Econometric Theory, **17**, 2001, 540–566. Printed in the United States of America.

A MARKOVIAN LOCAL RESAMPLING SCHEME FOR NONPARAMETRIC ESTIMATORS IN TIME SERIES ANALYSIS

EFSTATHIOS PAPARODITIS
University of Cyprus

DIMITRIS N. POLITIS
University of California, San Diego

In this paper we study the properties of a pth-order Markovian local resampling procedure in approximating the distribution of nonparametric (kernel) estimators of the conditional expectation $m(\mathbf{x};\phi) = E(\phi(X_{t+1})|\mathbf{Y}_{t,p} = \mathbf{x})$ where $\{\mathbf{X}_t, t \geq 1\}$ is a strictly stationary process, $\mathbf{Y}_{t,p} = (X_t, X_{t-1}, \ldots, X_{t-p+1})^\top$, and $\phi(\cdot)$ is a measurable real-valued function. Under certain regularity conditions, asymptotic validity of the proposed resampling scheme is established for a class of stochastic processes that is broader than the class of stationary Markov processes. Some simulations illustrate the finite sample performance of the proposed resampling procedure.

1. INTRODUCTION

Let $\{X_t, t = 1,2,\ldots\}$ be a (strictly) stationary real-valued stochastic process and denote by $F_{\mathbf{Y}_{t,p}}(\cdot)$ the marginal distribution of $\mathbf{Y}_{t,p} = (X_t, X_{t-1}, \ldots, X_{t-p+1})^\top$ and by $F_{X_{t+1}|\mathbf{Y}_{t,p}}(\cdot|\cdot)$ the conditional distribution of X_{t+1} given $\mathbf{Y}_{t,p}$. In this paper we are interested in estimating the sampling behavior of a nonparametric (kernel) estimator of the conditional expectation $m(\mathbf{x};\phi) = E(\phi(X_{t+1})|\mathbf{Y}_{t,p} = \mathbf{x})$ where $\phi(\cdot)$ is a real-valued, measurable function. Several conditional functionals of the process can be obtained by appropriately specifying the function ϕ. For instance, for $\phi_1(z) = z$ and $\phi_2(z) = z^2$ we have that $m(\mathbf{x};\phi_1)$ is the conditional mean, whereas $m(\mathbf{x};\phi_2) - m^2(\mathbf{x};\phi_1)$ is the conditional variance of X_{t+1} given that $\mathbf{Y}_{t,p} = \mathbf{x}$. Apart from the conditional moments $E(X_{t+1}^s|\mathbf{Y}_{t,p} = \mathbf{x})$ the function $m(\mathbf{x};\phi)$ can also be used to describe some other conditional functionals of the process that may be of interest. As an example, for $\phi(z) = \mathbf{1}_{(-\infty,y]}(z)$ and y a real number we have that $m(\mathbf{x};\phi) = F_{X_{t+1}|\mathbf{Y}_{t,p}}(y|\mathbf{x})$, that is, the conditional distribution function mentioned previously.

We are grateful to the co-editor Joel Horowitz and to both referees for a number of useful comments. This research has been supported in part by National Science Foundation Grant DMS-97-03964 and by a University of Cyprus Research Grant. Address correspondence to: Efstathios Paparoditis, Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, CY-1678 Nicosia, Cyprus; e-mail: stathisp@ucy.ac.cy.

 © 2001 Cambridge University Press 0266-4666/01 $9.50

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

Given observations $X_1, X_2, \ldots, X_T$ the nonparametric estimator of $m(\mathbf{x};\phi)$, which we consider in this paper, is given by

$$\hat{m}_h(\mathbf{x};\phi) = \frac{\sum_{t=p}^{T-1} \phi(X_{t+1}) K_h(\mathbf{x} - \mathbf{Y}_{t,p})}{\sum_{t=p}^{T-1} K_h(\mathbf{x} - \mathbf{Y}_{t,p})}. \tag{1.1}$$

In this notation $K_h(\cdot) = h^{-p}K(\cdot/h)$ where $K: \mathbf{R}^p \to \mathbf{R}$ is a nonnegative kernel satisfying $\int K(\mathbf{u})d\mathbf{u} = 1$, $\int \mathbf{u}K(\mathbf{u})\,d\mathbf{u} = 0$ and $h = h(T) > 0$ is the bandwidth used to smooth the observations. Note that (1.1) is the common Nadaraya–Watson estimator of $m(\mathbf{x};\phi)$. Nonparametric estimators such as the one considered in this paper are frequently used as an end product in time series analysis for modeling or predictive purposes or, perhaps more important, as a guide in identifying a parametric model to be used in a subsequent stage. For instance, a general univariate model that contains many of the linear and nonlinear classes discussed in the literature, including the popular nonlinear conditional heteroskedastic (ARCH) model, is given by $X_{t+1} = g(X_t, X_{t-1}, \ldots, X_{t-p+1}, \varepsilon_t)$ where $g: \mathbf{R}^{p+1} \to \mathbf{R}$ is a measurable function. Here $\{\varepsilon_t\}$ is an independent and identically distributed (i.i.d.) sequence of random variables with mean zero and finite variance and ε_t is assumed to be independent of X_s for $s \leq t$ (cf. Tong, 1990, Ch. 3). One way to gain insight into the nonlinearity features of such models given a set of observations is to estimate nonparametrically functionals such as the conditional mean or the conditional variance and to compare the estimates obtained with those expected under a particular hypothesis about the underlying model. Such an approach of detecting and modeling nonlinearity in time series analysis has been proposed by several authors; see among others Tong (1990), Auestad and Tjøstheim (1990), and Tjøstheim and Auestad (1994).

Under certain regularity conditions, strong consistency and asymptotic normality of the estimator $\hat{m}_h(\mathbf{x};\phi)$ have been established by several authors; we mention here among others the papers by Robinson (1983), Roussas (1990), and Masry and Tjøstheim (1995). The asymptotic normality of the estimator (1.1) is, for instance, useful in constructing pointwise confidence intervals for $m(\mathbf{x};\phi)$. However, there is an inherent difficulty in using the limiting Gaussian distribution for such purposes that is due to the fact that the mean and the variance of this distribution depend on unknown (and difficult to estimate) characteristics of the process.

The aim of this paper is to show that a Markovian local bootstrap procedure proposed by Paparoditis and Politis (1997) and which generates bootstrap replicates $X_1^*, X_2^*, \ldots, X_T^*$ that "reproduce" correctly the conditional distribution of X_{t+1} given the "past" $\mathbf{Y}_{t,p}$ leads to an asymptotically valid approximation of the distribution of

$$L_T(\mathbf{x}) = \sqrt{Th^p}\,(\hat{m}_h(\mathbf{x};\phi) - m(\mathbf{x};\phi)). \tag{1.2}$$

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

The proposed procedure provides an alternative (as compared to the normal) approximation of the distribution of $L_T(\mathbf{x})$ and at the same time leads to consistent estimation of quantities such as the mean and the variance of the distribution of interest. Although the local bootstrap procedure generates a pseudoseries $X_1^*, X_2^*, \ldots, X_T^*$ with a Markov dependence structure, we show in the present paper that the validity of this resampling scheme for the class of nonparametric estimators considered is *not* limited to the Markov case only. To understand this fact intuitively note that the information regarding estimation of $F_{X_{t+1}|\mathbf{Y}_{t,p}}(\cdot|\cdot)$ and its moments typically lies in the "scatterplot" of X_{t+1} vs. $\mathbf{Y}_{t,p}$ for $t = 1,2,\ldots,T-p$; the crucial feature of this scatterplot is the joint distribution of X_{t+1} and $\mathbf{Y}_{t,p}$ that is determined by the conditional distribution $F_{X_{t+1}|\mathbf{Y}_{t,p}}(\cdot|\cdot)$ and the marginal distribution $F_{\mathbf{Y}_{t,p}}(\cdot)$. Therefore, capturing $F_{X_{t+1}|\mathbf{Y}_{t,p}}(\cdot|\cdot)$ and $F_{\mathbf{Y}_{t,p}}(\cdot)$ by a bootstrap process may give valid distributional conclusions for general (not necessarily Markov) stationary processes.

Bootstrapping nonparametric estimators in time series analysis has received considerable interest in recent years. Franke, Kreiss, and Mammen (1996) propose some alternative approaches for bootstrapping the (pointwise) distribution of nonparametric estimators under the assumption that the process considered is generated by the nonlinear autoregressive process $X_{t+1} = f(X_t) + \sigma(X_t)\varepsilon_{t+1}$ with $\{\varepsilon_t\}$ an i.i.d. sequence. These approaches include a wild bootstrap and a residual-based bootstrap. Under the assumption that $\{X_t\}$ is a Markov process, Neumann and Kreiss (1998) show validity of the wild bootstrap for local polynomial estimators of the conditional mean $E(X_{t+1}|X_t = x)$ and for supremum type statistics. Compared to these approaches, our Markovian local resampling scheme avoids any kind of preliminary nonparametric estimation of unknown functions to generate the bootstrap pseudoseries. Furthermore, its realm of validity includes but is not restricted to the Markov class.

The paper is organized as follows. In Section 2 the Markovian local resampling scheme is briefly described and the bootstrap approximation to the statistic $L_T(\mathbf{x})$ is given. Section 3 is devoted to the ability of the Markovian local resampling procedure to approximate the distribution of $L_T(\mathbf{x})$. After stating the set of technical assumptions imposed on the class of stochastic process considered, the main result is given. The finite sample performance of the method proposed is illustrated by some simulated examples in Section 4, and all proofs are collected in Section 5.

2. THE MARKOVIAN LOCAL RESAMPLING SCHEME

The Markovian local resampling scheme introduced by Paparoditis and Politis (1997) generates bootstrap replicates by reshuffling the original data points according to a particular probability mechanism. This local resampling scheme can be described by the following three steps.

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

1. Select a resampling width $b = b(T) > 0$ and a set of p starting values $\mathbf{Y}^*_{p,p} = (X^*_p, X^*_{p-1}, \ldots, X^*_1)$. In practice, a simple choice is $\mathbf{Y}^*_{p,p} = \mathbf{Y}_{p,p}$. Select further a resampling kernel W that is a probability density on $\mathbf{R}^p$ satisfying $W > 0$ and $\int \mathbf{u} W(\mathbf{u}) d\mathbf{u} = 0$.
2. For any time point $t + 1 \in \{p + 1, p + 2, \ldots, T\}$ suppose that $\mathbf{Y}^*_{t,p}$ has been already generated. Let J be a discrete random variable taking its values in the set $\mathbf{N}_{p,T-1} = \{p, p + 1, \ldots, T - 1\}$ with probability mass function given by

$$P(J = s) = W_b(\mathbf{Y}^*_{t,p} - \mathbf{Y}_{s,p}) \Big/ \sum_{m \in \mathbf{N}_{p,T-1}} W_b(\mathbf{Y}^*_{t,p} - \mathbf{Y}_{m,p}) \tag{2.3}$$

   for $s \in \mathbf{N}_{p,T-1}$. The bootstrap replicate X^*_{t+1} is then defined by $X^*_{t+1} = X_{J+1}$.
3. If $t + 1 < T$ then go to step 2.

