Chapter 10
Resampling

10.1 General Introduction

Resampling or bootstrap methods refer to techniques where statistical inference is based on a simulated distribution of a statistic T_n obtained by resampling from an observed sample $X_1, \ldots, X_n$. Inference of this type is always conditional on the sample. In the most general version, no model assumptions are used except for global conditions such as stationarity, existence of some moments, etc. In the most restricted version, a parametric model is specified and resampling is used only as a simple way of obtaining an approximate distribution of T_n. Note that different terms such as 'bootstrap', 'resampling', 'subsampling', etc. are used in the literature for different variations of the same general idea. Since there does not seem to be a unified terminology, we use 'resampling' and 'bootstrap' as synonyms.

The original bootstrap (Efron 1979) was developed for i.i.d. data. Under the i.i.d. assumption, only the marginal distribution is unknown. Suppose, for instance, that we are interested in inference about the location parameter μ, given the observed data $\mathscr{Y}_n = (Y_1, \ldots, Y_n)$ where $Y_j = \mu + X_j \in \mathbb{R}$ and X_j are i.i.d. with distribution F_X. If we estimate μ by the sample mean $T_n = \bar{y}$, then we can write T_n as a functional $T_n(F_n) = \int u \, dF_n(u)$ of the empirical distribution function $F_n(x) = n^{-1} \sum 1\{Y_j \le x\}$. If the distribution function $F_Y(x) = F_X(x - \mu)$ of Y were known, then, in principle, the distribution of T_n could be calculated exactly by evaluating the n-dimensional integral $F_{T_n}(x) = P(T_n \le x) = \int_A dF_Y(y_1)\, dF_Y(y_2) \cdots dF_Y(y_n)$ where $A = \{y \in \mathbb{R}^n : y_1 + \cdots + y_n \le nx\}$. Usually, F_Y is unknown and is therefore replaced by an estimate $\hat{F}_Y$. One then has to evaluate $\hat{F}_{T_n}(x) = \hat{P}(T_n \le x) = \int_A d\hat{F}_Y(y_1)\, d\hat{F}_Y(y_2) \cdots d\hat{F}_Y(y_n)$. In most cases, the numerical evaluation of high dimensional integrals is difficult. The easiest alternative is Monte Carlo approximation which means that we approximate $\hat{F}_{T_n}$ by a simulated distribution, say $\hat{F}^*_{T_n}$, based on a sufficiently large sample of i.i.d. values $T^*_{n,1}, \ldots, T^*_{n,N}$ with $T^*_{n,j} \sim \hat{F}_{T_n}$. This can be done without actually computing $\hat{F}_{T_n}$ directly (after all that is what we wanted to avoid), namely by resampling. Independent samples $\mathscr{Y}^*_{n,j} = \{Y^*_{1,j}, \ldots, Y^*_{n,j}\}$ $(j = 1, 2, \ldots, N)$ are simulated and the

J. Beran et al., *Long-Memory Processes*,
DOI 10.1007/978-3-642-35512-7_10, © Springer-Verlag Berlin Heidelberg 2013

sample means $T^*_{n,j} = n^{-1}\sum_{i=1}^n Y^*_{i,j} = T_n(F^*_{n,j})$ (with $F^*_{n,j}$ denoting the empirical distribution function of $Y^*_{1,j}, \ldots, Y^*_{n,j}$) are computed. For each j, the values $Y^*_{i,j}$ $(i = 1, 2, \ldots, n)$ are obtained by simulating n independent realizations of a random variable $Y^* \sim \hat{F}_Y$. If $\hat{F}_Y$ is equal to the empirical distribution function F_n, then this is the same as drawing $Y^*_{i,j}$ $(i = 1, 2, \ldots, n)$ randomly with replacement (and equal probability n^{-1}) from the original set of observations $\{Y_1, \ldots, Y_n\}$.

Resampling procedures can thus be considered as a simulation device to obtain an approximate distribution function of a statistic T_n. It should be noted here that the sample mean is a relatively simple statistic because it can be expressed explicitly as a function of $Y_1, \ldots, Y_n$. Many estimators in statistics are defined by equations that do not lead to an explicit expression for T_n and F_{T_n}. For example, most non-Gaussian maximum likelihood estimators, M-estimators or minimum contrast estimators are defined as solutions of nonlinear equations for which no explicit solution exists. This makes resampling procedures even more useful because explicit expressions are not required.

The obvious question is how accurate a bootstrap approximation $\hat{F}^*_{T_n}$ of F_{T_n} is and, in fact, whether it works at all. Usually, if T_n is an appropriately standardized statistic, then it converges in distribution to a certain nondegenerate random variable $Z \sim F_Z$. For instance, in the i.i.d. example above, we may redefine T_n as $T_n = \sqrt{n}(\bar{y}_n - \mu)/\sigma$ which converges to a standard normal variable, provided that $\sigma^2 = \mathrm{var}(X_j)$ is finite. The asymptotic distribution F_Z is a natural competitor of the bootstrap approximation $\hat{F}^*_{T_n}$. Since F_Z is exactly correct asymptotically, the first requirement is that the same is true for $\hat{F}^*_{T_n}$. This is also called 'validity' of the bootstrap procedure. Thus, one needs to prove that $\hat{F}^*_{T_n}$ converges to F_Z as n tends to infinity. Once validity is shown, the next question is why we should prefer to use $\hat{F}^*_{T_n}$ instead of the asymptotic distribution F_Z. There are at least two possible reasons: (i) F_Z may be complicated or unknown, (ii) $\hat{F}^*_{T_n}$ may be more accurate than the asymptotic distribution F_Z.

The first reason is certainly relevant in the context of long-range dependence. For instance, under Gaussian subordination with Hermite rank two or higher, asymptotic distributions of normalized sums are marginals of non-Gaussian Hermite processes. These distributions are rather complicated and, in practice, we actually do not even know which one applies because the Hermite rank is an unknown quantity (in fact, we do not even know whether Gaussian subordination applies). Also, even in the case of a Gaussian limit (i.e. Hermite rank one), the exponent of n in the standardization is unknown and the normalizing constant (or even a slowly varying function) may be complicated. Resampling procedures based on self-normalized statistics that avoid explicit estimation of this exponent (and the constant or slowly varying function) provide a simple alternative to more explicit model based approaches. Other examples where F_Z may be complicated are encountered in the context of stable laws (see below).

To justify the second reason for using $\hat{F}^*_{T_n}$, namely improved accuracy, more refined asymptotic results are required since convergence of $\hat{F}^*_{T_n}$ to F_Z (which is

a basic prerequisite for considering $\hat{F}^*_{T_n}$ at all) does not automatically imply that, compared to F_Z, $\hat{F}^*_{T_n}$ is closer to the true finite sample distribution F_{T_n}. Suppose that $F_{T_n}(x) = F_Z(x) + a_n(x) + o(a_n)$ (with $a_n = o(1)$) and $F_{T_n}(x) = \hat{F}^*_{T_n}(x) + b_n(x) + o_p(b_n)$. (Note that in contrast to $a_n(x)$, $b_n(x)$ is random because $\hat{F}^*_{T_n}(x)$ is calculated conditionally on the observed sample.) For *validity* it is sufficient to show that $\tilde{b}_n = \sup_x |b_n(x)| = o_p(1)$. To prove that $\hat{F}^*_{T_n}$ is *more accurate* than F_Z, one needs to make a second order comparison. Such comparisons are usually based on Edgeworth expansions (see, e.g. Hall 1992). In many situations, it is indeed possible to show that $\tilde{b}_n = o_p(a_n)$ which means that the bootstrap error is of a smaller order than the one of the asymptotic approximation. The implications of such an improvement are often clearly visible. For instance, if F_X in the i.i.d. example above is highly skewed, then the distribution of $T_n = \sqrt{n}(\bar{y}_n - \mu)/\sigma$ can be highly skewed too, even for relatively large sample sizes. In such a case, an approximation by the standard normal distribution F_Z is inappropriate whereas a bootstrap distribution tends to mimic the asymmetry of F_{T_n} rather well.

The validity and accuracy of resampling techniques is fairly well understood in the i.i.d. case (see, e.g. Hall 1992; Politis et al. 1999; Lahiri 2003, and references therein). Once the assumption of independence is abandoned, further complications arise because the marginal distribution is not the only unknown quantity. In full generality, a statistic T_n is a functional of the complete joint n-dimensional distribution $F_{\mathscr{Y}_n}(y_1, \ldots, y_n) = P(Y_1 \le y_1, \ldots, Y_n \le y_n)$. The question how to resample from an observed series $\mathscr{Y}_n = (Y_1, \ldots, Y_n)$ is therefore much more difficult. First of all, we have one observation only (namely $\mathscr{Y}_n$ itself) from the n-dimensional distribution $F_{\mathscr{Y}_n}$ so that no consistent estimate of $F_{\mathscr{Y}_n}$ is available, unless certain assumptions are imposed. This is, of course, a general problem of statistical inference for stochastic processes, and led, already in the early days of time series analysis, to the introduction of properties such as stationarity and ergodicity. Most of the resampling theory for stochastic processes is concerned with the question under what kind of general conditions bootstrap works, which modifications are required to ensure validity and how to improve the second-order error. The original approach of drawing individual observations $Y^*_{i,j}$ $(i = 1, 2, \ldots, n)$ independently with replacement from $\{Y_1, \ldots, Y_n\}$ does not provide valid results in general because the dependence structure is removed completely by the resampling scheme.

