Sieve Bootstrap for Time Series

By

Peter Bühlmann

Technical Report No. 431
April 1995

Department of Statistics
University of California
Berkeley, California

Sieve Bootstrap for Time Series

Peter Bühlmann *
University of California
Berkeley

April 1995

Abstract

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.

We show its consistency for a class of nonlinear estimators and compare the procedure with the blockwise bootstrap, which has been proposed by Künsch (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 illustrating the advantages and disadvantages of the sieve compared to the blockwise bootstrap.

Key words and phrases. AIC, AR(∞), ARMA, autoregressive approximation, autoregressive spectrum, blockwise bootstrap, linear process, resampling, stationary sequence, threshold model.

Short title: Bootstrap for Time Series

*Supported by the Swiss National Science Foundation.

1 Introduction

The bootstrap proposed by Efron (1979) has become a powerful nonparametric method for estimating the distribution of a statistical procedure. 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. A model based approach, which resamples from approximately i.i.d. residuals, cf. Freedman (1984), Efron and Tibshirani (1986), Bose (1988), Franke and Kreiss (1992). Also Tsay (1992) uses this approach for diagnostics in the time series context. Obviously, these procedures are sensible to model misspecification and the attractive nonparametric feature 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 involved blocklength grows slowly with the sample size. By construction we call this procedure blockwise bootstrap (sometimes it is also called moving blocks bootstrap). After 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 (1994a, 1994b) and Bühlmann (1993, 1994a, 1994b). Generally, this blockwise bootstrap works satisfactory and enjoys the property of being robust against misspecified models. However, the resampled series exhibits artifacts 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 (1981), Geman and Hwang (1982)) and explains the name of our procedure. 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 the blockwise bootstrap, this resampling procedure is again nonparametric and moreover its bootstrap sample is (conditionally) stationary and does not exhibit additional artifacts of the dependence structure as above.

In an unpublished paper Kreiss (1988) also discusses the bootstrap for AR(∞) models. But his approach only covers linear processes

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

where $\{\psi_j\}_{j=0}^{\infty}$ decays exponentially and ε_t is an i.i.d. sequence with $\mathbf{E}[\varepsilon_t] = 0$. 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 to interpret the approximating autoregressive model as a sieve for a broader subclass of stationary processes.

Kreiss (1988) 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 periodogram values which yields a consistent procedure for smooth functions of the periodogram. This approach can be interpreted as approximating the modulus $|\Phi(e^{-i\lambda})|$ $(0 \le \lambda \le \pi)$, 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 $\Phi(z)$ $(|z| \le 1)$ 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 of a certain polynomial speed. For practice we argue here, that by fitting an autoregressive model first, using e.g. the Akaike information criterion (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 there a comparison with the blockwise bootstrap which indicates that the sieve bootstrap works better for very weak dependent processes, i.e., for processes with fast decaying coefficients $\{\psi_j\}_{j=0}^{\infty}$ in (1.1). In section 4 we present results of a simulation study. We compare the performance of the blockwise and the sieve bootstrap. To explore some limits of the sieve bootstrap we consider there 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 works also 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 $\mathbf{E}[X_t] = \mu_X$. 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

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

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of uncorrelated variables with $\mathbf{E}[\varepsilon_t] = 0$ 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

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

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 well known as a successful technique in different situations (cf. Berk (1974), An et al. (1982), Hannan (1987)).

We give now 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 $p = p(n) \to \infty$ $(n \to \infty)$ with $p(n) = o(n)$. 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

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

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

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

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

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

Then we can resample for any $t \in \mathbb{Z}$

$$\varepsilon_t^*\ i.i.d. \sim \hat{F}_{\varepsilon,n}$$

and define $\{X_t^*\}_{t\in\mathbb{Z}}$ by the recursion

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

In practice we construct a sieve bootstrap sample $X_1^*, \ldots X_n^*$ in the following way: choose starting values, e.g., equal to zero, 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 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 supply 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, i.e.,

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

This bootstrap construction exhibits some features which are different from Künsch's (1989) blockwise bootstrap. It yields again a (conditionally) stationary bootstrap sample and does not exhibit artifacts in the dependence structure like 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, there is no need of 'pre-vectorizing' 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. $F_n^{(m)}$,

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

Denote by $Y_t = (X_t, \ldots, X_{t+m-1})^T$ $(t = 1, \ldots, n-m+1)$ 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_t's, thus being a 'block of blocks' bootstrap scheme. For different dimensions m one has to use different vectorized observations. Our sieve bootstrap has the advantage of avoiding the construction of vectorized observations and enjoys the properties of a plug-in rule. Finally, our sieve bootstrap seems to be more promising for unequally spaced data or series with many missing values.

3 Main results

3.1 Assumptions

We consider now 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). Denote by

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

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

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

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

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 $\mathcal{F}_t = \sigma(\{\varepsilon_s; s \leq t\})$ the σ-field generated by $\{\varepsilon_s\}_{s=-\infty}^{t}$. Our main assumptions for the model are the following.

(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 $\mathbf{E}[\varepsilon_t|\mathcal{F}_{t-1}] \equiv 0$, $\mathbf{E}[\varepsilon_t^2|\mathcal{F}_{t-1}] \equiv \sigma^2 < \infty$, $\mathbf{E}|\varepsilon_t|^s < \infty$ for some $s \geq 4$.

(A2) $\Psi(z)$ is bounded away from zero for $|z| \leq 1$, $\sum_{j=0}^{\infty} j^r |\psi_j| < \infty$ 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 statistics based on model (A1) with non-independent variables $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. The arithmetic mean as a linear statistics is an exception in this respect. We therefore sometimes strengthen (A1) to

(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 $\mathbb{E}[\varepsilon_t] = 0,\ \mathbb{E}|\varepsilon_t|^s < \infty$ for some $s \geq 4$.

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 (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 $|z| \leq 1$ and $\sum_{j=0}^{\infty} j^r |\phi_j| < \infty$. 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 assumptions:

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

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

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

In the sequel we always denote by $R(j) = Cov(X_0, X_j)$.

3.2 Bootstrapping the mean

Our first result shows the 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 (A1) with non-independent innovations.

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

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

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

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

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

The proof is given in section 5.3.

