Ann. Inst. Statist. Math.
Vol. 52, No. 1, 139–159 (2000)

THE LOCAL BOOTSTRAP FOR KERNEL ESTIMATORS UNDER GENERAL DEPENDENCE CONDITIONS*

EFSTATHIOS PAPARODITIS[1] AND DIMITRIS N. POLITIS[2]

[1] *Department of Mathematics and Statistics, University of Cyprus, P.O. Box 537, CY-1678 Nicosia, Cyprus*
[2] *Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, Dept 0112, La Jolla, CA 92093-0112, U.S.A.*

(Received October 6, 1997; revised September 14, 1998)

Abstract. We consider the problem of estimating the distribution of a nonparametric (kernel) estimator of the conditional expectation $g(\boldsymbol{x};\phi) = E(\phi(X_{t+1}) \mid \boldsymbol{Y}_{t,m} = \boldsymbol{x})$ of a strictly stationary stochastic process $\{X_t, t \geq 1\}$. In this notation $\phi(\cdot)$ is a real-valued Borel function and $\boldsymbol{Y}_{t,m}$ a segment of lagged values, i.e., $\boldsymbol{Y}_{t,m} = (X_{t-i_1}, X_{t-i_2}, \dots, X_{t-i_m})$, where the integers i_j satisfy $0 \leq i_1 < i_2 < \cdots < i_m < \infty$. We show that under a fairly weak set of conditions on $\{X_t, t \geq 1\}$, an appropriately designed and simple bootstrap procedure that correctly imitates the conditional distribution of X_{t+1} given the selective past $\boldsymbol{Y}_{t,m}$, approximates correctly the distribution of the class of nonparametric estimators considered. The procedure proposed is entirely nonparametric, its main dependence assumption refers to a strongly mixing process with a polynomial decrease of the mixing rate and it is not based on any particular assumptions on the model structure generating the observations.

Key words and phrases: Resampling, confidence intervals, dependence, nonparametric estimators.

1. Introduction

Let $\{X_t, t \geq 1\}$ be a strictly stationary, real valued random process and let ϕ be a real-valued $\mathcal{B}$-measurable function, where $\mathcal{B}$ denotes the Borel σ-algebra over $\boldsymbol{R}$. For $m \in \boldsymbol{N}$, let $\boldsymbol{x} = (x_{i_1}, x_{i_2}, \dots, x_{i_m}) \in \boldsymbol{R}^m$ and assume that the conditional expectation

$$g(\boldsymbol{x};\phi) = E(\phi(X_{t+1}) \mid \boldsymbol{Y}_{t,m} = \boldsymbol{x}) \tag{1.1}$$

exists, where $\boldsymbol{Y}_{t,m} = (X_{t-i_1}, X_{t-i_2}, \dots, X_{t-i_m})$ and $0 \leq i_1 < i_2 < \cdots < i_m < \infty$ are integers. To give some examples, recall that $g(\boldsymbol{x};\phi_1)$ is the conditional mean and that $g(\boldsymbol{x};\phi_2) - g^2(\boldsymbol{x};\phi_1)$ is the conditional variance of $\{X_t\}$, where $\phi_1(z) = z$ and $\phi_2(z) = z^2$, while for $\phi(z) = 1_{(-\infty,y]}(z)$ the conditional distribution function $P(X_{t+1} \leq y \mid \boldsymbol{Y}_{t,m} = \boldsymbol{x})$ appears.

Given an observed stretch $X_1, X_2, \dots, X_T$ of the process, a nonparametric (kernel) estimator of $g(\boldsymbol{x};\phi)$ is given by

$$\hat{g}_h(\boldsymbol{x};\phi) = \frac{\sum_{j=i_m+1}^{T-1} \phi(X_{j+1}) K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m})}{\sum_{s=i_m+1}^{T-1} K_h(\boldsymbol{x} - \boldsymbol{Y}_{s,m})} \tag{1.2}$$

* Research supported by National Science Foundation Grant DMS-97-03964.

where $K_h(\cdot) = h^{-m}K(\cdot/h)$ and the kernel $K(\cdot)$ is a density function on $\boldsymbol{R}^m$. Nonparametric estimation of $g(\boldsymbol{x};\phi)$ for particular choices of ϕ has received considerable interest in the literature where such estimators have been also found useful in detecting and modeling nonlinearity in time series analysis; cf. among others Tong (1990), Auestad and Tjøstheim (1990) and Tjøstheim (1994). It is known, (see for instance Rosenblatt (1991)), that under some appropriate conditions on the process $\{X_t, t \geq 1\}$ which include some dependence assumptions, some conditions on the kernel K and the bandwidth h, the statistic $\sqrt{Th^m}(\hat{g}_h(\boldsymbol{x};\phi) - E[\hat{g}_h(\boldsymbol{x};\phi)])$ converges weakly towards a normal distribution the variance of which is given by

$$\sigma^2(\boldsymbol{x};\phi) = \frac{1}{f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})}\operatorname{Var}\{\phi(X_{t+1}) \mid \boldsymbol{Y}_{t,m} = \boldsymbol{x}\}\int K^2(\boldsymbol{u})d\boldsymbol{u}. \tag{1.3}$$

Here, $f_{\boldsymbol{Y}_{t,m}}(\cdot)$ stands for the stationary density of $\boldsymbol{Y}_{t,m}$. Furthermore, for $h^{(m+4)}T \to 0$ the bias $\sqrt{Th^m}E(\hat{g}_h(\boldsymbol{x};\phi) - g(\boldsymbol{x};\phi))$ is negligible while for $h^{(m+4)/2}T^{1/2} \to C_h > 0$ it is nonvanishing and (under appropriate differentiability conditions) converges weakly to

$$B(\boldsymbol{x};\phi) = \frac{1}{2}C_h\frac{1}{f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})}\sum_{k=1}^{m}K_2\Big\{\int\phi(y)f^{(x_{i_k},x_{i_k})}_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{x})dy - g(\boldsymbol{x};\phi)f^{(x_{i_k},x_{i_k})}_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})\Big\}; \tag{1.4}$$

cf. Auestad and Tjøstheim (1990) and Masry and Tjøstheim (1995). In the above notation $f^{(x_{i_k},x_{i_k})}_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x}) = \partial^2 f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})/\partial x_{i_k}\partial x_{i_k}$, $f^{(x_{i_k},x_{i_k})}_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{x}) = \partial^2 f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{x})/\partial x_{i_k}\partial x_{i_k}$ where $f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{x})$ denotes the (joint) density of $(X_{t+1},\boldsymbol{Y}_{t,m})$ and $K_2 = \int u_k^2 K(\boldsymbol{u})d\boldsymbol{u}$. The fact that the mean and the variance of the limiting Gaussian distribution depend on unknown (and difficult to estimate) characteristics of the process, leads to some inherent difficulties in using the above asymptotic normal approximation for the practical construction of confidence intervals for $g(\boldsymbol{x};\phi)$. Bootstrap methods offer, therefore, a potentially useful alternative.

The aim of this paper is to show that in the time series context, bootstrapping kernel estimators can be done in a simple, effective and model free way. For this we propose a bootstrap procedure that generates replicates $\{(X^*_{j+1},\boldsymbol{Y}_{j,m}); j = i_m + 1, i_m + 2, \ldots, T-1\}$ of the observed pairs $\{(X_{j+1},\boldsymbol{Y}_{j,m}); j = i_m + 1, i_m + 2, \ldots, T-1\}$ by correctly imitating the (unknown) conditional distribution function $F_{X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{x}) = P(X_{t+1} \leq \cdot \mid \boldsymbol{Y}_{t,m} = \boldsymbol{x})$, i.e., the distribution of X_{t+1} given the *selective* past $\boldsymbol{Y}_{t,m} = (X_{t-i_1}, X_{t-i_2}, \ldots, X_{t-i_m})$. Note that in this approach for each pair $(X_{j+1},\boldsymbol{Y}_{j,m})$ only the observation X_{j+1} is bootstrapped. We show that this bootstrap procedure which is based on a simple and consistent estimator of the above conditional distribution function and which also automatically mimics the stationary density $f_{\boldsymbol{Y}_{t,m}}(\cdot)$, leads to an asymptotically valid approximation of the distribution of $\hat{g}_h(\boldsymbol{x};\phi)$. We stress here the fact that our procedure is valid under a quite general set of dependence conditions on the process $\{X_t\}$ which includes several of the mixing processes discussed in the literature, and works without any particular (parametric or non parametric) assumptions on the model structure generating the observations. Loosely speaking, our procedure works for every weakly dependent stationary process for which the statistic $\hat{g}_h(\boldsymbol{x};\phi)$ has the asymptotic distributional behavior mentioned above (cf. also assumptions (A5) and (A6) in Section 2). The reason why our procedure 'works' lies in the fact that the sampling distribution of $\hat{g}_h(\boldsymbol{x};\phi)$ depends only on the conditional distribution $F_{X_{t+1}|Y_{t,m}}(\cdot \mid \boldsymbol{x})$

and the stationary (marginal) distribution of $\boldsymbol{Y}_{t,m}$. Therefore, in order for the bootstrap to work, it is not necessary to imitate the whole (and probably very complicated) process structure. A resampling procedure that "imitates" $F_{X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{x})$ and $F_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})$ with sufficient accuracy will have good chances of reproducing accurately the law of $\hat{g}_h(\boldsymbol{x};\phi)$. The bootstrap procedure discussed in this paper achieves this goal.

Bootstrapping time series data has received considerable interest in the last fifteen 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. We mention here the residual bootstrap (cf. Freedman (1984)), the block bootstrap (cf. Künsch (1989), Liu and Singh (1992)), the stationary bootstrap (cf. Politis and Romano (1994)) and the frequency domain bootstrap (cf. Franke and Härdle (1992)). For an overview of these different methods see Efron and Tibshirani (1993) and Shao and Tu (1995).

Nevertheless, the specific problem of approximating the distribution of nonparametric conditional moment estimators in a time series context has only very recently attracted the attention of bootstrap researchers. Franke *et al.* (1996) proposed an approach which is based on certain autoregressive model assumptions for the process considered. The bootstrap procedure proposed in the present paper does not rely on model assumptions for the original process, nor does it attempt to mimic the whole dependence structure of the process. The idea that the reproduction of the whole dependence structure is not necessary for the bootstrap to work in nonparametric estimation problems is used also in the (wild) bootstrap procedure proposed by Neumann and Kreiss (1998). However, their approach deals only with the autoregressive function $E(X_{t+1} \mid X_t = x)$ and, more importantly, they impose a Markovian assumption on the data generating process in order to generate the bootstrap replicates. No such model assumptions are needed here. Furthermore, in both approaches mentioned above, preliminary (nonparametric) estimators of certain conditional curves are required in order to generate the bootstrap replicates. Apart from avoiding any explicit estimation of nonparametric characteristics of the original process in order to carry out the bootstrap procedure, our approach is also more general as it is applicable to a larger class of stochastic process, to a greater class of estimation problems and it is totally model free. The latter aspect is of particular importance in nonlinear time series analysis since one of the frequent application of nonparametric estimators of conditional moments is for model identification purposes; cf. Tong (1990) and Tjøstheim (1994). Under a more restrictive set of dependence conditions, than those imposed here, Neumann (1997) shows validity of a nearest neighbor Markovian resampling procedure for supremum type statistics in a nonparametric estimation context.

