

# The local bootstrap for Markov processes<sup>☆</sup>

Efstathios Paparoditis<sup>a</sup>, Dimitris N. Politis<sup>b, \*</sup>

<sup>a</sup>*Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537,  
CY 1678 Nicosia, Cyprus*

<sup>b</sup>*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA*

---

## Abstract

A nonparametric bootstrap procedure is proposed for stochastic processes which follow a general autoregressive structure. The procedure generates bootstrap replicates by locally resampling the original set of observations reproducing automatically its dependence properties. It avoids an initial nonparametric estimation of process characteristics in order to generate the pseudo-time series and the bootstrap replicates mimic several of the properties of the original process. Applications of the procedure in nonlinear time-series analysis are considered and theoretically justified; some simulated and real data examples are discussed.

© 2002 Published by Elsevier Science B.V.

*MSC:* primary 62G09; 62M05; secondary 62M10

*Keywords:* Markov dependence; Nonlinear time series; Resampling; Time reversibility

---

## 1. Introduction

The present paper introduces a new resampling procedure, namely the local bootstrap, which is applicable in the setting of strictly stationary, real-valued stochastic processes  $\{X_t, t = 1, 2, \dots\}$  that follow a general (finite order) Markovian structure, i.e., a dependence structure where for every  $t \in \mathbb{N}$  and every Borel set  $A \subset \mathbf{R}$ ,  $X_t$  depends on the past only through the previous  $p$  values, i.e.,

$$P(X_{t+1} \in A | X_j, j \leq t) = P(X_{t+1} \in A | X_j, t - p + 1 \leq j \leq t).$$

Although the class of processes to which our procedure is applicable will be more precisely described in the next section, we mention here that it is large enough to

---

<sup>☆</sup> Research partially supported by NSF Grant DMS-97-03964.

\* Corresponding author.

*E-mail address:* politis@euclid.ucsd.edu (D.N. Politis).

include several of the commonly used nonlinear models in time-series analysis. An important member of this class is, for instance, the class of nonlinear, conditionally heteroscedastic, autoregressive processes defined by

$$X_t = m(X_{t-1}, X_{t-2}, \dots, X_{t-p}) + v(X_{t-1}, X_{t-2}, \dots, X_{t-p}) \epsilon_t, \quad t = 1, 2, \dots, \quad (1)$$

where the noise process  $\{\epsilon_t, t \geq 1\}$  is a sequence of i.i.d. zero mean random variables satisfying certain assumptions and  $m(\cdot)$  and  $v(\cdot)$  are smooth functions (generally unknown). Note that model (1) includes several of the familiar nonlinear parametric models; see Tong (1990) and Tjøstheim (1994) for details. However, (1) is not explicitly assumed in the following, i.e., the Markov class of processes to which our procedure is applicable is richer than model (1).

The local bootstrap scheme proposed in this paper involves resampling the *original* data points  $X_1, X_2, \dots, X_T$  to form pseudoserries  $X_1^*, X_2^*, \dots, X_T^*$  from which the statistics of interest are recalculated. To motivate the idea underlying the procedure proposed, consider the particular case where  $\{X_t\}$  indeed follows the general autoregressive structure (1) with  $p = 1$ , and let  $X_1, X_2, \dots, X_T$  be the observed series. Given the value assumed by the process at time  $t$ , i.e., given  $X_t$ , consider the problem of obtaining a replicate of the value assumed at  $t + 1$  when  $m(\cdot)$  and  $v(\cdot)$  are unknown. To proceed, assume that an observation  $X_s \in \{X_1, X_2, \dots, X_{T-1}\}$ , with  $s \neq t$  exists which is close to  $X_t$ , i.e.,  $|X_t - X_s| \leq b$ , for some small  $b > 0$ . Then for  $m(\cdot)$  and  $v(\cdot)$  smooth functions, differences between the subsequent values  $X_{t+1}$  and  $X_{s+1}$  are mainly due to differences in the unobserved errors, i.e.,  $\epsilon_{t+1}$  and  $\epsilon_{s+1}$ . To see this, suppose that  $m$  and  $v$  have bounded first-order derivatives and that  $\epsilon_t$  has finite variance; a Taylor series expansion yields

$$X_{s+1} = m(X_s) + v(X_s)\epsilon_{s+1} = m(X_t) + v(X_t)\epsilon_{s+1} + O_p(b),$$

i.e., for  $b$  close to zero,  $X_{s+1}$  can be thought to have been generated by the same stochastic difference equation as  $X_{t+1} = m(X_t) + v(X_t)\epsilon_{t+1}$  but with a different error term ( $\epsilon_{s+1}$  instead of  $\epsilon_{t+1}$ ). Since the above consideration is true for all observed values which have the same property as  $X_{s+1}$ , i.e., the property that their preceding value in the original data sequence is  $b$ -close to  $X_t$ , a replicate  $X_{t+1}^*$  of  $X_{t+1}$  can be obtained by randomly choosing a value from the set of observations

$$\{X_{s+1}: |X_t - X_s| \leq b \text{ and } 1 \leq s \leq T - 1\}.$$

In choosing  $X_{t+1}^*$  from the above set, an appropriate probability distribution, which we will term the *resampling kernel*, is used. Thus starting at  $t = 1$  by some appropriately chosen initial value  $X_1^*$ , subsequent replicates  $X_2^*, X_3^*, \dots, X_T^*$  can be obtained by applying the above local resampling scheme at each stage of the resampling process. It can then be shown that for an important class of statistical problems and under certain regularity conditions, asymptotically valid inference can be obtained by letting the *resampling width*  $b = b(T) \rightarrow 0$  as the sample size  $T$  increases but in such a way that the number of observations falling in the set from which bootstrap replicates are obtained goes to infinity.

Bootstrapping time-series data has received considerable interest in recent years. As a follow-up to the i.i.d. bootstrap of Efron (1979), several different approaches for

bootstrapping stationary observations have been proposed in the literature. Let us just mention here among others, the ‘residual bootstrap’ (cf. Freedman, 1984; Efron and Tibshirani, 1986), the block bootstrap (cf. Künsch, 1989; Liu and Singh, 1992), the blocks-of-blocks bootstrap (cf. Politis and Romano, 1992), the stationary bootstrap (cf. Politis and Romano, 1994), the frequency domain bootstrap (cf. Franke and Härdle, 1992); see Shao and Tu (1995) for an overview of those methods.

The local bootstrap proposal of this paper is perhaps more akin to a bootstrap procedure for (first order) Markov processes suggested by Rajarshi (1990). In that procedure bootstrap replicates are obtained by using the nonparametrically estimated one-step transition density function. To be more precise, for  $p = 1$  and  $X_t^*$  given, a subsequent value  $X_{t+1}^*$  is obtained as  $X_{t+1}^* \sim \hat{f}_{X_{t+1}|X_t}(\cdot | X_t^*)$  where  $\hat{f}_{X_{t+1}|X_t}(\cdot | X_t^*)$  is a nonparametric (kernel) estimator of the unknown conditional density function  $f_{X_{t+1}|X_t}(\cdot | X_t^*)$  of the process. Compared to the bootstrap procedure of Rajarshi (1990), the local resampling procedure proposed in this paper is simpler, easier to implement and avoids the nonparametric estimation of the conditional density  $f_{X_{t+1}|X_t}(\cdot | X_t^*)$ .

In essence, our local bootstrap *implicitly* performs an estimation of the one-step transition *distribution* as opposed to Rajarshi’s (1990) (explicit) estimation of the transition *density* by nonparametric smoothing of the transition distribution. We show that, similarly to the question “to smooth or not to smooth” in the classical i.i.d. bootstrap (cf. Efron and Tibshirani, 1993, and the references therein), this extra smoothing is typically *unnecessary* for the bootstrap to work in the Markovian case. As an added feature, our local bootstrap pseudo-series consist of (particular) reshufflings of the *original* data points, the reshuffling being done of course according to a particular probability mechanism. This constitutes a further similarity of the local bootstrap to Efron’s (1979) classical i.i.d. bootstrap; in effect, Efron’s (1979) i.i.d. bootstrap can be considered as a special case of our local bootstrap if we allow the Markovian order  $p$  to be zero.

Under model class (1), there exists a residual-based procedure investigated by Franke et al. (1996) that begins with nonparametric estimation of  $m(\cdot)$  and  $v(\cdot)$ , and a subsequent i.i.d. resampling of the estimated residuals; cf. also Auestad and Tjøstheim (1991). In contrast to this approach our local bootstrap scheme does *not* attempt to reduce the problem to i.i.d. resampling of residuals but involves resampling the original data points to form the pseudo-time series. Again, it has the advantage of side-stepping such initial smoothing problems, and at the same time it is valid for a class of stochastic processes that is wider than (1). As a final point, note that the local bootstrap is related to the resampling algorithms proposed by Shi (1991) in the context of i.i.d. regression, by Falk and Reiss (1992) in the context of bootstrapping conditional curves and in an ad hoc manner by Lall and Sharman (1996) in the context of hydrological time series, and by Chan et al. (1997) in the context of ecological series.

The paper proceeds as follows. In Section 2, the local resampling method is described. Its finite sample as well as its asymptotic properties are discussed in great detail in Section 3. The properties of the local bootstrap derived in this section make it potentially applicable to a variety of inference problems. Section 4 deals with such applications of the local bootstrap; here the resampling procedure proposed is used to conduct statistical inference pertaining to linear statistics, as well as analyze the lag-reversibility properties of a time series. Some practical aspects of selecting

the resampling parameters—in particular, the resampling width  $b$ —are discussed in Section 5. Applications to simulated and real data are given in this section as well. Section 6 contains some concluding remarks while all technical proofs are deferred to Appendix A.

## 2. The local bootstrap

### 2.1. Basic notation and assumptions

We begin with an assumption which specifies precisely the class of stochastic process  $\{X_t, t \geq 1\}$  considered. We denote by  $\mathcal{B}_p$  the Borel  $\sigma$ -algebra over  $\mathbf{R}^p$ .

(A1) There exists a positive integer  $p < \infty$  such that the stochastic process  $\mathcal{Y} = \{\mathbf{Y}_t, t \geq 1\}$  with  $\mathbf{Y}_t$  defined by  $\mathbf{Y}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})$ , forms an aperiodic, strictly stationary and geometrically ergodic Markov chain on  $(\mathbf{R}^p, \mathcal{B}_p)$ .

We assume that the integer  $p$  is the smallest possible for which (A1) is satisfied. Note that the aperiodicity assumption together with geometric ergodicity implies that a unique invariant probability measure  $\pi_{\mathbf{Y}}$  on  $\mathbf{R}^p$  exists, which is the stationary distribution of  $\mathcal{Y}$ . The assumption of strict stationarity implies then that  $\mathcal{L}(\mathbf{Y}_1) = \pi_{\mathbf{Y}}$  where  $\mathcal{L}(\mathbf{Y}_1)$  denotes the law of the starting value  $\mathbf{Y}_1 = (X_1, X_0, \dots, X_{2-p})$ . Furthermore, geometric ergodicity together with stationarity implies that the Markov chain  $\{X_t, t \geq 1\}$  is geometrically strong mixing (i.e.,  $\alpha$ -mixing), since it is geometrically absolutely regular (i.e.,  $\beta$ -mixing); cf. Doukhan (1995, Chapter 2.4).

Assumption (A1) is general enough in that it is fulfilled by several of the commonly used nonlinear time-series models. In fact, there is a large amount of literature dealing with stationarity and geometric ergodicity results for different nonlinear time-series models that admit a Markovian dependence structure; cf., among others, Tong (1990), Tjøstheim (1990), Ango Nze (1992) and Meyn and Tweedie (1993).

Denote in the following by  $F(\cdot)$  the stationary distribution of  $X_t$ , the existence of which is ensured by the fact that the sample paths of  $\{X_t, t \geq 1\}$  are those of the first component of  $\mathbf{Y}_t$ . Note that  $F(x) = F_{\mathbf{Y}}(x, \infty, \dots, \infty)$  where  $F_{\mathbf{Y}}(\mathbf{y}) = \pi_{\mathbf{Y}}((-\infty, \mathbf{y}])$  is the unique solution of

$$F_{\mathbf{Y}}(\mathbf{y}) = \int F_{\mathbf{Y}}(\mathbf{y}|\mathbf{x}) dF_{\mathbf{Y}}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R}^p \tag{2}$$

and  $F_{\mathbf{Y}}(\cdot|\mathbf{x})$  is the one-step transition distribution function of  $\mathcal{Y}$ , i.e.,

$$F_{\mathbf{Y}}(\mathbf{y}|\mathbf{x}) = P(\mathbf{Y}_{t+1} \leq \mathbf{y} | \mathbf{Y}_t = \mathbf{x}) \quad \text{for } \mathbf{y}, \mathbf{x} \in \mathbf{R}^p.$$

We denote in the following by  $F(y|\mathbf{x})$  the one-step transition distribution function of  $X_t$ , i.e.,

$$F(y|\mathbf{x}) = P(X_{t+1} \leq y | \mathbf{Y}_t = \mathbf{x}).$$

By (A1) the transition distribution function  $F(y|\mathbf{x})$  uniquely characterizes the law of the stationary process  $\{X_t, t \geq 1\}$ . Now, let  $f_{\mathbf{Y}}(\cdot)$  be the stationary density,  $f_{X_{t+1}|\mathbf{Y}_t}(\cdot|\mathbf{x})$  the one-step transition density,  $f_{X_{t+1}\mathbf{Y}_t}(\cdot, \mathbf{x}) = f_{X_{t+1}|\mathbf{Y}_t}(\cdot|\mathbf{x})f_{\mathbf{Y}_t}(\mathbf{x})$  and denote by  $\|\cdot\|$  the Euclidean norm.

