Biometrical Journal **47** (2005) 3, 346–357 DOI: 10.1002/bimj.200310112

On Studentising and Blocklength Selection for the Bootstrap on Time Series

M. Peifer[*,1], **B. Schelter**[1,2], **B. Guschlbauer**[2], **B. Hellwig**[2], **C. H. Lücking**[2], and **J. Timmer**[1]

[1] Freiburg Centre for Data Analysis and Modelling, Eckerstr. 1, 79104 Freiburg, Germany
[2] Neurologische Universitätsklinik Freiburg, Breisacher Str. 64, 79106 Freiburg, Germany

Received 18 December 2003, revised 24 May 2004, accepted 29 June 2004

Summary

For independent data, non-parametric bootstrap is realised by resampling the data with replacement. This approach fails for dependent data such as time series. If the data generating process is at least stationary and mixing, the blockwise bootstrap by drawing subsamples or blocks of the data saves the concept. For the blockwise bootstrap a blocklength has to be selected. We propose a method for selecting the optimal blocklength. To improve the finite size properties of the blockwise bootstrap, studentised statistics is considered. If the statistic can be represented as a smooth function model this studentisation can be approximated efficiently. The studentised blockwise bootstrap method is applied for testing hypotheses on medical time series.

Key words: Blockwise bootstrap; Studentising; Smooth function model; Tremor.

1 Introduction

Since the introduction of the bootstrap, Efron (1979), methods based on resampling have been applied on numerous statistical problems. The success of bootstrap may be explained by its easy implementation whenever the data are identical distributed and statistically independent. If time series of stochastic processes are taken into account, the observations are in the most cases not independent. Then, statistical inference using bootstrap methods becomes much more difficult and is often based on profound model assumptions, see e.g. model based bootstrap, Davison and Hinkley (1997) or the sieve bootstrap for ARMA-models, Bühlmann (1997).

If the model class or the model equations are unknown, a possible method of estimating the variance of a statistic was proposed by Hall (1985) and Carlstein (1986). This non-parametric method is based on building subsamples of the sequence of observations. Adapting this idea to the bootstrap methodology leads to blockwise bootstrap, Künsch (1989); Liu and Singh (1992). Here, the subsamples or data blocks are resampled, instead of the data points itself. For establishing a consistent approximation of the distribution or variance of the statistic, structural assumptions on the underlying process has to be posed. These assumptions are usually α-mixing and strong stationary. It is assumed that these basic conditions are valid throughout this paper. The free parameter – blocklength l – has to be adjusted to each specific problem. By minimising the mean squared error, Hall et al. (1995) showed that the optimal blocklength for a series of length n is proportional to $n^{\frac{1}{3}}$ for bias or variance approximation, $n^{\frac{1}{4}}$ for the approximation of one-sided distribution functions and $n^{\frac{1}{5}}$ for the approximation of two-sided distribution functions. The constant of proportionality depends on the process and on the statistic. In the following a method to select this constant is described. We show that this method gives appropriate estimates for the blocklength.

The second issue deals with the studentisation of the blockwise bootstrap method. If the statistic T_n can be described by a smooth function of means, the variance of T_n can be approximated by its first

[*] Corresponding author: e-mail: peifer@fdm.uni-freiburg.de

order Taylor expansion. A more general approach is to estimate this variance by linearising the statistic using the empirical influence function Hampel et al. (1986); Huber (1980). These two methods differ only by the way the statistic is linearised and therefore it is sufficient to regard only the smooth function model.

By using a formal Edgeworth expansion, it can be shown that the rate of convergence of e.g. confidence intervals increases if the statistic is studentised, Davison and Hall (1993); Götze and Künsch (1996); Hall (1988); Hall and Wilson (1991); Hall (1992), Fisher and Hall (1990). Note that for the validity of an Edgeworth expansion (see e.g. Bhattacharya and Ghosh (1978) for independent observations) in case of time dependent observations, the mixing coefficient has to decay exponentially, moment conditions have to be satisfied and a conditional Cramér condition has to be fulfilled, Götze and Hipp (1983). We conjecture that the convergence of bootstrap is increased even if there is no valid Edgeworth expansion, which might result from the scale correcting nature of the studentisation.

Recent contributions on blockwise bootstrap are mostly found in the econometric sector, such as Paparoditis and Politis (2003, 2001); Fitzenberger (1996). In case of biometrics, the contributions are rather sparse. For demonstrating the practical significance of the blockwise bootstrap in the area of medical statistics, a suitable application is discussed in Section 7.

2 Blockwise Bootstrap

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be an observed sample from a strongly stationary, α-mixing, p-variate sequence $(\boldsymbol{X}_t)_{t \in \mathbb{Z}}$. The real-valued statistics $T_n = T_n(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ is assumed to be invariant under permutations of the observations.

For the blockwise bootstrap, b subsamples or blocks of length l are formed from the observations. We further assume, without loss of generality, that $n/l \in \mathbb{N}$ (otherwise, the data sample is truncated until $n/l \in \mathbb{N}$ holds). In the framework of blockwise bootstrap, two kinds of building subsamples are predominating, the overlapping blocks and the non-overlapping blocks. The overlapping blocks are defined by

$$\boldsymbol{Y}_i = (\boldsymbol{X}_i, \ldots, \boldsymbol{X}_{i+l}) \quad i = 1, \ldots, b = n - l - 1\,, \tag{1}$$

and non-overlapping are defined as follows:

$$\boldsymbol{Y}_i = (\boldsymbol{X}_{(i-1)l+1}, \ldots, \boldsymbol{X}_{il}) \quad i = 1, \ldots, b = \frac{n}{l}\,. \tag{2}$$

It turns out that the blockwise bootstrap gives quite comparable results if either overlapping or non-overlapping blocks are used, Künsch (1989); Hall et al. (1995). Unless otherwise mentioned, the results of this paper remain the same whether overlapping or non-overlapping blocks are used.

