NONPARAMETRIC AUTOREGRESSION WITH MULTIPLICATIVE VOLATILITY AND ADDITIVE MEAN

By Lijian Yang, Wolfgang Härdle and Jens P. Nielsen

Michigan State University, Humboldt Universität zu Berlin and PFA Pension, Copenhagen

First version received June 1996

Abstract. For over a decade, nonparametric modelling has been successfully applied to studying nonlinear structures in financial time series. It is well known that the usual nonparametric models often have less than satisfactory performance when dealing with more than one lag. When the mean has an additive structure, however, better estimation methods are available which fully exploit such a structure. Although in the past such nonparametric applications had been focused more on the estimation of the conditional mean, it is equally if not more important to measure the future risk of the series along with the mean. For the volatility function, i.e. the conditional variance given the past, a multiplicative structure is more appropriate than an additive structure, as the volatility is a positive scale function and a multiplicative model provides a better interpretation of each lagged value's influence on such a function. In this paper we consider the joint estimation of both the additive mean and the multiplicative volatility. The technique used is marginally integrated local polynomial estimation. The procedure is applied to the deutschmark/US dollar daily exchange returns.

Keywords. Additive mean; geometric ergodicity; geometric mixing; local polynomial regression; marginal integration; multiplicative volatility; stationary probability density.

1. INTRODUCTION

The prediction of financial time series based on daily data is in general difficult, since after differencing most of the structure in the mean disappears. This is why random-walk-based models have been used in this context. The situation is different, though, for high frequency time series such as foreign exchange rates. Autoregressive models have been applied for such data with specific assumptions on the error distribution (see Engle, 1982; Engle and Ng, 1993). Some of the most common nonlinear autoregressive models were proposed by Tong (1978, 1983), Haggan and Ozaki (1981), Chan and Tong (1986) and Granger and Teräsvirta (1993). In particular it is important not only to predict future values but also to evaluate the risk, or the volatility, of the series. In the class of autoregressive conditional heteroskedastic (ARCH) models the volatility or the scale of innovative random shocks is a function of past values. Over the past 15 years, the strict parametric forms of these models have been questioned and more flexible nonparametric approaches have been studied as an alternative (see Robinson, 1983, 1984; Meese and Rose, 1991; Engle and Gonzalez-Rivera,

0143-9782/99/05 579–604 JOURNAL OF TIME SERIES ANALYSIS Vol. 20, No. 5
© 1999 Blackwell Publishers Ltd., 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

1991; Drost and Nijman, 1993). A more recent review is by Härdle and Chen (1995).

One of the models studied for foreign exchange rates, for example, is the conditional heteroskedastic autoregressive nonlinear (CHARN) model with one lag (Bossaerts *et al.*, 1996):

$$Y_i = m(Y_{i-1}) + s(Y_{i-1})\xi_i \tag{1.1}$$

where the $\{\xi_i\}_{i\geqslant 1}$ are independent and identically distributed (i.i.d.) random variables, $E(\xi_i) = E(\xi_i^3) = 0$, $E(\xi_i^2) = 1$ and $E(\xi_i^4) = m_4 < \infty$, and Y_0 is independent of the $\{\xi_i\}$. An analysis of the estimated residuals still revealed autocorrelation. Hence, more than one lagged variable in the modelling of the mean function $m(\cdot)$ and the scale function $s(\cdot)$ seems to be the necessary step in a further analysis.

We consider therefore in this paper the CHARN model of the form

$$Y_i = m(Y_{i-1}, Y_{i-2}, \ldots, Y_{i-d}) + s(Y_{i-1}, Y_{i-2}, \ldots, Y_{i-d})\xi_i \tag{1.2}$$