**(A2)**

- (i)  $F_{\mathbf{Y}}(\cdot)$  and  $F(\cdot|\mathbf{x})$  are absolutely continuous with respect to Lebesgue measure on  $\mathbf{R}^p$  and  $\mathbf{R}$  and have bounded densities  $f_{\mathbf{Y}_t}(\cdot)$  and  $f_{X_{t+1}|\mathbf{Y}_t}(\cdot|\mathbf{x})$ , respectively.
- (ii) For  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{R}^p$  and  $y \in \mathbf{R} \cup \{\infty\}$ ,

$$\left| \int_{-\infty}^y f_{X_{t+1},\mathbf{Y}_t}(z, \mathbf{x}_1) dz - \int_{-\infty}^y f_{X_{t+1},\mathbf{Y}_t}(z, \mathbf{x}_2) dz \right| \leq L(y) \|\mathbf{x}_1 - \mathbf{x}_2\|,$$

where  $\inf_y L(y) > 0$  and  $\sup_y L(y) = L < \infty$ .

- (iii) For  $y_1, y_2 \in \mathbf{R}$ ,  $|f_{X_{t+1}|\mathbf{Y}_t}(y_1|\mathbf{x}) - f_{X_{t+1}|\mathbf{Y}_t}(y_2|\mathbf{x})| \leq C(\mathbf{x})|y_1 - y_2|$  where  $C(\mathbf{x}) > 0$  and  $\sup_{\mathbf{x} \in \mathbf{R}^p} C(\mathbf{x}) = C < \infty$ .

The smoothness assumptions on the stationary and on the one-step transition densities imposed above are very common in nonparametric estimation and are needed here to ensure consistency of the empirical characteristics used to generate the bootstrap pseudo-series.

**(A3)** A compact subset  $S$  of  $\mathbf{R}$  exists such that  $X_t \in S$  a.s. Furthermore,  $f_{X_{t+1}|\mathbf{Y}_t}(\cdot|\mathbf{x}) > 0$  for every  $\mathbf{x} \in \mathbf{S}$ , where  $\mathbf{S}$  denotes the  $p$ -fold Cartesian product of  $S$ .

Note that the compactness of the support of  $f$  in (A3) is a rather technical assumption that can be weakened by simultaneously strengthening other requirements; see Remark 3.2 for a discussion of this point.

To set up the local resampling procedure, we first introduce a *resampling kernel*  $W(\cdot)$ , i.e., an appropriate probability density on  $\mathbf{R}^p$ , which is used in local resampling of the observations. This resampling kernel satisfies the following assumptions.

**(B1)**  $W(\cdot)$  is a bounded, (first order) Lipschitz continuous and symmetric probability density on  $\mathbf{R}^p$  satisfying  $W(\mathbf{u}) > 0$  for all  $\mathbf{u}$ . Furthermore,  $\int \mathbf{u}W(\mathbf{u}) d\mathbf{u} = 0$  and  $\int \|\mathbf{u}\|W(\mathbf{u}) d\mathbf{u} < \infty$ .

For any real number  $c > 0$  we set in the following  $W_c(\cdot) = c^{-p}W(\cdot/c)$ . Furthermore, the notations  $f(\cdot)$  and  $f(\cdot|\mathbf{x})$  for the stationary density  $f_{\mathbf{Y}}(\cdot)$  and the one-step transition density  $f_{X_{t+1}|\mathbf{Y}_t}(\cdot|\mathbf{x})$  will be also used in the sequel as convenient short-hands.

### 2.2. The resampling scheme

The resampling algorithm proposed generates bootstrap pseudo-series  $X_1^*, X_2^*, \dots, X_N^*$  where in general, the length  $N$  of the bootstrap series is not necessary equal to  $T$ , although  $N = T$  is a natural choice for statistical applications. Denote by  $\mathbf{Y}_t^*$  the row vector  $\mathbf{Y}_t^* = (X_t^*, X_{t-1}^*, \dots, X_{t-p+1}^*)$ . The local resampling algorithm generating the bootstrap replicates can then be described by the following two steps:

1. Select a *resampling width*  $b = b(T) > 0$  and a set of  $p$  starting values  $\mathbf{Y}_p^* = (X_1^*, X_2^*, \dots, X_p^*)$ .
2. For any  $t+1 \in \{p+1, p+2, \dots, N\}$  suppose that  $\mathbf{Y}_t^*$  has been generated already. Let  $J$  be a discrete random variable taking its values in the set  $\mathbf{N}_{p,T-1} = \{p, p+1, \dots, T-1\}$

with probability mass function given by

$$P(J = s) = W_b(\mathbf{Y}_t^* - \mathbf{Y}_s) \bigg/ \sum_{m \in \mathbf{N}_{p,T-1}} W_b(\mathbf{Y}_t^* - \mathbf{Y}_m) \tag{3}$$

for  $s \in \mathbf{N}_{p,T-1}$ . The bootstrap replicate  $X_{t+1}^*$  is then defined by  $X_{t+1}^* = X_{J+1}$ .

According to the above resampling scheme and based on the resampling kernel  $W$  and the resampling width  $b$ , the bootstrap procedure assigns at any time point  $t + 1$  and any assumed value of  $\mathbf{Y}_t^*$  different resampling probabilities to the data points  $X_{p+1}, X_{p+2}, \dots, X_T$ . These probabilities depend on how close to  $\mathbf{Y}_t^*$  is the set of each point’s preceding  $p$  values (in the original data sequence). In so doing the procedure actually defines for each time point  $t \in \{p + 1, p + 2, \dots\}$  a local neighborhood from which the value  $X_{t+1}^*$  is obtained; resampling from the successors of the values lying in this neighborhood justifies calling this procedure a *local* bootstrap.

**Remark 2.1.** Some variations of this procedure can be obtained by varying the way of choosing the starting values  $X_1^*, X_2^*, \dots, X_p^*$ . For instance, one possibility is to start resampling by setting  $\mathbf{Y}_p^* = \mathbf{Y}_p$  or  $\mathbf{Y}_p^* \in \{\mathbf{Y}_t: p \leq t \leq T - 1\}$ . Another possibility is to allow  $\mathbf{Y}_p^*$  to take, with a predetermined probability distribution any value from the set of values  $\{(X_{t_1}, X_{t_2}, \dots, X_{t_p}): p + 1 \leq t_j \leq T \text{ and } j = 1, 2, \dots, p\}$ ; this predetermined distribution could well be the stationary distribution of the corresponding bootstrap Markov chain in which case the bootstrap series would itself be *stationary* as Theorem 3.1 indicates.

**Example 2.1.** Before proceeding with the properties of the local resampling procedure and for the purposes of illustration we apply it to the well-known series of Wölfer sunspot numbers; cf. Tong (1990) for a description and discussion of this famous series. Fig. 1 shows the original series and a bootstrap replicate obtained using  $\mathbf{Y}_p^* = \mathbf{Y}_p$ , an autoregressive order of  $p = 9$ , the Gaussian product kernel as resampling kernel and a resampling width  $b$  according to Eq. (13) in Section 5. Note that an autoregressive order of  $p = 9$  has been suggested as appropriate for this series by Tjøstheim and Auestad (1994). As it is seen from this figure, the local bootstrap procedure manages to reproduce quite satisfactorily the main characteristics of this series.

### 3. Properties of the local bootstrap

#### 3.1. Case of fixed $T$

According to the proposed resampling scheme, for any fixed length  $T$  of the observed series, the sequence of bootstrap random variables  $\{X_t^*, t = 1, 2, \dots\}$  evolves through the ‘numbered states’  $X_1, X_2, \dots, X_T$  following a certain dependence (autoregressive) structure. In particular, we have for any positive integer  $N > p$  and any set

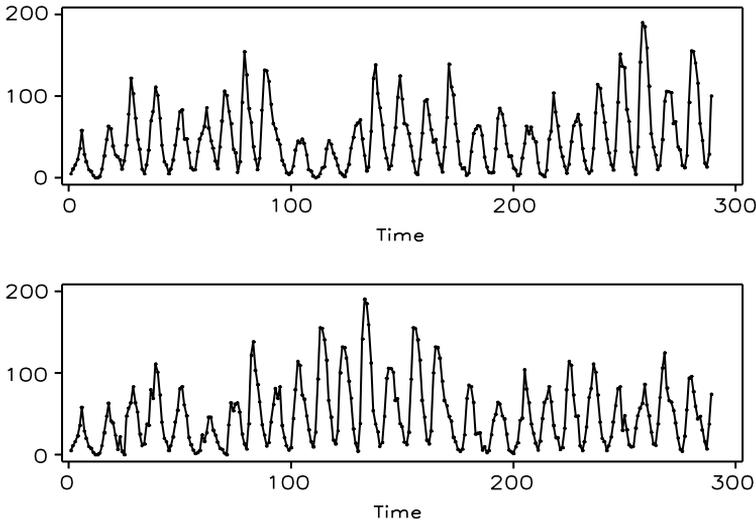


Fig. 1. Original series of sunspot numbers (top) and a bootstrap replication using the local resampling scheme with  $p = 9$ , the Gaussian product resampling kernel and a resampling width selected using Eq. (13).

$\{x_1, x_2, x_3, \dots, x_N\}$  with  $x_i \in \{X_1, X_2, \dots, X_T\}$  that

$$P(X_1^* = x_1, X_2^* = x_2, \dots, X_N^* = x_N) = P_0(\mathbf{Y}_p^* = \mathbf{x}_p) \prod_{t=p}^{N-1} P(X_{t+1}^* = x_{t+1} | \mathbf{Y}_t^* = \mathbf{x}_t),$$

where for  $\mathbf{x}_t = (x_t, x_{t-1}, \dots, x_{t-p+1})$  and  $j \in \{p, p+1, \dots, T-1\}$ ,  $P(X_{t+1}^* = X_{j+1} | \mathbf{Y}_t^* = \mathbf{x}_t)$  is the transition kernel given by

$$\begin{aligned} P(X_{t+1}^* = X_{j+1} | X_t^* = x_t, X_{t-1}^* = x_{t-1}, \dots, X_{t-p+1}^* = x_{t-p+1}) \\ = \frac{W_b(\mathbf{x}_t - \mathbf{Y}_j)}{\sum_{m=p}^{T-1} W_b(\mathbf{x}_t - \mathbf{Y}_m)}. \end{aligned} \tag{4}$$

Note that  $P_0$  denotes the probability function determining the choice of the initial values  $\mathbf{Y}_p^* = (X_1^*, X_2^*, \dots, X_p^*)$ . Thus, apart from the starting values and conditionally on the observed sequence  $X_1, X_2, \dots, X_T$ , the bootstrap series  $X_{p+1}^*, X_{p+2}^*, \dots$ , is the sample path of the first component of the  $p$ -dimensional process  $\mathcal{B}_T^* = \{\mathbf{Y}_t^*, t \geq p+1\}$ , where  $\mathbf{Y}_t^* = (X_t^*, X_{t-1}^*, \dots, X_{t-p+1}^*)$  is generated by a discrete and temporarily homogeneous first-order Markov chain that takes its values in the discrete state space  $\mathcal{E}_T = \mathcal{X}_T^p$ . Here  $\mathcal{X}_T^p$  denotes the Cartesian product of  $p$  copies of  $\mathcal{X}_T = \{X_{p+1}, X_{p+2}, \dots, X_T\}$ . Furthermore, the law of  $\{X_t^*, t \geq p+1\}$  is the law of the first component of  $\mathbf{Y}_t^*$ . The following theorem characterizes precisely the properties of this chain. It enables us to use Markov chain theory in order to investigate several of the properties of the probabilistic mechanism generating the local bootstrap replicates.

**Theorem 3.1.** For  $X_1, X_2, \dots, X_T$  given, and for every  $b > 0$  fixed and any resampling kernel  $W(\cdot)$  satisfying assumption (B1), there exists with probability one an integer

$t_0 \in \mathbb{N}$  such that for all  $t \geq t_0$  the Markov chain  $\mathcal{Y}_T^* = \{\mathbf{Y}_t^*; t \geq t_0\}$  generating the bootstrap values  $X_t^*$  is positive recurrent, irreducible and aperiodic.

Note that the particular value of  $t_0$  depends on the choice of the starting values  $\mathbf{Y}_p^*$ . Therefore, and for the sake of simplicity, we assume in the following that  $t_0 = p + 1$ .

**Remark 3.1.** Although not stated explicitly, the probabilistic behavior of the bootstrap Markov chain  $\mathcal{Y}_T^*$  depends on the resampling kernel  $W(\cdot)$  and more importantly on the resampling width  $b$ . In particular, for  $T$  fixed and  $b \rightarrow 0$ , we have  $P(X_{t+1}^* = X_{j+1} | \mathbf{Y}_t^* = \mathbf{x}) \rightarrow \delta_{j+1, l+1}$  where  $\delta_{j+1, l+1}$  is Kronecker’s delta and  $l$  the index of the observation in the original data sequence assumed by  $X_t^*$ , i.e.,  $X_t^* = X_l$ . In this case, the bootstrap process defined by the above local resampling scheme, reproduces (almost surely) the original sequence  $X_{p+1}, X_{p+2}, \dots, X_T$  of observations. On the other extreme, for  $b \rightarrow \infty$ ,  $P(X_{t+1}^* = X_{l+1} | \mathbf{Y}_t^* = \mathbf{x}) \rightarrow 1/(T - p)$ , i.e., in that case the local resampling scheme approaches an i.i.d. resampling from the set  $\{X_{p+1}, X_{p+2}, \dots, X_T\}$ . Clearly, to mimic the variability of the original series by simultaneously reproducing its (Markovian) dependence characteristics,  $b$  should lie between these two extremes.

