Sieve Bootstrap for Time Series

Author(s): Peter Bühlmann

Source: *Bernoulli*, Jun., 1997, Vol. 3, No. 2 (Jun., 1997), pp. 123-148

Published by: International Statistical Institute (ISI) and the Bernoulli Society for Mathematical Statistics and Probability

Stable URL: https://www.jstor.org/stable/3318584

Sieve bootstrap for time series

PETER BÜHLMANN

Department of Statistics, University of California, Berkeley CA 94720-3860, USA. e-mail: buhlmann@stat.berkeley.edu

We study a bootstrap method which is based on the method of sieves. A linear process is approximated by a sequence of autoregressive processes of order $p = p(n)$, where $p(n) \to \infty$, $p(n) = o(n)$ as the sample size $n \to \infty$. For given data, we then estimate such an AR($p(n)$) model and generate a bootstrap sample by resampling from the residuals. This sieve bootstrap enjoys a nice nonparametric property, being model-free within a class of linear processes.

We show its consistency for a class of nonlinear estimators and compare the procedure with the blockwise bootstrap, which has been proposed by Künsch in 1989. In particular, the sieve bootstrap variance of the mean is shown to have a better rate of convergence if the dependence between separated values of the underlying process decreases sufficiently fast with growing separation.

Finally, a simulation study helps to illustrate the advantages and disadvantages of the sieve compared to the blockwise bootstrap.

Keywords: Akaike information criterion; AR(∞); ARMA; autoregressive approximation; autoregressive spectrum; blockwise bootstrap; linear process; resampling; stationary sequence; threshold model

1. Introduction

The bootstrap proposed by Efron (1979) has become a powerful nonparametric method for estimating the distribution of a statistical quantity. However, by ignoring the order of the observations, it usually fails for dependent observations.

In the context of stationary time series two different bootstrap methods have been proposed. One is a model-based approach, which resamples from approximately i.i.d. residuals; cf. Freedman (1984); Efron and Tibshirani (1986); Bose (1988); and Franke and Kreiss (1992). Tsay (1992) uses this approach for diagnostics in the time-series context. Obviously, these procedures are sensitive to model misspecification and the attractive nonparametric features of Efron's bootstrap is lost. A nonparametric, purely model-free bootstrap scheme for stationary observations has been given by Künsch (1989); see also Liu and Singh (1992). Künsch's idea is to resample overlapping blocks of consecutive observations where the blocklength involved grows slowly with the sample size. By construction we call this procedure *blockwise bootstrap* (it is sometimes also called *moving blocks bootstrap*). Since the pioneering paper by Künsch (1989), the blockwise bootstrap and modifications thereof have been studied by Politis and Romano (1992; 1993), Shao and Yu (1993), Naik-Nimbalkar and Rajarshi (1994), Bühlmann and Künsch (1994; 1995) and Bühlmann (1993; 1994; 1995a). Generally, this blockwise bootstrap works satisfactorily and enjoys the property of being robust against misspecified models. However, the resampled

series exhibits artefacts which are caused by joining randomly selected blocks. The dependence between different blocks is neglected in the resampled series and the bootstrap sample is not (conditionally) stationary. Politis and Romano (1994) have given a modification of the blockwise bootstrap which yields a (conditionally) stationary bootstrap sample. However, their method depends on a tuning parameter which seems difficult to control.

Our approach here takes up the older idea of fitting parametric models first and then resampling from the residuals. But instead of considering a fixed finite-dimensional model we approximate an infinite-dimensional, nonparametric model by a sequence of finite-dimensional parametric models. This strategy is known as the method of *sieves* (cf. Grenander 1991; Geman and Hwang 1982) and explains the name given to our procedure. Implicitly, this approach is quite often used when choosing a model adaptively by a criterion such as the Akaike information criterion (AIC), rather than considering a pre-fixed model. To fix ideas, we approximate the true underlying stationary processes by an autoregressive model of order p, where $p = p(n)$ is a function of the sample size n with $p(n) \to \infty$, $p(n) = o(n)$ $(n \to \infty)$. Our definition of the bootstrap for a fixed model is the same as already given by Freedman (1984). However, we take here the point of view of approximating sieves. As with the blockwise bootstrap, this resampling procedure is again nonparametric and, moreover, its bootstrap sample is (conditionally) stationary and does not exhibit additional artefacts of the dependence structure as above.

Kreiss (1988; 1992) also discusses the bootstrap for AR(∞) models. But his approach only covers linear processes

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

where $\{\psi_j\}_{j=0}^{\infty}$ decays exponentially and $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is an i.i.d. sequence with $\mathrm{E}[\varepsilon_t] = 0$. This is not satisfactory, because it covers only linear processes with a very weak dependence, usually having exponentially decaying mixing coefficients. This does not allow the approximating autoregressive models to be interpreted as a sieve for a broader subclass of stationary processes. Kreiss (1988; 1992) shows, under the above conditions, consistency of the bootstrap for sample autocovariances and the linear part of a class of estimates for the unknown autoregressive parameters of the approximating autoregressive model. Our results in Section 3 are more general.

A related approach in the frequency domain of stationary time series has been given by Janas (1992). There one basically resamples from studentized periodogram values, yielding a consistent procedure for smooth functions of the periodogram. This approach can be interpreted as approximating the modulus $|\Phi(\mathrm{e}^{-\mathrm{i}\lambda})| (0 \leqslant \lambda \leqslant \pi)$, where $\Phi(z) = \sum_{j=0}^{\infty} \phi_j z^j$, $\phi_0 = 1$ $(z \in \mathbb{C})$ is the AR(∞) transfer function corresponding to the AR(∞) process $\sum_{j=0}^{\infty} \phi_j X_{t-j} = \varepsilon_t$, $t \in \mathbb{Z}$. Under some regularity conditions this model is equivalent to the linear model in (1.1). Our sieve bootstrap captures more in that we approximate the whole transfer function $\Phi(z)$ $(|z| \leqslant 1)$ instead of only its modulus as above.

We justify the sieve bootstrap by showing its consistency for statistics based on linear processes as in (1.1), where $\{\psi_j\}_{j=0}^{\infty}$ are allowed to decay at a certain polynomial speed. In

practice we argue here, that by fitting an autoregressive model first, using, for example, the AIC, and importantly, by taking the point of view of sieve approximation, this (model-based) bootstrap can still be regarded as a nonparametric method. Our results contribute in this direction.

In Section 2 we give the exact definition of the sieve bootstrap. In Section 3 we present the consistency of the sieve bootstrap for the arithmetic mean and a class of nonlinear statistics. In particular, we include a comparison with the blockwise bootstrap which indicates that the sieve bootstrap works better for very weak dependent processes, that is, for processes with fast decaying coefficients $\{\psi_j\}_{j=0}^{\infty}$ in (1.1). In Section 4 we present the results of a simulation study, comparing the performance of the blockwise and sieve bootstrap. To explore some limits of the sieve bootstrap we consider also nonlinear models which cannot be represented as in (1.1) with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ i.i.d. Surprisingly, the sieve bootstrap also works well for a non-Gaussian autoregressive threshold model which is beyond the linear theory. In Section 5 we include the proofs and some probabilistic properties of the sieve bootstrap.

2. Definition of the sieve bootstrap

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a real-valued, stationary process with expectation $\mathrm{E}[X_t] = \mu_X$. If $\{X_t\}_{t\in\mathbb{Z}}$ is purely stochastic, we know by Wold's theorem (cf. Anderson 1971) that $\{X_t - \mu_X\}_{t\in\mathbb{Z}}$ can be written as a one-sided infinite-order moving-average process

$$X_t - \mu_X = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \psi_0 = 1, \, t \in \mathbb{Z}, \tag{2.1}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of uncorrelated variables with $\mathrm{E}[\varepsilon_t] = 0$ and $\sum_{j=0}^{\infty}\psi_j^2 < \infty$. We require invertibility of the process in (2.1) which narrows the class of stationary processes a little bit. Under these additional assumptions of invertibility (cf. Anderson 1971, Theorem 7.6.9) we can represent $\{X_t\}_{t\in\mathbb{Z}}$ as a one-sided infinite-order autoregressive process

$$\sum_{j=0}^{\infty} \phi_j(X_{t-j} - \mu_X) = \varepsilon_t, \qquad \phi_0 = 1, \, t \in \mathbb{Z}, \tag{2.2}$$

with $\sum_{j=0}^{\infty}\phi_j^2 < \infty$.

The representation (2.2) motivates an autoregressive approximation as a sieve for the stochastic process $\{X_t\}_{t\in\mathbb{Z}}$. By (2.1) we could also use a moving-average approximation, but we rely on autoregressive approximation which, as a linear method, is much more popular, faster and better known as a successful technique in different situations (cf. Berk 1974; An *et al.* 1982; Hannan 1987).

We now give the definition of our sieve bootstrap. Denote by $X_1, \ldots, X_n$ a sample from the process $\{X_t\}_{t\in\mathbb{Z}}$. In a first step we fit an autoregressive process, with increasing order $p(n)$ as the sample size n increases. Let $p = p(n) \to \infty$ with $p(n) = o(n)$ $(n \to \infty)$. We then estimate the coefficients $\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n}$ corresponding to model (2.2), usually (but not necessarily) by the Yule–Walker estimates (cf. Brockwell and Davis 1987, Chapter 8.1).

Note that for this purpose we first have to subtract the sample mean $\bar{X}$. This procedure yields residuals

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

In a second step we construct the resampling based on this autoregressive approximation. We centre the residuals

$$\tilde{\varepsilon}_{t,n} = \hat{\varepsilon}_{t,n} - (n-p)^{-1} \sum_{t=p+1}^{n} \hat{\varepsilon}_{t,n} \qquad (t = p+1, \ldots, n)$$

and denote the empirical c.d.f. of $\{\tilde{\varepsilon}_{t,n}\}_{t=p+1}^{n}$ by

$$\hat{F}_{\varepsilon,n}(\cdot) = (n-p)^{-1} \sum_{t=p+1}^{n} 1_{[\tilde{\varepsilon}_{t,n} \leqslant \cdot]}.$$

Then we can resample, for any $t \in \mathbb{Z}$, ε_t^* i.i.d. $\sim \hat{F}_{\varepsilon,n}$, and define $\{X_t^*\}_{t\in\mathbb{Z}}$ by the recursion

$$\sum_{j=0}^{p(n)} \hat{\phi}_{jn}(X_{t-j}^* - \bar{X}) = \varepsilon_t^* \qquad \hat{\phi}_{0,n} = 1. \tag{2.3}$$

In practice, we construct a sieve bootstrap sample $X_1^*, \ldots, X_n^*$ in the following way: choose starting values, for example, equal to $\bar{X}$, generate an AR($p(n)$) process according to (2.3) until 'stationarity' is reached and then throw the first generated values away. Such an approach is implemented for example in the S-Plus function *arima.sim*. This bootstrap construction induces a conditional probability $\mathbb{P}^*$, given the sample $X_1, \ldots, X_n$. As usual, we denote quantities with respect to $\mathbb{P}^*$ with an asterisk *.

