

Bootstrap confidence regions computed from autoregressions of arbitrary order

Edwin Choi and Peter Hall

Australian National University, Canberra, Australia

[Received August 1998. Revised October 1999]

Summary. Given a linear time series, e.g. an autoregression of infinite order, we may construct a finite order approximation and use that as the basis for bootstrap confidence regions. The sieve or autoregressive bootstrap, as this method is often called, is generally seen as a competitor with the better-understood block bootstrap approach. However, in the present paper we argue that, for linear time series, the sieve bootstrap has significantly better performance than blocking methods and offers a wider range of opportunities. In particular, since it does not corrupt second-order properties then it may be used in a double-bootstrap form, with the second bootstrap application being employed to calibrate a basic percentile method confidence interval. This approach confers second-order accuracy without the need to estimate variance. That offers substantial benefits, since variances of statistics based on time series can be difficult to estimate reliably, and—partly because of the relatively small amount of information contained in a dependent process—are notorious for causing problems when used to Studentize. Other advantages of the sieve bootstrap include considerably greater robustness against variations in the choice of the tuning parameter, here equal to the autoregressive order, and the fact that, in contradistinction to the case of the block bootstrap, the percentile t version of the sieve bootstrap may be based on the ‘raw’ estimator of standard error. In the process of establishing these properties we show that the sieve bootstrap is second order correct.

Keywords: Akaike’s information criterion; Autoregressive bootstrap; Block bootstrap; Calibration; Coverage accuracy; Double bootstrap; Edgeworth expansion; Moving average; Percentile method; Percentile t method; Second-order accuracy; Sieve bootstrap; Studentization

1. Introduction

The autoregressive or sieve bootstrap for linear time series has been shown to produce consistent estimation of variances and distributions of linear statistics; see for example Kreiss (1992) and particularly Bühlmann (1995a, 1997). Nevertheless, in general terms the sieve bootstrap is at present viewed as little more than a competitor to the block bootstrap. For example, although the block bootstrap is known to be second order accurate (Götze and Künsch, 1996; Lahiri, 1996), the sieve bootstrap may also be shown to have this property, and so the two methods appear to compete on approximately equal terms.

In the present paper we argue, however, that for linear time series the sieve bootstrap has substantial advantages over blocking methods, to such an extent that block-based methods are not really competitive. The advantages include the following:

- (a) the percentile form of the sieve bootstrap readily admits calibration, giving high orders of accuracy without requiring variance estimation,

Address for correspondence: Peter Hall, Centre for Mathematics and Its Applications, Australian National University, Canberra, ACT 0200, Australia.
E-mail: halpstat@pretty.anu.edu.au

- (b) when using the percentile t form of the sieve bootstrap, the estimator of the standard error does not need to be adjusted to capture second-order effects and
- (c) the choice of autoregressive order for the sieve bootstrap is not nearly as critical as the selection of block length for the block bootstrap.

Direct attempts by the block bootstrap to replicate the dependence structure of the original time series are accurate only to first order. They fail to reproduce second-order effects because the dependence is corrupted at places where blocks join; see for example Davison and Hall (1993). As a result, when using the percentile t method in conjunction with the block bootstrap to capture second-order features of the distribution of a statistic, an indirect approach must be taken: the standard error estimate must be altered to counteract second-order inaccuracies. This is unnecessary in the case of the sieve bootstrap, however, since that method preserves the dependence structure to second order. That explains property (b) above. Property (a) follows for essentially the same reason: second-order features of a time series are not consistently estimated by the unadjusted block bootstrap, and so the double-bootstrapped percentile method does not give second-order accuracy in that context, whereas it does when used in conjunction with the sieve bootstrap.

The difficulty of estimating the variance for general statistics associated with a time series can make percentile t methods unattractive. For example, techniques based on variance approximation by Taylor expansion can be algebraically tedious to apply in the time series case, and so the enthusiasm for pivoting that we often find in the context of independent observations does not translate to time series data. Therefore, property (a) above is arguably the most telling advantage that the sieve bootstrap has over blocking methods. Property (b) is important when we can estimate variance, and property (c) confers obvious empirical advantages on the sieve bootstrap.

We shall describe and develop percentile and percentile t methods, and their properties, in the case of linear time series that are representable in terms of past disturbances, or errors, since this is the case that is most commonly encountered in practice. It will be clear that only minor modifications are needed to obtain a method for time series that are based on both past and future disturbances. An extension to vector-valued linear time series is similarly straightforward. Our numerical work will address the case of non-linear statistics, such as the median, as well as linear statistics.

Block bootstrap and related methods were introduced by Hall (1985), Carlstein (1986), Künsch (1989), Liu and Singh (1992), Politis and Romano (1994) and Hall and Jing (1996). Properties of the block bootstrap and its relatives have been developed by, among others, Politis and Romano (1992, 1993), Shao and Yu (1993), Naik-Nimbalkar and Rajarshi (1994), Bühlmann (1994, 1995b), Bühlmann and Künsch (1994, 1995), Hall *et al.* (1995), Götze and Künsch (1996), Lahiri (1996, 1997, 1998, 1999), Carlstein *et al.* (1998) and Hall *et al.* (1998). Bootstrap methods for time series based on structural models, and related to the sieve bootstrap, have been developed and discussed by, for example, Freedman (1984), Efron and Tibshirani (1986), Bose (1988), Franke and Kreiss (1992) and Davison and Hinkley (1997), chapter 8.

2. Methodology

2.1. Bootstrapping a linear time series

Let $\mathcal{T} = (X_j, -\infty < j < \infty)$ denote a stationary, linear time series, representable in the form

$$X_j = \mu + \sum_{i=0}^{\infty} \alpha_i \xi_{j-i} \tag{2.1}$$

for constants μ and α_i and independent and identically distributed (IID) random variables ξ_i . Inverting the series (2.1) we obtain an autoregression of potentially infinite order: $\xi_j = \sum_{i \geq 0} \beta_i Y_{j-i}$ for constants β_i , where $Y_j = X_j - \mu$, or equivalently

$$Y_j = \sum_{i=1}^{\infty} \omega_i Y_{j-i} + \epsilon_j, \tag{2.2}$$

where $\omega_i = -\beta_i/\beta_0$ and $\epsilon_j = \xi_j/\beta_0$. Throughout we shall define $\beta_0 = 1$, in which case μ and the constants $\alpha_i, i \geq 0$, and $\beta_i, i \geq 1$, are uniquely determined and are related by the formula $\sum_{i \geq 0} \alpha_i s^i = (\sum_{i \geq 0} \beta_i s^i)^{-1}$. The autocovariance function is then given by $\gamma(j) \equiv \text{cov}(X_0, X_j) = u^2 \sum_{i \geq 0} \alpha_i \alpha_{i+j}$, where $u^2 = \text{var}(\epsilon_0)$.

Assume that we observe a finite consecutive subsequence $\mathcal{X} = (X_1, \dots, X_n)$ of \mathcal{T} . Following Kreiss (1992) and Bühlmann (1995a, 1997), we suggest using \mathcal{X} to compute a finite m th-order approximation to the autoregression (2.2), as follows. Put $\bar{X} = n^{-1} \sum_{1 \leq i \leq n} X_i$, $\hat{Y}_i = X_i - \bar{X}$, $\tilde{\gamma}(j) = (n-j)^{-1} \sum_{1 \leq i \leq n-j} \hat{Y}_i \hat{Y}_{i+j}$, for $0 \leq j < n$, and $\tilde{\gamma}(-j) = \tilde{\gamma}(j)$. Let \hat{A} be the $m \times m$ matrix having $\tilde{\gamma}(j_1 - j_2)$ as its (j_1, j_2) th element, and let $\hat{a} = (\tilde{\gamma}(1), \dots, \tilde{\gamma}(m))^T$ be a column vector of length m . Employing the Yule-Walker equations, put $\hat{\omega} = (\hat{\omega}_1, \dots, \hat{\omega}_m)^T = \hat{A}^{-1} \hat{a}$, and compute residuals $\hat{\epsilon}_j$ using an m th-order empirical version of equation (2.2):

$$\hat{\epsilon}_j = \hat{Y}_j - \sum_{i=1}^m \hat{\omega}_i \hat{Y}_{j-i}, \quad m+1 \leq j \leq n.$$

Centre the residuals, defining $\tilde{\epsilon}_j = \hat{\epsilon}_j - (n-m)^{-1} \sum_{m+1 \leq i \leq n} \hat{\epsilon}_i$.