At completion of the algorithm, a new bootstrap pseudoseries $X^*_1, X^*_2, \ldots, X^*_T$ is created with distributional properties that "mimic" those of $F_{X_{t+1}|\mathbf{Y}_{t,p}}$ of the original series. Note that the Markovian local resampling scheme described earlier "works" by assigning at each step of the resampling process the resampling probabilities (2.3) to each of the observed original values $X_{s+1} \in \{X_{p+1}, X_{p+2}, \ldots, X_T\}$. By this algorithm the probability that the bootstrap random variable X^*_{t+1} takes the value X_{s+1} depends on how close $\mathbf{Y}_{s,p}$ is to the already generated segment of bootstrap values $\mathbf{Y}^*_{t,p}$. The closer $\mathbf{Y}_{s,p}$ is to $\mathbf{Y}^*_{t,p}$ the larger the resampling probability (2.3) will be provided the mass of the resampling kernel W is concentrated around its mean value zero. It is easily seen that by the positivity of the resampling kernel $P(X^*_{t+1} = X_{s+1}|\mathbf{Y}^*_{t,p}) > 0$. Furthermore, $\sum_{s=p}^{T-1} P(X^*_{t+1} = X_{s+1}|\mathbf{Y}^*_{t,p}) = 1$. Our procedure provides a 'natural' extension of Efron's (1979) classical i.i.d. bootstrap to the Markovian case. In effect, Efron's (1979) i.i.d. bootstrap can be considered a special case of our local bootstrap if we allow the Markovian order p to be zero, in which case all observations have the same resampling probability, that is, $P(J = s) = 1/(T - p)$.

Given a bootstrap series $X^*_1, X^*_2, \ldots, X^*_T$ generated by the Markovian local bootstrap, we propose to approximate the distribution of the statistic $L_T(\mathbf{x})$ by the distribution of the bootstrap statistic

$$L^*_T(\mathbf{x}) = \sqrt{Th^p}(\hat{m}^*_h(\mathbf{x}; \phi) - m^*_b(\mathbf{x}; \phi)), \tag{2.4}$$

where

$$\hat{m}^*_h(\mathbf{x}; \phi) = \frac{\sum_{j=p}^{T-1} \phi(X^*_{j+1}) K_h(\mathbf{x} - \mathbf{Y}^*_{j,p})}{\sum_{j=p}^{T} K_h(\mathbf{x} - \mathbf{Y}^*_{j,p})} \tag{2.5}$$

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

and

$$m_b^*(\mathbf{x};\phi) = \frac{\sum_{t=p}^{T-1} \phi(X_{t+1}) W_b(\mathbf{x} - \mathbf{Y}_{t,p})}{\sum_{t=p}^{T} W_b(\mathbf{x} - \mathbf{Y}_{t,p})}. \tag{2.6}$$

Note that if $\mathbf{x} \in \mathcal{X}_p$, where $\mathcal{X}_p$ is the p-times Cartesian product of the set of observed values $\mathcal{X} = \{X_1, X_2, \ldots, X_T\}$, then $m_b^*(\mathbf{x};\phi) = E^*(\phi(X_{t+1}^*)|\mathbf{Y}_{t,p}^* = \mathbf{x})$, that is, the conditional expectation of the bootstrap processes $\{X_t^*, t \geq 1\}$ is a kernel estimator of $m(\mathbf{x};\phi)$ with a "smoothing kernel" W and a "smoothing bandwidth" b. In the following discussion we assume throughout that the set of points $\mathbf{x}$ at which the unknown function $m(\mathbf{x};\phi)$ is estimated is a subset of $\mathcal{X}_p$. Note further that in obtaining $\hat{m}_h^*(\mathbf{x};\phi)$ the same formula, that is, the same kernel and the same bandwidth as in (1.1), is used with the only difference that the observations X_t, $t = 1,2,\ldots,T$ are replaced by the bootstrap replicates X_t^*, $t = 1,2,\ldots,T$.

3. VALIDITY OF THE MARKOVIAN LOCAL BOOTSTRAP FOR KERNEL ESTIMATORS

To state the main result of this paper the following set of technical assumptions is needed.

(A1).

(i) $F_{\mathbf{Y}_{t,p}}(\cdot)$ and $F_{X_{t+1}|\mathbf{Y}_{t,p}}(\cdot|\mathbf{x})$ are absolutely continuous with respect to Lebesgue measure on $\mathbf{R}^p$ and $\mathbf{R}$ and have bounded densities $f_{\mathbf{Y}_{t,p}}(\cdot)$ and $f_{X_{t+1}|\mathbf{Y}_{t,p}}(\cdot|\mathbf{x})$, respectively.

(ii) For all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{R}^p$, and $y \in \mathbf{R} \cup \{\infty\}$,

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

where $f_{X_{t+1}\mathbf{Y}_{t,p}}(z,\mathbf{x}) = f_{X_{t+1}|\mathbf{Y}_{t,p}}(z|\mathbf{x}) f_{\mathbf{Y}_{t,p}}(\mathbf{x})$, $\inf_y L(y) > 0$ and $\sup_y L(y) = L < \infty$.

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

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

Assumptions (A1)(ii) and (iii) are smoothness assumptions on the joint and conditional densities and are common in nonparametric estimation problems such as those discussed here (cf. Robinson, 1983; Masry and Tjøstheim, 1995). The assumption of compactness of the support of $f_{X_t}(\cdot)$ in (A1)(iv) is of a rather technical nature. It is imposed to simplify the technical arguments and

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

can be weakened by simultaneously strengthening other requirements (for details, cf. Paparoditis and Politis, 1997).

(A2).

(i) ϕ is a real-valued, measurable function such that $E|\phi(X_t)|^m < \infty$ for some $m > 4$.
(ii) For all $\mathbf{z}_1 = (y_1, \mathbf{x}_1)$, $\mathbf{z}_2 = (y_2, \mathbf{x}_2) \in \mathbf{R}^{p+1}$,

$$|\phi(y_1) f_{X_{t+1}\mathbf{Y}_{t,p}}(y_1, \mathbf{x}_1) - \phi(y_2) f_{X_{t+1}\mathbf{Y}_{t,p}}(y_2, \mathbf{x}_2)| \leq C\|\mathbf{z}_1 - \mathbf{z}_2\|,$$

where $0 < C < \infty$.

The next assumption deals with the dependence properties of $\{X_t\}$. Recall the definition of the strong mixing (α-mixing) coefficient, that is,

$$\alpha(n) = \sup_{A \in \mathcal{B}_1^j,\, B \in \mathcal{B}_{j+n}^{\infty}} |P(A \cap B) - P(A)P(B)|, \qquad n \in \mathbf{N},$$

where $\mathcal{B}_1^j$ and $\mathcal{B}_{j+n}^{\infty}$ denote the σ-algebras generated by the sets of random variables $\{X_1, X_2, \ldots, X_j\}$ and $\{X_{j+n}, X_{j+n+1}, \ldots\}$, respectively. Based on this definition, $\{X_t, t \geq 1\}$ is α-mixing if $\alpha(n) \to 0$ as $n \to \infty$. In the subsequent discussion we make the following assumptions.

(A3). $\{X_t, t \geq 1\}$ is α-mixing, and the mixing coefficient satisfies $\alpha(n) \leq C\rho^n$ for some positive constants C and $\rho \in (0,1)$.

Remark 3.1. Note that (A3) is fulfilled if the process $\{X_t, t \geq 1\}$ is a pth-order strictly stationary and geometrically ergodic Markov process and that this Markov class of stochastic processes contains several of the commonly used parametric and nonparametric models in time series analysis (for several examples, see Tong, 1990, Ch. 3). It is known that if $\{X_t, t \geq 1\}$ belongs to this Markov class then it is geometrically absolutely regular (i.e., β-mixing), which implies geometrically α-mixing (see Doukhan, 1994). However, the class of stochastic processes satisfying (A3) is broader than the Markov class described previously.

The following conditions are imposed on the smoothing and resampling parameters.

(A4).

(i) $K(\cdot)$ is a product kernel; that is, for every $\mathbf{x} = (x_1, x_2, \ldots, x_p) \in \mathbf{R}^p$ we have $K(\mathbf{x}) = \prod_{i=1}^p \kappa(x_i)$ where $\kappa(\cdot)$ is a bounded, Lipschitz continuous, and symmetric probability density the support of which is the interval $[-1,1]$.
(ii) The smoothing bandwidth h satisfies $h \to 0$ and $Th^p \to \infty$ as $T \to \infty$.

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

(A5).

(i) The resampling kernel W is a bounded, Lipschitz continuous, and symmetric probability density on $\mathbf{R}^p$ satisfying $W > 0$, $\int \mathbf{u} W(\mathbf{u}) d\mathbf{u} = 0$ and $\int \|\mathbf{u}\| W(\mathbf{u}) d\mathbf{u} < \infty$.
(ii) The resampling bandwidth b satisfies $b = O(T^{-\delta})$ for some $0 < \delta < 1/(2(p+2))$.

The next assumption is needed to get (asymptotic) expressions for the bias term in estimating $m(\mathbf{x};\phi)$ using the kernel estimator given in (1.1).

(A6).

(i) W is two times continuously differentiable. Furthermore, for $i,j \in \{1,2,\ldots,p\}$ we have $\int u_i^s W(\mathbf{u}) d\mathbf{u} = 0$ for s odd and $\int u_i u_j W(\mathbf{u}) d\mathbf{u} = \delta_{i,j} W_2$ where $d_{i,j}$ is Kronecker's δ and $0 < W_2 < \infty$ for all $i \in \{1,2,\ldots,p\}$.
(ii) $f_{X_{t+1}\mathbf{Y}_{t,p}}(y,\mathbf{x})$ is two times continuously differentiable with respect to $\mathbf{x}$, and the functions $\int \psi(y) f^{(x_{l_1})}_{X_{t+1}\mathbf{Y}_{t,p}}(y,\mathbf{x}) dy$ and $\int \psi(y) f^{(x_{l_1},x_{l_2})}_{X_{t+1}\mathbf{Y}_{t,p}}(y,\mathbf{x}) dy$ are Lipschitz continuous for $\psi = 1$ and $\psi = \phi$ and every $l_1, l_2 \in \{1,2,\ldots,p\}$. Here, $f^{(x_{l_1})}_{X_{t+1}\mathbf{Y}_{t,p}}(y,\mathbf{x})$ and $f^{(x_{l_1},x_{l_2})}_{X_{t+1}\mathbf{Y}_{t,p}}(y,\mathbf{x})$ denote first and second order partial derivatives of $f_{X_{t+1}\mathbf{Y}_{t,p}}(y,\cdot)$ evaluated at $\mathbf{x}$.

Given the preceding set of assumptions, large sample validity of the proposed Markovian local resampling scheme in approximating the distribution of $L_T(\mathbf{x})$ can be established. By "large sample validity" we mean that the law of $L_T^*(\mathbf{x})$ (conditionally on the data) is close to the law of $L_T(\mathbf{x})$ with high probability, if the sample size T is large enough. Our main theorem makes the preceding statement precise; here Kolmogorov's distance $d_0(\mathcal{P},\mathcal{Q}) = \sup_{x\in\mathbf{R}} |\mathcal{P}(X \le x) - \mathcal{Q}(X \le x)|$ between probability measures $\mathcal{P}$ and $\mathcal{Q}$ is employed.

