

Braumann, Alexander; Kreiss, Jens-Peter; Meyer, Marco

Article — Published Version

Simultaneous inference for autocovariances based on autoregressive sieve bootstrap

Journal of Time Series Analysis

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Braumann, Alexander; Kreiss, Jens-Peter; Meyer, Marco (2021) : Simultaneous inference for autocovariances based on autoregressive sieve bootstrap, Journal of Time Series Analysis, ISSN 1467-9892, John Wiley & Sons, Ltd, Oxford, UK, Vol. 42, Iss. 5-6, pp. 534-553,
<https://doi.org/10.1111/jtsa.12604>

This Version is available at:

<https://hdl.handle.net/10419/284790>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc-nd/4.0/>

SPECIAL ISSUE

SIMULTANEOUS INFERENCE FOR AUTOCOVARIANCES BASED ON AUTOREGRESSIVE SIEVE BOOTSTRAP

ALEXANDER BRAUMANN^a JENS-PETER KREISS^{a*} AND MARCO MEYER^a

^a*Technische Universität Braunschweig, Braunschweig, Germany*

In this article, maximum deviations of sample autocovariances and autocorrelations from their theoretical counterparts over an increasing set of lags are considered. The asymptotic distribution of such statistics for physically dependent stationary time series, which is of Gumbel type, only depends on second-order properties of the underlying time series. Since the autoregressive sieve bootstrap is able to mimic the second-order structure asymptotically correctly it is an obvious problem whether the autoregressive (AR) sieve bootstrap, which has been shown to work for a number of relevant statistics in time series analysis, asymptotically works for maximum deviations of autocovariances and autocorrelations as well. This article shows that the question can be answered positively. Moreover, potential applications including spectral density estimation and an investigation of finite sample properties of the AR-sieve bootstrap proposal by simulation are given.

Received 14 September 2020; Accepted 17 May 2021

Keywords: Autocovariance; autoregressive sieve bootstrap; stationary process, spectral density; simultaneous confidence bands

MOS subject classification: 62M10; 62G09; 62G15.

1. INTRODUCTION

For stationary time series $(X_t, t \in \mathbb{Z})$, where \mathbb{Z} denotes the set of all integers, the autocovariance function $\gamma(h) = \text{Cov}(X_{t+h}, X_t)$, $h = 0, 1, \dots$, and the autocorrelation function $\rho(h) = \gamma(h)/\gamma(0)$ are important characteristics. Obvious estimates of $\gamma(h)$ and $\rho(h)$ are their sample versions

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} X_{t+h} X_t \quad \text{and} \quad \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad h = 0, 1, \dots, n-1, \quad (1)$$

where we assume for the sake of simplicity $\text{EX}_t = 0$. Otherwise a centering with the sample mean of the observed stretch X_1, X_2, \dots, X_n of the underlying time series has to be added. Plots of sample autocorrelations, so-called correlograms, play an important role in time series analysis. Numerous statistical procedures have been developed for testing for serial correlation. A prominent example is the Box–Pierce portmanteau test, which uses $Q_K = n \sum_{h=1}^K \hat{\rho}(h)^2$ as a test statistic. It is not surprising that a lot of asymptotical theory has been developed around sample autocovariances and autocorrelations.

When investigating and proving properties of statistics in time series analysis important ingredients are the assumptions made for the underlying model. The assumption of a linear time series, that is

$$X_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad (2)$$

*Correspondence to: Jens-Peter Kreiss, Inst. f. Math. Stochastik, Technische Universität Braunschweig, Universitätsplatz 2, D-38106 Braunschweig, Germany. E-mail: j.kreiss@tu-bs.de

Dedicated to Professor Murray Rosenblatt to whom the second author owes a lot.

for absolutely summable coefficients (b_j) and i.i.d. random variables (ε_t) with mean zero and finite variance, serves for many results as a comfortable assumption, see for example the monograph Brockwell and Davis (1991). More flexible assumptions are mixing conditions (cf. Rosenblatt, 1961, 1972; see also Rosenblatt, 1985) or weak dependence assumptions in the sense of Doukhan and Louhichi (1999). For a comprehensive review of weak dependence we refer to Dedecker *et al.* (2007). Wu (2005) introduced the concept of physical dependence, which has proven to be a powerful tool for the derivation of far-reaching asymptotic theory for time series. Wu assumes that (X_t) is a (strictly) stationary time series of the form

$$X_t = g(u_t, u_{t-1}, \dots), \quad (3)$$

for i.i.d. random variables (u_t) and an arbitrary measurable function g . Such time series are denoted as *Bernoulli shifts*. An essential assumption now consists in quantifying the impact of a single innovation at some fixed time point on future values of the time series. For this let $(u'_t, t \in \mathbb{Z})$ be an i.i.d. copy of $(u_t, t \in \mathbb{Z})$ and let $X'_t := g(u_t, \dots, u_1, u'_0, u_{-1}, u_{-2}, \dots)$.

If $\|X_t\|_q := (E|X_t|^q)^{1/q} < \infty$ for some $q > 0$, the *physical dependence measure* is defined as

$$\delta_q(t) := \|X_t - X'_t\|_q. \quad (4)$$

Asymptotical results then are typically deduced under assumptions on the rate of decay of $\delta_q(t)$ as $t \rightarrow \infty$. See for example Wu (2011). For linear time series (X_t) according to (2) the representation (3) obviously holds with $(u_t) = (\varepsilon_t)$. It is not complicated to see that $\delta_q(t) = |b_t| \|\varepsilon_1 - \varepsilon'_1\|_q$ so that for linear time series assumptions on the decay of $\delta_q(t)$ one-to-one carry over to assumptions on the decay of the coefficients b_t .

Under all mentioned dependence assumptions on the underlying time series for sample autocovariances and sample autocorrelations as well asymptotic normality can be shown. Of particular interest are uniform confidence bands for sample autocovariances and sample autocorrelations. For this purpose asymptotic theory for

$$\sqrt{n} \max_{1 \leq h \leq d_n} |\hat{\gamma}(h) - \gamma(h)| \quad (5)$$

is necessary. First results concerning the asymptotic behavior of (5) have been deduced by Hannan and co-authors (see Hannan, 1974; An *et al.*, 1982 and also Hannan and Deistler, 1988). These papers allow d_n to grow logarithmically. Under the same assumption of logarithmically growing d_n Jirak (2011) proved under the assumption of a linear time series for a properly centered and rescaled version of (5) weak convergence towards a Gumbel distribution with cumulative distribution function $\exp(-\exp(-x))$. Xiao and Wu (2014), under much more general physical dependence assumptions, extended this result considerably in allowing the lags to grow at a rate $d_n \sim n^\eta$ for $\eta \in (0, 1)$ fulfilling some further restrictions depending on the physical dependence assumption. The result of Xiao and Wu (2014) reads as follows:

$$a_n^{-1} \left(\sigma_0^{-1/2} \max_{1 \leq h \leq d_n} \sqrt{n} |\hat{\gamma}(h) - \gamma(h)| - b_n \right) \xrightarrow{D} \text{Gumbel}, \quad (6)$$

where $\sigma_0 = \sum_{h \in \mathbb{Z}} \gamma^2(h)$ and a_n and b_n are defined in the beginning of Section 2, where also precise assumptions for such a result are stated.

The weak convergence results of Xiao and Wu (2014) and Jirak (2011) concerning the maximum deviation of sample autocovariances are extreme value results. It is often the case that satisfactory precision in approximating the finite sample distribution of extremes by asymptotic distributions, which is Gumbel in our case, typically

is reached for large sample sizes only. That is why Xiao and Wu (2014) suggested to apply a block-of-blocks bootstrap as a successful competitor to the asymptotic Gumbel distribution.

In this article, we propose another well-known bootstrap procedure for time series, the so-called autoregressive (AR) sieve bootstrap (cf. Kreiss, 1992; Bühlmann, 1997) and show its asymptotic validity for maximum deviation of sample autocovariances and autocorrelations. The AR-sieve bootstrap, which generates a pseudo time series based on an AR fit to the observations, is a well-established and investigated resampling procedure. Kreiss (1992) proved asymptotic validity for autocovariances and autocorrelations at fixed lags among other things. Paparoditis and Streitberg (1992) investigated asymptotic validity of the AR-sieve bootstrap for inference of high order autocorrelations. Bühlmann (1997) refines the AR-sieve bootstrap according to the rate of decay of AR coefficients and generalizes its asymptotic validity to so-called general mean statistics. The paper Kreiss *et al.* (2011) makes the abilities of the AR-sieve bootstrap very clear. To do so the authors defined a companion process, which indeed is of AR type with i.i.d. innovations, and showed for the same class of statistics as has been considered in Bühlmann (1997) that the AR-sieve bootstrap mimics the behavior of this companion process. Since the companion process always has the same second-order structure (e.g., the same autocovariance structure) as the underlying time series, which by far has neither to be an AR time series with i.i.d. innovations nor a linear time series, the paper Kreiss *et al.* (2011) within the class of generalized means culminates in a check criterion for the asymptotic validity of the AR-sieve bootstrap. Namely, the AR-sieve bootstrap for generalized means asymptotically works if and only if the asymptotic distribution of the statistic of interest solely depends on second-order properties of the underlying time series. This fact together with the asymptotic distribution result (6) for maximum deviations of sample autocovariances begs the question, whether or not AR-sieve bootstrap works for this statistic as well. This question is answered positively in this article. In addition we will indicate that the proposed bootstrap procedure is also suitable for simultaneous confidence intervals for the spectral density based on well-known lag window spectral density estimators.

It is worth mentioning that for (6) to hold it is important that the maximum is taken over an increasing number d_n of lags. If the maximum is only taken over a fixed number of lags, K say, then the asymptotic distribution would be completely different. Since asymptotic normality of a vector of fixed length K of sample autocovariances (and autocorrelations as well) can be shown we obtain from continuous mapping theorem that the maximum deviation of sample autocovariances for a fixed number of lags weakly converges to the maximum of a dependent Gaussian vector of fixed length K . For general stationary time series the asymptotic normal distribution of sample autocovariances depends on the fourth order cumulant structure of the underlying time series and this in general goes far beyond its second-order structure (see e.g., Rosenblatt, 1985, Chapter III, Corollary 2, for integrated periodograms for which sample autocovariances are special cases). By Kreiss *et al.* (2011) this immediately implies that AR-sieve bootstrap cannot work. However, for rather general stationary time series but still for the maximum deviation over a fixed number of lags, the hybrid periodogram bootstrap (HPB) suggested in Meyer *et al.* (2020) leads to an asymptotically valid resampling procedure. This procedure is valid even for fixed length vectors of autocovariances and autocorrelations as can be seen from Meyer and Paparoditis (2021+). Nevertheless, validity of the HPB method has been shown only for the case of fixed d_n , and an extension to the case of an increasing number of lags seems not immediate. We therefore propose the use of the AR-sieve bootstrap in this scenario. The latter has the additional advantage of being rather easy to implement.