Blockwise bootstrap is realised by resampling the blocks Y_i and gluing them together to form a kind of surrogate time series of length n. Finally, the statistic is applied on each bootstrapped series to estimate quantities like distribution functions, bias or variance. The algorithmic representation of this procedure is:

1. Drawing blocks with replacement from $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_b\}$ and forming $\boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_b^*$ by gluing the drawn blocks together.
2. Repeat 1. B times to generate B bootstrap samples $\boldsymbol{X}_1^*, \ldots, \boldsymbol{X}_B^*$.
3. Calculate $T_{n,k}^* = T_n(\boldsymbol{X}_{1,k}^*, \ldots, \boldsymbol{X}_{n,k}^*)$, $k = 1, \ldots, B$.
4. Finally, determine the bootstrap approximation of e.g.:
   - the distribution function $F^*(\xi) = B^{-1} \sum_{k=1}^{B} \theta(\xi - T_{n,k}^*)$
   - the bias $Bias^* = B^{-1} \sum_{k=1}^{B} T_{n,k}^* - T_n(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$
   - the variance $\sigma^{2*} = B^{-1} \sum_{k=1}^{B} \left(T_{n,k}^* - B^{-1} \sum_{j=1}^{B} T_{n,j}^* \right)^2$,

   where $\theta(\cdot)$ is the step function.

Some remarks on the consistency of the described blockwise bootstrap: Only demanding strong stationarity and that the mixing coefficient $\alpha(\tau)$ vanishes with respect to the time lag τ is too weak for a valid bootstrap approximation. The consistency can for example be achieved, as Naik-Nimbalkar and Rajarshi (1994) have shown, if

- $\sum_{i=0}^{\infty} (i+1)^7\, \alpha(i)^{1/2-\tau} < \infty$, for $\tau \in (0, \frac{1}{2})$
- X_i has a continuous distribution function on $\mathbb{R}$
- $l = l(n) = O(n^{1/2-\epsilon})$, with $0 < \epsilon < \frac{1}{2}$
- the statistic T_n has to be Hadamard or compactly differentiable within a sufficient space of distributions (see also Gill (1989)).

For the choice of the blocklength l, we assume that the mixing coefficient decays exponentially with respect to the time lag. This also assures under mild moment conditions and under some kind of Cramér condition the validity of the Edgeworth expansion, Götze and Hipp (1983). Here again, the compact differentiability cannot be dropped. The described smooth function model is always compact differentiable.

3 Studentising the Blockwise Bootstrap

3.1 Smooth function model

Suppose an i.i.d. sample of p-variate random vectors, $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n$. Let $\mu = E[\boldsymbol{W}_i]$ and $\bar{\boldsymbol{W}} = n^{-1} \sum_{i=1}^{n} \boldsymbol{W}_i$. A function A is called a smooth function model if

$$A : \mathbb{R}^p \to D \subset \mathbb{R}\,, \qquad A \in C^{\infty}(\mathbb{R}^p, D) \quad \text{and} \quad A(\mu) = 0\,. \tag{3}$$

This concept can be applied on statistical problems in which the statistic can be expressed by a smooth function g such as $T_n = g(\bar{\boldsymbol{W}})$. The associated smooth function model then yields: $A(\boldsymbol{w}) = g(\boldsymbol{w}) - g(\mu)$. The main advantage of the smooth function model is that the variance of the statistic T_n can be well approximated. Additionally, by using the chain rule, the existence of the Fréchet derivative with respect to the distribution is proven. In the following the i-th component of a vector, say $\boldsymbol{Y}$, is denoted by $Y^{(i)}$. To approximate the variance of T_n let

$$\boldsymbol{Z} = n^{\frac{1}{2}}(\bar{\boldsymbol{W}} - \mu)\,, \qquad a_i = \left.\frac{\partial A(\boldsymbol{w})}{\partial w^{(i)}}\right|_{\boldsymbol{w}=\mu}$$

$$b_{ij} = \left.\frac{\partial^2 A(\boldsymbol{w})}{\partial w^{(i)}\, \partial w^{(j)}}\right|_{\boldsymbol{w}=\mu} \quad \text{and} \quad C_{ij} = E[(\boldsymbol{W}_1 - \mu)^{(i)}\, (\boldsymbol{W}_1 - \mu)^{(j)}]\,.$$

Because of the independence of the $\boldsymbol{W}_i$'s, $E[Z^{(i)}] = 0$ and $E[Z^{(i)} Z^{(j)}] = C_{ij}$. Now, Taylor-expanding the function $S_n = n^{\frac{1}{2}} A(\bar{\boldsymbol{W}})$ at $\boldsymbol{w} = \mu$, leads to:

$$S_n = \sum_{i=1}^{p} a_i Z^{(i)} + n^{-\frac{1}{2}} \frac{1}{2} \sum_{i,j=1}^{p} b_{ij} Z^{(i)} Z^{(j)} + O_p(n^{-1})\,.$$

And therefore the expectations $E[S_n]$ and $E\left[S_n^2\right]$ are given by:

$$E[S_n] = n^{-\frac{1}{2}} \frac{1}{2} \sum_{i,j=1}^{p} b_{ij} C_{ij} + O(n^{-1})\,,$$

$$E\left[S_n^2\right] = \sum_{i,j=1}^{p} a_i a_j C_{ij} + O(n^{-1})\,.$$

Using $\mathrm{Var}\,(T_n) = n^{-1}\, \mathrm{Var}\,(S_n) = n^{-1}(E[S_n^2] - (E[S_n])^2) + O(n^{-2})$ gives the approximative variance of T_n:

$$\mathrm{Var}\,(T_n) = n^{-1} \sum_{i,j=1}^{p} a_i a_j C_{ij} + O(n^{-2})\,. \tag{4}$$