There are two main ideas how to solve this problem. The first approach is to resample whole blocks $B_r = (Y_r, \ldots, Y_{r+l-1})$ of adjacent observations instead of individual values. By letting the block length l tend to infinity such that at the same time $l/n \to 0$, an infinite time horizon is captured ultimately within each block while at the same time the number of blocks (and thus the number of items to resample from) also tends to infinity. Methods of this type are also called block or blockwise bootstrap or subsampling. The problem is, of course, that in general F_{T_n} depends on the complete n-dimensional distribution $F_{\mathscr{Y}_n}$ whereas the subsampling procedure essentially relies on estimating the lower-dimensional probability function $F_{\mathscr{Y}_l}$. Although l tends to infinity, we also have $l = o(n)$. It is therefore not clear

a priori whether information about the dependence structure beyond lag l is asymptotically negligible when characterizing the distribution of $T_n(F_{\mathscr{Y}_n})$, and in how far it matters that $F_{\mathscr{Y}_l}$ actually has to be estimated as well. As it turns out, the main dividing line is between short and long memory. The validity and second-order accuracy of relatively simple versions of blockwise subsampling can be established under short-memory assumptions (Carlstein 1986; Künsch 1989; Politis and Romano 1993). This is not the case in general for long-memory processes although some modifications of blockwise resampling work under certain specific assumptions (see below).

A second approach to adapting bootstrap to dependent data consists of removing all or some of the dependence *before* applying resampling. Resampling methods based on this principle are subsumed under the name 'sieve bootstrap'. For instance, under the assumption that a causal linear process $Y_t = \sum a_j \varepsilon_{t-j}$ (with ε_t i.i.d.) is observed, one may use a sequence of autoregressive filters $\Phi_n(B) = 1 - \varphi_{1,n}B - \cdots - \varphi_{p_n,n}B^{p_n}$ with $p_n \to \infty$ and $\varphi_{j,n}$ estimated by minimizing the least squares criterium $\sum(Y_t - \Phi_n(B)Y_t)^2$. Resampling is then applied to the residual process $e_{t,n} = \Phi_n(B)Y_t$. Under suitable short-memory conditions, it can be shown that with $p_n \to \infty$ it is possible to approximate the actual i.i.d. residuals ε_t with sufficient accuracy (for early literature on autoregressive fitting with $p_n \to \infty$, see, e.g. Parzen 1974; Berk 1974; Hannan and Deistler 1988; also see Shibata 1980 for the connection to optimal prediction and Akaike's information criterion). Note that, if the order p_n is kept fixed, then we are relying on the stronger assumption that Y_t is generated by a finite-order autoregressive process. This is a special case of a 'parametric bootstrap'. Validity and second-order accuracy of the sieve bootstrap have been established under short-memory conditions (see, e.g. Bühlmann 1997, 2002, and references therein). In general, sieve methods rely on more restrictive assumptions than blockwise bootstrap because the choice of the preprocessing device has to be appropriate. On the other hand, if the assumptions are correct, then the sieve bootstrap tends to provide more accurate approximations (see, e.g. Choi and Hall 2000).

While both approaches (blockwise and sieve) are quite well understood under short-memory conditions, the situation is more difficult in the presence of long memory. Generally, the validity of standard blockwise methods no longer holds, unless specific modifications are applied (see, e.g. Lahiri 1993, 2003; Hall et al. 1998; Nordman et al. 2006). The easiest situation is encountered for the parametric bootstrap where not only validity but also improved second-order accuracy has been established for certain classes of estimators under long-memory conditions (see, e.g. Andrews et al. 2006; Andrews and Lieberman 2005). Similar results are available for the sieve bootstrap based on autoregressive fitting as above with $p_n \to \infty$ such that $n^{\frac{1}{2}-d}(\log n)^{\frac{1}{2}-d}p_n \to 0$ (Poskitt 2007a, 2007b). Note that the results in Poskitt (2007a, 2007b) are also interesting from the point of view of parameter estimation for a long-memory process because it is shown that the fitted AR-coefficients $\varphi_{j,n}$ converge to the coefficients a_j in the Wold representation with a simultaneous bound on the estimation error $|\varphi_{j,n} - a_j|$ $(j = 1, 2, \ldots, p_n)$. This is achieved without using fractional differencing or direct estimation of the fractional differencing parameter

d (in contrast to comparable AR-fitting methods such as Bhansali et al. 2006; see Sect. 5.9.3).

In the following sections, a few selected resampling methods will be discussed in more detail in the context of long-range dependence. For further literature on resampling methods and Edgeworth expansions for long-memory processes, see, e.g. Lahiri (2003), and references given in Lieberman et al. (2001, 2003), Giraitis and Robinson (2003), Faÿ et al. (2004), Lieberman and Phillips (2004), Andrews and Lieberman (2005), Nordman and Lahiri (2005), Andrews et al. (2006), McElroy and Politis (2007), Poskitt (2007a, 2007b), Jach et al. (2012), Kim and Nordman (2011).

10.2 Some Basics on Bootstrap for i.i.d. Data

Let $\mathscr{Y}_n = \{Y_1, \ldots, Y_n\}$ be a sample from the distribution F. Note that at this moment we do not assume any particular dependence structure of the original sequence Y_j ($j \in \mathbb{N}$), except that the marginal distribution is the same. The simplest bootstrap procedure starts with drawing a sample $Y_1^*, \ldots, Y_n^*$ with replacement from $\mathscr{Y}_n$. Conditionally on $\mathscr{Y}_n$, the random variables $Y_1^*, \ldots, Y_n^*$ are i.i.d., no matter what the original model is. Moreover,

$$P_*\big(Y_1^* = Y_j\big) := P\big(Y_1^* = Y_j | \mathscr{Y}_n\big) = 1/n, \quad j = 1, \ldots, n,$$

which means that the common (random) distribution function of Y_j^* ($j = 1, 2, \ldots, n$) is equal to the empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} 1\{Y_j \le x\}.$$

To keep things simple, we consider estimation of the expected value $\mu = E(Y_1)$ by the sample mean $\bar{Y}_n$. Denote by $\bar{Y}_n^* = n^{-1} \sum_{j=1}^n Y_j^*$ the bootstrap sample mean. Also, let E_* be the expectation w.r.t. P_*. We have the following moment properties:

$$E_*\big(Y_i^*\big) = \int x \, dF_n(x) = \frac{1}{n} \sum_{j=1}^{n} Y_i = \bar{Y}_n,$$

$$E_*\big(\bar{Y}_n^*\big) = E\big(\bar{Y}_n^* | \mathscr{Y}_n\big) = \frac{1}{n} \sum_{j=1}^{n} E(Y_j | \mathscr{Y}_n) = \bar{Y}_n,$$

$$E\big(\bar{Y}_n^*\big) = E\big[E\big(\bar{Y}_n^* | \mathscr{Y}_n\big)\big] = E(\bar{Y}_n) = E(Y),$$

$$\mathrm{var}_*\big(Y_i^*\big) = \int x^2 \, dF_n(x) - \left(\int x \, dF_n(x)\right)^2 = \frac{1}{n} \sum_{j=1}^{n} Y_j^2 - \left(\frac{1}{n} \sum_{j=1}^{n} Y_j\right)^2 =: s^2,$$

and recalling that Y_j^* are conditionally independent,

$$\operatorname{var}_*\big(\bar{Y}_n^*\big) = \frac{1}{n}\operatorname{var}_*\big(Y_1^*\big) = \frac{s^2}{n}. \tag{10.1}$$

Let us now focus on the case where $Y_1, \ldots, Y_n$ are i.i.d. observations with a finite variance. The standardized sample mean is asymptotically standard normal, i.e.

$$T_n = \frac{\bar{Y}_n - \mu}{\sqrt{\operatorname{var}(\bar{Y}_n)}} = \sqrt{n}\frac{\bar{Y}_n - \mu}{\sqrt{\operatorname{var}(Y_1)}} \xrightarrow{\mathrm{d}} N(0,1). \tag{10.2}$$

In the bootstrap approach, the initial population one sampled from is replaced by $\mathscr{Y}_n$. Thus, the bootstrap version of T_n is obtained by replacing $\bar{Y}_n$ by the bootstrap sample mean $\bar{Y}_n^*$, the population mean μ by the bootstrap population mean $E_*(Y_1^*) = \bar{Y}_n$, and the population variance $\operatorname{var}(Y_1)$ by the bootstrap population variance $\operatorname{var}_*(Y_1^*) = s^2$. The bootstrap version of T_n is therefore given by

$$T_n^* = \frac{\bar{Y}_n^* - E_*(Y_1^*)}{\sqrt{\operatorname{var}_*(\bar{Y}_n^*)}} = \sqrt{n}\frac{\bar{Y}_n^* - E_*(Y_1^*)}{\sqrt{\operatorname{var}_*(Y_1^*)}} = \sqrt{n}\frac{\bar{Y}_n^* - \bar{Y}_n}{s}. \tag{10.3}$$

Since $\bar{Y}_n$ converges in probability to μ and the denominator converges in probability to $\sqrt{\operatorname{var}(Y)}$, T_n^* has the same behaviour as T_n asymptotically. More specifically, the following lemma justifies validity of the bootstrap for i.i.d. data with a finite variance (see, e.g. Lahiri 2003, Theorem 2.1).

Lemma 10.1 *Assume that $Y_1, \ldots, Y_n$ are i.i.d. with $\operatorname{var}(Y_i) < \infty$. Then*

$$\sup_x \big| P_*\big(T_n^* \le x\big) - \Phi(x) \big| = o_p(1),$$

where $\Phi(x)$ is the standard normal distribution.