Our bootstrap procedure is related to the local bootstrap proposal by Shi (1991). Like Shi's approach our bootstrap replicates X^*_{j+1} are obtained using a local version of the conditional distribution of X_{j+1} given the observed segment $\boldsymbol{Y}_{j,m}$. However, the context here is much more general in that we deal with time dependent data and $\boldsymbol{Y}_{j,m}$ is a segment of lagged observations of the same stochastic process. Furthermore, we do not assume any particular (e.g., regression type) model structure relating the set of random variables X_{j+1} and $\boldsymbol{Y}_{j,m}$. Note that bootstrap procedures related to that of Shi (1991) have been also proposed by Falk and Reiss (1992) and Cao-Abad and Gonzalez-Manteiga (1993).

The paper is organized as follows. Section 2 states the assumptions imposed on the process $\{X_t, t \geq 1\}$, discusses in detail the bootstrap procedure proposed to approximate the distribution of $\hat{g}_h(\boldsymbol{x};\phi)$, and establishes its asymptotic validity. To illustrate the procedure a real data example is discussed in this section too. A summary is given in

Section 3 while Section 4 contains all technical lemmas and proofs.

2. Assumptions and bootstrap approximations

The stochastic process $\{X_t, t \geq 1\}$ considered satisfies the following set of assumptions.

(A1) $\{X_t; t \geq 1\}$ is a strictly stationary process and ϕ is a Borel function on $\boldsymbol{R}$ such that $E|\phi(X_t)|^s < \infty$ for some $s > 8$.

(A2) i) The densities $f_{\boldsymbol{Y}_{t,m}}(\cdot)$ and $f_{\boldsymbol{Y}_{t+l,m}\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{x})$, $l \geq 1$ exist and are absolutely continuous with respect to Lebegue measure. (Note that here and the sequel if $l \leq i_m - i_1$ then $f_{\boldsymbol{Y}_{t+l,m}\boldsymbol{Y}_{t,m}}(\cdot)$ is the joint density of the distinct random variables in the set $\{\boldsymbol{Y}_{t+l,m}, \boldsymbol{Y}_{t,m}\}$.)

ii) $f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{z}) > 0$ for $\boldsymbol{z} \in \boldsymbol{D}_\eta(\boldsymbol{x})$ where $\boldsymbol{D}_\eta(\boldsymbol{x}) = \{\boldsymbol{z} : \|\boldsymbol{z} - \boldsymbol{x}\| < \eta\}$ for some $\eta > 0$ and $\|\cdot\|$ is the Euclidean norm. Furthermore, $f_{\boldsymbol{Y}_{t,m}}(\cdot)$ is assumed to be Lipschitz continuous.

iii) $f_{\boldsymbol{Y}_{t,m}|X_{t+1}}(\cdot \mid y) \leq C_1 < \infty$ and $f_{\boldsymbol{Y}_{t,m}|X_{t+1}}(\cdot \mid y)$ is Lipschitz continuous, i.e., $|f_{\boldsymbol{Y}_{t,m}|X_{t+1}}(\boldsymbol{x}_1 \mid y) - f_{\boldsymbol{Y}_{t,m}|X_{t+1}}(\boldsymbol{x}_2 \mid y)| \leq C\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$ for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{R}^m$.

iv) $f_{\boldsymbol{Y}_{t,m}\boldsymbol{Y}_{t+l,m}|X_{t+1}X_{t+l+1}}(\cdot \mid y_1, y_2) \leq C_2 < \infty$ for all $l \geq 1$.

The following assumptions are imposed on the smoothing bandwidth h and the smoothing kernel K in (1.2).

(A3) The bandwidth h satisfies $h \to 0$ and $Th^m \to \infty$ as $T \to \infty$.

(A4) K is a symmetric and square integrable density on $\boldsymbol{R}^m$ satisfying $|K(\boldsymbol{u}) - K(\boldsymbol{v})| \leq C\|\boldsymbol{u} - \boldsymbol{v}\|$ for all $\boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{R}^m$. Furthermore, for $i, j \in \{1, 2, \ldots, m\}$

$$\int u_i^s K(\boldsymbol{u}) d\boldsymbol{u} = 0 \quad \text{for } s \text{ odd} \quad \text{and} \quad \int u_i u_j K(\boldsymbol{u}) d\boldsymbol{u} = \delta_{i,j} K_2.$$

Here $\delta_{i,j}$ is Kronecker's δ and $0 < K_2 < \infty$ for all $i \in \{1, 2, \ldots, m\}$. Finally, K is a product kernel the support of which is the unit ball in $\boldsymbol{R}^m$.

To state our main assumption on the dependence properties of $\{X_t\}$, we first fix some additional notation. A strictly stationary process $\{X_t, t \geq 1\}$ is said to be strongly mixing if $\alpha(n) \to 0$ as $n \to \infty$ where

$$\alpha(n) = \sup_{A \in \mathcal{M}_1^t, B \in \mathcal{M}_{t+n}^\infty} |P(A \cap B) - P(A)P(B)|, \qquad n > 0,$$

cf. Rosenblatt (1956). Here $\mathcal{M}_1^t$ and $\mathcal{M}_{t+n}^\infty$ denote the σ-fields generated by the random variables $\{X_1, X_2, \ldots, X_t\}$ and $\{X_{t+n}, X_{t+n+1}, \ldots,\}$ respectively. We assume that

(A5) $\{X_t, t \geq 1\}$ is strongly mixing and the mixing coefficient $\alpha(\cdot)$ satisfies

$$\sum_{j=N}^\infty \alpha(j)^{1-2/s} = O(N^{-1}) \quad \text{as} \quad N \to \infty$$

for s such that the moment condition (A1) is satisfied.

Note that (A5) implies a polynomial decrease of the mixing coefficient which is satisfied for a huge class of stochastic processes commonly discussed in the literature. Robinson (1983) and Masry and Tjøstheim (1995) employ this kind of weak dependence assumption in order to investigate the asymptotic properties of kernel estimators for time series. We mention here that several of the commonly used parametric and nonparametric linear and nonlinear time series models belong to the class of strictly stationary and geometrically ergodic Markov processes, e.g., they are geometrically absolutely regular (i.e., β-mixing), a condition that is stronger than geometrically α-mixing, i.e., $\alpha(n) \leq C\rho^n$ for constants $C > 0$ and $\rho \in (0,1)$; see Doukhan (1994) for the different mixing concepts. For a more specific discussion of particular time series models that belong to the Markov class see among others Tong (1990), Tjøstheim (1990), Meyn and Tweedie (1993) and Doukhan (1994). However, we stress here the fact that, the class of processes for which (A5) is satisfied is not restricted to the Markov class.

Now, using (A1)–(A5) it can be shown by well-known arguments that, as $T \to \infty$

$$(2.1)\quad \frac{1}{T}\sum_{\substack{s,j=i_m+1\\ s\neq j}}^{T-1} |\,\mathrm{Cov}[h^{m/2}\phi(X_{s+1})K_h(\boldsymbol{x}-\boldsymbol{Y}_{s,m}), h^{m/2}\phi(X_{j+1})K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m})]| \to 0$$

and

$$\frac{1}{T}\sum_{\substack{s,j=i_m+1\\ s\neq j}}^{T-1} |\,\mathrm{Cov}[h^{m/2}K_h(\boldsymbol{x}-\boldsymbol{Y}_{s,m}), h^{m/2}K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m})]| \to 0;$$

cf. for instance Masry and Tjøstheim (1995), Lemmas 4.3 and 4.5. Under the same set of assumptions this implies that $\hat{g}_h(\boldsymbol{x};\phi) \to g(\boldsymbol{x};\phi)$ in probability. Furthermore, by straightforward calculations we get that

$$\mathrm{Var}(\hat{f}_{h,\boldsymbol{Y}_{t,m}}(\boldsymbol{x})) \to f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})\int K^2(\boldsymbol{u})d\boldsymbol{u}$$

and

$$\mathrm{Var}(\hat{g}_h(\boldsymbol{x};\phi)\hat{f}_h(\boldsymbol{x})) \to f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})\int K^2(\boldsymbol{u})d\boldsymbol{u}E[\phi^2(X_{t+1}) \mid \boldsymbol{Y}_{t,m}=\boldsymbol{x}]$$

as $T \to \infty$ where $\hat{f}_{h,\boldsymbol{Y}_{t,m}}(\boldsymbol{x}) = T^{-1}\sum_{j=i_m+1}^{T-1} K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m})$ is a kernel estimator of $f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})$.

As a careful examination of the proofs shows, the assumption (A5) is imposed in order to ensure that relations like (2.1) are fulfilled. In other words it will be possible to replace the assumption (A5) by some other set of conditions on the dependence structure of $\{X_t\}$ which ensure that relations similar to (2.1) hold. The strong mixing condition (A5) is a technically convenient way to achieve this.

We next state an assumption dealing with the asymptotic distribution of the kernel estimator (1.2). In this assumption '$\Rightarrow$' denotes weak convergence.

(A6) $\sqrt{Th^m}(\hat{g}_h(\boldsymbol{x};\phi) - E(\hat{g}_h(\boldsymbol{x};\phi))) \Rightarrow N(0,\sigma^2(\boldsymbol{x};\phi))$ as $T \to \infty$ where $\sigma^2(\boldsymbol{x};\phi)$ is given in (1.3).

Asymptotic normality of $\sqrt{Th^m}(\hat{g}_h(\boldsymbol{x};\phi) - E[\hat{g}_h(\boldsymbol{x};\phi)])$ has been established for (strong) mixing processes $\{X_t\}$ which satisfy (A5) and under some additional smoothness assumptions by Robinson (1983). Under a different set of conditions which however, also include a strong mixing assumption with a polynomial decrease of the mixing rate, the same result has been established by Masry and Tjøstheim (1995). See Rosenblatt (1991) for a general discussion.

To handle the bias term we finally introduce the following assumption.

(A7) i) The kernel $K(\boldsymbol{x})$ is two times continuously differentiable with respect to $\boldsymbol{x} \in \boldsymbol{R}^m$.

ii) $f_{X_{t+1}\boldsymbol{Y}_{t,m}}(\boldsymbol{x})$ is two times differentiable with respect to $\boldsymbol{x}$ and the functions $\int \psi(y) f^{(x_{i_k})}_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{x})dy$ and $\int \psi(y) f^{(x_{i_k},x_{i_s})}_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{x})dy$ are Lipschitz continuous for $\psi = 1$ and $\psi = \phi$.