We present now 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

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

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 (1994b) that a generalization of the blockwise bootstrap, the so-called correlated weights bootstrap, satisfies

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

where $w(.)$ is a window that is twice differentiable at zero with $w'(0) = 0,\ w''(0) \neq 0$. Formula (3.1) and (3.2) tell that the blockwise or correlated weights bootstrap variance $nVar^*(\bar{X}_n^*)$ is asymptotically equivalent to a lag-window spectral estimate at zero (multiplied by 2π) with the triangular- or a more smooth 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 get

for the blockwise bootstrap:

$$\begin{aligned} \mathbf{E}[nVar^*(\bar{X}_n^*)] - nVar(\bar{X}_n) &\sim -\ell^{-1} \sum_{k=-\infty}^{\infty} |k|R(k), \\ Var(nVar^*(\bar{X}_n^*)) &\sim \ell n^{-1} 4 (\sum_{k=-\infty}^{\infty} R(k))^2/3, \end{aligned} \tag{3.3}$$

for the correlated weights bootstrap:

$$\begin{aligned} \mathbf{E}[nVar^*(\bar{X}_n^*)] - nVar(\bar{X}_n) &\sim -\ell^{-2} w''(0) \sum_{k=-\infty}^{\infty} k^2 R(k)/2, \\ Var(nVar^*(\bar{X}_n^*)) &\sim \ell n^{-1} 2 \int_{\infty}^{\infty} w^2(x)dx\ (\sum_{k=-\infty}^{\infty} R(k))^2. \end{aligned} \tag{3.4}$$

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

$$\begin{aligned} &MSE(nVar^*(\bar{X}_n^*)) \sim const.n^{-2/3} \text{ for the blockwise bootstrap,} \\ &MSE(nVar^*(\bar{X}_n^*)) \sim const.n^{-4/5} \text{ for the correlated weights bootstrap.} \end{aligned}$$

In principle, one could obtain better rates for the correlated weights bootstrap under more restrictive conditions on the dependence structure, i.e., the smoothness of the spectral density at zero, and by taking a smoother window $w(.)$. 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 for the sieve bootstrap.

Theorem 3.2

(i) Assume that (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 $\sum_{t_1,t_2,t_3} |cum_4(X_0, X_{t_1}, X_{t_2}, X_{t_3})| < \infty$. Then

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

(ii) Assume that(A1) with $s = 4$, (A2) with $r = 1$ and (B) with $p(n) = o((n/log(n))^{1/4})$. Denote by

$$\hat{f}_{AR}(\lambda) = \frac{(n-p)^{-1}\sum_{t=p+1}^{n}\tilde{\varepsilon}_{t,n}^2}{2\pi|\sum_{j=0}^{p}\hat{\phi}_{j,n}e^{-ij\lambda}|^2}\ (-\pi \le \lambda \le \pi)$$

the autoregressive spectral estimator. Then

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

The proof is given in section 5.3

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, i.e., for some kind of 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 $p(n) = const.(n/log(n))^{1/(2r+2)}$ which yields for the sieve bootstrap

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

compare this with the results for the other bootstrap schemes above. 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 $nVar^*(\bar{X}_n^*)$ 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, $h \ge 2$ (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.,

$$T_n = f((n-m+1)^{-1}\sum_{t=1}^{n-m+1} g(X_t,\ldots,X_{t+m-1})), \tag{3.5}$$

where $g = (g_1,\ldots,g_q)^T$ and $f : \mathbb{R}^q \to \mathbb{R}^{\tilde{q}}$, $(q,\tilde{q} \ge 1)$. Denote by $\theta = \mathbb{E}[g(X_t,\ldots,X_{t+m-1})]$. This model class is also considered in Künsch (1989, Example 2.2). As examples 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 assumptions:

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

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

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

Theorem 3.3 *Assume that (C), (A1') with $s = 2(h+2)$, (A2) with $r = 1$ and (B) with $p(n) = o((n/log(n))^{1/4})$ hold. Then, denoting by $\theta^* = \mathbb{E}^*[g(X_t^*, \ldots, X_{t+m-1}^*)]$,*

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

The proof is given in section 5.3.

One possible extension of the model class as given in (3.5) would be

$$T_n = T(P_n), \tag{3.6}$$

where P_n is an empirical distribution of the data and T is a smooth functional.
To analyze the validity of the sieve bootstrap for estimators as in (3.6) we need results about the sieve bootstrapped empirical process. This route has been given 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, 1994a, 1994b). 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 (3.6), 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) (not being of finite order), then the AIC criterion leads to an asymptotically efficient choice $\hat{p}_{AIC}$ for the optimal order $p_{opt}(n)$ of some projected AR(∞), cf. Shibata (1980). As an example, suppose that the autoregressive coefficients in (2.2) decay like

$$\phi_j \sim const. j^{-v}, \; (j \to \infty).$$

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 $\hat{p}_{AIC} \sim const.n^{1/(2v)}$. Therefore $\hat{p}_{AIC} = o((n/log(n))^{1/(2r+2)}$ $(r = [v - 1 - \kappa])$ which is the assumption of Theorem 3.2(i). This then explains that $\hat{p}_{AIC}$ is at least a 'good' order for $nVar^*(\bar{X}_n^*)$ in Theorem 3.2(i): it is such that the error $nVar^*(\bar{X}_n^*) - nVar(\bar{X}_n)$ gets (automatically) smaller with faster decay of the coefficients $\{\psi_j\}_{j=0}^{\infty}$ in (A1). In other words, the sieve bootstrap solves the oracle problem. Shibata (1981) shows also the optimality of AIC for the global relative squared error of the autoregressive spectral estimator $\int_{-\pi}^{\pi}((\hat{f}_{AR}(\lambda) - f(\lambda))/f(\lambda))^2 d\lambda$. 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 of AIC for $nVar^*(\bar{X}_n^*)$.

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

(M4) SETAR(2;1,1), $X_t = (1.5 - 0.9X_{t-1} + \varepsilon_t)1_{[X_{t-1} \le 0]} + (-0.4 - 0.6X_{t-1} + \varepsilon_t)1_{[X_{t-1} > 0]}$, where ε_t *i.i.d.* $\sim \mathcal{N}(0, 4)$. This 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 the order 10^{-4}), whereas in (M3) and (M4) the autocorrelation function is 'damped-periodic', i.e., alternatively 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 consider here always 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 Akaike information criterion (AIC) in a range $0 \le p \le$

$10log_{10}(n)$ (this is the default value in S-Plus), cf. Shibata (1980). For the blockwise bootstrap we estimate the blocklength adaptively as in Bühlmann and Künsch (1994b) 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 (1994b)). Denote by $T_n = med\{X_1, \ldots, X_n\}$, $\sigma_n^2 = nVar(T_n)$, $(\sigma_n^2)^* = nVar^*(T_n^*)$, $RMSE = MSE((\sigma_n^2)^*)/\sigma_n^4$ (relative mean square error). We computed $\mathbb{E}$, $S.D.$ and $RMSE$ as sample moments over the 100 simulations, an estimated standard deviation of $RMSE$ is given in parentheses. The true variance σ_n^2 is based on 1000 simulations. The sample sizes are $n = 64$ and $n = 512$.