If one restricts to the class of linear time series (cf. (2)) then it is well known that the asymptotic distribution of sample autocorrelation only depends on second-order properties of the underlying model (this fact is known as Bartlett's formula, cf. Brockwell and Davis, 1991, Theorem 7.2.1). As a consequence AR-sieve bootstrap works within the class of linear time series in approximating the maximum deviation (over a fixed number of lags) of sample autocorrelations.

The article is organized as follows. Section 2 contains all necessary model assumptions and describes the proposed AR-sieve bootstrap procedure to approximate the distribution of maximum deviations of sample autocovariances and autocorrelations. It also states and proves the consistency result for the proposed bootstrap procedure. Section 3 sheds some light on finite sample size behavior of the proposed procedure based on simulations. Some technical proofs are deferred to Section 4.

2. MODEL ASSUMPTIONS AND BOOTSTRAP PROCEDURE

We focus on stationary time series $(X_t, t \in \mathbb{Z})$ generated according to the causal model

$$X_t = g(u_t, u_{t-1}, \dots), \quad (7)$$

and impose the following assumptions.

Assumption 1. For the time series given in (7) it holds

- (a) (u_t) is i.i.d. and g is a measurable function,
 (b) $EX_t = 0, E|X_t|^q < \infty$ for some $q > 4$.

For some $\alpha \geq \alpha' > 0$,

- (c) $\sum_{t=m}^{\infty} \delta_q(t) = \mathcal{O}(m^{-\alpha})$,
 (d) $\Delta_q(m) = \sum_{t=0}^{\infty} \min \left\{ (q-1)^{1/2} \left(\sum_{i=m}^{\infty} \delta_q(i)^2 \right)^{1/2}, \delta_q(t) \right\} = \mathcal{O}(m^{-\alpha'})$.

Under Assumption 1 it is known from Xiao and Wu (2014), Theorem 1, that the properly scaled statistic

$$T_n = a_n^{-1} \left(\sigma_0^{-1/2} \max_{1 \leq h \leq d_n} \sqrt{n} |\hat{\gamma}(h) - \gamma(h)| - b_n \right) \quad (8)$$

has an asymptotic standard Gumbel distribution, that is,

$$P(T_n \leq x) \rightarrow \exp(-\exp(-x)), \quad \forall x \in \mathbb{R}. \quad (9)$$

Here $a_n^{-1} = \sqrt{2 \log(2d_n)}$, $b_n = \sqrt{2 \log(2d_n)} - \frac{\log \log(2d_n) + \log(4\pi)}{\sqrt{8 \log(2d_n)}}$, $\sigma_0 = \sum_{h \in \mathbb{Z}} \gamma^2(h)$ and the maximum is taken over an increasing range of lags with $d_n \rightarrow \infty$, $d_n = \mathcal{O}(n^\eta)$, $0 < \eta < 1/2$ and

$$\eta < \alpha \frac{q}{2}, \quad \eta \min\{2(q-2-\alpha q), (1-2\alpha')q\} < q-4. \quad (10)$$

Observe that Xiao and Wu (2014) allow in their Theorem 1 that d_n may grow almost as fast as sample size n . We further restrict the rate of d_n since we use the centering $\gamma(h)$ instead of $E\hat{\gamma}(h) = (1 - \frac{h}{n})\gamma(h)$.

Since the asymptotic distribution of T_n given in (9) is completely determined by the probabilistic behavior of $\hat{\gamma}(h)$ for large lags, cf. the proof of Theorem 1 in Xiao and Wu (2014), Lemma 4, and (4.3), one can show that there is no difference in the asymptotic behavior of T_n (up to a factor of $\gamma(0)$) if we replace sample autocovariances by sample autocorrelations (see also Xiao and Wu, 2014, Corollary 1). That is we have as a consequence of (9) for all $x \in \mathbb{R}$

$$P \left(a_n^{-1} \left(\sigma_0^{-1/2} \gamma(0) \max_{1 \leq h \leq d_n} \sqrt{n} |\hat{\rho}(h) - \rho(h)| - b_n \right) \leq x \right) \rightarrow \exp(-\exp(-x)). \quad (11)$$

We further impose the following conditions on the second-order structure of the underlying process.

Assumption 2. $(X_t, t \in \mathbb{Z})$ is purely non-deterministic with autocovariance function γ fulfilling

$$\sum_{h \in \mathbb{Z}} (1 + |h|)^r |\gamma(h)| < \infty$$

for some $r \geq 0$ to be specified later. The spectral density f is therefore continuous and is further assumed to be uniformly bounded away from zero, that is, $f(\lambda) \geq c_f > 0$ for all frequencies λ .

To approximate the finite sample distribution of (8) we propose a version of the AR-sieve bootstrap. Before we precisely state the bootstrap proposal we will briefly discuss why an AR scheme is suitable to approximate certain features of the underlying time series. For a detailed presentation including proofs for the following statements we refer to Kreiss *et al.* (2011) or to the forthcoming monograph Kreiss and Paparoditis (2021), Section 2.3.1 (especially Theorems 2.14, 2.19, 2.21, and Corollary 2.15).

The success of the suggested application of the AR-sieve bootstrap is based on the fact that the underlying process under Assumptions 1 and 2 possesses an infinite order Wold-type AR representation of the form

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + e_t, \tag{12}$$

for an uncorrelated (and in general *not* i.i.d.) white noise process $(e_t, t \in \mathbb{Z})$.

The power series $A(z) := 1 - \sum_{k=1}^{\infty} a_k z^k$ is well defined and has no zeroes within the closed unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$ and therefore is always invertible with $B(z) := 1/A(z) = 1 + \sum_{j=1}^{\infty} b_j z^j$ for all $|z| \leq 1$.

Inversion of (12) leads to the well-known (moving average [MA]) Wold representation for purely non-deterministic, stationary and zero-mean time series

$$X_t = e_t + \sum_{j=1}^{\infty} b_j e_{t-j}. \tag{13}$$

The summability condition on the autocovariances from Assumption 2 carries over to the AR and MA coefficients, namely,

$$\sum_{k=1}^{\infty} (1 + |k|)^r |a_k| < \infty, \quad \text{and} \quad \sum_{j=1}^{\infty} (1 + |j|)^r |b_j| < \infty, \tag{14}$$

cf. Theorem 1.1 in Cheng and Pourahmadi (1993) or Theorem 3.8.4 in Brillinger (1981) and also pages 139/140 in Baxter (1962). The AR coefficients a_k can be estimated from fitted finite-order AR models to the underlying data. We denote the parameter estimates of a p th order AR fit by $\hat{a}(p) = (\hat{a}_1(p), \dots, \hat{a}_p(p))^T$. Typical choices for these estimators are Yule–Walker or least squares estimates but we do not have to specify this for our bootstrap procedure. The estimate $\hat{a}(p)$ is usually compared to the best theoretical fit of an AR model to the underlying time series. The best fit of a p th order AR model in the L^2 -sense is given by the L^2 -projection of X_t onto the *finite* past span $\{X_{t-1}, \dots, X_{t-p}\}$, $p \in \mathbb{N}$, that is, by

$$(a_1(p), \dots, a_p(p)) := \operatorname{argmin}_{c_1, \dots, c_p} \mathbb{E} \left(X_t - \sum_{k=1}^p c_k X_{t-k} \right)^2. \tag{15}$$

The quantities $a_1(p), \dots, a_p(p)$ are called the finite predictor coefficients, and we write $\underline{a}(p) = (a_1(p), \dots, a_p(p))^T$. Solving (15) leads to the well-known Yule–Walker equations

$$\Gamma(p) \left(a_1(p), \dots, a_p(p) \right)^T = \left(\gamma(1), \dots, \gamma(p) \right)^T, \tag{16}$$

where $\Gamma(p) = (\gamma(j-k))_{j,k=1, \dots, p}$. We further write $\gamma(p) := (\gamma(1), \dots, \gamma(p))^T$ for the right-hand side of (16). The finite predictors in turn converge by Baxter’s inequality with $p \rightarrow \infty$ towards the AR coefficients a_k . More precisely, there exists a constant $K < \infty$ such that we have under Assumptions 1 and 2 for all $s \in \mathbb{N}_0$ with $s \leq r$ and

for all $p \in \mathbb{N}$:

$$\sum_{k=1}^p (1+k)^s |a_k(p) - a_k| \leq K \sum_{k=p+1}^{\infty} (1+k)^s |a_k|. \quad (17)$$

The right-hand side converges to zero as p tends to infinity.

For our AR-sieve bootstrap scheme we will require $p = p(n)$ with $p(n) \rightarrow \infty$ as $n \rightarrow \infty$. We impose the following assumption on the rate with which the $\hat{a}_k(p)$ approach the finite predictor coefficients $a_k(p)$.

Assumption 3. It holds $p = p(n)$ with $p(n) \rightarrow \infty$ and

$$\max_{1 \leq k \leq p(n)} |\hat{a}_k(p(n)) - a_k(p(n))| = \mathcal{O}_p \left(\sqrt{\frac{\log n}{n}} \right), \quad (18)$$

as $n \rightarrow \infty$.

If one uses the Yule–Walker estimators for $\hat{a}(p)$, then (18) holds true, as the following result shows. However, since these specific estimators are otherwise not required for our bootstrap scheme we have formulated (18) as an assumption.

Lemma 1. If $p(n) \rightarrow \infty$ with $p(n) = \mathcal{O}(n^a)$ for $a < 1/2$, and if $\hat{a}(p)$ is given by the Yule–Walker estimator $\hat{\Gamma}(p)^{-1} \hat{\gamma}(p)$, where $\hat{\Gamma}(p)$ and $\hat{\gamma}(p)$ are defined as the corresponding quantities in (16) but with $\gamma(\cdot)$ replaced by $\hat{\gamma}(\cdot)$, then assertion (18) holds true.

Now the AR-sieve bootstrap procedure can be formulated.