10.3 Self-normalization

Consider $Y_j = \mu + X_j$ ($j \in \mathbb{N}$) with X_j a stationary zero-mean sequence and assume that after suitable standardization the sample mean converges to a nondegenerate random variable Z, or in other words,

$$T_n := \frac{\sum_{j=1}^n Y_j - n\mu}{v_n} = \frac{n}{v_n}(\bar{Y}_n - \mu) \xrightarrow{\mathrm{d}} Z \sim F_Z \tag{10.4}$$

where F_Z is a nondegenerate distribution. Usually, the choice of v_n is $v_n^2 = \operatorname{var}(\sum_{j=1}^n Y_j)$, provided that this quantity exists. In the i.i.d. case with finite variance, we have $v_n^2 = n \cdot \operatorname{var}(X_1)$ and Z standard normal. Usually, v_n has to be estimated. In some situations, v_n is not even computable or requires an additional

estimation step. For example, if the random variables X_j are i.i.d. with a regularly varying distribution with index $-\alpha$ ($\alpha \in (0,2)$), then $v_n = n^{1/\alpha} L(n)$ where $L(n)$ is a slowly varying function, and Z is a stable random variable. Thus, in principle, we would need to estimate α (and even the slowly varying function L) before computing T_n. Often, it is possible to replace v_n by a data-based normalizer V_n without explicit estimation of model specific quantities, such as α or L. For example, for i.i.d. data (both with finite and infinite variance), we can replace v_n by the square root of $V_n^2 = n^{-1} \sum_{j=1}^n (Y_j - \bar{Y}_n)^2$. Given a data-based normalizer V_n we then consider the 'self-normalized' statistic

$$U_n := \frac{\sum_{j=1}^n Y_j - n\mu}{V_n} = \frac{n}{V_n}(\bar{Y}_n - \mu). \tag{10.5}$$

The choice of the normalizer V_n has to be modified for dependent sequences to guarantee that V_n/v_n converges to one in probability.

Denote by Z_0 the limit of U_n. If Z in (10.4) is normal, then Z_0 is also a standard normal variable. In general, however, the distributions of Z and Z_0 can be quite complicated, and may even differ. For example, if the data are i.i.d. with infinite variance, then Z is a stable random variable, but Z_0 is different. To see this, assume that X_j ($j \in \mathbb{N}$) are i.i.d. and regularly varying with index $-\alpha$. Consider

$$W_n := \frac{n}{V_n}(\bar{Y}_n - \mu) \tag{10.6}$$

where

$$V_n^2 = \sum_{j=1}^n (Y_j - \bar{Y}_n)^2.$$

We note that the random variables Y_j^2 ($j \in \mathbb{N}$) are regularly varying with index $-\alpha/2$ and thus have an infinite mean. In particular, $n^{-2/\alpha} \sum_{j=1}^n (Y_j - \bar{Y}_n)^2$, and hence $n^{-1/\alpha} V_n$ converges to a stable random variable. This implies that

$$W_n = \frac{n^{-1/\alpha} \sum_{j=1}^n (Y_j - \mu)}{n^{-1/\alpha} V_n}$$

converges to a ratio R of two dependent stable random variables. In principle, we may use this information to construct confidence intervals for μ of the form

$$\left[\bar{Y}_n - z_{1-\frac{1}{2}p_0} n^{-1} V_n, \bar{Y}_n - z_{\frac{1}{2}p_0} n^{-1} V_n\right],$$

where z_p denotes the $(100p)$th percentile of R. However, these percentiles may not be easily computable. Resampling methods are useful to overcome this problem.

10.4 The Moving Block Bootstrap (MBB)

Lemma 10.1 provides validity of the bootstrap procedure in the case of i.i.d. data with existing second moments. Now we turn our attention to the case of dependent data. Assume that $Y_j = \mu + X_j$ ($j \in \mathbb{N}$) is a stationary sequence of random variables with short memory and $\sigma^2 := \mathrm{var}(Y) < \infty$. Then convergence (10.2) has to be replaced by

$$\sqrt{n}\frac{\bar{Y}_n - \mu}{\sigma_0} \xrightarrow{\mathrm{d}} N(0,1), \tag{10.7}$$

with

$$\sigma_0^2 = \mathrm{var}(Y) + 2\sum_{k=1}^{\infty} cov(Y_0, Y_k). \tag{10.8}$$

However, as mentioned above, sampling with replacement from $\mathscr{Y}_n$ produces conditionally independent random variables. Therefore, if we use T_n^* defined in (10.3), then the result in Lemma 10.1 still applies. This contradicts (10.7) so that the bootstrap procedure is no longer valid (except in the special case of uncorrelated observations). The asymptotic variance of bootstrap replicates is wrong by the factor $(\sigma_0/\sigma)^2$. The reason is that the bootstrap procedure cannot recreate $\mathrm{var}(\bar{Y}_n)$. More exactly, recall that $\mathrm{var}_*(\bar{Y}_n^*) = s^2/n$ (10.1). The expected value of the conditional variance is then equal to

$$E\big[\mathrm{var}_*\big(\bar{Y}_n^*\big)\big] = \frac{1}{n}\left\{E\big(Y^2\big) - \frac{1}{n^2}\sum_{j,j'=1}^{n} E(Y_j Y_{j'})\right\} = \frac{1}{n}\big\{\mathrm{var}(Y) - \mathrm{var}(\bar{X}_n)\big\}. \tag{10.9}$$

Since $\mathrm{var}(\bar{X}_n) \to 0$ (except for degenerate cases that are not of interest here), the expected variance is approximately equal to

$$E\big[\mathrm{var}_*\big(\bar{Y}_n^*\big)\big] \sim \frac{1}{n}\mathrm{var}(Y) = \frac{\sigma^2}{n}.$$

This is in contrast to

$$\mathrm{var}(\bar{Y}_n) \sim \frac{\sigma_0^2}{n}.$$

To obtain a valid bootstrap procedure, a suitable modification is required. One of the possible solutions is the so-called *Moving Block Bootstrap* (*MBB*) (Carlstein 1986; Künsch 1989). To preserve most of the dependence structure, we sample (with replacement) blocks $B_1^*, \ldots, B_k^*$ from the set of all available blocks $B_r = (Y_r, \ldots, Y_{r+l-1})$ ($r = 1, \ldots, N_b$; $N_b = n - l + 1$) instead of sampling single observations. A bootstrapped sample $Y_1^*, \ldots, Y_n^*$ is generated by pasting $k = [n/l]$ sampled blocks $B_1^*, \ldots, B_k^*$ next to each other. Note that, by definition, $B_r^* = (Y_{(r-1)l+1}^*, \ldots, Y_{rl}^*)$ ($r = 1, \ldots, k$). For example, if $k = 2$ and blocks, say, B_1 and

B_3 are selected, then the bootstrap sample is

$$\left(Y_1^*, \ldots, Y_l^*, Y_{l+1}^*, \ldots, Y_{2l}^*\right) = (Y_1, \ldots, Y_l, Y_3, \ldots, Y_{l+2}).$$

Also note that the actual length of the bootstrapped series is $\tilde{n} = kl = [n/l]l$ (where $[n/l]$ denotes the largest integer not exceeding n/l), but the difference between $\tilde{n}$ and n is negligible asymptotically. We will therefore write $n = kl$ for simplicity. Denote by

$$\zeta_r = \zeta_{r,l} = \sum_{j \in B_r} Y_j = \sum_{j=r}^{r+l-1} Y_j$$

$(r = 1, 2, \ldots, N_b)$ the block sums and by

$$\zeta_r^* = \zeta_{r,l}^* = \sum_{j \in B_r^*} Y_j^* = \sum_{j=(r-1)l+1}^{rl} Y_j^*$$

the corresponding bootstrapped quantities (the index l will be dropped unless it needs to be emphasized). The bootstrap mean is given by

$$\bar{Y}_n^* = n^{-1} \sum_{j=1}^{n} Y_j^* = \frac{1}{k} \sum_{r=1}^{k} \frac{1}{l} \zeta_{r,l}^* = \frac{1}{k} \sum_{r=1}^{k} \left(\frac{1}{l} \sum_{j=(r-1)l+1}^{rl} Y_j^* \right).$$

When drawing block B_r^*, each of the blocks B_s $(s = 1, \ldots, N_b)$ has the same probability of being chosen. Thus, for any $r \in \{1, \ldots, k\}$,

$$P_*\left(B_r^* = B_s\right) = \frac{1}{N_b} \quad (s = 1, \ldots, N_b) \tag{10.10}$$

so that

$$\begin{aligned} E_*\left(\bar{Y}_n^*\right) &= E_*\left[\frac{1}{k} \sum_{r=1}^{k} \left(\frac{1}{l} \sum_{Y_j^* \in B_r^*} Y_j^* \right) \right] \\ &= \frac{1}{N_b} \sum_{r=1}^{N_b} \left(\frac{1}{l} \sum_{Y_j \in B_r} Y_j \right) = \frac{1}{N_b l} \sum_{r=1}^{N_b} \sum_{j=r}^{r+l-1} Y_j \\ &= \frac{1}{N_b} \sum_{r=1}^{N_b} \frac{1}{l} \zeta_{r,l}. \end{aligned}$$

Note that, if $l/n \to 0$ fast enough, then $E_*(\bar{Y}_n^*)$ may be approximated by the sample mean $\bar{Y}_n$ because all variables Y_j occur in the sum l times except for l observations on the left and right border, respectively.