sieve bootstrap, $n = 64$	σ_n^2	$\mathbb{E}[(\sigma_n^2)^*]$	$S.D.((\sigma_n^2)^*)$	$RMSE$
(M1)	16.40	13.16	8.57	0.312 (0.061)
(M2)	14.11	8.14	8.18	0.515 (0.063)
(M3)	3.13	5.03	5.17	3.089 (0.891)
(M4)	8.85	7.81	1.99	0.065 (0.008)
blockw. bootstrap, $n = 64$				
(M1)	16.40	9.86	6.80	0.331 (0.025)
(M2)	14.11	8.26	9.34	0.610 (0.072)
(M3)	3.13	7.11	11.69	15.520 (6.872)
(M4)	8.85	11.44	15.75	3.255 (1.600)
sieve bootstrap, $n = 512$				
(M1)	16.68	16.10	4.25	0.066 (0.009)
(M2)	14.22	12.45	6.49	0.223 (0.046)
(M3)	2.62	2.87	0.68	0.076 (0.020)
(M4)	9.79	8.03	1.09	0.045 (0.004)
blockw. bootstrap, $n = 512$				
(M1)	16.68	14.24	5.10	0.115 (0.015)
(M2)	14.22	11.03	4.28	0.141 (0.013)
(M3)	2.62	3.47	2.37	0.928 (0.265)
(M4)	9.79	9.93	5.75	0.345 (0.058)

The results can be classified as follows. For processes with positive autocorrelation function, both procedures have roughly about the same performance. There might be a small advantage for the sieve bootstrap for small sample sizes.

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

(M3') ARMA(1,1), $X_t = -0.8X_{t-1} - 0.5\varepsilon_{t-1} + \varepsilon_t$, where ε_t $i.i.d. \sim t_6$.

sieve bootstrap, $n = 64$	σ_n^2	$\mathbb{E}[(\sigma_n^2)^*]$	$S.D.((\sigma_n^2)^*)$	$RMSE$
(M3')	2.39	2.28	0.72	0.093 (0.013)
blockw. bootstrap, $n = 64$				
(M3')	2.39	2.93	2.26	0.945 (0.198)
sieve bootstrap, $n = 512$				
(M3')	2.24	2.19	0.28	0.016 (0.002)
blockw. bootstrap, $n = 512$				
(M3')	2.24	2.94	2.10	0.977 (0.430)

Now the sieve bootstrap performs very well. The fact that the blockwise bootstrap does not gain performance with the larger sample size is due to one 'extraordinary' realization out of the 100 simulations. Without this realization the $RMSE$ for the blockwise bootstrap with $n = 512$ decreases to 0.559.

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 (A1'). The approximating series does not even asymptotically capture the 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 brakes down we finally considered a similar threshold model as (M4) but now with smaller innovations ε_t

(M4') SETAR(2;1,1), $X_t = (1.5 - 0.9X_{t-1} + \varepsilon_t)1_{[X_{t-1} \le 0]} + (-0.4 - 0.6X_{t-1} + \varepsilon_t)1_{[X_{t-1} > 0]}$, 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 MAD as estimators based on the 100 simulations for the expectation and standard deviation.

sieve bootstrap, $n = 64$	σ_n^2	$\mathbb{E}[(\sigma_n^2)^*]$	$S.D.((\sigma_n^2)^*)$	$RMSE$
(M4')	7.47	3.45	0.76	0.300 (0.011)
blockw. bootstrap, $n = 64$				
(M4')	7.47	3.49	3.88	0.553 (0.049)
sieve bootstrap, $n = 512$				
(M4')	12.47	3.60	0.54	0.508 (0.006)
blockw. bootstrap, $n = 512$				
(M4')	12.47	9.96	8.45	0.499 (0.040)

The sieve bootstrap has a bias which does not decrease with increasing sample size. This exhibits the fact that the model (M4') cannot be represented as a linear process. As expected the standard deviation decreases with larger sample size.

For the blockwise bootstrap, the relative bias is getting smaller with larger sample size. This reflects the theory that the blockwise is asymptotically working for general mixing processes which can be strongly nonlinear. However, the blockwise bootstrap is 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 over 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 (B), we set $\hat{\Phi}_n(z) = \sum_{j=0}^{p(n)} \hat{\phi}_{j,n} z^j$, $\hat{\phi}_{0,n} = 1$ ($z \in \mathbb{C}$, $|z| \le 1$). It is known that $\hat{\Phi}_n(z)$ is invertible for $|z| \le 1$, i.e., $1/\hat{\Phi}_n(z) = \hat{\Psi}_n(z) = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} z^j$ ($|z| \le 1$) (cf. Brockwell and Davis (1987), p. 233). Hence using the definition (2.3) of the sieve bootstrap we represent

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

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

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

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

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

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

Prof: See Bühlmann (1995), formula (3.2) and (3.3). □

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

$$\mathbb{E}^*[\varepsilon_t^*] = 0. \tag{5.1}$$

The next Lemma gives results about higher moments.

Lemma 5.3 *Assume that (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*

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

Proof: It is

$$\mathbb{E}^*[(\varepsilon_t^*)^{2w}] = (n-p)^{-1} \sum_{t=p+1}^{n} (\hat{\varepsilon}_{t,n} - \hat{\varepsilon}_n^{(\cdot)})^{2w}, \tag{5.2}$$

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

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

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

$$\begin{aligned} \hat{\varepsilon}_n^{(\cdot)} &= (n-p)^{-1} (\sum_{t=p+1}^{n} (\varepsilon_t - (\bar{X} - \mu_X) \sum_{j=0}^{\infty} \phi_j) + \sum_{t=p+1}^{n} Q_{t,n} + \sum_{t=p+1}^{n} R_{t,n}) \\ &= I + II + III, \end{aligned} \tag{5.4}$$

where $Q_{t,n} = \sum_{j=0}^{p} (\hat{\phi}_{j,n} - \phi_{j,n})(X_{t-j} - \bar{X})$, $R_{t,n} = \sum_{j=0}^{\infty} (\phi_{j,n} - \phi_j)(X_{t-j} - \bar{X})$. By (A1) and (A2) we have

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

By the Cauchy-Schwarz inequality

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

In An et al. (1982), proof of Theorem 5, 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)