Bootstrap Proposal. (*AR-sieve bootstrap*)

Step 1: Select an order $p = p(n) \in \mathbb{N}$, $p \ll n$ and fit a p th order AR model to X_1, X_2, \dots, X_n . Denote by $\hat{a}(p) = (\hat{a}_k(p), k = 1, 2, \dots, p)$ parameter estimates which fulfill Assumption 3.

Step 2: Let (v_t^*) be a sequence of i.i.d. random variables with zero mean, unit variance, and finite fourth moment and denote by $\tilde{\sigma}^2(p)$ the sample variance of the residuals $\tilde{e}_t(p) = X_t - \sum_{j=1}^p \hat{a}_j(p) X_{t-j}$, $t = p+1, p+2, \dots, n$, of the AR fit (or some other consistent estimator of $\text{Var}(e_t)$ in (12)).

Then define bootstrap innovations $e_t^* := \tilde{\sigma}(p) \cdot v_t^*$, that is as i.i.d. white noise with variance $\tilde{\sigma}^2(p)$ and finite fourth order moment. Let $(X_1^*, X_2^*, \dots, X_n^*)$ be a set of pseudo observations from the time series $(X_t^*, t \in \mathbb{Z})$ where

$$X_t^* = \sum_{j=1}^p \hat{a}_j(p) X_{t-j}^* + e_t^*. \quad (19)$$

Step 3: Let $T_n^* = T_n(X_1^*, X_2^*, \dots, X_n^*)$ be the same estimator as T_n (cf. (8)) based on the pseudo-time series $X_1^*, X_2^*, \dots, X_n^*$ with $\gamma(h)$ replaced by $\gamma^*(h) = E^* X_{t+h}^* X_t^*$, $h \geq 1$, and σ_0 replaced by $\sigma_0^* = \sum_{h \in \mathbb{Z}} (\gamma(h)^*)^2$. That is

$$T_n^* = a_n^{-1} \left((\sigma_0^*)^{-1/2} \max_{1 \leq h \leq d_n} \sqrt{n} |\hat{\gamma}^*(h) - \gamma^*(h)| - b_n \right), \quad (20)$$

where $\hat{\gamma}^*(h) = \frac{1}{n} \sum_{t=1}^{n-h} X_{t+h}^* X_t^*$, $h \geq 0$.

The AR-sieve bootstrap approximation of $\mathcal{L}(T_n)$ finally is given by $\mathcal{L}^*(T_n^*)$.

Remark 1. The classical AR-sieve bootstrap as presented in Kreiss *et al.* (2011) uses in (19) innovations (e_t^*) drawn with replacement from the empirical distribution of centered residuals $\hat{e}_t(p) = \tilde{e}_t(p) - \bar{e}$, where $\bar{e} = (n - p)^{-1} \sum_{t=p+1}^n \tilde{e}_t(p)$.

In principle we could do the same here. As can be seen from the proof of Theorem 1 in this case we would need some slow but still polynomial rate of convergence for the Mallows distance $d_2(F_n, F)$. Here F_n denotes the empirical distribution of n innovations e_1, \dots, e_n from the white noise of the Wold-type AR representation (12) and F denotes the true marginal cumulative distribution function of e_1 .

The distance d_r , for $r > 0$ is defined according to

$$d_r(F, G) = \inf (E|X - Y|^r)^{1/r}, \tag{21}$$

where the infimum is over all possible two-dimensional distributions (X, Y) with $X \sim F$ and $Y \sim G$, but arbitrary dependence between X and Y , cf. Bickel and Freedman (1981), beginning of Section 8. Especially the following identity applies

$$d_r(F, G) := \left(\int_0^1 |F^{-1}(u) - G^{-1}(u)|^r du \right)^{1/r}. \tag{22}$$

Moreover, pointwise convergence of F_n towards F together with consistency of second-order sample moments (which both could be ensured under rather mild additional assumptions) is sufficient to guarantee $d_2(F_n, F) \rightarrow 0$, cf. Bickel and Freedman (1981) Lemma 8.3. However, this result does not give the needed rate of convergence.

Much stronger assumptions, like i.i.d. structure or some kind of mixing, on the white noise (e_t) in (12) would lead to desired convergence rates of $d_2(F_n, F)$. Under the restrictive assumption of i.i.d. innovations in (12) we even obtain a \sqrt{n} -rate convergence of Mallows distance, cf. Samworth and Johnson (2005), Theorem 2.4. For the i.i.d. case and a lower but still sufficient rate of convergence see Horowitz and Karandikar (1994). Dedecker and Merlevède (2017) give convergence rates of $d_2(F_n, F)$ for a class of α -mixing innovations (e_t) , while Fournier and Guillin (2015) assumed ρ -mixing for their result.

For a general discussion of Mallows distance (also known as Wasserstein distance) see Panaretos and Zemel (2019). Be that as it may, mixing –or even more i.i.d. –assumptions on the innovations in the Wold-type AR representation (12) of the underlying time series (X_t) both are rather restrictive. For example, assuming an i.i.d. structure for the innovations (e_t) in (12) would imply that the underlying time series is linear, which would mean a massive restriction of the basic assumption (7) for the underlying time series. Since we want to present results which go (far) beyond linear time series, we decided to define the innovations (e_t^*) in the AR-sieve bootstrap procedure in the described simpler way. In Section 3 we report on simulation results for both choices of distributions for the bootstrap innovations. At least for the considered models only slight differences are observed.

Our main result reads as follows:

Theorem 1. Let $(X_t, t \in \mathbb{Z})$ be a process fulfilling Assumption 1(a)–(d) and Assumption 2 with $r > 1/2$, and furthermore,

$$\left| Ee_1^2 - \frac{1}{n} \sum_{t=1}^n e_t^2 \right| = \mathcal{O}_p(n^{-\beta}), \tag{23}$$

for some $\beta > 0$. For the maximum lag d_n we assume that $d_n = \mathcal{O}(n^\eta)$, $0 < \eta < 1/2$, η fulfilling the conditions given in (10) as well as the conditions

$$\eta < 2ar, \quad 2a < \frac{1 - \eta}{2}, \quad \eta < \beta, \tag{24}$$

where a controls the rate of $p(n)$, that is, $p(n) \sim n^a$.

Then, for T_n^* defined in (20) and T_n defined in (8), it holds

$$d_K(\mathcal{L}^*(T_n^*), \mathcal{L}(T_n)) \rightarrow 0 \text{ in probability,} \quad (25)$$

where $d_K(P, Q) = \sup_x |P(-\infty, x] - Q(-\infty, x]|$ denotes the Kolmogorov distance of two probability measures P and Q .

The conditions (23) and (24) are discussed in Remark 2. A detailed proof of this theorem is given in Section 4, but we want to point out the main idea of the proof at this point. The AR-sieve bootstrap time series (X_t^*) defined in (19) asymptotically approximates the so-called companion process, which is an imaginary time series given through

$$\tilde{X}_t = \sum_{k=1}^{\infty} a_k \tilde{X}_{t-k} + \tilde{\varepsilon}_t, \quad t \in \mathbb{Z}, \quad (26)$$

where the (a_k) are the coefficients of the Wold-type AR representation of the underlying time series given in (12) and $(\tilde{\varepsilon}_t)$ denotes an i.i.d. white noise with $\tilde{\varepsilon}_t = \sqrt{\text{Var } e_t} \cdot v_t^*$, where the i.i.d. sequence (v_t^*) is defined in Step 2 of the Bootstrap Proposal. Here (e_t) is the uncorrelated white noise of the AR representation (12). Be aware that the decisive difference between the innovation sequences $(\tilde{\varepsilon}_t)$ and (e_t) is that the $\tilde{\varepsilon}_t$'s are i.i.d. while the e_t 's are in general only uncorrelated. This important difference makes the companion process (\tilde{X}_t) to be a linear time series which by far is not true for the underlying time series (X_t) . However, the autocovariance structure of the underlying time series (X_t) and of its companion (\tilde{X}_t) is identical. This last property of (\tilde{X}_t) together with its simple linear structure suffices to ensure that the maximum deviation of sample autocovariances from the companion process have the same limit distribution as the underlying statistic T_n itself (cf. (48)). Finally, to show that AR-sieve bootstrap asymptotically works it suffices to prove that the AR-sieve bootstrap time series approximates the companion process well enough to obtain that the distance between the distributions $\mathcal{L}^*(T_n^*)$ and its companion counterpart converges to zero (cf. Lemma 2).

Since in applications usually sample autocorrelations instead of autocovariances are considered the following corollary to Theorem 1 is relevant.

Corollary 1. Under the same assumption as in Theorem 1 we obtain for

$$R_n := a_n^{-1} \left(\sigma_0^{-1/2} \gamma(0) \max_{1 \leq h \leq d_n} \sqrt{n} |\hat{\rho}(h) - \rho(h)| - b_n \right) \quad (27)$$

and its AR-sieve bootstrap counterpart R_n^* , in which $\hat{\rho}(h)$ is replaced by $\hat{\rho}^*(h) = \frac{\hat{\gamma}^*(h)}{\hat{\gamma}^*(0)}$, cf. AR-Sieve Bootstrap Proposal,

$$d_K(\mathcal{L}^*(R_n^*), \mathcal{L}(R_n)) \rightarrow 0 \text{ in probability,} \quad (28)$$

that is AR-sieve bootstrap works for maximum deviations of sample autocorrelations as well.

Remark 2. A more basic assumption which guarantees (23) with $\beta = 1/2$ would be to assume that the so-called fourth order cumulants of the uncorrelated innovations (e_t) in (12) are absolutely summable. More exactly this means for

$$\text{cum}(e_r, e_s, e_t, e_u) = E(e_r e_s e_t e_u) - E(e_r e_s)E(e_t e_u) - E(e_r e_t)E(e_s e_u) - E(e_r e_u)E(e_s e_t)$$

that $\sum_{h_1, h_2, h_3} |\text{cum}(e_0, e_{h_1}, e_{h_2}, e_{h_3})| < \infty$.

If the underlying time series (X_t) would be assumed to be α -mixing with suitably summable mixing coefficients we then obtain absolute summability of the cumulants of (X_t) , which by means of the AR representation (12) carries over to absolute summability of the cumulants of (e_t) .

Note that a strong mixing property of (X_t) does not automatically transfer to (e_t) .

The conditions given in (24) relate two basic properties of the process (X_t) , namely r which represents the ‘smoothness’ of the spectral density of (X_t) and the rate β from (23) to the rate of the maximum lag d_n . For example, given that $\beta = 1/2$, for $r = 1$ the maximum lag d_n may grow with a rate close to $n^{1/3}$, while for $r = 2$ it may grow with a rate close to $n^{1/2}$. Moreover the inequalities in (24) give sufficient lower and upper bounds for the rate of the AR order $p(n)$.