THEOREM 3.1. *Suppose (A1)–(A6) hold and let* $T^{1/2}h^{(p+4)/2} \to C_h \ge 0$. *We then have that, as* $T \to \infty$,

$$d_0\{\mathcal{L}(L_T^*(\mathbf{x})|X_1,X_2,\ldots,X_T), \mathcal{L}(L_T(\mathbf{x}))\} \to 0 \quad \textit{in probability.}$$

Note that if $T^{1/2}h^{(p+4)/2} \to 0$ then the bias term $\sqrt{Th^p}(E\hat{m}_h(\mathbf{x};\phi) - m(\mathbf{x};\phi))$ is asymptotically negligible, whereas for $T^{1/2}h^{(p+4)/2} \to C_h > 0$ this term converges weakly to

$$B(\mathbf{x};\phi) = C_h \int u^2 \kappa(u) du \sum_{l=1}^{p} \left\{ m^{(x_l)}(\mathbf{x};\phi) f^{(x_l)}_{\mathbf{Y}_{t,p}}(\mathbf{x}) + \frac{1}{2} m^{(x_l x_l)}(\mathbf{x};\phi) f_{\mathbf{Y}_{t,p}}(\mathbf{x}) \right\}, \tag{3.7}$$

where $m^{(x_l)}(\mathbf{x};\phi)$, $f^{(x_l)}_{\mathbf{Y}_{t,p}}(\mathbf{x})$ and $m^{(x_l x_l)}(\mathbf{x};\phi)$ denote first and second order partial derivatives of $m(\mathbf{x};\phi)$ and $f_{\mathbf{Y}_{t,p}}(\mathbf{x})$ with respect to x_l (cf. Auestad and Tjøstheim, 1990). The case $T^{1/2}h^{(p+4)/2} \to C_h$ actually corresponds to optimal smoothing (minimizing the mean square error of $\hat{m}_h(\mathbf{x},\phi)$), where the case

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

$T^{1/2}h^{(p+4)/2} \to 0$ corresponds to undersmoothing that is suboptimal. In most cases where resampling is used to estimate the distribution of a kernel smoothed estimator, suboptimal smoothing is used so that the bias term becomes negligible. Quite remarkably, the Markovian local bootstrap procedure also approximates correctly the asymptotic mean (bias term) of the statistic $L_T(\mathbf{x})$.

4. NUMERICAL EXAMPLES

Two simple numerical examples are presented in this section to demonstrate the finite sample performance of the bootstrap method proposed.

We consider realizations of length $T = 500$ from the random coefficient autoregressive model (cf. Tong, 1990, p. 111)

$$X_t = [0.1 + \beta_t]X_{t-1} + \varepsilon_t \tag{4.8}$$

and from the moving average model

$$X_t = \varepsilon_t + 0.7\varepsilon_{t-1}, \tag{4.9}$$

where $\{\varepsilon_t\}$ and $\{\phi_t\}$ are i.i.d. sequences independent from each other and satisfying $\varepsilon_t \sim N(0,1)$ and $\phi_t \sim N(0,0.81)$.

For model (4.8) we are interested in estimating the distribution of the kernel estimator of the conditional variance $v(x) = \mathrm{Var}(X_{t+1}|X_t = x) = 1 + 0.81x^2$. Using a smoothing bandwidth $h = 0.7$ and Epanechnikov's kernel, such an estimator of $v(x)$ is given by $\hat{v}_h(x) = \hat{m}_h(x;\phi_2) - \hat{m}_h^2(x;\phi_1)$. For model (4.9) we are interested in estimating the distribution of the kernel estimator $\hat{m}_h(x)$ of the conditional mean $m(x) = E(X_{t+1}|X_t = x)$. Note that under the Gaussianity assumption the conditional mean of $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$ is linear and given by $m(x) = \{\theta/(1 + \theta^2)\}x$.

The 5%, 50%, and 95% percentage points of the exact distributions of the statistics considered have been estimated using the corresponding percentage points of the bootstrap statistics $\sqrt{Th}(\hat{v}_h^*(x) - v_b^*(x))$ and $\sqrt{Th}(\hat{m}_h^*(x) - m_b^*(x))$ for several values of x in the interval $(-3,3)$. Note that $\hat{v}_h^*(x) = \hat{m}_h^*(x,\phi_2) - [\hat{m}_h^*(x,\phi_1)]^2$, $v_b^*(x) = m_b^*(x;\phi_2) - [m_b^*(x,\phi_1)]^2$, $\hat{m}_h^*(x) = \hat{m}_h^*(x,\phi_1)$, and $m_b^*(x) = m_b^*(x,\phi_1)$. The results of this simulation example are presented in Figure 1 for model (4.8) and in Figure 2 for model (4.9). The bootstrap estimates, shown in these figures by crosses, are mean estimates over 100 independent repetitions of the corresponding models. For each repetition the distribution of the bootstrap statistic has been evaluated using 1,000 bootstrap replications, the Gaussian resampling kernel, and resampling width $b = 1.0$. The percentage points of the exact distribution, shown in these figures by circles, have been estimated using 10,000 replications of each of the models considered.

As these figures show, the bootstrap procedure proposed gives a very satisfactory estimation of the exact percentage points of the distribution of interest over the whole range of x-values considered. We mention here that the accu-

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

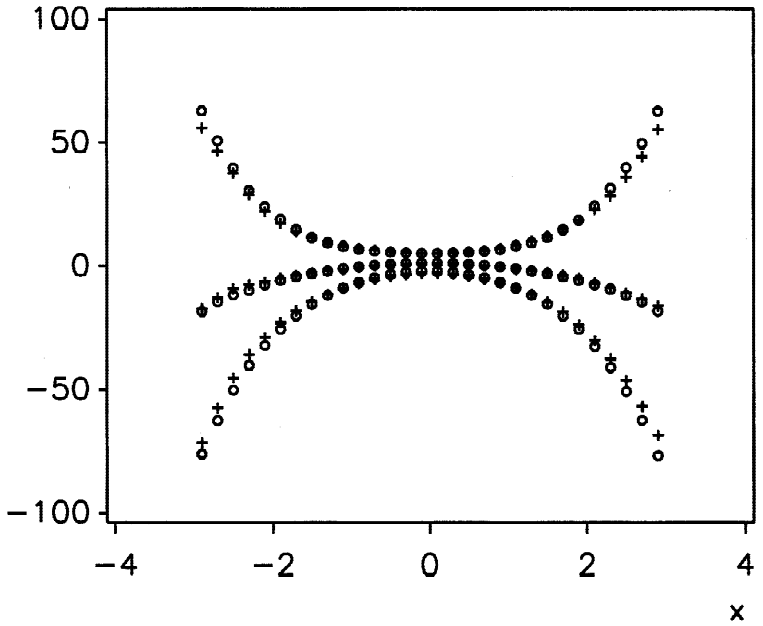


FIGURE 1. Estimated exact and bootstrap estimates of the 5%, 50%, and 95% percentage points of the distribution of $\sqrt{Th}(\hat{v}_h(x) - v(x))$ for model (4.8).

racy of the bootstrap estimator decreases as the value of $|x|$ increases. This can be seen in Table 1, where the estimated standard deviation of the bootstrap estimators is reported for both models considered and for some different values of x.

Table 2 gives (for both models considered) the empirical coverage probabilities of 90% bootstrap confidence intervals and the corresponding mean lengths

TABLE 1. Standard errors of the bootstrap estimates of the 5%, 50%, and 95% percentage points of the distributions considered in Figures 1 and 2

Model (4.8)				Model (4.9)			
X	5%	50%	95%	X	5%	50%	95%
−2.5	13.171	5.245	15.026	−2.4	1.048	0.722	1.110
−1.5	3.050	1.294	3.411	−1.4	0.473	0.296	0.400
0.0	0.545	0.408	0.720	0.0	0.315	0.216	0.287
1.5	3.202	1.372	3.186	1.4	0.406	0.315	0.493
2.5	12.737	5.344	11.028	2.4	1.306	0.801	1.123

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

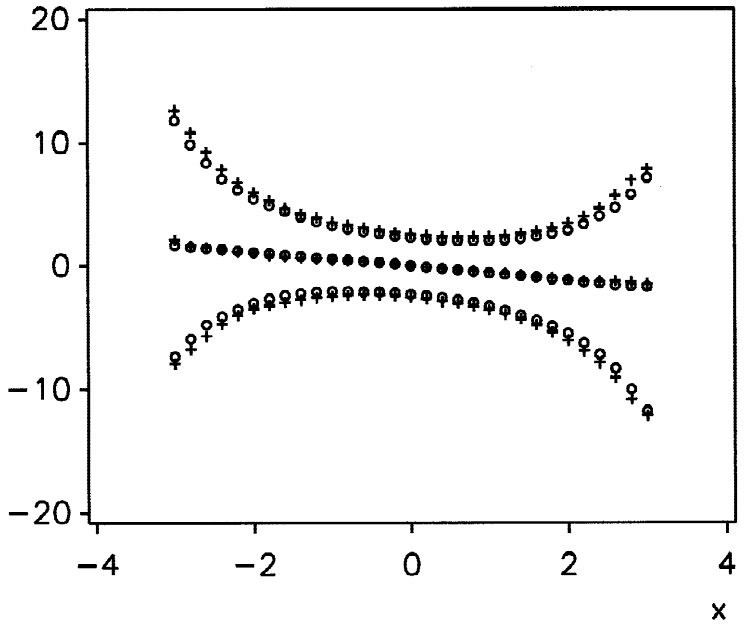


FIGURE 2. Estimated exact and bootstrap estimates of the 5%, 50%, and 95% percentage points of the distribution of $\sqrt{Th}(\hat{m}_h(x) - m(x))$ for model (4.9).

of these intervals based on 100 independent repetitions. For model (4.8) we calculate the bootstrap confidence intervals for the conditional variance $v(x)$ and for model (4.9) the bootstrap confidence intervals for the conditional mean $m(x)$ using the values $x = -1.5$, $x = 0$, and $x = 1.5$. Also included in this table is a comparison of the Markovian local bootstrap with the block bootstrap method (cf. Künsch, 1989; Liu and Singh, 1992). Note that a $(1 - \alpha)100\%$ bootstrap confidence interval for the conditional variance $v(x)$ is defined here by $[\hat{v}_h(x) - \hat{q}_{1-\alpha/2}, \hat{v}_h(x) + \hat{q}_{\alpha/2}]$ where $\hat{q}_p$ denotes the pth quantile of the bootstrap estimate of the distribution of $\hat{v}_h(x) - v(x)$. For the Markovian local resampling scheme this distribution is estimated using the distribution of $\hat{v}_h^*(x) - v_b^*(x)$ and for the block bootstrap method using the distribution of $\hat{v}_h^+(x) - \hat{v}_h(x)$. Here $\hat{v}_h^+(x)$ is the kernel estimator of $v(x)$ based on the replicate $X_1^+, X_2^+, \ldots, X_T^+$ obtained using the block bootstrap method. The bootstrap confidence intervals for the conditional mean $m(x)$ have been obtained in an analogous way, that is, by replacing in the preceding formulas $v(x)$, $\hat{v}_h(x)$, $\hat{v}_h^*(x)$, $v_b^*(x)$, and $\hat{v}_h^+(x)$ by $m(x)$, $\hat{m}_h(x)$, $\hat{m}_h^*(x)$, $m_b^*(x)$, and $\hat{m}_h^+(x)$, respectively, where $\hat{m}_h^+(x)$ is the kernel estimate of the conditional mean based on the block bootstrap series. Finally, and to see the effects of varying the resampling parameters, both methods have been applied using several values of the resampling width b and of the block size B.