Consider now any statistics $T_n = T_n(X_1, \ldots, X_n)$, where T_n is a measurable function of n observations. We define the bootstrapped statistics T_n^* by the plug-in principle:

$$T_n^* = T_n(X_1^*, \ldots, X_n^*).$$

This bootstrap construction exhibits some features which are different from Künsch's (1989) blockwise bootstrap. It again yields a (conditionally) stationary bootstrap sample and does not exhibit artefacts in the dependence structure as in the blockwise bootstrap, where the dependence between different blocks is neglected. The sieve bootstrap sample is not a subset of the original sample.

Moreover, the sieve bootstrap is often more simple to apply. There is no need to 'prevectorize' the original observations. Let us explain the vectorizing of observations for the blockwise bootstrap. Suppose the statistic of interest T_n can be written as a functional T at an m-dimensional empirical c.d.f. $F_n^{(m)}$,

$$T_n = T(F_n^{(m)}) \qquad (m \geqslant 1).$$

Denote by $Y_t = (X_t, \ldots, X_{t+m-1})^{\mathrm{T}}$ $(t = 1, \ldots, n-m+1)$ the m-dimensional vectorized

observations. Then $F_n^{(m)}$ is the empirical c.d.f. of $\{Y_t\}_{t=1}^{n-m+1}$. The blockwise bootstrap is now applied to the Y_ts, thus being a 'block of blocks' bootstrap scheme. For different dimensions m one has to use different vectorized observations. As an example, for blockwise bootstrapping the autocorrelation function $\rho(k)$ in some range $0 \leqslant k \leqslant M$ we either have to use $m = k + 1$ for every individual k, which results in $M + 1$ different bootstrap samples, or use $m = M + 1$ for all k, which seems quite crude when k is small. Our sieve bootstrap has the advantage of avoiding the construction of vectorized observations and enjoys the properties of a plug-in rule.

Also, our sieve bootstrap is simpler for unequally spaced data or series with many missing values. Imagine that we have an underlying stationary process $\{Y_t\}_{t \in \mathbb{Z}}$ from which we have observed $X_t = Y_{z(t)t}$, $t = 1, \ldots, n$ where $z_1, \ldots, z_n$ are binary non-random variables in $\{0, 1\}$. A missing value at time point t is then given by $z_t = 0$, by setting Y_0 as 'NA' (not available). The variables $z_1, \ldots, z_n$ are known. With the sieve bootstrap we can easily mimic this situation by constructing the sample $X_1^*, \ldots, X_n^* = Y_{z(1)1}^*, \ldots, Y_{z(n)n}^*$, where Y_0^* is set to be 'NA' and $Y_1^*, \ldots, Y_n^*$ is the sieve bootstrap sample based on an AR(p), which is estimated by the observed time series with missing values $X_1, \ldots, X_n$. In the case of the blockwise bootstrap, an additional complication arises. One has to use blocks either with non-equal numbers of observations or with non-constant blocklengths.

3. Main results

3.1. ASSUMPTIONS

We now consider more carefully the models (2.1) and (2.2) and give the precise assumptions about the stationary process $\{X_t\}_{t \in \mathbb{Z}}$ from which a sample $X_1, \ldots, X_n$ is drawn. We prefer the formulation in the MA(∞) representation (2.1) rather than in the AR(∞) representation (2.2). Write

$$\Phi(z) = \sum_{j=0}^{\infty} \phi_j z^j, \qquad \phi_0 = 1, z \in \mathbb{C},$$

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j, \qquad \psi_0 = 1, z \in \mathbb{C}.$$

Then the models (2.1) and (2.2) can be written as

$$\Phi(B)(X - \mu_X) = \varepsilon, \qquad X - \mu_X = \Psi(B)\varepsilon,$$

where B denotes the back-shift operator $(Bx)_t = x_{t-1}$, $x \in \mathbb{R}^{\mathbb{Z}}$. At least formally we can see that $\Psi(z) = 1/\Phi(z)$. Denote by $\mathscr{F}_t = \sigma(\{\varepsilon_s; s \leqslant t\})$, the σ-field generated by $\{\varepsilon_s\}_{s=-\infty}^{t}$. Our main assumptions for the model are the following.

Assumption A1. $X_t - \mu_X = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, $\psi_0 = 1$ $(t \in \mathbb{Z})$ *with* $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ *stationary, ergodic and* $\mathrm{E}[\varepsilon_t | \mathscr{F}_{t-1}] \equiv 0$, $\mathrm{E}[\varepsilon_t^2 | \mathscr{F}_{t-1}] \equiv \sigma^2 < \infty$, $\mathrm{E}|\varepsilon_t|^s < \infty$ *for some* $s \geqslant 4$.

Assumption A2. $\Psi(z)$ *is bounded away from zero for* $|z| \leqslant 1$, $\sum_{j=0}^{\infty} j^r |\psi_j| < \infty$ *for some* $r \in \mathbb{N}$.

Since our sieve bootstrap scheme draws independently from the residuals, it is usually unable to catch the probability structure of a statistic based on a model satisfying assumption A1 with non-independent variables $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. The arithmetic mean as a linear statistic is an exception in this respect. We therefore sometimes strengthen assumption A1 to

Assumption A1'. $X_t - \mu_X = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, $\psi_0 = 1$ $(t \in \mathbb{Z})$ *with* $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ *i.i.d. and* $\mathrm{E}[\varepsilon_t] = 0$, $\mathrm{E}|\varepsilon_t|^s < \infty$ *for some* $s \geqslant 4$.

Assumption A1 is basically the same as in An *et al.* (1982). Assumption A2 includes models with polynomial decay of the coefficients $\{\psi_j\}_{j=0}^{\infty}$ or equivalently $\{\phi_j\}_{j=0}^{\infty}$. ARMA$(p, q)$ models $(p < \infty, q < \infty)$ satisfy assumption A2 with an exponential decay of $\{\psi_j\}_{j=0}^{\infty}$. Note that assumption A2 implies that $\Phi(z)$ is bounded away from zero for $|z| \leqslant 1$ and $\sum_{j=0}^{\infty} j^r |\phi_j| < \infty$. In particular, assumption A1' is more restrictive than the conditions in Künsch (1989) for the blockwise bootstrap, which is shown to be generally valid for strong-mixing sequences (cf. Künsch 1989).

We now specify our autoregressive approximation and make the following widely used assumption:

Assumption B. $p = p(n) \to \infty$, $p(n) = o(n)$ $(n \to \infty)$ *and* $\hat{\phi}_p = (\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n})^{\mathrm{T}}$ *satisfy the empirical Yule–Walker equations*

$$\hat{\Gamma}_p \hat{\phi}_p = -\hat{\gamma}_p,$$

where $\hat{\Gamma}_p = [\hat{R}(i - j)]_{i,j=1}^{p}$, $\hat{\gamma}_p = (\hat{R}(1), \ldots, \hat{R}(p))^{\mathrm{T}}$, $\hat{R}(j) = n^{-1} \sum_{t=1}^{n-|j|} (X_t - \bar{X})(X_{t+|j|} - \bar{X})$.

Below we write $R(j)$ for $\mathrm{cov}(X_0, X_j)$.

3.2. BOOTSTRAPPING THE MEAN

Our first result, proved in Section 5.3, shows consistency in the simple case of the arithmetic mean. As mentioned in Section 3.1, the sieve bootstrap in this case will be shown to be consistent even for processes as in assumption A1 with non-independent innovations.

Theorem 3.1. *Let assumptions A1 with* $s = 4$, *A2 with* $r = 1$ *and B with* $p(n) = o((n/\log(n))^{1/4})$ *hold. Then:*

(i) $\mathrm{var}^* (n^{-1/2} \sum_{t=1}^{n} X_t^*) - \mathrm{var} (n^{-1/2} \sum_{t=1}^{n} X_t) = o_P(1)$ $(n \to \infty)$.

(ii) *If, in addition,* $n^{-1/2} \sum_{t=1}^{n} (X_t - \mu_X) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sum_{k=-\infty}^{\infty} R(k))$, *then*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}^* \left[n^{-1/2} \sum_{t=1}^{n} (X_t^* - \bar{X}) \leqslant x \right] - \mathbb{P} \left[n^{-1/2} \sum_{t=1}^{n} (X_t - \mu_X) \leqslant x \right] \right| = o_P(1) \; (n \to \infty).$$

Remark. Assumption A1 only guarantees consistency of second moments; the additional assumption in (ii) is needed for consistency of the distribution function.