Theorem 3.1 implies that for every  $T$ , a unique, discrete, invariant probability measure  $\pi_T^*$  on  $\mathcal{E}_T$  exists, which is the stationary distribution of the bootstrap chain  $\mathcal{Y}_T^*$ . Furthermore, a constant  $\eta_T^* \in (0, 1)$ , that in general depends on the values  $X_1, \dots, X_T$  but not on anything else, exists such that  $\|P^{(k)}(\mathbf{x}, \cdot) - \pi_T^*(\cdot)\|_V = o(\eta_T^{*k})$  where  $P^{(k)}(\mathbf{x}, \cdot)$  denotes the  $k$ -step transition kernel of  $\mathbf{Y}_t^*$ , i.e.,  $P^{(k)}(\mathbf{x}, A) = P(\mathbf{Y}_{t+k}^* \in A | \mathbf{Y}_t^* = \mathbf{x})$  for  $A \in \mathcal{B}(\mathcal{E}_T)$ . In this notation  $\|\cdot\|_V$  states for the total variation norm while  $\mathcal{B}(\mathcal{E}_T)$  denotes the  $\sigma$ -field of all countable subsets of  $\mathcal{E}_T$ ; cf. Meyn and Tweedie (1993). Note that the discrete bootstrap Markov chain  $\mathcal{Y}_T^*$  generating the bootstrap replicates satisfies Doeblin’s condition  $D$ ; cf. Doob (1953, p. 192). Together with the aperiodicity of  $\mathcal{Y}_T^*$  this implies that for every  $T$ ,  $\{X_t^*, t \geq 1\}$  is uniformly mixing, i.e.,  $\phi_T^*(n) \rightarrow 0$  as  $n \rightarrow \infty$  where

$$\phi_T^*(n) = \sup_{A \in \mathcal{F}_1^{*p}, B \in \mathcal{F}_{p+n}^{*\infty}} |P^*(B|A) - \pi_T^*(B)| \tag{5}$$

and  $\mathcal{F}_1^{*p} = \sigma(X_1^*, X_2^*, \dots, X_p^*)$  and  $\mathcal{F}_{p+n}^{*\infty} = \sigma(X_{p+n}^*, X_{p+n+1}^*, \dots)$  are the  $\sigma$ -algebras generated by  $X_1^*, X_2^*, \dots, X_p^*$  and  $X_{p+n}^*, X_{p+n+1}^*, \dots$ , respectively; see Doukhan (1994, p. 88).

Clearly, if  $\mathcal{L}(\mathbf{Y}_1^*) = \pi_T^*$  the bootstrap process  $\{X_t^*, t \geq 1\}$  is strictly stationary, the stationary distribution function of which is denoted in the following by  $F_T^*(\cdot)$ . From Theorem 3.1 we then have  $F_T^*(y) = F_{\mathbf{Y}, T}^*(y, \infty, \dots, \infty)$  where  $F_{\mathbf{Y}, T}^*(\mathbf{y}) = \pi_T^*((-\infty, \mathbf{y}])$  and  $F_{\mathbf{Y}, T}^*(\mathbf{y})$  is uniquely determined by the equation

$$F_{\mathbf{Y}, T}^*(\mathbf{y}) = \int F_{\mathbf{Y}, T}^*(\mathbf{y} | \mathbf{x}) dF_{\mathbf{Y}, T}^*(\mathbf{x}). \tag{6}$$

$F_{\mathbf{Y}, T}^*(\cdot | \mathbf{x})$  appearing in the above equation is the one-step transition distribution function given by

$$F_{\mathbf{Y}, T}^*(\mathbf{y} | \mathbf{x}) = P(\mathbf{Y}_{t+1}^* \leq \mathbf{y} | \mathbf{Y}_t^* = \mathbf{x}) \quad \text{for } \mathbf{x} \in \mathcal{E}_T.$$

### 3.2. Case of increasing $T$

Our discussion was so far restricted to the case of a fixed length  $T$  of the observed series. We now turn to the probabilistic properties of the bootstrap process  $\{X_t^*, t \geq 1\}$  as the length  $T$  of the observed series increases. The asymptotic results obtained in the following are derived under the assumption that the local neighborhood from which observations are resampled is getting narrower as the sample sizes increases but in such a way that the number of observations ‘falling’ in this neighborhood increases to infinity. To be more precise, we assume that

**(B2)**  $b \asymp T^{-\delta}$  with  $0 < \delta < 1/(2p)$ .

The notation  $b \asymp T^{-\delta}$  means that constants  $c_1$  and  $c_2$  exist such that  $0 < c_1 \leq bT^\delta \leq c_2 < \infty$  for all  $T$  large enough.

Consider first the behavior of the one-step transition distribution function  $F_T^*(y|\mathbf{x}) = P(X_{t+1}^* \leq y | \mathbf{Y}_t^* = \mathbf{x})$  governing the law of the bootstrap process and note that

$$\begin{aligned}
 F_T^*(y|\mathbf{x}) &= \sum_{j=p}^{T-1} 1_{(-\infty, y]}(X_{j+1}) P(X_{t+1}^* = X_{j+1} | \mathbf{Y}_t^* = \mathbf{x}) \\
 &= \frac{\sum_{j=p}^{T-1} 1_{(-\infty, y]}(X_{j+1}) W_b(\mathbf{x} - \mathbf{Y}_j)}{\sum_{m=p}^{T-1} W_b(\mathbf{x} - \mathbf{Y}_m)} \tag{7}
 \end{aligned}$$

is used to mimic the one-step transition distribution function

$$F(y|\mathbf{x}) = \int 1_{(-\infty, y]}(z) f_{X_{t+1}|\mathbf{Y}_t}(z|\mathbf{x}) dz \tag{8}$$

The following result which is essential in obtaining several of the properties of the local bootstrap replicates can then be established.

**Theorem 3.2.** *Under assumptions (A1)–(A3) and (B1)–(B2) we have as  $T \rightarrow \infty$*

$$\sup_{y \in \mathbf{R}} \sup_{\mathbf{x} \in \mathbf{S}} |F_T^*(y|\mathbf{x}) - F(y|\mathbf{x})| \rightarrow 0 \quad a.s.$$

The importance of the above result lies in the fact that, apart from the choice of the initial values, the statistical model associated with  $\{X_t, t \geq 1\}$  is completely specified by means of the one-step transition distribution function  $F(y|\mathbf{x})$ . In fact,  $F(y|\mathbf{x})$  together with the Markov property determines uniquely all the finite-dimensional distributions and therefore the whole dependence structure of  $\{X_t, t \geq 1\}$ . Similarly, the whole dependence structure of the local bootstrap series is determined by the bootstrap one-step transition distribution function  $F_T^*(\cdot|\mathbf{x})$ . Now, since under our assumptions,  $F_T^*(y|\mathbf{x}) \rightarrow F(y|\mathbf{x})$  strongly and uniformly in both arguments as  $T \rightarrow \infty$ , the above result justifies theoretically the use of the local bootstrap series in order to reproduce the dependence characteristics of the stochastic process generating the observations.

**Remark 3.2.** From a technical point of view, assumption (A3) on a bounded support of  $f(\cdot)$  can be avoided by extending the result of Theorem 3.2 to a strong uniform

convergence of  $F_T^*(\cdot | \mathbf{x})$  to  $F(\cdot | \mathbf{x})$  over compact subsets  $\mathbf{S}_T$  of  $\mathbf{R}^p$  which expand to all of  $\mathbf{R}^p$  as the sample size  $T$  increases. In particular, let  $\mathbf{S}_T$  be the ball in  $\mathbf{R}^p$  with center the origin and radius  $R_T$  and let  $R_T = T^\gamma$  for some  $\gamma > 0$ . Assume further that  $\inf_{\mathbf{x} \in \mathbf{S}_T} f(\mathbf{x}) \geq T^{-\eta}$  for some  $\eta > 0$ . It can then be shown along the same lines as in the proof of Theorem 3.2 (cf. also Theorem 3.2 of Roussas, 1988) that if the resampling width  $b$  is chosen so that  $b \asymp T^{-\delta}$  with  $0 < \eta + \delta p < 1/2$ , then as  $T \rightarrow \infty$

$$\sup_{y \in \mathbf{R}} \sup_{\mathbf{x} \in \mathbf{S}_T} |F_T^*(y | \mathbf{x}) - F(y | \mathbf{x})| \rightarrow 0 \quad \text{a.s.}$$

This implies that in case of an unbounded support of  $f(\cdot)$  and in order to obtain an asymptotically valid procedure, resampling should be done from a ‘truncated’ sequence of observations, denoted by  $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_T$ , where  $\tilde{X}_t = X_t$  if  $|X_t| \leq R_T$  and  $\tilde{X}_t = \text{sgn}(X_t)R_T$  if  $|X_t| > R_T$  and where the truncation level  $R_T$  tends to  $\infty$  at the rate given above as  $T \rightarrow \infty$ . We stress here the fact that such a truncation changes solely the state space of the bootstrap Markov chain and does not affect its Markov properties and characteristics.

The properties of the bootstrap stationary distribution function  $F_T^*$  are summarized in the following theorem.

**Theorem 3.3.** *Under the same assumptions as in Theorem 3.2, we have*

$$\sup_{y \in \mathbf{R}} |F_T^*(y) - F(y)| \rightarrow 0 \quad \text{a.s.}$$

Note that once a bootstrap realization has been obtained, the pseudo-observations  $X_1^*, X_2^*, \dots, X_N^*$  can be used to recalculate statistics of interest. In order to be able to obtain asymptotic bounds of covariances and to establish the large sample validity of the local bootstrap for a large class of statistical inference problems, we characterize in the following the asymptotic behavior of the weak dependence of the local bootstrap series by means of the  $\rho$ -mixing coefficient. This coefficient is defined as follows: Let  $\{Z_t, t \in \mathbf{N}\}$  be a sequence of random variables and let  $\mathcal{F}_1^J = \sigma(Z_k, 1 \leq k \leq J)$  and  $\mathcal{F}_{J+n}^\infty = \sigma(Z_k, J+n \leq k < \infty)$  be the  $\sigma$ -algebras generated by  $Z_1, Z_2, \dots, Z_J$  and  $Z_{J+n}, Z_{J+n+1}, \dots$ , respectively. The  $\rho$ -mixing coefficient is then defined by

$$\rho(n) = \sup\{|\text{Corr}(\xi_1, \xi_2)|; \xi_1 \text{ being } \mathcal{F}_1^J \text{-measurable, } \xi_2 \text{ being } \mathcal{F}_{J+n}^\infty \text{-measurable with } E|\xi_1|^2 < \infty, E|\xi_2|^2 < \infty \text{ and } J \geq 1\}.$$

The random sequence  $\{Z_k\}$  is called  $\rho$ -mixing if  $\rho(n) \rightarrow 0$  as  $n \rightarrow \infty$ . Note that  $\rho$ -mixing implies  $\alpha$ -mixing; cf. Doukhan (1994, p. 20). The following theorem can then be established, where as before we define:

$$\rho^*(n) = \sup\{|\text{Corr}(\xi_1, \xi_2)|; \xi_1 \text{ being } \mathcal{F}_1^{*J} \text{-measurable, } \xi_2 \text{ being } \mathcal{F}_{J+n}^{*\infty} \text{-measurable with } E|\xi_1|^2 < \infty, E|\xi_2|^2 < \infty \text{ and } J \geq 1\},$$

where  $\mathcal{F}_1^{*J}, \mathcal{F}_{J+n}^{*\infty}$  were defined in Eq. (5).

**Theorem 3.4.** Under the assumptions of Theorem 3.2 and conditionally on  $X_1, X_2, \dots, X_T$ ,  $T_0 \in \mathbf{N}$  exists such that for all  $T \geq T_0$  the  $\rho^*$ -mixing coefficient of the bootstrap series  $\{X_t^*, t \geq 1\}$  satisfies  $\rho_T^*(n) \leq \varrho^n$ , where the constant  $\varrho \in (0, 1)$  is independent of  $T$ .

#### 4. Applications

##### 4.1. Linear statistics

In this section we deal with applications of the local bootstrap procedure in estimating the sampling properties of linear statistics defined by

$$\hat{\theta}_T = \frac{1}{T - m + 1} \sum_{t=1}^{T-m+1} g(X_t, X_{t+1}, \dots, X_{t+m-1}), \tag{9}$$

where  $g: \mathbf{R}^m \rightarrow \mathbf{R}$  is a continuous function and  $m \in \mathbf{N}$ . To simplify notation, let  $Z_t = g(X_t, X_{t+1}, \dots, X_{t+m-1})$ , denote by  $\theta_g$  the mean  $\theta_g = E[Z_t]$  which is assumed to exist and set  $\theta = \theta_g$ . Under our assumptions on the process  $\{X_t\}$  and if  $E|Z_t|^{2+\kappa} < \infty$  for some  $\kappa > 0$  it follows by Theorem 1.7 of Ibragimov (1962) that

$$\sigma_g^2 = \lim_{T \rightarrow \infty} \text{Var}(\sqrt{T}\hat{\theta}_T) = \text{Var}(Z_1) + 2 \sum_{\tau=1}^{\infty} \text{Cov}(Z_1, Z_{1+\tau}) > 0$$

exists. Furthermore,

$$\mathcal{L}(\sqrt{T}(\hat{\theta}_T - \theta)) \Rightarrow N(0, \sigma_g^2),$$

where  $\mathcal{L}(Y)$  denotes the law of the random variable  $Y$  and  $\Rightarrow$  stands for weak convergence. Now, as an alternative to the normal approximation to  $\sqrt{T}(\hat{\theta}_T - \theta)$ , or just to be able to estimate the limiting variance  $\sigma_g^2$ , we propose in the following to use the bootstrap distribution of the statistic  $\sqrt{T}(\hat{\theta}_T^* - \theta^*)$  where  $\hat{\theta}_T^*$  is defined by

$$\hat{\theta}_T^* = \frac{1}{T - m + 1} \sum_{t=1}^{T-m+1} g(X_t^*, X_{t+1}^*, \dots, X_{t+m-1}^*) \tag{10}$$

and  $\theta^* = E^*[g(X_t^*, X_{t+1}^*, \dots, X_{t+m-1}^*)]$ . Here  $E^*$  denotes expectation with respect to the bootstrap distribution.

