrap. Notice that the basic assumptions imposed on the underlying process \mathbf{X} for such a frequency domain bootstrap procedure to be valid are that the underlying process satisfies (1.1) with a strictly positive spectral density f_X . Furthermore, validity of such a frequency domain procedure has been established only for those statistics for which their limiting distribution does not depend on the fourth order moment structure of the innovation process e_t in (1.1). Thus such a frequency domain bootstrap essentially works for statistics like ratio statistics or nonparametric estimators of the spectral density like (3.19). The results of this section, that is Theorem 3.2, Corollary 3.1 and Theorem 3.3, imply that if the underlying stationary process satisfies (1.1) and if the spectral density is strictly positive then the AR-sieve bootstrap works in the same cases in which the frequency domain bootstrap procedures work.

4. CONCLUSIONS

In this paper we have investigated the range of validity of the AR-sieve bootstrap. Based on a quite general Wold-type autoregressive representation we provided a simple and effective tool for verifying whether or not the AR-sieve bootstrap asymptotically works. The central question is to what extent the complex dependence structure of the underlying stochastic process shows up in the (asymptotic) distribution of the relevant statistical quantities. If the asymptotic behavior of the statistic of interest based on our data series is identical to that of the same statistic based on data generated from the companion autoregressive process, then the AR-sieve bootstrap leads to asymptotically correct results.

The family of estimators that have been considered ranges from simple arithmetic means and sample autocorrelations to quite general sample means of functions of the observations as well as spectral density estimators and integrated periodograms. Our concrete findings concerning validity of the AR-sieve bootstrap are different for different statistics. Generally speaking, if the asymptotic distribution of a relevant statistic is determined solely by the first and second order moment structure, then the AR-sieve bootstrap is expected to work. Thus, validity of the AR-sieve bootstrap does not require that the underlying stationary process obeys a linear $AR(\infty)$ representation (with i.i.d. or martingale difference errors) as was previously thought. Indeed, for many statistics of interest, the range of the validity of the AR-sieve bootstrap goes far beyond this subclass of linear processes. In contrast, we point out the possibility that the AR-sieve bootstrap may fail even though the data series *is* linear; a prominent example is the sample autocovariance in the case of the data arising from a non-causal $AR(p)$ or a non-invertible $MA(q)$ model.

Finally, our results bear out an interesting analogy between frequency domain bootstrap methods and the AR-sieve method. In the past, both of these methodologies have been thought to work only in the linear time series setting. Nevertheless, we have just shown the validity of the AR-sieve bootstrap for many statistics of interest without the assumption of linearity, e.g. under the general assumption (A3) and some extra conditions. In recent literature, some examples have been found where the frequency domain bootstrap also works without the assumption of a linear process; see e.g. the case of spectral density estimators studied by Shao and Wu (2007). By analogy to the AR-sieve results of the paper, it can be conjectured that frequency domain bootstrap methods might also be valid without the linearity assumption as long as the statistic in question has a large-sample distribution depending only on first and second order moment properties; cf. Kirch and Politis (2011) for some results in that direction.

5. APPENDIX: AUXILIARY RESULTS AND PROOFS

PROOF of Lemma 2.2. From Baxter (1962), Theorem 2.2, in a slightly more general version given in Baxter (1963), Theorem 1.1, we obtain for arbitrary submultiplicative weight or norm functions $\nu(k) \geq 1$, i.e. $\nu(n) \leq \nu(m) \cdot \nu(n - m)$ for all n, m , that the following bound holds true for all $p \in \mathbb{N}$ and a constant $C > 0$

$$(5.1) \quad \sum_{k=0}^p \nu(k) \left| \frac{a_k(p)}{\sigma^2(p)} - \frac{a_k}{\sigma_\varepsilon^2} \right| \leq C \cdot \sum_{k=p+1}^{\infty} \nu(k) \left| \frac{a_k}{\sigma_\varepsilon^2} \right|.$$

Here $\sigma^2(p) = E(X_t - \sum_{k=1}^p a_k(p) X_{t-k})^2 \leq \sigma^2(0)$ for all $p \in \mathbb{N}$ and $\sigma^2(p) \rightarrow \sigma_\varepsilon^2$ (cf. (2.4)) as $p \rightarrow \infty$.