We now present a comparison of the sieve bootstrap with the blockwise bootstrap in the case of the mean $\bar{X}_n$. It is shown in Künsch (1989, Theorem 3.1) that

$$n \operatorname{var}^* (\bar{X}_n^*) \approx \sum_{k=-\ell}^{\ell} (1 - |k|/\ell) \hat{R}(k), \tag{3.1}$$

where $\ell = \ell(n) \to \infty$, $\ell(n) = o(n) \; (n \to \infty)$ is the blocklength. More generally it is shown in Bühlmann and Künsch (1994) that a generalization of the blockwise bootstrap, the so-called correlated weights bootstrap, satisfies

$$n \operatorname{var}^* (\bar{X}_n^*) \approx \sum_{k=-\ell}^{\ell} w(|k|/\ell) \hat{R}(k), \tag{3.2}$$

where $w(\cdot)$ is a window that is twice differentiable at zero, with $w'(0) = 0$, $w''(0) \neq 0$. Formulae (3.1) and (3.2) tell us that the blockwise or correlated weights bootstrap variance $n \operatorname{var}^* (\bar{X}_n^*)$ is asymptotically equivalent to a lag-window spectral estimate at zero (multiplied by 2π) with the triangular or a smoother window, respectively. Parzen (1957) has given asymptotic expressions for the mean square error of lag-window spectral estimators (see also Priestley 1981). Thus under suitable conditions we obtain, for the blockwise bootstrap

$$\mathrm{E}[n \operatorname{var}^* (\bar{X}_n^*)] - n \operatorname{var}(\bar{X}_n) \sim -\ell^{-1} \sum_{k=-\infty}^{\infty} |k| R(k),$$

$$\operatorname{var}[n \operatorname{var}^* (\bar{X}_n^*)] \sim \ell n^{-1} 4 \left(\sum_{k=-\infty}^{\infty} R(k) \right)^2 \Big/ 3; \tag{3.3}$$

and for the correlated weights bootstrap

$$\mathrm{E}[n \operatorname{var}^* (\bar{X}_n^*)] - n \operatorname{var}(\bar{X}_n) \sim -\ell^{-2} w''(0) \sum_{k=-\infty}^{\infty} k^2 R(k)/2,$$

$$\operatorname{var}[n \operatorname{var}^* (\bar{X}_n^*)] \sim \ell n^{-1} 2 \int_{\infty}^{\infty} w^2(x) \, \mathrm{d}x \left(\sum_{k=-\infty}^{\infty} R(k) \right)^2. \tag{3.4}$$

By choosing $\ell(n) = \mathrm{const.} n^{1/3}$ in (3.3) or $\ell(n) = \mathrm{const.} n^{2/5}$ in (3.4), we obtain the best order for the error of the bootstrap variance, namely

$$n \operatorname{var}^* (\bar{X}_n^*) - n \operatorname{var}(\bar{X}_n) = O_P(n^{-1/3}) \qquad \text{for the blockwise bootstrap,} \tag{3.5}$$

$$n \operatorname{var}^* (\bar{X}_n^*) - n \operatorname{var}(\bar{X}_n) = O_P(n^{-2/5}) \qquad \text{for the correlated weights bootstrap.} \tag{3.6}$$

In principle, one could obtain better rates for the correlated weights bootstrap under more restrictive conditions on the dependence structure, that is, the smoothness of the spectral density at zero, and by taking a smoother window $w(\cdot)$. However, we have to deal here with an unsolved 'oracle' problem: since we do not know a priori the smoothness of the spectral density we cannot choose the optimal weights for the correlated weights bootstrap. On the other hand, we can show the following result, proved in Section 5.3, for the sieve bootstrap.

Theorem 3.2

(i) *Let assumptions A1 with $s = 4$, A2 with $r \geq 1$ and B with $p(n) = o((n/\log(n))^{1/(2r+2)})$ hold. Moreover, assume that $\sum_{t_1,t_2,t_3} |\mathrm{cum}_4(X_0, X_{t_1}, X_{t_2}, X_{t_3})| < \infty$. Then*

$$n\,\mathrm{var}^*(\bar{X}_n^*) - n\,\mathrm{var}(\bar{X}_n) = O_P((p/n)^{1/2}) + O_P(p^{-r}).$$

(ii) *Let assumptions A1 with $s = 4$, A2 with $r = 1$ and B with $p(n) = o((n/\log(n))^{1/4})$ hold. Denote by*

$$\hat{f}_{\mathrm{AR}(\lambda)} = \frac{(n-p)^{-1} \sum_{t=p+1}^{n} \tilde{\varepsilon}_{t,n}^2}{2\pi \left| \sum_{j=0}^{p} \hat{\phi}_{j,n}\, \mathrm{e}^{-\mathrm{i}j\lambda} \right|^2} \qquad (-\pi \leq \lambda \leq \pi)$$

the autoregressive spectral estimator. Then

$$n\,\mathrm{var}^*(\bar{X}_n^*) - 2\pi\hat{f}_{\mathrm{AR}(0)} = O(n^{-1}) \text{ almost surely.}$$

The sieve bootstrap yields a better variance estimate than the blockwise or correlated weights bootstrap if the coefficients $\{\psi_j\}_{j=0}^{\infty}$ decay sufficiently fast, that is to say, for some weak form of weak dependence. As an example, we consider an ARMA(p, q) model ($p < \infty$, $q < \infty$), where the coefficients $\{\psi_j\}_{j=0}^{\infty}$ decay exponentially. Then for any $0 < \kappa < 1/2$ we can choose $r > 1/(2\kappa) - 1$ and $p(n) = \mathrm{const.}\, n^{1/(2r+2)} \log(n)^{-1/(2r+2)-1}$, which yields for the sieve bootstrap

$$n\,\mathrm{var}^*(\bar{X}_n^*) - n\,\mathrm{var}(\bar{X}_n) = O_P(n^{-1/2+\kappa});$$

compare this with the results for the other bootstrap schemes in (3.5) and (3.6). We mention here that the 'oracle' problem can now be solved (at least in some non-optimal sense). For further discussion, see Section 3.4.

By Theorem 3.2(ii), the sieve bootstrap variance $n\,\mathrm{var}^*(\bar{X}_n^*)$ is asymptotically equivalent to the autoregressive spectral estimate at zero, multiplied by 2π. Under additional conditions the autoregressive spectral estimate has the same asymptotic distribution as the lag-window estimate with a rectangular window (cf. Berk 1974). Our comparison is now completed by interpreting the different bootstrap variances as lag-window estimates at zero with different windows, namely rectangular (sieve bootstrap), triangular (blockwise bootstrap), smooth at

zero with some non-vanishing derivative of order h, $h \geq 2$ (correlated weights bootstrap). This comparison should only be considered as an additional interpretation since the sieve bootstrap should be seen as a sieve rather than a kernel (window) method.

3.3. BOOTSTRAP FOR A CLASS OF NONLINEAR ESTIMATORS

We focus now on estimators that are functions of linear statistics, i.e.,

$$T_n = f\{(n-m+1)^{-1} \sum_{t=1}^{n-m+1} g(X_t, \ldots, X_{t+m-1})\}, \tag{3.7}$$

where $g = (g_1, \ldots, g_q)^\mathrm{T}$ and $f \colon \mathbb{R}^q \to \mathbb{R}^{\tilde{q}}$, $(q, \tilde{q} \geq 1)$. Let $\theta = \mathrm{E}[g(X_t, \ldots, X_{t+m-1})]$. This model class is also considered in Künsch (1989, Example 2.2). It includes versions of the sample autocovariances, autocorrelations, partial autocorrelations and Yule–Walker estimators in autoregressive processes. We usually require that f and g satisfy some smoothness properties and make the following assumption:

Assumption C. $f = (f_1, \ldots, f_{\tilde{q}})^\mathrm{T}$ *has continuous partial derivatives* $\mathbf{y} \mapsto (\partial f_u / \partial x_i)|_{\mathbf{x}=\mathbf{y}}$ $(u = 1, \ldots, \tilde{q};\ i = 1, \ldots, m)$ *for* $\mathbf{y}$ *in a neighbourhood* $U(\theta)$ *of* θ, *and the differentials at* θ, $\mathbf{y} \mapsto Df_u(\theta; \mathbf{y}) = \sum_{i=1}^{m} (\partial f_u / \partial x_i)|_{\mathbf{x}=\theta} y_i$ $(u = 1, \ldots, \tilde{q})$ *do not vanish. The function* g *has continuous partial derivatives of order* h $(h \geq 1)$, $\mathbf{y} \mapsto (\partial^h g_u / \partial x_{i_1} \ldots \partial x_{i_h})|_{\mathbf{x}=\mathbf{y}}$, *which satisfy the Lipschitz condition: for every* $\mathbf{y}$, $\mathbf{z}$,

$$\left| \frac{\partial^h g_u(\mathbf{x})}{\partial x_{i_1} \ldots \partial x_{i_h}} \Big|_{\mathbf{x}=\mathbf{y}} - \frac{\partial^h g_u(\mathbf{x})}{\partial x_{i_1} \ldots \partial x_{i_h}} \Big|_{\mathbf{x}=\mathbf{z}} \right| \leq C_u \|\mathbf{y} - \mathbf{z}\| \qquad u = 1, \ldots, q;\ 1 \leq i_i, \ldots i_h \leq m,$$

where $\|\cdot\|$ *denotes the Euclidean norm, and* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^m$.

Theorem 3.3. *Let assumptions C, A1′ with* $s = 2(h+2)$, *A2 with* $r = 1$ *and B with* $p(n) = o((n/\log(n))^{1/4})$ *hold. Then, writing* $\theta^* = \mathrm{E}^*[g(X_t^*, \ldots, X_{t+m-1}^*)]$, *we have*

$$\sup_{\mathbf{x} \in \mathbb{R}^{\tilde{q}}} |\mathbb{P}^*[n^{1/2}(T_n^* - f(\theta^*)) \leq \mathbf{x}] - \mathbb{P}[n^{1/2}(T_n - f(\theta)) \leq \mathbf{x}]| = o_P(1) \qquad (n \to \infty).$$

The proof is given in Section 5.3.

One possible extension of the model class as given in (3.7) would be

$$T_n = T(P_n), \tag{3.8}$$

where P_n is an empirical distribution of the data and T is a smooth functional. To analyse the validity of the sieve bootstrap for estimators as in (3.8) we need results about the sieve bootstrapped empirical process. This route has been mapped out in the i.i.d. set-up by Bickel and Freedman (1981) and for the blockwise bootstrap by Naik-Nimbalkar and Rajarshi (1994) and Bühlmann (1993; 1994; 1995a). At present no results in this direction exist for the sieve bootstrap. However, for the linear part $(n-m+1)^{-1} \sum_{t=1}^{n-m+1} IF(X_t, \ldots, X_{t+m-1}; P)$

of an estimator in (3.8), Theorem 3.3 usually yields consistency; here $IF(\mathbf{x}, P)$ denotes the influence function of the functional T at the m-dimensional distribution P of $(X_t, \ldots, X_{t+m-1})$ (cf. Hampel *et al.* 1986).

3.4. CHOICE OF THE ORDER p

Our main results in Section 3.2 and 3.3 require some regularity conditions for the order $p = p(n)$ of the approximating autoregressive process which cover quite general situations. This order acts as some kind of smoothing parameter. We briefly address now the question of a 'good' or even some kind of optimal choice of p. Our Theorem 3.2 and its discussion indicate that the choice of p determines the accuracy of the procedure. We suggest two concepts which can be combined for choosing the parameter p.