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

TABLE 2. Empirical coverage probabilities (C.P.) and mean lengths of 90% bootstrap confidence intervals for the conditional variance of model (4.8) and for the conditional mean of model (4.9)

	$x = -1.5$		$x = 0.0$		$x = 1.5$	
	C.P.	Length	C.P.	Length	C.P.	Length
			Model (4.8)			
Block bootstrap						
$B = 10$	0.86	2.131	0.83	1.475	0.83	2.115
$B = 15$	0.85	2.028	0.96	1.296	0.81	2.049
$B = 20$	0.79	1.888	0.88	1.309	0.79	2.013
$B = 25$	0.79	1.567	0.93	0.771	0.79	1.530
Markov local bootstrap						
$b = 0.7$	0.82	1.435	0.80	0.408	0.87	1.412
$b = 1.0$	0.86	1.424	0.92	0.447	0.84	1.429
$b = 1.2$	0.84	1.419	0.95	0.480	0.84	1.404
$b = 1.5$	0.87	1.457	0.95	0.518	0.85	1.444
			Model (4.9)			
Block bootstrap						
$B = 10$	0.76	0.352	0.85	0.244	0.86	0.345
$B = 15$	0.82	0.349	0.89	0.248	0.89	0.352
$B = 20$	0.85	0.349	0.91	0.240	0.91	0.340
$B = 25$	0.93	0.340	0.89	0.239	0.85	0.345
Markov local bootstrap						
$b = 0.7$	0.88	0.387	0.90	0.271	0.88	0.384
$b = 1.0$	0.87	0.386	0.92	0.272	0.92	0.389
$b = 1.2$	0.93	0.393	0.93	0.271	0.94	0.388
$b = 1.5$	0.93	0.395	0.93	0.274	0.94	0.388

As this table shows the local bootstrap procedure compares favorably with the block bootstrap procedure regarding the coverage probabilities of the confidence intervals for both classes of models considered. In fact, both methods seem to behave very similarly taking into account the standard error $\sqrt{\hat{p}(1-\hat{p})/100}$ of the estimated coverage probability $\hat{p}$. Apart from this overall behavior, it seems that in the case of the moving average model (4.9) the block bootstrap method is slightly more efficient than the local resampling method in terms of the average length of the bootstrap confidence intervals obtained. On the other hand, and in terms of the same quantity, the local bootstrap method is quite a bit more efficient than the block bootstrap method in the case of the random coefficient autoregressive process (4.8); this is not surprising because model (4.8) is a true first order Markov process.

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

5. DERIVATIONS

To prove Theorem 3.1 we make use of some lemmas that are established in what follows. We begin with two lemmas concerning some asymptotic properties of the bootstrap series $\{X_t^*\}$ and that are reproduced from and proved in Paparoditis and Politis (1997). To state these lemmas we fix some additional notation. For $\mathbf{x} = (x_1, x_2, \ldots, x_p) \in \mathbf{R}^p$ let $F_T^*(\mathbf{x}) = P(X_t^* \leq x_1, X_{t-1}^* \leq x_2, \ldots, X_{t-p+1}^* \leq x_p)$ be the stationary distribution function of the bootstrap Markov chain and $F(\mathbf{x}) = P(X_t \leq x_1, X_{t-1} \leq x_2, \ldots, X_{t-p+1} \leq x_p)$ the stationary distribution function of the original process $\{X_t, t \geq 1\}$; that is, we set $F(\cdot) = F_{\mathbf{Y}_{t,p}}(\cdot)$.

LEMMA 5.1. *Suppose that Assumptions (A1)–(A3) and (A5) hold. Then conditionally on $X_1, X_2, \ldots, X_T$ we have that, as $T \to \infty$,*

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

LEMMA 5.2. *Under the same assumptions as in Lemma 5.1 and conditionally on $X_1, X_2, \ldots, X_T$, $T_0 \in \mathbf{N}$ exists such that for all $T \geq T_0$ the bootstrap series $\{X_t^*, t \geq 1\}$ is strongly mixing where the strong mixing coefficient α^* satisfies $\alpha^*(n) \leq C\varrho^n$ for some constants $C > 0$ and $\varrho \in (0,1)$ that are independent of T.*

LEMMA 5.3. *Suppose that Assumptions (A1)–(A5) hold and let ϕ be a real-valued, continuous function, y a fixed real number, and s a positive integer. We then have that, as $T \to \infty$,*

$$\sup_{\mathbf{x} \in \mathbf{S}} \left| \frac{h^{s-1} \sum_{l=p}^{T-1} \phi(X_{l+1}) \kappa_h^s(y - X_{l+1}) W_b(\mathbf{x} - \mathbf{Y}_{l,p})}{\sum_{n=p}^{T-1} W_b(\mathbf{x} - \mathbf{Y}_{n,p})} - \phi(y) f_{X_{t+1}|\mathbf{Y}_{t,p}}(y|\mathbf{x}) \int \kappa^s(u) du \right| \to 0$$

in probability.

Proof. Let $N(\mathbf{x}; \phi) = \phi(y) f_{X_{t+1}\mathbf{Y}_{t,p}}(y, \mathbf{x}) \int \kappa^s(u) du$,

$$N_T(\mathbf{x}; \phi) = \frac{h^{s-1}}{T} \sum_{l=p}^{T-1} \phi(X_{l+1}) \kappa_h^s(y - X_{l+1}) W_b(\mathbf{x} - \mathbf{Y}_{l,p}),$$

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

$\hat{f}_b(\mathbf{x}) = T^{-1}\sum_{n=p}^{T-1} W_b(\mathbf{x} - \mathbf{Y}_{n,p})$, and $f(\mathbf{x}) = f_{\mathbf{Y}_{t,p}}(\mathbf{x})$. Note that

$$\left| \frac{h^{s-1}\sum_{l=p}^{T-1} \phi(X_{l+1})\kappa_h^s(y - X_{l+1})W_b(\mathbf{x} - \mathbf{Y}_{l,p})}{\sum_{n=p}^{T-1} W_b(\mathbf{x} - \mathbf{Y}_{n,p})} - \phi(y) f_{X_{t+1}|\mathbf{Y}_{t,p}}(y|\mathbf{x}) \int \kappa^s(u)du \right|$$

$$\leq \frac{1}{\hat{f}_b(\mathbf{x})} \left\{ |N_T(\mathbf{x};\phi) - N(\mathbf{x};\phi)| + \frac{|N(\mathbf{x};\phi)|}{f(\mathbf{x})} |\hat{f}_b(\mathbf{x}) - f(\mathbf{x})| \right\}. \tag{5.10}$$

In the following discussion we show that $\sup_{\mathbf{x}\in\mathbf{S}}|N_T(\mathbf{x};\phi) - N(\mathbf{x};\phi)| \to 0$ in probability. Because $\sup_{\mathbf{x}\in\mathbf{S}}|\hat{f}_b(\mathbf{x}) - f(\mathbf{x})| \to 0$ follows by the same arguments, the lemma is established using (5.10) and (A1)(iv).

Consider $|N_T(\mathbf{x};\phi) - N(\mathbf{x};\phi)|$. For this term we have

$$\begin{aligned}
&|N_T(\mathbf{x};\phi) - N(\mathbf{x};\phi)| \\
&\quad \leq \left| h^{s-1} \int \phi(z_1)\kappa_h^s(y - z_1)W_b(\mathbf{x} - \mathbf{z}_2) f_{X_{t+1}\mathbf{Y}_{t,p}}(z_1, \mathbf{z}_2)dz_1\, d\mathbf{z}_2 \right. \\
&\qquad \left. - \phi(y) f_{X_{t+1}\mathbf{Y}_{t,p}}(y, \mathbf{x}) \int \kappa^s(u)du \right| \\
&\qquad + h^{s-1} \left| \frac{1}{T}\sum_{l=p}^{T-1} \phi(X_{l+1})\kappa_h^s(y - X_{l+1})W_b(\mathbf{x} - \mathbf{Y}_{l,p}) \right. \\
&\qquad\qquad \left. - \int \phi(z_1)\kappa_h^s(y - z_1)W_b(\mathbf{x} - \mathbf{z}_2) f_{X_{t+1}\mathbf{Y}_{t,p}}(z_1, \mathbf{z}_2)dz_1\, d\mathbf{z}_2 \right| \\
&\quad = T_1(\mathbf{x}) + T_2(\mathbf{x})
\end{aligned}$$

with an obvious notation for $T_1(\mathbf{x})$ and $T_2(\mathbf{x})$.

Consider first $T_1(\mathbf{x})$. Using the substitutions $z_1 = y - u_1 h$ and $\mathbf{z}_2 = \mathbf{x} - \mathbf{u}_2 b$ we have

$$\begin{aligned}
T_1(\mathbf{x}) &= \left| \int (\phi(y - u_1 h) f_{X_{t+1},\mathbf{Y}_{t,p}}(y - u_1 h, \mathbf{x} - \mathbf{u}_2 b) \right. \\
&\qquad \left. - \phi(y) f_{X_{t+1},\mathbf{Y}_{t,p}}(y, \mathbf{x}))\kappa^s(u_1)W(\mathbf{u}_2)du_1\, d\mathbf{u}_2 \right| \\
&\leq \int \left| \phi(y - u_1 h) f_{X_{t+1},\mathbf{Y}_{t,p}}(y - u_1 h, \mathbf{x} - \mathbf{u}_2 b) - \phi(y) f_{X_{t+1},\mathbf{Y}_{t,p}}(y, \mathbf{x}) \right| \\
&\qquad \times \kappa^s(u_1)W(\mathbf{u}_2)du_1\, d\mathbf{u}_2 \\
&= O(\max\{h, b\})
\end{aligned}$$

uniformly in $\mathbf{x}$ by (A2)(ii).

Consider next $T_2(\mathbf{x})$ and observe that because ϕ is continuous and K is bounded and compactly supported, $\phi(X_{l+1})\kappa^s((y - X_{l+1})/h)$ is bounded. Fur-

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

thermore, instead of $h^{s-1}T^{-1}\sum_{l=p}^{T-1}\phi(X_{l+1})\kappa_h^s(y-X_{l+1})W_b(\mathbf{x}-\mathbf{Y}_{l,p})$ we consider in the following discussion the asymptotically equivalent statistic $R_T(y,\mathbf{x}) = h^{s-1}(T-p)^{-1}\sum_{l=p}^{T-1}\phi(X_{l+1})\kappa_h^s(y-X_{l+1})W_b(\mathbf{x}-\mathbf{Y}_{l,p})$. Observe that

$$E(R_T(y,\mathbf{x})) = h^{s-1}\int \phi(z_1)\kappa_h^s(y-z_1)W_b(\mathbf{x}-\mathbf{z}_2)f_{X_{t+1}\mathbf{Y}_{t,p}}(z_1,\mathbf{z}_2)dz_1\,d\mathbf{z}_2$$

and that

$$\sup_{\mathbf{x}\in\mathbf{S}} T_2(\mathbf{x}) \le \sup_{\mathbf{x}\in\mathbf{S}}|R_T(y,\mathbf{x}) - E(R_T(y,\mathbf{x}))| + o_P(1).$$

Thus we have to show that $\sup_{\mathbf{x}\in\mathbf{S}}|R_T(y,\mathbf{x}) - E(R_T(y,\mathbf{x}))| \to 0$ in probability. To deal with this term divide $\mathbf{S}$ in a number of N_T cubes denoted by $\mathbf{I}_{i,T}$ with centers $\mathbf{x}_i$ and length L_T. We then get

$$\begin{aligned}
&\sup_{\mathbf{x}\in\mathbf{S}}|R_T(y,\mathbf{x}) - E(R_T(y,\mathbf{x}))| \\
&\quad\le \max_{1\le i\le N_T}\sup_{\mathbf{x}\in\mathbf{S}\cap\mathbf{I}_{i,T}}|R_T(y,\mathbf{x}) - R_T(y,\mathbf{x}_i)| \\
&\qquad + \max_{1\le i\le N_T}|R_T(y,\mathbf{x}_i) - E(R_T(y,\mathbf{x}_i))| \\
&\qquad + \max_{1\le i\le N_T}\sup_{\mathbf{x}\in\mathbf{S}\cap\mathbf{I}_{i,T}}|E(R_T(y,\mathbf{x}_i)) - E(R_T(y,\mathbf{x}))| \\
&\quad = M_{1,T} + M_{2,T} + M_{3,T}.
\end{aligned}$$