Let $\{\epsilon_i^*, -\infty < i < \infty\}$ be chosen by sampling independently and uniformly, with replacement, from $\{\tilde{\epsilon}_{m+1}, \dots, \tilde{\epsilon}_n\}$, and compute a finite order sieve bootstrap approximation \mathcal{T}^* to \mathcal{T} : $\mathcal{T}^* = (\mathcal{X}_j^*, -\infty < j < \infty)$, where $X_j^* = Y_j^* + \bar{X}$ and

$$Y_j^* = \sum_{i=1}^m \hat{\omega}_i Y_{j-i}^* + \epsilon_j^*, \quad -\infty < j < \infty. \tag{2.3}$$

In practice we start the series \mathcal{T}^* in a non-stationary way and then ‘burn it in’, using the recursion (2.3).

The theoretical results that we shall give in Section 4 are valid for this construction of \mathcal{T}^* , and for many other constructions besides. In particular, recursive definitions of the autoregressive gradient estimators $\hat{\omega}_i$ are possible, or alternatively \mathcal{T}^* may be based on an m th-order moving average, rather than an m th-order autoregression, without affecting our main results (4.2) and (4.3).

2.2. Different approximate models

When employing either the percentile t or the iterated percentile bootstrap method there are at least two ways of using the bootstrap to simulate the procedure of making inference from a time series of potentially infinite autoregressive order. We may use a finite autoregression of length m throughout, taking m so large that we have a good approximation to the infinite length case, or we may simulate an autoregression of relatively large order, m_1 say, at each step of the bootstrap, and use order $m < m_1$ when estimating the variance (for the percentile t bootstrap). The latter approach attempts to mimic the procedure in which a finite autoregression of relatively low order is used to approximate an infinite order process. Although

it is important to take m_1 large, the size of m is not so critical for the second method; to a significant extent the bootstrap corrects for values of m that are too small, which can be expected to be more of a problem for relatively long-range-dependent autoregressions. In this sense the second method can have advantages over the first.

For brevity we shall consider only the first approach in detail. The second method is a little more complex to apply in practice, but it can perform well and has interesting properties. In particular, depending on the values of m and m_1 , the second approach can fail to be second order accurate in the usual sense, since errors of second order are introduced through the noise that is inherent in estimating high order autoregressive parameters. In this sense the m_1 th order autoregression is often not such a good approximation to the original, infinite order process. Nevertheless, coverage accuracy can be almost as good as for the first approach, since the aforementioned noise is virtually uncorrelated with the statistic of interest.

2.3. Empirical choice of m

It may be proved from results of Shibata (1976, 1980) that, under appropriate regularity conditions, the value \hat{m}_A of m chosen by the Akaike information criterion (AIC) satisfies $n^{1/2} \gamma(\hat{m}_A) \rightarrow C$ in probability, for a positive constant C . We shall argue in Section 4 that, to obtain second-order accurate confidence regions, an empirical order selector \hat{m} should satisfy $\gamma(\hat{m}) = o_p(n^{-1/2})$ in probability. Therefore, \hat{m}_A provides a lower bound to \hat{m} . An empirical selector that is an order of magnitude larger than m would generally be adequate.

In the case of exponentially decaying mixing coefficients, treated by Götze and Künsch (1996) and Lahiri (1996) in the context of the blockwise bootstrap, we require $\gamma(\hat{m}) = O_p(n^{-1})$ to obtain particularly high orders of accuracy. This is achieved, in asymptotic terms, if \hat{m} equals the integer part of $2\hat{m}_A + C'$, for any constant C' .

Likewise, if the mixing coefficients decrease only polynomially fast then taking \hat{m} equal to the integer part of $C'\hat{m}_A^2$, for any positive constant C' , is more than adequate. The fact that C' is not completely determined in either case is a reflection of the relative latitude that is available for selecting the autoregressive order in applications of the sieve bootstrap to constructing confidence regions. See property (c) in Section 1. This is in marked contrast with the block bootstrap, where the selection of the tuning parameter, i.e. the block length, is more critical. There, the block length plays the role of a smoothing parameter and adjusts the balance between the bias and variance in a crucial way. In the context of block bootstrap estimation of variance this property was discussed by Hall (1985). For block bootstrap construction of confidence intervals it has been addressed by Hall *et al.* (1995), Götze and Künsch (1996) and Lahiri (1996), among others. In many cases, the optimal block length is constrained to be asymptotic to $c_1 n^{c_2}$, where $c_2 > 0$ may be known but depends on the problem and $c_1 > 0$ is unknown and must be estimated from data.

The AIC choice of m provides a valuable practical guide to selecting the autoregressive order; see Section 3 for numerical examples.

2.4. Linear statistics

A linear statistic of lag l , computed from \mathcal{X} , is a sample mean of the form

$$\bar{X}_g = (n - l + 1)^{-1} \sum_{i=1}^{n-l+1} g(X_i, \dots, X_{i+l-1}),$$

where g is a known function of l variables. Let $\mu_g = E\{g(X_1, \dots, X_l)\}$ denote the expected

value of \bar{X}_g , and write $(n - l + 1)^{-1}\sigma_g^2$ for the variance, where

$$\sigma_g^2 = \gamma_g(0) + 2 \sum_{j=1}^{n-l} \{1 - j(n - l + 1)^{-1}\} \gamma_g(j)$$

and $\gamma_g(j) = \text{cov}\{g(X_1, \dots, X_l), g(X_{j+1}, \dots, X_{j+l})\}$. Under appropriate regularity conditions (Bühlmann, 1997), σ_g^2 is estimated root n consistently by

$$\hat{\sigma}_g^2 = \hat{\gamma}_g(0) + 2 \sum_{j=1}^{n-l} \{1 - j(n - l + 1)^{-1}\} \hat{\gamma}_g(j),$$

where $\hat{\gamma}_g(j) = \text{cov}\{g(X_1^*, \dots, X_l^*), g(X_{j+1}^*, \dots, X_{j+l}^*) | \mathcal{X}\}$, i.e. $\hat{\sigma}_g^2 - \sigma_g^2 = O_p(n^{-1/2})$ as $n \rightarrow \infty$. Among other matters, the regularity conditions involve the rate at which the order m of the fitted autoregression increases with n .

We shall consider bootstrap methods for constructing confidence intervals for μ_g , based on \bar{X}_g . Likewise, if f is a function of a single variable and has a continuous derivative in a neighbourhood of μ_g , with $f'(\mu_g) \neq 0$, we may use those methods to construct confidence intervals for $f(\mu_g)$ based on $f(\bar{X}_g)$. Again, the double-sieve-bootstrap produces second-order accuracy.

2.5. Percentile and percentile t methods

The percentile method amounts to approximating the unconditional distribution of \bar{X}_g or of $\bar{X}_g - \mu_g$ by the conditional distribution (given \mathcal{X}) of \bar{X}_g^* or of $\bar{X}_g^* - \bar{X}_g$ respectively. For example, using the second of these two approaches to construct a confidence interval (s, ∞) which covers μ_g with probability π , we would take $s = \bar{X}_g - \hat{r}_\pi$ where

$$\hat{r}_\pi = \inf\{r: P(\bar{X}_g^* - \bar{X}_g \leq r | \mathcal{X}) \geq \pi\}.$$

There are numerous ways to calibrate this interval, and one is to consider an additive correction of the original nominal coverage level. First compute the version \mathcal{T}^{**} of \mathcal{T}^* that arises if we replace \mathcal{X} by \mathcal{X}^* in the definition of \mathcal{T}^* , let \mathcal{X}^{**} be a segment of length n in \mathcal{T}^{**} , let \bar{X}_g^{**} be the analogue of \bar{X}_g^* computed from \mathcal{X}^{**} , let

$$\hat{r}_\pi^* = \inf\{r: P(\bar{X}_g^{**} - \bar{X}_g^* \leq r | \mathcal{X}^*) \geq \pi\}$$

denote the version of \hat{r}_π that arises if we replace $(\mathcal{X}, \mathcal{X}^*)$ by $(\mathcal{X}^*, \mathcal{X}^{**})$ when calculating \hat{r}_π , and put $\hat{p}(q) = P(\bar{X}_g^* - \hat{r}_q^* \leq \bar{X}_g | \mathcal{X})$ and

$$\hat{q}_\pi = \hat{p}^{-1}(\pi) = \inf\{q: \hat{p}(q) \geq \pi\}.$$

Then, $(\bar{X}_g - \hat{r}_{\hat{q}_\pi}, \infty)$ is a double-bootstrap, or calibrated bootstrap, percentile method confidence interval for μ_g , with nominal coverage π . Note that the autoregressions used in the two levels of the double bootstrap both have the same order, m .