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

Furthermore by the extended Baxter inequality,

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

(see Bühlmann (1995), proof of Theorem 3.1). Thus

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

By (5.4), (5.5), (5.7) and (5.8) we have shown (5.3).
Next we show that

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

Write

$$\hat{\varepsilon}_{t,n} = \varepsilon_t - (\bar{X} - \mu_X) \sum_{j=0}^{\infty} \phi_j + Q_{t,n} + R_{t,n} \tag{5.10}$$

(for the notation see (5.4)). Analogously as for proving (5.7), (5.8) and using that $\mathbb{E}|\varepsilon_t|^{2w} < \infty$, we arrive at

$$(n-p)^{-1} \sum_{t=p+1}^{n} |Q_{t,n}|^{2w} = o_P((p^{1/2}(log(n)/n)^{1/4})^{2w}) = o_P(1), \tag{5.11}$$

$$(n-p)^{-1} \sum_{t=p+1}^{\infty} |R_{t,n}|^{2w} = O_P((\sum_{j=p+1}^{\infty} |\phi_j|)^{2w}) = o_P(1). \tag{5.12}$$

Now expand the right hand side of (5.10). Then by (5.11), (5.12), the ergodicity of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and by using Hölder's inequality we can show (5.9). Finally by a binomial expansion in (5.2) and using (5.3), (5.9) and again Hölder's inequality we complete the proof.

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

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

Proof: Denote by $F_{\varepsilon,n}(x) = (n-p)^{-1} \sum_{t=p+1}^{n} 1_{[\varepsilon_t \leq x]}$, $F_\varepsilon(x) = \mathbb{P}[\varepsilon_t \leq x]$ and 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

$$d_2(\hat{F}_{\varepsilon,n}, F_{\varepsilon,n}) = o_P(1),\ \ \hat{F}_{\varepsilon,n} \text{ defined as in section 2.} \tag{5.13}$$

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

$$\begin{aligned} & d_2(\hat{F}_{\varepsilon,n}, F_{\varepsilon,n})^2 \leq \mathbb{E}|Z_2 - Z_1|^2 \\ = \ & (n-p)^{-1} \sum_{t=p+1}^{n} (\tilde{\varepsilon}_{t,n} - \varepsilon_t)^2 = (n-p)^{-1} \sum_{t=p+1}^{n} (Q_{t,n} + R_{t,n} - (\bar{X} - \mu_X) \sum_{j=0}^{\infty} \phi_j - \hat{\varepsilon}_n^{(.)})^2, \end{aligned}$$

where we used the notation as in the proof of Lemma 5.3. But the last expression converges to zero in probability by (5.3), (5.11), (5.12) and $(\bar{X} - \mu_X) = o_P(1)$. Hence (5.13) holds.□

In the next step we extend Lemma 5.4 for the innovations to the observations.

Lemma 5.5 *Assume that (A1') with $s = 4$, (A2) with $r = 0$ and (B) with $p(n) = o((log(n)/n)^{1/2})$ hold. Then*

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

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

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

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 and let $\gamma > 0$ be arbitrary. Then as for proving Slutsky's Theorem

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

By Markov's inequality

$$\mathbb{P}^*[|U_{t,n}^*| > \gamma/2] \leq 2 \sum_{j=0}^{M} |\hat{\psi}_{j,n} - \psi_j| \mathbb{E}^* |\varepsilon_t^*| / \gamma$$
$$\mathbb{P}^*[|V_{t,n}^*| > \gamma/2] \leq 2 \sum_{j=M+1}^{\infty} |\hat{\psi}_{j,n}| \mathbb{E}^* |\varepsilon_t^*| / \gamma,$$

By Lemma 5.1, $\sum_{j=M+1}^{\infty} |\hat{\psi}_{j,n}| \leq M^{-1} \sum_{j=M+1}^{\infty} j |\hat{\psi}_{j,n}| \leq const. M^{-1}$ almost surely. Moreover by Lemma 5.2, for any finite M and any $\xi > 0$ there exists an $n_0 = n_0(M, \xi)$ such that for $n \geq n_0$

$$\sum_{j=0}^{M} |\hat{\psi}_{j,n} - \psi_j| \leq \xi.$$

Let $\kappa > 0$ be arbitrary. We bound $\mathbb{E}^* |\varepsilon_t^*| \leq (\mathbb{E}^* |\varepsilon_t^*|^2)^{1/2}$ and use Lemma 5.3. Thus we can choose $M = M(\gamma, \kappa)$ such that for n sufficiently large

$$\mathbb{P}^*[|U_{t,n}^*| > \gamma/2] \leq \kappa/2 \text{ in probability,}$$
$$\mathbb{P}^*[|V_{t,n}^*| > \gamma/2] \leq \kappa/2 \text{ in probability.} \tag{5.14}$$

Therefore

$$\mathbb{P}^*[X_t^* \leq x] \leq \mathbb{P}^*[\sum_{j=0}^{M} \psi_j \varepsilon_{t-j}^* \leq x + \gamma] + \kappa \text{ in probability} \tag{5.15}$$

and analogously

$$\mathbb{P}^*[X_t^* \leq x] \geq \mathbb{P}^*[\sum_{j=0}^{M} \psi_j \varepsilon_{t-j}^* \leq x - \gamma] - \kappa \text{ in probability.} \tag{5.16}$$

By the Cramér-Wold device and 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

$$|\mathbb{P}^*[\sum_{j=0}^{M}\psi_j\varepsilon_{t-j}^* \leq x+\gamma] - \mathbb{P}[\sum_{j=0}^{M}\psi_j\varepsilon_{t-j} \leq x+\gamma]| \leq \kappa \text{ in probability.} \tag{5.17}$$

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

$$\mathbb{P}[\sum_{j=0}^{M}\psi_j\varepsilon_{t-j} \leq x+\gamma] \leq \mathbb{P}[X_t \leq x+\gamma+\zeta] + \kappa, \tag{5.18}$$

$$\mathbb{P}[\sum_{j=0}^{M}\psi_j\varepsilon_{t-j} \leq x-\gamma] \geq \mathbb{P}[X_t \leq x-\gamma-\zeta] - \kappa. \tag{5.19}$$

(We use here that $\sum_{j=M+1}^{\infty}|\psi_j| = o(1)$ $(M\to\infty)$; if the M chosen in (5.14) is too small we increase M which would not affect the results above).
By (5.15)-(5.20) we have for n sufficiently large

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

Since γ, ζ and κ are arbitrary and x is a continuity point of the c.d.f. of X_t we complete the proof. □

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

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

Proof: We use the Cramér-Wold device and show that

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

For this we decompose $X_{t_i}^*$ as in the proof of Lemma 5.5 and follow its lines. □