Now, recalling that the blocks are conditionally independent, the conditional variance of the bootstrap mean is

$$\begin{aligned}\operatorname{var}_*\big(\bar{Y}_n^*\big) &= \operatorname{var}_*\left(\frac{1}{k}\sum_{r=1}^{k}\frac{1}{l}\zeta_r^*\right)\\ &= \frac{k}{(kl)^2}\operatorname{var}_*\big(\zeta_r^*\big) = \frac{k}{(kl)^2}\operatorname{var}_*\left(\sum_{j=1}^{l} Y_j^*\right)\\ &= \frac{k}{n^2}\left\{\frac{1}{N_b}\sum_{r=1}^{N_b}\left(\sum_{j=r}^{r+l-1} Y_j\right)^2 - \left(\frac{1}{N_b}\sum_{r=1}^{N_b}\sum_{j=r}^{r+l-1} Y_j\right)^2\right\}.\end{aligned}$$

For the unconditional expected value of the variance, we may assume, without loss of generality, that $\mu = 0$. Then the second term does not contribute asymptotically, and we obtain

$$E\big[\operatorname{var}_*\big(\bar{Y}_n^*\big)\big] \sim \frac{k}{n^2}E\left[\left(\sum_{j=1}^{l} Y_j\right)^2\right] = \frac{1}{nl}\operatorname{var}\left(\sum_{j=1}^{l} Y_j\right).$$

If the stationary sequence Y_j has short memory and $n, l \to \infty$ such that $l/n \to 0$, this leads to

$$E\big[\operatorname{var}_*\big(\bar{Y}_n^*\big)\big] \sim \frac{1}{nl}\sigma_0^2 l = \frac{\sigma_0^2}{n},$$

where σ_0 is given in (10.8). Therefore, the bootstrap variance of the bootstrap mean is asymptotically the same as $\operatorname{var}(\bar{Y}_n)$ and the MBB bootstrap statistic

$$T_n^* = \frac{\bar{Y}_n^* - E_*(Y_1^*)}{\sqrt{\operatorname{var}_*(\bar{Y}_n^*)}}$$

has the same asymptotic distribution as $T_n = (\bar{Y}_n - \mu)/\sigma_0$.

However, if the random variables X_j ($j \in \mathbb{N}$) are Gaussian with autocovariance function $\gamma_X(k) \sim L_\gamma k^{2d-1}$ ($0 < d < \frac{1}{2}$), then

$$\operatorname{var}(\bar{Y}_n) \sim n^{-2}v_n^2$$

and

$$T_n = \frac{n(\bar{Y}_n - \mu)}{v_n} \xrightarrow{\mathrm{d}} N(0,1)$$

where $v_n^2 = n^{2d+1} L_S$ with $L_S = C_1 L_\gamma$ (Sect. 4.2.2). On the other hand,

$$E\big[\mathrm{var}_*(\bar{Y}_n^*)\big] \sim \frac{k}{n^2} E\left(\sum_{j=1}^{l} Y_j\right)^2 = \frac{1}{nl}\,\mathrm{var}\left(\sum_{j=1}^{l} Y_j\right)$$
$$\sim C\frac{1}{nl} l^{2d+1} = C\frac{l^{2d}}{n}.$$

Thus

$$\frac{E[\mathrm{var}_*(\bar{Y}_n^*)]}{\mathrm{var}(\bar{Y}_n)} \sim \mathrm{const}\left(\frac{l}{n}\right)^{2d} \to 0$$

and

$$\frac{\bar{Y}_n^* - E_*(\bar{Y}_n^*)}{\sqrt{\mathrm{var}(\bar{Y}_n)}} = \frac{\bar{Y}_n^* - E_*(\bar{Y}_n^*)}{\sqrt{\mathrm{var}_*(\bar{Y}_n^*)}} \sqrt{\frac{\mathrm{var}_*(\bar{Y}_n^*)}{\mathrm{var}(\bar{Y}_n)}} = T_n^* \sqrt{\frac{\mathrm{var}_*(\bar{Y}_n^*)}{\mathrm{var}(\bar{Y}_n)}} \to 0.$$

This means that the MBB bootstrap heavily underestimates the variability of the sample mean $\bar{Y}_n$ such that the asymptotic coverage probabilities of bootstrap confidence intervals for μ are zero. The reason is that too much of the long-memory property is lost by pasting together independent blocks. In the short-memory case, the rate of $\sum_{t=1}^{n} Y_t$ is $O_p(\sqrt{n})$ which is the same as for i.i.d. data, and therefore also the same as for $\sum_{r=1}^{k} \zeta_r^* = O_p(\sqrt{kl})$ with $kl = n$. The error in the standardization is only a multiplicative constant that can be made arbitrarily small by letting l tend to infinity. This is no longer the case under long memory because independent sampling of blocks changes the *rate* of the original sum $\sum_{t=1}^{n} Y_t = O_p(n^d \cdot n^{\frac{1}{2}})$ to the smaller rate of the bootstrapped sum given by $\sum_{r=1}^{k} \zeta_r^* = O_p(k^{\frac{1}{2}} l^{d+\frac{1}{2}}) = O_p(l^d \cdot n^{\frac{1}{2}})$.

A simple remedy to make the MBB bootstrap work in the long-memory context is suggested in Lahiri (1993). Instead of using the sample mean directly, we consider a statistic that takes into account independence introduced by blockwise resampling. This can be done by adjusting the standardization accordingly. As before $k = [n/l]$ blocks $B_1^*, \ldots, B_k^*$ are sampled independently with replacement, but we now consider the correctly standardized statistic

$$\tilde{T}_n^* = k^{-\frac{1}{2}} \sum_{r=1}^{k} \frac{\zeta_r^* - l \cdot E_*(Y_1^*)}{v_l}$$
$$= k^{-\frac{1}{2}} \sum_{r=1}^{k} l^{-d-\frac{1}{2}} \frac{\zeta_r^* - l \cdot E_*(Y_1^*)}{\sqrt{C_1 L_\gamma}},$$

or

$$\tilde{T}_n^* = k^{-\frac{1}{2}} \sum_{r=1}^{k} l^{-d-\frac{1}{2}} \frac{\zeta_r^* - l \cdot \bar{Y}_n}{\sqrt{C_1 L_\gamma}}.$$

Since $\tilde{T}_n^*$ is equal to $k^{-\frac{1}{2}}$ times a sum of k independent equally distributed standardized variables, the central limit theorem holds and one can even show uniform convergence (Lahiri 1993)

$$\sup_{x \in \mathbb{R}} \left| P_* \left(\tilde{T}_n^* \leq x \right) - \Phi(x) \right| = o_p(1).$$

This result has to be interpreted with care, however, because we are dealing with the case of long memory. For instance, consider the Gaussian subordination model $Y_j = \mu + G(X_j)$ where X_j is a stationary Gaussian process with $E(X_j) = 0$, $\mathrm{var}(X_j) = 1$, $E[G(X_j)] = 0$ and autocovariance function $\gamma_X(k) \sim L_\gamma(k)|k|^{2d_X - 1}$ (as $k \to \infty$) for some $0 < d_X < \frac{1}{2}$. If G has Hermite rank one, then the standardized sample mean converges to a standard normal variable and the standardization is the same as in $\tilde{T}_n^*$. In this sense, validity of the modified MBB procedure is established. However, if G has a Hermite rank m higher than one and $d_X > \frac{1}{2}(1 - m^{-1})$, then the asymptotic limit of the standardized sample mean is non-Gaussian. This means that the modified MBB is no longer valid. The question then arises why the modified MBB should be used at all. The reason is obviously not a complicated asymptotic distribution since validity holds only in the case where the asymptotic distribution is normal. As discussed previously, another possible motivation for using resampling is a better approximation of finite sample distributions. In how far the conditional distribution of $\tilde{T}_n^*$ does indeed provide a better approximation of the distribution of T_n has not yet been fully explored in the long-memory context. However, the idea of a modified MBB can be extended to other problems where the definition of a bootstrap based statistic with known asymptotic distribution is useful in its own right. For instance, Beran and Shumeyko (2012b) develop an MBB based test of the null hypothesis that a nonparametric trend function is continuous (see Sect. 10.7.2 below).

10.5 The Sampling Window Bootstrap (SWB)

As we saw above, the modified MBB is not valid under Gaussian subordination unless the Hermite rank of G is one. The reason is that independent sampling of blocks automatically entails the central limit theorem, independently of the Hermite rank. A natural idea to solve this problem is to avoid independent resampling. In the so-called sampling window (SW) approach, independent sampling of blocks is replaced by including all available blocks with equal weight in an empirical distribution function.

To be specific, we consider as before estimation of μ for the process $Y_j = \mu + G(X_j)$ where G has Hermite rank m, and X_j ($j \in \mathbb{N}$) is a stationary Gaussian

sequence with $E(X_j) = 0$, $\mathrm{var}(X_j) = 1$ and autocovariances $\gamma_X(k) \sim L_\gamma(k)k^{2d_X-1}$ with $L_\gamma(k) = c_\gamma > 0$, $\frac{1}{2}(1 - m^{-1}) < d < \frac{1}{2}$. From Theorem 4.4 we have

$$n^{-(1-m(\frac{1}{2}-d))}L_S^{-1/2}\left(\sum_{j=1}^{n} Y_j - n\mu\right) \xrightarrow{\mathrm{d}} \frac{J(m)}{m!}Z_{m,H}(1), \tag{10.11}$$

where $L_S = J^2(m)/m!C_m c_\gamma^m$, $v_n = n^{1-m(\frac{1}{2}-d)}L_S^{1/2}$ and $Z = Z_{m,H}(1)$. As before, the replicates $T^*_{n,1}, \ldots, T^*_{n,N_b}$ are based on standardized sums over blocks $B_r = (Y_r, \ldots, Y_{r+l-1})$ $(r = 1, 2, \ldots, N_b)$ of length l. However, instead of resampling blocks independently and pasting them together, we use all N_b (partially overlapping) blocks to obtain the empirical distribution function

$$F^*_{T_n}(x) = \frac{1}{N_b}\sum_{r=1}^{N_b} 1\{T^*_{n,r} \le x\} = \frac{1}{N_b}\sum_{r=1}^{N_b} 1\left\{\frac{S_{n,l,r} - l\bar{Y}_n}{v_l} \le x\right\}$$

with

$$T^*_{n,r} := T^*_{n,l,r} := \frac{\sum_{j=r}^{r+l-1} Y_j - l\bar{Y}_n}{v_l} = \frac{S_{n,l,r} - l\bar{Y}_n}{v_l} \quad (r = 1, 2, \ldots, N_b).$$