By the Lipschitz continuity of K it is easily seen that $M_{1,T} = O_P(L_T b^{-(p+1)})$ and $M_{3,T} = O_P(L_T b^{-(p+1)})$. Furthermore,

$$\begin{aligned}
&P\left(\max_{1\le i\le N_T}|R_T(y,\mathbf{x}_i) - E(R_T(y,\mathbf{x}_i))| > L_T b^{-(p+1)}\right) \\
&\quad\le \sum_{i=1}^{N_T} P\left(\frac{1}{(T-p)hb^p}\left|\sum_{l=p}^{T-1}\omega_{l,T}(y,\mathbf{x}_i)\right| > L_T b^{-(p+1)}\right),
\end{aligned}$$

where

$$\begin{aligned}
\omega_{l,T}(y,\mathbf{x}_i) &= \phi(X_{l+1})\kappa^s\left(\frac{y-X_{l+1}}{h}\right)W\left(\frac{\mathbf{x}_i-\mathbf{Y}_{l,p}}{b}\right) \\
&\quad - E\left[\phi(X_{l+1})\kappa^s\left(\frac{y-X_{l+1}}{h}\right)W\left(\frac{\mathbf{x}_i-\mathbf{Y}_{l,p}}{b}\right)\right]
\end{aligned}$$

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

and $E(\omega_{l,T}^2(y,\mathbf{x})) \leq \sigma^2 = C^2 h b^p$ with $C > 0$ a generic constant independent of T. To bound this probability let first $\tilde{\omega}_{l,T}(y,\mathbf{x}_i) = C^{-1}\omega_{l,T}(\mathbf{x}_i)$ and

$$L_T = \frac{(\log(T-p))^3\sqrt{b^{2+p+\epsilon}}}{\sqrt{T-p}\sqrt{h^{1-\epsilon}}}$$

for some $0 < \epsilon < 1$. Applying Proposition 1 of Doukhan (1994, p. 33), with $\sigma = \sqrt{hb^p}$ and $\log\sigma^{-1} = \log\sqrt{hb^p}^{-1}$ and using $h/h^{1-\epsilon} > 1$ and $b^{2+p+\epsilon}/b^{2+p} > 1$, we then have that constants $\alpha > 0$ and $\beta > 0$ exist such that

$$\begin{aligned}
&\sum_{i=1}^{N_T} P\left(\frac{1}{(T-p)hb^p}\left|\sum_{l=p}^{T-1}\tilde{\omega}_{l,T}(y,\mathbf{x}_i)\right| > \frac{1}{C}L_T b^{-(p+1)}\right)\\
&\quad\leq O(N_T)P\left(\left|\sum_{l=p}^{T-1}\tilde{\omega}_{l,T}(y,\mathbf{x}_i)\right| > \frac{(\log(T-p))^3}{\log\sqrt{hb^p}^{-1}}\sqrt{T-p}\sqrt{hb^p}\log\sqrt{hb^p}^{-1}\right)\\
&\quad\leq O(N_T)P\left(\left|\sum_{l=p}^{T-1}\tilde{\omega}_{l,T}(y,\mathbf{x}_i)\right| > \frac{1}{C}(\log(T-p))^2\sqrt{T-p}\sqrt{hb^p}\log\sqrt{hb^p}^{-1}\right)\\
&\quad\leq O(N_T)\alpha\exp\{-\beta\log(T-p)/\sqrt{C}\}\\
&\quad= o(1). \qquad \textbf{(5.11)}
\end{aligned}$$

Thus for $b = O(T^{-\delta})$ and $\delta < 1/p$ we get that $L_T/b^{(p+1)} \to 0$ and $L_T \to 0$ as $T \to \infty$. ■

The following result deals with a well known property of the kernel estimator $\hat{m}_h(\mathbf{x};\phi)$. Its proof is omitted because it closely follows that of Lemma 5.3.

LEMMA 5.4. *Suppose that Assumptions (A1)–(A4) hold and let φ be a real-valued continuous function. Then as $T \to \infty$*

$$\sup_{\mathbf{x}\in\mathbf{S}}|\hat{m}_h(\mathbf{x};\varphi) - m(\mathbf{x};\varphi)| \to 0$$

in probability.

Let $f(\cdot|\mathbf{x})$ denote the conditional density $f_{X_{t+1}|\mathbf{Y}_{t,p}}(\cdot|\mathbf{x})$ and let $\hat{f}(y|\mathbf{x})$ be the kernel estimator of $f(y|\mathbf{x})$ given by

$$\hat{f}(y|\mathbf{x}) = \frac{\sum_{l=p}^{T-1}\kappa_h(y - X_{l+1})W_b(\mathbf{x} - \mathbf{Y}_{l,p})}{\sum_{n=p}^{T-1}W_b(\mathbf{x} - \mathbf{Y}_{n,p})}, \qquad \textbf{(5.12)}$$

where $y \in S$. Note that in the notation of Lemma 5.3 and if we set $\phi(\cdot) \equiv 1$ and $s = 1$ we immediately get that $\sup_{\mathbf{x}\in\mathbf{S}}|\hat{f}(y|\mathbf{x}) - f(y|\mathbf{x})| \to 0$ in probability as $T \to \infty$.

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

To simplify calculations and because the results obtained are asymptotic for $T \to \infty$, we assume in the following that $\mathbf{Y}^*_{p,p} \sim F^*_T$, that is, we assume that the bootstrap series starts with its stationary distribution. Now, for $k \in \mathbf{N}$ let $\mathbf{x}_{j_k} = (x_{j_k}, x_{j_{k+1}}, \ldots, x_{j_{k+p-1}}) \in \mathcal{X}_p$ and recall that $\{X^*_t, t \geq 1\}$ is a discrete Markov chain with state space $\mathcal{X}_p$. Using iteratively the relation

$$P(\mathbf{Y}^*_{t,p} = \mathbf{x}_{j_1}) = \sum_{j_{p+1}=p+1}^{T} P(X^*_t = x_{j_1} | \mathbf{Y}^*_{t-1,p} = \mathbf{x}_{j_2}) P(\mathbf{Y}^*_{t-1,p} = \mathbf{x}_{j_2})$$

and the transition probability (2.3) we get the following useful expression for the stationary probability mass function of $\mathbf{Y}^*_{t,p}$.

$$\begin{aligned}
P(\mathbf{Y}^*_{t,p} = \mathbf{x}_{j_1}) &= \sum_{j_{p+1}} \sum_{j_{p+2}} \cdots \sum_{j_{2p}} P(X^*_t = x_{j_1} | X^*_{t-1} = x_{j_2}, X^*_{t-2} = x_{j_3}, \ldots, X^*_{t-p} = x_{j_{p+1}}) \\
&\quad \times P(X^*_{t-1} = x_{j_2} | X^*_{t-2} = x_{j_3}, X^*_{t-3} = x_{j_4}, \ldots, X^*_{t-p-1} = x_{j_{p+2}}) \\
&\quad \times \cdots \times P(X^*_{t-p+1} = x_{j_p} | X^*_{t-p} = x_{j_{p+1}}, \\
&\qquad\qquad X^*_{t-p-1} = x_{j_{p+2}}, \ldots, X^*_{t-2p+1} = x_{j_{2p}}) \\
&\quad \times P(X^*_{t-p} = x_{j_{p+1}}, X^*_{t-p-1} = x_{j_{p+2}}, \ldots, X^*_{t-2p+1} = x_{j_{2p}}) \\
&= \int \prod_{i=1}^{p} P(X^*_{t-i+1} = x_{j_i} | \mathbf{Y}^*_{t-i,p} = \mathbf{x}_{j_{i+1}}) \, dF^*_T(\mathbf{x}_{j_{p+1}}) \\
&= \int \prod_{i=1}^{p} \frac{W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{j_i - 1, p})}{\sum_{n=p}^{T-1} W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{n,p})} \, dF^*_T(\mathbf{x}_{j_{p+1}}). \qquad \textbf{(5.13)}
\end{aligned}$$

To proceed with the proof of Theorem 3.1 we first note that $L^*_T(\mathbf{x})$ can be expressed as

$$\begin{aligned}
L^*_T(\mathbf{x}) &= \frac{1}{\hat{f}^*_h(\mathbf{x})} \sqrt{\frac{h^p}{T}} \sum_{j=p}^{T-1} (\phi(X^*_{j+1}) - E^*[\phi(X^*_{j+1}) | Y^*_{j,p}]) K_h(\mathbf{x} - \mathbf{Y}^*_{j,p}) \\
&\quad + \frac{1}{\hat{f}^*_h(\mathbf{x})} \sqrt{\frac{h^p}{T}} \sum_{j=p}^{T-1} (E^*[\phi(X^*_{j+1}) | Y^*_{j,p}] - m^*(\mathbf{x}; \phi)) K_h(\mathbf{x} - \mathbf{Y}^*_{j,p}) \\
&= \frac{1}{\hat{f}^*_h(\mathbf{x})} \{L^*_{1,T}(\mathbf{x}) + L^*_{2,T}(\mathbf{x})\},
\end{aligned}$$

where $\hat{f}^*_h(\mathbf{x}) = T^{-1} \sum_{j=p}^{T-1} K_h(\mathbf{x} - \mathbf{Y}^*_{j,p})$ and an obvious notation for $L^*_{1,T}(\mathbf{x})$ and $L^*_{2,T}(\mathbf{x})$. Using the preceding expression, the assertion of the theorem is established if we show that $\hat{f}^*_h(\mathbf{x}) \to f(\mathbf{x})$ weakly, that $L^*_{1,T}(\mathbf{x})$ converges to a Gaussian distribution with mean zero and the appropriate variance, and that $L^*_{2,T}(\mathbf{x})/\hat{f}^*_h(\mathbf{x})$ is a consistent estimator of the mean (bias term) of $L_T(\mathbf{x})$. This is established in the following three lemmas.