The following theorem shows that the bootstrap approximation proposed ‘works’ asymptotically in the sense that  $\sigma_{g,T}^{*2} = \text{Var}^*(\sqrt{T}\hat{\theta}_T^*)$  converges to the correct limit variance  $\sigma_g^2$ , and that the law of  $\sqrt{T}(\hat{\theta}_T^* - \theta^*)$  (conditionally on the data) is close to the law of  $\sqrt{T}(\hat{\theta}_T - \theta)$  with high probability, if the sample size  $T$  is large enough. In this theorem, Kolmogorov’s distance, defined by  $d_0(\mathcal{P}, \mathcal{Q}) = \sup_{x \in \mathbf{R}} |\mathcal{P}(X \leq x) - \mathcal{Q}(X \leq x)|$ , is used in order to estimate the distance between probability measures  $\mathcal{P}$  and  $\mathcal{Q}$ .

**Theorem 4.1.** *Assume that (A1)–(A3) and (B1)–(B2) are fulfilled. For  $g$  continuous and  $E|Z_t|^{2+\kappa} < \infty$  for some  $\kappa > 0$ , we have as  $T \rightarrow \infty$*

1.  $\sigma_{g,T}^{*2} \rightarrow \sigma_g^2$   
and
2.  $d_0\{\mathcal{L}(\sqrt{T}(\hat{\theta}_T^* - \theta^*)|X_1, X_2, \dots, X_T), \mathcal{L}(\sqrt{T}(\hat{\theta}_T - \theta))\} \rightarrow 0$ ,  
both convergences being valid with probability one.

The large sample validity of the local bootstrap procedure can be also established for multivariate versions of (9), i.e., for linear statistics where  $g$  is a continuous mapping from  $\mathbf{R}^m$  into  $\mathbf{R}^s$ , and  $s > 1$ . In particular, for  $x \in \mathbf{R}^m$ , let  $g(x) = (g_1(x), \dots, g_s(x))^\top$ , where  $g_u: \mathbf{R}^m \rightarrow \mathbf{R}$  for  $u = 1, 2, \dots, s$  are univariate functions. Denote by  $\hat{\theta}_T = (\hat{\theta}_{1,T}, \hat{\theta}_{2,T}, \dots, \hat{\theta}_{s,T})^\top$  the corresponding multivariate version of (9) where

$$\hat{\theta}_{u,T} = \frac{1}{T - m + 1} \sum_{t=1}^{T-m+1} g_u(X_t, X_{t+1}, \dots, X_{t+m-1}). \tag{11}$$

To approximate the distribution of  $\sqrt{T}(\hat{\theta}_T - \theta)$ , where  $\theta = E(\hat{\theta}_T)$ , the bootstrap statistic  $\sqrt{T}(\hat{\theta}_T^* - \theta^*)$  can be used where  $\hat{\theta}_T^* = (\hat{\theta}_{1,T}^*, \hat{\theta}_{2,T}^*, \dots, \hat{\theta}_{s,T}^*)^\top$  and

$$\hat{\theta}_{u,T}^* = \frac{1}{T - m + 1} \sum_{t=1}^{T-m+1} g_u(X_t^*, X_{t+1}^*, \dots, X_{t+m-1}^*).$$

Furthermore,  $\theta^* = E^*(\hat{\theta}_T^*)$ . If we define  $\sigma_{g,i,j}^2 = \lim_{T \rightarrow \infty} \text{Cov}(\sqrt{T}\hat{\theta}_{i,T}, \sqrt{T}\hat{\theta}_{j,T})$  and  $\sigma_{g,i,j,T}^{*2} = \text{Cov}^*(\sqrt{T}\hat{\theta}_{i,T}^*, \sqrt{T}\hat{\theta}_{j,T}^*)$ , for  $i, j = 1, \dots, s$ , we have the following result.

**Theorem 4.2.** *Assume that (A1)–(A3), (B1)–(B2) are fulfilled and that  $g: \mathbf{R}^m \rightarrow \mathbf{R}^s$  is continuous. If  $E|g_u(X_t, X_{t+1}, \dots, X_{t+m-1})|^{2+\kappa} < \infty$  for  $u = 1, 2, \dots, s$  and some  $\kappa > 0$ , then as  $T \rightarrow \infty$*

1.  $\sigma_{g,i,j,T}^{*2} \rightarrow \sigma_{g,i,j}^2$ , for any  $i, j = 1, \dots, s$ , and
2.  $d_0\{\mathcal{L}(\sqrt{T}(\hat{\theta}_T^* - \theta^*)|X_1, X_2, \dots, X_T), \mathcal{L}(\sqrt{T}(\hat{\theta}_T - \theta))\} \rightarrow 0$ ,  
both convergences being valid with probability one.

In the above, an obvious extension of Kolmogorov’s distance  $d_0(\mathcal{P}, \mathcal{Q})$  for probability measures  $\mathcal{P}$  and  $\mathcal{Q}$  on  $\mathbf{R}^s$  was used.

**Remark 4.1.** Theorem 4.2 and a simple application of the delta method leads to the asymptotic validity of the local bootstrap for smooth functions of linear statistics too, i.e., for statistics defined by  $\psi(\hat{\theta}_T)$  where  $\psi(\cdot): \mathbf{R}^s \rightarrow \mathbf{R}^l$  is a real-valued function satisfying certain smoothness conditions; see Shao and Tu (1995), or Brockwell and Davis (1991, Proposition 6.4.3) for more details. A prime example of such a smooth functions of linear statistics that is very useful in time-series analysis is given by the sample autocorrelations  $\hat{r}_s = \sum_{t=1}^{T-s} (X_t - \bar{X})(X_{t+s} - \bar{X}) / \sum_{t=1}^T (X_t - \bar{X})^2$ , for  $s = 1, 2, \dots, T - 1$ , and where  $\bar{X} = T^{-1} \sum_{t=1}^T X_t$  is the sample mean; we address this important example in the simulations of Section 5.

### 4.2. Detecting time reversibility

A characterization of stationary linear Gaussian time series  $\{X_t, t \in \mathbf{Z}\}$  is time reversibility, i.e., the property that for every integer  $r$  and for all integers  $t_1, t_2, \dots, t_n$ , the random vectors  $(X_{t_1}, X_{t_2}, \dots, X_{t_r})^\top$  and  $(X_{-t_1}, X_{-t_2}, \dots, X_{-t_r})^\top$  have the same joint probability distributions. In fact, Weiss (1975) showed that autoregressive processes satisfying  $X_t = \sum_{i=1}^p a_i X_{t-i} + \epsilon_t$  for some white-noise sequence  $\{\epsilon_t\}$  are time reversible if and only if they are Gaussian, i.e., if and only if  $\{\epsilon_t\}$  is a Gaussian sequence. Thus detecting time reversibility is an important aspect of time-series analysis. To avoid the complexity in the definition of time reversibility which is due to the fact that properties of a sequence of multivariate distributions are involved, statistics of less general form have been considered in the literature; cf. Tong (1990), Lawrence (1991), Tsay (1992). In the following we concentrate on bivariate distributions and on the so-called lag reversibility, i.e., the property that in the presence of time reversibility the joint distributions of  $(X_t, X_{t-r})$  and  $(X_{t-r}, X_t)$  are equal for all  $t$  and all  $r = 1, 2, \dots$ . Note that by (A2) the random vector  $(X_t, X_{t-r})$  has an absolutely continuous distribution. Since in the presence of lag reversibility the random variable  $D_{t,r} = X_t - X_{t-r}$  has a symmetric distribution, this implies  $P(D_{t,r} > 0) = P(D_{t,r} < 0) = 1/2$ .

Given observations  $X_1, X_2, \dots, X_T$  a simple estimator of  $P_r \equiv P(D_{t,r} > 0)$  is given by  $\hat{P}_r = (T - r)^{-1} \sum_{t=r+1}^T I(D_{t,r})$  where  $I(D_{t,r}) = 1$  if  $D_{t,r} > 0$  and  $I(D_{t,r}) = 0$  if  $D_{t,r} < 0$ . Note here that, since the  $X_t$  random variables have densities, we can safely ignore the case  $D_{t,r} = 0$  since it happens with probability zero; otherwise, the divisor of  $\hat{P}_r$  should be corrected to  $\tilde{T} - r$  where  $\tilde{T} = T - \#\{t : D_{t,r} = 0\}$ . Note also that since  $\{X_t\}$  is ergodic,  $\hat{P}_r \rightarrow E(\hat{P}_r) = P_r$  a.s. as  $T \rightarrow \infty$ . To discuss in more detail the sampling properties of  $\hat{P}_r$  consider the random variable  $I_{t,r} = I(D_{t,r})$ . Straightforward calculations yield

$$\begin{aligned} (T - r)\text{Var}(\hat{P}_r) &= \text{Var}[I_{r+1,r}] + 2(T - r)^{-1} \sum_{s=1}^{T-r-1} (T - r - s) \text{Cov}(I_{r+1,r}, I_{r+1+s,r}) \\ &= P(D_{r+1,r} > 0) \left\{ P(D_{r+1,r} \leq 0) + 2(T - r)^{-1} \sum_{s=1}^{T-r-1} (T - r - s) \right. \\ &\quad \left. \times [P(D_{r+1+s,r} > 0 | D_{r+1,r} > 0) - P(D_{r+1,r} > 0)] \right\}. \end{aligned}$$

By the properties of  $\{X_t\}$ , the process  $\{I_{t,r}, t \geq r + 1\}$  is strictly stationary and ergodic. Furthermore,  $\{I_{t,r}, t \geq r + 1\}$  is strongly mixing with geometrically decreasing mixing coefficient, i.e. for  $n \geq r + 1$  we have  $\alpha_I(n) = \alpha_X(n - r)$  where  $\alpha_I$  and  $\alpha_X$  denote the strong mixing coefficient of  $\{I_{t,r}\}$  and  $\{X_t\}$ , respectively. Thus for  $0 < P(D_{r+1,r} > 0) < 1$  a simple application of Theorem 1.6 of Ibragimov (1962) yields that  $(T - r)\text{Var}(\hat{P}_r) \rightarrow \sigma_{r,\infty}^2 > 0$  as  $T \rightarrow \infty$  with

$$\begin{aligned} \sigma_{r,\infty}^2 &= P(D_{r+1,r} > 0) \left\{ P(D_{r+1,r} \leq 0) \right. \\ &\quad \left. + 2 \sum_{s=1}^{\infty} [P(D_{r+1+s,r} > 0 | D_{r+1,r} > 0) - P(D_{r+1,r} > 0)] \right\}. \end{aligned}$$

Furthermore,

$$\mathcal{L}(\sqrt{T-r}(\hat{P}_r - P_r)) \Rightarrow N(0, \sigma_{r,\infty})^2.$$

It is worth mentioning here that despite the above asymptotic result and for  $T$  given, the statistic  $\hat{P}_r$  has a discrete distribution the support of which is the set  $\{i/(T-r), i = 0, 1, \dots, T-r\}$ .

The conditional distributions involved in the expression of the asymptotic variance  $\sigma_{r,\infty}^2$  given above, are difficult to calculate in practice, making the established asymptotic result of limited value for practical applications, e.g., for the construction of large sample confidence intervals for  $P_r$ . Consequently, to approximate the distribution of  $\sqrt{T-r}(\hat{P}_r - P_r)$  we propose in the following to use the bootstrap statistic  $\sqrt{T-r}(\hat{P}_r^* - P_r^*)$  where  $\hat{P}_r^* = (T-r)^{-1} \sum_{t=r+1}^T I(D_{t,r}^*)$ ,  $P_r^* = P^*(D_{t,r+1}^* > 0)$  and  $D_t^* = X_t^* - X_{t-r}^*$ . Moreover, the bootstrap variance  $\sigma_{r,T}^{*2} = (T-r) \text{Var}^*(\hat{P}_r^*)$  can be used to estimate the unknown variance  $\sigma_{r,T}^2$ .

**Proposition 4.1.** *Under assumptions (A1)–(A3) and (B1)–(B2) we have as  $T \rightarrow \infty$*

1.  $\sigma_{r,T}^{*2} \rightarrow \sigma_{r,\infty}^2$   
and
2.  $d_0 \{ \mathcal{L}((T-r)^{1/2}(\hat{P}_r^* - P_r^*) | X_1, X_2, \dots, X_T), \mathcal{L}((T-r)^{1/2}(\hat{P}_r - P_r)) \} \rightarrow 0,$   
both convergences being valid with probability one.