where the $\{\xi_i\}_{i\geqslant 1}$ are as in (1.1) and $Y_0, Y_1, \ldots, Y_{d-1}$ are random variables independent of the $\{\xi_i\}$. The conditional volatility function is $\upsilon(Y_{i-1}, Y_{i-2}, \ldots, Y_{i-d}) = s^2(Y_{i-1}, Y_{i-2}, \ldots, Y_{i-d})$. This form of the CHARN model in financial time series has been studied by Gouriéroux and Monfort (1992) and Masry and Tjøstheim (1995a). The estimation problem for the functions $m(\cdot)$ and $\upsilon(\cdot)$ has been treated by Härdle and Tsybakov (1997) for the case $d = 1$ with the local polynomial regression method. Härdle *et al.* (1998) studied vector autoregression with an arbitrary number of lags and arbitrary dimension. We define the CHARN model for general dimensions; however, from a practical point of view, the method can be expected to suffer from the statistical imprecision introduced by a large number of lags, in particular in the small-sample case. We illustrate the method with a foreign exchange rate application. Through lag selection (see Tschernig and Yang, 1999), we ended up using the first lag and the third lag of the time series.

Stone (1982) showed in the i.i.d. regression case that, if the mean function $m(\cdot)$ is a sum of univariate functions, then the one-dimensional convergence rate can be achieved for the estimation of $m(\cdot)$'s component functions. Tools for analysis of additive models in this context have been developed by Hastie and Tibshirani (1990), including the BRUTO algorithm for nonparametric modelling which Chen and Tsay (1993a, 1993b) applied to autoregressive time series. The 'integration method' (but not the term marginal integration) was introduced by Auestad and Tjøstheim (1991) and further explored by Tjøstheim and Austad (1994) for the precise analysis, previously unavailable, of additive model estimators. It provides closed form bias and various expressions of the one-dimensional function estimator. The term marginal integration was introduced by Linton and Nielsen (1995), who worked in the i.i.d. regression setting. Marginal integration has recently been employed in the autoregression setting by Masry and Tjøstheim (1995a, 1995b) and in the i.i.d. regression setting by Linton and Härdle (1996) and Severance-Lossin and Sperlich (1995).

© Blackwell Publishers Ltd 1999

The idea of the integration method is quite straightforward: in the regression setting for instance, if the mean function $m(x_1, x_2, \ldots, x_d)$ is a sum of univariate functions, say

$$m(x_1, x_2, \ldots, x_d) = c + \sum_{\beta=1}^{d} m_\beta(x_\beta) \tag{1.3}$$

then

$$m_\beta(x_\beta) = \int m(x_1, x_2, \ldots, x_d)\, dF(x_1, \ldots, \widehat{x_\beta}, \ldots, x_d) - C$$

where $F(x_1, \ldots, \widehat{x_\beta}, \ldots, x_d)$ is the joint distribution function of all the variables $X_1, \ldots, X_d$ with the βth X_β removed, and C is an additive constant. Hence each component function m_β is identified from $m(x_1, x_2, \ldots, x_d)$ through a simple integration procedure. Linton and Nielsen (1995) introduced the idea of applying integration estimation to multiplicative structures in dimension 2; in this paper we extend the integration formula to multiplicative volatility functions of any dimension.

To estimate the parameters in the CHARN model, we have to estimate the conditional mean function $m(\cdot)$ and the conditional variance or volatility function $v(\cdot)$ at the same time. The flexibility of our CHARN model is important in a number of economic applications, e.g. the prediction of financial time series, where the volatility function often plays an even more important role than the mean function. It is therefore beneficial to obtain the joint estimation of both $m(\cdot)$ and $v(\cdot)$ for model (1.2). The volatility function $v(\cdot)$ measures the scale and is always positive; therefore it seems more appropriate to model its changes multiplicatively rather than additively, as in the EGARCH model of Nelson (1991). In this paper we jointly estimate the additive (mean) and the multiplicative (volatility) functions with the integration method.