By assigning equal weights to all available blocks and avoiding any kind of random reshuffling of the sequence, the complete dependence structure can essentially be preserved. Why this is so can be seen in more detail as follows. Recall that, as $n \to \infty$, $F_{T_n}(x) = P(T_n \le x) \to F_Z(x) := P(Z_{m,H}(1) \le x)$ for all $x \in \mathbb{R}$ and note that $E[F^*_{T_n}(x)] = P(T^*_{n,l,1} \le x)$. We will prove (Hall et al. 1998):

Theorem 10.1 *Let X_j be as defined above, and $l, n \to \infty$ such that $l/n \to 0$. Then*

$$\sup_{x\in\mathbb{R}} \left|F^*_{T_n}(x) - F_{T_n}(x)\right| \xrightarrow{\mathrm{p}} 0. \tag{10.12}$$

Proof In the first step, we will replace $\bar{Y}_n$ by μ in the definition of $F^*_{T_n}(x)$. To justify this, we note that with $\tilde{T}_{n,l,r} = (S_{n,l,r} - l\mu)/v_l$ we have

$$\tilde{T}_{n,l,r} - T^*_{n,l,r} = \frac{l}{v_l}(\bar{Y}_n - \mu) = \frac{lv_n}{nv_l}T_n.$$

On account of (10.11), T_n converges in distribution to the finite random variable $Z = Z_{m,H}(1)$. Furthermore,

$$\frac{lv_n}{nv_l} \to 0$$

since it was assumed that $l, n \to \infty$ and $l/n \to 0$.

The next useful fact is that both $F_{T_n}(x)$ and $F_{T_l}(x)$ converge to $F_Z(x)$ as $n, l \to \infty$. It is therefore sufficient to prove

$$\sup_{x\in\mathbb{R}} \big|\tilde{F}_{T_n}(x) - F_{T_l}(x)\big| \xrightarrow{\mathrm{p}} 0,$$

where

$$\tilde{F}_{T_n}(x) = \frac{1}{N_b}\sum_{r=1}^{N_b} 1\{\tilde{T}_{n,l,r} \le x\} = \frac{1}{N_b}\sum_{r=1}^{N_b} 1\big\{(S_{n,l,r} - l\mu)/v_l \le x\big\}.$$

We note that $E[\tilde{F}_{T_n}(x)] = P(T_l \le x) = F_{T_l}(x)$. Therefore,

$$\begin{aligned} E\big[\big(\tilde{F}_{T_n}(x) - F_{T_l}(x)\big)^2\big] &= \mathrm{var}\big(\tilde{F}_{T_n}(x)\big) = \frac{1}{N_b}\,\mathrm{var}\big(1\{\tilde{T}_{n,l,1} \le x\}\big) \\ &\quad + \frac{2}{N_b}\sum_{r=2}^{l} cov\big(1\{\tilde{T}_{n,l,1} \le x\}, 1\{\tilde{T}_{n,l,r} \le x\}\big) \\ &\quad + \frac{2}{N_b}\sum_{r=l+1}^{N_b} cov\big(1\{\tilde{T}_{n,l,1} \le x\}, 1\{\tilde{T}_{n,l,r} \le x\}\big) \\ &\le \frac{1}{N_b} + \frac{2l}{N_b} + \frac{2}{N_b}\sum_{r=l+1}^{N_b} cov\big(1\{\tilde{T}_{n,l,1} \le x\}, 1\{\tilde{T}_{n,l,r} \le x\}\big). \end{aligned}$$

Now, let us consider the case $m = 1$ only, so that v_l^2 is proportional to l^{2d_X+1}. Then the random variables $\tilde{T}_{n,l,r}$, $r = l+1, \ldots, N_b$, are centred Gaussian and w.l.o.g. we can assume that they have unit variance (formally, $\mathrm{var}(\tilde{T}_{n,l,r}) \sim 1$ as $l \to \infty$). Note that for a standardized bivariate normal vector $Z = (Z_1, Z_2)$ we have

$$\big|cov(Z_1, Z_2)\big| = \big|corr(Z_1, Z_2)\big| \ge \big|cov\big(1\{Z_1 \le x\}, 1\{Z_2 \le x\}\big)\big|.$$

Moreover, the separation between blocks B_1 and B_r is $r - l$. Therefore,

$$\begin{aligned} &\frac{2}{N_b}\sum_{r=l+1}^{N_b} Cov\big(1\{\tilde{T}_{n,l,1} \le x\}, 1\{\tilde{T}_{n,l,r} \le x\}\big) \\ &\quad \le \frac{2}{N_b}\sum_{r=l+1}^{N_b}\sum_{j=1}^{l}\sum_{j'=r}^{r+l-1} \gamma_X(j' - j) \\ &\quad \le \frac{2l^2}{N_b v_l^2}\sum_{r=l+1}^{N_b} \gamma_X(r-l) \sim C\frac{2l^2}{N_b v_l^2} N_b^{2d} \sim C\frac{l^{1-2d}}{N_b^{1-2d}} = C\left(\frac{l}{n-l+1}\right)^{1-2d} \to 0 \end{aligned}$$

as $l, n \to \infty$ such that $l/n \to 0$.

The arguments for $m > 1$ are analogous, but covariances between Hermite polynomials of higher order have to be considered. □

We conclude that the empirical distribution $F^*_{T_n}(x)$ is a consistent estimator of the limiting distribution $F_Z(x)$ so that the SW bootstrap is a valid procedure under Gaussian subordination with arbitrary Hermite rank. This is in contrast to the MBB bootstrap which is valid for Hermite rank one only. Since the SW approach preserves non-Gaussianity, one may also hope that it will provide better finite sample approximations even in the case of a Gaussian limit. Some examples in the next section illustrate this conjecture.

Remark 10.1 This theorem is adapted from Hall et al. (1998); see also Lahiri (2003, Theorem 10.4). We note that the authors consider a general form of $\gamma_X(k)$ with a possible slowly varying function. It requires slightly modified assumptions on the length l of the blocks. Furthermore, Theorem 2.4 in Hall et al. (1998) implies that it is enough to prove (10.12) for a fixed x.

Remark 10.2 The proof above also works for weakly dependent random variables (informally, when $d = 0$), and under Gaussian subordination with $0 < d_X < \frac{1}{2}(1 - m^{-1})$.

So far, we assumed that the standardization sequence v_n is known. In practice, this is, of course, not the case because $v_n = n^{(1-m(\frac{1}{2}-d_X))} L_S^{1/2}$ depends on the long-memory parameter d_X and the constant $L_\gamma(n) \equiv c_\gamma$. There are at least two possible solutions to this problem. The first one is to estimate the parameters d_X and c_γ directly by fitting a parametric or semiparametric model (see Sects. 5.5, 5.6, 5.7, 5.8 and 5.9). The standardization v_n is then replaced by $\hat{v}_n = n^{\hat{d}_X + \frac{1}{2}} \hat{L}_S^{1/2}$. Note, however, that in general the true Hermite rank m is not known. Nevertheless, if m is larger than one, then the exponent of n can also be estimated by the same methods. The difference is that we are then not estimating d_X but rather $\tilde{d} = (1 - m(\frac{1}{2} - d_X)) - \frac{1}{2}$. The other solution is to replace v_n by a direct fully nonparametric estimate V_n. Thus, we consider the statistics

$$U_n := \frac{\sum_{j=1}^n Y_j - n\mu}{V_n} = \frac{n(\bar{Y}_n - \mu)}{V_n},$$

and, with the blocks defined as before,

$$U^*_{n,r} := U^*_{n,l,r} := \frac{\sum_{j=r}^{r+l-1} Y_j - l\bar{Y}_n}{V_l} = \frac{S_{n,l,r} - l\bar{Y}_n}{V_l} \quad (r = 1, \ldots, N_b).$$

Note that, compared to the previous parametric or semiparametric estimation of v_n, direct estimators of v_n are more general, but at the same also less efficient, if the model assumptions needed for estimating d and L_γ by parametric or semiparametric

methods hold. A possible, though somewhat arbitrary, choice is, for instance,

$$V_l^2 = V_{n,l}^2 = \frac{E_{n,l,m_1}^4}{E_{n,l,m_2}^2}$$

where

$$E_{n,l,m_i}^2 = \frac{1}{l - m_i + 1} \sum_{j=1}^{l-m_i+1} (S_{n,m_i,j} - m_i \bar{Y}_n)^2$$

and

$$S_{n,m_i,j} = \frac{1}{m_i} \sum_{h=j}^{j+m_i-1} Y_h.$$

The crucial part of this construction is that

$$\frac{V_n^2}{\operatorname{var}(\sum_{j=1}^n Y_j)} = \frac{E_{n,n,m_1}^4}{v_n^2 E_{n,n,m_2}^2} \xrightarrow{\text{p}} 1$$

as $n \to \infty$. Therefore, the limiting distribution of $U_n = nV_n^{-1}(\bar{Y}_n - \mu)$ is the same as that of $T_n = nv_n^{-1}(\bar{Y}_n - \mu)$, namely $F_Z(x) = P(Z_{m,H}(1) \le x)$. We state the following result without proof (see Hall et al. 1998 or Lahiri 2003, Theorem 10.5).