Since $\nu(k) = (1+k)^r$ is submultiplicative for all $r \geq 0$, cf. Gröchenig (2007), Lemma 2.1, we obtain from (5.1)

$$(5.2) \quad \begin{aligned} & \sum_{k=0}^p (1+k)^r |a_k(p) - a_k| \\ & \leq \sum_{k=0}^p (1+k)^r \left| \frac{a_k(p)}{\sigma^2(p)} - \frac{a_k}{\sigma_\varepsilon^2} \right| \cdot \sigma^2(p) + \sum_{k=0}^p (1+k)^r |a_k| \left| \frac{1}{\sigma_\varepsilon^2} - \frac{1}{\sigma^2(p)} \right| \cdot \sigma^2(p) \\ & \leq \frac{C \sigma^2(0)}{\sigma_\varepsilon^2} \left(1 + \sum_{k=0}^{\infty} (1+k)^r |a_k| \right) \cdot \sum_{k=p+1}^{\infty} (1+k)^r |a_k|, \end{aligned}$$

which is the assertion of Lemma 2.2. To see the last bound observe that because of $a_0(p) = a_0 = 1$ we can bound $\left| \frac{1}{\sigma^2(p)} - \frac{1}{\sigma_\varepsilon^2} \right|$ by the right hand side of (5.1) as well. \square

PROOF of Lemma 2.3. As mentioned just in front of the statement of Lemma 2.3 we have $A_p(z) = 1 - \sum_{k=1}^p a_k(p) z^k$ for all $|z| \leq 1$. Now assume that (2.16) is false. Then there exists a sequence $\{p(k) : k \in \mathbb{N}\} \subset \mathbb{N}$, $p(k) \rightarrow \infty$ and a sequence $\{z_k : k \in \mathbb{N}\}$ of complex numbers with $|z_k| \leq 1 + 1/p(k)$ such that

$$(5.3) \quad A_{p(k)}(z_k) \rightarrow_{k \rightarrow \infty} 0.$$

Let us further assume that we can find a subsequence of $\{z_k : k \in \mathbb{N}\}$ which completely stays within the closed unit disk. Without loss of generality assume that $\{z_k\}$ itself has this property. Since we have $A_p(z) \neq 0 \forall |z| \leq 1$ and because A_p is holomorphic the minimum principle of holomorphic functions leads to

$$(5.4) \quad |A_p(z)| \geq \min_{|z|=1} |A_p(z)| \quad \forall |z| \leq 1.$$

The set $\{z \in \mathbb{C} \mid |z| = 1\}$ is compact and $|A_p|$ is continuous, thus there exists a z_p^* with $|z_p^*| = 1$ and $|A_p(z_p^*)| = \min_{|z|=1} |A_p(z)|$.

Without loss of generality assume that z_p^* converges to a complex number z^* with $|z^*| = 1$. From the above we have

$$(5.5) \quad |A_{p(k)}(z_{p(k)}^*)| \leq |A_{p(k)}(z_k)| \rightarrow_{k \rightarrow \infty} 0.$$

Writing

$$(5.6) \quad A(z^*) = A(z^*) - A(z_{p(k)}^*) + A(z_{p(k)}^*) - A_{p(k)}(z_{p(k)}^*) + A_{p(k)}(z_{p(k)}^*),$$

having in mind that $A_p(z)$ converges to $A(z)$ uniformly on the closed unit disk because of Lemma 2.2 and regarding (5.5) as well as the continuity of $A(z)$ we finally obtain $A(z^*) = 0$ which is a contradiction to $A(z) \neq 0$ for all $|z| = 1$ (cf. below (2.5)).