Similarly, the percentile t bootstrap may be used to estimate the distribution of $(\bar{X}_g - \mu_g)/\hat{\sigma}_g$, and hence to calculate a confidence interval for μ_g . In general this can require application of the double bootstrap, with the second stage computing the bootstrap form $\hat{\sigma}_g^{*2} = \text{var}(\bar{X}_g^{**} | \mathcal{X}^*)$ of $\hat{\sigma}_g^2$. Virtually the same amount of effort is required to calculate a double-bootstrap percentile interval, and it has the advantage of being robust against stochastic fluctuations of the standard error estimators $\hat{\sigma}_g$ and $\hat{\sigma}_g^*$. The double bootstrap can sometimes be avoided in the context of percentile t methods, at the expense of relatively intricate algebraic calculations, but that can be daunting to a practitioner who is not inclined

theoretically. Together, these difficulties render percentile t methods relatively unattractive in general problems involving time series.

However, in the case of the sample mean, where $g(x) \equiv x$, we may compute $\hat{\sigma} = \hat{\sigma}_g$ (and so also $\hat{\sigma}^* = \hat{\sigma}_g^*$) by using only matrix algebra, without resort to Monte Carlo simulation. Thus, the percentile t method is here competitive in terms of computational labour. Specifically, assume that $1 - \sum_{1 \leq i \leq m} \hat{\omega}_i z^i$ has no zeros in the closed unit disc, which would always hold if the empirical covariances $\tilde{\gamma}_j$ were defined with factors n^{-1} instead of $(n-j)^{-1}$, put $\hat{\beta}_0 = 1$, $\hat{\beta}_i = -\hat{\omega}_i$, for $1 \leq i \leq m$, and $\hat{\beta}_i = 0$ otherwise, and define \hat{B} to be the infinite matrix with (j_1, j_2) th element $\sum_{i \geq 0} \hat{\beta}_i \hat{\beta}_{j_1+j_2-i}$. Let \tilde{u}^2 denote the variance of the sample $\tilde{\epsilon}_{m+1}, \dots, \tilde{\epsilon}_n$, and put $(\hat{\gamma}(0), \hat{\gamma}(1), \dots)^T = \tilde{u}^2 \hat{B}^{-1} (1, 0, 0, \dots)^T$. (In practice we employ a finite dimensional approximation.) Then, in the case $g(x) \equiv x$, $\hat{\gamma}$ is identical with $\hat{\gamma}_g$.

Moreover, the percentile t confidence interval for μ_g , with nominal coverage π , is $(\bar{X}_g - \hat{\sigma}_g \hat{s}_\pi, \infty)$ where

$$\hat{s}_\pi = \inf\{s: p\{(\bar{X}_g^* - \bar{X}_g)/\hat{\sigma}_g^* \leq s | \mathcal{X}\} \geq \pi\}.$$

3. Numerical properties

We undertook a simulation study to compare the performances of sieve and block bootstrap methods for constructing confidence intervals in linear problems, represented here by the mean and slope in simple linear regression, and non-linear problems, represented by the median and interquartile range. We employed four time series models:

$$X_i = \epsilon_i + 0.5\epsilon_{i-1} - 0.4\epsilon_{i-2},$$

where ϵ_i are IID and have a centred χ_1^2 -distribution (model T1);

$$X_i = 0.7X_{i-1} - 0.3X_{i-2} + \epsilon_i,$$

where ϵ_i are IID and have a centred χ_1^2 -distribution (model T2);

$$X_i = -0.8X_{i-1} - 0.5\epsilon_{i-1} + \epsilon_i,$$

where ϵ_i are IID with distribution $0.95 N(0, 1) + 0.05 N(0, 100)$ (model T3);

$$X_i = -0.8X_{i-1} - 0.5\epsilon_{i-1} + \epsilon_i,$$

where ϵ_i are IID and have a t_6 -distribution (model T4).

We have avoided discussing the case of normal errors since it would usually be treated by using parametric, rather than nonparametric, bootstrap methods. However, our techniques give good results in that case, since the symmetry and particularly light tails of the normal distribution significantly enhance the performance of the nonparametric bootstrap.

Models T3 and T4 have previously been considered by Bühlmann (1997). Our time series had lengths $n = 128$ and $n = 512$, although we report here in detail on only the case $n = 128$. The results in each setting were based on 500 simulations, with 199 and 200 resamples drawn at the first and second stages of resampling operations. The orders of autoregression employed in the sieve bootstrap were $m = 1, \dots, 15$. We chose the moving blocks version of the block bootstrap (see for example Künsch (1989)) in our study, with block lengths $l = 2, \dots, 48$.

The mean and median were approximated from 5000 simulations. For brevity we confine attention to one-sided 90% percentile method confidence intervals, of the form (s, ∞) , computed using the double bootstrap. Results for other linear and non-linear problems will be discussed briefly later in this section.

The AIC estimate \hat{m}_A of autoregressive order was computed by minimizing the AIC in the range $1 \leq m \leq [10 \log_{10}(n)]$. It had a particularly long-tailed distribution. In the most extreme case, that of model T2, the median of \hat{m}_A was 2.0 (for both $n = 128$ and $n = 512$), but the means and standard deviations were in each case more than twice this value. For models T2–T4 the standard deviation of \hat{m}_A was in all cases approximately equal to the mean, although for model T1 the standard deviation was about half the mean. (All these results are based on 5000 simulations in each setting.) However, since the performance of the sieve bootstrap is insensitive to an overestimation of m then these fluctuations seldom cause practical difficulties.

Fig. 1 depicts coverage errors of percentile method block bootstrap confidence intervals for the mean, as a function of block length, and also shows analogous results for the calibrated form. (The dotted and unbroken curves represent single- and double-bootstrap cases respectively.) Two features are immediately clear: the coverage errors are particularly sensitive to the choice of block length, and calibration does not in general offer an improved performance. The latter property is a consequence of the fact that the block bootstrap implicitly corrupts the dependence structure of the time series at those time points where blocks are joined together; the second bootstrap level further subdivides the series, producing further corruption.

Fig. 2 depicts analogous results for the sieve bootstrap. Here, the choice of autoregressive order has relatively little effect on performance, which is typically better than for the block bootstrap; and, although bootstrap calibration does not appreciably affect performance, we would expect this to be the case in this example since the coverage accuracy is comparatively good in the single-bootstrap setting.

The analogue of Fig. 1 in the case of the median is similar to that for the mean and so is not given here. However, Fig. 3 (the analogue of Fig. 2) is of interest. The uncalibrated form of the percentile method has particularly poor coverage performance for models T3 and T4, but calibration provides a dramatic improvement. As before, performance is relatively insensitive to the choice of autoregressive order.

Table 1 gives Monte Carlo approximations to the true coverages of nominal 90% intervals in the case of the sieve bootstrap, when $n = 128$ or $n = 512$, $\hat{m} = 2\hat{m}_A$ (as suggested by the discussion in Section 2.3), and \hat{m}_A was chosen using the AIC in each individual simulation. We truncated \hat{m} at $[10 \log_{10}(n)]$ to avoid excessively large \hat{m} . The results are much as predicted by Figs 2 and 3, reflecting the robustness of coverage against variations in m . In the case of the median, calibrated sieve bootstrap methods perform well for models T3 and T4, but poorly for models T1 and T2.

For non-linear problems, we also considered confidence intervals for quantiles and lengths of the interquartile range, obtaining results resembling those in Figs 2 and 3. Although beyond the theory presented in this paper, the coverages of calibrated intervals for the length of interquartile range with nominal 90% coverage are 90.6%, 82.0%, 89.0% and 86.2% respectively for models T1–T4. The sample size was $n = 128$, and the autoregressive order in each simulation was again chosen by the AIC, although the truncation step was found to be unnecessary here.