## 5. Practical aspects and examples

### 5.1. Some remarks on choosing the resampling parameters

The asymptotic results in the previous section have been obtained under certain assumptions on the resampling kernel  $W$  and the assumption that the resampling width  $b$  goes to zero as  $T \rightarrow \infty$ . However, and for a given sample size  $T$  at hand, some decisions have to be made before applying the local bootstrap procedure in practice. In particular, the practitioner is faced with choosing the resampling kernel  $W$  and the resampling width  $b$ . Recall that in the local resampling scheme, the essential quantity is  $F_T^*(\cdot | \mathbf{x})$  which is used implicitly to mimic the behavior of  $F(\cdot | \mathbf{x})$ . One can, therefore, argue that the quality of the bootstrap approximations proposed depends heavily on how close the function  $F_T^*(\cdot | \mathbf{x})$  is to the unknown  $F(\cdot | \mathbf{x})$ . Observing that

$$F_T^*(\cdot | \mathbf{x}) - F(\cdot | \mathbf{x}) = \{E(F_T^*(\cdot | \mathbf{x})) - F(\cdot | \mathbf{x})\} + \{F_T^*(\cdot | \mathbf{x}) - E(F_T^*(\cdot | \mathbf{x}))\},$$

and assuming that  $W(\mathbf{u})$  is a product kernel given by  $W(\mathbf{u}) = \prod_{i=1}^p W(u_i)$  with  $0 < W_2 = \int u_i^2 W(u_i) du_i < \infty$ , it can be shown under appropriate smoothness conditions on  $F(\cdot | \mathbf{x})$  and by the same arguments as in Auestad and Tjøstheim (1990), that

$$\begin{aligned} \text{Bias}(F_T^*(\cdot | \mathbf{x})) &= E(F_T^*(\cdot | \mathbf{x})) - F(\cdot | \mathbf{x}) \\ &\approx \frac{b^2 W_2}{2} \sum_{i=1}^p \left( F_i^{(2)}(\cdot | \mathbf{x}) + 2F_i^{(1)}(\cdot | \mathbf{x}) \frac{f_i^{(1)}(\mathbf{x})}{f(\mathbf{x})} \right) \end{aligned}$$

and

$$\begin{aligned} \text{Var}(F_T^*(\cdot|\mathbf{x})) &= E(F_T^*(\cdot|\mathbf{x}) - E(F_T^*(\cdot|\mathbf{x})))^2 \\ &\approx \frac{1}{Tb^p f(\mathbf{x})} \int W^2(\mathbf{u}) \mathbf{d}\mathbf{u} F(\cdot|\mathbf{x})(1 - F(\cdot|\mathbf{x})), \end{aligned}$$

where  $F_i^{(k)}(\cdot|\mathbf{x})$  and  $f_i^{(k)}(\mathbf{x})$  denote the  $k$ th partial derivatives of  $F(\cdot|\mathbf{x})$  and  $f(\mathbf{x})$  with respect to  $x_i$ ,  $i \in \{1, 2, \dots, p\}$ , respectively. Observe that increasing  $b$  increases the bias, while decreasing  $b$  increases the variance. A ‘good’ choice of  $b$  can therefore be achieved by minimizing a measure of the stochastic ‘distance’ of  $F_T^*(\cdot|\mathbf{x})$  to  $F(\cdot|\mathbf{x})$ ; a commonly used such measure is the mean squared error of  $F_T^*(\cdot|\mathbf{x})$ . Choosing the  $b$  that minimizes the mean squared error of  $F_T^*(\cdot|\mathbf{x})$  effectively ‘balances’ the bias (squared) and the variance terms given above. In this case the ‘optimal’  $b$  minimizes

$$\int \{[\text{Bias}(F_T^*(y|\mathbf{x}))]^2 + \text{Var}(F_T^*(y|\mathbf{x}))\} \mathbf{d}y. \tag{12}$$

Since the above expression depends on unknown quantities further work is needed for achieving a practical width selection procedure. One approach, although computationally cumbersome, will be to substitute the unknown quantities in (12) by some nonparametric pilot estimators. An alternative approach will be to consider a cross-validation rule for selecting  $b$  as in Härdle and Vieu (1992) and Kim and Cox (1996).

A different approach for selecting  $b$  that may prove quite appealing in practice can be obtained by the following considerations. Although time series  $\{X_t\}$  is generally nonlinear we may attempt—for the purposes of getting a rough estimate of the value of  $b$  which minimizes (12)—to fit a linear model to our data as an approximation. A simple rule-of-thumb approach for selecting  $b$ —that also simplifies the calculations considerably—is then obtained if we assume that  $X_t = \delta + \sum_{i=1}^p a_i X_{t-i} + \epsilon_t$  with  $\{\epsilon_t\}$  an i.i.d. sequence. For this particular case and under the simplifying assumption that  $\epsilon_t \sim N(0, \sigma^2)$  it is shown in Appendix B of the paper, that the ‘optimal’ resampling width  $b = b(\mathbf{x})$  that minimizes (12) is given by

$$b_{AR}(\mathbf{x}) = \left[ \frac{\sigma^4 W_1}{Tf(\mathbf{x})W_2^2 \{2\sigma^2 C_1^2(\mathbf{x}) + 0.25C_2^2\}} \right]^{1/(p+4)}, \tag{13}$$

where  $W_1 = \int W^2(\mathbf{u}) \mathbf{d}\mathbf{u}$ ,  $C_1(\mathbf{x}) = \mathbf{a}^\top \mathbf{\Gamma}_p^{-1}(\mathbf{x} - \boldsymbol{\mu})$  and  $C_2 = \mathbf{a}^\top \mathbf{a}$ . Here  $\boldsymbol{\mu} = (\mu, \mu, \dots, \mu)^\top$ ,  $\mu = E(X_t)$ ,  $\mathbf{a} = (a_1, a_2, \dots, a_p)^\top$  and  $\mathbf{\Gamma}_p$  is the  $p \times p$  matrix having the process autocovariance  $\gamma(i - j) = E(X_{t-j} - \mu)(X_{t-i} - \mu)$  in the  $(i, j)$ th position. Note that a sample version  $\hat{b}_{AR}(\mathbf{x})$  of  $b_{AR}(\mathbf{x})$  is easily obtained by fitting the imposed linear structure to the data and replacing the unknown quantities in (13) by the corresponding sample estimates. We stress here the fact that the resampling width  $b_{AR}(\mathbf{x})$  depends on the particular state  $\mathbf{x}$  of the process, i.e., it changes with each value resampled. The resampling width  $b_{AR}(\mathbf{x})$  given in (13) can now be used in general situations as a crude approximation to the minimizer of (12). Compared to the approaches mentioned previously, such a rule-of-thumb selection of the resampling width  $b$  has the advantage

of possessing simplicity and of requiring little programming effort; cf. Fan and Gijbels (1996) for similar approaches in a different context.

### 5.2. Numerical examples

In this section, some simulated and real data examples are used in order to demonstrate the finite sample behavior of the local bootstrap proposal.

**Example 5.1.** Samples of size  $T = 100$  and  $200$  have been generated according to the following four Markov models:

$$\text{AR: } X_t = 0.8X_{t-1} - 0.6X_{t-2} + \epsilon_t,$$

$$\text{ARC: } X_t = 0.8X_{t-1} - 0.6X_{t-2} + u_t,$$

$$\text{NLAR: } X_t = 0.8 \log(1 + 3X_{t-1}^2) - 0.6 \log(1 + 3X_{t-3}^2) + \epsilon_t,$$

$$\text{EXP: } X_t = -0.5 \exp\{-50X_{t-1}^2\}X_{t-1} - (0.9 - 1.3 \exp\{-50X_{t-1}^2\})X_{t-2} + \epsilon_t,$$

where the errors  $\{\epsilon_t\}$  are i.i.d., and the  $\{u_t\}$  are i.i.d. as well. The distribution of the  $\epsilon_t$  is Gaussian  $(0, 1)$ , while that of  $u_t$  is a mixture of 90% Gaussian  $(-1, 1)$  and 10% Gaussian  $(9, 1)$ . The AR and NLAR model have been used by Chen and Tsay (1993) while the EXP model by Auestad and Tjøstheim (1990). Note that the AR model is time reversible while the ARC model not.

To gauge the ability of the local resampling scheme to estimate the standard deviation of the sample lag reversibility coefficients  $\hat{P}_r$ , Table 1 is presented that shows the results obtained for the standard deviations  $\sigma_{1,T}$  and  $\sigma_{2,T}$  of  $\hat{P}_1$  and  $\hat{P}_2$ , respectively. The “exact” value of these standard deviations has been estimated using 5000 replications of the models considered while the bootstrap estimates are based on  $B = 250$  bootstrap replications. The Gaussian product kernel and a resampling width selection based on Eq. (13) have been used to obtain the bootstrap replicates. Reported in Table 1, are mean values and standard errors of the bootstrap estimates based on 400 trials.

Table 2 shows the results obtained in using the local bootstrap to estimate the standard deviation of the sample lag-1 autocorrelation coefficient  $\hat{r}_1 = \frac{\sum_{t=1}^{T-1} (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}$  where  $\bar{X} = T^{-1} \sum_{t=1}^T X_t$ . Sample sizes of  $T = 100$  and  $200$  have been used for the four Markov models considered. Again, the calculation of the “exact” values of the standard deviation of  $\hat{r}_1$  are based on 5000 replications of each model. To evaluate the performance of the local bootstrap, 400 trials are used. The bootstrap approximations reported in this table are based on  $B = 250$  bootstrap replications in each trial where the resampling parameters used are identical to those of the first part of the simulation experiment.

**Example 5.2.** The series of sunspot numbers is a well-known series showing several features of nonlinearity. For this series, we apply the local resampling scheme with  $p = 9$ , the Gaussian resampling kernel and a resampling width selected according to (13); cf. also Example 2.1. Note that to calculate (13), a linear autoregressive process of order  $p = 9$  has been fitted to the data using the Durbin–Levinson algorithm, cf. Brockwell and Davis (1987). To get an overall impression of the lag reversibility

Table 1

Estimated “exact” and bootstrap estimates of the standard deviation  $\sigma_{r,T}$  of the sample lag-reversibility coefficient  $\sqrt{T-r}\hat{P}_r$ ,  $r = 1, 2$ ; here,  $\sigma_{r,T}$  denotes the estimated “exact” standard deviation,  $E(\sigma_{r,T}^*)$  the mean, and  $STD(\sigma_{r,T}^*)$  the standard error of the bootstrap estimates

	$\sigma_{r,T}$	$E(\sigma_{r,T}^*)$	$E(\sigma_{r,T}^*)/\sigma_{r,T}$	$STD(\sigma_{r,T}^*)$
	$T = 100$			
	$\sqrt{T-1}\hat{P}_1$			
AR	0.291	0.286	0.983	0.0233
ARC	0.271	0.279	1.029	0.0249
NLAR	0.306	0.299	0.977	0.0209
EXP	0.237	0.244	1.029	0.0398
	$\sqrt{T-2}\hat{P}_2$			
AR	0.281	0.273	0.971	0.0207
ARC	0.258	0.269	1.043	0.0247
NLAR	0.274	0.276	1.001	0.0185
EXP	0.195	0.199	1.027	0.0357
	$T = 200$			
	$\sqrt{T-1}\hat{P}_1$			
AR	0.286	0.285	0.996	0.0192
ARC	0.265	0.273	1.030	0.0157
NLAR	0.304	0.299	0.983	0.0173
EXP	0.240	0.244	1.017	0.0308
	$\sqrt{T-2}\hat{P}_2$			
AR	0.276	0.274	0.993	0.0171
ARC	0.251	0.267	1.063	0.0161
NLAR	0.269	0.275	1.022	0.0172
EXP	0.193	0.198	1.025	0.0265

Table 2

Estimated “exact” and bootstrap estimates of the standard deviation  $\sigma_T$  of the first-order sample autocorrelation coefficient  $\hat{r}_1$ ; here,  $\sigma_T$  denotes the estimated “exact” standard deviation,  $E(\sigma_T^*)$  the mean, and  $STD(\sigma_T^*)$  the standard error of the bootstrap estimates

	$\sigma_T$	$E(\sigma_T^*)$	$E(\sigma_T^*)/\sigma_T$	$STD(\sigma_T^*)$
	$T = 100$			
AR	0.0451	0.0470	1.0421	0.00911
ARC	0.0433	0.0446	1.0300	0.01297
NLAR	0.0845	0.0815	0.9645	0.01452
EXP	0.0289	0.0302	1.0449	0.00813
	$T = 200$			
AR	0.0312	0.0324	1.0385	0.00497
ARC	0.0312	0.0306	0.9807	0.00650
NLAR	0.0597	0.0589	0.9859	0.00834
EXP	0.0192	0.0203	1.0573	0.00441

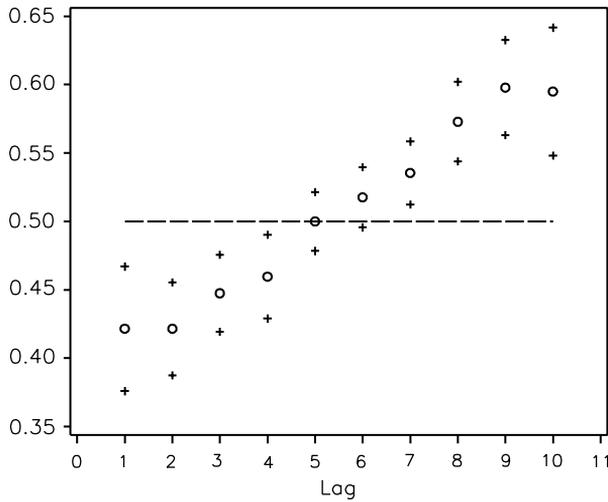


Fig. 2. Lag reversibility plot of the series of sunspot number together with pointwise 95% bootstrap confidence intervals for  $P_r$ ,  $r = 1, 2, \dots, 10$ .

behavior of this series and to demonstrate the directionality apparent in this series, Fig. 2 shows a *lag reversibility plot*, i.e., a plot of the estimated coefficients  $P_r$  for the values  $r = 1, 2, \dots, 10$  together with their corresponding 95% points confidence intervals. The dashed horizontal line drawn in this plot corresponds to the values  $P_r = 0.5$  which are the expected values of  $\hat{P}_r$  in the presence of lag reversibility. The confidence intervals for  $P_r$  shown in Fig. 2 have been calculated using the formula  $[\hat{P}_r - 1.96\sigma_{r,T}^*, \hat{P}_r + 1.96\sigma_{r,T}^*]$  where  $\sigma_{r,T}^* = \sqrt{\sigma_{r,T}^{*2}}$  is the bootstrap estimate of the standard error, and using  $B = 500$  bootstrap replications of the series considered.