Finally, the estimators $\hat{a}_i = \left.\dfrac{\partial A(\boldsymbol{w})}{\partial w^{(i)}}\right|_{\boldsymbol{w}=\bar{\boldsymbol{W}}}$ and $\hat{C}_{ij} = n^{-1} \sum_{k=1}^{n} (\boldsymbol{W}_k - \bar{\boldsymbol{W}})^{(i)} (\boldsymbol{W}_k - \bar{\boldsymbol{W}})^{(j)}$ are plugged into Eq. (9) to estimate the variance:

$$\hat{\sigma}_n^2 = \widehat{\mathrm{Var}}\,(T_n) = n^{-1} \sum_{i,j=1}^{p} \hat{a}_i \hat{a}_j \hat{C}_{ij}\,. \tag{5}$$

Extending these results to blockwise defined statistics leads to variance estimators which are sufficient for studentising the described bootstrap method.

3.2 The procedure of studentising blockwise bootstrap

Since Eq. (5) does not give consistent results if the observation $\boldsymbol{W}_i$ are not i.i.d., a modification of the smooth function A is needed before applying these results on time series. To approximate the variance of T_n under conditions outlined in Section 2, the observations $\boldsymbol{X}_i$ are transformed into a new series $\boldsymbol{W}_i$ in which $T_n = g(\bar{\boldsymbol{W}})$, as in Section 3.1. Again, the series $\boldsymbol{W}_i$ is not i.i.d.. To incorporate the dependence structure into $\hat{\sigma}_n^2$, the blocking scheme is implemented into the statistic. For this purpose define

$$\tilde{W}_i = (W^{(1)}_{(i-1)\,l+1}, \ldots, W^{(1)}_{il}, \ldots, W^{(p)}_{(i-1)\,l+1}, \ldots, W^{(p)}_{il})\,, \quad i = 1, \ldots, \frac{n}{l}$$

for non-overlapping blocks and for overlapping blocks

$$\tilde{W}_i = (W^{(1)}_i, \ldots, W^{(1)}_{i+l}, \ldots, W^{(p)}_i, \ldots, W^{(p)}_{i+l})\,, \quad i = 1, \ldots, n - l - 1\,.$$

The statistic is then rewritten into

$$T_n = \tilde{g}\left(b^{-1} \sum_{i=1}^{b} \tilde{\boldsymbol{W}}_i\right),$$

where $\tilde{g}(w) = g\left(l^{-1} \sum_{i=1}^{l} (w^{(i)}, \ldots, w^{((p-1)\,l+i)})\right)$. And hence again, $A(\boldsymbol{w}) = \tilde{g}(\boldsymbol{w}) - \tilde{g}(\mu)$. Suppose that the blocklength is adequately chosen and therefore the blocks are approximately independent, such that the approximation in Eq. (5) can be used.

To give an example of the described procedure, let T_n be the sample variance $T_n = n^{-1} \sum_{i=1}^{n} X_i^2 - \left(n^{-1} \sum_{i=1}^{n} X_i\right)^2$. Then

$$\begin{aligned}\hat{\sigma}_n^2 = {} & n^{-1} l^{-2} 4 \Bigg[(\hat{E}[X_1])^2 \left(l\hat{g}_1(0) + 2 \sum_{i=1}^{l-1} (l-i)\, \hat{g}_1(i) \right) + l\hat{g}_2(0) + 2 \sum_{i=1}^{l-1} (l-i)\, \hat{g}_2(i) \\ & + 2\hat{E}[X_1] \left(l\hat{g}_3(0) + 2 \sum_{i=1}^{l-1} (l-i)\, \hat{g}_3(i) + l\hat{g}_4(0) + 2 \sum_{i=1}^{l-1} (l-i)\, \hat{g}_4(i) \right) \Bigg],\end{aligned}$$

where

$$\hat{E}[X_1] = \bar{X} = n^{-1} \sum_{i=1}^{n} X_i\,,$$

$$\hat{g}_1(\tau) = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} (X_i - \bar{X})\,(X_{i+\tau} - \bar{X})\,, \qquad \hat{g}_2(\tau) = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} \left(X_i^2 - \tilde{X}^2\right) \left(X_{i+\tau}^2 - \tilde{X}^2\right),$$

$$\hat{g}_3(\tau) = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} (X_i - \bar{X}) \left(X_{i+\tau}^2 - \tilde{X}^2\right), \qquad \hat{g}_4(\tau) = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} \left(X_i^2 - \tilde{X}^2\right) (X_{i+\tau} - \bar{X})\,,$$

and $\tilde{X}^2 = n^{-1} \sum_{i=1}^{n} X_i^2\,.$

Studentised bootstrap is realised by studentising each bootstrap sample $T^*_{\text{stud},n} = \dfrac{T^*_n - T_n}{\hat{\sigma}^*_n}$. The bootstrap approximation of the studentised statistic is now determined similar to Section 2 and can be used e.g. to approximate confidence intervals of level α

$$\hat{\mathcal{I}}_\alpha = [t_- \cdot \hat{\sigma}_n + T_n,\, t_+ \cdot \hat{\sigma}_n + T_n]\,,$$

where

$$t_- = \sup_{x \in \mathbb{R}} \left\{ \left(\frac{\alpha}{2}\right) \geq F^*_{\text{stud}}(x) \right\} \quad \text{and} \quad t_+ = \inf_{x \in \mathbb{R}} \left\{ \left(1 - \frac{\alpha}{2}\right) \leq F^*_{\text{stud}}(x) \right\}.$$

The distribution function $F^*_{\text{stud}}(x)$ has to be calculated as in Section 2. The only difference is that T^*_n is replaced by its studentised bootstrap samples $T^*_{\text{stud},n}$.