Other linear problems that we investigated include confidence intervals for the lag 1 autocorrelation parameter, as well as the slope parameter β , in simple linear regression with time series errors η_i , where $Y_i = \alpha + \beta x_i + \eta_i$, $x_i = i/n$ and $i = 1, \dots, n$. Again, the results are qualitatively similar to those in Figs 1 and 2—the performance of the block bootstrap is very sensitive to the choice of block size, and calibration offers little improvement. In contrast, with $(\alpha, \beta) = (2.0, 5.0)$ and $n = 128$, the coverages of percentile sieve bootstrap intervals for

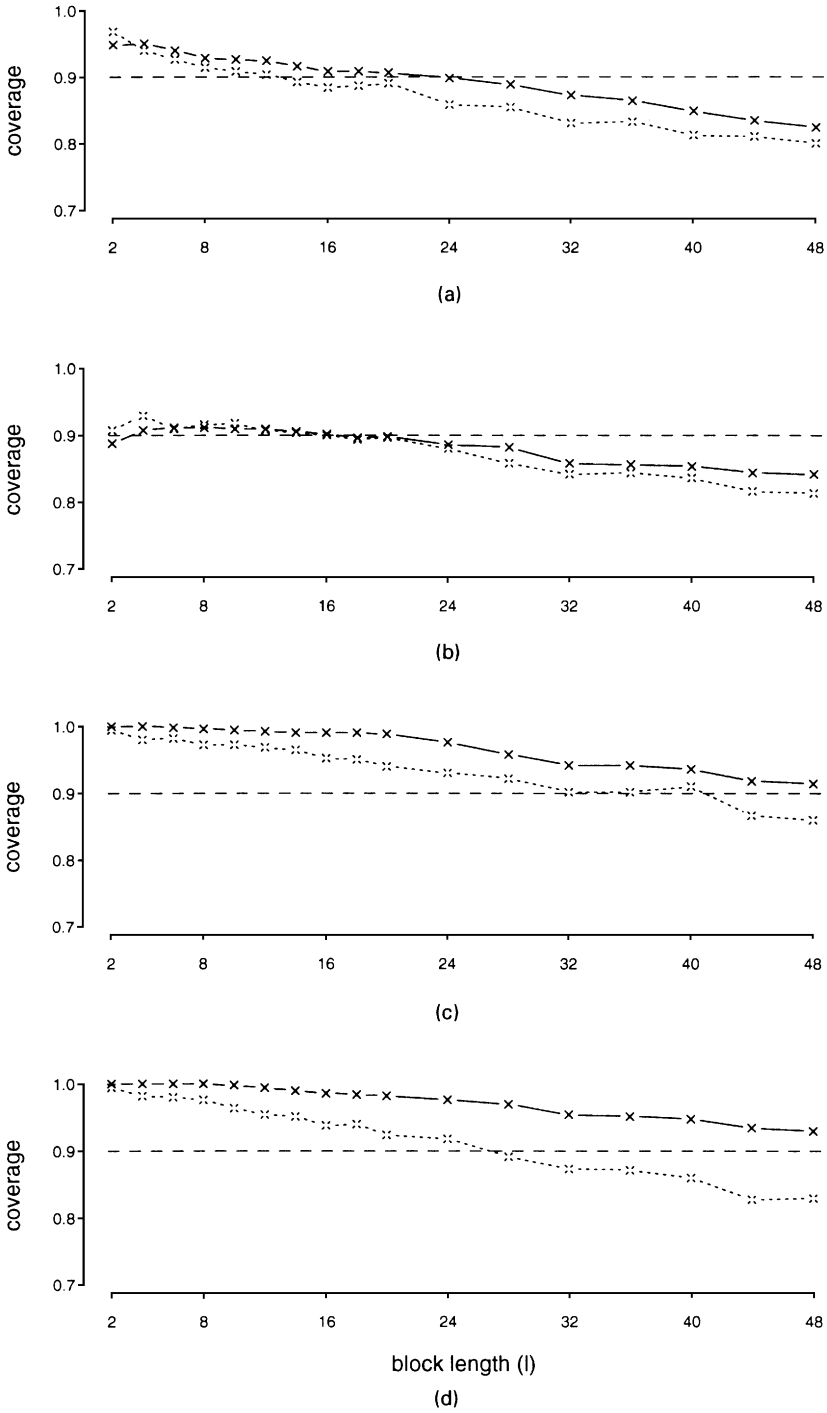


Fig. 1. Coverages of ordinary percentile (.....) and calibrated percentile (—) confidence intervals for the mean, constructed by using the block bootstrap (sample size $n = 128$; nominal coverage 90%): (a) model T1; (b) model T2; (c) model T3; (d) model T4

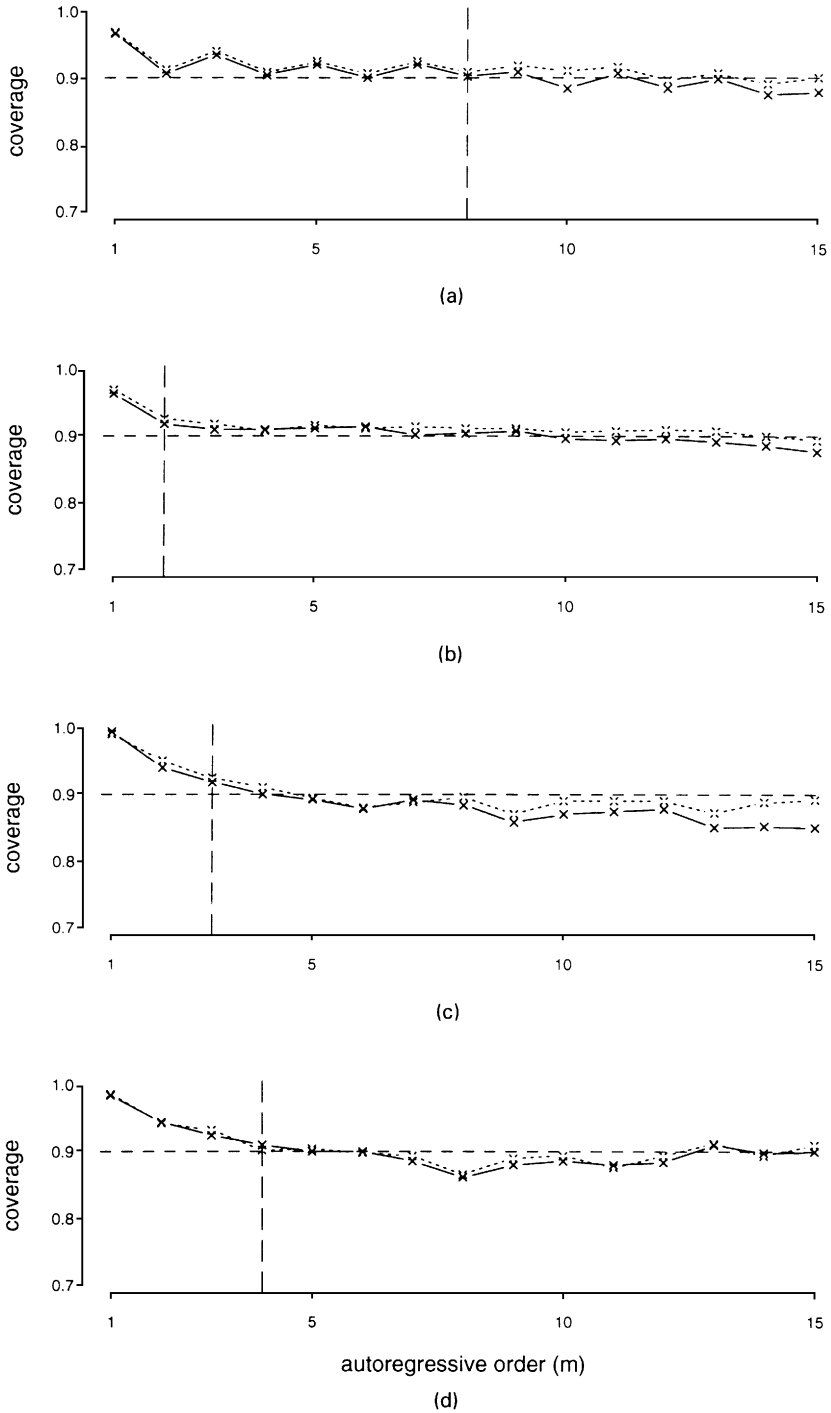


Fig. 2. Coverages of ordinary percentile (.....) and iterated percentile (—) confidence intervals for the mean, constructed by using the sieve bootstrap (sample size $n = 128$; nominal coverage 90%; |, median of AIC estimates of the autoregressive order): (a) model T1; (b) model T2; (c) model T3; (d) model T4