The local bootstrap procedure proposed in this paper is based on the simple idea to obtain bootstrap replicates of the sequence of observed pairs $\{(X_{j+1}, \boldsymbol{Y}_{j,m}), j = i_m + 1, i_m+2, \ldots, T-1\}$ by resampling the observation X_{j+1} given the segment $\boldsymbol{Y}_{j,m}$ and using a simple and consistent estimator of the conditional distribution function $F_{X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{Y}_{j,m})$. In particular, for $j = i_m + 1, i_m + 2, \ldots, T-1$, denote by $(X^*_{j+1}, \boldsymbol{Y}_{j,m})$ the bootstrap replicate of the pair $(X_{j+1}, \boldsymbol{Y}_{j,m})$ where

$$X^*_{j+1} \sim \tilde{F}_{b,X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{Y}_{j,m}) \tag{2.2}$$

and $\tilde{F}_{b,X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{x})$ is a version of the empirical conditional distribution function given by

$$\tilde{F}_{b,X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{x}) = \frac{\sum_{l=i_m+1}^{T-1} 1_{(-\infty,\cdot]}(X_{l+1}) K_b(\boldsymbol{x} - \boldsymbol{Y}_{l,m})}{\sum_{s=i_m+1}^{T-1} K_b(\boldsymbol{x} - \boldsymbol{Y}_{s,m})} \tag{2.3}$$

where $K_b(\cdot) = b^{-m} K(\cdot/b)$ and $b > 0$ is called the resampling width.

Note that by (2.2) the distribution of X^*_{j+1} varies with the index j and that *conditionally* on the observed sequence $X_1, X_2, \ldots, X_T$, the bootstrap replicates $(X^*_{j+1}, \boldsymbol{Y}_{j,m})$ and $(X^*_{i+1}, \boldsymbol{Y}_{i,m})$ are independent for $i \neq j$. Thus for every index j, X^*_{j+1} is a random variable taking values in the set $\{X_{i_m+2}, X_{i_m+3}, \ldots, X_T\}$ with probability mass function given by

$$P(X^*_{j+1} = X_{s+1} \mid \boldsymbol{Y}_{j,m}) = K_b(\boldsymbol{Y}_{j,m} - \boldsymbol{Y}_{s,m}) / \sum_{l=i_m+1}^{T-1} K_b(\boldsymbol{Y}_{j,m} - \boldsymbol{Y}_{l,m}) \tag{2.4}$$

and $s = i_m+1, i_m+2, \ldots, T-1$. It is easily seen that for every $\boldsymbol{Y}_{j,m}$, $\sum_{s=i_m+1}^{T-1} P(X^*_{t+1} = X_{s+1} \mid \boldsymbol{Y}_{j,m}) = 1$ and that (2.4) provides a practically simple rule for generating the X^*_{j+1}'s. Thus, in contrast to the i.i.d. case, the probability that X^*_{j+1} assumes any of the observed values $X_{i_m+2}, X_{i_m+3}, \ldots, X_T$ is not constant but depends on the closeness of the corresponding preceding segments $\boldsymbol{Y}_{i_m+1,m}, \boldsymbol{Y}_{i_m+2,m}, \ldots, \boldsymbol{Y}_{T-1,m}$ to the segment $\boldsymbol{Y}_{j,m}$. Furthermore, to choose X^*_{j+1} we only consider those X_{s+1} for which $\boldsymbol{Y}_{s,m}$ lies in a "small" neighborhood of $\boldsymbol{Y}_{j,m}$, where the "size" of this neighborhood is controlled by the resampling bandwidth b.

Remark 2.1. An alternative approach to generate the bootstrap replicates will be to use instead of the discrete version (2.3) a smoothed version of $F_{X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{x})$. In particular, let $\hat{f}_{b,X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{x})$ be the kernel estimator of the conditional density $f_{X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{x})$ given by

$$\hat{f}_{b,X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{x}) = \frac{\sum_{j=i_m+1}^{T-1} \kappa_b(\cdot - X_{j+1}) K_b(\boldsymbol{x} - \boldsymbol{Y}_{j,m})}{\sum_{s=i_m+1}^{T-1} K_b(\boldsymbol{x} - \boldsymbol{Y}_{s,m})}. \tag{2.5}$$

Here, $\kappa_b(\cdot) = b^{-1}\kappa(\cdot/b)$ and the kernel $\kappa : \boldsymbol{R} \to \boldsymbol{R}$ satisfies the conditions given in (A4) for $m = 1$. Now, for $j = i_m + 1, i_m + 2, \ldots, T-1$, bootstrap replicates $(X^+_{j+1}, \boldsymbol{Y}_{j,m})$ can be obtained using

$$X^+_{j+1} \sim \hat{f}_{b,X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{Y}_{j,m}). \tag{2.6}$$

We mention here that in order to generate the X^+_{j+1}'s the explicit estimation of the conditional density $\hat{f}_{b,X_{t+1}|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{Y}_{j,m})$ can be avoided because X^+_{j+1} can be also expressed as

$$X^+_{j+1} = X^*_{j+1} + bW^+_{j+1} \tag{2.7}$$

where W^+_{j+1} is an i.i.d. sequence with $W^+_{j+1} \sim \kappa(\cdot)$ and independent from X^*_{j+1}.

Now, using the generated bootstrap sequence $\{(X^*_{j+1}, \boldsymbol{Y}_{j,m}), j = i_m + 1, i_m + 2, \ldots, T-1\}$ the distribution of the statistic $\sqrt{Th^m}(\hat{g}_h(\boldsymbol{x};\phi) - g(\boldsymbol{x};\phi))$ can be approximated by means of the bootstrap statistic $\sqrt{Th^m}(\hat{g}^*_h(\boldsymbol{x};\phi) - g^*(\boldsymbol{x};\phi))$. Here

$$\hat{g}^*_h(\boldsymbol{x};\phi) = \frac{\sum_{j=i_m+1}^{T-1} \phi(X^*_{j+1}) K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m})}{\sum_{s=i_m+1}^{T-1} K_h(\boldsymbol{x} - \boldsymbol{Y}_{s,m})} \tag{2.8}$$

and

$$g^*(\boldsymbol{x};\phi) = E^*[\phi(X^*_{j+1}) \mid \boldsymbol{Y}_{j,m} = \boldsymbol{x}] = \hat{g}_b(\boldsymbol{x};\phi). \tag{2.9}$$

To study the asymptotic behavior of our procedure we let the bandwidth b used to resample the observations in (2.3) get narrower as the sample size increases.

(A8) $b = O(T^{-\delta})$ where $0 < \delta < 1/\{m(m+2)\}$.

The following result establishes the large sample validity of the bootstrap approximation proposed. It uses Kolmogorov's distance, defined by $d_0(\mathcal{P}, \mathcal{Q}) = \sup_{x\in\boldsymbol{R}} |\mathcal{P}(X \le x) - \mathcal{Q}(X \le x)|$, in order to estimate the distance between two probability measures $\mathcal{P}$ and $\mathcal{Q}$.

THEOREM 2.1. *Suppose that assumptions* (A1)–(A6) *hold.*

i) *If* $T^{1/2}h^{(m+4)/2} \to 0$ *and* (A8) *is fulfilled then we have that, as* $T \to \infty$

$$d_0\{\mathcal{L}(\sqrt{Th^m}(\hat{g}_h(\boldsymbol{x};\phi) - g(\boldsymbol{x};\phi))), \mathcal{L}(\sqrt{Th^m}(\hat{g}^*_h(\boldsymbol{x};\phi) - g^*(\boldsymbol{x};\phi)) \mid X_1, \ldots, X_T)\} \to 0$$

in probability.

ii) *If* $T^{1/2}h^{(m+4)/2} \to C_h > 0$, (A7) *holds and* $b = O(T^{-\delta})$ *for* $0 < \delta < 1/\{(m+2)^2\}$ *then the same result as in part* i) *of the Theorem is true.*

Remark 2.2. Note that the above assumption on the asymptotic behavior of the resampling width b compared to the behavior of the smoothing bandwidth h is needed in order for the bootstrap to get a correct approximation of the asymptotic bias term $B(\boldsymbol{x};\phi)$. Such an 'oversmoothing assumption', i.e., the bootstrap conditional mean function $g^*_b(\boldsymbol{x};\phi)$ is somewhat smoother than $\hat{g}_h(\boldsymbol{x};\phi)$, is common in applications of the bootstrap to problems similar to that considered here; see among others Romano (1988), Härdle and Bowman (1988) and Franke and Härdle (1992). An alternative is to restrict the bootstrap in approximating the distribution of the 'bias free' statistic $\hat{g}_h(\boldsymbol{x};\phi) - E[\hat{g}_h(\boldsymbol{x};\phi)]$ and to estimate the bias term explicitly by estimating (nonparametrically) the unknown quantities appearing in (1.4).

Remark 2.3. Using the bootstrap pairs $\{(X^+_{j+1}, \boldsymbol{Y}_{j,m}), j = i_m+1, \ldots, T-1\}$ given in Remark 2.1, the distribution of $\sqrt{Th^m}(\hat{g}_h(\boldsymbol{x};\phi) - g(\boldsymbol{x};\phi))$ can be also approximated by that of $\sqrt{Th^m}(\hat{g}^+_h(\boldsymbol{x};\phi) - g^+(\boldsymbol{x};\phi))$ where

$$\hat{g}^+_h(\boldsymbol{x};\phi) = \frac{\sum_{j=i_m+1}^{T-1} \phi(X^+_{j+1}) K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m})}{\sum_{s=i_m+1}^{T-1} K_h(\boldsymbol{x} - \boldsymbol{Y}_{s,m})} \tag{2.10}$$

and $g^{+}(\boldsymbol{x};\phi) = \int \phi(y)\hat{f}_{g,X_{t+1}|\boldsymbol{Y}_{t,m}}(y \mid \boldsymbol{x})dy$. Relation (2.7) suggests that the (asymptotic) properties of this bootstrap alternative in approximating the distribution of $\sqrt{Th^m}(\hat{g}_h(\boldsymbol{x};\phi) - g(\boldsymbol{x};\phi))$ will be identical to those of $\sqrt{Th^m}(\hat{g}_h^*(\boldsymbol{x};\phi) - g^*(\boldsymbol{x};\phi))$ stated in Theorem 2.1. However, this bootstrap alternative will not be discussed further here.