Theorem 10.2 *Assume that X_j ($j \in \mathbb{N}$) is a stationary sequence of standard normal random variables, such that $\gamma_X(k) \sim L_\gamma k^{2d-1}$, $d \in (0, 1/2)$. Let*

$$F_{U_n}^*(x) = \frac{1}{N} \sum_{r=1}^{N} 1\{U_{n,l,r}^* \le x\}$$

and $F_{U_n}(x) = P(U_n \le x)$. If $l, n \to \infty$ such that $l/n \to 0$, then, as $n \to \infty$,

$$V_n^2 / \operatorname{var}\left(\sum_{j=1}^n Y_j\right) = V_n^2 / v_n \xrightarrow{\text{p}} 1$$

and

$$\sup_{x \in \mathbb{R}} \left| F_{U_n}^*(x) - F_{U_n}(x) \right| \xrightarrow{\text{p}} 0. \tag{10.13}$$

Combining Theorems 10.1 and 10.2 implies that the empirical distribution function $F_{U_n}^*(x)$ approximates $F_{U_n}(x)$ which in turn approximates $F_Z(x) = \lim_{n\to\infty} P(T_n \le x)$. Thus, validity of the SW bootstrap based on U_n is also established.

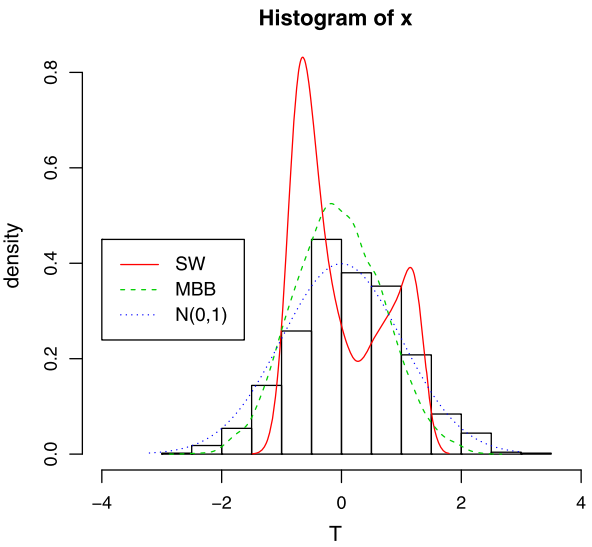

Fig. 10.1 Histogram of a simulated series $Y_t = G(X_t)$ of length $n = 1000$, where X_t is a FARIMA$(0, 0.4, 0)$ process with variance one and $G(x) = x + 0.005(x^3 - x)$. Also plotted are distributions obtained by blockwise bootstrap with block length $l = 177$, and by an analogous SW bootstrap

10.6 Some Practical Issues

The main practical problem with the bootstrap procedures above is that it is not clear how to choose the tuning parameters for an observed data set with a finite number of observations and unknown data generating process. For both bootstrap procedures, the block length is to be chosen such that l tends to infinity at a slower rate than n. Even if we restrict attention to block lengths proportional to $n^{1-\varepsilon}$ for some $0 < \varepsilon < 1$, one needs to specify ε and the proportionality constant. For the block bootstrap, there is an additional tuning parameter k.

As a general rule, the block length should be neither too small nor too large, compared to n. If l is very small, then the computed statistics fail to capture the asymptotic effect of long-range dependence. On the other hand, if l is too large, then the number of blocks to choose from is small so that there is not enough variability among the (highly dependent) block statistics, and the results may heavily depend on spurious features of the observed series. The latter problem is more likely to occur for the SW bootstrap because there the whole shape of the sample path plays a role. This is illustrated in Figs. 10.1, 10.2 and 10.3. The figures are based on a simulated series of the process $Y_t = G(X_t)$ where X_t is a FARIMA$(0, 0.4, 0)$ process with variance one and $G(x) = x + 0.005(x^3 - x)$. Since the Hermite rank of G is one, both bootstrap procedures are valid. Given the dominant linear part and the relatively large sample size of $n = 1000$, one would expect a good approximation by any reasonable bootstrap method. In Fig. 10.1, l is chosen to be equal to $n^{1-\varepsilon}$ with $\varepsilon = \frac{1}{4}$ so that $l = 177$. While the block bootstrap and even the asymptotic standard normal approximation are close to the simulated histogram, the SW bootstrap yields a completely wrong bimodal distribution. The reason for the bimodal shape can be seen in Figs. 10.2(a)–(d). Due to strong long memory (with $d = 0.4$), the simulated sample path stays below zero for a relatively long time in the beginning and towards

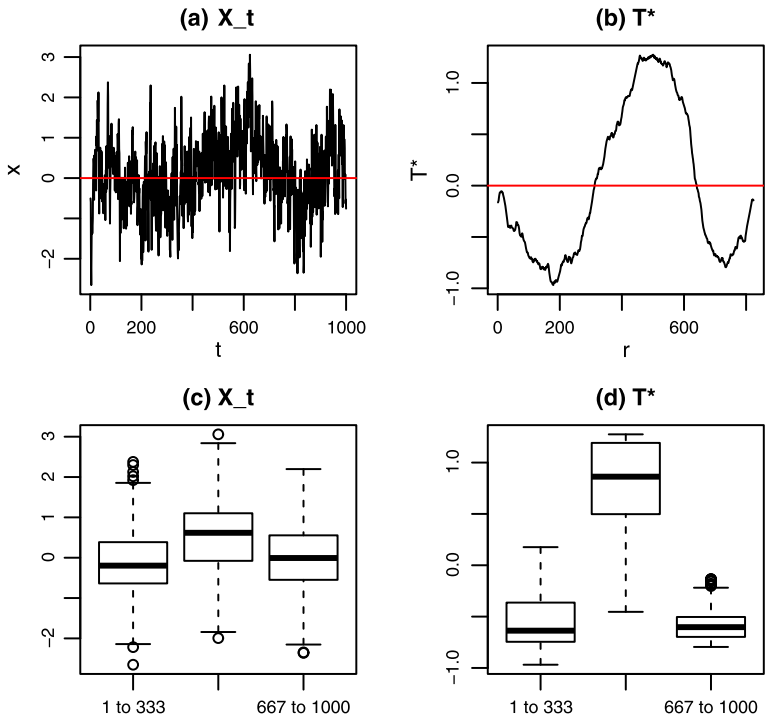

Fig. 10.2 Same simulated series as for the histogram in Fig. 10.1 (**a**), together with values of $T^*_{n,r}$ for blocks moving from left to right (**b**), and boxplots of X_t and $T^*_{n,r}$ for three different regions ((**c**) and (**d**))

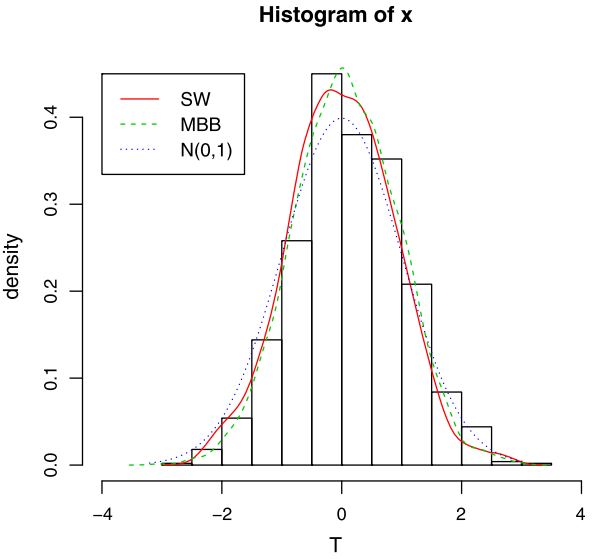

Fig. 10.3 Histogram of a simulated series $Y_t = G(X_t)$ of length $n = 1000$, where X_t is a FARIMA$(0, 0.4, 0)$ process with variance one and $G(x) = x + 0.005(x^3 - x)$. Also plotted are distributions obtained by blockwise bootstrap with block length $l = 5$, and by an analogous SW bootstrap

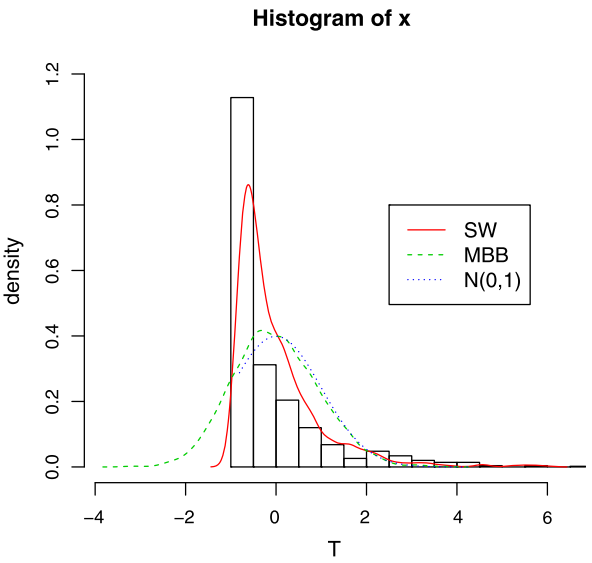

Fig. 10.4 Histogram of a simulated series $Y_t = G(X_t)$ of length $n = 1000$, where X_t is a FARIMA(0, 0.4, 0) process with variance one and $G(X_t) = H_2(X_t) = X_t^2 - 1$. Also plotted are distributions obtained by blockwise bootstrap with block length $l = 5$, and by an analogous SW bootstrap

the end whereas it is above zero most of the time in the middle period. As a result, conditionally on the observed sample path, block sums and hence the values of $T^*_{n,r}$ exhibit a bimodal distribution (Figs. 10.1 and 10.2(b), (d)). In contrast, for the block bootstrap the long wave in the observed series does not influence the result because blocks are resampled randomly. The dependence of the SW bootstrap on spurious features can be alleviated by choosing a smaller block length. This illustrated Fig. 10.3 where $\varepsilon = \frac{3}{4}$ and hence $l = 5$ was used.