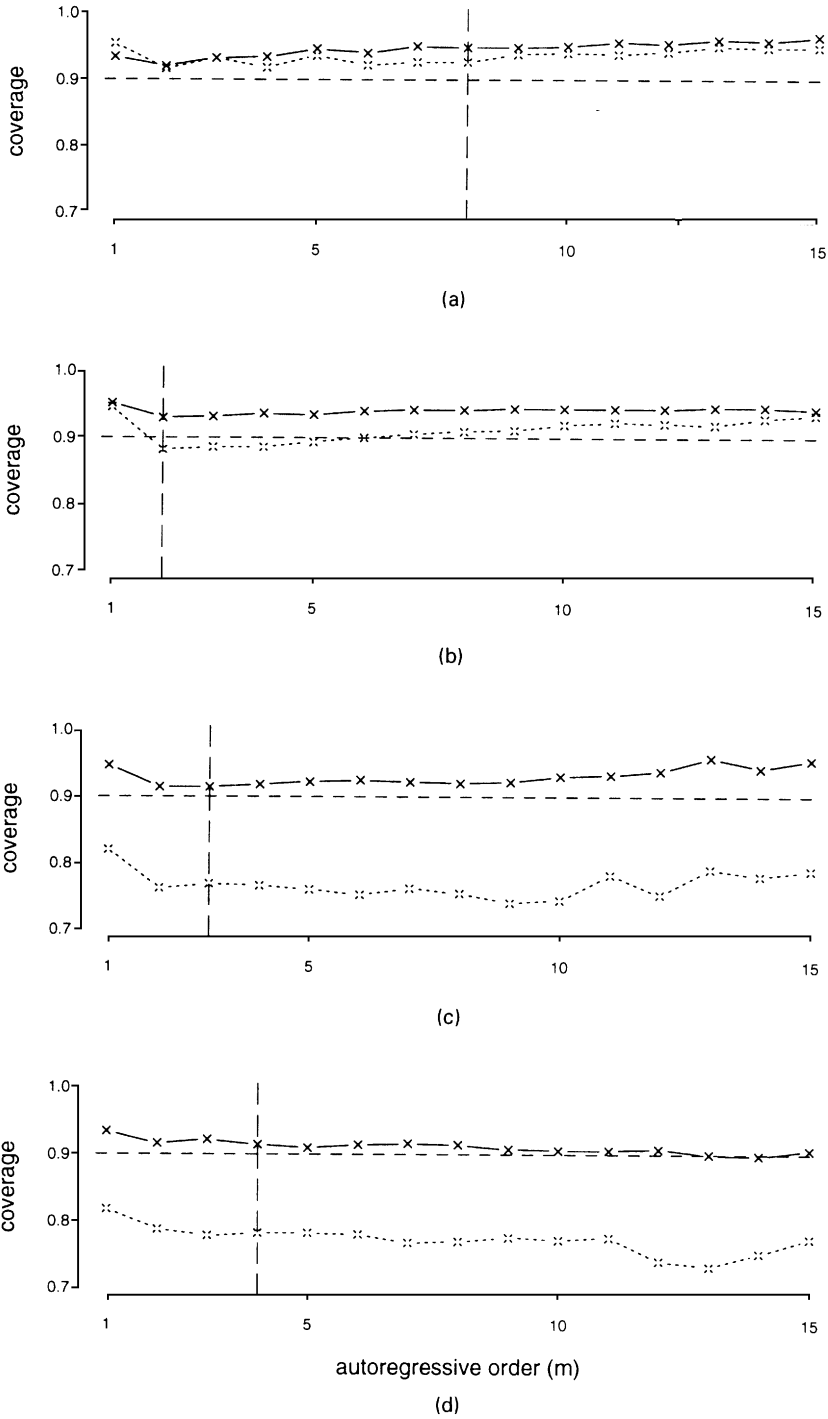


Fig. 3. Coverages of ordinary percentile (.....) and iterated percentile (—) confidence intervals for the median, constructed by using the sieve bootstrap (sample size $n = 128$; nominal coverage 90%; |, median of AIC estimates of the autoregressive order): (a) model T1; (b) model T2; (c) model T3; (d) model T4

Table 1. Coverage accuracy of ordinary and calibrated one-sided sieve bootstrap confidence intervals (s, ∞) , when the autoregressive order is taken equal to twice the AIC estimate†

Parameter	Method	Coverage accuracies for the following models:			
		T1	T2	T3	T4
<i>n</i> = 128					
Mean	Ordinary	0.894	0.892	0.882	0.902
	Iterated	0.884	0.888	0.868	0.900
Median	Ordinary	0.946	0.908	0.762	0.778
	Iterated	0.954	0.938	0.930	0.920
<i>n</i> = 512					
Mean	Ordinary	0.904	0.902	0.888	0.894
	Iterated	0.884	0.904	0.890	0.890
Median	Ordinary	0.970	0.926	0.728	0.726
	Iterated	0.974	0.946	0.928	0.894

†Sample sizes *n* were 128 and 512, and the nominal coverage was 90%.

β are 81.0%, 85.6%, 92.0% and 93.0% respectively when the errors η_i were generated from models T1–T4. Note here that the actual coverage is independent of the choice of (α, β) . The improvements are slight for calibrated methods in the present settings, with coverages equal to 82.2%, 86.6%, 92.0% and 92.8% respectively. Nevertheless, their performances generally improve as the sample size increases. For example, using $n = 512$, the calibrated methods have respective coverages of 91.0%, 92.0%, 90.8% and 91.4%.

Overall, our results suggest that the sieve bootstrap generally performs better, in terms of coverage accuracy, for linear statistics than for non-linear statistics. Its robustness against the choice of autoregressive order is clear for either class of estimators, however.

4. Theoretical properties

We address only the context of a linear statistic, introduced in Section 2.4. Similar results may be derived for smooth functions of multivariate linear statistics computed from multivariate linear time series.

Assume the following conditions:

- (a) all moments of the distribution of ϵ_0 are finite,
- (b) there exists $\rho \in (0, 1)$ such that $|\alpha_i| + |\beta_i| = O(\rho^i)$ as $i \rightarrow \infty$,
- (c) $g(x_1, \dots, x_l)$ is a non-degenerate finite linear form in products $p_1(x_1) \dots p_l(x_l)$, where each p_j is a polynomial, and
- (d) the distribution of ϵ_0 is absolutely continuous with a bounded probability density.

Condition (a) allows us to give relatively short proofs and may be relaxed if we are studying only Edgeworth and Cornish–Fisher expansions to two or three terms. Condition (b) implies the exponentially decaying mixing context of Götze and Künsch (1996) and so allows us to make a direct comparison with their results. Nevertheless, our proofs have versions in cases where $|\alpha_i|$ decreases only polynomially fast as $i \rightarrow \infty$, although there we do not obtain an order of accuracy of $O(n^{\eta-1})$ for all $\eta > 0$; see equations (4.2) and (4.3). Indeed, in the polynomial case the value of $\eta < \frac{1}{2}$ depends on the size of λ such that $|\alpha_i| = O(i^{-\lambda})$ and decreases to 0 as $\lambda \uparrow \infty$.

Conditions (a)–(c) imply that for even integers $r \geq 2$ the *r*th cumulant of $n^{1/2}(\bar{X}_g - \mu_g)$ is of order $n^{-(r-2)/2}$ and may be represented as a power series in n^{-1} . The first cumulant is of course

0, and for odd $r \geq 3$ the r th cumulant is of order $n^{-(r-2)/2}$ and can be represented as a series in odd powers of $n^{-1/2}$. Therefore, at least formally we may express the distribution of $n^{1/2}(\bar{X}_g - \mu_g)$ in an Edgeworth expansion:

$$P\{n^{1/2}(\bar{X}_g - \mu_g) \leq \sigma_g x\} = \Phi(x) + \sum_{j=1}^r n^{-j/2} \psi_j(x) \phi(x) + O(n^{-(r+1)/2}) \tag{4.1}$$

uniformly in x , where Φ and ϕ are respectively the standard normal distribution and density functions, and ψ_j is a polynomial of degree no more than $3j - 1$, not depending on n and with parity opposite to that of j .