Since we cannot find a subsequence of z_k completely staying in the unit disk it exists a subsequence $(z_{k'})$ that completely stays in the region $1 < |z| \leq 1 + 1/p(k')$. Again assume without loss of generality that $k' = k$ and that z_k converges to some z_o which necessarily must fulfill $|z_o| = 1$.

We will show that $A(z_o) = 0$ holds, which again is a contradiction to $A(z) \neq 0$ for all $|z| = 1$. To this end let us write $A(z_o)$ in the following way

$$(5.7) \quad A(z_o) = A_{p(k)}(z_k) + \sum_{j=1}^{p(k)} (a_j(p(k)) - a_j) z_k^j + \sum_{j=1}^{p(k)} a_j (z_k^j - z_o^j) - \sum_{j=p(k)+1}^{\infty} a_j z_o^j.$$

The first summand on the right hand side converges to zero by (5.3) and the last summand is bounded through $\sum_{j=p(k)+1}^{\infty} |a_j| \rightarrow 0$ as $k \rightarrow \infty$. The second summand in turn is bounded by

$$(5.8) \quad \sup_{|z| \leq 1 + \frac{1}{p}} \left| \sum_{j=1}^p (a_j(p) - a_j) z^j \right| \leq \sum_{j=1}^p |a_j(p) - a_j| \sup_{|z| \leq 1 + \frac{1}{p}} |z|^p \rightarrow_{p \rightarrow \infty} 0.$$

For the third and last summand, which reads

$$(5.9) \quad \sum_{j=1}^{\infty} a_j (z_k^j - z_o^j) 1\{j \leq p(k)\},$$

one obtains by dominated convergence (recall that $|z_k|^j$ for $j \leq p(k)$ is bounded by 3 and that $z_k \rightarrow z_o$) also convergence to zero.

This concludes the proof of Lemma 2.3. \square

PROOF of Lemma 2.4. Under the assumptions the autoregressive coefficients a_k have the following property:

$$(5.10) \quad \mathbf{a} := (1, -a_1, -a_2, \dots) \in \ell_1^v := \left\{ (z_j : j \in \mathbb{N}_0) \subset \mathbb{C} \mid \sum_{j=0}^{\infty} (1+j)^r |z_j| < \infty \right\}.$$

Because of $1 - \sum_{j=1}^{\infty} a_j z^j \neq 0 \forall |z| \leq 1$, cf. Corollary 2.1, we have from Gröchenig (2007), Theorem 6.2, that a multiplicative inverse $\mathbf{a}^{-1} \in \ell_1^v$ exists. For this result observe that our weight function $\nu(k) = (1+k)^r$ satisfies the so-called Gelfand-Raikov-Shilov (GRS) condition

$$(5.11) \quad \nu(nk)^{1/n} \rightarrow 1 \text{ as } n \rightarrow \infty,$$

cf. Gröchenig (2007), Lemma 2.1.

Since multiplication here is the usual convolution of sequences we have that $\mathbf{a}^{-1} = (1, \alpha_1, \alpha_2, \dots)$, where the coefficients α_k coincide with the power series coefficients of $(1 - \sum_{k=1}^{\infty} a_k z^k)^{-1}$. The assertion $\mathbf{a}^{-1} \in \ell_1^v$ then just means that we have for the coefficients α_k in (2.19)

$$(5.12) \quad \sum_{j=0}^{\infty} (1+j)^r |\alpha_j| < \infty.$$

Exactly along the same lines we obtain (cf. (2.19))

$$(5.13) \quad \begin{aligned} \mathbf{a}(p)^{-1} &= (1, -a_1(p), -a_2(p), \dots, -a_p(p), 0, \dots)^{-1} \\ &= (1, \alpha_1(p), \alpha_2(p), \dots) \in \ell_1^v. \end{aligned}$$