If the process $\{X_t\}_{t \in \mathbb{Z}}$ is an AR(∞) as in (2.2), the AIC criterion leads to an asymptotically efficient choice $\hat{p}_{\text{AIC}}$ for the optimal order $p_{\text{opt}}(n)$ of some projected AR(∞); cf. Shibata (1980). The related BIC criterion has an optimality property when the order of the true underlying autoregressive process is finite. We prefer AIC, by taking the view that the true model is complex and not of finite dimension. As an example of the behaviour of the AIC, suppose that the autoregressive coefficients in (2.2) decay as

$$\phi_j \sim \text{const.} j^{-v} \text{ as } j \to \infty \qquad (v > 1).$$

Then $\sum_{j=0}^{\infty} j^r |\phi_j| < \infty$ for $r = [v - 1 - \kappa]$ ($\kappa > 0$), where $[x]$ denotes the integer part of $x \in \mathbb{R}$, and equivalently $\sum_{j=0}^{\infty} j^r |\psi_j| < \infty$. On the other hand, Shibata (1980) has shown that $\hat{p}_{\text{AIC}} \sim \text{const.} n^{1/(2v)}$. Therefore $\hat{p}_{\text{AIC}} = o((n/\log(n))^{1/(2r+2)})$ ($r = [v - 1 - \kappa]$), which is the assumption of Theorem 3.2(i). This then explains that $\hat{p}_{\text{AIC}}$ is at least a 'good' order for $n \text{var}^*(\bar{X}_n^*)$ in Theorem 3.2(i): it is such that the error $n \text{var}^*(\bar{X}_n^*) - n \text{var}(\bar{X}_n)$ gets (automatically) smaller with faster decay of the coefficients $\{\psi_j\}_{j=0}^{\infty}$ in assumption A1. In other words, the sieve bootstrap solves the 'oracle' problem. Shibata (1981) shows also the optimality of the AIC for the global relative squared error of the autoregressive spectral estimator $\int_{-\pi}^{\pi} \{(\hat{f}_{\text{AR}}(\lambda) - f(\lambda))/f(\lambda)\}^2 \, \text{d}\lambda$. But by Theorem 3.2(ii) we know that the corresponding AR spectral estimator should be considered locally at zero. At present we have no optimality result for the AIC for $n \text{var}^*(\bar{X}_n^*)$.

The other concept relies on the idea of prewhitening, as a graphical device. For some candidates p, we fit the autoregressive model, obtain the residuals and compute some spectral density estimate based on the residuals. We would choose p such that this estimated spectrum is close to a constant. This method can detect autocorrelation but is not able to distinguish between uncorrelated and independent innovations (compare assumptions A1 and A1').

These two concepts seem to be nicer than the adaptive choice of a blocklength in the blockwise bootstrap (cf. Bühlmann and Künsch 1994). There, the optimal blocklength depends not only on the dependence structure of the observation process but also on the structure of the estimator to be bootstrapped.

4. Simulations

We study and compare the performance of the sieve and blockwise bootstrap. We consider the following models.

(M1) AR(48), $X_t = \sum_{j=1}^{48} \phi_j X_{t-j} + \varepsilon_t$, $\phi_j = (-1)^{j+1} 7.5/(j+1)^3$ ($j = 1, \ldots, 48$), where ε_t i.i.d. $\sim \mathcal{N}(0, 1)$.

(M2) ARMA(1, 1), $X_t = 0.8 X_{t-1} - 0.5\varepsilon_{t-1} + \varepsilon_t$, where ε_t i.i.d. $\sim 0.95\mathcal{N}(0, 1) + 0.05\mathcal{N}(0, 100)$. Models with these ARMA parameters have been considered in Glasbey (1982).

(M3) ARMA(1, 1), $X_t = -0.8 X_{t-1} - 0.5\varepsilon_{t-1} + \varepsilon_t$, where ε_t i.i.d. $\sim 0.95\mathcal{N}(0, 1) + 0.05\mathcal{N}(0, 100)$.

(M4) SETAR(2; 1, 1), $X_t = (1.5 - 0.9 X_{t-1} + \varepsilon_t)1_{[X_{t-1} \leqslant 0]} + (-0.4 - 0.6 X_{t-1} + \varepsilon_t)1_{[X_{t-1} > 0]}$, where ε_t i.i.d. $\sim \mathcal{N}(0, 4)$. This self-exciting threshold autoregressive (SETAR) model is considered in Moeanaddin and Tong (1990).

Models (M1)–(M3) satisfy our assumption A1′ for any $s \in \mathbb{N}$. This is not true for model (M4), which represents a nonlinear process with non-Gaussian marginal distribution (cf. Moeanaddin and Tong 1990). In models (M1) and (M2) the autocorrelation function is positive (in (M1), there are at lag 31 and some bigger lags slightly negative autocorrelations of order 10^{-4}), whereas in (M3) and (M4) the autocorrelation function is 'damped periodic', that is, alternately changing signs and decreasing.

Since the sieve bootstrap relies on a linear approximation we do not want to give advantage to the sieve bootstrap, and always consider here the sample median as the estimator to be bootstrapped.

For the sieve bootstrap we choose the order $p(n)$ of the approximating autoregressive process by minimizing the AIC in a range $0 \leqslant p \leqslant 10 \log_{10}(n)$ (this is the default value in S-PLUS); cf. Shibata (1980). In Table 4.1 we give the summary statistics for the data-driven choices of $\hat{p}_{\mathrm{AIC}}$ based on 100 simulations. For the blockwise bootstrap we estimate the blocklength adaptively as in Bühlmann and Künsch (1994), where we make the additional truncation of very large blocklengths at size $n/2$ (this was used in (M3) and (M4)).

Our results are based on 100 simulations; the number of bootstrap replicates is always 300. We only report the bootstrap estimates for the variance; the estimates for higher cumulants are not very accurate (cf. Bühlmann and Künsch 1994). Let $T_n = \mathrm{med}\{X_1, \ldots, X_n\}$, $\sigma_n^2 = n\,\mathrm{var}(T_n)$, $(\sigma_n^2)^* = n\,\mathrm{var}^*(T_n^*)$, RMSE $= \mathrm{MSE}((\sigma_n^2)^*)/\sigma_n^4$ (relative mean square error). We computed the mean, standard deviation and RMSE as sample moments over the 100 simulations; an estimated standard error of the RMSE is given in parentheses. The true variance σ_n^2 is based on 1000 simulations. The sample sizes are $n = 64$ and $n = 512$.

The results, given in Tables 4.2 and 4.3, can be classified as follows. For processes with positive autocorrelation function, both procedures exhibit roughly the same performance. There might be a small advantage for the sieve bootstrap. In a pairwise comparison of RMSE over the 100 simulations for models (M1) and (M2), the only significant difference at the 5% level was in the case of (M1), $n = 512$, in favour of the sieve bootstrap.

If the autocorrelation function of the model is 'damped periodic', the sieve bootstrap outperforms the blockwise bootstrap. This can be explained by the equivalence of the

Table 4.1. Autoregressive order estimation by AIC

	$E[\hat{p}_{AIC}]$	$SD(\hat{p}_{AIC})$	$\min(\hat{p}_{AIC})$	$\max(\hat{p}_{AIC})$
$n = 64$				
(M1)	1.8	1.0	1	5
(M2)	1.8	1.4	0	9
(M3)	2.8	1.7	1	14
(M3′)	2.8	1.1	1	8
(M4)	1.6	1.2	1	6
(M4′)	1.6	1.2	1	6
$n = 512$				
(M1)	3.6	2.1	2	13
(M2)	4.7	3.7	2	21
(M3)	5.2	2.6	3	21
(M3′)	4.9	2.7	3	20
(M4)	2.6	2.7	1	17
(M4′)	5.0	3.1	1	19

Table 4.2. Sieve bootstrap variance estimation

	σ_n^2	$E[(\sigma_n^2)^*]$	$SD((\sigma_n^2)^*)$	RMSE
$n = 64$				
(M1)	16.4	13.1	8.6	0.31 (0.061)
(M2)	14.1	8.1	8.2	0.52 (0.063)
(M3)	3.1	5.0	5.2	3.09 (0.891)
(M3′)	2.4	2.3	0.7	0.09 (0.013)
(M4)	8.9	7.8	2.0	0.07 (0.008)
(M4′)	7.5	3.6	0.8	0.30 (0.011)
$n = 512$				
(M1)	16.7	16.1	4.3	0.07 (0.009)
(M2)	14.2	12.5	6.5	0.22 (0.046)
(M3)	2.6	2.9	0.7	0.08 (0.020)
(M3′)	2.2	2.2	0.3	0.02 (0.002)
(M4)	9.8	8.0	1.1	0.05 (0.004)
(M4′)	12.5	3.6	0.5	0.51 (0.006)

bootstrap variance to the corresponding spectral estimators at zero. It is known from spectral estimation that lag-window estimation is harder for 'damped periodic' autocorrelation functions, whereas the autoregressive estimate is usually more reliable. In (M3) with $n = 64$ both procedures perform badly. This is mainly due to the influential innovation

Table 4.3. Blockwise bootstrap variance estimation

	σ_n^2	$E[(\sigma_n^2)^*]$	$SD((\sigma_n^2)^*)$	RMSE
$n = 64$				
(M1)	16.4	9.9	6.8	0.33 (0.025)
(M2)	14.1	8.3	9.3	0.61 (0.072)
(M3)	3.1	7.1	11.7	15.52 (6.872)
(M3′)	2.4	2.9	2.3	0.95 (0.198)
(M4)	8.9	11.4	15.8	3.26 (1.600)
(M4′)	7.5	3.5	3.9	0.55 (0.049)
$n = 512$				
(M1)	16.7	14.2	5.1	0.12 (0.015)
(M2)	14.2	11.0	4.3	0.14 (0.013)
(M3)	2.6	3.5	2.4	0.93 (0.265)
(M3′)	2.2	2.9	2.1	0.98 (0.430)
(M4)	9.8	9.9	5.8	0.35 (0.058)
(M4′)	12.5	10.0	8.5	0.50 (0.040)

outliers in the series: one such outlier is followed by approximately 10 contaminated values until the series stabilizes. Therefore we also considered the model

(M3′) ARMA(1, 1), $X_t = -0.8X_{t-1} - 0.5\varepsilon_{t-1} + \varepsilon_t$, where ε_t i.i.d. $\sim t_6$.

Now the sieve bootstrap performs very well (see Table 4.2). The fact that the blockwise bootstrap does not gain performance with the larger sample size (see Table 4.3) is due to one 'extraordinary' occurrence out of the 100 simulations. Without this occurrence the RMSE for the blockwise bootstrap with $n = 512$ decreases to 0.56.