As already mentioned in the Introduction, any bootstrap procedure that correctly imitates the conditional distribution function $F_{X_{t+1},|\boldsymbol{Y}_{t,m}}(\cdot \mid \boldsymbol{x})$ and the stationary distribution $F_{\boldsymbol{Y}_{t,m}}(\cdot)$, will have good chances in approximating the distribution of the kernel estimator $\hat{g}_h(\boldsymbol{x};\phi)$. Therefore, some other bootstrap alternatives that achieve this goal could be also considered. One possibility in this direction will be to generate a whole bootstrap process, say, $\{X_t^*, t \geq 1\}$, for which the conditional distribution of X_{t+1}^* given $\boldsymbol{Y}_{t,m}^* = (X_{t-i_1}^*, X_{t-i_2}^*, \ldots, X_{t-i_m}^*)^\top$ and the stationary distribution of $\boldsymbol{Y}_{t,m}^*$ mimic correctly that of X_{t+1} given $\boldsymbol{Y}_{t,m}$ and that of $\boldsymbol{Y}_{t,m}$. One way to achieve this is by means of a Markovian bootstrap procedure in which the one step transition distribution function is based on an estimator of $f_{X_{t+1}|\boldsymbol{Y}_{t,m}}$ or of $F_{X_{t+1}|\boldsymbol{Y}_{t,m}}$. For purposes different to that discussed here, such a Markovian bootstrap procedure has been proposed by Rajarshi (1990) and Paparoditis and Politis (1997).

3. Numerical examples

Example 1. Consider a data generating process that follows the model

$$X_t = \sin(X_{t-2}) + \varepsilon_t,$$

where ε_t is an i.i.d. sequence of $N(0,1)$ random variables. For this model, we are interested in estimating the distribution of the conditional mean $g(x) = E(X_{t+1} \mid X_{t-1} = x)$ based on realizations of length $T = 300$. To obtain this estimator, Epanechnikov's kernel with a smoothing bandwidth $h = 0.7$ has been applied. In this simulated example and in order to reduce computations, the smoothing bandwidth h was chosen as the one which was judged to give best visual fit to the theoretical curve $g(x)$. Clearly, a objective rule for selecting h, like one based on a cross-validation criterion, can be also applied to select the smoothing bandwidth h. Such an application is demonstrated in the real data example below. To resample the observations, the same kernel and a resampling width of $b = 1.2$ has been used. Note that according to the theory developed, b should be larger than h in order for the bootstrap to estimate consistently the bias. Table 1 presents some of the results obtained for this particular values of h and b while Table 2 shows some results for two different values of the smoothing bandwidth h and some different values of the resampling width b. Note that the percentage points of the distribution of $\sqrt{Th}(\hat{g}_h(x) - g(x))$ presented in these tables has been approximated using the corresponding percentage points of the bootstrap statistic $\sqrt{Th}(\hat{g}_h^*(x) - g^*(x))$. The results of the presented bootstrap approximations are based on 100 trials where for each simulated trial the distribution of the bootstrap statistic has been evaluated using 1,000 bootstrap replications. Finally, the estimation of the percentage points of the exact distribution is based on 10,000 replications of the model considered. As these tables show, the bootstrap provides a reasonable estimation of the percentage points of the distribution of the kernel estimator considered. Furthermore, the bootstrap approximations did not seem to be very sensitive with respect to the choice of the resampling width b. Nevertheless, further investigations are required in order to provide some objective rule for selecting b in practice.

Table 1. Estimated exact (Exa) and bootstrap (Boo) estimates of some percentage points of the distribution of the kernel estimator $\hat{g}_h(x)$ for different values of x. "Boo" refers to the mean value while "STD" to the standard deviation of the bootstrap estimates over 100 independent replications of the model considered.

		$x =$ −2.4	−1.6	−0.8	0.0	0.8	1.6	2.4
2.5%	Exa	−7.308	−3.198	−2.215	−3.058	−4.285	−5.080	−6.187
	Boo	−7.058	−3.473	−2.529	−2.998	−3.821	−4.688	−6.466
	STD	1.287	0.514	0.387	0.362	0.479	0.606	1.340
5%	Exa	−6.198	−2.594	−1.679	−2.583	−3.747	−4.239	−4.942
	Boo	−5.985	−2.809	−1.988	−2.509	−3.319	−4.016	−5.309
	STD	1.108	0.452	0.345	0.333	0.429	0.538	1.173
10%	Exa	−5.067	−1.857	−1.109	−1.995	−3.103	−3.404	−3.510
	Boo	−4.737	−2.005	−1.375	−1.952	−2.714	−3.277	−4.014
	STD	0.940	0.385	0.308	0.312	0.399	0.475	0.976
90%	Exa	3.648	3.546	3.133	1.959	1.093	1.886	5.125
	Boo	4.009	3.367	2.834	1.935	1.391	1.972	4.595
	STD	0.990	0.594	0.438	0.369	0.325	0.402	0.969
95%	Exa	5.085	4.382	3.715	2.528	1.679	2.627	6.309
	Boo	5.355	4.114	3.437	2.480	1.983	2.713	5.982
	STD	1.153	0.678	0.493	0.389	0.339	0.486	1.147
97.5%	Exa	6.353	5.092	4.203	3.003	2.243	3.236	7.270
	Boo	6.394	4.774	3.947	2.965	2.474	3.334	6.870
	STD	1.316	0.760	0.552	0.424	0.383	0.551	1.308

Example 2. In this example realizations of length $T = 500$ from the random coefficient autoregressive process

$$X_t = \theta_t X_{t-1} + \varepsilon_t,$$

are considered where $\theta_t \sim N(0, 0.9^2)$, $\varepsilon_t \sim N(0,1)$ and $\{\theta_t\}$ independent of $\{\varepsilon_t\}$; cf. Tong (1990), p. 111. Here we are interested in estimating the distribution of a kernel estimator of $s^2(x) = E(X_{t+1}^2 \mid X_t = x) = 1 + 0.81x^2$. Note that in this example $s^2(x)$ is the conditional variance of X_{t+1} given that $X_t = x$. Denote by $\hat{s}^2(x)$ a kernel estimator of $s^2(x)$ using Epanechnikov's kernel and $h = 0.7$. Figure 1 shows the simulated exact density of $\sqrt{Th}(\hat{s}^2(x) - s^2(x))$ for $x = 0$ based on 1000 replications of this model. Also shown in this figure are four bootstrap estimates of this density based on different original time series. The bootstrap approximations presented are based on the same kernel, a resampling width of $b = 1.0$ and 1000 bootstrap replications. All estimated densities shown in this figure have been obtained using the Gaussian smoothing kernel and a bandwidth selection according to Silverman's rule. Finally, plotted in these exhibits are also the corresponding large sample Gaussian approximations of the distribution of $\sqrt{Th}(\hat{s}^2(x) - s^2(x))$ with bias and the variance estimated from the data. As this figure shows, the bootstrap works very satisfactorily and reproduces quite accurately the overall behavior including some nonsymmetric features of the true distribution of the statistic considered.

Table 2. Estimated exact (Exa) and bootstrap (Boo) estimates of some percentage points of the distribution of the kernel estimator $\hat{g}_h(x)$ for some different values of h and b. "Boo" refers to the mean value of the bootstrap estimates over 100 independent replications of the model considered. The numbers in parentheses are the standard deviations of the bootstrap estimates.

			Percentage					
			5%			95%		
			$x=$ −1.0	0.0	1.0	$x=$ −1.0	0.0	1.0
$h=0.5$	Exa		−2.237	−2.402	−3.125	3.247	2.352	2.242
	Boo	$b=1.0$	−2.401	−2.443	−3.044	3.133	2.444	2.378
			(0.349)	(0.284)	(0.394)	(0.448)	(0.362)	(0.375)
		$b=1.2$	−2.427	−2.534	−3.108	3.071	2.554	2.466
			(0.303)	(0.329)	(0.379)	(0.397)	(0.307)	(0.366)
		$b=1.4$	−2.540	−2.558	−3.088	3.106	2.534	2.552
			(0.395)	(0.267)	(0.408)	(0.383)	(0.282)	(0.349)
$h=0.7$	Exa		−1.938	−2.583	−3.883	3.908	2.528	1.720
	Boo	$b=1.0$	−1.851	−2.469	−3.578	3.618	2.458	1.993
			(0.339)	(0.349)	(0.423)	(0.547)	(0.408)	(0.384)
		$b=1.2$	−2.081	−2.509	−3.510	3.458	2.480	2.081
			(0.295)	(0.333)	(0.449)	(0.447)	(0.389)	(0.381)
		$b=1.4$	−2.104	−2.566	−3.456	3.464	2.540	2.120
			(0.386)	(0.309)	(0.448)	(0.446)	(0.307)	(0.350)

A real data example. Theorem 2.1 enables us to use the proposed bootstrap procedure in order to obtain asymptotically valid confidence intervals for $g(\boldsymbol{x};\phi)$. Consider for instance, the estimation of lower and upper bounds of a 90% pointwise confidence interval for the 'lag three' conditional mean $E(X_{t+1} \mid X_{t-2} = x)$, where $X_t = \log_{10}(Z_t)$ and Z_t is the well-known MacKenzie river series of the annual Canadian lynx trappings. The original series consists of $T = 114$ observations. In Tong (1990), p. 11, a kernel estimator of $E(X_{t+1} \mid X_{t-2} = x)$ has been used in order to demonstrate some nonlinearity features of this series. We discuss this example here again and demonstrate how the bootstrap proposal of this paper can be used to obtain confidence intervals for the unknown mean function $E(X_{t+1} \mid X_{t-2} = x)$. For simplicity denote by $\hat{g}_h(x)$ the kernel estimator (1.2) of $E(X_{t+1} \mid X_{t-2} = x)$ obtained using Epanechnikov's kernel with the smoothing bandwidth $h = 0.44$. This bandwidth has been obtained as the minimizer of the function $CV(h)$ over a set of values of h in the interval $(0,3)$, where $CV(h)$ is given by

$$CV(h) = \frac{1}{T-3}\sum_{j=1}^{T-3}\{X_{j+3} - \hat{g}_{h,j}(X_j)\}^2$$

with $\hat{g}_{h,j}(x)$ the leave-one-out estimator of $g(x)$; see Härdle and Vieu (1992) and Kim and Cox (1996) for details on this cross-validation rule. The bootstrap replicates $\{(X^*_{j+1}, X_{j-2}), j = 3,4,\ldots,T-1\}$ have been generated using in (2.3) the same kernel and the bandwidth $b \approx 2\times h \approx 0.9$. Recall that in order for the bootstrap to capture correctly the bias of this nonparametric estimator, the resampling width g should be somewhat larger than the smoothing bandwidth h. The pointwise confidence intervals has

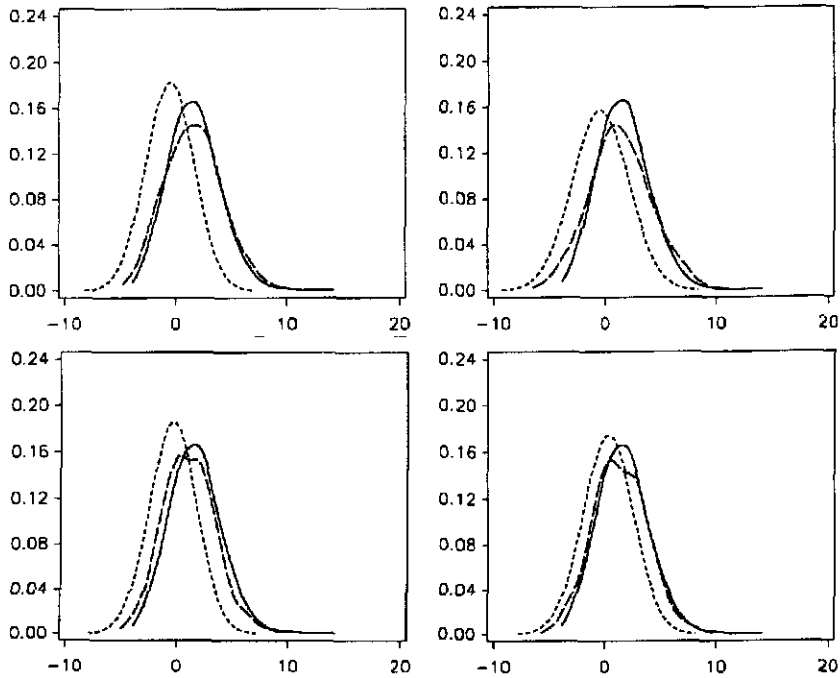

Fig. 1. Estimated exact density (solid line) of $\sqrt{Th}(\hat{s}^2(x) - s(x))$, $x = 0$, for the process of Example 2. The dashed line is the corresponding bootstrap approximation and the dotted line the asymptotic Gaussian approximation.