Figure 10.4 shows an example where only the SW bootstrap is a valid resampling procedure. The simulated series is $Y_t = G(X_t)$ with $G(X_t) = H_2(X_t) = X_t^2 - 1$. Since the Hermite rank is two, the asymptotic distribution is given by the marginal of the Hermite–Rosenblatt process. This distribution is skewed to the right. The simulated histogram of T_n with $n = 1000$ is indeed highly skewed. In contrast, the distribution obtained by the MBB is symmetric and very close to the standard normal density. The SW bootstrap provides a much better approximation with a skewed shape. As before, however, the concrete choice of the block length is crucial. The good approximation in Fig. 10.4 with $l = 5$ ($\varepsilon = \frac{1}{4}$) is in sharp contrast to the disastrous result in Fig. 10.5 with $l = 177$ ($\varepsilon = \frac{3}{4}$).

Generally, one may conclude that the SW method is quite flexible since it is able to capture non-Gaussian limits. This is very useful even for large sample sizes because the distribution of Hermite processes is rather complicated except for Hermite rank one. On the other hand, the flexibility of the SW method comes at a price. Since almost the complete dependence structure of the observed series is preserved, results may heavily depend on the particular sample path. This lack of 'robustness' can lead to artefacts. A good choice of the block length l plays an important role. On the one hand, l needs to be large enough to come as close as possible to the situation with n observations. On the other hand, if l is too large, then some spurious properties

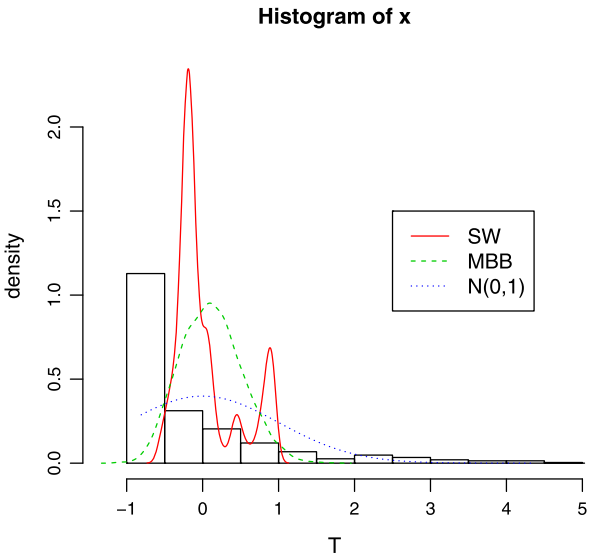

Fig. 10.5 Histogram of a simulated series $Y_t = G(X_t)$ of length $n = 1000$, where X_t is a FARIMA$(0, 0.4, 0)$ process with variance one and $G(X_t) = H_2(X_t) = X_t^2 - 1$. Also plotted are distributions obtained by blockwise bootstrap with block length $l = 177$, and by an analogous SW bootstrap

of the observed sample path may have an undue influence on the result (see, e.g. Fig. 10.1). Thus, as is so often in nonparametric statistics, a suitable balance has to be achieved between two conflicting aims.

10.7 More Complex Models

10.7.1 Bootstrap for the Heavy-Tailed SV Model

10.7.1.1 The HTLM Model

We consider a stochastic volatility model $X_t = \xi_t \sigma_t$, where the random variables ξ_t are i.i.d., strictly positive and regularly varying with index $-\alpha$, $\alpha \in (1, 2)$, that is,

$$P(\xi_1 > x) \sim Ax^{-\alpha}.$$

The sequence $\sigma_t = \exp(\zeta_t)$ is stationary and ergodic, and independent of the sequence ξ_t. Furthermore, ζ_t is a Gaussian long-memory process with parameter d. Suppose that $E[\xi_1] \neq 0$. We saw in Example 4.17 that, if $1/2 + d < 1/\alpha$, then

$$n^{-1/\alpha} S_n(u) \Rightarrow A^{1/\alpha} C_\alpha^{-1/\alpha} \left(E\left[\sigma_1^\alpha\right]\right)^{1/\alpha} \tilde{Z}_\alpha(u), \tag{10.14}$$

where $\tilde{Z}_\alpha(\cdot)$ is an α-stable Lévy process such that $\tilde{Z}_\alpha(1) \stackrel{\mathrm{d}}{=} S_\alpha(1, 1, 0)$. On the other hand, if $1/2 + d > 1/\alpha$, then

$$n^{-(1/2+d)} L_1^{-1/2}(n) S_n(u) \Rightarrow J(1) E[\xi_1] B_H(u)\,, \tag{10.15}$$

where $B_H(\cdot)$ is a fractional Brownian motion, $H = d + \frac{1}{2}$, $L_1(n) = C_1 L_\gamma(n)$ and $J(1) = E(\zeta_1 \exp(\zeta_1))$. In Example 4.17, we called this model LMSD. A very similar model was considered in McElroy and Politis (2007). There, $X_t = \xi_t \sigma_t$ with $\sigma_t = \sigma(\zeta_t)$. The function $\sigma(\cdot)$ is supposed to have Hermite rank 1, and furthermore $E[\sigma_1] = 0$. For this model, we have the same dichotomy as in (10.14)–(10.15), only the constants of the limiting distributions change. McElroy and Politis coined the term "HTLM (Heavy Tailed with Long Memory)".

10.7.1.2 Subsampling for the HTLM Model

We consider $Y_t = \mu + X_t$, where X_t ($t \in \mathbb{N}$) is the HTLM model described above, with $E[\xi] \neq 0$ (but $E[X_t] = 0$ since the subordinated Gaussian sequence σ_t is centred). We noted above that the limiting distribution F is either stable or normal. Furthermore, the scaling v_n is the maximum of $n^{1/\alpha}$ and $n^{d+1/2}L(n)$, where $L(n)$ is a slowly varying function.

Recall the self-normalized statistics W_n from (10.6). Since our data are dependent, we have to change the self-normalizer. It can be constructed as

$$V_n^2 = \sum_{j=1}^{n} (Y_j - \bar{Y}_n)^2 + nLM_n(\rho),$$

where

$$LM_n(\rho) = \left| \sum_{|k|=1}^{[n^\rho]} \frac{1}{n - |k|} \sum_{j=1}^{n-k} \left(Y_j Y_{j+k} - \bar{Y}_n^2 \right) \right|^{1/\rho}, \quad \rho \in (0, 1).$$

To get an idea about the behaviour of V_n^2, we note that Y_j^2 ($j \in \mathbb{N}$) are regularly varying with index $-\alpha/2$ and thus they have an infinite mean. This implies that the behaviour of Y_j^2 is free of long memory. In particular, $\sum_{j=1}^n Y_j^2$ grows at rate $n^{2/\alpha}$, $n^{-2/\alpha} \sum_{j=1}^n (Y_j - \bar{Y}_n)^2$ converges to a stable random variable and $n^{-(2d+1)} L^{-2}(n) \sum_{j=1}^n (Y_j - \bar{Y}_n)^2$ converges in probability to 0. As for $LM_n(\rho)$, we recognize $(n - |h|)^{-1} \sum_{j=1}^{n-h} (Y_j Y_{j+k} - \bar{Y}_n^2)$ as the sample covariance at lag k associated with the sequence Y_j ($j \in \mathbb{N}$) which is the same as the sample covariance of X_j ($j \in \mathbb{N}$). We expect that they converge in probability to $\gamma_X(k) = E^2[\xi_0] E[\sigma_0 \sigma_k]$. If we assume that $\gamma_X(k) \sim L_\gamma k^{2d-1}$, $d \in (0, 1/2)$, then we expect $LM_n(\rho)$ to grow at the rate

$$C \left| \sum_{|k|=1}^{[n^\rho]} k^{2d-1} \right|^{1/\rho} \approx C n^{2d},$$

since the Hermite rank is one. Thus, with $v_n := \max\{n^{1/\alpha}, n^{d+1/2}L(n)\}$ we may conclude that

$$W_n := \frac{n(\bar{Y}_n - \mu)}{V_n} = \frac{v_n^{-1}\sum_{j=1}^{n}(Y_j - \mu)}{v_n^{-1}V_n}$$

converges to a non-degenerate random variable.

Now, using the blocks $B_r = (Y_r, \ldots, Y_{r+l-1})$, $r = 1, \ldots, N_b$, we construct replicates of W_n as

$$W^*_{n,l,r} = l\frac{\bar{Y}_{n,l,r} - \bar{Y}_n}{V_{n,l,r}}, \quad r = 1, \ldots, N_b,$$

where

$$\bar{Y}_{n,l,r} = \frac{1}{l}\sum_{j=r}^{r+l-1} Y_j$$

and

$$LM_{n,l,r}(\rho) = \left|\sum_{|k|=1}^{[l^\rho]} \frac{1}{l-|k|}\sum_{j=r}^{r+l-1-|k|}\left(Y_jY_{j+k} - \bar{Y}^2_{n,l,r}\right)\right|^{1/\rho}.$$

A $(1-\theta)$-confidence interval can be constructed as

$$[\bar{Y}_n - z_{1-\frac{\theta}{2}}V_n, \bar{Y}_n - z_{\frac{\theta}{2}}V_n],$$

where $z_{\frac{\theta}{2}}$ is the $(1-\theta)$-percentile of the empirical distribution function

$$F^*_n(x) = \frac{1}{n-l+1}\sum_{r=1}^{n-l+1} 1\{W^*_{n,l,r} \le x\}.$$

For details, we refer to McElroy and Politis (2007) and Jach et al. (2012).