### 6. Conclusions

In this paper, we introduced a new nonparametric resampling procedure, i.e., the local bootstrap, which is applicable to stationary stochastic processes that are essentially Markov of general order  $p$ . The local bootstrap manages to create bootstrap pseudo-series by implicitly imitating the Markov property of the process *without* explicit nonparametric estimation of the transition probability measure. Similarly, if the process in question is assumed to be generated by the nonlinear autoregression (1) with conditional mean and variance given by the (unknown) functions  $m$  and  $v$ , the local bootstrap “works” *without* explicit nonparametric estimation of  $m$  and  $v$  and residual resampling. Also noteworthy is the fact that, as in the classical i.i.d. bootstrap of Efron (1979), the bootstrap pseudo-series consist of (particular) reshufflings of the *original* data points  $X_1, \dots, X_T$ .

The assumption of a Markov dependence of order  $p$  is satisfied by a great class of nonlinear models that are of interest in the time-series literature; see e.g. Tong

(1990). Thus, our procedure is potentially applicable to a large variety of statistical inference problems in nonlinear time-series analysis. In this paper, we have focused on some specific applications for which the asymptotic validity of the local bootstrap procedure has been established. We have also presented a limited number of real data and finite-sample simulations indicating the performance of the local bootstrap in such situations.

**Acknowledgements**

The authors thank the editor and, referees and Michael Neumann for very helpful comments.

**Appendix A. Proofs**

**Proof of Theorem 3.1.** For  $X_1, X_2, \dots, X_T$  given, let  $B(\tilde{\mathcal{E}}_T)$  be the  $\sigma$ -field of all finite subsets of  $\tilde{\mathcal{E}}_T = \tilde{\mathcal{X}}_T^p$  where  $\tilde{\mathcal{X}}_T^p$  denotes the Cartesian product of  $p$  copies of  $\tilde{\mathcal{X}}_T \equiv \{X_1, X_2, \dots, X_T\}$ . Let  $\Omega_T = \prod_{k=1}^\infty \tilde{\mathcal{E}}_{k,T}$  where  $\tilde{\mathcal{E}}_{k,T}$  is a copy of  $\tilde{\mathcal{E}}_T$  equipped with a copy of the  $\sigma$ -field  $B(\tilde{\mathcal{E}}_T)$  and denote by  $\mathcal{A}_T$  the product  $\sigma$ -field on  $\Omega_T$ . Consider now the Markov chain  $\mathbf{Y}^* = \{\mathbf{Y}_t^*, t \geq 1\}$  on the path space  $(\Omega_T, \mathcal{A}_T, P_{\pi_0}^*)$  defined as follows: For any  $A \in \mathcal{A}_T$  and for  $\pi_0$  an appropriate initial distribution,  $P_{\pi_0}^*(\mathbf{Y}^* \in A)$  describes the probability of the event  $\mathbf{Y}^* \in A$  when  $\mathcal{L}(\mathbf{Y}_1^*) = \pi_0$ . Apart from the initial distribution  $\pi_0$ ,  $P_{\pi_0}^*(\cdot)$  is determined by the temporarily homogeneous transition probability kernel  $P(\mathbf{x}, \mathbf{z}) = P(\mathbf{Y}_{t+1}^* = \mathbf{z} | \mathbf{Y}_t^* = \mathbf{x})$  which for  $\mathbf{Y}_t^* = \mathbf{x}$  and  $\mathbf{x} = (x_1, x_2, \dots, x_p) \in \tilde{\mathcal{E}}_T$  is given by

$$P(\mathbf{x}, \mathbf{z}) = \begin{cases} \frac{W_b(\mathbf{x} - \mathbf{Y}_j)}{\sum_{m=p}^{T-1} W_b(\mathbf{x} - \mathbf{Y}_m)} & \text{if } \mathbf{z} = (X_{j+1}, x_1, x_2, \dots, x_{p-1}), p \leq j \leq T - 1, \\ 0 & \text{else.} \end{cases} \tag{A.1}$$

Now, verify that by (B1) for every  $b > 0$ ,  $\mathcal{E}_T \subset \tilde{\mathcal{E}}_T$  with  $\mathcal{E}_T = \mathcal{X}_T^p$  and  $\mathcal{X}_T = \{X_{p+1}, X_{p+2}, \dots, X_T\}$  is an absorbing communicating class; i.e.,  $\sum_{\mathbf{z} \in \mathcal{E}_T} P(\mathbf{x}, \mathbf{z}) = 1$  for all  $\mathbf{x} \in \mathcal{E}_T$ . Furthermore, by Theorem 2.3 of Doob (1953, p. 180),  $t_0 \in \mathbf{N}$  exists such that  $P(\mathbf{Y}_{t_0}^* \in \mathcal{E}_T | \mathbf{Y}_1^* = \mathbf{x}) = 1$  for all  $\mathbf{x} \in \tilde{\mathcal{E}}_T \setminus \mathcal{E}_T$ . By Proposition 4.1.2 of Meyn and Tweedie (1993, p. 84), it follows then that an irreducible Markov chain denoted by  $\mathcal{Y}_T^*$  exists whose state space is restricted to  $\mathcal{E}_T$  and with transition probability matrix  $\mathbf{P}_{\mathcal{E}_T}$ , i.e., the restriction of the matrix  $\mathbf{P} = (P(\mathbf{x}, \mathbf{z}))_{\mathbf{x}, \mathbf{z} \in \tilde{\mathcal{E}}_T}$  to the absorbing and communicating class  $\mathcal{E}_T$ . The positive recurrence of the chain is a simple consequence of  $|\mathcal{E}_T| < \infty$  while the aperiodicity follows from (A.1) and the fact that by (B1),  $W_b(\mathbf{x} - \mathbf{Y}_j) > 0$  for every  $b > 0$  and  $p \leq j \leq T - 1$ .  $\square$

Since we are dealing in the sequel with the asymptotic properties of the local re-sampling procedure we ignore in what follows the effects of the starting values and assume that the bootstrap chain  $\mathcal{Y}_T^*$  starts with its stationary distribution (6), i.e., that  $\mathbf{Y}_p^* \sim \pi_T^*$ .

**Lemma 6.1.** Under the assumptions of Theorem 3.2 we have for every  $y \in \mathbf{R}$ ,

$$\sup_{\mathbf{x} \in \mathbf{S}} \left| \frac{1}{T} \sum_{j=p}^{T-1} 1_{(-\infty, y]}(X_{j+1}) W_b(\mathbf{x} - \mathbf{Y}_j) - \int 1_{(-\infty, y]}(z) f_{X_{T+1}, \mathbf{Y}_T}(z, \mathbf{x}) dz \right| \rightarrow 0 \quad a.s.$$

as  $T \rightarrow \infty$ .

**Proof.** In the following we use the shorthand  $f(z, \mathbf{x})$  for  $f_{X_{T+1}, \mathbf{Y}_T}(z, \mathbf{x})$ . The result is established by showing that

$$\sup_{\mathbf{x} \in \mathbf{S}} |E[r_b^*(\mathbf{x})] - r(\mathbf{x})| \rightarrow 0 \quad a.s. \tag{A.2}$$

and

$$\sup_{\mathbf{x} \in \mathbf{S}} |r_b^*(\mathbf{x}) - E[r_b^*(\mathbf{x})]| \rightarrow 0 \quad a.s., \tag{A.3}$$

where

$$r_b^*(\mathbf{x}) = \frac{1}{T-p} \sum_{j=p}^{T-1} 1_{(-\infty, y]}(X_{j+1}) W_b(\mathbf{x} - \mathbf{Y}_j)$$

and

$$r(\mathbf{x}) = \int 1_{(-\infty, y]}(z) f(z, \mathbf{x}) dz.$$

Note that  $r_b^*(\mathbf{x})$  is asymptotically equivalent to  $T^{-1} \sum_{j=p}^{T-1} 1_{(-\infty, y]}(X_{j+1}) W_b(\mathbf{x} - \mathbf{Y}_j)$ . For (A.2) we have

$$\begin{aligned} \sup_{\mathbf{x} \in \mathbf{S}} |E[r_b^*(\mathbf{x})] - r(\mathbf{x})| &= \sup_{\mathbf{x} \in \mathbf{S}} \left| \int_{-\infty}^y \int (f(z, \mathbf{x} - \mathbf{u}b) - f(z, \mathbf{x})) W(\mathbf{u}) d\mathbf{u} dz \right| \\ &\leq \sup_{\mathbf{x} \in \mathbf{S}} \int \left| \int_{-\infty}^y (f(z, \mathbf{x} - \mathbf{u}b) - f(z, \mathbf{x})) dz \right| W(\mathbf{u}) d\mathbf{u} \\ &= b L(y) \int \|\mathbf{u}\| W(\mathbf{u}) d\mathbf{u} \\ &= O(b), \end{aligned}$$

where the last inequality follows using (A.2)(ii) and (B1).

To handle (A.3) we apply the technique of splitting the supremum over  $\mathbf{x} \in \mathbf{S}$ . Since  $\mathbf{S}$  is compact it can be covered by a finite number  $N_T$  of cubes  $I_{i,T}$  with centers  $\mathbf{x}_i$  the sides of which are of length  $L_T$ . From (B1) and the mean value theorem we then have

$$\begin{aligned} &\sup_{\mathbf{x} \in \mathbf{S}} |r_b^*(\mathbf{x}) - E[r_b^*(\mathbf{x})]| \\ &= \max_{1 \leq i \leq N_T} \sup_{\mathbf{x} \in \mathbf{S} \cap I_i} \left| \frac{1}{T} \sum_{j=p}^{T-1} 1_{(-\infty, y]}(X_{j+1}) W_b(\mathbf{x} - \mathbf{Y}_j) \right| \end{aligned}$$

$$\begin{aligned}
 & -E[1_{(-\infty, y]}(X_{j+1})W_b(\mathbf{x} - \mathbf{Y}_j)] \\
 & \leq \max_{1 \leq i \leq N_T} \left| \frac{1}{T} \sum_{j=p}^{T-1} Z_{j,T}(\mathbf{x}_i) \right| + C^* L_T b^{-(p+1)},
 \end{aligned}$$

where  $C^* > 0$  and

$$Z_{j,T}(\mathbf{x}_i) = 1_{(-\infty, y]}(X_{j+1})W_b(\mathbf{x}_i - \mathbf{Y}_j) - E[1_{(-\infty, y]}(X_{j+1})W_b(\mathbf{x}_i - \mathbf{Y}_j)].$$

Now, by (B1),  $|Z_{j,T}(\mathbf{x}_i)| \leq Mb^{-p}$ . For  $\epsilon > 0$  arbitrary, let  $L_T = \epsilon b^{p+1}/2C^*$  and note that  $N_T = O(L_T^{-p})$ . Applying inequality (1.4) of Roussas (1988, p. 136), we get

$$\begin{aligned}
 P \left( \max_{1 \leq i \leq N_T} \left| \frac{1}{T} \sum_{j=p}^{T-1} Z_{j,T}(\mathbf{x}_i) \right| \geq \frac{\epsilon}{2} \right) & \leq N_T \max_{1 \leq i \leq N_T} P \left( \frac{1}{T} \left| \sum_{j=p}^{T-1} Z_{j,T}(\mathbf{x}_i) \right| \geq \frac{\epsilon}{2} \right) \\
 & \leq O(L_T^{-p}) \tilde{C} T^{-l/2 + l\delta p}
 \end{aligned}$$

for any  $l > 2$  such that  $\sum_{j=1}^{\infty} (j+1)^{l/2-1} \alpha(j) < \infty$ , where  $\alpha(\cdot)$  is the (strong) mixing coefficient of  $\{X_t, t \geq 1\}$  and  $\tilde{C} > 0$ . Now, by the above choice of  $L_T$  we get that in order for the Borel–Cantelli Lemma to apply, the condition  $l/2 - l\delta p - \delta p(p+1) > 1$  should be satisfied, i.e.,

$$0 < \delta < \frac{l}{2lp + 2p(p+1)} - \frac{1}{lp + p(p+1)}.$$

Finally,  $0 < \delta < 1/2p$  follows because by the exponential decrease of the strong mixing coefficient  $\alpha(j)$  we have  $\sum_{j=1}^{\infty} (j+1)^{l/2-1} \alpha(j) < \infty$  for every  $l > 2$ .  $\square$

For  $\mathbf{x} \in \mathbf{S}$ , let

$$f_b^*(\mathbf{x}) = \frac{1}{T} \sum_{j=p}^{T-1} W_b(\mathbf{x} - \mathbf{Y}_j).$$

Setting  $y = \infty$  in the proof of the above lemma we immediately get  $\sup_{\mathbf{x} \in \mathbf{S}} |f_b^*(\mathbf{x}) - f(\mathbf{x})| \rightarrow 0$  a.s.; cf. also Theorem 3.2 of Roussas (1988).