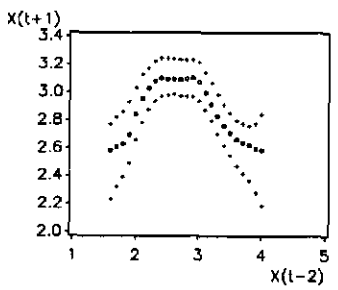

Fig. 2. A kernel estimate of the conditional mean $E(X_{t+1} \mid X_{t-2} = x)$ together with the a pointwise 90% bootstrap confidence interval for the logarithmically transformed MacKenzie River series of annual Canadian lynx trappings.

been calculated using the formula $[\hat{g}_h(x) - (Th)^{-1/2}U^*_{0.95}(x), \hat{g}_h(x) - (Th)^{-1/2}U^*_{0.05}(x)]$ where $U^*_\alpha(x)$ is the α percentage point of the distribution of the bootstrap statistic $\sqrt{Th}(\hat{g}^*_h(x) - g^*(x))$ and which has been approximated by the corresponding empirical percentage point based on 1000 bootstrap replications. The results obtained are summarized in Fig. 2.

4. Conclusions

In this paper we have seen that for a bootstrap procedure to "work" in approximating the distribution of nonparametric estimators in a time series context, it is not necessary to reproduce the whole dependence structure of the stochastic process considered but it suffices to correctly imitate the conditional distribution of which the nonparametric estimator can be considered as a "parameter". This can be interpreted as a consequence for the bootstrap of what Hart (1996) called the *whitening by windowing principle* in the context of nonparametric estimation. We proposed here a rather simple and flexible bootstrap procedure which implicitly uses a consistent estimator of the appropriate conditional distribution to generate the bootstrap replicates. Our procedure works without any particular model assumptions on the data generating process and under a quite general set of dependence conditions (cf. assumption (A5)). Furthermore, it side steps all the well-known initial estimation problems associated with applications of the bootstrap in nonparametric contexts similar to that considered here. It is worth mentioning that the idea underlying the bootstrap procedure proposed is not restricted to the case of a univariate time series discussed here but it is also applicable to nonparametric estimation problems in which the segment $\boldsymbol{Y}_{t,m}$ contains not only lagged values of X_t but also those of other stochastic processes that might influence the motion of X_{t+1}; cf. for instance Chen and Tsay (1993).

5. Proofs

The proof of Theorem 2.1 will be given in the following steps. We first establish Lemma 5.1 which deals with the uniform convergence properties of bootstrap conditional means $g^*(\boldsymbol{x};\varphi)$ to $g(\boldsymbol{x};\varphi)$ for real-valued Borel functions φ and for values of $\boldsymbol{x}$ in the with sample size decreasing interval $[\boldsymbol{x}-d;\boldsymbol{x}+d]$. According to equation (5.5) the bootstrap statistic $\sqrt{Th^m}(\hat{g}_h^*(\boldsymbol{x};\phi)-g^*(\boldsymbol{x};\phi))$ can be partioned in two terms. Lemma 5.2 shows that the first term converges weakly to the desired Gaussian distribution while Lemma 5.3 shows that the second term correctly approximates the (asymptotic) bias of the estimator $\hat{g}_h(\boldsymbol{x};\phi)$.

LEMMA 5.1. *For $d_T>0$ let $\boldsymbol{D}_d(\boldsymbol{x})=[\boldsymbol{x}-d_T;\boldsymbol{x}+d_T]$. Let further φ be a Borel function such that $E|\varphi(X_{t+1})|<\infty$. If assumptions* (A2), (A4), (A5) *and* (A8) *are fulfilled and if $d_T\to 0$ as $T\to\infty$ then*

$$\sup_{\boldsymbol{z}\in\boldsymbol{D}_d(\boldsymbol{x})}|g^*(\boldsymbol{z};\varphi)-g(\boldsymbol{z};\varphi)|\to 0$$

in probability.

PROOF. Note that $\hat{g}^*(\boldsymbol{z};\varphi)=\hat{g}_b(\boldsymbol{z};\varphi)$ is a (Nadaraya-Watson) kernel estimator of $g(\boldsymbol{z};\phi)$ and that

$$|g^*(\boldsymbol{z};\varphi)-g(\boldsymbol{z};\varphi)|\le\frac{1}{\hat{f}_{b,\boldsymbol{Y}_{t,m}}(\boldsymbol{z})}\Bigg\{|g(\boldsymbol{z};\varphi)(\hat{f}_{b,\boldsymbol{Y}_{t,m}}(\boldsymbol{z})-f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{z}))| + \left|\frac{1}{T}\sum_{j=i_m+1}^{T-1}\varphi(X_{j+1})K_b(\boldsymbol{z}-\boldsymbol{Y}_{j,m}) - \int\varphi(y)f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{z})dy\right|\Bigg\}$$

where $\hat{f}_{b,\boldsymbol{Y}_{t,m}}(\boldsymbol{x})=T^{-1}\sum_{j=i_m+1}^{T-1}K_b(\boldsymbol{x}-\boldsymbol{Y}_{j,m})$. Since $g(\boldsymbol{z};\varphi)$ is continuous, it is also bounded in $\boldsymbol{D}_d(\boldsymbol{x})$. Furthermore, by (A2) ii) $\inf_{\boldsymbol{z}\in\boldsymbol{D}_d(\boldsymbol{x})}f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{z})\ge\eta>0$. Therefore, to establish the assertion of the lemma it suffices to show that both differences on the right hand side of the above inequality goes to zero in probability and uniformly in $\boldsymbol{D}_d(\boldsymbol{x})$. In the following we show this for the term

$$\sup_{\boldsymbol{z}\in \boldsymbol{D}_d(\boldsymbol{x})}\left|T^{-1}\sum_{j=i_m+1}^{T-1}\varphi(X_{j+1})K_b(\boldsymbol{z}-\boldsymbol{Y}_{j,m})-\int\varphi(y)f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{z})dy\right|$$

only, since $\sup_{\boldsymbol{z}\in \boldsymbol{D}_d(\boldsymbol{x})}|\hat{f}_{b,\boldsymbol{Y}_{t,m}}(\boldsymbol{z})-f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{z})|\to 0$ in probability, can be shown using similar arguments. We have

$$\begin{aligned}
(5.1)\qquad &\sup_{\boldsymbol{z}\in \boldsymbol{D}_d(\boldsymbol{x})}\left|T^{-1}\sum_{j=i_m+1}^{T-1}\varphi(X_{j+1})K_b(\boldsymbol{z}-\boldsymbol{Y}_{j,m})-\int\varphi(y)f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{z})dy\right|\\
&\le \sup_{\boldsymbol{z}\in \boldsymbol{D}_d(\boldsymbol{x})}\left|T^{-1}\sum_{j=i_m+1}^{T-1}\varphi(X_{j+1})K_b(\boldsymbol{z}-\boldsymbol{Y}_{j,m})\right.\\
&\qquad\left.-\int\varphi(y)K_b(\boldsymbol{z}-\boldsymbol{v})f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{v})dyd\boldsymbol{v}\right|\\
&\quad+\sup_{\boldsymbol{z}\in \boldsymbol{D}_d(\boldsymbol{x})}\left|\int\varphi(y)K_b(\boldsymbol{z}-\boldsymbol{v})f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{v})dyd\boldsymbol{v}\right.\\
&\qquad\left.-\int\varphi(y)f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{z})dy\right|.
\end{aligned}$$

By (A2) ii) we get for the second term above

$$\begin{aligned}
&\left|\int\varphi(y)K_b(\boldsymbol{z}-\boldsymbol{v})f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{v})dyd\boldsymbol{v}-\int\varphi(y)f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{z})dy\right|\\
&\le\left|\int\varphi(y)(f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{z}-b\boldsymbol{u})-f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{z}))K(\boldsymbol{u})d\boldsymbol{u}dy\right|\\
(5.2)\qquad&\le\int|\varphi(y)||f_{\boldsymbol{Y}_{t,m}|X_t}(\boldsymbol{z}-b\boldsymbol{u}\mid y)-f_{\boldsymbol{Y}_{t,m}|X_{t+1}}(\boldsymbol{z}\mid y)|K(\boldsymbol{u})f_{X_t}(y)d\boldsymbol{u}dy\\
&\le bC\int\|\boldsymbol{u}\|K(\boldsymbol{u})d\boldsymbol{u}\int|\varphi(y)|f_{X_t}(y)dy\\
(5.3)\qquad&=O(b).
\end{aligned}$$

Consider next the first term on the right hand side of (5.1) and replace $\int\varphi(y)K_b(\boldsymbol{z}-\boldsymbol{v})f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{v})dyd\boldsymbol{v}$ by the asymptotically equivalent term $\frac{T-i_m}{T}\int\varphi(y)K_b(\boldsymbol{z}-\boldsymbol{v})f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{v})dyd\boldsymbol{v}$. Furthermore, note that $\boldsymbol{D}_d(\boldsymbol{x})$ can be covered by a finite number N_T of cubes $\boldsymbol{I}_{i,T}$ with centers $\boldsymbol{x}_i$ the sides of which have length L_T. Note that $N_T=(2d/L_T)^m$. By the Lipschitz continuity of K we then have

$$\begin{aligned}
&\sup_{\boldsymbol{z}\in \boldsymbol{D}_d(\boldsymbol{x})}\left|\frac{1}{T}\sum_{j=i_m+1}^{T-1}\varphi(X_{j+1})K_b(\boldsymbol{z}-\boldsymbol{Y}_{j,m})\right.\\
&\qquad\left.-\frac{T-i_m}{T}\int\varphi(y)K_b(\boldsymbol{z}-\boldsymbol{v})f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{v})dyd\boldsymbol{v}\right|
\end{aligned}$$

$$\leq \max_{1\leq x_i \leq N_T} \left| \frac{1}{T} \sum_{j=i_m+1}^{T-1} W_{j,T}(\boldsymbol{x}_i) \right| + O(L_T b^{-(m+1)})$$

where

$$W_{j,T}(\boldsymbol{x}_i) = \varphi(X_{j+1}) K_b(\boldsymbol{x}_i - \boldsymbol{Y}_{j,m}) - E(\varphi(X_{j+1}) K_b(\boldsymbol{x}_i - \boldsymbol{Y}_{j,m})).$$