We have

$$\begin{aligned} & \sum_{k=0}^{\infty} (1+k)^r |\alpha_k(p) - \alpha_k| \\ &= |\mathbf{a}(\mathbf{p})^{-1} - \mathbf{a}^{-1}|_{\ell_1^v} \\ &= |\mathbf{a}(\mathbf{p})^{-1}(\mathbf{a} - \mathbf{a}(\mathbf{p}))\mathbf{a}^{-1}|_{\ell_1^v} \\ &= |\mathbf{a}(\mathbf{p})^{-1} - \mathbf{a}^{-1}|_{\ell_1^v} |\mathbf{a} - \mathbf{a}(\mathbf{p})|_{\ell_1^v} + |\mathbf{a}^{-1}|_{\ell_1^v} + |\mathbf{a}^{-1}|_{\ell_1^v} |\mathbf{a} - \mathbf{a}(\mathbf{p})|_{\ell_1^v} \end{aligned}$$

Simple algebra finally leads to

$$\sum_{k=0}^{\infty} (1+k)^r |\alpha_k(p) - \alpha_k| = |\mathbf{a}(\mathbf{p})^{-1} - \mathbf{a}^{-1}|_{\ell_1^v} \leq \frac{|\mathbf{a}^{-1}|_{\ell_1^v}^2 |\mathbf{a} - \mathbf{a}(\mathbf{p})|_{\ell_1^v}}{1 - |\mathbf{a}^{-1}|_{\ell_1^v} |\mathbf{a} - \mathbf{a}(\mathbf{p})|_{\ell_1^v}}.$$

Recall that $|\mathbf{a} - \mathbf{a}(\mathbf{p})|_{\ell_1^v} = \sum_{k=1}^p (1+k)^r |a_k - a_k(p)| + \sum_{k=p+1}^{\infty} (1+k)^r |a_k|$ in order to obtain from Lemma 2.2 the desired assertion. \square

PROOF of Lemma 2.5. For simplicity we write p instead of $p(n)$. From Lemma 2.3 we have that the polynomial $A_p(z)$ has no zeroes with magnitude less than or equal to $1 + 1/p$. Since we easily get from assumption **(B)** convergence of $\widehat{A}_p(z) = 1 - \sum_{k=1}^p \widehat{a}_k(p) z^k$ to $A_p(z)$ uniformly on the closed disk with radius $1 + 1/p(n)$ the polynomial $\widehat{A}_p(z)$ does not possess zeroes with magnitude less than or equal to $1 + 1/p(n)$. Therefore Cauchy's inequality for holomorphic functions applies and

yields

$$\begin{aligned}
|\widehat{\alpha}_k(p) - \alpha_k(p)| &\leq \frac{1}{(1 + 1/p)^k} \max_{|z|=1+1/p} \left| A_p(z)^{-1} - \widehat{A}_p(z)^{-1} \right| \\
&= \frac{1}{(1 + 1/p)^k} \max_{|z|=1+1/p} \frac{|A_p(z) - \widehat{A}_p(z)|}{|A_p(z)\widehat{A}_p(z)|} \\
&\leq \frac{1}{(1 + 1/p)^k} \max_{|z|=1+1/p} \frac{\sum_{k=1}^p |\widehat{a}_k(p) - a_k(p)|(1 + \frac{1}{p})^k}{|A_p(z)\widehat{A}_p(z)|} \\
&= \left(1 + \frac{1}{p}\right)^{-k} \cdot \frac{1}{p^2} \cdot \mathcal{O}_P(1).
\end{aligned}$$

□

PROOF of Theorem 3.1. A careful inspection of the proof of Theorem 3.3 in Bühlmann (1997) (see also the corresponding technical report Bühlmann (1995)) shows that only the following properties of the underlying, the companion and the fitted autoregressive process really are needed:

- (i) $\varepsilon_t^* \xrightarrow{\mathcal{D}^*} \widetilde{\varepsilon}_t$ in probability
- (ii) $(X_{t_1}^*, \dots, X_{t_d}^*) \xrightarrow{\mathcal{D}^*} (\widetilde{X}_{t_1}, \dots, \widetilde{X}_{t_d})$ in probability
- (iii) $\sum_{j=0}^{\infty} |\widehat{\alpha}_j(p(n)) - \alpha_j| \rightarrow_{n \rightarrow \infty} 0$ in probability
- (iv) $\sum_{j=0}^{\infty} j |\widehat{\alpha}_j(p(n))|$ is uniformly bounded in probability
- (v) $\sum_{j=0}^{\infty} j |\alpha_j| < \infty$
- (vi) the empirical moments of $\widehat{\varepsilon}_t(p(n))$ converge for orders up to $2(h+2)$ to the moments of $\widetilde{\varepsilon}_1$
- (vii) the autoregressive representation of infinite order of the process (\widetilde{X}_t) is invertible
- (viii) Yule-Walker parameter estimators are used for the autoregressive fit of order $p(n)$ to the data X_1, \dots, X_n .

Because of **(A4)** and **(A5)** and the easily obtained fact that for the Mallows metric d_2

$$(5.14) \quad d_2 \left(\widehat{F}_n, F_n \right) \rightarrow 0 \quad \text{in probability,}$$

where \widehat{F}_n denotes the empirical distribution function of the centered residuals $\widehat{\varepsilon}_t(p(n))$, $t = p(n) + 1, \dots, n$ of the autoregressive fit and F_n denotes the empirical distribution function of fictitious observations $\varepsilon_t(p(n))$, $t = p(n) + 1, \dots, n$, we obtain (i).

(ii) is obtained exactly along the lines as in Corollary 5.6 of Bühlmann (1997).

To see (iii) recall from Section 2 that we have $|\widehat{\alpha}_j(p(n)) - \alpha_j(p(n))| = (1 + 1/p(n))^{-j} \cdot 1/p^2 \mathcal{O}_P(1)$ as well as $\sum_{j=1}^{\infty} j |\alpha_j(p(n)) - \alpha_j| \leq C \cdot \sum_{j=p(n)}^{\infty} j |a_j|$ which converges to

0 as n goes to infinity. These two assertions ensure (iii).

(iv) is obtained exactly along these lines by using the fact that $\sum_{j=1}^{\infty} j |\alpha_j| \leq \infty$, cf. (2.29), which also is (v).

Furthermore it is easy to see that the difference of the empirical moments (up to the necessary order) of $\widehat{\varepsilon}_t(p(n))$ and of ε_t converge to zero due to the bounds for $\widehat{\alpha}_j(p) - \alpha_j(p)$, $\alpha_j(p) - \alpha_j$ and α_j , cf. (2.24), (2.21) and (5.12). Together with **(A5)** we obtain (vi).

Finally we use for the autoregressive fit Yule-Walker parameter estimators and the autoregressive representation of (\widetilde{X}_t) is invertible (cf. Section 2). This concludes the proof of Theorem 3.1. \square

PROOF of Theorem 3.2. Also due to the results of Dahlhaus (1985), we obtain that the distribution of $\sqrt{n}(M(\widetilde{I}_n, \varphi) - M(f_X, \varphi))$, where \widetilde{I}_n denotes the periodogram of n observations of the autoregressive companion process (\widetilde{X}_t) , cf. (3.2), asymptotically is normal with mean zero and variance

$$(5.15) \quad (E\varepsilon_t^4 / (E(\varepsilon_t)^2)^2 - 3) \left(\int_0^\pi \varphi(\lambda) f_X(\lambda) d\lambda \right)^2 + 2\pi \int_0^\pi \varphi^2(\lambda) f_X^2(\lambda) d\lambda.$$

Thus it suffices to show that the distribution of the bootstrap approximation of $\sqrt{n} \left(M(\widetilde{I}_n^*, \varphi) - M(\widehat{f}_{AR}, \varphi) \right)$ shares the same asymptotic distribution. Exactly along the lines of proof of Theorem 4.1 in Kreiss and Paparoditis (2003) (without the additional nonparametric correction considered therein) which makes use of Proposition 6.3.9 of Brockwell and Davis (1991) we obtain the the desired result. \square