**Proof of Theorem 3.2.** It suffices to show the strong uniform convergence of  $F_T^*(y|\mathbf{x})$  to  $F(y|\mathbf{x})$  over  $\mathbf{S}$  for every  $y \in \mathbf{R}$ . The asserted uniform convergence over  $y \in \mathbf{R}$  follows then by the same arguments as in the proof of the Glivenko–Cantelli Theorem; cf. also the proof of Theorem 3.1 in Roussas (1969, p. 1390). Now, since

$$|F_T^*(y|\mathbf{x}) - F(y|\mathbf{x})| \leq \frac{1}{f_b^*(\mathbf{x})} \{ |r_b^*(\mathbf{x}) - r(\mathbf{x})| + F(y|\mathbf{x}) |f_b^*(\mathbf{x}) - f(\mathbf{x})| \},$$

the result follows by (A.3) and Lemma 6.1.  $\square$

**Proof of Theorem 3.3.** As in the proof of Theorem 3.2 it suffices to show the strong convergence of  $F_T^*(y)$  to  $F(y)$  for every  $y \in \mathbf{R}$ . Now, for almost all sample paths  $\{X_t, t \geq 1\}$  we have by Theorem 3.2 that  $F_T^*(y|\mathbf{x}) \rightarrow F(y|\mathbf{x})$ . So take any one of the

sample paths for which the above holds and consider the sequence of bootstrap stationary distribution functions  $\{F_{\mathbf{Y},T}^*(\mathbf{y})\}_{T \in \mathbb{N}}$  corresponding to the sequence of bootstrap Markov chains  $\{\mathcal{Q}_T^*\}_{T=1,2,\dots}$  where for  $\mathbf{y} = (y_1, y_2, \dots, y_p)$  and  $\mathbf{x} = (x_1, x_2, \dots, x_p)$ ,

$$\begin{aligned} F_{\mathbf{Y},T}^*(\mathbf{y}) &= \int F_{\mathbf{Y},T}^*(\mathbf{y}|\mathbf{x}) dF_{\mathbf{Y},T}^*(\mathbf{x}) \\ &= \int \prod_{i=2}^p 1_{(-\infty, y_i]}(x_{i-1}) F_T^*(y_1|\mathbf{x}) dF_{\mathbf{Y},T}^*(\mathbf{x}). \end{aligned}$$

By Helly’s selection theorem, Billingsley (1995, p. 336), a subsequence  $\{F_{\mathbf{Y},T_k}^*(\mathbf{y})\}_{k=1,2,\dots}$  exists such that for all continuity points  $\mathbf{y} \in \mathbf{R}^p$  of  $G$ ,

$$\lim_{k \rightarrow \infty} F_{\mathbf{Y},T_k}^*(\mathbf{y}) = G(\mathbf{y}),$$

where  $G$  is a right continuous, nondecreasing function satisfying  $0 \leq G(\mathbf{y}) \leq 1$ . Since by (A3) the sequence  $\{F_{\mathbf{Y},T}^*(\mathbf{y})\}$  is tight,  $G$  is also a distribution function. Let  $g : \mathbf{R}^p \rightarrow \mathbf{R}$  be any bounded and continuous function and note first that  $\lim_{k \rightarrow \infty} \int g(\mathbf{y}) dF_{\mathbf{Y},T_k}^*(\mathbf{y}) \rightarrow \int g(\mathbf{y}) dG(\mathbf{y})$ . Using the notation  $\mathbf{y}_{p-1} = (y_1, y_2, \dots, y_{p-1})$  we further get

$$\begin{aligned} \int g(z, \mathbf{y}_{p-1}) dF_{\mathbf{Y},T_k}^*(z, \mathbf{y}_{p-1}) &= \int g(z, \mathbf{y}_{p-1}) F_{T_k}^*(dz|\mathbf{y}) dF_{\mathbf{Y},T_k}^*(\mathbf{y}) \\ &= \int g(z, \mathbf{y}_{p-1}) F(dz|\mathbf{y}) dF_{\mathbf{Y},T_k}^*(\mathbf{y}) \\ &\quad + \int g(z, \mathbf{y}_{p-1}) (F_{T_k}^*(dz|\mathbf{y}) - F(dz|\mathbf{y})) dF_{\mathbf{Y},T_k}^*(\mathbf{y}) \\ &\rightarrow \int g(z, \mathbf{y}_{p-1}) F(dz|\mathbf{y}) dG(\mathbf{y}) \end{aligned}$$

as  $k \rightarrow \infty$ , where the last convergence follows by Theorem 3.2, the continuity of  $\tilde{g}(\mathbf{y}) = \int g(z, \mathbf{y}_{p-1}) F(dz|\mathbf{y})$  and the weak convergence  $F_{\mathbf{Y},T_k}^* \Rightarrow G$ . Thus  $G$  is a stationary distribution of  $\mathbf{Y}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})$  from which  $G = F_{\mathbf{Y}}$  follows by the uniqueness of  $F_{\mathbf{Y}}$ . Since the above is true for every subsequence  $\{F_{\mathbf{Y},T_k}^*(\mathbf{y})\}$  of  $\{F_{\mathbf{Y},T}^*(\mathbf{y})\}$  that converge at all, we conclude by a corollary in Billingsley (1995, p. 381), that

$$\lim_{T \rightarrow \infty} F_{\mathbf{Y},T}^*(\mathbf{y}) = F_{\mathbf{Y}}(\mathbf{y}). \tag{A.4}$$

Since  $F_{\mathbf{Y},T}^*(\mathbf{y}) \rightarrow F_{\mathbf{Y}}(\mathbf{y})$  occurs for any sample path in a set of sample paths that has probability one, we conclude that  $F_T^*(y) \rightarrow F(y)$  a.s.  $\square$

**Proof of Theorem 3.4.** Consider the process  $\{X_t^+\}$  defined by  $X_t^+ = X_t^* + u_t$  where  $\{u_t\}$  is an i.i.d. sequence of random variables having the uniform distribution  $U[-h_T, h_T]$ .  $\{u_t\}$  is assumed to be independent of  $\{X_t^*\}$  while  $h_T$  is some sequence depending on  $T$  such that  $h_T \rightarrow 0$  as  $T \rightarrow \infty$  at some rate to be specified later. Denote by  $\rho_T^+$

the  $\rho$ -mixing coefficient of  $\{X_t^+\}$ . By a theorem of Csaki and Fisher, cf. Theorem 3.1 in Bradley (1986), we have that  $\rho_T^* = \rho_T^+$  and therefore the assertion of the Theorem is established if we show the required  $\rho$ -mixing property for the sequence  $\{X_t^+\}$ . In fact more is true, i.e., it can be shown that  $\{X_t^+\}$  is uniformly ergodic from which the required exponential decrease of the  $\rho^+$ -mixing coefficient follows by Theorem 1 of Doukhan (1994, p. 88). To see the uniform ergodicity of  $\{X_t^+\}$  note first that the one-step transition distribution function of  $\{X_t^+\}$ , denoted by  $F_T^+(y|\mathbf{x})$ , is given by  $F_T^+(y|\mathbf{x}) = \int F_{u,T}(y-z) dF_T^*(z|\mathbf{x})$  with  $F_{u,T}(\cdot)$  the distribution function of  $u_t$ . The one-step transition density of  $\{X_t^+\}$  is given by

$$\begin{aligned} f_T^+(y|\mathbf{x}) &= \frac{1}{2h_T} \sum_{j=p+1}^T 1_{(y-h_T, y+h_T)}(X_j^*) P(X_{t+1}^* = X_j | \mathbf{Y}_t^* = \mathbf{x}) \\ &= \frac{1}{2h_T} [F_T^*(y+h_T|\mathbf{x}) - F_T^*(y-h_T|\mathbf{x})]. \end{aligned}$$

Thus we have

$$\begin{aligned} &\sup_{y \in \mathbf{R}} \sup_{\mathbf{x} \in \mathbf{S}} |f^+(y|\mathbf{x}) - f(y|\mathbf{x})| \\ &\leq \frac{1}{2h_T} \sup_{y \in \mathbf{R}} \sup_{\mathbf{x} \in \mathbf{S}} |F_T^*(y+h_T|\mathbf{x}) - F(y+h_T|\mathbf{x})| \\ &\quad + \frac{1}{2h_T} \sup_{y \in \mathbf{R}} \sup_{\mathbf{x} \in \mathbf{S}} |F_T^*(y-h_T|\mathbf{x}) - F(y-h_T|\mathbf{x})| \\ &\quad + \sup_{y \in \mathbf{R}} \sup_{\mathbf{x} \in \mathbf{S}} \left| \frac{1}{2h_T} \{F(y+h_T|\mathbf{x}) - F(y-h_T|\mathbf{x})\} - f(y|\mathbf{x}) \right| \\ &= S_1 + S_2 + S_3. \end{aligned}$$

Now, let  $h_T \sim (\log(T))^{-1}$  and verify that by the same arguments as in the proof of Lemma 6.1 and Theorem 3.2,  $S_1 \rightarrow 0$  and  $S_2 \rightarrow 0$  a.s. as  $T \rightarrow \infty$  and  $b \rightarrow 0$  such that (B2) is satisfied. Furthermore, for  $S_3$  we get by the mean value theorem for some  $y < y_1 < y + h_T$  and  $y - h_T < y_2 < y$  that

$$\begin{aligned} S_3 &= \sup_y \sup_{\mathbf{x} \in \mathbf{S}} \left| \frac{1}{2} \{f(y_1|\mathbf{x}) + f(y_2|\mathbf{x})\} - f(y|\mathbf{x}) \right| \\ &\leq \frac{1}{2} \sup_y \sup_{\mathbf{x} \in \mathbf{S}} |f(y_1|\mathbf{x}) - f(y|\mathbf{x})| + \frac{1}{2} \sup_y \sup_{\mathbf{x} \in \mathbf{S}} |f(y_2|\mathbf{x}) - f(y|\mathbf{x})| \\ &= O(h_T), \end{aligned}$$

where the last inequality follows using the Lipschitz continuity in (A.2)(iii).

To conclude the proof of the uniform ergodicity of  $\{X_t^+\}$  we use Theorem 16.2.4 of Meyn and Tweedie (1993) and show that a nonnegative measure  $\mu$  exists such that for  $\mathbf{Y}_t^+ = (X_t^+, X_{t-1}^+, \dots, X_{t-p}^+)$ ,  $P(\mathbf{Y}_{t+p}^+ \in A | \mathbf{Y}_t^+ = \mathbf{x}) \geq \mu(A)$  for all  $\mathbf{x} \in \mathbf{S}$ . We

demonstrate this for the case  $p = 1$  the general case being handled in exactly the same manner. Thus let  $p = 1$  and note that by (A.3),  $\varepsilon > 0$  exists such that  $f(\cdot|x) \geq \varepsilon > 0$  for all  $x \in S$ . Furthermore,  $T_0 \in \mathbf{N}$  exists such that  $\sup_y \sup_x |f_T^+(y|x) - f(y|x)| < \varepsilon/2$  for all  $T \geq T_0$ . Hence,

$$\begin{aligned} P(X_{t+1}^+ \in A | X_t^+ = x) &\geq \int_A f(z|x) dz - \int_A |f_T^+(z|x) - f(z|x)| dz \\ &\geq \varepsilon \lambda(A) - \frac{\varepsilon}{2} \lambda(A) \\ &= \frac{\varepsilon}{2} \lambda(A), \end{aligned}$$

where  $\lambda$  denotes Lebesgue measure.  $\square$

**Proof of Theorem 4.1.** By the Markov property, Theorems 3.2 and 3.3 we have that for every fixed  $k \in \mathbf{N}$  and in a set of sample paths of  $\{X_t, t \geq 1\}$  that has probability one

$$F_{X_t^*, X_{t+1}^*, \dots, X_{t+k+m-1}^*}^* \Rightarrow F_{X_t, X_{t+1}, \dots, X_{t+k+m-1}} \tag{A.5}$$

in the weak sense as  $T \rightarrow \infty$ . Here  $F_{X_t^*, X_{t+1}^*, \dots, X_{t+k+m-1}^*}^*$  denotes the joint distribution function of  $X_t^*, X_{t+1}^*, \dots, X_{t+k+m-1}^*$  while  $F_{X_t, X_{t+1}, \dots, X_{t+k+m-1}}$  the joint distribution function of  $X_t, X_{t+1}, \dots, X_{t+k+m-1}$ . From this and by the continuity of  $g$  and the definition of  $Z_t$ , (resp.  $Z_t^*$ ), we have with probability one,

$$F_{Z_t^*, Z_{t+1}^*, \dots, Z_{t+k}^*}^* \Rightarrow F_{Z_t, Z_{t+1}, \dots, Z_{t+k}}. \tag{A.6}$$

Furthermore, recall that by (A3)  $X_t^*$  takes its values in the compact set  $S$ . This fact, together with the continuity of  $g$ , implies that the sequences  $|Z_t^*|^2$  and  $|Z_t^* Z_{t+h}^*|$  are uniformly integrable. Therefore, using (A.6) we get for every integer  $h, 1 \leq h \leq k$  that

$$E^* |Z_t^*|^2 \rightarrow E |Z_t|^2 \quad \text{a.s.} \tag{A.7}$$

and

$$E^* (Z_t^* Z_{t+h}^*) \rightarrow E (Z_t Z_{t+h}) \quad \text{a.s.} \tag{A.8}$$