5.3 Proofs of main results

Proof of Theorem 3.1

(i) By using successively Lemma 5.3, Lemma 5.1 with $r = 1$ and Lemma 5.2 we have

$$\begin{aligned} nVar^*(\bar{X}_n^*) &= \sum_{k=-n+1}^{n-1}\sum_{j=0}^{\infty}\hat{\psi}_{j,n}\hat{\psi}_{j+|k|,n}(1-|k|/n)\mathbf{E}^*|\varepsilon_t^*|^2 \\ &= \sum_{k=-n+1}^{n-1}\sum_{j=0}^{\infty}\hat{\psi}_{j,n}\hat{\psi}_{j+|k|,n}(1-|k|/n)\mathbf{E}|\varepsilon_t|^2(1+o_P(1)) \\ &= (\sum_{j=0}^{\infty}\hat{\psi}_{j,n})^2\mathbf{E}|\varepsilon_t|^2(1+o_P(1)) \\ &= (\sum_{j=0}^{\infty}\psi_{j,n})^2\mathbf{E}|\varepsilon_t|^2 + o_P(1). \end{aligned}$$

Since $nVar(\bar{X}_n)) = (\sum_{j=0}^{\infty} \psi_{j,n})^2 \mathbb{E}|\varepsilon_t|^2$ we have shown (i).

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

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

Then as for proving (i)

$$nVar^*(\bar{X}_{n,M}^*) = (\sum_{j=0}^{M} \psi_j)^2 \mathbb{E}|\varepsilon_t|^2 + o_P(1).$$

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

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

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

$$n^{-1/2} \sum_{i=1}^{N} B_{n,i} = o_{P^*}(1) \text{ in probability.} \tag{5.20}$$

It is $\mathbb{E}^*[n^{-1/2} \sum_{i=1}^{N} B_{n,i}] = 0$. Since the $X_{t,M}^*$'s are M-dependent with respect to $\mathbb{P}^*$, $\{B_{n,i}\}_{i=1}^{N}$ are (conditionally) independent for n sufficiently large and hence, as for proving (i)

$$Var^*(n^{-1/2} \sum_{i=1}^{N} B_{n,i}) = n^{-1} N Var^*(B_{n,1})$$
$$= \quad n^{-1} N b (\sum_{j=0}^{M} \psi_j)^2 \mathbb{E}|\varepsilon_t|^2 + o_P(1) = o_P(1).$$

Therefore (5.20) holds.

We next show that

$$n^{-1/2} \sum_{i=1}^{N} A_{n,i} \xrightarrow{d^*} \mathcal{N}(0, (\sum_{j=0}^{M} \psi_j)^2 \mathbb{E}|\varepsilon_t|^2) \text{ in probability.} \tag{5.21}$$

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

$$Var^*(n^{-1/2}\sum_{i=1}^N A_{n,i}) = n^{-1}NaVar^*(A_{n,1})$$
$$= (\sum_{j=0}^M \psi_j)^2 \mathbb{E}|\varepsilon_t|^2 + o_P(1). \quad (5.22)$$

Then we check Lindeberg's condition

$$N\mathbb{E}^*[\frac{A_{n,1}^2}{\sigma_n^2} 1_{[|A_{n,1}/\sigma_n|>\kappa]}] = o_P(1) \text{ for } \kappa > 0, \quad (5.23)$$

where $\sigma_n^2 = Var^*(\sum_{i=1}^N A_{n,i}) \sim const.n$ in probability.
But by reasoning as for Chebychev's inequality

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

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

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

which proves (5.23).
Thus by (5.22), (5.23) and the M-dependence of the $X_{t,M}^*$'s we have shown (5.21) and hence by (5.20)

$$n^{1/2}(\bar{X}_{n,M}^* - \bar{X}) \xrightarrow{d^*} \mathcal{N}(0, (\sum_{j=0}^M \psi_j)^2 \mathbb{E}|\varepsilon_t|^2) \text{ in probability.} \quad (5.24)$$

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

$$Var^*(n^{1/2}\bar{Y}_{n,M}^*) = \sum_{k=-n+1}^{n-1} \sum_{j=M+1}^\infty \hat{\psi}_{j,n}\hat{\psi}_{j+|k|,n}(1-|k|/n)\mathbb{E}^*|\varepsilon_t^*|^2$$
$$\leq const. \sum_{j=M+1}^\infty |\hat{\psi}_{j,n}| \leq const.M^{-1}\sum_{j=M+1}^\infty j|\hat{\psi}_{j,n}| \text{ in probability.} \quad (5.25)$$

By (5.24) and (5.25) 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 $Cov^*(X_0^*, X_k^*) = \hat{R}(k)\frac{\mathbb{E}^*|\varepsilon_t^*|^2}{\hat{\sigma}}$ for $|k| \leq p$, where $\hat{\sigma}^2 = \hat{R}(0) + \hat{\phi}_p^T\hat{\gamma}_p$ is the Yule-Walker estimate of $\sigma^2 = \mathbb{E}|\varepsilon_t|^2$. The difference between $\mathbb{E}^*|\varepsilon_t^*|^2$ and $\hat{\sigma}^2$ is due to initial

conditions like $X_{p-1} = \ldots = X_0 = \bar{X}$. These edge effects are negligible, i.e., $\frac{\mathbb{E}^*|\varepsilon_t^*|^2}{\hat{\sigma}^2} = 1 + O_P(pn^{-1})$. We have by using Lemma 5.1 and Lemma 5.3

$$\begin{aligned}
& nVar^*(\bar{X}_n^*) \\
= \; & \sum_{k=-p}^{p} \hat{R}(k)(1-|k|/n)(1+O_P(pn^{-1}) + 2\sum_{k=p+1}^{n-1}\sum_{j=0}^{\infty} \hat{\psi}_{j,n}\hat{\psi}_{j+k,n}(1-k/n)\mathbb{E}^*|\varepsilon_t^*|^2 \\
= \; & \sum_{k=-p}^{p} \hat{R}(k)(1-|k|/n)(1+O_P(pn^{-1}) + O_P(\sum_{j=p+1}^{\infty}|\hat{\psi}_{j,n}|) \\
= \; & \sum_{k=-p}^{p} \hat{R}(k)(1-|k|/n)(1+O_P(pn^{-1}) + o_P(p^{-r}).
\end{aligned}$$

Now (i) follows by Theorem 9.3.4 in Anderson (1971).