4 Blocklength Selection

The success of the method depends on the choice of the blocklength l. Taking the mean squared error as objective measure to be minimised, the blocklength selection is similar to the choice of the smoothing parameter in non-parametric function estimation, Peifer et al. (2003); Hall et al. (1995). In both cases a tradeoff between bias and variance is present. For the sample mean, the mean squared error can be calculated and is in case of the bootstrap variance approximation, Hall et al. (1995),

$$\text{MSE}\,(l) = E[(\text{Var}^*\,(\bar{X}) - \text{Var}\,(\bar{X}))^2] \approx \frac{1}{n^2 l^2} C_1 + \frac{l}{n^3} C_2\,, \tag{6}$$

where

$$C_1 = \left(\sum_{k=-\infty}^{\infty} |k|\, \gamma(k) \right)^2, \qquad C_2 = \begin{cases} 2 \left(\sum\limits_{k=-\infty}^{\infty} \gamma(k) \right)^2 & \text{for non-overlapping blocks} \\ \dfrac{4}{3} \left(\sum\limits_{k=-\infty}^{\infty} \gamma(k) \right)^2 & \text{for overlapping blocks}\,. \end{cases}$$

Here, $\gamma(k)$ is the auto-covariance function of the process. The optimal blocklength is therefore given by $l_{\text{opt}} = (2C_1/C_2)^{1/3} n^{1/3}$. To generalise this concept, the statistic T_n is linearised before, using either the smooth function model or the empirical influence function. The constants C_1 and C_2 are then determined by the covariance structure of the linearised statistic. A first idea to estimate the blocklength is to plug-in the empirical auto-covariance function into C_1 and C_2. Since the correlated errors of the estimated auto-covariance function are amplified by the factor $|k|$ in $C_1 = \left(\sum_{k=-\infty}^{\infty} |k|\, \gamma(k) \right)^2$ this method fails. To avoid this problem, the global behaviour of the auto-covariance function is parameterised and fitted to the empirical auto-covariance function. For this, it is assumed that the mixing coefficient decays exponentially and hence the auto-covariance function decays in maximum exponentially, too. In making use of this extra assumption, some points of the correlation function are estimated: $\hat{\gamma}(k), k = 0, \ldots, m < n$, and the function $f(k) = \phi^k, 0 \leq \phi < 1$ is fitted to the envelope of $\hat{\gamma}(k)$. Then the estimated parameter ϕ contains the characteristic time scale of the process and the blocklength is finally estimated by replacing $\gamma(k)$ in C_1, C_2 with ϕ^k. The procedure of selecting the blocklength is therefore the following:

- Linearise the statistic T_n by transforming the data points to V_i such that $T_n = g(\bar{\boldsymbol{W}}) \approx n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \left.\frac{\partial g(\boldsymbol{w})}{\partial w^{(j)}}\right|_{\boldsymbol{w}=\boldsymbol{W}} W_i^{(j)} = n^{-1} \sum_{i=1}^{n} V_i \Rightarrow V_i = \sum_{j=1}^{p} \left.\frac{\partial g(\boldsymbol{w})}{\partial w^{(j)}}\right|_{\boldsymbol{w}=\boldsymbol{W}} W_i^{(j)}$ for the smooth function model of Section 3.1. Or alternatively linearise the statistic using the influence function approach, Hampel et al. (1986); Huber (1980).
- Estimate the auto-covariance function of the transformed series V_i.

- Determine the envelope of the estimated auto-covariance function. In case of oscillatory processes we propose using the Hilbert transform, Oppenheim and Schafer (1975). This method gives reliable results if the auto-covariance function $\hat{\gamma}(k)$ of V_i is having a narrow frequency band or equivalently, if the power-spectrum of process V_i is having pronounced peaks, Gabor (1946); Boashash (1992). The Hilbert transform for the discrete signal $\hat{\gamma}(k)$ is then carried out by

$$\tilde{\gamma}(k) = \mathcal{F}^{-1}\{2\theta(\omega)\,\mathcal{F}\{\hat{\gamma}\}\,(\omega)\}\,(k)\,,$$

  where $\mathcal{F}, \mathcal{F}^{-1}$ denotes the discrete Fourier transformation and its reverse transformation. Again, $\theta(\cdot)$ is the step function, truncating the negative frequencies of the signal. Since $\tilde{\gamma}(k) \in \mathbb{C}$, the signal can be decomposed into $\tilde{\gamma}(k) = A(k)\,e^{i\varphi(k)}$, in which $A(k)$ is interpreted as envelope of the original signal $\hat{\gamma}$.
- Fit $f(k) = \phi^k$ to the envelope by using well elaborated methods such as the Levenberg-Marquardt algorithm, Press et al. (1992).
- The blocklength $\hat{l}$ is then:

$$\hat{l} = \left(\frac{4v\left(\frac{\phi}{1-\phi} + \frac{\phi^2}{(1-\phi)^2}\right)^2}{\left(1 + 2\frac{\phi}{1-\phi}\right)^2}\, n \right)^{\frac{1}{3}}, \tag{7}$$

  where $v = 1$ for non-overlapping and $v = 3/2$ for overlapping blocks.