Remark 3. It is worth mentioning, that if we restrict for the underlying data to linear time series (2) with i.i.d. innovations, then it follows from Kreiss *et al.* (2011), Example 3.3, that the proposed AR-sieve bootstrap works for fixed-length vectors of sample autocorrelations. This implies that in the case of linear time series the proposed AR-sieve bootstrap procedure also works if d_n , that is the number of lags considered for the maximum deviation, does not grow but stays fixed. So for linear time series the AR-sieve bootstrap covers a broader situation as is described in Corollary 1.

Application 1. One application of Corollary 1 are simultaneous confidence bands for the autocorrelation function. If we denote by $q_{n,1-\alpha}^*$ the $(1 - \alpha)$ -quantile of the bootstrap distribution $\mathcal{L}^*(R_n^*)$, which easily can be obtained from simulations, then

$$\hat{\rho}(h) \pm \frac{a_n q_{n,1-\alpha}^* + b_n}{\sqrt{n\hat{\gamma}(0)\tilde{\sigma}_0^{-1/2}}}, \quad h = 1, 2, \dots, d_n, \tag{29}$$

is an asymptotically $1 - \alpha$ confidence band for the autocorrelation function $\rho(\cdot)$. Here $\tilde{\sigma}_0$ is an arbitrary consistent estimator for σ_0 .

Applications of the AR-sieve bootstrap to testing problems are considered in Section 3.

Application 2. Another application of the bootstrap proposal are asymptotically $1 - \alpha$ simultaneous confidence intervals for the spectral density f of the underlying time series. If one is interested in statistical inference for the entire sequence of autocovariances then one may switch to the frequency domain and may prefer tests or confidence bands for the spectral density. To elaborate, consider the following well-known class of lag window spectral density estimators

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{h=-B_n}^{B_n} \hat{\gamma}(h)w\left(\frac{h}{B_n}\right) e^{-ih\lambda}, \tag{30}$$

where $w : [-1, 1] \rightarrow [0, \infty)$, $w(0) = 1$, denotes a continuous and symmetric weight function and B_n is the truncation point with $B_n \rightarrow \infty$ as $n \rightarrow \infty$. To avoid additional bias considerations assume that $n/B_n^5 \rightarrow 0$ as $n \rightarrow \infty$. From Theorems 3–5 in Liu and Wu (2010) we obtain under physical dependence assumptions that

$$S_n := \alpha_n^{-1} \left[\sqrt{\frac{n}{B_n}} \|w\|_2^{-1} \max_{j=0,1,\dots,B_n} \frac{|\hat{f}(\lambda_{j,n}) - f(\lambda_{j,n})|}{f(\lambda_{j,n})} - \beta_n \right] \tag{31}$$

converges weakly to the standard Gumbel distribution. Here $\alpha_n^{-1} = \sqrt{2 \log(2B_n)}$, $\beta_n = \sqrt{2 \log(2B_n)} - \frac{\log \log(2B_n) + \log(4\pi)}{\sqrt{8 \log(2B_n)}}$, $\lambda_{j,n} = \frac{\pi j}{B_n}$, $j = 0, 1, 2, \dots, B_n$, and $\|w\|_2^2 = \int_{-1}^1 w^2(u) du$.

For the special case of linear time series of the form (2) a similar result has been obtained in Woodrooffe and Van Ness (1967) (cf. Theorem 2.3).

It should be mentioned that in Liu and Wu (2010) the weak convergence of (31) is formulated for the square of the maximum deviation of the lag window estimator and therefore appears slightly different compared to Woodrooffe and Van Ness (1967). But it can be directly shown that both statements are equivalent.

However, our method of proof for Theorem 1 yields that for $B_n \sim n^\eta$, $\eta > 1/5$, $\beta = 1/2$, and properly chosen further rates a and r the following result is valid

$$d_K(S_n^*, S_n) \rightarrow 0 \text{ in probability,} \quad (32)$$

where the bootstrap counterpart S_n^* of S_n is defined as

$$S_n^* := \alpha_n^{-1} \left[\sqrt{\frac{n}{B_n}} \|w\|_2^{-1} \max_{j=0,1,\dots,B_n} \frac{|\hat{f}^*(\lambda_{j,n}) - f^*(\lambda_{j,n})|}{f^*(\lambda_{j,n})} - \beta_n \right], \quad (33)$$

with $f^*(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma^*(h) e^{-ih\lambda}$ and $\hat{f}^*(\lambda)$ is defined as in (30) with $\hat{\gamma}^*(h)$ replacing $\hat{\gamma}(h)$. We sketch the argument leading to (32) in Section 4 in Remark 6.

If we denote the $(1 - \alpha)$ -quantile of the bootstrap distribution $\mathcal{L}^*(S_n^*)$ by $Q_{n,1-\alpha}^*$ then

$$\left[\hat{f}(\lambda_{j,n}) \frac{1}{1+c^*}, \hat{f}(\lambda_{j,n}) \frac{1}{1-c^*} \right], \quad j = 0, 1, \dots, B_n, \quad (34)$$

with $c^* = \sqrt{\frac{B_n}{n}} \|w\|_2 (\alpha_n Q_{n,1-\alpha}^* + \beta_n)$, is an asymptotically $1 - \alpha$ simultaneous confidence interval for $f(\lambda_{j,n})$, $j = 0, 1, \dots, B_n$.

3. SIMULATIONS

We give simulation evidence that the AR-sieve bootstrap works well when testing the hypothesis $H_0 : \rho(h) = \rho_h^0$, $h = 1, \dots, d_n$. Here d_n grows with the number of observations n and ρ_h^0 are given autocorrelations. We know from Corollary 1 that for the sample autocorrelations not centered with the sample mean the AR-sieve bootstrap is asymptotically valid. In this section we consider a scaled version of the test statistic

$$M_n = \max_{1 \leq h \leq d_n} \sqrt{n} |\check{\rho}(h) - \rho_h^0|, \quad (35)$$

where $\check{\rho}(h) = \check{\gamma}(h)/\check{\gamma}(0)$, $\check{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X})$, $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$.

A bootstrap test for H_0 is based on the sample X_1, \dots, X_n and uses the test statistic $M_n^* = \max_{1 \leq h \leq d_n} \sqrt{n} |\check{\rho}^*(h) - \rho_h^*|$, $\rho_h^* = \sum_{j=0}^{\infty} \hat{b}_{n,j+h} \hat{b}_{n,j} / \sum_{j=0}^{\infty} \hat{b}_{n,j}^2$.

In the reported simulations we focus on the difference between the true and nominal rejection probability and show that the bootstrap test is overall less conservative than the test using asymptotic theory.

Consider the following linear and nonlinear time series:

$$X_t = u_t, \quad (36a)$$

$$X_t = u_t + b u_{t-1}, \quad (36b)$$

$$X_t = a X_{t-1} + u_t, \quad (36c)$$

$$X_t = (a + b u_t) X_{t-1} + u_t, \quad (36d)$$

$$X_t = \sqrt{a + b X_{t-1}^2} \cdot u_t, \quad (36e)$$

$$X_t = a \sin(X_{t-1}) + u_t \quad (36f)$$

$u_t \sim \mathcal{N}(0, 1)$ and $\mathbb{E}u_t u_s = 0$ for $s \neq t$. Note that models (36a), (36b), and (36c) are linear time series in the strict sense, that is, with i.i.d. innovations in its Wold representations since the u_t are independent. The parameters for the models above are given by $b = 0.5$ for (36b), $a = 0.5$ for (36c), $a = b = 0.25, 0.4$ for (36d), $a = b = 0.25$ for (36e), $a = 0.6$ for (36f).

For each time series (X_t) we generate 4000 times data $X_1, \dots, X_n, n = 600, 1800$, and perform the asymptotic test by comparing M_n with $\left(\sum_{k=-t_n}^{t_n} \check{\rho}(k)^2\right)^{1/2} (a_n q_{1-\alpha} + b_n)$, where $q_{1-\alpha}$ is the $1 - \alpha$ quantile of the Gumbel distribution. For the bootstrap tests we use the AR-sieve bootstrap. Option A is implemented with bootstrap innovations $e_t^* = \bar{\sigma}(p) \cdot v_t^*, v_t^* \sim \mathcal{N}(0, 1)$, while for option B we obtained bootstrap innovations from drawing with replacement from the empirical distribution of centered residuals of the AR fit as is described in Remark 1. Note that the distribution of v_t^* can be freely chosen under the conditions of the bootstrap proposal for option A. In fact the simulation results reported below do not change if we use for example a simple two-point distribution for v_t^* , that is, $P(v_t^* = -1) = P(v_t^* = 1) = 1/2$.

We reject H_0 if the bootstrap p -value $\#(M_n^* > M_n)/B, B = 1000$ is smaller than a given $\alpha \in (0, 1)$.

The following tables show the empirical rejection rates for the different models. We set $d_n = \lfloor \log n \rfloor, \lfloor n^{1/3} \rfloor, 25, \lfloor n^{1/2} \rfloor, t_n = \min(d_n, n^{1/3})$. The simulation setting is close to the setting in Xiao and Wu (2014) to compare our results with their simulation results.

It can be seen from Tables I and II that AR-sieve bootstrap works rather well (even for a sample size of $n = 600$) for the time series models (36a)–(36c), which all are linear time series and therefore advantageous for the AR-sieve bootstrap. For the nonlinear time series models (36d)–(36f) the behavior of the proposed AR-sieve bootstrap is mixed. While for model (36f) and the ARCH model (36e) the behavior of the AR-sieve bootstrap is quite satisfying and reasonable respectively, the finite sample behavior for the bilinear time series (36d) depends on the selection of parameters. For $a = b = 0.25$ (BILIN) bootstrap outperforms the asymptotic situation, whereas the choice $a = b = 0.4$ (BILIN2) leads to much worse results also compared to the simulation results presented in Xiao and Wu (2014) for the block of blocks bootstrap.