The surprise is the extremely high performance of the sieve bootstrap in (M4), though this model is beyond the theory of linear processes as in assumption A1′. The approximating series does not even asymptotically capture model (M4). However, it seems that the AR approximation is in some sense close enough; the marginal distribution of X_t is not too far away from Gaussianity (cf. Moeanaddin and Tong 1990). The blockwise bootstrap, which does not seem to be restricted to linear processes as in (2.1), yields a poor result.

To see where the sieve bootstrap breaks down, we also considered a similar threshold model as (M4) but now with smaller innovations ε_t:

(M4′) SETAR(2; 1, 1), $X_t = (1.5 - 0.9X_{t-1} + \varepsilon_t)1_{[X_{t-1}\leq 0]} + (-0.4 - 0.6X_{t-1} + \varepsilon_t)1_{[X_{t-1}>0]}$, where ε_t i.i.d. $\sim \mathcal{N}(0, 1)$ (see Moeanaddin and Tong 1990).

The marginal distribution of X_t in (M4′) is now strongly bimodal and much further away from Gaussianity than in (M4). Since the blockwise bootstrap behaves wildly, we use the median and the median absolute deviation (MAD) as estimators based on the 100 simulations for the expectation and standard deviation of the bootstrap variance. The sieve bootstrap has a bias which does not decrease with increasing sample size. This shows that the model (M4′)

cannot be represented as a linear process. As expected, the standard deviation decreases with larger sample size (compare with Table 4.2). For the blockwise bootstrap, the bias is getting smaller with larger sample size (see Table 4.3). This reflects the theory that the blockwise bootstrap is asymptotically working for general mixing processes which can be strongly nonlinear. However, the blockwise bootstrap is, in this situation, no better than the sieve bootstrap. We wondered if we could use a fixed blocklength and improve the blockwise procedure: by trying the blocklengths $\ell = 4, 8, 16, 32, 64$ for the sample size $n = 512$, we could not find any significantly better result.

We draw the final conclusion that in the framework of linear processes, the sieve bootstrap is generally superior to the blockwise bootstrap.

5. Structural properties of the sieve bootstrap and proofs

5.1. AUTOREGRESSIVE APPROXIMATION

We first cite here two results which serve as important tools in our analysis. Using the estimation procedure as in assumption B, we set $\hat{\Phi}_n(z) = \sum_{j=0}^{p(n)} \hat{\phi}_{j,n} z^j$, $\hat{\phi}_{0,n} = 1$ ($z \in \mathbb{C}$, $|z| \leqslant 1$). It is known that $\hat{\Phi}_n(z)$ is invertible for $|z| \leqslant 1$, i.e., $1/\hat{\Phi}_n(z) = \hat{\Psi}_n(z) = \sum_{j=0}^{\infty} \hat{\phi}_{j,n} z^j$ ($|z| \leqslant 1$); cf. Brockwell and Davis (1987, p. 233). Hence, using the definition (2.3) of the sieve bootstrap we write

$$X_t^* - \bar{X} = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \varepsilon_{t-j}^*, \qquad t \in \mathbb{Z}.$$

The next result can be seen as a generalization of Wiener's theorem (cf. Wiener 1933; Zygmund 1959) for the estimation case.

Lemma 5.1. *Let assumptions A1 with $s = 4$, A2 with $r \in \mathbb{N}$ and B with $p(n) = o((n/\log(n))^{1/(2r+2)})$ hold. Then there exists a random variable n_1 such that*

$$\sup_{n \geqslant n_1} \sum_{j=0}^{\infty} j^r |\hat{\psi}_{j,n}| < \infty \text{ almost surely.}$$

Proof. This is essentially Theorem 3.1 in Bühlmann (1995c), which covers slightly more general situations.   □

Lemma 5.2. *Let assumptions A1 with $s = 4$, A2 with $r = 1$ and B with $p(n) = o((n/\log(n))^{1/4})$ hold. Then*

$$\sup_{0 \leqslant j < \infty} |\hat{\psi}_{j,n} - \psi_j| = o(1) \ (n \to \infty) \text{ almost surely.}$$

Proof. This follows from Theorem 3.2 in Bühlmann (1995c).   □

5.2. PROPERTIES OF THE SIEVE BOOTSTRAP SAMPLE

We first present some results about the resampled innovations ε_t^* i.i.d. $\hat{F}_{\varepsilon,n}$. By the definition of $\hat{F}_{\varepsilon,n}$ (see Section 2) we have

$$E^*[\varepsilon_t^*] = 0.$$

The next lemma gives results about higher moments.

Lemma 5.3. *Let assumptions A1 with $s = \max\{2w, 4\}$ $w \in \mathbb{N}$, A2 with $r = 0$ and B with $p(n) = o((n/\log(n))^{1/2})$ hold. Then*

$$E^*[(\varepsilon_t^*)^{2w}] = E[(\varepsilon_t)^{2w}] + o_P(1).$$

Proof. To prove the lemma, we need the equality

$$E^*[(\varepsilon_t^*)^{2w}] = (n-p)^{-1} \sum_{t=p+1}^{n} (\hat{\varepsilon}_{t,n} - \hat{\varepsilon}_n^{(\cdot)})^{2w}, \tag{5.1}$$

where $\hat{\varepsilon}_n^{(\cdot)} = (n-p)^{-1} \sum_{t=p+1}^{n} \hat{\varepsilon}_{t,n}$.

Denote by $\phi_p = (\phi_{1,n}, \ldots, \phi_{p,n})^{\mathsf{T}}$ the solutions of the theoretical Yule–Walker equations $\Gamma_p \phi_p = -\gamma_p$ (compare with assumption B and replace the sample moments by true moments). For ease of notation we set $\hat{\phi}_{j,n} = \phi_{j,n} = 0$ for $j > p$, $\hat{\phi}_{0,n} = \phi_{0,n} = 1$. We write

$$\hat{\varepsilon}_{t,n} = \varepsilon_t - (\bar{X} - \mu_X) \sum_{j=0}^{\infty} \phi_j + Q_{t,n} + R_{t,n}, \tag{5.2}$$

where $Q_{t,n} = \sum_{j=0}^{p} (\hat{\phi}_{j,n} - \phi_{j,n})(X_{t-j} - \bar{X})$, $R_{t,n} = \sum_{j=0}^{\infty} (\phi_{j,n} - \phi_j)(X_{t-j} - \bar{X})$. We first show

$$\hat{\varepsilon}_n^{(\cdot)} = o_P(1). \tag{5.3}$$

We have

$$\hat{\varepsilon}_n^{(\cdot)} = (n-p)^{-1} \left\{ \sum_{t=p+1}^{n} \left(\varepsilon_t - (\bar{X} - \mu_X) \sum_{j=0}^{\infty} \phi_j \right) + \sum_{t=p+1}^{n} Q_{t,n} + \sum_{t=p+1}^{n} R_{t,n} \right\}$$

$$= S_1 + S_2 + S_3. \tag{5.4}$$

By assumptions A1 and A2 we have

$$S_1 = O_P(n^{-1/2}). \tag{5.5}$$

By the Cauchy-Schwarz inequality

$$|S_2| \leqslant \left(\sum_{j=0}^{p} (\hat{\phi}_{j,n} - \phi_{j,n})^2 \right)^{1/2} \left((n-p)^{-1} \sum_{t=p+1}^{n} \sum_{j=0}^{p} (X_{t-j} - \bar{X})^2 \right)^{1/2}. \tag{5.6}$$

In the proof of Theorem 5 in An *et al.* (1982), it is shown under the assumption about $p(n)$ that

$$\sum_{j=0}^{p}(\hat{\phi}_{j,n} - \phi_{j,n})^2 = o((\log(n)/n)^{1/2}) \text{ almost surely.}$$

Thus by (5.6)

$$S_2 = o((\log(n)/n)^{1/4})O_P(p^{1/2}) = o_P(1). \tag{5.7}$$

Furthermore, by the extended Baxter inequality,

$$\sum_{j=0}^{\infty}|\phi_{j,n} - \phi_j| \leq \text{const.} \sum_{j=p+1}^{\infty}|\phi_j|$$

(see the proof of Bühlmann's (1995c) Theorem 3.1). Thus

$$\mathrm{E}|S_3| \leq \mathrm{E}|X_t - \bar{X}|\sum_{j=0}^{p}|\phi_{j,n} - \phi_j| \leq \mathrm{E}|X_t - \bar{X}|\sum_{j=p+1}^{\infty}|\phi_j| = o_P(1). \tag{5.8}$$

By virtue of (5.4), (5.5), (5.7) and (5.8) we have shown (5.3).

Next we show that

$$(n-p)^{-1}\sum_{t=p+1}^{n}(\hat{\varepsilon}_{t,n})^{2w} = \mathrm{E}[(\varepsilon_t)^{2w}] + o_P(1). \tag{5.9}$$

Analogously as for proving (5.7), (5.8) and using the fact that $\mathrm{E}|\varepsilon_t|^{2w} < \infty$, we arrive at

$$(n-p)^{-1}\sum_{t=p+1}^{n}|Q_{t,n}|^{2w} = o_P((p^{1/2}(\log(n)/n)^{1/4})^{2w}) = o_P(1), \tag{5.10}$$

$$(n-p)^{-1}\sum_{t=p+1}^{\infty}|R_{t,n}|^{2w} = O_P\left(\left(\sum_{j=p+1}^{\infty}|\phi_j|\right)^{2w}\right) = o_P(1). \tag{5.11}$$

Now expand the right-hand side of (5.2). Then by (5.10), (5.11), the ergodicity of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and using Hölder's inequality, we can show (5.9). Finally, by a binomial expansion in (5.1) and using (5.3), (5.9) and again Hölder's inequality, we complete the proof. $\square$

Lemma 5.4. *Let assumptions A1 with $s = 4$, A2 with $r = 1$ and B with $p(n) = o((n/\log(n))^{1/2})$ hold. Then*

$$\varepsilon_t^* \xrightarrow{d^*} \varepsilon_t \text{ in probability.}$$

Proof. Let $F_{\varepsilon,n}(x) = (n-p)^{-1}\sum_{t=p+1}^{n}1_{[\varepsilon_t \leq x]}$, $F_\varepsilon(x) = \mathbb{P}[\varepsilon_t \leq x]$ and denote by $d_2(.,.)$ the Mallows metric (cf. Bickel and Freedman 1981). Then it is known that

$$d_2(F_{\varepsilon,n}, F_\varepsilon) = o(1) \text{ almost surely}$$

(see Bickel and Freedman 1981, Lemma 8.4). Thus it remains to prove that

$$d_2(\hat{F}_{\varepsilon,n}, F_{\varepsilon,n}) = o_P(1), \tag{5.12}$$

for $\hat{F}_{\varepsilon,n}$ defined as in Section 2.