For a AR[1]-process, the parameter ϕ asymptotically coincides the process parameter in $X_{t-1} = \phi X_t + \epsilon_t$, where ϵ_t is Gaussian white noise. It should be remarked that the formula determining the blocklength, Eq. (7), slightly differs from the formula derived in Carlstein (1986). This is due to the chosen resampling scheme, whereas Carlstein straightly uses the subsamples to estimate the variance of the regarded statistic. In general, due to the calculation of the envelope of the auto-covariance function the proposed approach differs from fitting a single autoregressive process of order 1 to the process and determining the blocklength from the process parameter. Such a procedure was e.g. studied in Sherman (1998) within the class of autoregressive-moving average processes. In the following, a simulation study is performed to investigate the proposed blocklength selection method and the effect of the studentisation.

5 Simulations

To investigate the effect of studentisation and the choice of the blocklength, three different data generating processes are chosen:

- Autoregressive process of order 1 (AR[1]), $X_t = a_1 X_{t-1} + \sigma\epsilon_t$, $a_1 = \exp(-1/\tau)$, $\tau > 0$ and $(\epsilon_t)_{t\in\mathbb{Z}}$ i.i.d. sequence of $\mathcal{N}(0,1)$ random variables.
- Autoregressive process of order 2 (AR[2]), $X_t = a_1 X_{t-1} + a_2 X_{t-2} + \sigma\epsilon_t$, where $a_1 = 2\exp(-1/\tau)\cos\left(\frac{2\pi}{T}\right)$, $a_2 = -\exp(-2/\tau)$, τ, $T > 0$ and $(\epsilon_t)_{t\in\mathbb{Z}}$ i.i.d. sequence of $\mathcal{N}(0,1)$ random variables.
- Stochastic van der Pol oscillator (SVDP), given by the stochastic differential equation

$$dX_1 = X_2\,dt$$

$$dX_2 = \left\{\mu(1 - X_1^2)\,X_2 - X_1\right\}dt + \sigma\,dB_t\,, \quad \mu > 0\,,$$

  where dB_t is the increment of the Brownian motion.

The motivation of these three processes is the following: The two autoregressive processes are linear but their dynamical behaviour is different – whereas the AR[1]-process can be seen as a stochastically driven relaxator with relaxation time τ. Since this process is non-oscillatory, the calculation of the envelope for selecting the blocklength is skipped for each simulation concerning the AR[1]. The AR[2]-process can be interpreted as a stochastically driven, damped oscillator with relaxation time τ and period T. Finally, the SVDP shows non-linear oscillating behaviour with a mean period of approximately 9, Kurrer and Schulten (1991); Leung (1995).

Throughout the simulation the variance σ was set equal one. For the AR[1]-process the parameter was chosen $\tau = 5$ and for the AR[2]-process $\tau = 10$, $T = 5$. The stochastic differential equation of the SVDP was integrated by an Euler integration scheme, where $\mu = 3$, the integration step size $\delta t = 0.001$ and the sampling time was chosen to be $\Delta t = 0.5$. More details and the theoretical justification of the chosen parameters for the SVDP can be found in Timmer (2000).

Beside the data generating processes, a specific statistic has to be chosen. Motivated by the application, Section 7, the sample variance is used throughout this section. In order to give a measure of the accuracy for the following bootstrap approximations, two-sided equally tailed 95%-confidence intervals are calculated. The coverage-error of the confidence intervals is then estimated by the relative frequency, in which the variance of the process is falling into the interval over 1000 independent runs. The sign of the coverage error was chosen to be negative for conservative and positive for anti-conservative confidence intervals. Since the variance of the SVDP is not known theoretically, rather long datasets of 10^5 data points are used to approximate the variance.

The results of the simulations are shown in Figure 1, where the coverage error is determined in dependence on the amount of data. This is done for either studentised or non-studentised blockwise bootstrap, in which the blocklength is selected by the described method of Section 4. For both, the AR[1]-process and the AR[2]-process the asymptotic coverage is approached for only 1500 data points, when the studentised method is used. In contrast to the non-studentised method, which is still showing a small coverage-error at $n = 5000$ data points. The situation is different for the SVDP, where the blockwise bootstrap shows a slight conservative behaviour, which is independent of studentisation. We therefore propose using the studentised bootstrap to enhance the rate of convergence, which is in accordance of many theoretical results concerning the blockwise bootstrap and the "ordinary" bootstrap, see e.g. Davison and Hall (1993); Götze and Künsch (1996); Beran (1987); Hall and Martin (1988); Hall (1992); Hall and Wilson (1991); Fisher and Hall (1990); Timmer et al. (1999).

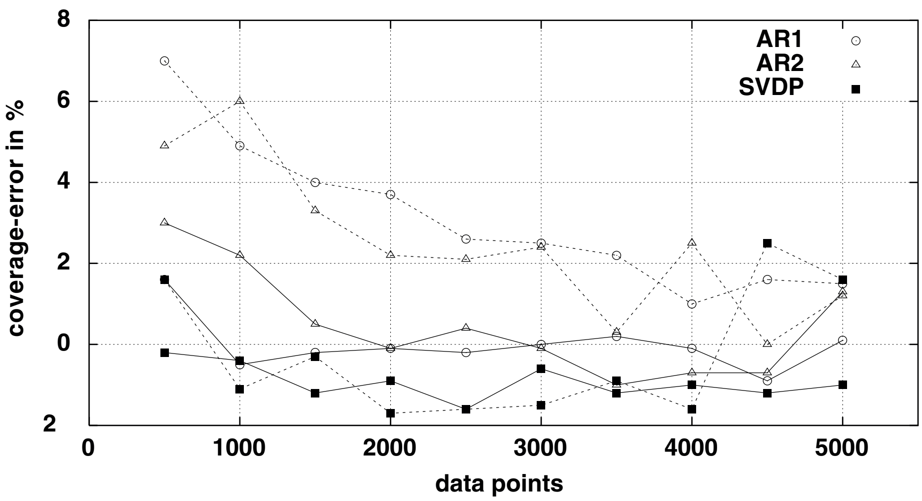

Figure 1 Coverage error of 95%-confidence intervals on dependence of the number of data points, where both, the non-studentised blockwise bootstrap (dotted lines) and the studentised blockwise bootstrap (solid lines) are considered. The data generating processes are AR[1]-process (○), AR[2]-process (△) and the stochastic van der Pol oscillator (■). Negative values of the coverage error denote conservative confidence intervals while positive values are corresponding to anti-conservative confidence intervals.