In fact, if conditions (a)–(d) hold then equation (4.1) is true uniformly in x , for any $r \geq 1$. When $r = 1$, equation (4.1) is identical with its counterpart for the mean of a sum of n independent random variables, provided that we take standardized skewness to equal $\tau_g^0/(\sigma_g^0)^3$, where $(\sigma_g^0)^2 = \gamma_g(0) + 2 \sum_{j=1}^\infty \gamma_g(j)$ and

$$\tau_g^0 = \sum_{-\infty < i_1, i_2 < \infty} \sum E[\{g(X_{i_1}, \dots, X_{i_1}) - \mu_g\}\{g(X_{i_1}, \dots, X_{i_1+l-1}) - \mu_g\}\{g(X_{i_2}, \dots, X_{i_2+l-1}) - \mu_g\}].$$

To describe second-order accuracy of the calibrated percentile method we adjoin an extra condition to (a)–(d):

- (e) the order m of the fitted autoregression satisfies $m = O(n^\eta)$ and $\rho^m = O(n^{\eta-1})$ for all $\eta > 0$.

Then, assuming conditions (a)–(e), we have

$$P\{\mu_g \in (\bar{X}_g - \hat{r}_{\hat{q}_\pi}, \infty)\} = \pi + O(n^{\eta-1}) \tag{4.2}$$

uniformly in $0 < \pi < 1$, for all $\eta > 0$. The analogous formula for the other type of percentile interval, suggested in Section 2.5, also holds. Together these results establish property (a) in Section 1. We may weaken condition (e) to $m = O(n^\eta)$ for all $\eta > 0$, and $\rho^m = O(n^{-\eta-1/2})$ for some $\eta > 0$, if in equation (4.3) below we seek only the weaker form of second-order accuracy, where $O(n^{\eta-1})$ is replaced by $o(n^{-1/2})$.

Finally we state the version of equation (4.2) for percentile t intervals: assuming conditions (a)–(e),

$$P\{\mu_g \in (\bar{X}_g - \hat{\sigma}_g \hat{s}_\pi, \infty)\} = \pi + O(n^{\eta-1}) \tag{4.3}$$

uniformly in $0 < \pi < 1$, for all $\eta > 0$. This establishes property (b) in Section 1. Note that we can standardize for scale by using the ‘raw’ standard error $\hat{\sigma}_g$, rather than a modified version which would be necessary if we were to attain second-order accuracy for the percentile t version of the block bootstrap. The percentile t method may also be calibrated, reducing the error in equation (4.3) to $O(n^{\eta-3/2})$ for all $\eta > 0$. However, if we do not have a simple formula for $\hat{\sigma}_g$, and must estimate it by using the bootstrap, then implementing the calibrated percentile t method can require four levels of bootstrap iteration.

The closest analogue of equation (4.3) in the case of the block bootstrap percentile t method also has $O(n^{\eta-1})$, for all $\eta > 0$, on the right-hand side. See Götze and Künsch (1996) and Lahiri (1996). However, in this instance the adjusted variance estimator that is used takes negative values with positive probability. If instead we employ a variance estimator computed using lag weights that confer positivity, then the order of error is only $O(n^{\eta-2/3})$. These difficulties were discussed by Götze and Künsch (1996) and occur because of the need to correct in a somewhat unnatural way for the failure of the ‘naïve’ or ‘unadjusted’ block

bootstrap method to approximate all second-order effects accurately. These problems do not arise for the sieve bootstrap.

Property (c) in Section 1 is reflected in the fact that condition (e) involves only bounds on m . In marked contrast, good performance in the case of the block bootstrap requires the relatively demanding condition $m \sim c_1 n^{c_2}$, where $c_1, c_2 > 0$, c_1 is unknown and m now denotes the block length. Bühlmann and Künsch (1994) and Hall *et al.* (1995) have discussed the relatively tricky problem of estimating c_1 . The need to determine m in this concise manner makes the block bootstrap inherently more sensitive to the choice of the tuning parameter.

Under more restrictive conditions on the manner in which α_i and β_i vary as i increases, it may be shown that results (4.2) and (4.3) hold for the empirical, AIC-based choice of m , i.e. $m = \lfloor 2\hat{m}_A + C' \rfloor$ for any constant C' , considered in Section 2.3.

Appendix A: Technical arguments

A.1. Outline proof of equation (4.1)

Assume conditions (a)–(d) and that $N = O\{\log(n)\}$. Define $X_{Nj} = \sum_{0 \leq i \leq N} \alpha_i \epsilon_{j-i}$ and $\Delta_j = \sum_{i > N} \alpha_i \epsilon_{j-i}$, and put $g_k(x_1, \dots, x_l) = (\partial/\partial x_k) g(x_1, \dots, x_l)$ for $1 \leq k \leq l$. Then, for all i ,

$$\left| g(X_i, \dots, X_{i+l-1}) - \left\{ g(X_{Ni}, \dots, X_{N,i+l-1}) + \sum_{k=1}^l \Delta_{i+k-1} g_k(X_{Ni}, \dots, X_{N,i+l-1}) \right\} \right| \leq \sum_{k=1}^l \Delta_{i+k-1}^2 U_{ik}^2,$$

where the non-negative random variables U_{ik} satisfy $\sup_{ik} \{E(U_{ik}^\lambda)\} < \infty$ for all $\lambda > 0$. Given a random variable Z , write $(1 - E)Z$ for $Z - E(Z)$, and define

$$\Delta = (1 - E) \sum_{i=1}^{n-l+1} \sum_{k=1}^l \Delta_{i+k-1} g_k(X_{Ni}, \dots, X_{N,i+l-1}).$$

Since $|\alpha_i| = O(\rho^i)$ then, by direct calculation, we have $\sup_i \{E(\Delta_i^{2\nu})\} = O(\rho^{2N\nu})$ and $E(\Delta^{2\nu}) = O(n^\nu \rho^{2N\nu})$ for all integers $\nu \geq 1$. Hence, by Markov’s inequality,

$$P \left[\left| (1 - E) \sum_{i=1}^{n-l+1} \{g(X_i, \dots, X_{i+l-1}) - g(X_{Ni}, \dots, X_{N,i+l-1})\} \right| > (n^{1/2} \rho^N + n \rho^{2N}) n^\eta \right] = O(n^{-\lambda})$$

for all $\eta, \lambda > 0$. Similarly it may be proved that any cumulant of $n^{-1/2} S_N$, where

$$S_N = \sum_{1 \leq i \leq n-l+1} g(X_{Ni}, \dots, X_{N,i+l-1}),$$

differs from its counterpart for $n^{-1/2} \bar{X}_g$ only in terms of order ρ^N . Hence, if we can show that the formal r -term Edgeworth expansion of the distribution of $n^{-1/2} S_N$ is valid up to a remainder of order $n^{-(r+1)/2}$, then equation (4.1) will follow on taking N equal to the integer part of a sufficiently large constant multiple of $\log(n)$.

The desired expansion for $n^{-1/2} S_N$ follows from results of Heinrich (1985). We take p , in Heinrich’s notation, arbitrarily large, and we avoid his integrability assumption on the characteristic function of S_N —see Heinrich’s condition (1.3)—by adding an independent normal $N(0, n^{-\lambda})$ random variable to $n^{-1/2} S_N$, for $\lambda > r + 1$. This ‘perturbation’ produces an error that is of smaller order than the $n^{-(r+1)/2}$ -term in the remainder and provides enough smoothing to confer the desired integrability.

A.2. Outline proof of equation (4.2)

The methods used to derive equation (4.1) may be extended to give its bootstrap form, which in the case $r = 1$ implies that result uniformly in x . Therefore, if \hat{r}_π is as defined in Section 2.5, and if r_π is the solution of $P(\bar{X}_g - \mu_g \leq r_\pi) = \pi$, then $\hat{r}_\pi / \hat{\sigma}_g = r_\pi / \sigma_g + O_p(n^{\eta-3/2})$ for all $\eta > 0$. Therefore, since (as we shall shortly show) $\hat{\sigma}_g^2 - \sigma_g^2 = O_p(n^{\eta-1/2})$ for all $\eta > 0$, then

$$n^{1/2}(\hat{r}_\pi - r_\pi) = \frac{1}{2} r_\pi \sigma_g^{-2} (\hat{\sigma}_g^2 - \sigma_g^2) + O_p(n^{\eta-1}). \tag{A.1}$$

The delta method for Edgeworth expansions now gives that $P\{\mu_g \in (\bar{X}_g - \hat{r}_\pi, \infty)\} = p(\pi) + O(n^{\eta-1})$, for all $\eta > 0$, where we define

$$p(\pi) = P\{n^{1/2}(\bar{X}_g - \mu_g) - \frac{1}{2}\sigma_g^{-2}(\hat{\sigma}_g^2 - \sigma_g^2) \leq n^{1/2}r_\pi\}.$$

The delta method may be justified by using the fact that the term $R(\pi)$, represented by $O_p(n^{\eta-1})$ in equation (A.1), may be shown to satisfy $P\{|R(\pi)| > n^{\eta-1}\} = O(n^{-\lambda})$ for all $\eta, \lambda > 0$, and by using the Edgeworth expansion that we consider next.