For (ii) we observe that

$$\begin{aligned}
& |nVar^*(\bar{X}_n^*) - 2\pi \hat{f}_{AR}(0)| \\
\leq \; & |n^{-1}2\sum_{k=1}^{n-1} kCov^*(X_0^*, X_k^*)| + 2|\sum_{k=n}^{\infty} Cov^*(X_0^*, X_k^*)(1-k/n)| \\
\leq \; & n^{-1}2\sum_{j=0}^{\infty}|\hat{\psi}_{j,n}|\sum_{j=0}^{\infty} j|\hat{\psi}_{j,n}| + n^{-1}2\sum_{j=0}^{\infty}\sum_{j=n}^{\infty} j|\hat{\psi}_{j,n}| = O(n^{-1}) \text{ almost surely,}
\end{aligned}$$

where the last bound follows from Lemma 5.1. □

Proof of Theorem 3.3
In the sequel we denote by $\mathbf{X}_t = (X_t, \ldots, X_{t+m-1})^T$, $\mathbf{X}_t^* = (X_t^*, \ldots, X_{t+m-1}^*)^T$. The strategy is to show that

$$(n-m+1)^{-1/2}\sum_{t=1}^{n-m+1}(g(\mathbf{X}_t^*) - \mathbb{E}^*[g(\mathbf{X}_t^*)]) \xrightarrow{d^*} \mathcal{N}(\mathbf{0}, \Sigma_{m\times m}) \text{ in probability,} \quad (5.26)$$

where $(\Sigma)_{u,v} = \sum_{k=-\infty}^{\infty} Cov(g_u(\mathbf{X}_0), g_v(\mathbf{X}_k))$ is the asymptotic covariance matrix of $(n-m+1)^{-1/2}\sum_{t=1}^{n-m+1}(g(\mathbf{X}_t) - \mathbb{E}[g(\mathbf{X}_t)])$. We leave the proof away that, under the conditions of Theorem 3.3,

$$(n-m+1)^{-1/2}\sum_{t=1}^{n-m}(g(\mathbf{X}_t) - \mathbb{E}[g(\mathbf{X}_t)]) \xrightarrow{d} \mathcal{N}(0, \Sigma_{m\times m}).$$

This can be shown in a similar (easier) way as the following proof for the bootstrap quantities.
Then we will use the Delta technique.

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

In a first step we show that

$$\begin{aligned} & Cov^*((n-m+1)^{-1/2}\sum_{t=1}^{n-m+1} g_u(\mathbf{X}^*_{t,M}), (n-m+1)^{-1/2}\sum_{t=1}^{n-m+1} g_v(\mathbf{X}^*_{t,M})) \\ = & \ (\Sigma_M)_{u,v} + o_P(1)\ (1 \le u, v \le q), \end{aligned} \tag{5.27}$$

where $(\Sigma_M)_{u,v} = \sum_{k=-M-m+1}^{M+m-1} Cov(g_u(\mathbf{X}_0), g_v(\mathbf{X}_k))$.
We immediately have by the M-dependence of the $X^*_{t,M}$'s

$$\begin{aligned} & Cov^*((n-m+1)^{-1/2}\sum_{t=1}^{n-m+1} g_u(\mathbf{X}^*_{t,M}), (n-m+1)^{-1/2}\sum_{t=1}^{n-m+1} g_v(\mathbf{X}^*_{t,M})) \\ = & \sum_{k=-M-m+1}^{M+m-1} Cov^*(g_u(\mathbf{X}^*_{0,M}), g_v(\mathbf{X}^*_{k,M}))(1-|k|/(n-m+1)). \end{aligned} \tag{5.28}$$

We truncate the g_u's,

$$\tilde{g}_u(\mathbf{x}) = g_u(\mathbf{x})1_{[|g_u(\mathbf{x})| \le K]} + K sign(g_u(\mathbf{x})),\ \ K > 0.$$

Since $\tilde{g}_u$ is continuous and bounded we get by Corollary 5.6, which also holds for the truncated versions $X^*_{t,M}$ and $X_{t,M}$,

$$\begin{aligned} & \sum_{k=-M-m+1}^{M+m-1} Cov^*(\tilde{g}_u(\mathbf{X}^*_{0,M}), \tilde{g}_v(\mathbf{X}^*_{k,M})) \\ = & \sum_{k=-M-m+1}^{M+m-1} Cov(\tilde{g}_u(\mathbf{X}_{0,M}), \tilde{g}_v(\mathbf{X}_{k,M})) + o_P(1). \end{aligned} \tag{5.29}$$

Now we discuss that the the effect of truncating the g_u's is negligible. We have by Hölder's inequality

$$\begin{aligned} & \mathbb{E}^*|g_u(\mathbf{X}^*_{0,M})1_{[|g_u(\mathbf{X}^*_{0,M})|>K]}|^2 \\ \le & \ (\mathbb{E}^*|g_u(\mathbf{X}^*_{0,M})|^{2(h+2)/(h+1)})^{(h+1)/(h+2)}(\mathbb{P}^*[|g_u(\mathbf{X}^*_{0,M})| > K])^{1/(h+2)} = O_P(1)K^{-2/(h+1)} \end{aligned}$$

(we show $\mathbb{E}^*|g_u(\mathbf{X}^*_{0,M})|^{2(h+2)/(h+1)} = O_P(1)$ in the same way as later (5.34)).
Let $\kappa > 0$ be arbitrary. Hence we can choose $K = K(\kappa, M(\kappa)) = K(\kappa)$ such that for n sufficiently large

$$\begin{aligned} & |Cov^*(g_u(\mathbf{X}^*_{0,M}), g_v(\mathbf{X}^*_{k,M})) - Cov^*(\tilde{g}_u(\mathbf{X}^*_{0,M}), \tilde{g}_v(\mathbf{X}^*_{k,M}))| \\ \le & \ \kappa/(M+m-1) \text{ in probability.} \end{aligned} \tag{5.30}$$

Analogously we can do the same for the original observations

$$|Cov(g_u(\mathbf{X}_{0,M}), g_v(\mathbf{X}_{k,M})) - Cov(\tilde{g}_u(\mathbf{X}_{0,M}), \tilde{g}_v(\mathbf{X}_{k,M}))| \le \kappa/(M+m-1). \tag{5.31}$$

Putting together (5.28)-(5.31) and using that $\kappa > 0$ is arbitrary we have shown (5.27).