Table 1 Comparison of the selected blocklength $\hat{l}$ with its optimal value l_{opt} minimising the MSE. The estimated blocklength is averaged over 1000 independent realisations of the three exemplary processes. For getting an impression of the distribution of the selected blocklength, the standard deviation of $\hat{l}$ is calculated. Additionally, the MSE of the selected blocklength is compared with the optimal mean-squared-error MSE_{opt}.

		$\hat{l}$	MSE	l_{opt}	MSE_{opt}
AR[1]	$n = 1000$	18.02 ± 2.71	0.079	18	0.078
	$n = 3000$	26.33 ± 2.26	0.028	26	0.028
	$n = 5000$	31.19 ± 2.20	0.017	31	0.017
AR[2]	$n = 1000$	25.18 ± 3.24	0.090	21	0.085
	$n = 3000$	35.88 ± 2.91	0.032	30	0.031
	$n = 5000$	42.42 ± 2.71	0.020	35	0.019
SVDP	$n = 1000$	30.04 ± 2.67	0.0017	73	0.0016
	$n = 3000$	42.28 ± 2.15	0.0006	104	0.0005
	$n = 5000$	49.87 ± 1.91	0.0004	124	0.0003

To compare the estimated blocklength $\hat{l}$, according to Section 4, with the optimal l_{opt}, realisations having 1000, 3000 and 5000 data points are regarded for each exemplary process. The selected blocklengths are averaged over an ensemble of 1000 of such realisations. In case of the AR[1]-process, the optimal blocklength can be calculated analytically. For the AR[2]-process and the SVDP the optimal blocklength is determined by averaging the the auto-covariance function up to time lag 100 over 10000 independent realisations. The results are shown in Table 1. As expected, the estimated blocklengths coincides perfectly for the AR[1]-process. For the AR[2]-process, the estimated blocklengths are slightly larger than the optimal, but the calculated mean-squared-errors are nearly identical. In case of the SVDP the estimated blocklengths are substantially smaller than the optimal and again, the mean-squared-errors are roughly matching. This indicates that the minimum of the MSE is very broad for the SVDP. The proposed method for selecting the blocklength therefore yielded reliable blocklengths within the given simulations. Furthermore, inspecting the variance of the estimated blocklengths reveals a decrease while increasing the amount of data which must be satisfied because of the consistency of the auto-covariance estimation under the assumed properties of the process.

6 Alternative Methods for Selecting the Blocklength

Hall et al. (1995) have used a kind of cross-validation to estimate the optimal blocklength. They form subsamples from the observed realisation to approximate the mean-squared-error for a given blocklength. Then, the blocklength is varied such that the approximated mean-squared-error is minimised. Finally, the known scaling law is used to extrapolate the calculated blocklength to the original number of data points. The main disadvantage of this procedure is the immense expense of computation time, which makes this method almost non-applicable for large datasets.

The second alternative is due to Bühlmann and Künsch (1999). This method is quite similar to the proposed method. The main difference is that the sums in Eq. (6) are calculated by using spectral methods, in contrast to the time domain approach we use.

To compare the distribution of the selected blocklength for the three methods, the moving average process

$$X_i = \theta + (Y_{i+1} + Y_{i+2})\, 2^{-\frac{1}{2}},$$

$Y_i \sim \chi_1^2$ i.i.d. and $\theta \in R$, also given in Hall et al. (1995); Bühlmann and Künsch (1999), is considered. The distribution of the blocklength is estimated by the relative frequency over 1000 independent rea-

Table 2 Distributions of the empirically chosen blocklength, calculated by the relative frequencies of 1000 simulations. All time series are of length 100 and are generated according to the moving average process in Sec. 6. The distributions of the alternative methods are due to Hall et al. (1995), Bühlmann and Künsch (1999) and the theoretically optimal blocklength is $l_{\text{opt}} = 3$.

l	1	2	3	4	5	6	7	8	10	11	13
Hall et al.	0.00	0.27	0.52	0.00	0.06	0.07	0.00	0.02	0.02	0.02	0.02
Bühlm et al.	0.08	0.08	0.45	0.31	0.07	0.01	0.00	0.00	0.00	0.00	0.00
$\hat{l}$	0.00	0.00	0.10	0.48	0.30	0.09	0.03	0.00	0.00	0.00	0.00

lisations, where the statistic T_n is now the sample mean. The results are shown in Table 2. Since the theoretically optimal blocklength is $l_{\text{opt}} = 3$, our method tends to have slightly larger blocklength than the others but has in contrast no mass at $\hat{l} < l_{\text{opt}}$. We therefore conclude that the results of our method are comparable within this simulation.