Now, to prove the first part of the theorem, note that for  $\varepsilon > 0$  arbitrary,  $k_1 = k_1(\varepsilon) \in \mathbf{N}$  exists such that

$$2 \sum_{\tau=k_1+1}^{\infty} |\text{Cov}(Z_1, Z_{1+\tau})| < \varepsilon/3. \tag{A.9}$$

Furthermore, for  $\varrho \in (0, 1)$  as in Theorem 3.4,  $k_2 = k_2(\varepsilon) \in \mathbf{N}$  exists such that

$$\frac{2\varrho^{k_2+1}}{1-\varrho} < \frac{\varepsilon}{6 \text{Var}(Z_t)}. \tag{A.10}$$

Let  $k = \max\{k_1, k_2\}$ . We then have

$$\begin{aligned}
 |\sigma_{g,T}^{*2} - \sigma_g^2| &\leq \left| \text{Var}^*(Z_1^*) - \text{Var}(Z_1) + 2 \sum_{\tau=1}^k (\text{Cov}^*(Z_1^*, Z_{1+\tau}^*) - \text{Cov}(Z_1, Z_{1+\tau})) \right| \\
 &\quad + 2 \sum_{\tau=k+1}^{T-1} |\text{Cov}^*(Z_1^*, Z_{1+\tau}^*)| + 2 \sum_{\tau=k+1}^{\infty} |\text{Cov}(Z_1, Z_{1+\tau})|. \tag{A.11}
 \end{aligned}$$

Because of (A.7) and (A.8),  $T_1 \in \mathbf{N}$  exists such that

$$|\text{Var}^*(Z_1^*) - \text{Var}(Z_1)| + 2 \sum_{\tau=1}^k |\text{Cov}^*(Z_1^*, Z_{1+\tau}^*) - \text{Cov}(Z_1, Z_{1+\tau})| < \varepsilon/3. \tag{A.12}$$

Furthermore,  $T_2 \in \mathbf{N}$  exists such that

$$|\text{Var}^*(Z_1^*) - \text{Var}(Z_1)| < \frac{\varepsilon(1 - \varrho)}{12\varrho^{k+1}}. \tag{A.13}$$

Now, let  $T^* = \max\{T_0, T_1, T_2\}$  where  $T_0$  is given in Theorem 3.4. For  $T \geq T^*$  we then have using (A.10) and (A.13) and the inequality for  $\rho$ -mixing sequences given in Theorem 3 of Doukhan (1994, p. 9) that

$$\begin{aligned}
 &2 \sum_{\tau=k+1}^{T-1} |\text{Cov}^*(Z_1^*, Z_{1+\tau}^*)| \\
 &\leq 2 \sum_{\tau=k+1}^{T-1} \varrho^\tau \text{Var}^*(Z_1^*) \\
 &\leq \frac{2\varrho^{k+1}}{1 - \varrho} \text{Var}(Z_1) + \frac{2\varrho^{k+1}}{1 - \varrho} |\text{Var}^*(Z_1^*) - \text{Var}(Z_1)| < \varepsilon/3. \tag{A.14}
 \end{aligned}$$

Thus from (A.11) and using (A.9), (A.12) and (A.11) we conclude that  $|\sigma_{g,T}^{*2} - \sigma_g^2| < \varepsilon$ .

To prove the second part of the theorem note first that by (A3) and (A.6) we have for some  $\kappa > 0$  that  $E^*|Z_t^*|^{2+\kappa} \rightarrow E|Z_t|^{2+\kappa}$  a.s. Now, since  $\rho$ -mixing implies  $\alpha$ -mixing, the assertion follows by Theorems 3.4 and 1.7 of Ibragimov (1962).  $\square$

**Proof of Theorem 4.2.** The proof of weak convergence follows from Theorem 4.1 by means of the Cramer–Wold Device; cf. Brockwell and Davis (1991). Convergence of second moments follows from weak convergence and Theorem (viii-B) of Rao (1973, p. 121).  $\square$

**Proof of Proposition 4.1.** Assertion (1) of the proposition follows by the same arguments as in the proof of the first part of Theorem 4.1. Part (2) is an immediate consequence of the fact that  $0 \leq I_{i,r}^* \leq 1$ , Theorems 3.4 and 1.6 of Ibragimov (1962).  $\square$

**Appendix B. Derivation of Formula (13)**

Since  $f(\mathbf{x})$  is the density of  $N(\mathbf{0}, \Gamma_p)$  and

$$F(y|\mathbf{x}) = \Phi\left(\frac{y - \delta - \sum_{i=1}^p a_i x_i}{\sigma}\right),$$

where  $\Phi(\cdot)$  denotes the distribution function of the standard normal, we get

$$\begin{aligned} \int \text{Var}[F_T^*(y|\mathbf{x})] dy &= \frac{\sigma}{Tb^p f(\mathbf{x})} \int W^2(\mathbf{u}) d\mathbf{u} \int \Phi(z)(1 - \Phi(z)) dz \\ &= \frac{\sigma}{Tb^p f(\mathbf{x})\sqrt{\pi}} \int W^2(\mathbf{u}) d\mathbf{u}. \end{aligned}$$

Furthermore,

$$F_i^{(1)}(y|\mathbf{x}) = -\frac{a_i}{\sigma} \phi\left(\frac{y - \delta - \sum_{i=1}^p a_i x_i}{\sigma}\right)$$

and

$$F_i^{(2)}(y|\mathbf{x}) = \frac{a_i^2}{\sigma^2} \phi'\left(\frac{y - \delta - \sum_{i=1}^p a_i x_i}{\sigma}\right),$$

where  $\phi$  denotes the density of the standard normal and  $\phi'$  its first derivative. Thus by straightforward calculations we get

$$\begin{aligned} &\int [\text{Bias}(F_T^*(y|\mathbf{x}))]^2 dy \\ &= b^4 W_2^2 \int \left\{ \frac{1}{2} \sum_{i=1}^p \left( F_i^{(2)}(y|\mathbf{x}) + 2F_i^{(1)}(y|\mathbf{x}) \frac{f_i^{(1)}(\mathbf{x})}{f(\mathbf{x})} \right) \right\}^2 dy \\ &= b^4 W_2^2 \sum_{i=1}^p \sum_{j=1}^p \left\{ C_1 \frac{a_i a_j f_i^{(1)}(\mathbf{x}) f_j^{(1)}(\mathbf{x})}{\sigma f^2(\mathbf{x})} \right. \\ &\quad \left. - C_2 \frac{a_i a_j^2 f_i^{(1)}(\mathbf{x})}{\sigma^2 f(\mathbf{x})} + C_3 \frac{1}{4\sigma^3} a_i^2 a_j^2 \right\}, \end{aligned}$$

where  $C_1 = \int \phi^2(u) du$ ,  $C_2 = -\int u\phi^2(u) du$  and  $C_3 = \int u^2\phi^2(u) du$ . Now, since  $C_1 = 1/(2\sqrt{\pi})$ ,  $C_2 = 0$  and  $C_3 = 1/(4\sqrt{\pi})$ , we get using

$$\sum_{i=1}^p \sum_{j=1}^p \frac{a_i a_j f_i^{(1)}(\mathbf{x}) f_j^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} = \{\mathbf{a}^\top \Gamma_p^{-1}(\mathbf{x} - \boldsymbol{\mu})\}^2,$$

where  $\mathbf{a} = (a_1, a_2, \dots, a_p)^\top$  and  $\boldsymbol{\mu} = (\mu, \mu, \dots, \mu)^\top$  that

$$\int [\text{Bias}(F_T^*(y|\mathbf{x}))]^2 dy = b^4 W_2^2 \left\{ \frac{1}{2\sqrt{\pi}\sigma} (\mathbf{a}^\top \Gamma_p^{-1}(\mathbf{x} - \boldsymbol{\mu}))^2 + \frac{1}{16\sqrt{\pi}\sigma^3} (\mathbf{a}^\top \mathbf{a})^2 \right\}.$$

## References

- Ango Nze, P., 1992. Critères d'ergodicité de quelques modèles à représentation markovienne. *C. R. Acad. Sci. Paris* 315 (1), 1301–1304.
- Auestad, B., Tjøstheim, D., 1990. Identification of nonlinear time series: first order characterization and order determination. *Biometrika* 77, 669–687.
- Auestad, B., Tjøstheim, D., 1991. Functional identification in nonlinear time series. In: Roussas, G. (Ed.), *Nonparametric Functional Estimation and Related Topics*. Kluwer Academic Publishers, Dordrecht.
- Billingsley, P., 1995. *Probability and Measure*. Wiley, New York.
- Bradley, R.C., 1986. Basic properties of strong mixing conditions. In: Eberlein, E., Taqqu, M.S. (Eds.), *Dependence in Probability and Statistics*. Birkhauser, Boston, MA, pp. 165–192.
- Brockwell, P.J., Davis, R.A., 1991. *Time Series: Theory and Methods*, 2nd Edition. Springer, Berlin.
- Chan, K.S., Tong, H., Stenseth, N.C., 1997. Analyzing short time series data from periodically fluctuating rodent populations by threshold models: a nearest block bootstrap approach. Technical Report, No. 258, Department of Statistics and Actuarial Science, University of Iowa.
- Chen, R., Tsay, R.S., 1993. Nonlinear additive ARX models. *J. Amer. Statist. Assoc.* 88, 955–967.
- Doob, J.L., 1953. *Stochastic Processes*. Wiley, New York.
- Doukhan, P., 1994. *Mixing. Properties and Examples*. Lecture Notes in Statistics, Vol. 85. Springer, New York.
- Efron, B., 1979. Bootstrap methods: another look at the jackknife. *Ann. Statist.* 7, 1–26.
- Efron, B., Tibshirani, R., 1986. Bootstrap measures for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.* 1, 54–77.
- Efron, B., Tibshirani, R., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Falk, M., Reiss, R.-D., 1992. Bootstrapping conditional curves. In: Jöckel, K.H., Rothe, G., Sandler, W. (Eds.), *Bootstrapping and Related Techniques*, Lecture Notes in Economics and Mathematical Systems, Vol. 376. Springer, New York.
- Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and its Applications*. Chapman & Hall, New York.
- Franke, J., Härdle, W., 1992. On bootstrapping kernel spectral estimates. *Ann. Statist.* 20, 121–145.
- Franke, J., Kreiss, J.-P., Mammen, E., 1996. Bootstrap of kernel smoothing in nonlinear time series, preprint.
- Freedman, D.A., 1984. On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.* 12, 827–842.
- Härdle, W., Vieu, P., 1992. Kernel regression smoothing of time series. *J. Time Series Anal.* 13, 209–232.
- Ibragimov, I.A., 1962. Some limit theorems for stationary processes. *Theory Probab. and Appl.* 7, 349–382.
- Kim, T.Y., Cox, D., 1996. Bandwidth selection in kernel smoothing of time series. *J. Time Series Anal.* 17, 49–63.
- Künsch, H., 1989. The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* 17, 1217–1241.
- Lall, U., Sharman, A., 1996. A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Res.* 32, 679–693.
- Lawrence, A.J., 1991. Directionality and reversibility in time series. *Int. Statist. Rev.* 59, 67–79.
- Liu, R.Y., Singh, K., 1992. Moving blocks jackknife and bootstrap capture weak dependence. In: R. LePage, L. Billard (Eds.), *Exploring the Limits of the Bootstrap*. Wiley, New York, pp. 225–248.
- Meyn, S.P., Tweedie, R.L., 1993. *Markov Chains and Stochastic Stability*. Springer, London.
- Politis, D.N., Romano, J.P., 1992. A general resampling scheme for triangular arrays of  $\alpha$ -mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.* 20, 1985–2007.
- Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. *J. Amer. Statist. Assoc.* 89, 1303–1313.
- Rajarshi, M.B., 1990. Bootstrap in Markov-sequences based on estimates of transition density. *Ann. Institut. Statist. Math.* 42, 253–268.
- Rao, C.R., 1973. *Linear Statistical Inference and its Applications*, 2nd Edition. Wiley, New York.
- Roussas, G.G., 1969. Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Statist.* 40, 1386–1400.
- Roussas, G.G., 1988. Nonparametric estimation in mixing sequences of random variables. *J. Statist. Plann. Infer.* 18, 135–149.
- Shao, J., Tu, D., 1995. *The Jackknife and the Bootstrap*. Springer, New York.

- Shi, S.G., 1991. Local Bootstrap. *Ann. Institut. Statist. Math.* 43, 667–676.
- Tong, H., 1990. *Nonlinear Time Series: A Dynamical Approach*. Oxford University Press, New York.
- Tjøstheim, D., 1990. Nonlinear time series and Markov chains. *Adv. in Appl. Probab.* 22, 587–611.
- Tjøstheim, D., 1994. Nonlinear time series: a selective review. *Scandinavian J. Statist.* 21, 97–130.
- Tjøstheim, D., Auestad, B., 1994. Nonparametric identification of nonlinear time series: selecting significant lags. *J. Amer. Statist. Assoc.* 89, 1410–1419.
- Tsay, R., 1992. Model checking via parametric bootstraps in time series analysis. *Appl. Statist.* 41, 1–15.
- Weiss, G., 1975. Time reversibility of linear stochastic processes. *J. Appl. Probab.* 12, 831–836.