By developing an Edgeworth expansion of the distribution of the 2-vector

$$V = (n^{1/2}(\bar{X}_g - \mu_g), n^{1/2}(\hat{\sigma}_g^2 - \sigma_g^2))^T \tag{A.2}$$

we may construct an Edgeworth expansion of $p(\pi)$,

$$p(\pi) = \pi + n^{-1/2} \psi_{(1)}(z_\pi) \phi(z_\pi) + O(n^{\eta-1}) \tag{A.3}$$

for all $\eta, \lambda > 0$, where z_π is the π -level quantile of the standard normal distribution and $\psi_{(1)}$ is a polynomial. Then, using arguments that are now standard in asymptotic theory for the double bootstrap, but are more complex on account of m -dependence, we may prove that the contribution $n^{-1/2} \psi_{(1)}(z_\pi) \phi(z_\pi)$ is corrected by the double bootstrap, up to $O(n^{\eta-1})$ for all $\eta > 0$. We shall content ourselves with an outline of the proof of the Edgeworth expansion of the distribution of V , going as far as terms that are necessary to identify the non-remainder components in equation (A.3). Condition (c) is essential for our proof; it allows us to reduce calculations to no more than a more complex version of those in the case $g(x) \equiv x$.

Put $\hat{\beta}_0 = 1, \hat{\beta}_i = -\hat{\omega}_i$ for $1 \leq i \leq m$, and $\hat{\beta}_i = 0$ for $i > m$. Let $\hat{\alpha}_0, \hat{\alpha}_1, \dots$ be defined by the identity $\sum_{i \geq 0} \hat{\alpha}_i s^i = (\sum_{i \geq 0} \hat{\beta}_i s^i)^{-1}$, for $|s| < 1$. Using the moment assumptions on ϵ_0 , and the fact that $m = O(n^\eta)$ for all $\eta > 0$, we may prove that

$$\max_{0 \leq i \leq m} |\hat{\beta}_i - \beta_i| = O_p(n^{\eta-1/2})$$

for all $\eta > 0$. See also Hannan and Kavalieris (1986). For any $\eta > 0$ there exists $\rho_1 \in [0, 1)$ such that $\rho_1^m = O(n^{\eta-1/2})$. Then,

$$\max_{0 \leq i \leq m} (|\hat{\beta}_i|/\rho_1^i) = O_p(1),$$

which in view of the definition of $\hat{\alpha}_i$ implies that ρ_1 may be chosen to satisfy the aforesaid conditions and also $\sup_{i \geq 1} (|\hat{\alpha}_i|/\rho_1^i) = O_p(1)$.

From this point the proof for general g , satisfying condition (c), is similar to that in the simplest case, where $l = 1$ and $g(x) \equiv x$. The first step is to express each $\hat{\gamma}_g(j)$ as a series in polynomials in the $\hat{\alpha}_i$, and the next to write $\hat{\alpha}_{i_1} \dots \hat{\alpha}_{i_k}$ as

$$\alpha_{i_1} \dots \alpha_{i_k} + \sum_j (\hat{\alpha}_{i_j} - \alpha_{i_j}) \prod_{r \neq j} \alpha_{i_r},$$

plus terms of smaller order. For brevity we treat only the case $g(x) \equiv x$. There, $\hat{\gamma}_g(j) = \hat{\gamma}(j) = \tilde{u}^2 \sum_{i \geq 0} \hat{\alpha}_i \hat{\alpha}_{i+j}$ and $\gamma_g(j) = \gamma(j) = u^2 \sum_{i \geq 0} \alpha_i \alpha_{i+j}$, where \tilde{u}^2 denotes the sample variance of $\tilde{\epsilon}_{m+1}, \dots, \tilde{\epsilon}_n$ and estimates $u^2 = \text{var}(\epsilon_0)$. Now, $|\hat{\gamma}(j)| + |\gamma(j)| = O_p(\rho_1^j)$, and so

$$\hat{\sigma}_g^2 - \sigma_g^2 = \sum_{-\infty < j < \infty} \{\hat{\gamma}(j) - \gamma(j)\} + O_p(n^{-1}) = \tilde{u}^2 \left(\sum_{i=0}^{\infty} \hat{\alpha}_i \right)^2 - u^2 \left(\sum_{i=0}^{\infty} \alpha_i \right)^2 + O_p(n^{-1}).$$

For brevity, in the argument below we choose not to develop an explicit approximation to the $O_p(n^{-1})$ term here through sums of m -dependent random variables, but in view of condition (c) such an approximation is possible.

We may prove that $\tilde{u}^2 - u^2 = \hat{u}^2 - u^2 + O_p(n^{\eta-1})$, where $\hat{u}^2 = n^{-1} \sum_{1 \leq i \leq n} \epsilon_i^2, \hat{u}^2 - u^2 = O_p(n^{-1/2})$, and

$$\begin{aligned} \sum_{i=0}^{\infty} \hat{\alpha}_i - \sum_{i=0}^{\infty} \alpha_i &= - \left\{ \sum_{i=0}^{\infty} (\hat{\beta}_i - \beta_i) \right\} / \left(\sum_{i=0}^{\infty} \beta_i \right) \sum_{i=0}^{\infty} \hat{\beta}_i, \\ \sum_{i=0}^{\infty} |\hat{\beta}_i - \beta_i| &\leq m \max_{0 \leq i \leq m} |\hat{\beta}_i - \beta_i| + \sum_{i=m+1}^{\infty} |\beta_i| = O_p(n^{\eta-1/2}), \end{aligned}$$

for all $\eta > 0$. Therefore, writing $s = \sum_{i \geq 0} \beta_i$ we have that

$$\hat{\sigma}_g^2 - \sigma_g^2 = (\hat{u}^2 - u^2)s^{-2} - 2u^2 \left(\sum_{i=0}^m \hat{\beta}_i - s \right) s^{-3} + O_p(n^{\eta-1}). \tag{A.4}$$

We may assume without loss of generality that $\mu = E(X_0) = 0$. Put

$$\tilde{\gamma}(j) = n^{-1} \sum_{1 \leq i \leq n} X_i X_{i+j},$$

let \bar{A} and A be the $m \times m$ matrices with $\tilde{\gamma}(j_1 - j_2)$ and $\gamma(j_1 - j_2)$ respectively as their (j_1, j_2) th components, and define $\bar{a} = (\tilde{\gamma}(1), \dots, \tilde{\gamma}(m))^T$, $a = (\gamma(1), \dots, \gamma(m))^T$ and $\mathbf{1} = (1, \dots, 1)^T$. Then, $\bar{A} = A + O_p(n^{\eta-1})$, and so $-(\hat{\beta}_1, \dots, \hat{\beta}_m)^T = \bar{A}^{-1} \hat{a} = \bar{A}^{-1} \bar{a} + O_p(n^{\eta-1})$, for all $\eta > 0$. Put $\Delta_A = \bar{A} - A$ and $\Delta_a = \bar{a} - a$. Then, $\bar{A}^{-1} \bar{a} = A^{-1} a + A^{-1} \Delta_a - A^{-1} \Delta_A A^{-1} \bar{a} + O_p(n^{\eta-1})$, whence

$$\sum_{i=1}^m \hat{\beta}_i = -\mathbf{1}^T \bar{A}^{-1} \hat{a} = -\mathbf{1}^T A^{-1} a + (A^{-1} \mathbf{1})^T (\Delta_A A^{-1} a - \Delta_a) + O_p(n^{\eta-1}) \tag{A.5}$$

for all $\eta > 0$.