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

LEMMA 5.5. *Under Assumptions (A1)–(A5) we have conditionally on* $X_1, X_2, \ldots, X_T$ *that for all* $\mathbf{x} \in \mathbf{S}$,

$\hat{f}_h^*(\mathbf{x}) \to f(\mathbf{x})$ *in probability.*

Proof. Let $\mathbf{x} = (x_1, x_2, \ldots, x_p)$. Using (5.13) we have

$$
\begin{aligned}
&E^*(\hat{f}_h^*(\mathbf{x})) \\
&= \frac{T-p}{T} \sum_{j_1, j_2, \ldots, j_p} K_h(\mathbf{x} - \mathbf{x}_{j_1}) P(\mathbf{Y}_{t,p}^* = \mathbf{x}_{j_1}) \\
&= \frac{T-p}{T} \int \sum_{j_1, j_2, \ldots, j_p} \prod_{m=1}^{p} \prod_{i=1}^{p} \frac{\kappa_h(x_m - x_{j_m}) W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{j_i - 1, p})}{\sum_{n=p}^{T-1} W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{n,p})} \, dF_T^*(\mathbf{x}_{j_{p+1}}) \\
&= \frac{T-p}{T} \int \sum_{j_2, \ldots, j_p} \prod_{m=2}^{p} \prod_{i=2}^{p} \hat{f}(x_1 | \mathbf{x}_{j_2}) \frac{\kappa_h(x_m - x_{j_m}) W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{j_i - 1, p})}{\sum_{n=p}^{T-1} W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{n,p})} \, dF_T^*(\mathbf{x}_{j_{p+1}}) \\
&= \frac{T-p}{T} \int \sum_{j_2, \ldots, j_p} \prod_{m=2}^{p} \prod_{i=2}^{p} f(x_1 | \mathbf{x}_{j_2}) \frac{\kappa_h(x_m - x_{j_m}) W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{j_i - 1, p})}{\sum_{n=p}^{T-1} W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{n,p})} \, dF_T^*(\mathbf{x}_{j_{p+1}}) \\
&\quad + o_P(1) \\
&= \frac{T-p}{T} \int \sum_{j_3, \ldots, j_p} \prod_{m=3}^{p} \prod_{i=3}^{p} \frac{\sum_{j_2} f(x_1 | \mathbf{x}_{j_2}) \kappa_h(x_2 - x_{j_2}) W_b(\mathbf{x}_{j_3} - \mathbf{Y}_{j_2 - 1, p})}{\sum_{n=p}^{T-1} W_b(\mathbf{x}_{j_3} - \mathbf{Y}_{n,p})} \\
&\quad \times \frac{\kappa_h(x_m - x_{j_m}) W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{j_i - 1, p})}{\sum_{n=p}^{T-1} W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{n,p})} \, dF_T^*(\mathbf{x}_{j_{p+1}}) + o_P(1) \\
&= \frac{T-p}{T} \int \sum_{j_3, \ldots, j_p} \prod_{m=3}^{p} \prod_{i=3}^{p} f(x_1 | x_2, x_{j_3}, \ldots, x_{j_{p+1}}) f(x_2 | \mathbf{x}_{j_3}) \\
&\quad \times \frac{\kappa_h(x_m - x_{j_m}) W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{j_i - 1, p})}{\sum_{n=p}^{T-1} W_b(\mathbf{x}_{j_{i+1}} - \mathbf{Y}_{n,p})} \, dF_T^*(\mathbf{x}_{j_{p+1}}) + o_P(1),
\end{aligned}
$$

where the preceding $o_P(1)$ terms are due to the uniform (over $\mathbf{x}$) convergence of $\hat{f}(\cdot|\mathbf{x})$ given in (5.12) to $f(\cdot|\mathbf{x})$ and Lemma 5.3. Continuing in this way we end up with the expression

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

$$E^*(\hat{f}_h^*(\mathbf{x})) = \frac{T-p}{T}\int f(x_1|x_2,\ldots,x_p,x_{j_{p+1}})f(x_2|x_3,\ldots,x_p,x_{j_{p+1}},x_{j_{p+2}})$$
$$\times\cdots\times f(x_p|x_{j_{p+1}},\ldots,x_{j_{2p}})\,dF_T^*(\mathbf{x}_{j_{p+1}}) + o_P(1)$$
$$\to \int f(x_1|x_2,\ldots,x_p,x_{j_{p+1}})f(x_2|x_3,\ldots,x_p,x_{j_{p+1}},x_{j_{p+2}})$$
$$\times\cdots\times f(x_p|x_{j_{p+1}},\ldots,x_{j_{2p}})\,dF(\mathbf{x}_{j_{p+1}})$$
$$= f(\mathbf{x}), \tag{5.14}$$

where the last convergence is due to Lemma 5.1 and Helly–Bray's theorem.

To evaluate the variance of $\hat{f}_h^*(\mathbf{x})$, let $Z_{h,j}^* = K_h(\mathbf{x} - \mathbf{Y}_{j,p}^*)$ and note that

$$\mathrm{Var}^*[\,\hat{f}_h^*(\mathbf{x})] = (T-p)T^{-2}\mathrm{Var}^*(Z_{h,1}^*)$$
$$+ 2T^{-1}\sum_{\tau=1}^{T-p-1}(1-\tau/T)\mathrm{Cov}^*(Z_{h,1}^*, Z_{h,1+\tau}^*). \tag{5.15}$$

Now, $h^p\,\mathrm{Var}^*[Z_{h,j}^*] \le h^p\sum_{j_1,j_2,\ldots,j_p}K_h^2(\mathbf{x}-\mathbf{x}_{j_1})P(\mathbf{Y}_{t,p}^* = \mathbf{x}_{j_1})$ and by the same arguments as those leading to (5.14) we get $h^p\mathrm{Var}^*[Z_{h,j}^*] \to f(\mathbf{x})\int K^2(\mathbf{u})d\mathbf{u}$, that is, $T^{-1}\,\mathrm{Var}^*[Z_{h,j}^*] = O_P(T^{-1}h^{-p})$. To handle the second term on the right hand side of (5.15) note first that by Davydov's lemma (Hall and Heyde, 1980, App.) we have

$$|\mathrm{Cov}(Z_{h,1}^*, Z_{h,1+\tau}^*)| \le 8\alpha^*(\tau)^{1-2/\nu}(E^*|Z_{h,1}^*|^\nu)^{2/\nu}$$

for some real $\nu > 2$. Because $E^*|Z_{h,1}^*|^\nu = \sum_{j_1,j_2,\ldots,j_p}K_h^\nu(\mathbf{x}-\mathbf{x}_{j_1})P(\mathbf{Y}_{t,p}^* = \mathbf{x}_{j_1}) = O_P(h^{(1-\nu)p})$ by the same arguments as in establishing that (5.14) is true, we get by Lemma 5.2 and arguments similar to those in Masry and Tjøstheim (1995, pp. 271–272) that $|T^{-1}\sum_{\tau=1}^{T-p-1}(1-\tau/T)\mathrm{Cov}^*(Z_{h,1}^*, Z_{h,1+\tau}^*)| = o_P(T^{-1}h^{-p})$. ■

LEMMA 5.6. *Under Assumptions (A1)–(A5) we have conditionally on* $X_1, X_2, \ldots, X_T$ *that, as* $T \to \infty$,

$$\sqrt{\frac{h^p}{T}}\sum_{j=p}^{T-1}(\phi(X_{j+1}^*) - m_b^*(\mathbf{Y}_{j,p}^*;\phi))K_h(x-\mathbf{Y}_{j,p}^*) \Rightarrow N(0,\tau^2(\mathbf{x};\phi)) \quad \textit{in probability,}$$

where $\tau^2(\mathbf{x};\phi) = \int K^2(\mathbf{u})d\mathbf{u}\, f(\mathbf{x})\,\mathrm{Var}[\phi(X_{t+1})|\mathbf{Y}_{t,p} = \mathbf{x}]$.

Proof. Define

$$D_{j+1}^* \equiv \sqrt{\frac{h^p}{T}}(\phi(X_{j+1}^*) - m_b^*(\mathbf{Y}_{j,p}^*;\phi))K_h(\mathbf{x}-\mathbf{Y}_j^*)$$

and denote by $\mathcal{F}_j^* = \sigma(X_1^*, X_2^*, \ldots, X_j^*)$ the σ-algebra generated by $\{X_1^*, X_2^*, \ldots, X_j^*\}$. Because for $\mathbf{x} \in \mathcal{X}_p$, $m_b^*(\mathbf{x};\phi) = E^*[\phi(X_{t+1}^*)|\mathbf{Y}_{t,p}^* = \mathbf{x}]$ we have

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

by the Markov property that $\{D_{j+1}^*, \mathcal{F}_j^*\}_{j=p,\,p+1,\ldots}$ forms a martingale difference scheme. Thus to establish the desired result, we apply Theorem 1 of Brown (1971). By this theorem we have to show that

$$\sum_{j=p}^{T-1} E^*[D_{j+1}^{*2}|\mathcal{F}_j^*] \to \int K^2(\mathbf{u})d\mathbf{u}\, f(\mathbf{x})\mathrm{Var}[\phi(X_{t+1})|\mathbf{Y}_{t,p} = \mathbf{x}] \quad \text{in probability} \tag{5.16}$$

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

$$\sum_{j=p}^{T-1} E^*[D_{j+1}^{*2} I(D_{j+1}^{*2} \geq \delta)|\mathcal{F}_j^*] \to 0 \quad \text{in probability} \tag{5.17}$$

is satisfied. To simplify notation and to stress the essentials, we focus in what follows on the case $p = 1$. The case of general p can be handled along the same lines.

Consider first (5.16). Because

$$E^*[D_{j+1}^{*2}|\mathcal{F}_j^*] = \frac{h}{T}\sum_{l=2}^{T}(\phi(X_l) - m_b^*(X_j^*;\phi))^2 K_h^2(x - X_j^*)P(X_{j+1}^* = X_l|X_j^*)$$

we have

$$\sum_{j=1}^{T-1} E^*[D_{j+1}^{*2}|\,\mathcal{F}_j^*] = \frac{1}{Th}\sum_{j=1}^{T} K^2\left(\frac{x - X_j^*}{h}\right)\hat{v}_b^*(X_j^*;\phi),$$

where

$$\hat{v}_b^*(z;\phi) = \frac{\sum_{l=2}^{T}(\phi(X_l) - m_b^*(z;\phi))^2 W_b(z - X_{l-1})}{\sum_{s=1}^{T-1} W_b(z - X_s)}.$$

Let $v(z;\phi) = E[(\phi(X_{t+1}) - m(z;\phi))^2|X_t = z]$ and note that as in the proof of Lemma 5.5, $(Th)^{-1}\sum_{j=1}^{T} K^2((x - X_j^*)/h) = O_P(1)$. Then

$$\sum_{j=1}^{T} E^*[D_{j+1}^{*2}|\mathcal{F}_j^*] = \frac{1}{Th}\sum_{j=1}^{T} K^2\left(\frac{x - X_j^*}{h}\right)v(X_j^*;\phi)$$
$$+ \frac{1}{Th}\sum_{j=1}^{T} K^2\left(\frac{x - X_j^*}{h}\right)(\hat{v}_b^*(X_j^*;\phi) - v(X_j^*;\phi)).$$

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

Now, because

$$\hat{v}_b^*(z;\phi) = m_b^*(z;\phi^2) - (m^*(z;\phi))^2$$

is a Nadaraya–Watson kernel estimator of $v(z;\phi) = E[(\phi(X_{t+1}) - m(z;\phi))^2 | X_t = z]$ we have for $D_h = [x-h, x+h]$ that

$$\begin{aligned}\sup_{z\in D_h} |\hat{v}_b^*(z;\phi) - v(z;\phi)| &\le \sup_{z\in D_h} |m_b^*(z;\phi^2) - m(z;\phi^2)| \\ &\quad + \sup_{z\in D_h} |m_b^*(z;\phi) - m(z;\phi)| \sup_{z\in D_h} |m_b^*(z;\phi) + m(z;\phi)| \\ &\to 0\end{aligned}$$

in probability by Lemma 5.4. Thus

$$\begin{aligned}&\frac{1}{Th}\sum_{j=1}^{T-1} K^2\left(\frac{x - X_j^*}{h}\right)|\hat{v}_b^*(X_j^*;\phi) - v(X_j^*;\phi)| \\ &\quad \le \sup_{z\in D_h} |\hat{v}_b^*(z;\phi) - v(z;\phi)| O_P(1) \to 0.\end{aligned}$$

To conclude the proof of (5.16) we show that

$$V_T^*(x;\phi) \equiv \frac{1}{Th}\sum_{j=1}^{T} K^2\left(\frac{x - X_j^*}{h}\right) v(X_j^*;\phi) \to \tau^2(x;\phi) \qquad \textbf{(5.18)}$$

in probability. Taking expectation of this term with respect to the bootstrap distribution we get using the same arguments as in establishing (5.14) that

$$E^*[V_T^*(x;\phi)] = \int \frac{h^{-1}\sum_{l=2}^{T} v(X_l;\phi)(K^2((x - X_l)/h))W_b(X_m - X_{l-1})}{\sum_{s=1}^{T-1} W_b(X_m - X_s)} dF_T^*(X_m)$$

$$\to \int K^2(u)du f(x)\mathrm{Var}[\phi(X_{t+1})|X_t = x],$$

where the last convergence follows using Lemma 5.3. Furthermore, along the same lines as in the proof of (5.15) we get that $\mathrm{Var}^*[V_T^*(x;\phi)] = O_P(T^{-1}h^{-p})$.