7 Application

The proposed method is applied to test a time of the day (TOD) dependency in the variance of physiological (healthy) hand tremor. For each data set the recorded time series are of length 30000 data points and are sampled with 1 kHz. The proband reported not to have drunken any coffee or alcohol, and there was no medication for a period of 24 hours before the recordings. At 4 different time of the

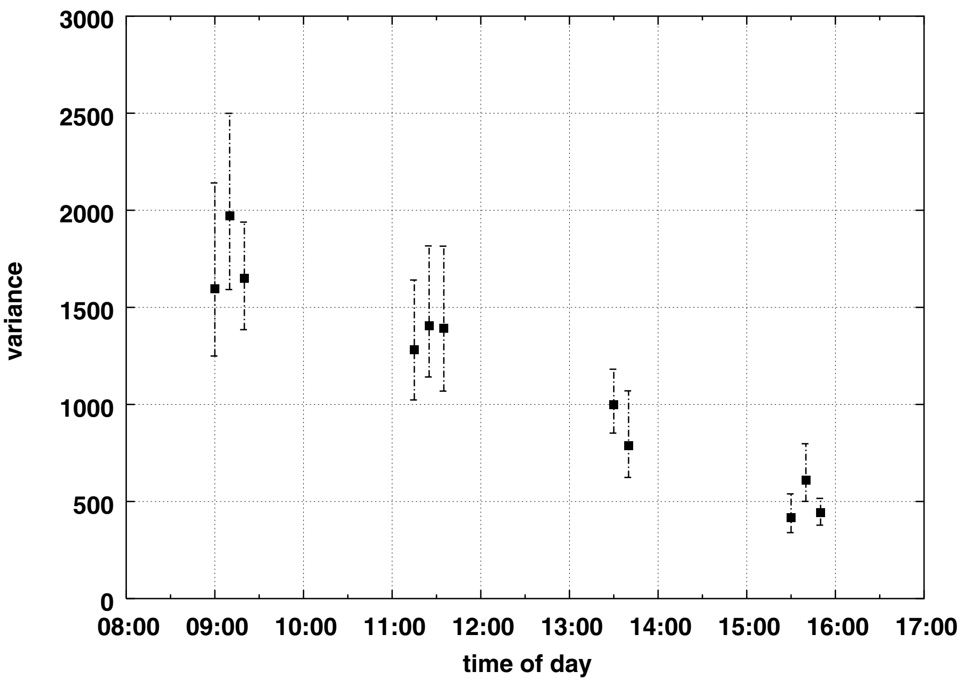

Figure 2 To detect a possible time of day dependency of the healthy hand tremor, the left hand acceleration was recorded at 4 different times. At each time, 3 repeated measurements were recorded whereas one measurement at 13.30 was excluded because of the presence of a drift. The 95%-confidence intervals of the variance suggests a time of day dependency and the consistency of the repeated measurements. The temporal distance of the repeated measurements are stretched for sake of clarity.

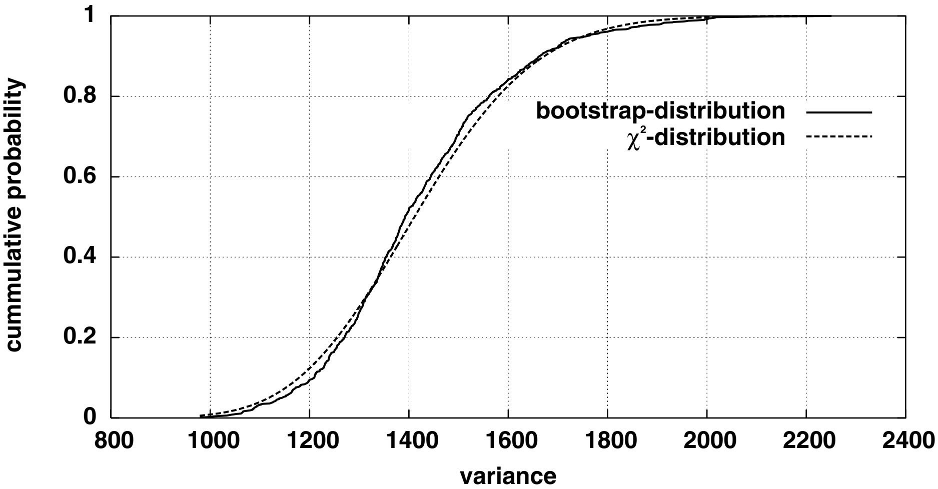

Figure 3 Comparison of the bootstrap-distribution (solid line) and its fitted χ^2-distribution (dashed line) for the dataset having the largest Kolmogrov-distance.

day, 9.00, 11.15, 13.30 and 15.30 the tremor of the outstretched left hand was recorded. For each TOD, 3 data sets were recorded to test the consistency of the measurements. The recording at 13.30 was excluded from the analysis because of the presence of a drift, which is due to a slow hand movement.

For a first inspection, 95%-confidence intervals of the sample variance for all datasets are estimated using the studentised blockwise bootstrap method. The results, Figure 2, clearly show a TOD dependence. The confidence intervals further suggest, that the repeated measurements are consistent.

For testing the hypothesis of the TOD dependence and the consistency of the repeated measurements statistically, we choose to parameterise the bootstrap distributions of the variance estimators by χ^2-distributions. The degrees of freedom of the χ^2-distributions are estimated by minimising the Kolmogorov distance. This parameterisation is motivated by the asymptotic behaviour of the variance estimator, when having an i.i.d. sample. Note, that there is no practicable statistical test procedure for testing the goodness of the chosen parametrisation, e.g. Kolmogorov-Smirnov test. Since the underlying bootstrap distribution is itself approximative and the amount of data is the chosen number of bootstrap samples, the test result can be controlled. Nevertheless, to demonstrate the usefulness of the chosen parametrisation, a comparison of the bootstrap distribution function and its fitted χ^2-distribution for the dataset having the largest Kolmogorov distance is given in Figure 3.

Finally, the parameterised distributions are log-transformed to yield a Gaussian error-model in good approximation. Now, a two factorial ANOVA has been carried out to test the hypotheses:

- No over-all-effect is present.
- There is no TOD dependence.
- The repeated measurements are consistent.