It is worth mentioning that for BILIN2, $X_t = aX_{t-1} + (1 + bX_{t-1})u_t$, a simple adaptation of the AR-sieve bootstrap improves finite sample results substantially. In particular, if we choose e_t^* to capture conditional heteroscedasticity, the empirical rejection rates for BILIN2 change to 2.0($\alpha = 1\%$), 8.0($\alpha = 5\%$), 14.4($\alpha = 10\%$) if $d_n = 6$ and to 2.0($\alpha = 1\%$), 6.6($\alpha = 5\%$), 12.4($\alpha = 10\%$) if $d_n = 24$, for $n = 600$. Similar improvements can be seen for $d_n = 8, 25$. This adapted AR-sieve bootstrap method computes $e_t^* = \hat{s}(X_{t-1}^*) \cdot v_t^*, \hat{s}^2(X_{t-1}^*) = \sum_{s=p+1}^n K_b((X_{t-1}^* - X_{s-1}^*)/b) \hat{e}_s^2$, with K_b being some kernel function, to capture conditional heteroscedasticity. We have chosen for K_b the biweight kernel with $b = 0.75$.

One of the reasons for the good finite sample performance of the proposed bootstrap for linear time series models may be that the AR-sieve bootstrap also asymptotically works if d_n stays fixed (see Remark 3). Another reason maybe that the rate of convergence is of polynomial order (see Remark 4). This suggests that it may be worth the effort to develop more sophisticated bootstrap methods that are able to correctly reproduce the distribution of sample autocorrelations both for fixed and increasing number of lags h also for nonlinear time series. Such bootstrap methods could be the block of blocks bootstrap applied in Xiao and Wu (2014) or the HPB suggested in Meyer *et al.* (2020). Both have been shown to be valid for a fixed d_n .

4. PROOFS

Proof of Lemma 1. First note that under the stated assumptions it holds

$$\max_{0 \leq h \leq p(n)} \sqrt{n} |\hat{\gamma}(h) - E\hat{\gamma}(h)| = \mathcal{O}_p(\sqrt{\log n}),$$

as can be seen from Xiao and Wu (2014), Theorem 1. Under $p(n) = \mathcal{O}(n^a)$ for $a \leq 1/2$ one can obviously replace the centering $E\hat{\gamma}(h) = (1 - \frac{h}{n})\gamma(h)$ by $\gamma(h)$, which leads to an additional $\mathcal{O}(1)$ term that is negligible compared

Table I. Empirical rejection rates for $n = 600$ and nominal level α , both reported in percentages. For each model, the first line gives rejection rates using Gumbel quantiles, the even lines give results using the AR-sieve bootstrap option A, the odd lines (except the first one) gives results using the AR-sieve bootstrap option B

$n = 600$		$d_n = 6$			$d_n = 8$			$d_n = 24$			$d_n = 25$		
α		1	5	10	1	5	10	1	5	10	1	5	10
	WN	0.2	2.6	6.3	0.1	2.7	6.5	0.2	2.5	6.7	0.2	2.5	6.6
A	$p = 2$	0.8	4.7	9.4	0.8	4.7	9.3	1.1	5.2	10.1	1.1	5.2	9.8
B	2	1.1	4.8	9.6	1.1	4.9	9.7	0.8	4.8	9.7	0.8	4.7	10.0
A	$p = 6$	0.7	4.5	9.3	0.6	4.2	9.0	1.0	4.9	9.4	1.1	5.0	9.2
B	6	1.0	4.6	9.4	1.0	4.6	9.2	0.5	4.2	9.3	0.6	4.5	9.3
	MA(0.5)	0.2	1.6	4.7	0.2	1.7	4.6	0.2	2.4	5.9	0.2	2.4	5.8
A	$p = 2$	1.0	5.3	11.1	1.1	5.4	11.2	1.2	4.6	9.8	1.2	4.5	9.6
B	2	1.0	5.1	10.8	1.0	4.8	10.2	1.0	4.9	9.5	1.1	4.9	9.5
A	$p = 6$	1.0	5.0	10.5	0.9	5.1	10.5	1.2	4.3	9.2	1.0	4.3	9.4
B	6	0.9	4.7	10.3	1.0	4.5	9.4	1.0	4.7	9.3	1.0	4.5	9.5
	AR(0.5)	0.1	1.7	3.7	0.1	1.6	4.1	0.2	1.7	4.1	0.2	1.7	4.2
A	$p = 2$	1.1	4.9	10.2	1.5	5.3	10.9	1.3	5.6	11.0	1.5	5.9	10.9
B	2	1.3	5.4	10.2	1.2	5.2	10.3	0.9	4.9	9.7	1.0	4.8	9.9
A	$p = 6$	1.2	4.9	10.1	1.2	5.5	10.5	1.2	5.5	10.3	1.1	5.5	10.9
B	6	1.5	5.2	10.4	1.4	5.3	10.3	1.0	4.7	9.3	0.9	4.7	9.6
	BILIN	0.3	2.9	7.2	0.3	2.8	6.9	0.2	2.6	6.4	0.2	2.8	6.5
A	$p = 2$	1.8	7.2	12.9	1.5	6.8	12.6	0.9	5.0	10.7	0.9	5.1	10.8
B	2	1.8	6.2	11.8	1.8	5.9	11.5	1.2	5.2	10.3	1.4	5.4	10.4
A	$p = 6$	1.4	6.9	12.7	1.2	6.5	12.2	0.8	4.9	9.9	0.8	4.7	9.7
B	6	1.7	6.2	11.9	1.7	5.8	11.3	1.1	4.9	9.9	1.1	5.0	10.0
	BILIN2	1.6	7.8	14.1	1.5	6.9	12.6	1.2	5.0	9.6	1.3	4.9	9.5
A	$p = 2$	5.7	14.3	23.0	5.0	12.9	20.4	3.2	9.2	15.7	3.2	9.1	15.7
B	2	5.6	14.6	22.9	4.7	13.0	21.3	2.9	9.4	15.8	3.0	9.4	15.8
A	$p = 6$	5.3	14.3	23.2	4.7	12.8	20.7	3.0	8.9	15.0	2.9	8.8	15.0
B	6	5.8	14.6	22.8	4.5	13.1	20.9	2.8	9.1	15.2	2.8	8.9	15.3
	ARCH	0.7	4.7	10.8	0.7	4.5	10.2	0.6	4.3	8.1	0.6	4.3	8.2
A	$p = 2$	2.2	7.9	14.1	2.1	7.3	13.8	1.6	5.9	11.5	1.5	6.0	11.6
B	2	2.5	8.6	14.8	2.4	7.8	14.3	2.1	6.5	11.9	2.2	6.3	11.9
A	$p = 6$	2.1	7.4	14.2	2.1	7.3	13.3	1.3	5.7	10.9	1.3	5.6	11.1
B	6	2.2	8.2	14.6	2.3	7.8	13.7	2.2	5.9	11.2	2.0	6.0	11.1
	SIN	0.2	1.7	4.9	0.2	1.7	5.0	0.1	2.3	5.3	0.1	2.3	5.2
A	$p = 2$	0.9	4.6	10.5	0.8	4.6	9.8	1.2	4.9	8.6	1.3	4.9	8.8
B	2	1.1	5.1	9.9	0.8	5.3	9.9	1.2	6.0	10.7	1.2	6.0	11.1
A	$p = 6$	0.9	4.2	10.4	1.1	4.2	9.7	1.0	4.4	8.3	0.9	4.6	8.1
B	6	1.1	4.7	9.6	1.0	5.0	9.6	0.9	5.5	9.9	0.9	5.5	10.4

to $\mathcal{O}_p(\sqrt{\log n})$. This implies

$$\max_{0 \leq h \leq p(n)} \sqrt{n} |\hat{\gamma}(h) - \gamma(h)| = \mathcal{O}_p(\sqrt{\log n}). \tag{37}$$

In the following we denote for any matrix A with entries $a_{r,s}$ by $\|A\|_1 = \max_s \sum_r |a_{r,s}|$ the column sum norm. Note that this norm is submultiplicative, and that for column vectors it is simply the sum norm. Abbreviate $p(n)$ by p . We then have

$$\begin{aligned} \sum_{k=1}^p |\hat{a}_k(p) - a_k(p)| &= \|\hat{\underline{a}}(p) - \underline{a}(p)\|_1 \\ &\leq \|\hat{\Gamma}(p)^{-1}(\hat{\underline{\gamma}}(p) - \underline{\gamma}(p))\|_1 + \|(\hat{\Gamma}(p)^{-1} - \Gamma(p)^{-1})\underline{\gamma}(p)\|_1. \end{aligned} \tag{38}$$

Table II. Empirical rejection rates for $n = 1800$ and nominal level α , both reported in percentages. For each model, the first line gives rejection rates using Gumbel quantiles, the even lines give results using the AR-sieve bootstrap option A, the odd lines (except the first one) gives results using the AR-sieve bootstrap option B