Let S be uniformly distributed on $\{p+1, \ldots, n\}$ and let $Z_1 = \varepsilon_S$, $Z_2 = \tilde{\varepsilon}_{S,n}$, where $\tilde{\varepsilon}_{t,n}$ is defined as in Section 2. Then

$$d_2(\hat{F}_{\varepsilon,n}, F_{\varepsilon,n})^2 \leq \mathrm{E}_S|Z_2 - Z_1|^2 = (n-p)^{-1} \sum_{t=p+1}^{n} (\tilde{\varepsilon}_{t,n} - \varepsilon_t)^2$$

$$= (n-p)^{-1} \sum_{t=p+1}^{n} \left(Q_{t,n} + R_{t,n} - (\bar{X} - \mu_X) \sum_{j=0}^{\infty} \phi_j - \hat{\varepsilon}_n^{(\cdot)} \right)^2,$$

where we have used the same notation as in the proof of Lemma 5.3. But the last expression converges to zero in probability by (5.3), (5.10), (5.11) and $\bar{X} - \mu_X = o_P(1)$. Hence (5.12) holds. $\qquad\square$

In the next step we extend Lemma 5.4 for the innovations ε_t^* to the observations X_t^*.

Lemma 5.5. *Let assumptions A1' with $s = 4$, A2 with $r = 0$ and B with $p(n) = o((n/\log(n)^{1/4})$ hold. Then*

$$X_t^* \xrightarrow{\mathrm{d}^*} X_t \text{ in probability.}$$

Proof. Let $M > 0$; we specify its value later. We decompose

$$X_t^* - \bar{X} = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \varepsilon_{t-j}^* = \sum_{j=0}^{M} \psi_j \varepsilon_{t-j}^* + U_{t,n}^* + V_{t,n}^*,$$

where $U_{t,n}^* = \sum_{j=0}^{M} (\hat{\psi}_{j,n} - \psi_j) \varepsilon_{t-j}^*$, $V_{t,n}^* = \sum_{j=M+1}^{\infty} \hat{\psi}_{j,n} \varepsilon_{t-j}^*$.

Let $x \in \mathbb{R}$ be a continuity point of the c.d.f. of $X_t - \mu_X$ and let $\gamma > 0$ be arbitrary. Then, as for proving Slutsky's theorem,

$$\mathbb{P}^*[X_t^* - \bar{X} \leq x] \leq \mathbb{P}^*\left[\sum_{j=0}^{M} \psi_j \varepsilon_{t-j}^* \leq x + \gamma \right] + \mathbb{P}^*[|U_{t,n}^*| > \gamma/2] + \mathbb{P}^*[|V_{t,n}^*| > \gamma/2].$$

Let $\kappa > 0$ be arbitrary. Applying Lemma 5.2 and 5.3 and Lemma 5.1 and 5.3, respectively, we can choose $M = M(\gamma, \kappa)$ such that, for n sufficiently large,

$$\mathbb{P}^*[|U_{t,n}^*| > \gamma/2] \leq \kappa/2 \text{ in probability,}$$

$$\mathbb{P}^*[|V_{t,n}^*| > \gamma/2] \leq \kappa/2 \text{ in probability.}$$

Therefore

$$\mathbb{P}^*[X_t^* - \bar{X} \leq x] \leq \mathbb{P}^*\left[\sum_{j=0}^{M} \psi_j \varepsilon_{t-j}^* \leq x + \gamma\right] + \kappa \text{ in probability,} \qquad (5.13)$$

and analogously

$$\mathbb{P}^*[X_t^* - \bar{X} \leq x] \geq \mathbb{P}^*\left[\sum_{j=0}^{M} \psi_j \varepsilon_{t-j}^* \leq x - \gamma\right] - \kappa \text{ in probability.} \qquad (5.14)$$

By Lemma 5.4, combined with the i.i.d property of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and the conditional i.i.d. property of $\{\varepsilon_t^*\}_{t\in\mathbb{Z}}$, we have, for n sufficiently large,

$$\left|\mathbb{P}^*\left[\sum_{j=0}^{M} \psi_j \varepsilon_{t-j}^* \leq x \pm \gamma\right] - \mathbb{P}\left[\sum_{j=0}^{M} \psi_j \varepsilon_{t-j} \leq x \pm \gamma\right]\right| \leq \kappa \text{ in probability.} \qquad (5.15)$$

Analogously, as before, we can show, for an arbitrary $\zeta > 0$,

$$\mathbb{P}\left[\sum_{j=0}^{M} \psi_j \varepsilon_{t-j} \leq x + \gamma\right] \leq \mathbb{P}[X_t - \mu_X \leq x + \gamma + \zeta] + \kappa, \qquad (5.16)$$

$$\mathbb{P}\left[\sum_{j=0}^{M} \psi_j \varepsilon_{t-j} \leq x - \gamma\right] \geq \mathbb{P}[X_t - \mu_X \leq x - \gamma - \zeta] - \kappa. \qquad (5.17)$$

By (5.13)–(5.17) we have, for n sufficiently large,

$$\mathbb{P}^*[X_t^* - \bar{X} \leq x] \leq \mathbb{P}[X_t - \mu_X \leq x + \gamma + \zeta] + 3\kappa \text{ in probability,}$$
$$\mathbb{P}^*[X_t^* - \bar{X} \leq x] \geq \mathbb{P}[X_t - \mu_X \leq x - \gamma - \zeta] - 3\kappa \text{ in probability.}$$

Since γ, ζ and κ are arbitrary, x is a continuity point of the c.d.f. of $X_t - \mu_X$ and $\bar{X} - \mu_X = o_P(1)$, the proof is complete. $\qquad\square$

Corollary 5.6. *Suppose that the assumptions of Lemma 5.5 hold. Then for every $d \in \mathbb{N}$, $t_1, \ldots, t_d \in \mathbb{Z}$*

$$(X_{t_1}^*, \ldots, X_{t_d}^*) \xrightarrow{\mathrm{d}^*} (X_{t_1}, \ldots, X_{t_d}) \text{ in probability.}$$

Proof. We use the Cramér–Wold device and show that

$$\sum_{i=1}^{d} c_i X_{t_i}^* \xrightarrow{\mathrm{d}^*} \sum_{i=1}^{d} c_i X_{t_i} \text{ in probability} \qquad (c_i \in \mathbb{R}).$$

For this we decompose $X_{t_i}^*$ as in the proof of Lemma 5.5 and proceed along the same lines. $\qquad\square$

5.3. PROOFS OF MAIN RESULTS

Proof of Theorem 3.1. (i) By successively using Lemma 5.3, Lemma 5.1 with $r = 1$ and Lemma 5.2 we have

$$n \operatorname{var}^* (\bar{X}_n^*) = \sum_{k=-n+1}^{n-1} \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \hat{\psi}_{j+|k|,n} (1 - |k|/n) \mathrm{E}^* |\varepsilon_t^*|^2$$

$$= \sum_{k=-n+1}^{n-1} \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \hat{\psi}_{j+|k|,n} (1 - |k|/n) \mathrm{E} |\varepsilon_t|^2 (1 + o_P(1))$$

$$= \left(\sum_{j=0}^{\infty} \hat{\psi}_{j,n} \right)^2 \mathrm{E} |\varepsilon_t|^2 (1 + o_P(1))$$

$$= \left(\sum_{j=0}^{\infty} \psi_j \right)^2 \mathrm{E} |\varepsilon_t|^2 + o_P(1).$$

Since $n \operatorname{var} (\bar{X}_n) = (\sum_{j=0}^{\infty} \psi_j)^2 \mathrm{E} |\varepsilon_t|^2 + o(1)$ we have shown (i).

(ii) We truncate the MA(∞) representation of $X_t^* - \bar{X}$ and set

$$X_{t,M}^* - \bar{X} = \sum_{j=0}^{M} \hat{\psi}_{j,n} \varepsilon_{t-k}^*, \qquad \bar{X}_{n,M}^* = n^{-1} \sum_{t=1}^{n} X_{t,M}^*.$$

Then, as for proving (i),

$$n \operatorname{var}^* (\bar{X}_{n,M}^*) = \left(\sum_{j=0}^{M} \psi_j \right)^2 \mathrm{E} |\varepsilon_t|^2 + o_P(1).$$

We now use a blocking technique with 'small, negligible' and 'large, dominating' blocks. Let

$$A_{n,i} = \sum_{t=(i-1)(a+b)+1}^{ia+(i-1)b} (X_{t,M}^* - \bar{X}), \qquad i = 1, \ldots, [n/(a+b)],$$

$$B_{n,i} = \sum_{t=ia+(i-1)b+1}^{i(a+b)} (X_{t,M}^* - \bar{X}), \qquad i = 1, \ldots, [n/(a+b)],$$

where $a = a(n) \to \infty$, $b = b(n) \to \infty$, $a(n) = o(n)$, $b(n) = o(a(n))$. Let $N(a+b) = n$ and assume without loss of generality that $N \in \mathbb{N}$. Then

$$n^{1/2} (\bar{X}_{n,M}^* - \bar{X}) = n^{-1/2} \sum_{i=1}^{N} A_{n,i} + n^{-1/2} \sum_{i=1}^{N} B_{n,i}.$$

We first show that

$$n^{-1/2} \sum_{i=1}^{N} B_{n,i} = o_{P^*}(1) \text{ in probability.} \tag{5.18}$$

We have that $\mathrm{E}^*[n^{-1/2} \sum_{i=1}^{N} B_{n,i}] = 0$. Since the $X_{t,M}^*$s are M-dependent with respect to $\mathbb{P}^*$, the $\{B_{n,i}\}_{i=1}^{N}$ are (conditionally) independent for n sufficiently large and, hence, as for proving (i)

$$\mathrm{var}^* \left(n^{-1/2} \sum_{i=1}^{N} B_{n,i} \right) = n^{-1} N \, \mathrm{var}^* \left(B_{n,1} \right)$$

$$= n^{-1} Nb \left(\sum_{j=0}^{M} \psi_j \right)^2 \mathrm{E}|\varepsilon_t|^2 + o_P(1) = o_P(1).$$

Therefore (5.18) holds.