We may assume without loss of generality that $u^2 = 1$. Then we may write $\gamma(j) = \sum_{i \geq 0} \alpha_i \alpha_{i+j}$ and $\sum_{i \geq 0} \beta_i \alpha_{j-i} = \delta_{0j}$, the Kronecker delta. Hence, for $i \geq 1$,

$$\sum_{j=1}^{\infty} \gamma(i-j) \beta_j = \sum_{k=1}^{\infty} \alpha_k \left(\sum_{j=0}^{\infty} \beta_j \alpha_{k+i-j} - \alpha_{k+i} \right) = -\gamma(i). \tag{A.6}$$

If we take A_{∞} and a_{∞} to be the versions of A and a respectively that arise when $m = \infty$, and if we define $b_{\infty} = (\beta(1), \beta(2), \dots)^T$ and $\mathbf{1}_{\infty} = (1, 1, \dots)^T$, then equation (A.6) is equivalent to $A_{\infty} b_{\infty} = -a_{\infty}$, i.e. $b_{\infty} = -A_{\infty}^{-1} a_{\infty}$. Therefore, $\sum_{i \geq 1} \beta_i = -\mathbf{1}_{\infty}^T A_{\infty}^{-1} a_{\infty}$. A similar argument, writing A_{∞} as a 2×2 block matrix with A in the top left-hand corner, and invoking condition (b), may be used to prove that

$$\mathbf{1}^T A^{-1} a = -\sum_{i=1}^m \beta_i + O(\rho^m) = 1 - s + O(\rho^m).$$

Hence, since $\hat{\beta}_0 = \beta_0 = 1$ and $\rho^m = O(n^{\eta-1})$, we deduce from equation (A.5) that

$$1 + \sum_{i=1}^m \hat{\beta}_i - s = (A^{-1} \mathbf{1})^T (\Delta_A A^{-1} a - \Delta_a) + O_p(n^{\eta-1}) \tag{A.7}$$

for all $\eta > 0$.

Each element of Δ_A or of Δ_a is of the form $\tilde{\gamma}(j) - \gamma(j)$ for some $|j| \leq m - 1$. Put

$$\Gamma(j) = (1 - E)n^{-1} \sum_{1 \leq i \leq n} X_{mi} X_{m,i+j},$$

and observe that $\tilde{\gamma}(j) - \gamma(j) = \Gamma(j) + O_p(n^{\eta-1})$, uniformly in $|j| \leq m$. Therefore, if we replace $\tilde{\gamma}(j) - \gamma(j)$ by $\Gamma(j)$ whenever the former appears in either Δ_A or Δ_a , result (A.7) continues to hold. It has the form $\sum_{0 \leq i \leq m} \hat{\beta}_i - s = T_1 + O_p(n^{\eta-1})$, where T_1 is now a linear function in $2m$ -dependent random variables with zero mean. Trivially, both $\bar{X}_g - u_g$ and $T_2 = \hat{u}^2 - u^2$ also have this form; and, by equation (A.4),

$$\hat{\sigma}_g^2 - \sigma_g^2 = T_2 s^{-2} - 2u^2 T_1 s^{-3} + O_p(n^{\eta-1}).$$

The techniques of Heinrich (1982, 1984, 1985), and the delta method, may now be used to derive an Edgeworth expansion of the distribution of $V = (V_1, V_2)^T$, defined at equation (A.2). Contributions of size $n^{-1/2}$ from V_2 may be included as ‘dummy variables’; as noted earlier, they may be identified if desired, but they enter only into the remainder term in the expansion of $p(\pi)$ at equation (A.3). Arguing

thus, the expansion of the distribution of V may be taken up to a remainder of $O(n^{\eta-1})$, for all $\eta > 0$, leading to result (4.2).

In the double bootstrap form of these arguments we approximate an m th-order autoregression, with coefficients $\hat{\omega}_1, \dots, \hat{\omega}_m$, by another autoregression of the same order, with coefficients $\hat{\omega}_1^*, \dots, \hat{\omega}_m^*$. Therefore, the problem of quantifying the error arising from approximating an infinite order autoregression by one of finite order does not arise at this level.

A.3. Outline proof of equation (4.3)

The arguments in Appendix A.2 show how to derive an Edgeworth expansion of the joint distribution of $n^{1/2}(\bar{X}_g - \mu_g)$ and $n^{1/2}(\hat{\sigma}_g^2 - \sigma_g^2)$, up to and including terms that are necessary to identify contributions of smaller order than $n^{\eta-1}$, for any $\eta > 0$, to the Edgeworth expansion of the distribution of $T = n^{1/2}(\bar{X}_g - \mu_g)/\hat{\sigma}_g$. The analogous expansion of the bootstrap distribution of $T^* = n^{1/2}(\bar{X}_g^* - \bar{X}_g)/\hat{\sigma}_g^*$ may be derived in the same way. Moreover, the terms in the expansion for T^* are within $O(n^{\eta-1})$ of their counterparts for T . Result (4.3) now follows via the delta method.

References

- Bose, A. (1988) Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.*, **16**, 1709–1722.
- Bühlmann, P. (1994) Blockwise bootstrapped empirical processes for stationary sequences. *Ann. Statist.*, **22**, 995–1012.
- (1995a) Moving-average representation for autoregressive approximations. *Stochast. Process. Applic.*, **60**, 331–342.
- (1995b) The blockwise bootstrap for general empirical processes of stationary sequences. *Stochast. Process. Applic.*, **58**, 247–265.
- (1997) Sieve bootstrap for time series. *Bernoulli*, **3**, 123–148.
- Bühlmann, P. and Künsch, H. R. (1994) Block length selection in the bootstrap for time series. *Research Report 72*. Seminar für Statistik, Eidgenössische Technische Hochschule Zürich, Zürich.
- (1995) The blockwise bootstrap for general parameters of a stationary time series. *Scand. J. Statist.*, **22**, 35–54.
- Carlstein, E. (1986) The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Ann. Statist.*, **14**, 1171–1179.
- Carlstein, E., Do, K.-A., Hall, P., Hesterberg, T. and Künsch, H. R. (1998) Matched-block bootstrap for dependent data. *Bernoulli*, **4**, 305–328.
- Davison, A. C. and Hall, P. (1993) On Studentizing and blocking methods for implementing the bootstrap with dependent data. *Aust. J. Statist.*, **35**, 215–224.
- Davison, A. C. and Hinkley, D. N. (1997) *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.
- 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 two-stage least-squares estimates in stationary linear 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.
- Götze, F. and Künsch, H. R. (1996) Second-order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.*, **24**, 1914–1933.
- Hall, P. (1985) Resampling a coverage pattern. *Stochast. Process. Applic.*, **20**, 231–246.
- Hall, P., Horowitz, J. L. and Jing, B. (1995) On blocking rules for the block bootstrap with dependent data. *Biometrika*, **82**, 561–574.
- Hall, P. and Jing, B. (1996) On sample reuse methods for dependent data. *J. R. Statist. Soc. B*, **58**, 727–737.
- Hall, P., Jing, B.-Y. and Lahiri, S. N. (1998) On the sampling window method under long range dependence. *Statist. Sin.*, **8**, 1189–1204.
- Hannan, E. J. and Kavalieris, L. (1986) Regression, autoregression models. *J. Time Ser. Anal.*, **7**, 27–49.
- Heinrich, L. (1982) A method for the derivation of limit theorems for sums of m -dependent random variables. *Z. Wahrsch. Ver. Geb.*, **64**, 211–239.
- (1984) Non-uniform estimates and asymptotic expansions of the remainder in the central limit theorem for m -dependent random variables. *Math. Nachschr.*, **115**, 7–20.
- (1985) Some remarks on asymptotic expansions in the central limit theorem for m -dependent random variables. *Math. Nachschr.*, **122**, 151–155.
- Kreiss, J.-P. (1992) Bootstrap procedures for AR(∞)-processes. *Lect. Notes Econ. Math. Syst.*, **376**, 107–113.

- Künsch, H. R. (1989) The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, **17**, 1217–1261.
- Lahiri, S. N. (1996) On Edgeworth expansion and moving block bootstrap for studentized M -estimators in multiple linear regression models. *J. Multiv. Anal.*, **56**, 42–59.
- (1997) On empirical choice of the optimal block size for block bootstrap methods.
- (1998) Effects of block lengths on the validity of block resampling methods.
- (1999) Theoretical comparison of block bootstrap methods. *Ann. Statist.*, **27**, 386–404.
- Liu, R. and Singh, K. (1992) Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (eds R. Lepage and L. Billard), pp. 225–248. New York: Wiley.
- 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.
- 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.
- (1993) Nonparametric resampling for homogeneous strong mixing random fields. *J. Multiv. Anal.*, **47**, 301–328.
- (1994) The stationary bootstrap. *J. Am. Statist. Ass.*, **89**, 1303–1313.
- Shao, Q.-M. and Yu, H. (1993) Bootstrapping the sample means for stationary mixing sequences. *Stochast. Process. Applic.*, **48**, 175–190.
- Shibata, R. (1976) Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, **63**, 117–126.
- (1980) Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.*, **8**, 147–164.