$n = 1,800$		$d_n = 7$			$d_n = 12$			$d_n = 25$			$d_n = 42$		
α		1	5	10	1	5	10	1	5	10	1	5	10
	WN	0.2	2.5	7.5	0.3	2.6	7.2	0.4	3.0	7.1	0.4	2.9	6.9
A	$p = 2$	1.1	5.4	10.8	0.8	4.9	10.2	1.0	5.2	10.5	0.8	5.2	10.4
B	2	0.9	4.7	10.2	1.0	4.5	10.1	1.0	4.8	9.4	0.8	4.5	8.9
A	$p = 7$	1.1	5.2	10.6	0.9	4.9	10.0	0.9	5.0	10.3	0.8	5.0	9.6
B	7	0.9	4.6	10.2	1.0	4.1	9.9	1.0	4.7	9.0	0.8	4.4	8.6
	MA(0.5)	0.1	2.1	5.6	0.1	2.2	5.8	0.2	2.3	6.4	0.2	2.2	6.5
A	$p = 2$	1.1	5.5	11.1	1.0	4.8	10.1	1.0	5.3	10.5	0.9	4.7	10.2
B	2	1.0	5.5	10.6	0.8	4.9	10.0	0.9	4.5	9.3	1.0	4.2	9.4
A	$p = 7$	1.0	5.1	10.4	1.0	4.9	9.4	0.9	5.1	10.4	0.9	4.9	10.1
B	7	0.8	4.9	10.2	0.8	4.7	9.5	0.8	4.5	9.1	1.0	4.3	9.3
	AR(0.5)	0.3	1.9	4.3	0.3	2.2	4.9	0.5	2.9	5.7	0.4	3.0	6.1
A	$p = 2$	1.1	5.3	10.1	1.0	5.2	10.6	1.1	5.5	10.8	1.2	5.0	10.5
B	2	1.5	5.4	10.3	1.4	5.4	10.5	1.5	5.6	11.1	1.3	5.9	10.8
A	$p = 7$	1.1	5.5	10.2	1.0	5.3	10.5	1.1	5.5	10.7	1.4	4.9	10.2
B	7	1.4	5.5	10.0	1.4	5.4	10.2	1.5	5.4	10.9	1.3	5.8	10.6
	BILIN	0.2	3.2	8.0	0.2	3.0	7.8	0.2	3.0	7.3	0.4	3.1	7.2
A	$p = 2$	1.4	5.9	11.0	1.1	5.3	10.8	0.8	4.9	9.6	0.8	4.9	9.5
B	2	1.3	6.4	12.1	1.1	5.5	11.3	1.0	5.1	10.5	1.0	5.1	10.0
A	$p = 7$	1.4	5.8	11.2	1.2	5.1	10.7	0.9	4.6	9.7	0.8	4.8	9.0
B	7	1.4	6.6	12.2	1.0	5.5	11.2	1.1	4.9	10.1	0.9	5.1	9.9
	BILIN2	2.0	8.1	15.0	1.8	7.0	13.2	1.3	6.2	11.1	1.4	5.4	10.7
A	$p = 2$	6.2	14.7	23.6	4.2	12.0	19.7	3.5	10.6	16.6	3.1	8.6	14.8
B	2	5.7	14.7	23.0	4.3	12.2	19.4	3.3	9.8	16.7	2.7	8.8	15.2
A	$p = 7$	5.8	15.0	23.4	4.0	11.9	19.5	3.4	10.4	16.5	2.9	8.4	14.9
B	7	5.6	14.4	22.9	4.4	12.2	19.2	2.9	9.4	16.3	2.5	8.6	14.9
	ARCH	0.8	5.2	10.4	0.8	4.5	9.8	0.9	4.5	9.4	0.7	4.4	8.8
A	$p = 2$	2.0	7.9	13.8	1.8	6.8	13.0	1.4	5.7	11.3	1.3	5.3	10.4
B	2	2.3	7.4	13.2	2.0	7.1	12.2	2.0	6.3	11.5	1.8	6.0	11.1
A	$p = 7$	2.2	7.9	13.4	1.8	6.9	12.8	1.2	5.7	10.8	1.3	5.4	10.2
B	7	2.3	7.3	13.1	1.9	6.9	12.5	1.8	6.0	11.5	1.7	6.1	10.8
	SIN	0.1	2.1	5.4	0.0	2.7	6.1	0.2	2.8	6.1	0.1	2.7	5.3
A	$p = 2$	1.2	4.6	9.7	1.2	4.9	10.0	1.1	4.9	10.1	1.6	4.5	9.4
B	2	0.8	4.5	11.0	0.9	5.4	10.4	0.9	4.9	9.7	0.5	4.3	8.5
A	$p = 7$	1.1	4.7	9.5	1.4	4.7	9.3	1.0	4.5	9.6	1.2	4.3	9.7
B	7	1.0	4.3	10.7	0.5	5.4	10.2	0.5	4.4	9.3	0.5	4.0	8.1

For the first summand on the right-hand side it follows from (37)

$$\begin{aligned} \|\widehat{\Gamma}(p)^{-1}(\widehat{\gamma}(p) - \underline{\gamma}(p))\|_1 &\leq \|\widehat{\Gamma}(p)^{-1}\|_1 \cdot \sum_{h=1}^p |\widehat{\gamma}(h) - \gamma(h)| \\ &\leq \|\widehat{\Gamma}(p)^{-1}\|_1 \cdot \mathcal{O}_p \left(p(n) \sqrt{\frac{\log n}{n}} \right). \end{aligned}$$

For the $\|\widehat{\Gamma}(p)^{-1}\|_1$ factor we consider as a preliminary

$$\begin{aligned} \|\widehat{\Gamma}(p)^{-1} - \Gamma(p)^{-1}\|_1 &= \|(\widehat{\Gamma}(p)^{-1} - \Gamma(p)^{-1} + \Gamma(p)^{-1})(\widehat{\Gamma}(p) - \Gamma(p))\Gamma(p)^{-1}\|_1 \\ &\leq \|\widehat{\Gamma}(p)^{-1} - \Gamma(p)^{-1}\|_1 \|\widehat{\Gamma}(p) - \Gamma(p)\|_1 \|\Gamma(p)^{-1}\|_1 \\ &\quad + \|\Gamma(p)^{-1}\|_1 \|\widehat{\Gamma}(p) - \Gamma(p)\|_1 \|\Gamma(p)^{-1}\|_1, \end{aligned}$$

which implies

$$\|\widehat{\Gamma}(p)^{-1} - \Gamma(p)^{-1}\|_1 \cdot (1 - \|\widehat{\Gamma}(p) - \Gamma(p)\|_1 \|\Gamma(p)^{-1}\|_1) \leq \|\Gamma(p)^{-1}\|_1^2 \|\widehat{\Gamma}(p) - \Gamma(p)\|_1. \quad (39)$$

From Theorem 6.6.11 in Hannan and Deistler (1988) (or assertion (xxii) in Hannan and Kavalieris, 1984) we have $\|\Gamma(p)^{-1}\|_1 = \mathcal{O}(1)$. Moreover, it holds with (37)

$$\|\widehat{\Gamma}(p) - \Gamma(p)\|_1 \leq 2 \sum_{k=0}^p |\widehat{\gamma}(h) - \gamma(h)| = \mathcal{O}_p \left(p(n) \sqrt{\frac{\log n}{n}} \right) = o_p(1)$$

under our assumption of $p(n) = \mathcal{O}(n^a)$ for $a < 1/2$. Therefore the $(1 - \dots)$ factor on the left-hand side of (39) is (in probability and for n large enough) positive and we get the bound

$$\|\widehat{\Gamma}(p)^{-1} - \Gamma(p)^{-1}\|_1 \leq \frac{\|\Gamma(p)^{-1}\|_1^2 \|\widehat{\Gamma}(p) - \Gamma(p)\|_1}{1 - \|\widehat{\Gamma}(p) - \Gamma(p)\|_1 \|\Gamma(p)^{-1}\|_1} = \mathcal{O}_p \left(p(n) \sqrt{\frac{\log n}{n}} \right). \quad (40)$$

This gives $\|\widehat{\Gamma}(p)^{-1}\|_1 \leq \|\Gamma(p)^{-1}\|_1 + \|\widehat{\Gamma}(p)^{-1} - \Gamma(p)^{-1}\|_1 = \mathcal{O}(1) + o_p(1)$ and the first summand on the right-hand side of (38) is of the rate stated in (18). The second summand on the right-hand side of (38) is of the same rate due to (40) and $\|\underline{\gamma}(p)\|_1 \leq \sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$. \square

We next present a result which transfers the convergence of the AR parameters from Baxter's inequality to the MA parameters ($b_k : k \in \mathbb{N}$), cf. (13). To elaborate, we define the z -transform of the finite predictor coefficients ($a_k(p) : k = 1, \dots, p$) by

$$A_p(z) = 1 - \sum_{k=1}^p a_k(p) z^k \quad \forall z \in \mathbb{C}. \quad (41)$$

Under our assumptions this function is uniformly bounded away from zero on the unit disk plus a small ring around the disk. To be precise, there exists $\delta > 0$ such that $|A_p(z)| \geq \delta$ uniformly for all $|z| \leq 1 + 1/p$ and all p large enough, cf. Kreiss *et al.* (2011), Lemma 2.3. For those p , $B_p(z) := 1/A_p(z)$ again can be written as a power series of the form

$$B_p(z) = 1/A_p(z) = 1 + \sum_{k=1}^{\infty} b_k(p) z^k \quad \forall |z| \leq 1 + 1/p, \quad (42)$$

for suitable coefficients ($b_j(p) : j \in \mathbb{N}$).

Then a constant $C < \infty$ exists such that for all p large enough and for all $s \in \mathbb{N}_0$ with $s \leq r$:

$$\sum_{j=1}^{\infty} (1+j)^s |b_j(p) - b_j| \leq C \cdot \sum_{k=p+1}^{\infty} (1+k)^s |a_k|, \quad (43)$$

cf. Kreiss *et al.* (2011), Lemma 2.4. The right-hand side converges to zero as p tends to infinity.

In our bootstrap procedure the estimated AR coefficients ($\widehat{a}_k(p)$, $k = 1, \dots, p$) are used to construct bootstrap random variables. Let

$$\widehat{A}_p(z) := 1 - \sum_{k=1}^p \widehat{a}_k(p) z^k \quad \forall z \in \mathbb{C}. \quad (44)$$

Since $\widehat{A}_p(z) \neq 0$ for all $|z| \leq 1$, cf. Brockwell and Davis (1991), p. 419, its reciprocal can be expanded as a power series

$$\widehat{B}_p(z) := \frac{1}{\widehat{A}_p(z)} = 1 + \sum_{k=1}^{\infty} \widehat{b}_k(p) z^k \quad \forall |z| \leq 1 \tag{45}$$

for suitable coefficients $(\widehat{b}_k(p), k \in \mathbb{N})$. The bootstrap time series resulting from our proposal possess an infinite-order MA representation with exactly these coefficients $(\widehat{b}_k(p))$. As for the difference of $\widehat{b}_k(p)$ and $b_k(p)$, one can follow along the lines of the proof of Lemma 2.5 in Kreiss *et al.* (2011) –merely replacing the bound for the expression $\sum_{k=1}^{p(n)} |\widehat{a}_k(p(n)) - a_k(p(n))|$ used there by the bound from (18) –to obtain uniformly for all $k \in \mathbb{N}$

$$|\widehat{b}_k(p(n)) - b_k(p(n))| \leq \left(1 + \frac{1}{p(n)}\right)^{-k} \cdot \mathcal{O}_p\left(p(n) \sqrt{\frac{\log n}{n}}\right). \tag{46}$$

We define the auxiliary statistic \widetilde{T}_n based on the imaginary companion process (\widetilde{X}_t) as follows:

$$\widetilde{T}_n = a_n^{-1} \left(\sigma_0^{-1/2} \max_{1 \leq h \leq d_n} \sqrt{n} |\widehat{\gamma}(h) - \gamma(h)| - b_n \right), \tag{47}$$

where $\widehat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} \widetilde{X}_{t+k} \widetilde{X}_t$.