Applying Markov's inequality we get

$$\begin{aligned}
(5.4) \qquad & P\left(\max_{1\leq x_i \leq N_T} \left| \frac{1}{T} \sum_{j=i_m+1}^{T-1} W_{j,T}(\boldsymbol{x}_i) \right| > L_T b^{-(m+1)} \right) \\
& \leq \sum_{i=1}^{N_T} P\left(\left| \frac{1}{T} \sum_{j=i_m+1}^{T-1} W_{j,T}(\boldsymbol{x}_i) \right| > L_T b^{-(m+1)} \right) \\
& \leq \frac{b^{2(m+1)}}{L_T^2} \sum_{i=1}^{N_T} E\left(\frac{1}{T} \sum_{j=i_m+1}^{T-1} W_{j,T}(\boldsymbol{x}_i) \right)^2 \\
& = O\left(\frac{b^{(m+2)} d_T^m}{L_T^{(m+2)} T} \right)
\end{aligned}$$

where the last equation follows because uniformly in $\boldsymbol{x}_i$, $E(T^{-1} \sum_{j=i_m+1}^{T-1} W_{j,T}(\boldsymbol{x}_i))^2 = O(T^{-1} b^{-m})$. Now, for $L_T = b/T^{1/(m+2)}$ and $b = O(T^{-\delta})$ with δ as in (A8) we get that, $b^{(m+2)} d_T^m / L_T^{(m+2)} T = d_T^m \to 0$ and $L_T b^{-(m+1)} \to 0$.

PROOF OF THEOREM 2.1. It suffices to show ii) since assertion i) follows by the same arguments but by ignoring the bias term. To prove this assertion note first that

$$\begin{aligned}
(5.5) \qquad & \sqrt{Th^m}(\hat{g}_h^*(\boldsymbol{x};\phi) - g^*(\boldsymbol{x};\phi)) \\
& = \frac{1}{\hat{f}_{h,\boldsymbol{Y}_{t,m}}(\boldsymbol{x})} \sqrt{\frac{h^m}{T}} \sum_{j=i_m+1}^{T-1} (\phi(X_{j+1}^*) - g^*(\boldsymbol{x};\phi)) K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) \\
& = \frac{1}{\hat{f}_{h,\boldsymbol{Y}_{t,m}}(\boldsymbol{x})} \sqrt{\frac{h^m}{T}} \sum_{j=i_m+1}^{T-1} (\phi(X_{j+1}^*) - g^*(\boldsymbol{Y}_{j,m};\phi)) K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) \\
& \quad + \frac{1}{\hat{f}_{h,\boldsymbol{Y}_{t,m}}(\boldsymbol{x})} \sqrt{\frac{h^m}{T}} \sum_{j=i_m+1}^{T-1} (m^*(\boldsymbol{Y}_{j,m};\phi) - g^*(\boldsymbol{x};\phi)) K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m}).
\end{aligned}$$

Now, since under the assumptions made $\hat{f}_{h,\boldsymbol{Y}_{t,m}}(\boldsymbol{x}) \to f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})$ in probability, the desired result follows immediately from Lemmas 5.2 and 5.3.

LEMMA 5.2. *Under assumptions* (A1)–(A5) *and* (A8) *we have conditionally on* $X_1, X_2, \ldots, X_T$ *that, as* $T \to \infty$

$$\sqrt{\frac{h^m}{T}} \sum_{j=i_m+1}^{T-1} (\phi(X_{j+1}^*) - g^*(\boldsymbol{Y}_{j,m};\phi)) K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) \Rightarrow N(0, \tau^2(\boldsymbol{x};\phi))$$

where $\tau^2(\boldsymbol{x};\phi) = f_{\boldsymbol{Y}_{t,m}}^2(\boldsymbol{x}) \sigma^2(\boldsymbol{x};\phi)$ *and* $\sigma^2(\boldsymbol{x};\phi)$ *is given in* (1.3).

PROOF. Note that conditionally on $X_1, X_2, \ldots, X_T$,

$$\{(\phi(X^*_{j+1}) - g^*(\boldsymbol{Y}_{j,m};\phi))K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m}),$$
$$j = i_m+1, i_m+2, \ldots, T-1; T = i_m+1, i_m+2, \ldots, \}$$

forms a triangular array of independent random variables. Define $U^*_{j+1} = \phi(X^*_{j+1}) - g^*(\boldsymbol{Y}_{j,m};\phi)$ and observe that conditionally on $\boldsymbol{Y}_{j,m}$ we have for the s-order moment of U^*_{j+1}

$$E^*(U^*_{j+1})^s = \frac{\sum_{l=i_m+1}^{T-1}(\phi(X_{l+1}) - g^*(\boldsymbol{Y}_{j,m};\phi))^s K_b(\boldsymbol{Y}_{j,m} - \boldsymbol{Y}_{l,m})}{\sum_{n=i_m+1}^{T-1} K_b(\boldsymbol{Y}_{j,m} - \boldsymbol{Y}_{n,m})}.$$

Clearly,

$$E^*(U^*_{j+1}) = E^*(\phi(X^*_{j+1}) \mid \boldsymbol{Y}_{j,m}) - g^*(\boldsymbol{Y}_{j,m};\phi) = 0.$$

Now, to establish the theorem we have to show that

$$\text{(5.6)} \qquad \frac{h^m}{T}\sum_{j=i_m+1}^{T-1} E^*(U^*_{j+1})^2 K_h^2(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) \to \tau^2(\boldsymbol{x};\phi)$$

and that for every $\eta > 0$ the Linderberg condition

$$\text{(5.7)} \qquad \frac{h^m}{T}\sum_{j=i_m+1}^{T-1} E^*[(U^*_{j+1}K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m}))^2 I(h^m T^{-1} U^{*^2}_{j+1} K_h^2(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) > \eta)] \to 0$$

in probability, is satisfied.

Consider first (5.6) and let

$$V(\boldsymbol{x};\phi) = \int (\phi(y) - g(\boldsymbol{x};\phi))^2 f_{X_{t+1}|\boldsymbol{Y}_{t,m}}(y \mid \boldsymbol{x})dy$$

and $V^*(\boldsymbol{x};\phi) = E^*(U^*_{j+1})^2$ given that $\boldsymbol{Y}_{j,m} = \boldsymbol{x}$. We then have

$$\begin{aligned}
&\frac{h^m}{T}\sum_{j=i_m+1}^{T-1} E^*(U^*_{j+1})^2 K_h^2(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) \\
&= \frac{h^m}{T}\sum_{j=i_m+1}^{T-1} V(\boldsymbol{Y}_{j,m};\phi) K_h^2(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) \\
&\quad + \frac{h^m}{T}\sum_{j=i_m+1}^{T-1} (V^*(\boldsymbol{Y}_{j,m},\phi) - V(\boldsymbol{Y}_{j,m};\phi)) K_h^2(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) \\
&= T_1 + T_2
\end{aligned}$$

with an obvious notation for T_1 and T_2. Now by standard arguments and by the continuity of $V(\cdot;\phi)$ we have $T_1 \to \tau^2(\boldsymbol{x};\phi)$ in probability. Furthermore, using $V(\boldsymbol{z};\phi) = g(\boldsymbol{z};\phi(x) = x^2) - g^2(\boldsymbol{z};\phi(x) = x)$ and $V^*(\boldsymbol{z};\phi) = g^*(\boldsymbol{z};\phi(x) = x^2) - (g^*(\boldsymbol{z};\phi(x) = x))^2$, we get for the second term and for $\varphi = \phi^2$ resp. $\varphi = \phi$ in Lemma 5.1 that

$$\begin{aligned}
|T_2| &\le \sup_{\boldsymbol{z}\in[\boldsymbol{x}-h,\boldsymbol{x}+h]} |V^*(\boldsymbol{z};\phi) - V(\boldsymbol{z};\phi)| \frac{1}{T}\sum_{j=i_m+1}^{T-1} K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) \\
&\le O_P(1) \sup_{\boldsymbol{z}\in[\boldsymbol{x}-h,\boldsymbol{x}+h]} |g^*(\boldsymbol{z};\phi^2) - g(\boldsymbol{z};\phi^2)| \\
&\quad + O_P(1) \sup_{\boldsymbol{z}\in[\boldsymbol{x}-h,\boldsymbol{x}+h]} |g^*(\boldsymbol{z};\phi) - g(\boldsymbol{z};\phi)| \sup_{\boldsymbol{z}\in[\boldsymbol{x}-h,\boldsymbol{x}+h]} |g^*(\boldsymbol{z};\phi) + g(\boldsymbol{z};\phi)| \\
&\to 0 \quad \text{in probability.}
\end{aligned}$$

To establish the Lindeberg condition (5.7) note that

$$\sum_{j=i_m+1}^{T-1} E^*[h^m T^{-1}(U_{j+1}^* K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m}))^2 I(h^m T^{-1} U_{j+1}^{*^2} K_h^2(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) > \eta)]$$
$$\leq \frac{h^{2m}}{\eta T^2} \sum_{j=i_m+1}^{T-1} E^*(U_{j+1}^*)^4 K_h^4(\boldsymbol{x} - \boldsymbol{Y}_{j,m}).$$

To evaluate this term we proceed as in the proof of (5.6) and first show that

$$\frac{h^{3m}}{T} \sum_{j=i_m+1}^{T-1} E^*(U_{j+1}^*)^4 K_h^4(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) = O_P(1).$$

For this let

$$M(\boldsymbol{x};\phi) = \int (\phi(y) - g(\boldsymbol{x};\phi))^4 f_{X_{t+1}|\boldsymbol{Y}_{t,m}}(y \mid \boldsymbol{x}) dy$$

and $M^*(\boldsymbol{x};\phi) = E^*(U_{j+1}^*)^4$ conditionally on $\boldsymbol{Y}_{j,m} = \boldsymbol{x}$. We then have

$$\text{(5.8)} \qquad \frac{h^{3m}}{T} \sum_{j=i_m+1}^{T-1} M(\boldsymbol{Y}_{j,m}) K_h^4(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) \to M(\boldsymbol{x};\phi) f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x}) \int K^4(\boldsymbol{u}) d\boldsymbol{u}$$

and

$$\text{(5.9)} \qquad \frac{h^{3m}}{T} \sum_{j=i_m+1}^{T-1} |M^*(\boldsymbol{Y}_{j,m};\phi) - M(\boldsymbol{Y}_{j,m};\phi)| K_h^4(\boldsymbol{x} - \boldsymbol{Y}_{j,m})$$
$$\leq \sup_{\boldsymbol{z} \in [\boldsymbol{x}-h, \boldsymbol{x}+h]} |M^*(\boldsymbol{z};\phi) - M(\boldsymbol{z};\phi)| O_p(1).$$