To prove (5.17) we use the inequality $E[|X|I(|X| > C)] \le C^{-1}E|X|^2$ where $C > 0$ is a constant provided the random variable X satisfies $E|X|^2 < \infty$. We then have for every $\delta > 0$

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

$$\sum_{j=1}^{T-1} E^*[D_{j+1}^{*2} I(D_{j+1}^{*2} \geq \delta) | \mathcal{F}_j^*]$$

$$\leq \frac{1}{\delta} \sum_{j=1}^{T-1} E^*[D_{j+1}^{*4} | \mathcal{F}_j^*]$$

$$= \frac{1}{\delta h^2 T^2} \sum_{j=1}^{T} E^* \left[\{\phi(X_{j+1}^*) - m_b^*(X_j^*; \phi)\}^4 K^4 \left(\frac{x - X_j^*}{h} \right) | \mathcal{F}_j^* \right]$$

$$= \frac{1}{\delta h^2 T^2} \sum_{j=1}^{T-1} K^4((x - X_j^*)/h) \frac{\sum_{l=2}^{T} (\phi(X_l) - m_b^*(X_j^*; \phi))^4 W_b(X_j^* - X_{l-1})}{\sum_{s=1}^{T-1} W_b(X_j^* - X_s)}.$$

Let

$$R_T(z; \phi) = \frac{\sum_{l=2}^{T} (\phi(X_l) - m_b^*(z; \phi))^4 W_b(z - X_l)}{\sum_{n=1}^{T-1} W_b(z - X_s)},$$

which is a Nadaraya–Watson estimator of $R(z; \phi) = E[(\phi(X_{t+1}) - m(z; \phi))^4 | X_t = x]$, and note that $\sup_{z \in D_h} |R_T(z; \phi) - R(z; \phi)| \to 0$ in probability. Thus

$$\sum_{j=1}^{T-1} E^*[D_{j+1}^{*2} I(D_{j+1}^{*2} \geq \delta) | \mathcal{F}_j^*] \leq \frac{1}{\delta T^2 h^2} \sum_{j=1}^{T-1} K^4((x - X_j^*)/h) R_T(X_j^*; \phi) + o_P(1)$$

$$= O_P(\delta^{-1} T^{-1} h^{-1}),$$

where the last equality follows by the same arguments as those used to prove (5.18). ■

LEMMA 5.7. *Under Assumptions (A1)–(A6) and if* $T^{1/2} h^{(p+4)/2} \to C_h > 0$, *we have conditionally on* $X_1, X_2, \ldots, X_T$ *that, as* $T \to \infty$,

$$\sqrt{\frac{h^p}{T}} \sum_{j=p}^{T-1} (E^*[\phi(X_{j+1}^*) | \mathbf{Y}_{j,p}^*] - m_b^*(\mathbf{x}; \phi)) K_h(\mathbf{x} - \mathbf{Y}_{j,p}^*)$$

$$\to C_h \int u^2 \kappa(u) du \sum_{l=1}^{p} \left\{ m^{(x_l)}(\mathbf{x}; \phi) f_{\mathbf{Y}_{t,p}}^{(x_l)}(\mathbf{x}) + \frac{1}{2} m^{(x_l x_l)}(\mathbf{x}; \phi) f_{\mathbf{Y}_{t,p}}(\mathbf{x}) \right\} \tag{5.19}$$

in probability.

Proof. Let $m_b^*(\mathbf{Y}_{t,p}^*; \phi) = E^*[\phi(X_{t+1}^*) | \mathbf{Y}_{t,p}^*]$ and $\mathbf{x} = (x_1, x_2, \ldots, x_p)$. By a Taylor series expansion we get

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

$$\sqrt{\frac{h^p}{T}}\sum_{j=p}^{T-1}(m_b^*(\mathbf{Y}_{j,p}^*;\phi)-m_b^*(\mathbf{x};\phi))K_h(\mathbf{x}-\mathbf{Y}_{j,p}^*)$$

$$=\sum_{l=1}^{p}m_b^{*(x_l)}(\mathbf{x};\phi)\sqrt{\frac{h^p}{T}}\sum_{j=p}^{T-1}(X_{j-l+1}^*-x_l)K_h(\mathbf{x}-\mathbf{Y}_{j,p}^*)$$

$$+\frac{1}{2}\sum_{l_1=1}^{p}\sum_{l_2=1}^{p}m_b^{*(x_{l_1}x_{l_2})}(\tilde{\mathbf{x}};\phi)\sqrt{\frac{h^p}{T}}\sum_{j=p}^{T-1}(X_{j-l_1+1}^*-x_{l_1})(X_{j-l_2+1}^*-x_{l_2})$$

$$\times K_h(\mathbf{x}-\mathbf{Y}_{j,p}^*)$$

for some $\tilde{\mathbf{x}}$ between $\mathbf{x}$ and $\mathbf{Y}_{j,p}^*$. Because under the assumptions made $\sqrt{Th^p}(E\hat{m}_h(\mathbf{x};\phi)-m(\mathbf{x};\phi))\to B(\mathbf{x};\phi)$ in probability, the assertions of the lemma will be established if we show that, in probability,

$$m_b^{*(x_l)}(\mathbf{x};\phi)\to m^{(x_l)}(\mathbf{x};\phi), \tag{5.20}$$

$$\sqrt{\frac{h^p}{T}}\sum_{j=p}^{T-1}(X_{j-l+1}^*-x_l)K_h(\mathbf{x}-\mathbf{Y}_{j,p}^*)\to C_h\int u^2\kappa(u)du f_{\mathbf{Y}_{t,p}}^{(x_l)}(\mathbf{x}), \tag{5.21}$$

$$\sqrt{\frac{h^p}{T}}\sum_{j=p}^{T-1}(X_{j-l_1+1}^*-x_{l_1})(X_{j-l_2+1}^*-x_{l_2})K_h(\mathbf{x}-\mathbf{Y}_{j,p}^*)\to\delta_{l_1,l_2}C_h\int u^2\kappa(u)du f_{\mathbf{Y}_{t,p}}(\mathbf{x}), \tag{5.22}$$

and

$$\sup_{\mathbf{z}\in[\mathbf{x}-h,\mathbf{x}+h]}|m_b^{*(x_{l_1}x_{l_2})}(\mathbf{z};\phi)-\tilde{m}^{(x_{l_1}x_{l_2})}(\mathbf{z};\phi)|\to 0, \tag{5.23}$$

where

$$\tilde{m}^{(x_ix_j)}(\mathbf{z};\phi)=\frac{1}{f_{\mathbf{Y}_{t,p}}(\mathbf{z})}\left\{\frac{1}{2}(1+\delta_{i,j})\int\phi(y)f_{X_{t+1}\mathbf{Y}_{t,p}}^{(x_ix_j)}(y,\mathbf{z})dy-m(\mathbf{z};\phi)f_{\mathbf{Y}_{t,p}}^{(x_ix_j)}(\mathbf{z})\right.$$

$$\left.-m^{(x_i)}(\mathbf{z};\phi)f_{\mathbf{Y}_{t,p}}^{(x_j)}(\mathbf{z})-m^{(x_j)}(\mathbf{z};\phi)f_{\mathbf{Y}_{t,p}}^{(x_i)}(\mathbf{z})\right\}. \tag{5.24}$$

The rest of the proof is devoted to showing that (5.20)–(5.23) are true.

Consider (5.20) and note that

$$m_b^{*(x_l)}(\mathbf{x};\phi)=\frac{1}{\hat{f}_b(\mathbf{x})}\left\{\frac{1}{Tb^{p+1}}\sum_{j=p}^{T-1}\phi(X_{j+1})W^{(x_l)}((\mathbf{x}-\mathbf{Y}_{j,p})/b)\right.$$

$$\left.-m_b^*(\mathbf{x};\phi)\frac{1}{Tb^{p+1}}\sum_{j=p}^{T-1}W^{(x_l)}((\mathbf{x}-\mathbf{Y}_{j,p})/b)\right\}.$$

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

Recall that $\hat{f}_b(\mathbf{x}) \to f_{\mathbf{Y}_{t,p}}(\mathbf{x})$ and $m_b^*(\mathbf{x};\phi) \to m(\mathbf{x};\phi)$ in probability. Furthermore, by standard arguments we have

$$\frac{1}{Tb^{p+1}} \sum_{j=p}^{T-1} \phi(X_{j+1}) W^{(x_l)}((\mathbf{x} - \mathbf{Y}_{j,p})/b) \to \int \phi(y) f^{(x_l)}_{X_{t+1}\mathbf{Y}_{t,p}}(y, \mathbf{x})\, dy \tag{5.25}$$

and

$$\frac{1}{Tb^{p+1}} \sum_{j=p}^{T-1} W^{(x_l)}((\mathbf{x} - \mathbf{Y}_{j,p})/b) \to f^{(x_l)}_{\mathbf{Y}_{t,p}}(\mathbf{x}). \tag{5.26}$$

Therefore, $m_b^{*(x_l)}(\mathbf{x};\phi) \to m^{(x_l)}(\mathbf{x};\phi)$ in probability.

The proof of (5.21) and (5.22) is very similar and makes explicit use of (5.13) and arguments like those applied in the proof of Lemma 5.3. To avoid tedious manipulations of formulas we focus on the case $p = 1$. Consider (5.21) so we have

$$E^*\left[\sqrt{\frac{h}{T}} \sum_{j=1}^{T-1} (X_j^* - x) K_h(x - X_j^*)\right]$$

$$= (T-1)\sqrt{\frac{h}{T}} \sum_{l=2}^{T} (X_l - x) K_h(x - X_l) P(X_t^* = X_l)$$

$$= (T-1)\sqrt{\frac{h}{T}} \int \frac{\sum_{l=2}^{T} (X_l - x) K_h(x - X_l) W_b(X_s - X_{l-1})}{\sum_{n=2}^{T} W_b(X_s - X_n)}\, dF_T^*(X_s).$$