Proof of Theorem 1. Both T_n and \widetilde{T}_n converge in distribution to the Gumbel distribution under Assumptions 1 and 2, as long as $r > 1/2$. While the weak convergence for T_n follows from Xiao and Wu (2014), Theorem 1, directly, the same theorem also yields the result for \widetilde{T}_n since the companion process (\widetilde{X}_t) is a simpler linear time series which fulfills the assumptions of Theorem 1 as well. Because convergence in Kolmogorov distance is equivalent to convergence in distribution if the target distribution is continuous, we obtain by using the triangular inequality for d_K that $d_K(\mathcal{L}(\widetilde{T}_n), \mathcal{L}(T_n)) \rightarrow 0, n \rightarrow \infty$. In Lemma 2 we argue that $d_K(\mathcal{L}^*(T_n^*), \mathcal{L}(\widetilde{T}_n)) \rightarrow 0$ as $n \rightarrow \infty$. \square

Lemma 2. Under the same assumption as in Theorem 1 we obtain for the AR-sieve bootstrap

$$d_1(\mathcal{L}^*(T_n^*), \mathcal{L}(\widetilde{T}_n)) \rightarrow 0, \quad \text{in probability, } n \rightarrow \infty, \tag{48}$$

where d_1 is defined in (21). Equation (48) implies convergence in Kolmogorov distance.

Remark 4. As can be seen from the following proof, the rate of convergence to zero in (48) is of polynomial order. This means that the proposed bootstrap procedure approximates the distribution of the statistic of interest \widetilde{T}_n , cf. (47), based on the companion process (\widetilde{X}_t) defined in (26) with a better rate compared to the asymptotic Gumbel distribution. Note that it is well known, cf. the theorem in Hall (1979), that even for i.i.d. Gaussian random variables the approximation rate of the limiting Gumbel distribution is only of logarithmic order. But it must be said that this does not mean that our bootstrap proposal outperforms the approximation rate of the asymptotic Gumbel distribution. Namely, in the proof of Theorem 1 we differentiate between the approximation of $\mathcal{L}(\widetilde{T}_n)$ by $\mathcal{L}^*(T_n^*)$, which indeed is of polynomial order, and the approximation of $\mathcal{L}(T_n)$, the distribution of the statistic of interest, by $\mathcal{L}(\widetilde{T}_n)$. This last approximation relies on the asymptotic Gumbel distribution and therefore is only of logarithmic order.

However, if we consider the classical AR-sieve bootstrap for causal $AR(\infty)$ time series and make assumptions that ensure polynomial convergence to zero of $d_2(F_n, F)$ (cf. Remark 1 for details), the situation is different. In this case the companion process (\widetilde{X}_t) from (26) has to be slightly modified to describe what the bootstrap actually

imitates. (\tilde{X}_t) then has to be defined as the autoregression given in (26) but with innovations $\tilde{\varepsilon}_t$ defined as i.i.d. random variables with marginal distribution (and this is the difference to (26)) given by the marginal distribution of the true innovations e_t from the Wold-type representation (12) (cf. Remark 5). In this scenario, the companion process (\tilde{X}_t) coincides with the underlying time series (X_t) . Then we indeed obtain for this specific situation that the classical AR-sieve bootstrap achieves a polynomial convergence rate in approximating the distribution of T_n and therefore outperforms the approximation rate by the asymptotic Gumbel distribution.

Proof. It suffices to show that $E^*|T_n^* - \tilde{T}_n| = o_p(1)$. We split into two parts,

$$E^*|T_n^* - \tilde{T}_n| \leq a_n^{-1} (\sigma_0^*)^{-1/2} E^* \max_{1 \leq h \leq d_n} \sqrt{n} |\hat{\gamma}^*(h) - \hat{\gamma}(h) + \gamma(h) - \gamma^*(h)| \\ + a_n^{-1} \left| (\sigma_0^*)^{-1/2} - \sigma_0^{-1/2} \right| E^* \max_{1 \leq h \leq d_n} \sqrt{n} |\hat{\gamma}(h) - \gamma(h)| = I + II.$$

Let us introduce $X_t^+ = \sum_{j=1}^{\infty} b_j e_{t-j}^* + e_t^*$ and use $\hat{\gamma}^+(h) = \frac{1}{n} \sum_{t=1}^{n-h} X_{t+h}^+ X_t^+$, $\gamma^+(h) = E^* X_{t+h}^+ X_t^+$ to split term I even further:

$$I \leq a_n^{-1} (\sigma_0^*)^{-1/2} E^* \max_{1 \leq h \leq d_n} \sqrt{n} |\hat{\gamma}^*(h) - \hat{\gamma}^+(h) + \gamma^+(h) - \gamma^*(h)| \\ + a_n^{-1} (\sigma_0^*)^{-1/2} E^* \max_{1 \leq h \leq d_n} \sqrt{n} |\hat{\gamma}^+(h) - \hat{\gamma}(h) + \gamma(h) - \gamma^+(h)| = I_a + I_b.$$

The following bounds for I_a , I_b , and II are obtained after substituting the MA(∞) representations of the processes \tilde{X}_t , X_t^+ and X_t^* into the sample autocovariance terms. Let $S_{n,h,i,j} = \sum_{t=1}^{n-h} (e_{t-j}^* e_{t-(i-h)}^*) - \tilde{\sigma}(p(n))^2 \mathbb{1}_{[j=i-h]}$. Applying Pollard's lemma (Horowitz and Karandikar, 1994, Lemma 2.3) gives

$$I_a \leq a_n^{-1} (\sigma_0^*)^{-1/2} \sum_{j=0}^{\infty} |\hat{b}_{n,j} - b_j| \sqrt{d_n} \left(\max_{1 \leq h \leq d_n} E^* \left(|S_{n,h,i,j}| / \sqrt{n} \right)^2 \right)^{1/2} \mathcal{O}_p(1). \quad (49)$$

By straightforward calculations we obtain $E^* (|S_{n,h,i,j}| / \sqrt{n})^2 = \mathcal{O}_p(1)$ because we assumed finite fourth moments of (v_t^*) in our Bootstrap Proposal. Assumption 3 implies $\sum_{j=0}^{\infty} |\hat{b}_{n,j} - b_j| = \mathcal{O}_p \left(p(n)^2 \sqrt{\log n/n} + p(n)^{-r} \right)$ (in particular see (43) and (46)). Using $d_n = \mathcal{O}(n^n)$ we can bound I_a by

$$I_a \leq [p(n)^2 n^{-1/2} n^{n/2} + p(n)^{-r} n^{n/2}] \mathcal{O}_p(\log n). \quad (50)$$

By tedious but straightforward calculations, applying Pollard's lemma and by using $E(v_t^*)^2 = 1$ we obtain

$$I_b \leq \sqrt{d_n} \sqrt{E^*(\tilde{\varepsilon}_1 - e_1^*)^2} \mathcal{O}_p(\sqrt{\log n}) = n^{n/2} \left| \sqrt{E e_1^2} - \tilde{\sigma}(p(n)) \right| \mathcal{O}_p(\sqrt{\log n}). \quad (51)$$

In order that the upper bound in (51) converges to 0, we need a certain rate of decay for $\left| \sqrt{E e_1^2} - \tilde{\sigma}(p(n)) \right| \leq |\tilde{\sigma}_e - \tilde{\sigma}(p(n))| + \left| \sqrt{E e_1^2} - \tilde{\sigma}_e \right|$, $\tilde{\sigma}_e^2 = \frac{1}{n-p(n)} \sum_{t=p(n)+1}^n e_t^2$. The expectation of the first summand can be bounded up to a constant by the square root of

$$E \left(\sum_{j=p(n)+1}^{\infty} a_j X_{t-j} + \sum_{j=1}^{p(n)} (a_j - \hat{a}_j(p(n))) X_{t-j} \right)^2 \leq \left(p(n)^{-2r} + p(n)^2 \frac{\log n}{n} \right) \mathcal{O}(1).$$

The last inequality is obtained by using Assumption 3. Together with (23) this gives

$$I_b \leq (n^{-\beta/2} n^{n/2} + p(n)^{-r} n^{n/2} + p(n) n^{-1/2} n^{n/2}) \mathcal{O}_p(\log n). \tag{52}$$

Finally we consider term II. Note that $E \max_{1 \leq h \leq d_n} \sqrt{n} |\widehat{\gamma}(h) - \gamma(h)| = \mathcal{O}(d_n^{1/2}) = \mathcal{O}(n^{n/2})$ holds along the same lines as for I_a as long as $r > 1/2$. This means that term II can be bounded by $|(\sigma_0^*)^{-1/2} - \sigma_0^{-1/2}| \mathcal{O}_p(\sqrt{\log n} n^{n/2})$. An upper bound for $|(\sigma_0^*)^{-1/2} - \sigma_0^{-1/2}|$ can be obtained by calculating

$$|(\sigma_0^*)^{-1/2} - \sigma_0^{-1/2}|^2 \leq (|\sigma_0^* - \sigma_0|) \mathcal{O}_p(1) \leq \left(\sum_{j=0}^{\infty} |b_j - \widehat{b}_j(p(n))| + \left| \sqrt{Ee_1^2} - \widetilde{\sigma}(p(n)) \right| \right) \mathcal{O}_p(1).$$

Using the rates for both summands which have been derived above, we obtain

$$II \leq (p(n)^2 n^{-1/2} n^{n/2} + n^{(n-\beta)/2} + p(n)^{-r} n^{n/2}) \mathcal{O}_p(\log n). \tag{53}$$

Under the conditions given in (24) the upper bounds given in (50), (52), and (53) are $o_p(n^{-\gamma})$ for some suitable $\gamma > 0$. □

Remark 5. In (51) the definition of $\widetilde{\varepsilon}_t$ and e_t^* via the same i.i.d. sequence (v_t^*) ensures that $\sqrt{E(\widetilde{\varepsilon}_1 - e_1^*)^2} = d_2(\widetilde{\varepsilon}_1, e_1^*)$, where d_2 is the Mallows distance, cf. (22). In case that we would have used in the bootstrap process innovations e_t^* drawn from the empirical distribution of centered residuals of the AR fit (cf. Remark 1), we have to change the definition of the companion process (\widetilde{X}_t) in the way that the i.i.d. innovations $\widetilde{\varepsilon}_t$ possess the marginal distribution, F say, of the residuals e_t of the AR Wold representation (12). In this case we would arrive in (51) with the bound

$$\sqrt{d_n} d_2(F, F_n) \mathcal{O}_p(\sqrt{\log n}),$$

where F_n denotes the empirical distribution of a sequence e_1, \dots, e_n from the only uncorrelated innovations of the representation (12). To complete the proof, we then would need a convergence rate for $d_2(F, F_n)$. As already stated in Remark 1 such rates are available in the literature, but need further and typically quite restrictive additional assumptions on the innovations on the AR representation, that we want to avoid.