Now, since

$$M^*(\boldsymbol{x};\phi) = g^*(\boldsymbol{x};\phi^4) - 4g^*(\boldsymbol{x};\phi^3) g^*(\boldsymbol{x};\phi) - 2g^*(\boldsymbol{x};\phi^2) g^{*^2}(\boldsymbol{x};\phi) - 3g^{*^4}(\boldsymbol{x};\phi)$$

and

$$M(\boldsymbol{x};\phi) = g(\boldsymbol{x};\phi^4) - 4g(\boldsymbol{x};\phi^3) g(\boldsymbol{x};\phi) - 2g(\boldsymbol{x};\phi^2) g^2(\boldsymbol{x};\phi) - 3g^4(\boldsymbol{x};\phi),$$

it follows using Lemma 5.1 that $\sup_{\boldsymbol{z} \in [\boldsymbol{x}-h, \boldsymbol{x}+h]} |M^*(\boldsymbol{z};\phi) - M(\boldsymbol{z};\phi)| \to 0$. Because of this and equations (5.8) and (5.9) we, finally, conclude that

$$\sum_{j=i_m+1}^{T-1} E^*[h^m T^{-1}(U_{j+1}^* K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m}))^2 I(h^m T^{-1} U_{j+1}^{*^2} K_h^2(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) > \eta)]$$
$$= O_P(\eta^{-1} T^{-1} h^{-m}).$$

LEMMA 5.3. *Under assumptions* (A1)–(A7) *and for* $g \sim T^{-\delta}$ *with* $0 < \delta < 1/\{(m+2)^2+1\}$ *we have that, as* $T \to \infty$

$$\frac{1}{\hat{f}_{h,\boldsymbol{Y}_{t,m}}(\boldsymbol{x})} \sqrt{\frac{h^m}{T}} \sum_{j=i_m+1}^{T-1} (g^*(\boldsymbol{Y}_{j,m};\phi) - g^*(\boldsymbol{x};\phi)) K_h(\boldsymbol{x} - \boldsymbol{Y}_{j,m}) \to B(\boldsymbol{x};\phi)$$

in probability, where $B(\boldsymbol{x};\phi)$ *is given in* (1.4).

PROOF. Using (A7) we have

$$\sqrt{\frac{h^m}{T}}\sum_{j=i_m+1}^{T-1}(g^*(\boldsymbol{Y}_{j,m};\phi)-g^*(\boldsymbol{x};\phi))K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m})$$
$$=\sum_{k=1}^{m} m^{*^{(x_{i_k})}}(\boldsymbol{x};\phi)\sqrt{\frac{h^m}{T}}\sum_{j=i_m+1}^{T-1}(X_{j-i_k}-x_{i_k})K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m})$$
$$+\frac{1}{2}\sum_{k,s=1}^{m}\sqrt{\frac{h^m}{T}}\sum_{j=i_m+1}^{T-1}(X_{j-i_k}-x_{i_k})(X_{j-i_s}-x_{i_s})$$
$$\times g^{*^{(x_{i_k}x_{i_s})}}(\tilde{\boldsymbol{Y}}_{j,m};\phi)K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m})$$

where $\tilde{\boldsymbol{Y}}_{j,m}$ is a suitable value between $\boldsymbol{x}$ and $\boldsymbol{Y}_{j,m}$ and $g^{*^{(x_{i_k})}}(\boldsymbol{x};\phi)$ and $g^{*^{(x_{i_k}x_{i_s})}}(\boldsymbol{x};\phi)$ denote the first and second order partial derivatives of $g^*(\boldsymbol{x};\phi)$ with respect to x_{i_k} and x_{i_k} and x_{i_s} respectively. To prove the lemma it suffices to show that in probability,

$$\sqrt{\frac{h^m}{T}}\sum_{j=i_m+1}^{T-1}(X_{j-i_k}-x_{i_k})K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m})\to C_h f_{\boldsymbol{Y}_{t,m}}^{(x_{i_k})}(\boldsymbol{x})K_2, \tag{5.10}$$

$$g^{*^{(x_{i_k})}}(\boldsymbol{x};\phi)\to g^{(x_{i_k})}(\boldsymbol{x};\phi), \tag{5.11}$$

$$\frac{1}{2}\sum_{k,s=1}^{m}\tilde{g}^{(x_{i_k}x_{i_s})}(\boldsymbol{x};\phi)\sqrt{\frac{h^m}{T}}\sum_{j=i_m+1}^{T-1}(X_{j-i_k}-x_{i_k})(X_{j-i_s}-x_{i_s})K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m}) \tag{5.12}$$
$$\to\frac{1}{2}C_hK_2f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})\sum_{k=1}^{m}g^{(x_{i_k}x_{i_k})}(\boldsymbol{x};\phi)$$

and

$$\sup_{\boldsymbol{z}\in[\boldsymbol{x}-h,\boldsymbol{x}+h]}|g^{*^{(x_{i_k}x_{i_k})}}(\boldsymbol{z};\phi)-\tilde{g}^{(x_{i_k}x_{i_k})}(\boldsymbol{z};\phi)|\to 0 \tag{5.13}$$

where

$$\tilde{g}^{(x_{i_k}x_{i_s})}(\boldsymbol{z};\phi) \tag{5.14}$$
$$=\frac{1}{f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{z})}\left\{\frac{1}{2}(1+\delta_{k,s})\left(\int\phi(y)f_{X_{t+1}\boldsymbol{Y}_{t,m}}^{(x_{i_k}x_{i_s})}(y,\boldsymbol{z})dy-g(\boldsymbol{z};\phi)f_{\boldsymbol{Y}_{t,m}}^{(x_{i_k}x_{i_s})}(\boldsymbol{z})\right)\right.$$
$$\left.-g^{(x_{i_k})}(\boldsymbol{z};\phi)f_{\boldsymbol{Y}_{t,m}}^{(x_{i_s})}(\boldsymbol{z})-g^{(x_{i_s})}(\boldsymbol{z};\phi)f_{\boldsymbol{Y}_{t,m}}^{(x_{i_k})}(\boldsymbol{z})\right\}.$$

Note that $\tilde{g}^{(x_{i_k}x_{i_k})}(\boldsymbol{z};\phi)=g^{(x_{i_k}x_{i_k})}(\boldsymbol{z};\phi)$.

Establishing (5.10) to (5.13) the assertion of the lemma follows then because

$$\frac{1}{2}\left\{\int\phi(y)f_{X_{t+1}\boldsymbol{Y}_{t,m}}^{(x_{i_k},x_{i_k})}(y,\boldsymbol{x})dy-g(\boldsymbol{x};\phi)f_{\boldsymbol{Y}_{t,m}}^{(x_{i_k},x_{i_k})}(\boldsymbol{x})\right\}$$
$$=g^{(x_{i_k})}(\boldsymbol{x};\phi)f_{\boldsymbol{Y}_{t,m}}^{(x_{i_k})}(\boldsymbol{x})+\frac{1}{2}g^{(x_{i_k}x_{i_k})}(\boldsymbol{x};\phi)f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x}).$$

Since the proofs of (5.10) to (5.13) include several and tedious manipulations of formulaes we omit in what follows the details and stress only the essentials. In the sequel $K_b^{(x_{i_k})}(\boldsymbol{x}-\boldsymbol{Y}_{j,m})$ and $K_b^{(x_{i_k}x_{i_s})}(\boldsymbol{x}-\boldsymbol{Y}_{j,m})$ denote the first and second order partial derivatives of $K_b(\boldsymbol{x}-\boldsymbol{Y}_{j,m})$ with respect to the variables x_{i_k} and x_{i_s}.

To prove (5.10) note that for $\boldsymbol{v}=(v_{i_1},v_{i_2},\ldots,v_{i_m})$ we have

$$\begin{aligned}\sqrt{\frac{h^m}{T}}E\left(\sum_{j=i_m+1}^{T-1}(X_{j-i_k}-x_{i_k})K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m})\right)&=\sqrt{h^mT}\int(v_{i_k}-x_{i_k})K_h(\boldsymbol{x}-\boldsymbol{v})f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{v})d\boldsymbol{v}\\&=T^{1/2}h^{(m+4)/2}\int u_k^2K(\boldsymbol{u})d\boldsymbol{u}f_{\boldsymbol{Y}_{t,m}}^{(x_{i_k})}(\boldsymbol{x})+o(1).\end{aligned}$$

Furthermore,

$$\begin{aligned}\frac{h^m}{T}\sum_{j=i_m+1}^{T-1}\operatorname{Var}((X_{j-i_k}-x_{i_k})K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m}))&\le h^{m+2}\operatorname{Var}(K_h(\boldsymbol{x}-\boldsymbol{Y}_{t,m}))\\&=O(h^2)\end{aligned}$$

while by (A5) and as for (2.1)

$$\frac{h^m}{T}\sum_{\substack{s,j=i_m+1\\ s\neq j}}^{T-1}|\operatorname{Cov}((X_{s-i_k}-x_{i_k})K_h(\boldsymbol{x}-\boldsymbol{Y}_{s,m}),(X_{j-i_k}-x_{i_k})K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m}))|\to 0.$$

To prove (5.11) express the partial derivative as

$$\begin{aligned}g^{*(x_{i_k})}(\boldsymbol{x};\phi)=\frac{1}{\hat{f}_g(\boldsymbol{x})}\Bigg\{&\frac{1}{T}\sum_{j=i_m+1}^{T-1}\phi(X_{j+1})K_b^{(x_{i_k})}(\boldsymbol{x}-\boldsymbol{Y}_{j,m})\\&-g^*(\boldsymbol{x};\phi)\frac{1}{T}\sum_{j=i_m+1}^{T-1}K_b^{(x_{i_k})}(\boldsymbol{x}-\boldsymbol{Y}_{j,m})\Bigg\}\end{aligned}$$

and verify that

$$\frac{1}{T}\sum_{j=i_m+1}^{T-1}\phi(X_{j+1})K_b^{(x_{i_k})}(\boldsymbol{x}-\boldsymbol{Y}_{j,m})\to\int\phi(y)f_{X_{t+1}\boldsymbol{Y}_{t,m}}^{(x_{i_k})}(y,\boldsymbol{x})dy \tag{5.15}$$

and

$$\frac{1}{T}\sum_{j=i_m+1}^{T-1}K_b^{(x_{i_k})}(\boldsymbol{x}-\boldsymbol{Y}_{j,m})\to f_{\boldsymbol{Y}_{t,m}}^{(x_{i_k})}(\boldsymbol{x}) \tag{5.16}$$

in probability. Consider for instance (5.15). Using $\int K^{(u_{i_k})}(\boldsymbol{u})d\boldsymbol{u}=0$ and $\int u_{i_s}K^{(u_{i_k})}(\boldsymbol{u})d\boldsymbol{u}=-\delta_{s,k}$ we get by straightforward calculations

$$\frac{1}{T}E\left(\sum_{j=i_m+1}^{T-1}\phi(X_{j+1})K_b^{(x_{i_k})}(\boldsymbol{x}-\boldsymbol{Y}_{j,m})\right)=\int\phi(y)f_{X_{t+1}\boldsymbol{Y}_{t,m}}^{(x_{i_k})}(y,\boldsymbol{x})dy+o(1),$$

$$\frac{1}{T^2}\sum_{j=i_m+1}^{T-1}\operatorname{Var}(\phi(X_{j+1})K_b^{(x_{i_k})}(\boldsymbol{x}-\boldsymbol{Y}_{j,m}))=O((Tb^{m+2})^{-1}),$$

and

$$\frac{1}{T^2}\sum_{\substack{s,j=i_m+1\\ s\neq j}}^{T-1} \mathrm{Cov}(\phi(X_{s+1})K_b^{(x_{i_k})}(\boldsymbol{x}-\boldsymbol{Y}_{s,m}),\phi(X_{j+1})K_b^{(x_{i_k})}(\boldsymbol{x}-\boldsymbol{Y}_{j,m})) \to 0.$$