Now, to establish the desired result we have to show that

$$\sup_{y \in S} \left| \frac{\sqrt{Th} \sum_{l=2}^{T} (X_l - x) K_h(x - X_l) W_b(y - X_{l-1})}{\sum_{n=1}^{T-1} W_b(y - X_{n-1})} - C_h \frac{f^{(x)}_{X_t X_{t-1}}(x,y)}{f_{X_t}(y)} \int u^2 K(u)\, du \right|$$

$$\to 0. \tag{5.27}$$

By (A1)(iv) and the uniform convergence of $\hat{f}_b(y)$ to $f(y)$, it suffices to show that

$$\sup_{y \in S} \left| \sqrt{\frac{h}{T}} \sum_{l=2}^{T} (X_l - x) K_h(x - X_l) W_b(y - X_{l-1}) - C_h f^{(x)}_{X_t X_{t-1}}(x,y) \int u^2 K(u)\, du \right|$$

$$\to 0. \tag{5.28}$$

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

For this note that

$$E\left[\sqrt{\frac{h}{T}}\sum_{l=2}^{T}(X_l-x)K_h(x-X_l)W_b(y-X_{l-1})\right]$$
$$=-(T-1)\sqrt{\frac{h^3}{T}}\int uK(u)W(r)f_{X_tX_{t-1}}(x-uh,y-rb)drdu$$
$$=(T-1)\sqrt{\frac{h^5}{T}}\int u^2K(u)W(r)f^{(x)}_{X_tX_{t-1}}(\tilde{x},y-rb)drdu,$$

where $|\tilde{x}-x|<h$. Thus by the Lipschitz continuity of $f^{(x)}(\cdot,\cdot)$ we get that

$$(T-1)\sqrt{\frac{h^5}{T}}\int u^2K(u)W(r)f^{(x)}(\tilde{x},y-rb)drdu\to C_hf^{(x)}_{X_tX_{t-1}}(x,y)\int u^2K(u)du$$

uniformly in y; that is, we have shown that

$$\sup_{y\in S}\left|E\left[\sqrt{\frac{h}{T}}\sum_{l=2}^{T}(X_l-x)K_h(x-X_l)W_b(y-X_{l-1})\right]-C_hf^{(x)}_{X_tX_{t-1}}(x,y)\int u^2K(u)du\right|$$
$$\to 0$$

in probability. Thus to establish (5.28) it remains to show that

$$\sup_{y\in S}\left|\sqrt{\frac{h}{T}}\sum_{l=2}^{T}\{(X_l-x)K_h(x-X_l)W_b(y-X_{l-1})\right.$$
$$\left.-E(X_l-x)K_h(x-X_l)W_b(y-X_{l-1})\}\right|\to 0$$

in probability. To handle this term, we proceed as in the proof that $M_{1,T}\to 0$ in probability in Lemma 5.3. In particular, let

$$Z_{l,T}(y)=(X_l-x)K\left(\frac{x-X_l}{h}\right)W\left(\frac{y-X_{l-1}}{b}\right)$$
$$-E(X_l-x)K\left(\frac{x-X_l}{h}\right)W\left(\frac{y-X_{l-1}}{b}\right)$$

and note that $|Z_{l,T}|\le C_1h$ and that $EZ^2(y)\le C_2^2h^3b$ for some generic positive constants C_1 and C_2. Using the splitting device applied in the proof of Lemma 5.3 we get by the Lipschitz continuity of W that

$$\sup_{y\in S}\left|\frac{1}{\sqrt{Th}b}\sum_{l=2}^{T}Z_{l,T}(y)\right|\le\max_{1\le i\le N_T}\left|\frac{1}{\sqrt{Th}b}\sum_{l=2}^{T}Z_{l,T}(y_i)\right|+O(h^{1/2}L_Tb^{-2}).$$

Along the same lines as in dealing with (5.11) we get for the first term on the right hand side of the preceding inequality using the notation $\tilde{Z}_{l,T}(y_i)=h^{-1}C_2^{-1}Z_{l,T}(y_i)$ and for $0<\epsilon<1$ and $L_T=(\log(T))^3b^{3/2+\epsilon}h^{1/2+\epsilon}$ that

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

$$
\begin{aligned}
&P\left(\max_{1\le i\le N_T}\left|\frac{1}{\sqrt{Thb}}\sum_{l=2}^{T} Z_{l,T}(y_i)\right| \ge \frac{L_T h^{1/2}}{b^2}\right)\\
&\quad\le \sum_{i=1}^{N_T} P\left(\left|\sum_{l=2}^{T}\tilde{Z}_{l,T}(y_i)\right| \ge \frac{L_T T^{1/2}}{C_2 b}\right)\\
&\quad\le \sum_{i=1}^{N_T} P\left(\left|\sum_{l=2}^{T}\tilde{Z}_{l,T}(y_i)\right| \ge \frac{(\log(T))^2\sqrt{T}\sqrt{hb}\log\sqrt{hb}^{-1}}{C_2}\right)\\
&\quad\le O(N_T)\alpha\exp\{-\beta\log(T)/\sqrt{C_2}\}
\end{aligned}
$$

for some $\alpha > 0$ and $\beta > 0$. Thus by the assumptions made on b and h we get that $L_T \to 0$, $L_T h^{1/2}/b^2 \to 0$, and $O(N_T)\alpha\exp\{-\beta\log(T)/C_2\} \to 0$ as $T \to \infty$.

The proof of (5.22) follows exactly the same lines as the proof of (5.21) and is therefore omitted. To establish (5.23) verify first that

$$
\begin{aligned}
&m_b^{*(x_{l_1},x_{l_2})}(\mathbf{z};\phi)\\
&\quad= \frac{1}{\hat{f}_b(\mathbf{z})}\left\{\frac{1}{Tb^{p+2}}\sum_{j=p}^{T-1}\phi(X_{j+1})W^{(x_{l_1},x_{l_2})}((\mathbf{z}-\mathbf{Y}_{j,p})/b)\right.\\
&\qquad\qquad - m_b^{*(x_{l_2})}(\mathbf{z};\phi)\,\frac{1}{Tb^{p+2}}\sum_{j=p}^{T-1}W^{(x_{l_1},x_{l_2})}((\mathbf{z}-\mathbf{Y}_{j,p})/b)\\
&\qquad\qquad \left. - m_b^{*(x_{l_1})}(\mathbf{z};\phi)\hat{f}_b^{(x_{l_2})}(\mathbf{z}) - m_b^{*(x_{l_2})}(\mathbf{z};\phi)\hat{f}_b^{(x_{l_1})}(\mathbf{z})\right\}. \qquad \textbf{(5.29)}
\end{aligned}
$$

We then have to show that every element on the right hand side of (5.29) converges uniformly to the corresponding elements on the right hand side of (5.24). We show this for the first term on the right hand side of (5.29), the other terms being handled in exactly the same manner. For this note that

$$
\begin{aligned}
&\sup_{\mathbf{z}\in\mathbf{S}}\left|\frac{1}{T}\sum_{j=p}^{T-1}\phi(X_{j+1})W_b^{(x_{l_1}x_{l_2})}(\mathbf{z}-\mathbf{Y}_{j,p}) - \frac{1}{2}(1+\delta_{x_{l_1}x_{l_2}})\int\phi(y)f^{(x_{l_1}x_{l_2})}_{X_{t+1}\mathbf{Y}_{t,p}}(y,\mathbf{z})\,dy\right|\\
&\quad\le \sup_{\mathbf{z}\in\mathbf{S}}\left|\frac{1}{T}\sum_{j=p}^{T-1}(\phi(X_{j+1})W_b^{(x_{l_1}x_{l_2})}(\mathbf{z}-\mathbf{Y}_{j,p}) - E\phi(X_{j+1})W_b^{(x_{l_1}x_{l_2})}(\mathbf{z}-\mathbf{Y}_{j,p}))\right|\\
&\qquad + \sup_{\mathbf{z}\in\mathbf{S}}\left|\frac{T-p}{T}\int\phi(y)W_b^{(x_{l_1}x_{l_2})}(\mathbf{z}-\mathbf{r})f_{X_{t+1}\mathbf{Y}_{t,p}}(y,\mathbf{r})\,dyd\mathbf{r}\right.\\
&\qquad\qquad \left. - \frac{1+\delta_{x_{l_1}x_{l_2}}}{2}\int\phi(y)f^{(x_{l_1}x_{l_2})}_{X_{t+1}\mathbf{Y}_{t,p}}(y,\mathbf{z})\,dy\right|\\
&\quad= B_{1,T} + B_{2,T},
\end{aligned}
$$

with an obvious notation for $B_{1,T}$ and $B_{2,T}$. Now, by the same arguments used to handle the term $T_1(\mathbf{x})$ in the proof of Lemma 5.3 and by (A6), we get that $B_{2,T} = O(b)$. Furthermore, by the mean value theorem we have applying the

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

same splitting device as the one used in handling the term $T_2(\mathbf{x})$ in the lemma mentioned earlier, that

$$\sup_{\mathbf{z}\in\mathbf{S}}\left|\frac{1}{T}\sum_{j=p}^{T-1}(\phi(X_{j+1})W_b^{(x_{l_1}x_{l_2})}(\mathbf{z}-\mathbf{Y}_{j,p})-E\phi(X_{j+1})W_b^{(x_{l_1}x_{l_2})}(\mathbf{z}-\mathbf{Y}_{j,p}))\right|$$
$$\leq \max_{1\leq i\leq N_T}\left|\frac{1}{T}\sum_{j=p}^{T-1}(\phi(X_{j+1})W_b^{(x_{l_1}x_{l_2})}(\mathbf{z}_i-\mathbf{Y}_{j,p})-E\phi(X_{j+1})W_b^{(x_{l_1}x_{l_2})}(\mathbf{z}_i-\mathbf{Y}_{j,p}))\right|$$
$$+ O_P(L_T b^{-(p+3)}).$$

To conclude the proof verify that $\mathrm{Var}(\phi(X_{j+1})W^{(x_{l_1}x_{l_2})}((\mathbf{z}_i-\mathbf{Y}_{j,p})/b))\leq Cb^p$ and let $L_T=(\log T)^3 b^{(p/2)+1+\epsilon}/\sqrt{T}$ for some $0<\epsilon<1$. Using the notation

$$\Psi_{j,T}(\mathbf{z}_i)=\phi(X_{j+1})W^{(x_{l_1}x_{l_2})}((\mathbf{z}_i-\mathbf{Y}_{j,p})/b)-E(\phi(X_{j+1})W^{(x_{l_1}x_{l_2})}((\mathbf{z}_i-\mathbf{Y}_{j,p})/b))$$

and applying the exponential inequality for strongly mixing sequences used in the proof of Lemma 5.3 we then get that for some constants $\alpha>0$ and $\beta>0$

$$\sum_{i=1}^{N_T}P\left(\frac{1}{Tb^{p+2}}\sum_{j=p}^{T-1}\Psi_{j,T}(\mathbf{z}_i)\geq L_T b^{-(p+3)}\right)$$
$$\leq\sum_{i=1}^{N_T}P\left(\sum_{j=p}^{T-1}C^{-1}\Psi_{j,T}(\mathbf{z}_i)\geq\frac{\sqrt{T}L_T}{Cb^{(p/2)+1}\log\sqrt{b^{p/2}}^{-1}}\sqrt{T}b^{p/2}\log\sqrt{b^{p/2}}^{-1}\right)$$
$$\leq O(N_T)\alpha\exp\{-\beta\log(T)/\sqrt{C}\}$$
$$=o(1).$$ ■

REFERENCES

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

Brown, B.M. (1971) Martingale central limit theorems. *Annals of Mathematical Statistics* 42, 59–66.

Doukhan, P. (1994) *Mixing: Properties and Examples*. Lecture Notes in Statistics 85. New York: Springer-Verlag.

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

Franke, J., J.-P. Kreiss, & E. Mammen (1996) Bootstrap of Kernel Smoothing in Nonlinear Time Series. Preprint.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Applications*. San Diego: Academic Press.

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

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

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

Neumann, M. & J.-P. Kreiss (1998) Regression-type inference in nonparametric autoregression. *Annals of Statistics* 26, 1570–1613.

Paparoditis, E. & D.N. Politis (1997) The local bootstrap for Markov processes. *Journal of Statistical Planning and Inference*, forthcoming.

Robinson, P.M. (1983) Nonparametric estimation for time series models. *Journal of Time Series Analysis* 4, 185–208.

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press

Roussas, G. (1990) Nonparametric regression estimation under mixing conditions. *Stochastic Processes and Their Applications* 36, 107–116.

Tjøstheim, D. & B. Auestad (1994) Nonparametric identification of nonlinear time series: Projections. *Journal of the American Statistical Association* 89, 1398–1409.

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

https://doi.org/10.1017/S0266466601173020 Published online by Cambridge University Press