PROOF of Theorem 3.3. Let $\widetilde{f}_n(\lambda) = \int_{-\pi}^\pi K_h(\omega - \lambda) \widetilde{I}_n(\lambda) d\lambda$ and consider $\widetilde{Y}_n = \sqrt{nh}(\widetilde{f}_n(\lambda) - f_X(\lambda))$. Since \widetilde{Y}_n converges to a Gaussian distribution with mean and variance as in (3.21) and (3.22) respectively, it suffices to show that $\sqrt{nh}(f_n^*(\lambda) - f_{AR}(\lambda))$ shares exactly the same asymptotic behavior as Y_n . This however, follows exactly along the same lines as in the proof of Theorem 5.1 in Kreiss and Paparoditis (2003), again without the additional nonparametric correction considered therein. \square

Acknowledgements. The authors would like to express their gratitude to two anonymous referees. Their careful reading of a previous version of the paper and their thorough reports led to a considerable improvement of the present paper.

REFERENCES

- Andrews, B., Davis, R. A. and Breidt, F. J. (2007): Rank-Based Estimation for All-Pass Time Series Models. *Annals of Statistics* **35**, 844–869.
 Baxter, G. (1962): An Asymptotic Result for the Finite Predictor *Math. Scand.* **10**, 137–144.

- Baxter, G. (1963): A Norm Inequality for a Finite-Section Wiener-Hopf Equation. *Illinois J. Math.* **7**, 97–103.
- Bickel, P.J. and Bühlmann, P. (1999): A new Mixing Notion and Functional Central Limit Theorems for a Sieve Bootstrap in Time Series. *Bernoulli* **5**, 413–446.
- Breidt, F.J. and Davis, R. A. (1991): Time-Reversibility, Identifiability and Independence of Innovations for Stationary Time Series. *Journal of Time Series Analysis* **13**, 377–390.
- Breidt, F.J. and Davis, R. A. and Dunsmuir, W. T. M (1995): Improved Bootstrap Prediction Intervals for Autoregressions. *Journal of Time Series Analysis* **16**, 177–200.
- Brockwell, P.J. and Davis, R.A. (1991): *Time Series: Theory and Methods (2nd Ed.)*. Springer, New York.
- Bühlmann, P. (1995): Sieve Bootstrap for Time Series. *Technical Report 431*, Department of Statistics, University of California, Berkeley.
- Bühlmann, P. (1997): Sieve Bootstrap for Time Series. *Bernoulli* **3**, 123–148.
- Bühlmann, P. (2002): Bootstraps for Time Series. *Statistical Science* **17**, 52–72.
- Choi, E. and Hall, P. (2000): Bootstrap confidence regions computed from autoregressions of arbitrary order *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**, 461–477.
- Dahlhaus, R. (1985): Asymptotic Normality of Spectral Estimates. *Journal of Multivariate Analysis* **16**, 412–431.
- Dahlhaus, R. and Janas, D. (1996): A Frequency Domain Bootstrap for Ratio Statistics in Time Series Analysis. *The Annals of Statistics* **24**, 1934–1963.
- Doukhan, P. and Louhichi, S. (1999): A New Weak Dependence Condition and Application to Moment Inequalities. *Stoch. Proc. Appl.* **84**, 313–1507.
- Franke, J. and Härdle, W. (1992): On Bootstrapping Kernel Spectral Estimates. *The Annals of Statistics* **20**, 121–145.
- Freedman, D. (1984): On bootstrapping two-stage least squares estimates in stationary linear models. *The Annals of Statistics* **12**, 827–842.
- Gröchenig, K. (2007): Weight functions in time-frequency analysis. In: *Pseudodifferential Operators: Partial Differential Equations and Time-Frequency Analysis*. Rodino, L. et al. (Eds.). *Fields Institute Comm.* **52**, 343–366.
- Kirch, C. and Politis, D.N. (2011). TFT-Bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *The Annals of Statistics*, to appear.
- Kokoszka, P. and Politis, D. N. (2011): Nonlinearity of ARCH and stochastic volatility models and Bartlett’s formula. To appear in *Probability and Mathematical Statistics*
- Kreiss, J.-P. (1988): *Asymptotical Inference for a Class of Stochastic Processes*. Habilitationsschrift, Universität Hamburg.
- Kreiss, J.-P. (1992): Bootstrap Procedures for $AR(\infty)$ Processes. In: *Bootstrapping and Related Techniques*. Jöckel, K.H., Rothe, G. and Sendler, W. (Eds.). *Lecture Notes in Economics and Mathematical Systems* **376** Springer, Heidelberg, 107–113.
- Kreiss, J.-P. und Neuhaus, G. (2006): *Einführung in die Zeitreihenanalyse*. Springer, Heidelberg.
- Kreiss, J.-P. and Paparoditis, E. (2003): Autoregressive Aided Periodogram Bootstrap for Time Series. *The Annals of Statistics* **31**, 1923–1955.
- Künsch, H.R. (1989): The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics* **17**, 1217–1241.
- Lahiri, S.N. (2003): *Resampling Methods for Dependent Data*. Springer, New York.
- Lii, K. S. and Rosenblatt, M. (1982): Deconvolution and Estimation of Transfer Function Phase and Coefficients for NonGaussian Linear Processes. *The Annals of Statistics* **10**, 1195–1208.