We next show that

$$n^{-1/2} \sum_{i=1}^{N} A_{n,i} \overset{\mathrm{d}^*}{\to} \mathscr{N} \left(0, \left(\sum_{j=0}^{M} \psi_j \right)^2 \mathrm{E}|\varepsilon_t|^2 \right) \text{ in probability.} \tag{5.19}$$

Again $\mathrm{E}^*[n^{-1/2} \sum_{i=1}^{N} A_{n,i}] = 0$. As above, and by using $Na \sim n$,

$$\mathrm{var}^* \left(n^{-1/2} \sum_{i=1}^{N} A_{n,i} \right) = n^{-1} N \, \mathrm{var}^* \left(A_{n,1} \right)$$

$$= \left(\sum_{j=0}^{M} \psi_j \right)^2 \mathrm{E}|\varepsilon_t|^2 + o_P(1). \tag{5.20}$$

Then we check Lindeberg's condition

$$N \mathrm{E}^* \left[\frac{A_{n,1}^2}{\sigma_n^2} 1_{[|A_{n,1}/\sigma_n| > \kappa]} \right] = o_P(1) \qquad \text{for } \kappa > 0, \tag{5.21}$$

where $\sigma_n^2 = \mathrm{var}^* \left(\sum_{i=1}^{N} A_{n,i} \right) \sim \mathrm{const.} n$ in probability.

But by reasoning as for Chebyshev's inequality

$$N \mathrm{E}^* \left[\frac{A_{n,1}^2}{\sigma_n^2} 1_{[|A_{n,1}/\sigma_n| > \kappa]} \right] \leq N \kappa^{-2} \sigma_n^{-4} \mathrm{E}^* |A_{n,1}|^4.$$

A direct calculation using Lemmas 5.1 and 5.3 leads then to $\mathrm{E}^*|A_{n,1}|^4 = O_P(a^2)$ and hence

$$N \mathrm{E}^* \left[\frac{A_{n,1}^2}{\sigma_n^2} 1_{[|A_{n,1}/\sigma_n| > \kappa]} \right] = O_P(Nn^{-2}a^2) = O_P(n^{-1}a) = o_P(1),$$

which proves (5.21).

Thus by (5.20), (5.21) and the M-dependence of the $X^*_{t,M}$s we have shown (5.19) and hence by (5.18)

$$n^{1/2}(\bar{X}^*_{n,M} - \bar{X}) \xrightarrow{\mathrm{d}^*} \mathcal{N}\left(0, \left(\sum_{j=0}^{M} \psi_j\right)^2 \mathrm{E}|\varepsilon_t|^2\right) \text{ in probability.} \qquad (5.22)$$

Finally we show that the effect of truncation is negligible. Let

$$n^{1/2}\bar{Y}^*_{n,M} = n^{1/2}(\bar{X}^*_n - \bar{X}^*_{n,M}) = n^{-1/2}\sum_{t=1}^{n}\sum_{j=M+1}^{\infty} \hat{\psi}_{j,n}\varepsilon^*_{t-j}.$$

Then, by Lemmas 5.1 and 5.3,

$$\mathrm{var}^*(n^{1/2}\hat{Y}^*_{n,M}) = \sum_{k=-n+1}^{n-1}\sum_{j=M+1}^{\infty} \hat{\psi}_{j,n}\hat{\psi}_{j+|k|,n}(1-|k|/n)\mathrm{E}^*|\varepsilon^*_t|^2$$

$$\leqslant \text{const.}\sum_{j=M+1}^{\infty}|\hat{\psi}_{j,n}| \leqslant \text{const.}M^{-1}\sum_{j=M+1}^{\infty}j|\hat{\psi}_{j,n}| \text{ in probability.} \qquad (5.23)$$

By (5.22) and (5.23) we complete the proof for (ii) (for this kind of reasoning cf. Anderson 1971, Corollary 7.7.1). □

Proof of Theorem 3.2. Note that $\mathrm{cov}^*(X^*_0, X^*_k) = \hat{R}(k)(\mathrm{E}^*|\varepsilon^*_t|^2)/\hat{\sigma}^2$ for $|k| \leqslant p$, where $\hat{\sigma}^2 = \hat{R}(0) + \hat{\phi}^{\mathrm{T}}_p\hat{\gamma}_p$ is the Yule–Walker estimate of $\sigma^2 = \mathrm{E}|\varepsilon_t|^2$. The difference between $\mathrm{E}^*|\varepsilon^*_t|^2$ and $\hat{\sigma}^2$ is due to initial conditions such as $X_{p-1} = \cdots = X_0 = \bar{X}$. These edge effects are negligible, i.e., $(\mathrm{E}^*|\varepsilon^*_t|^2)/\hat{\sigma}^2 = 1 + O_P(pn^{-1})$. We have, by using Lemmas 5.1 and 5.3,

$$n\,\mathrm{var}^*(\bar{X}^*_n) = \sum_{k=-p}^{p}\hat{R}(k)(1-|k|/n)(1+O_P(pn^{-1})) + 2\sum_{k=p+1}^{n-1}\sum_{j=0}^{\infty}\hat{\psi}_{j,n}\hat{\psi}_{j+k,n}(1-k/n)\mathrm{E}^*|\varepsilon^*_t|^2$$

$$= \sum_{k=-p}^{p}\hat{R}(k)(1-|k|/n)(1+O_P(pn^{-1})) + O_P\left(\sum_{j=p+1}^{\infty}|\hat{\psi}_{j,n}|\right)$$

$$= \sum_{k=-p}^{p}\hat{R}(k)(1-|k|/n)(1+O_P(pn^{-1})) + O_P(p^{-r}).$$

Now (i) follows by Theorem 9.3.4 in Anderson (1971).
For (ii) we observe that

$$|n\,\mathrm{var}^*(\bar{X}^*_n) - 2\pi\hat{f}_{\mathrm{AR}}(0)| \leqslant \left|n^{-1}2\sum_{k=1}^{n-1}k\,\mathrm{cov}^*(X^*_0, X^*_k)\right| + 2\left|\sum_{k=n}^{\infty}\mathrm{cov}^*(X^*_0, X^*_k)\right|$$

$$\le n^{-1} 2 \sum_{j=0}^{\infty} |\hat{\psi}_{j,n}| \sum_{j=0}^{\infty} j |\hat{\psi}_{j,n}| + n^{-1} 2 \sum_{j=0}^{\infty} \sum_{j=n}^{\infty} j |\hat{\psi}_{j,n}|$$

$$= O(n^{-1}) \text{ almost surely,}$$

where the last bound follows from Lemma 5.1.   □

Proof of Theorem 3.3. We only sketch the main steps of the proof; details can be found in Bühlmann (1995b). Write $\mathbf{X}_t = (X_t, \dots, X_{t+m-1})^{\mathsf{T}}$, $\mathbf{X}_t^* = (X_t^*, \dots, X_{t+m-1}^*)^{\mathsf{T}}$. The strategy is to show that

$$(n - m + 1)^{-1/2} \sum_{t=1}^{n-m+1} (g(\mathbf{X}_t^*) - \mathrm{E}^*[g(\mathbf{X}_t^*)]) \xrightarrow{\mathrm{d}^*} \mathcal{N}(\mathbf{0}, \Sigma_{q \times q}) \text{ in probability ,} \qquad (5.24)$$

where $(\Sigma)_{u,v} = \sum_{k=-\infty}^{\infty} \mathrm{cov}(g_u(\mathbf{X}_0), g_v(\mathbf{X}_k))$ is the asymptotic covariance matrix of $(n - m + 1)^{-1/2} \sum_{t=1}^{n-m+1} (g(\mathbf{X}_t) - \mathrm{E}[g(\mathbf{X}_t)])$. Then we will use the Delta technique.

We proceed similarly as for proving Theorem 3.1(ii). We write $X_{t,M}^* = \bar{X} + \sum_{j=0}^{M} \hat{\psi}_{j,n} \varepsilon_{t-j}^*$, $X_{t,M} = \mu_X + \sum_{j=0}^{M} \psi_j \varepsilon_{t-j}$ and define $\mathbf{X}_{t,M}^*$ and $\mathbf{X}_{t,M}$ analogously as for $\mathbf{X}_t^*$ and $\mathbf{X}_t$.

Since the function g is smooth, satisfying a Lipschitz condition for the hth derivative, one can show, by using Taylor expansions and Minkowski's inequality, that

$$\mathrm{E}^* |g_u(\mathbf{X}_{0,M}^*)|^{2+2/(h+1)} = O_P(1), \qquad 1 \le u \le q. \qquad (5.25)$$

This implies uniform integrability of $|g_u(\mathbf{X}_{0,M}^*)|^2$ for $1 \le u \le q$. Then, by the M-dependence of the $X_{t,M}^*$s and by Corollary 5.6,

$$\mathrm{cov}^* \left((n - m + 1)^{-1/2} \sum_{t=1}^{n-m+1} g_u(\mathbf{X}_{t,M}^*), (n - m + 1)^{-1/2} \sum_{t=1}^{n-m+1} g_v(\mathbf{X}_{t,M}^*) \right) = (\Sigma_M)_{u,v} + o_P(1),$$

$$1 \le u, v \le q, \qquad (5.26)$$

where $(\Sigma_M)_{u,v} = \sum_{k=-M-m+1}^{M+m-1} \mathrm{cov}(g_u(\mathbf{X}_0), g_v(\mathbf{X}_k))$.