Choosing a level of significance of 1%, we can infer that an over-all-effect is present, which is the TOD dependence (both having p-values less than 10^{-5}). The hypothesis of the consistency of the repeated measurements cannot be rejected. Hence, the test results are in perfect accordance of the intuition imparted by the visual inspection of the confidence intervals shown in Figure 2.

8 Discussion and Conclusion

The problem of estimating distributions of real valued statistics when the observed data are statistically dependent rises in many areas of applied statistics. It is often not possible to derive a suitable model of the data generating process. In these cases non-parametric methods like bootstrap are widely used.

The proposed blockwise bootstrap is such a method, which can cope with the dependence structure of the observations. But some structural assumptions has to be fulfilled to achieve consistent results. Beside these assumptions, the free parameter of the blocklength has to be adjusted in order to minimise the bias and the variance of the approximation. The discussed method for selecting the blocklength is easy to apply and fast to compute. On three exemplary processes, the simulation study shows that the chosen blocklength gives suitable approximations. In comparing our blocklength selection method with existing alternatives by Hall et al. (1995) and by Bühlmann and Künsch (1999), we conclude that all methods are giving quite comparable results, whereby the distribution of the selected blocklength is sharper for proposed method.

The effect of studentising the statistic has also been studied. It turns out to use the studentised blockwise bootstrap is highly recommended to enhance the convergence rate. This result is in accordance with many past investigations of the bootstrap method.

To show the practical significance of the proposed method, we gave an application, in which a time-of-day dependence of the human hand tremor is tested. The test clearly confirms this dependence for the investigated subject. The discussed method seems to be appropriate for further studies in this area.

Acknowledgements We like to thank H. U. Voss and E. Mammen for suggestions on the manuscript. M. P. received financial support form the Deutsche Forschungsgemeinschaft (DFG). We also like to thank the anonymous referees for their careful review of the manuscript, which led to an improvement of the paper.

References

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74**, 457–468.

Bhattacharya, R., Ghosh, J. (1978). On the validity of the formal Edgeworth expansion. *Annals of Statistics* **6**, 434–451.

Boashash, B. (1992). Interpreting and estimating the instantaneous frequency of a signal-part I: Fundamentals; part II: Algorithms and applications. *Proceedings of the IEEE* **80**, 519–568.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3**, 123–148.

Bühlmann, P., Künsch, H. (1999). Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis* **31**, 295–310.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics* **14**, 1171–1179.

Davison, A., Hall, P. (1993). On studentizing and blocking methods for implementing the bootstrap with dependent data. *Australian Journal of Statistics* **35**, 215–224.

Davison, A., Hinkley, D. (1997). *Bootstrap methods and their application.* Cambridge University Press.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* **7**, 1–26.

Fisher, H., Hall, P. (1990). On bootstrap hypothesis testing. *Australian Journal of Statistics* **32**, 177–190.

Fitzenberger, B. (1996). The moving block bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics* **82**, 235–287.

Gabor, D. (1946). Theory of communication. *Journal of IEE* **93**, 429–457.

Gill, R. (1989). Non- and semi-parametric maximum likelihood estimators and the Mises method (part I). *Scandinavian Journal of Statistics* **16**, 97–128.

Götze, F., Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlichkeitstheorie verwandte Gebiete* **64**, 211–239.

Götze, F., Künsch, H. (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics* **24**, 1914–1933.

Hall, P. (1985). Resampling a coverage pattern. *Stochastic Processes and its Applications* **20**, 231–246.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Annals of Statistics* **16**, 927–985.

Hall, P. (1992). *The Bootstrap and the Edgeworth Expansion.* Springer.

Hall, P., Horowitz, J., Jing, B. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82**, 561–574.

Hall, P., Martin, M. (1988). On bootstrap resampling and iteration. *Biometrika* **75**, 661–671.

Hall, P., Wilson, S. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics* **47**, 757–762.

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

Huber, P. (1980). *Robust Statistics.* Wiley.

Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* **17**, 1217–1241.

Kurrer, C., Schulten, K. (1991). Effect of noise and perturbations on limit cycle systems. *Physica D* **50**, 311–320.

Leung, H. (1995). Stochastic transient of a noisy van der Pol oscillator. *Physica A* **221**, 340–347.

Liu, R., Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In: Lepage, R., Billard, L. (Eds.), *Exploring the Limits of Bootstrap.* Wiley, New York, pp. 225–248.

Naik-Nimbalkar, U., Rajarshi, M. (1994). Validity of blockwise bootstrap for empirical processes with stationary observations. *Annals of Statistics* **22**, 980–994.

Oppenheim, A., Schafer, R. (1975). *Digital Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ.

Paparoditis, E., Politis, D. (2001). The tapered block bootstrap for general statistics from stationary sequences. *Econometrics Journal* **5**, 131–148.

Paparoditis, E., Politis, D. (2003). Residual-based block bootstrap for unit root testing. *Econometrica* **3**, 813–855.

Peifer, M., Timmer, J., Voss, H. (2003). Non-parametric identification on non-linear oscillating systems. *Journal of Sound & Vibration* **267**, 1157–1167.

Press, W., Flannery, B., Saul, S., Vetterling, W. (1992). *Numerical Recipes.* Cambridge University Press, Cambridge.

Sherman, M. (1998). Efficiency and robustness in subsampling for dependent data. *Journal Statistical Planning & Inference* **75**, 133–146.

Timmer, J. (2000). Estimating parameters in nonlinear stochastic differential equations. *Chaos, Solitons & Fractals* **11**, 2571–2578.

Timmer, J., Lauk, M., Vach, W., Lücking, C. (1999). A test for a difference between spectral peak frequencies. *Computational Statistics & Data Analysis* **30**, 45–55.