- Lii, K. S. and Rosenblatt, M. (1996): Maximum Likelihood Estimation for NonGaussian NonMinimum Phase ARMA Sequences. *Statistica Sinica* **6**, 1–22.
- Paparoditis, E. and Streitberg, B. (1991): Order Identification Statistics in Stationary Autoregressive Moving Average Models: Vector Autocorrelations and the Bootstrap. *Journal of Time Series Analysis* **13**, 415–435.
- Paparoditis, E. (1996): Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes. *Journal of Multivariate Analysis* **57**, 277–296.
- Paparoditis, E. and Politis, D.N. (2009): Resampling and Subsampling for Financial Time Series. In: *Handbook of Financial Time Series*. Andersen, T.G., Davis, R.A., Kreiss, J.-P., and Mikosch, T. (Eds.). Springer, Berlin-Heidelberg, 983–999.
- Politis, D.N., Romano, J.P. and Wolf, M. (1999): Weak Convergence of Dependent Empirical Measures with Application to Subsampling in Function Spaces. *Journal of Statistical Planning and Inference* **79**, 179–190.
- Politis, D.N. (2003): The Impact of Bootstrap Methods on Time Series Analysis. *Statistical Science* **18**, 219–230.
- Poskitt, D.S. (2007): Properties of the Sieve Bootstrap for Fractionally Integrated and Non-Invertible Processes. *J. Time Ser. Analysis* **29**, 224–250.
- Pourahmadi, M. (2001): *Foundations of Time Series Analysis and Prediction Theory*. Wiley, 2001.
- Priestley, M. B. (1981): *Spectral Analysis and Time Series*. Academic Press, New York.
- Romano, J. P. and Thombs, L. A. (1996): Inference for Autocorrelations Under Weak Assumptions. *Journal of the American Statistical Association* **91**, 590–600.
- Shao, X. and Wu W. B. (2007): Asymptotic Spectral Theory for nonlinear Time Series. *Annals of Statistics* **35**, 1773–1801. Wiley, 2001.
- Taniguchi, M. and Kakizawa, Y. (2000): *Asymptotic theory of statistical inference for time series*. Springer-Verlag, New York.

TECHNISCHE UNIVERSITÄT BRAUNSCHWEIG, INSTITUT FÜR MATHEMATISCHE STOCHASTIK, POCKELSTRASSE 14, D-38106 BRAUNSCHWEIG, GERMANY.

UNIVERSITY OF CYPRUS, DEPARTMENT OF MATHEMATICS AND STATISTICS, 1678 NICOSIA, CYPRUS.

UNIVERSITY OF CALIFORNIA, SAN DIEGO, DEPARTMENT OF MATHEMATICS, LA JOLLA, CA 92093-0112, USA.