Remark 6. Concerning validity of the bootstrap for the maximum deviation of the lag window spectral density estimator (cf. (32)) we first state that from Theorems 3–5 in Liu and Wu (2010) or Theorem 2.3 in Woodrooffe and Van Ness (1967) (because \widetilde{X}_t is a linear time series) we have that

$$\widetilde{S}_n := \alpha_n^{-1} \left[\sqrt{\frac{n}{B_n}} \|w\|_2^{-1} \max_{j=0,1,\dots,B_n} \frac{|\widehat{f}(\lambda_{j,n}) - f(\lambda_{j,n})|}{f(\lambda_{j,n})} - \beta_n \right] \xrightarrow{D} \text{Gumbel}, \tag{54}$$

where $\widehat{f}(\lambda)$ denotes the lag window estimator based on n observations from the companion process (\widetilde{X}_t) (cf. (26)). Note that the spectral density of (\widetilde{X}_t) coincides with the spectral density f of the underlying time series.

Next, the assumption $n/B_n^5 \rightarrow 0$ guarantees that the biases of \widehat{f} and of \widehat{f}^* are sufficiently small so that we only need to compare

$$S_n^{',*} := \alpha_n^{-1} \left[\sqrt{\frac{n}{B_n}} \|w\|_2^{-1} \max_{j=0,1,\dots,B_n} |\widehat{f}^*(\lambda_{j,n}) - E\widehat{f}^*(\lambda_{j,n})| / f^*(\lambda_{j,n}) - \beta_n \right]$$

and

$$\tilde{S}'_n := \alpha_n^{-1} \left[\sqrt{\frac{n}{B_n}} \|w\|_2^{-1} \max_{j=0,1,\dots,B_n} |\hat{f}(\lambda_{j,n}) - \widehat{\mathbb{E}}\hat{f}(\lambda_{j,n})| / f(\lambda_{j,n}) - \beta_n \right]$$

with each other. We show $S'_n{}^* - \tilde{S}'_n = o_p(1)$. Because the spectral density f^* of the bootstrap time series (X_t^*) (cf. (19)) is uniformly consistent for the underlying spectral density f the necessary consideration is further reduced to

$$\alpha_n^{-1} \sqrt{\frac{n}{B_n}} \max_{j=0,1,\dots,B_n} |\hat{f}^*(\lambda_{j,n}) - \widehat{\mathbb{E}}\hat{f}^*(\lambda_{j,n}) - \hat{f}(\lambda_{j,n}) + \widehat{\mathbb{E}}\hat{f}(\lambda_{j,n})|.$$

But this expression is bounded through

$$\alpha_n^{-1} \sqrt{B_n} \max_{0 \leq h \leq B_n} \sqrt{n} |\hat{\gamma}^*(h) - \hat{\gamma}(h) + \gamma(h) - \gamma^*(h)|$$

and for exactly this term we have shown in the proof of Lemma 2 that it is bounded by $\mathcal{O}_p(\sqrt{B_n} n^{-\gamma})$. For $\beta = 1/2$, $r = 2$, a suitable choice for the rate a would lead to $\gamma > 0.1$ so that we end up with $S'_n{}^* - \tilde{S}'_n \rightarrow 0$ in probability, as $n \rightarrow \infty$. Together with (54) this yields validity of the bootstrap for maximum deviation of lag window spectral density estimators.

5. CONCLUSION

We showed that the AR-sieve bootstrap procedure successfully can be applied to maximum deviations of sample autocovariances and autocorrelations as well as lag window spectral density estimators. These results extend the range of validity of AR sieve bootstrap beyond the class of generalized means as has been considered in Kreiss *et al.* (2011). The proof is achieved as in this article and makes again use of the companion process (\tilde{X}_t) , cf. (26), which indeed is the process which the AR-sieve bootstrap time series is able to mimic. Since the companion behaves according to maximum deviations of sample autocovariances and sample autocorrelations asymptotically exactly behaves like the underlying time series, which indeed stems from a much more general class of time series (7), it is obtained that AR-sieve bootstrap works for that case.

Simulation results demonstrate that the AR-sieve bootstrap works especially well for underlying linear time series but is also able to compete with the asymptotic distribution and the block of blocks bootstrap as was considered in Xiao and Wu (2014).

ACKNOWLEDGEMENTS

The authors are extremely grateful to numerous personal discussions with Efstathios Papanoditis (University of Cyprus) on bootstrap procedures in general and with Han Xiao (Rutgers University) on the special situation of maximum deviations of sample autocovariances and autocorrelations. Moreover, the authors very much appreciate the reports of two anonymous reviewers. Their thorough reports lead to a considerable improvement of the paper. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study. However, the R-code developed for the simulations can be provided upon reasonable request.

REFERENCES

- An HZ, Chen ZG, Hannan EJ. 1982. Autocorrelation, autoregression and autoregressive approximation. *Annals of Statistics* **10**: 926–936.
- Baxter G. 1962. An asymptotic result for the finite predictor. *Mathematica Scandinavica* **10**: 137–144.
- Bickel PJ, Freedman DA. 1981. Some asymptotic theory for the bootstrap. *Annals of Statistics* **9**: 1196–1217.
- Brillinger D. R. 1981. *Time Series: Data Analysis and Theory*. San Francisco, CA: Holden Day. See also: SIAM Classics in Applied Mathematics (2001).
- Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*, 2nd ed. New York: Springer.
- Bühlmann P. 1997. Sieve bootstrap for time series. *Bernoulli* **3**: 123–148.
- Cheng R, Pourahmadi M. 1993. Baxter's inequality and convergence of finite predictors of multivariate stochastic processes. *Probability Theory and Related Fields* **95**: 115–124.
- Dedecker J., Doukhan P., Lang G., Leon J.R., Louhichi S., Prieur C. 2007. Weak dependence: with examples and applications. In *Springer Lecture Notes in Statistics* Heidelberg: Springer.
- Dedecker J, Merlevède F. 2017. Behavior of the Wasserstein distance between the empirical and the marginal distributions of stationary α -dependent sequences. *Bernoulli* **23**: 2083–2127.
- Doukhan P, Louhichi S. 1999. A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications* **84**: 313–342.
- Efron B. 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**: 1–16.
- Fournier N, Guillin A. 2015. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields* **162**: 707–738.
- Hall P. 1979. On the rate of convergence of normal extremes. *Journal of Applied Probability* **16**: 433–439.
- Hannan EJ. 1974. The uniform convergence of autocovariances. *Annals of Statistics* **2**: 803–806.
- Hannan E. J., Deistler M. 1988. *The Statistical Theory of Linear Systems*. Chichester: John Wiley & Sons. Republished 2012 by the Society for Industrial and Applied Mathematics (SIAM).
- Hannan EJ, Kavalieris L. 1983. The convergence of autocorrelations and autoregressions. *Australian Journal of Statistics* **25**: 287–297.
- Hannan EJ, Kavalieris L. 1984. Multivariate linear time series models. *Advances in Applied Probability* **16**: 492–561.
- Horowitz J, Karandikar RL. 1994. Mena rates of convergence of empirical measures in the Wasserstein metric. *Journal of Computational and Applied Mathematics* **55**: 261–273.
- Jirak M. 2011. On the maximum of covariance estimators. *Journal of Multivariate Analysis* **102**: 1032–1046.
- Kreiss J-P. 1992. Bootstrap procedures for AR(∞)-processes. In *Lecture Notes in Economics and Mathematical Systems No. 376 (Proc. Bootstrapping and Related Techniques, Trier)* Heidelberg: Springer; 107–113.
- Kreiss J-P, Paparoditis E, Politis DN. 2011. On the range of validity of the autoregressive sieve bootstrap. *Annals of Statistics* **39**: 2103–2130.
- Kreiss J.-P., Paparoditis E. 2021. *Bootstrap for time series: theory and applications*. Heidelberg: Springer, Forthcoming monograph.
- Lahiri SN. 2003. *Resampling Methods for Dependent Data*. New York: Springer.
- Liu W, Wu WB. 2010. Asymptotics of spectral density estimates. *Econometric Theory* **26**: 1218–1245.
- Meyer M, Paparoditis E, Kreiss J-P. 2020. Extending the validity of frequency domain bootstrap methods to general stationary processes. *Annals of Statistics* **48**: 2404–2427.
- Meyer M., Paparoditis E. 2021+. A frequency domain bootstrap for general multivariate stationary processes. Preprint: <https://arxiv.org/abs/2102.01943v1>.
- Panaretos VM, Zemel Y. 2019. Statistical aspects of Wasserstein distances. *Annual Review of Statistics and its Applications* **6**: 405–431.
- Paparoditis E, Politis DN. 2012. Nonlinear spectral density estimation thresholding the correlogram. *Journal of Time Series Analysis* **33**: 386–397.
- Paparoditis E, Streitberg B. 1992. Order identification statistics in stationary autoregressive moving average models: vector autocorrelations and the bootstrap. *Journal of Time Series Analysis* **13**: 415–434.
- Politis DN. 2003. The impact of bootstrap methods on time series analysis. *Statistical Science* **18**: 219–230.
- Rosenblatt M. 1956. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America* **42**: 43–47.
- Rosenblatt M. 1961. Independence and dependence. In *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob. Vol. II* Berkeley: Univ. of California Press; 431–443.
- Rosenblatt M. 1972. Central limit theorems for stationary processes. In *Proc. Sixth Berkeley Symp. on Probability and Statistics Vol. II* Los Angeles: Univ. of California Press; 551–561.
- Rosenblatt M. 1985. *Stationary Sequences and Random Fields*. Basel: Birkhäuser.
- Samworth R., Johnson O. 2005. Convergence of the empirical process in Mallows distance, with an application to bootstrap performance. Preprint arXiv:math/0406603.

- Woodroffe MB, Van Ness JW. 1967. The maximum deviation of sample spectral densities. *The Annals of Mathematical Statistics* **38**: 1558–1569.
- Wu WB. 2005. Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 14150–14154.
- Wu WB. 2009. An asymptotic theory for sample covariances of Bernoulli shifts. *Stochastic Processes and their Applications* **120**: 2412–2431.
- Wu WB. 2011. Asymptotic theory for stationary time series. *Statistics and its Interface* **4**: 207–226.
- Xiao H, Wu WB. 2014. Portmanteau test and simultaneous inference for serial covariances. *Statistica Sinica* **24**: 577–599.