10.7.2 Testing for Jumps in a Trend Function

In some situations, the modified MBB approach can be useful for defining test statistics whose distribution under the null hypothesis is asymptotically normal due to the resampling device. For instance, consider a model with a nonparametric trend function given by

$$Y_i = m(t_i) + e_i \tag{10.16}$$

where $m \in L^2[0,1]$ and e_i a Gaussian process with autocovariance function $\gamma(k) \sim L_\gamma|k|^{-\alpha}$ for some $\alpha = 2d - 1 \in (0,1)$. Beran and Shumeyko (2012b) derive an MBB-based test for

$$H_0 : m \in C[0,1]$$

against the alternative H_1 that m has at least one isolated jump. The idea is to use the wavelet estimator

$$\hat{m}(t) = \hat{m}_{\text{low}}(t) + \hat{m}_{\text{high}}(t)$$

given in Sect. 7.5. The low resolution component $\hat{m}_{\text{low}}(t)$ is an optimal estimator of m, if m is continuous whereas the high resolution part $\hat{m}_{\text{high}}(t)$ captures departures from continuity. A natural idea is therefore to test H_0 against H_1 by designing a test statistic that compares two types of residuals, $\hat{e}_i = Y_i - \hat{m}(t) = Y_i - \hat{m}_{\text{low}}(t) - \hat{m}_{\text{high}}(t)$ and $\hat{e}_{i,\text{low}} = Y_i - \hat{m}(t) = Y_i - \hat{m}_{\text{low}}(t)$. This can be done, for instance, as follows. For a given block size l, define block sums

$$\zeta_r = \hat{e}_r + \cdots + \hat{e}_{r+l-1} = \sum_{\hat{e}_j \in B_r} \hat{e}_j$$

and

$$\zeta_{r,\text{low}} = \hat{e}_{r,\text{low}} + \cdots + \hat{e}_{r+l-1,\text{low}} = \sum_{\hat{e}_j \in B_r} \hat{e}_{j,\text{low}}$$

($1 \le r \le N_b = n - l + 1$). Then, k blocks $B_1^*, \ldots, B_k^*$ are sampled independently with replacement and bootstrap samples $\zeta_1^*, \ldots, \zeta_k^*$ and $\zeta_{1,\text{low}}^*, \ldots, \zeta_{k,\text{low}}^*$ are computed. The corresponding bootstrap statistics are

$$T_{kl}^* = k^{-1/2} \sum_{r=1}^{k} \frac{\zeta_r^*}{v_l}, \qquad T_{kl,\text{low}}^* = k^{-1/2} \sum_{r=1}^{k} \frac{\zeta_{r,\text{low}}^*}{v_l}$$

with $v_l = L_\gamma^{1/2} l^{d+\frac{1}{2}}$. Extending the proofs in Lahiri (1993) and Beran and Shumeyko (2012a), the following result can be derived (Beran and Shumeyko 2012b):

Theorem 10.3 *Suppose that $m \in L^2[0,1]$, m' exists except for a finite set $\mathcal{N} \subset [0,1]$ and is piecewise continuous outside of $\mathcal{N}$. Moreover, let*

$$l = O\left(n^{\delta}\right)$$

where

$$\frac{1}{2r+\alpha} < \delta < \frac{2}{2r+\alpha}$$

and define $\tilde{\sigma}^2 = 2\sigma^2(1-\alpha)^{-1}(2-\alpha)^{-1}$ where $\sigma^2 = \text{var}(e_t)$. Then, under $H_0 : m \in C[0,1]$, we have

$$\begin{aligned} E_*\left(T_{kl,\text{low}}^*\right) &= E_*\left(T_{kl}^*\right) + o_p\left(n^{0.5\alpha\delta - \ln n}\right) = o_p(1), \\ \text{Var}_*\left(T_{kl,\text{low}}^*\right) &= \text{Var}_*\left(T_{kl}^*\right) + o_p\left(n^{\alpha\delta - 2\ln n}\right) = \tilde{\sigma}^2 + o_p(1), \\ T_{kl,\text{low}}^* &= T_{kl}^* + O_p\left(n^{0.5\alpha\delta - \ln n}\right) \end{aligned}$$

and

$$\sup_{x\in\mathbb{R}}\left|P_*\big(T^*_{kl,\mathrm{low}}\le x\big)-\Phi\left(\frac{x}{\tilde\sigma}\right)\right|=o_p(1),$$

$$\sup_{x\in\mathbb{R}}\left|P_*\big(T^*_{kl}\le x\big)-\Phi\left(\frac{x}{\tilde\sigma}\right)\right|=o_p(1),$$

$$\sup_{x\in\mathbb{R}}\big|P_*\big(T^*_{kl}\le x\big)-P_*\big(T^*_{kl,\mathrm{low}}\le x\big)\big|=o_p(1).$$

Thus, under H_0, the two statistics are asymptotically equivalent and converge uniformly in distribution to the $N(0,\tilde\sigma^2)$ distribution. This is no longer the case under H_1:

Theorem 10.4 *Suppose that the same assumptions as in the previous theorem hold except that m has at least one isolated jump. Then the first two moments and the distribution of T^*_{kl} as well as $E_*(T^*_{kl,\mathrm{low}})$ are the same asymptotically as under H_0. However,*

$$\mathrm{Var}_*\big(T^*_{kl,\mathrm{low}}\big)=\tilde\sigma^2+w_n+o_p(1)$$

where

$$w_n=C^*n^\beta$$

with

$$0<\beta=\alpha\delta-\frac{\alpha}{2r+\alpha}<\frac{\alpha}{2r+\alpha}.$$

Moreover,

$$\sup_{x\in\mathbb{R}}\left|P_*\big(T^*_{kl,\mathrm{low}}\le x\big)-\Phi\left(\frac{x}{\sqrt{\tilde\sigma^2+w_n}}\right)\right|=o_p(1),$$

$$\sup_{x\in\mathbb{R}}\left|P_*\big(T^*_{kl}\le x\big)-\Phi\left(\frac{x}{\tilde\sigma}\right)\right|=o_p(1).$$

Note in particular that under H_1 the ratio of the variances $\mathrm{var}(T^*_{kl,\mathrm{low}})/\mathrm{var}(T^*_{kl})$ diverges to infinity. We may therefore test

$$H_0:\mathrm{var}_*\big(T^*_{kl,\mathrm{low}}\big)=\mathrm{var}\big(T^*_{kl}\big)$$

against

$$H_1:\mathrm{var}_*\big(T^*_{kl,\mathrm{low}}\big)>\mathrm{var}\big(T^*_{kl}\big).$$

Repeating the bootstrap procedure described so far, say N_T times, we calculate

$$W_{\mathrm{low}}=\tilde\sigma^{-2}\sum_{i=1}^{N_T}\big(T^{*(i)}_{kl,\mathrm{low}}-\bar T^*_{kl,\mathrm{low}}\big)^2$$

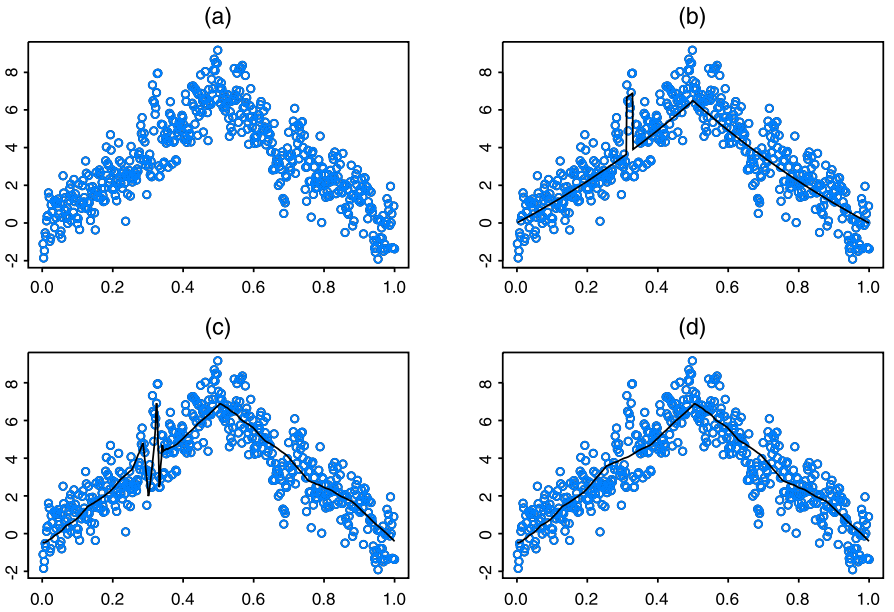

Fig. 10.6 (**a**) Simulated series $Y_i = m(t_i) + e_i$ with e_i generated by a FARIMA process and (**b**) a trend function with a local jump. The wavelet estimate of $m(t_i)$ is shown in (**c**), a kernel estimate in (**d**). The bootstrap based test (using the trend estimate in (**c**)) detects the jump at the 5 %-level of significance

and reject H_0, if W_{low} is too large. Conditionally on the sample, the simulated statistics $T^{*(i)}_{kl,\text{low}}$ $(i = 1, 2, \ldots, N_T)$ are independent. Moreover, under H_0 they are asymptotically $N(0, \tilde{\sigma}^2)$-distributed so that W_{low} is approximately $\chi^2_{N_T-1}$-distributed. Approximate critical values for W_{low} are therefore given by corresponding quantiles of the $\chi^2_{N_T-1}$-distribution. To obtain more exact finite sample quantiles, one can instead simulate the distribution of

$$W = \tilde{\sigma}^{-2} \sum_{i=1}^{N_T} \left(T^{*(i)}_{kl} - \bar{T}^{*}_{kl}\right)^2$$

via resampling. This approach is adopted in Beran and Shumeyko (2012b).

Figure 10.6 shows a typical example where the wavelet decomposition and the test based on W_{low} enables us to detect a very local discontinuity in the trend function. In spite of the presence of local spurious trends caused by strong long memory in the residuals, the local disturbance in the trend function (Fig. 10.6(b)) is captured by the high resolution component (Fig. 10.6(c)). This is in contrast to other nonparametric regression methods such as kernel or local polynomial regression (Fig. 10.6(d)).