Now we invoke the Cramér-Wold device for showing the convergence of the random vector $(n-m+1)^{-1/2}\sum_{t=1}^{n-m+1}(g(\mathbf{X}^*_{t,M}) - \mathbb{E}^*[g(\mathbf{X}^*_{t,M})])$. Denote by $\ell(\mathbf{x}) = \sum_{u=1}^{q} c_u g_u(\mathbf{x}),\ c_u \in$

$\mathbb{R}$. Then we use the same blocking technique as in the proof of Theorem 3.1(ii) (for notation see also there). We will show now the Lindeberg condition as in (5.23), where now $A_{n,1} = \sum_{t=1}^{a}(\ell(\mathbf{X}_{t,M}^*) - \mathbf{E}^*[\ell(\mathbf{X}_{t,M}^*)])$. We bound

$$N\mathbf{E}^*[\frac{A_{n,1}^2}{\sigma_n^2} 1_{[|A_{n,1}/\sigma_n|>\kappa]}] \leq N\kappa^{-\delta}\sigma_n^{-2-\delta}\mathbf{E}^*|A_{n,1}|^{2+\delta} \ (\delta > 0). \tag{5.32}$$

By using the M-dependence of the $X_{t,M}^*$'s with respect to $\mathbb{P}^*$ we get

$$\mathbf{E}^*|A_{n,1}|^{2+\delta} \leq const.a(n)^{1+\delta/2} \ (\delta > 0), \text{ if } \mathbf{E}^*|\ell(\mathbf{X}_{t,M}^*)|^{2+2\delta} = O_P(1), \tag{5.33}$$

cf. Yokoyama (1980). Choose $\delta = 1/(h+1)$. We will show now that

$$\mathbf{E}^*|\ell(\mathbf{X}_{t,M}^*)|^{2+2/(h+1)} = O_P(1). \tag{5.34}$$

We use the following notation for the linear operator for the i-th derivative:

$$D^i\ell(\mathbf{y})(\mathbf{z}) = \sum_{d_1,\ldots,d_i=1}^{m} \frac{\partial^i \ell}{\partial x_{d_1}\ldots\partial x_{d_i}}(\mathbf{y}) z_{d_1}\ldots z_{d_i}, \ \frac{\partial^i \ell}{\partial x_{d_1}\ldots\partial x_{d_i}}(\mathbf{y}) = \frac{\partial^i \ell(\mathbf{x})}{\partial x_{d_1}\ldots\partial x_{d_i}}|_{\mathbf{x}=\mathbf{y}},$$

where $\mathbf{x} = (x_1,\ldots,x_m)^T$, $\mathbf{y} = (y_1,\ldots,y_m)^T$, $\mathbf{z} = (z_1,\ldots,z_m)^T$.
Denote by $\|.\|_{*p}$ the usual $\mathcal{L}_p$-norm with respect to $\mathbb{P}^*$. By a Taylor expansion

$$\ell(\mathbf{X}_{t,M}^*) = \sum_{i=0}^{h-1} D^i\ell(\bar{X})(\mathbf{X}_{t,M}^* - \bar{X}) + D^h\ell(\tau)(\mathbf{X}_{t,M}^* - \bar{X}),$$

where $\|\tau - \bar{X}\| \leq \|\mathbf{X}_{t,M}^* - \bar{X}\|$, $\|.\|$ denoting the Euclidean norm in $\mathbb{R}^m$.
Hence by Minkowski's inequality

$$\begin{aligned} &\|\ell(\mathbf{X}_{t,M}^*)\|_{*2+2/(h+1)} \\ \leq\ & \sum_{i=0}^{h-1} \|D^i\ell(\bar{X})(\mathbf{X}_{t,M}^* - \bar{X})\|_{*2+2/(h+1)} + \|D^h\ell(\tau)(\mathbf{X}_{t,M}^* - \bar{X})\|_{*2+1/(h+1)}. \end{aligned} \tag{5.35}$$

Now by the definitions of $D^i\ell(.)$ and $X_{t,M}^*$ and by using Lemma 5.3

$$\begin{aligned} &\|D^i\ell(\bar{X})(\mathbf{X}_{t,M}^* - \bar{X})\|_{*2+2/(h+1)} \\ \leq\ & const. \sum_{j=0}^{M} |\hat{\psi}_{j,n}|\mathbf{E}^*|\varepsilon_t^*|^{(2+2/(h+1))i} = O_P(1) \ (0 \leq i \leq h-1). \end{aligned} \tag{5.36}$$

On the other hand, by using the Lipschitz property of $\frac{\partial^h \ell(.)}{\partial x_{i_1}\ldots x_{i_h}}$

$$|D^h\ell(\tau) - D^h\ell(0)| \leq const.\|\tau - \bar{X}\| \leq const.\|X_{t,M}^* - \bar{X}\|,$$

and therefore, similarly as above,

$$\|D^h\ell(\tau)(\mathbf{X}_{t,M}^* - \bar{X})\|_{*2+2/(h+1)} \leq const. \sum_{j=0}^{M} |\hat{\psi}_{j,n}|\mathbf{E}^*|\varepsilon_t^*|^{(2+2/(h+1))(h+1)} = O_P(1). \tag{5.37}$$

By (5.35)-(5.37) we have shown (5.34) and hence (5.33). Therefore by (5.32), the Lindeberg condition holds. Hence by invoking (5.27)

$$(n-m+1)^{-1/2}\sum_{t=1}^{n-m+1}(g(\mathbf{X}_{t,M}^*)-\mathbb{E}^*[g(\mathbf{X}_{t,M}^*)])\stackrel{d^*}{\longrightarrow}\mathcal{N}(0,\Sigma_M)\text{ in probability.}\quad(5.38)$$

Next we show that

$$\lim_{M\to\infty}(\Sigma_M)_{u,v}=(\Sigma)_{u,v},\ \ u,v=1,\ldots,m.\quad(5.39)$$

First we prove in a straightforward way by using Taylor expansions (in a similar spirit as for proving (5.34)) and by using the smoothness of $g(.)$, $\mathbb{E}|\varepsilon_t|^{2(h+2)}<\infty$ and the assumption $\sum_{j=0}^{\infty}j|\psi_j|<\infty$

$$\begin{aligned}&|\sum_{k=-M-m+1}^{M+m-1}Cov(g_u(\mathbf{X}_{0,M}),g_v(\mathbf{X}_{k,M}))-\sum_{k=-M-m+1}^{M+m-1}Cov(g_u(\mathbf{X}_0),g_v(\mathbf{X}_k))|\\ \le\ \ &const.\sum_{j=M}^{\infty}j|\psi_j|.\end{aligned}$$

Furthermore we can show, again using the smoothness of $g(.)$,

$$\sum_{k=M+m}^{\infty}|Cov(g_u(\mathbf{X}_0),g_v(\mathbf{X}_k))|\to 0\ (M\to\infty).$$

(This would follow directly if additionally $\{X_t\}_{t\in\mathbb{Z}}$ would be strong-mixing, but we can prove it by using similar techniques as for showing (5.40) below).
The last two estimates then prove that (5.39) holds.