To establish (5.12) note that

$$E\left(\frac{1}{2}\sum_{k,s=1}^{m}\tilde{g}^{(x_{i_k}x_{i_s})}(\boldsymbol{x};\phi)\sqrt{\frac{h^m}{T}}\sum_{j=i_m+1}^{T-1}(X_{j-i_k}-x_{i_k})(X_{j-i_s}-x_{i_s})K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m})\right)$$
$$=\frac{(T-i_m)h^{(m+4)/2}}{2\sqrt{T}}\sum_{k,s=1}^{m}\tilde{g}^{(x_{i_k}x_{i_s})}(\boldsymbol{x};\phi)\int u_{i_k}u_{i_s}K(\boldsymbol{u})f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x}-\boldsymbol{u}h)d\boldsymbol{u}$$
$$=\frac{1}{2}C_hK_2f_{\boldsymbol{Y}_{t,m}}(\boldsymbol{x})\sum_{k=1}^{m}g^{(x_{i_k}x_{i_k})}(\boldsymbol{x};\phi)+o(1)$$

and that

$$\frac{h^m}{T}\mathrm{Var}\left(\sum_{i=i_m+1}^{T-1}(X_{j-i_k}-x_{i_k})(X_{j-i_s}-x_{i_s})K_h(\boldsymbol{x}-\boldsymbol{Y}_{j,m})\right)=O(h^4).$$

Finally, to establish (5.13) verify first that

$$\begin{aligned}(5.17)\qquad g^{*(x_{i_k}x_{i_s})}(\boldsymbol{z};\phi)=\frac{1}{\hat{f}_{b,\boldsymbol{Y}_{t,m}}(\boldsymbol{z})}\Bigg\{&\frac{1}{T}\sum_{j=i_m+1}^{T-1}\phi(X_{j+1})K_b^{(x_{i_k}x_{i_s})}(\boldsymbol{z}-\boldsymbol{Y}_{j,m})\\ &-g^*(\boldsymbol{z};\phi)\frac{1}{T}\sum_{j=i_m+1}^{T-1}K_b^{(x_{i_k}x_{i_s})}(\boldsymbol{z}-\boldsymbol{Y}_{j,m})\\ &-g^{*(x_{i_k})}(\boldsymbol{z};\phi)\frac{1}{T}\sum_{j=i_m+1}^{T-1}K_b^{(x_{i_s})}(\boldsymbol{z}-\boldsymbol{Y}_{j,m})\\ &-g^{*(x_{i_s})}(\boldsymbol{z};\phi)\frac{1}{T}\sum_{j=i_m+1}^{T-1}K_b^{(x_{i_k})}(\boldsymbol{z}-\boldsymbol{Y}_{j,m})\Bigg\}.\end{aligned}$$

By the same arguments as in Lemma 5.1 it can be seen that to establish the desired uniform convergence it suffices to show that every component on the right hand side of (5.17) converges uniformly on $[\boldsymbol{x}-h,\boldsymbol{x}+h]$ to the corresponding component on the right hand side of (5.14). Since the arguments used are very similar for all terms considered, we demonstrate this for the term $T^{-1}\sum_{j=i_m+1}^{T-1}\phi(X_{j+1})K_b^{(x_{i_k}x_{i_s})}(\boldsymbol{z}-\boldsymbol{Y}_{j,m})$. Let $S_T(\boldsymbol{x})=\sum_{j=i_m+1}^{T-1}\phi(X_{j+1})K_b^{(x_{i_k}x_{i_s})}(\boldsymbol{z}-\boldsymbol{Y}_{j,m})$. We then have

$$\begin{aligned}(5.18)\qquad &\left|\frac{1}{T}E(S_T(\boldsymbol{z}))-\frac{1}{2}(1+\delta_{k,s})\int\phi(y)f^{(x_{i_k}x_{i_s})}_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{z})dy\right|\\ &=\Bigg|\frac{T-i_m}{T}\int\phi(y)K_b^{(x_{i_k}x_{i_s})}(\boldsymbol{z}-\boldsymbol{v})f_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{v})dyd\boldsymbol{v}\\ &\qquad-\frac{1+\delta_{k,s}}{2}\int\phi(y)f^{(x_{i_k}x_{i_s})}_{X_{t+1}\boldsymbol{Y}_{t,m}}(y,\boldsymbol{z})dy\Bigg|.\end{aligned}$$

Using (A7), $\int K^{(u_{i_k}u_{i_s})}(u)du = \int u_{i_s}K^{(u_{i_k}u_{i_s})}(u)du = 0$, $\int u_{i_s}u_{i_k}K^{(u_{i_s},u_{i_k})}(u)du = 1+\delta_{s,k}$ the substitution $z_i - v_i = u_i b$ and a Taylor series expansion of $f_{X_{t+1}Y_{t,m}}(y, z-ub)$, we get similarly as in (5.2) that the expression on the right hand side of (5.18) is $O(b)$ uniformly in z. Furthermore, by the mean value theorem and applying the splitting device of the supremum used in the proof of Lemma 5.1 we get

$$\sup_{z\in[x-h,x+h]}\left|\frac{1}{T}(S_T(z) - E(S_T(z)))\right| \leq \max_{1\leq i\leq N_T}\left|\frac{1}{T}(S_T(x_i) - ES_T(x_i))\right| + O(L_T b^{-(m+3)}).$$

Since $\mathrm{Var}(T^{-1}(S_T(z))) = O(T^{-1}b^{-(m+4)})$ uniformly in z, we get using Markov's inequality and along the same lines as in (5.4) that

$$P\left(\max_{1\leq i\leq N_T}\left|\frac{1}{T}(S_T(x_i) - E(S_T(x_i)))\right| > L_T b^{-(m+3)}\right) = O\left(\frac{b^{(m+2)}h^m}{L_T^{(m+2)}T}\right).$$

Thus for L_T as in Lemma 5.1 and $b = O(T^{-\delta})$ with $0 < \delta < 1/(m+2)^2$ we have that, $b^{(m+2)}h^m/(L_T^{(m+2)}T) = O(h^m) \to 0$ and $L_T b^{-(m+3)} \to 0$.

Acknowledgements

The authors are very grateful to the referees for their valuable comments.

REFERENCES

Auestad, B. and D. Tjøstheim (1990). Identification of nonlinear time series: first order characterization and order determination, *Biometrika*, **77**, 669–687.

Cao-Abad, R. and Gonzalez-Manteiga, W. (1993). Bootstrap methods in regression smoothing, *J. Nonparametr. Statist.*, **2**, 379–388.

Chen, R. and Tsay, R. S. (1993). Nonlinear additive ARX models, *J. Amer. Statist. Assoc.*, **88**, 955–967.

Doukhan, P. (1994). *Mixing. Properties and Examples*, Lecture Notes in Statist., No. 85, Springer, New York.

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

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Falk, M. and Reiss, R.-D. (1992). Bootstrapping conditional curves, *Bootstrapping and Related Techniques* (eds. K. H. Jöckel, G. Rothe and W. Sendler), Lecture Notes in Econom. and Math. Systems, No. 376, Springer, New York.

Franke, J. and Härdle, W. (1992). On bootstrapping kernel spectral estimates, *Ann. Statist.*, **20**, 121–145.

Franke, J., Kreiss, J.-P. and 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. and Bowmann, A. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands, *J. Amer. Statist. Assoc.*, **83**, 100–110.

Härdle, W. and Vieu, P. (1992). Kernel regression smoothing of time series, *J. Time Ser. Anal.*, **13**, 209–232.

Hart, J. (1996). Some automated methods of smoothing time-dependent data, *J. Nonparametr. Statist.*, **6**, 115–142.

Kim, T. Y. and Cox, D. D. (1996). Bandwidth selection in kernel smoothing of time series, *J. Time Ser. Anal.*, **17**, 49–63.

Künsch, H. R. (1989). The Jackknife and the bootstrap for general stationary observations, *Ann. Statist.*, **17**, 1217–1241.

Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence, *Exploring the Limits of the Bootstrap* (eds. R. LePage and L. Billard), 225–248, Wiley, New York.

Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series, *Econometric Theory*, **11**, 258–289.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*, Springer, London.

Neumann, M. (1997). On robustness of model-based bootstrap schemes in nonparametric time series analysis, Discussion Paper No. 88/97, SFB 373, Humbolt University, Berlin.

Neumann, M. and Kreiss, J.-P. (1998). Regression-type inference in nonparametric autoregressions, *Ann. Statist.*, **26**, 1570–1613.

Paparoditis, E. and Politis, D. N. (1997). The local bootstrap for Markov processes (preprint).

Politis, D. N. and 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. Inst. Statist. Math.*, **42**, 253–268.

Robinson, P. M. (1983). Nonparametric estimation for time series models, *J. Time Ser. Anal.*, **4**, 185–208.

Romano, J. P. (1988). Bootstrapping the mode, *Ann. Inst. Statist. Math.*, **40**, 565–586.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition, *Proc. Nat. Acad. Sci. U.S.A.*, **42**, 43–47.

Rosenblatt, M. (1991). *Stochastic Curve Estimation*, NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 3, Institute of Mathematical Statistics, Alexandria, Virginia.

Shao, J. and Tu, D. (1995). *The Jackknife and the Bootstrap*, Springer, New York.

Shi, S. G. (1991). Local bootstrap, *Ann. Inst. Statist. Math.*, **43**, 667–676.

Tjøstheim, D. (1990). Nonlinear time series and Markov chains, *Advances in Applied Probability*, **22**, 587–611.

Tjøstheim, D. (1994). Nonlinear time series: a selective review, *Scand. J. Statist.*, **21**, 97–130.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical Approach*, Oxford University Press, New York.