We now invoke the Cramér–Wold device for showing the convergence of the random vector $(n - m + 1)^{-1/2} \sum_{t=1}^{n-m+1} (g(\mathbf{X}_{t,M}^*) - \mathrm{E}^*[g(\mathbf{X}_{t,M}^*)])$. Let $\ell(\mathbf{x}) = \sum_{u=1}^{q} c_u g_u(\mathbf{x})$, $c_u \in \mathbb{R}$. We use the same blocking technique (and notation) as in the proof of Theorem 3.1(ii) and outline how to show the Lindeberg condition as in (5.21), where now $A_{n,1} = \sum_{t=1}^{a} (\ell(\mathbf{X}_{t,M}^*) - \mathrm{E}^*[\ell(\mathbf{X}_{t,M}^*)])$. We bound

$$N \mathrm{E}^* \left[\frac{A_{n,1}^2}{\sigma_n^2} \mathbf{1}_{[|A_{n,1}/\sigma_n| > \kappa]} \right] \le N \kappa^{-\delta} \sigma_n^{-2-\delta} \mathrm{E}^* |A_{n,1}|^{2+\delta}, \qquad \delta > 0. \qquad (5.27)$$

By using the M-dependence of the $X_{t,M}^*$s with respect to $\mathbb{P}^*$ we obtain

$$\mathrm{E}^* |A_{n,1}|^{2+2/(h+1)} \le \mathrm{const.} a(n)^{1+2/(2h+2)},$$

since one can show $\mathrm{E}^* |\ell(\mathbf{X}_{t,M}^*)|^{2+2/(h+1)} = O_P(1)$ (compare with (5.25)) and by using moment bounds for ϕ-mixing variables. Choose $\delta = 2/(h + 1)$ in (5.27). Therefore by (5.27), the Lindeberg condition holds. Hence, by invoking (5.26),

$$(n - m + 1)^{-1/2} \sum_{t=1}^{n-m+1} (g(\mathbf{X}_{t,M}^*) - \mathrm{E}^*[g(\mathbf{X}_{t,M}^*)]) \xrightarrow{\mathrm{d}^*} \mathcal{N}(0, \Sigma_M) \text{ in probability.} \qquad (5.28)$$

We can show in a straightforward way, by using Taylor expansions for the smooth functions g, Minkowski's inequality and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$, that

$$\lim_{M \to \infty} (\Sigma_M)_{u,v} = (\Sigma)_{u,v}, \qquad u, v = 1, \ldots, m. \qquad (5.29)$$

To show that the effect of truncating the MA(∞) representation of $\mathbf{X}_t^*$ is negligible, we will prove

$$\mathrm{var}^* \left((n - m + 1)^{-1/2} \sum_{t=1}^{n-m+1} (\ell(\mathbf{X}_t^*) - \ell(\mathbf{X}_{t,M}^*)) \right) \leqslant \mathrm{const.} M^{-1} \text{ in probability.} \qquad (5.30)$$

To see this, let $Z_t^* = \ell(\mathbf{X}_t^*) - \ell(\mathbf{X}_{t,M}^*)$. Then

$$\mathrm{var}^* \left((n - m + 1)^{-1/2} \sum_{t=1}^{n-m+1} (\ell(\mathbf{X}_t^*) - \ell(\mathbf{X}_{t,M}^*)) \right) \leqslant \sum_{k=-\infty}^{\infty} |\mathrm{cov}^* (Z_0^*, Z_k^*)|.$$

Let $k > m$ be fixed. Write

$$\tilde{\mathbf{X}}_k^* - \bar{X} = \left(\sum_{j=0}^{k-m} \hat{\psi}_{j,n} \varepsilon_{k-j}^*, \ldots, \sum_{j=0}^{k-m} \hat{\psi}_{j,n} \varepsilon_{k+m-1-j}^* \right)^{\mathrm{T}},$$

$$\tilde{\mathbf{X}}_{k,M}^* - \bar{X} = \left(\sum_{j=0}^{M \wedge (k-m)} \hat{\psi}_{j,n} \varepsilon_{k-j}^*, \ldots, \sum_{j=0}^{M \wedge (k-m)} \hat{\psi}_{j,n} \varepsilon_{k+m-1-j}^* \right)^{\mathrm{T}}.$$

Then

$$Z_k^* = \ell(\tilde{\mathbf{X}}_k^*) - \ell(\tilde{\mathbf{X}}_{k,M}^*) + V_1 - V_2,$$

where $V_1 = \ell(\mathbf{X}_k^*) - \ell(\tilde{\mathbf{X}}_k^*)$, $V_2 = \ell(\mathbf{X}_{k,M}^*) - \ell(\tilde{\mathbf{X}}_{k,M}^*)$. By the independence of $(\ell(\tilde{\mathbf{X}}_k^*) - \ell(\tilde{\mathbf{X}}_{k,M}^*))$ and Z_0^* with respect to $\mathbb{P}^*$, we have

$$|\mathrm{cov}^* (Z_0^*, Z_k^*)| \leqslant |\mathrm{E}^*[(Z_0^* - \mathrm{E}^*[Z_0^*])V_1]| + |\mathrm{E}^*[(Z_0^* - \mathrm{E}^*[Z_0^*])V_2]|$$

$$\leqslant \|Z_0^* - \mathrm{E}^*[Z_0^*]\|_{*2}(\|V_1\|_{*2} + \|V_2\|_{*2}),$$

where $\|\cdot\|_{*p}$ denotes the usual $\mathcal{L}_p$-norm with respect to $\mathbb{P}^*$. By using Taylor expansions for the smooth function $\ell(\cdot)$, Minkowski's and Hölder's inequalities, we obtain

$$\|Z_0\|_{*2} \leqslant \mathrm{const.} \sum_{j=M+1}^{\infty} |\hat{\psi}_{j,n}| \text{ in probability,}$$

$$\|V_1\|_{*2} \leqslant \mathrm{const.} \sum_{j=k-m+1}^{\infty} |\hat{\psi}_{j,n}| \text{ in probability,}$$

$$\|V_2\|_{*2} \leqslant \mathrm{const.} \sum_{j=(M \wedge (k-m))+1}^{M} |\hat{\psi}_{j,n}| \text{ in probability,}$$

For $0 \leqslant k \leqslant m$ the bound for $\|Z_0^*\|_{*2}$ applies. Hence

$$\sum_{k=-\infty}^{\infty} |\text{Cov}^*(Z_0^*, Z_k^*)| \leqslant \text{const.} \sum_{j=M+1}^{\infty} |\hat{\psi}_{j,n}| \sum_{j=0}^{\infty} j|\hat{\psi}_{j,n}|$$

$$\leqslant \text{const.} M^{-1} \text{ in probability,}$$

where the last bound follows by Lemma 5.1 with $r = 1$. This proves (5.30) and therefore, by (5.28), (5.29) and by applying Corollary 7.7.1. in Anderson (1971), we have shown (5.24).

Finally, we use the Delta technique. Similarly as for (5.26), we can show $\|\theta^* - \theta\| = o_P(1)$. Using this and the continuous differentiability of f, we can show, along the same lines as in the proof of Theorem A in Serfling (1980, p. 122), that $n^{1/2}(T_n^* - f(\theta^*))$ has the same asymptotic normal distribution as $n^{1/2}(T_n - f(\theta))$. This completes the proof.   □

Acknowledgements

References

An, H.-Z., Chen, Z.-G. and Hannan, E.J. (1982) Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.*, **10**, 926–936. (Correction: **11**, 1018.)

Anderson, T.W. (1971) *The Statistical Analysis of Time Series*. New York: Wiley.

Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Ann. Statist.*, **2**, 489–502.

Bickel, P.J. and Freedman, D.A. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, 1196–1217.

Bose, A. (1988) Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.*, **16**, 1709–1722.

Brockwell, P.J. and Davis, R.A. (1987) *Time Series: Theory and Methods*. New York: Springer-Verlag.

Bühlmann, P. (1993) The blockwise bootstrap in time series and empirical processes. Doctoral thesis, ETH Zürich.

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

Bühlmann, P. (1995a) The blockwise bootstrap for general empirical processes of stationary sequences. *Stochastic Process. Appl.*, **58**, 247–265.

Bühlmann, P. (1995b) Sieve bootstrap for time series. Technical Report No. 431, Dept. of Statistics, University of California, Berkeley.

Bühlmann, P. (1995c) Moving-average representation for autoregressive approximations. *Stochastic Process. Appl.* **60**, 331–342.

Bühlmann, P. and Künsch, H.R. (1994) Block length selection in the bootstrap for time series. Research Report No. 72. Seminar für Statistik, ETH Zürich.

Bühlmann, P. and Künsch, H.R. (1995) The blockwise bootstrap for general parameters of a stationary time series. *Scand. J. Statist.*, **22**, 35–54.

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

Efron, B. and Tibshirani, R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.*, **1**, 54–77.

Franke, J. and Kreiss, J.-P. (1992) Bootstrapping stationary autoregressive moving average models. *J. Time Ser. Anal.*, **13**, 297–317.

Freedman, D.A. (1984) On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.*, **12**, 827–842.

Geman, S. and Hwang, C.-R. (1982) Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.*, **10**, 401–414.

Glasbey, C.A. (1982) A generalization of partial autocorrelations useful in identifying ARMA models. *Technometrics*, **24**, 223–228.

Grenander, U. (1981) *Abstract Inference*. New York: Wiley.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986) *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley.

Hannan, E.J. (1987) Rational transfer function approximation. *Statist. Sci.*, **5**, 105–138.

Janas, D. (1992) Bootstrap procedures for time series. Doctoral thesis, Universität Heidelberg.

Kreiss, J.-P. (1988) Asymptotic statistical inference for a class of stochastic processes. *Habilitationsschrift*, Universität Hamburg.

Kreiss, J.-P. (1992) Bootstrap procedures for AR(∞)-processes. In K.H. Jöckel, G. Rothe and W. Sender (eds), *Bootstrapping and Related Techniques*. Lecture Notes in Economics and Mathematical Systems 376. Heidelberg: Springer.

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

Liu, R. and Singh, K. (1992) Moving blocks jackknife and bootstrap capture weak dependence. In R. LePage and L. Billard (eds), *Exploring the Limits of Bootstrap*. New York: Wiley.

Moeanaddin, R. and Tong, H. (1990) Numerical evaluation of distributions in nonlinear auto-regression. *J. Time Ser. Anal.*, **11**, 33–48.

Naik-Nimbalkar, U.V. and Rajarshi, M.B. (1994) Validity of block-wise bootstrap for empirical processes with stationary observations. *Ann. Statist.*, **22**, 980–994.

Parzen, E. (1957) On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.*, **28**, 329–348.

Politis, D.N. and Romano, J.P. (1992) A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.*, **20**, 1985–2007.

Politis, D.N. and Romano, J.P. (1993) Nonparametric resampling for homogeneous strong mixing random fields. *J. Multivariate Anal.*, **47**, 301–328.

Politis, D.N. and Romano, J.P. (1994) The stationary bootstrap. *J. Amer. Statist. Assoc.*, **89**, 1303–1313.

Priestley, M.B. (1981) *Spectral Analysis and Time Series*, Vol. 1. New York: Academic Press.

Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Shao, Q.-M. and Yu, H. (1993) Bootstrapping the sample means for stationary mixing sequences. *Stochastic Process. Appl.*, **48**, 175–190.

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

Shibata, R. (1981) An optimal autoregressive spectral estimate. *Ann. Statist.*, **9**, 300–306.

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

Wiener, N. (1933) *The Fourier Integral and Certain of its Applications*. Cambridge: Cambridge University Press.

Zygmund, A. (1959) *Trigonometric Series*, Vol. 1. Cambridge: Cambridge University Press.

Received April 1995; Revised March 1996