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

$$Var^*((n-m+1)^{-1/2}\sum_{t=1}^{n-m+1}(\ell(\mathbf{X}_t^*)-\ell(\mathbf{X}_{t,M}^*))\le const.M^{-1}\text{ in probability.}\quad(5.40)$$

Denote by $Z_t^*=\ell(\mathbf{X}_t^*)-\ell(\mathbf{X}_{t,M}^*)$. Then

$$Var^*((n-m+1)^{-1/2}\sum_{t=1}^{n-m+1}(\ell(\mathbf{X}_t^*)-\ell(\mathbf{X}_{t,M}^*))\le\sum_{k=-\infty}^{\infty}|Cov^*(Z_0^*,Z_k^*)|.$$

Let $k>m$ be fixed. Denote by

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

Then

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

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

$$\begin{aligned} |Cov^*(Z_0^*, Z_k^*)| &\leq |\mathbb{E}^*[(Z_0^* - \mathbb{E}^*[Z_0^*])I]| + |\mathbb{E}^*[(Z_0^* - \mathbb{E}^*[Z_0^*])II]| \\ &\leq \|Z_0^* - \mathbb{E}^*[Z_0^*]\|_{*2}(\|I\|_{*2} + \|II\|_{*2}). \end{aligned}$$

By using Taylor expansions (τ_0, τ_k denote the appropriate mid-points), Hölder's and Minkowski's inequality we obtain the following bounds (similarly as in (5.36) and (5.37))

$$\begin{aligned} \|Z_0\|_{*2} &\leq \sum_{i=1}^{h-1} \|D^i \ell(\mathbf{X}_{t,M}^*)\|_{*2(h+1)/(h-i+1)} \|(\mathbf{X}_t^* - \mathbf{X}_{t,M}^*)_1\|_{*2(h+1)}^i \\ &+ \|D^h(\tau_0)\|_{*2(h+1)} \|(\mathbf{X}_t^* - \mathbf{X}_{t,M}^*)_1\|_{*2(h+1)}^h \\ &\leq const. \sum_{j=M+1}^{\infty} |\hat{\psi}_{j,n}| \text{ in probability,} \end{aligned}$$

$$\begin{aligned} \|I\|_{*2} &\leq \sum_{i=1}^{h-1} \|D^i \ell(\tilde{\mathbf{X}}_{t,M}^*)\|_{*2(h+1)/(h-i+1)} \|(\mathbf{X}_t^* - \tilde{\mathbf{X}}_t^*)_1\|_{*2(h+1)}^i \\ &+ \|D^h(\tau_k)\|_{*2(h+1)} \|(\mathbf{X}_t^* - \tilde{\mathbf{X}}_t^*)_1\|_{*2(h+1)}^h \\ &\leq const. \sum_{j=k-m+1}^{\infty} |\hat{\psi}_{j,n}| \text{ in probability,} \end{aligned}$$

and similarly

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

where the right hand side disappears for $k \geq M + m$.
For $0 \leq k \leq m$ the bound for $\|Z_0^*\|_{*2}$ applies. Hence

$$\begin{aligned} \sum_{k=-\infty}^{\infty} |Cov^*(Z_0^*, Z_k^*)| &\leq const. \sum_{j=M+1}^{\infty} |\hat{\psi}_{j,n}| \sum_{j=0}^{\infty} j |\hat{\psi}_{j,n}| \\ &\leq const. M^{-1} \text{ in probability,} \end{aligned}$$

where the last bound follows by Lemma 5.1 with $r = 1$. This proves (5.40) and therefore by (5.38), (5.39) and by applying Corollary 7.7.1 in Anderson (1971) we have shown (5.26).

Finally we use the Delta technique. Similarly as for (5.27) 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 Serfling (1980) (proof of Theorem A, p.122) that $n^{1/2}(T_n^* - f(\theta^*))$ has the same asymptotic distribution as $n^{1/2}(T_n - f(\theta))$. This completes the proof. □

Acknowledgments: I would like to thank Peter Bickel for suggesting the method of sieves in connection with the bootstrap and for his helpful comments. I also greatly appreciate interesting comments from John Rice and many discussions with Ritei Shibata.

References

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

[2] Anderson,T.W. (1971). The Statistical Analysis of Time Series. Wiley, New York.

[3] Berk,K.N. (1974). Consistent autoregressive spectral estimates. Ann. Statist. **2** 489-502.

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

[5] Bose,A. (1988). Edgeworth correction by bootstrap in autoregressions. Ann. Statist. **16** 1709-1722.

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

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

[8] Bühlmann,P. (1994b). The blockwise bootstrap for general empirical processes of stationary sequences. To appear in Stoch. Proc. and Appl.

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

[10] Bühlmann,P. and Künsch,H.R. (1994a). The blockwise bootstrap for general parameters of a stationary time series. To appear in Scand. J. Statist. **22**.

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

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

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

[14] Efron,B. and Tibshirani,R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Science **1** 54-77.

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

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

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

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

[19] Grenander,U. (1981). Abstract Inference. Wiley, New York.

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

[21] Hannan,E.J. (1987). Rational transfer function approximation. Stat. Science **5** 105-138.

[22] Janas,D. (1992). Bootstrap Procedures for Time Series. PhD thesis, Universität Heidelberg.

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

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

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

[26] Moeanaddin,R. and Tong,H. (1990). Numerical evaluation of distributions in nonlinear autoregression. J. Time Ser. Anal. **11**, 33-48.

[27] Naik-Nimbalkar,U.V. and Rajarshi,M.B. (1992). Validity of block-wise bootstrap for empirical processes with stationary observations. Ann. Statist. **22** 980-994.

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

[29] 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.

[30] Politis,D.N. and Romano,J.P. (1993). Nonparametric resampling for homogeneous strong mixing random fields. J. Mult. Anal. **47** 301-328.

[31] Politis,D.N. and Romano,J.P. (1994). The Stationary Bootstrap. J. Amer. Statist. Assoc. **89** 1303-1313.

[32] Priestley,M.B. (1981). Spectral Analysis and Time Series **1**. Academic, New York.

[33] Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics. Wiley, New York.

[34] Shao,Q.-M. and Yu,H. (1993). Bootstrapping the sample means for stationary mixing sequences. Stoch. Proc. and Appl. **48** 175-190.

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

[36] Shibata,R. (1981). An optimal autoregressive spectral estimate. Ann. Statist. **9** 300-306.

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

[38] Wiener,N. (1933). The Fourier Integral and Certain of its Applications. Cambridge Univ. Press, Cambridge.

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

[40] Zygmund,A. (1959). Trigonometric Series, Vol.1. Cambridge Univ. Press, Cambridge.

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