We therefore assume that the mean function $m(\cdot)$ is additive while the volatility function $v(Y_{i-1}, Y_{i-2}, \ldots, Y_{i-d}) = s(Y_{i-1}, Y_{i-2}, \ldots, Y_{i-d})^2$ is multiplicative:

$$m(Y_{i-1}, Y_{i-2}, \ldots, Y_{i-d}) = c_m + \sum_{\beta=1}^{d} m_\beta(Y_{i-\beta}) \tag{1.4}$$

$$v(Y_{i-1}, Y_{i-2}, \ldots, Y_{i-d}) = c_v \prod_{\beta=1}^{d} v_\beta(Y_{i-\beta}) \tag{1.5}$$

where c_m and c_v are constants, and $\{m_\beta(\cdot)\}_{\beta=1}^{d}$ and $\{v_\beta(\cdot)\}_{\beta=1}^{d}$ are sets of unknown functions. Besides the better rate of convergence for the estimation of $\{m_\beta(\cdot)\}_{\beta=1}^{d}$ and $\{v_\beta(\cdot)\}_{\beta=1}^{d}$ as discussed above, these univariate functions also allow us to quantify the impact of each lagged variable $Y_{i-\beta}$ on the mean and volatility more directly.

To formulate the identifiability conditions for the functions $\{m_\beta(\cdot)\}_{\beta=1}^{d}$ and

© Blackwell Publishers Ltd 1999

$\{v_\beta(\cdot)\}_{\beta=1}^d$, the process Y_i has to converge to a stationary distribution. If we denote by $\boldsymbol{X}_i$ the vector $(Y_{i-1}, Y_{i-2}, \ldots, Y_{i-d})^{\mathrm{T}}$, then $\{\boldsymbol{X}_i\}$ is a d-dimensional Markov process. Many authors, such as Tweedie (1975), Nummelin and Tuominen (1982), Mokkadem (1987), Tjøstheim (1990) and Diebolt and Guégan (1993), have developed geometric ergodicity criteria for Markov processes. Here we state some general assumptions.

(A1) The random variable ξ_i has a density function $p(\cdot)$. This density $p(\cdot)$ and the volatility function $v(\cdot)$ are strictly positive in a neighborhood of x.

(A2) There exists an $r > 0$ such that for $\sum_{\beta=1}^d |y_{i-\beta}| > r$, the functions $m(\cdot)$ and $s(\cdot)$ satisfy

$$|m(y_{i-1}, y_{i-2}, \ldots, y_{i-d})| \leqslant C_1 \left(1 + \sum_{\beta=1}^{d} |y_{i-\beta}|\right)$$

$$|s(y_{i-1}, y_{i-2}, \ldots, y_{i-d})| \leqslant C_2 \left(1 + \sum_{\beta=1}^{d} |y_{i-\beta}|\right)$$

with $C_1 + C_2 E|\xi_1| < 1/d$.

These assumptions are standard in this context in order to prevent the process from either dying our or exploding. Ango Nze (1992) proved the following.

LEMMA 1.1. *Under assumptions (A1) and (A2) the process $\{\boldsymbol{X}_i\}$ is geometrically ergodic, i.e. it is ergodic with stationary probability measure $\pi(\cdot)$ such that, for almost every $\boldsymbol{x}$,*

$$\|P^n(\cdot|\boldsymbol{x}) - \pi(\cdot)\|_{\mathrm{TV}} = \mathrm{O}(\rho^n)$$

for some $0 \leqslant \rho < 1$, where $P^n(\cdot|\boldsymbol{x})$ is the probability measure of $\boldsymbol{X}_n$ given $\boldsymbol{X}_d = \boldsymbol{x}$ and $\|\cdot\|_{\mathrm{TV}}$ is the total variation distance.

This lemma ensures that the process $\{\boldsymbol{X}_i\}$ is asymptotically stationary. We denote by $F(\cdot)$ the stationary distribution function. For all $1 \leqslant \alpha \leqslant d$, we denote by $F_\alpha(\cdot)$ the stationary distribution function of the αth variable, and $\overline{F}(\cdot)$ the stationary distribution function with the αth variable deleted. We allow ourselves to use the short-hand notation Y_β for $Y_{i-\beta}$. Let x_β denote the deterministic version of $Y_{i-\beta}$. We can now state the identifiability conditions.

(A3) $Em_\beta(Y) = \int m_\beta(x_\beta)\, dF_\beta(x_\beta) = 0$, for any Y that has distribution $F_\beta(\cdot)$ and for all $1 \leqslant \beta \leqslant d$.

(A4) $E\prod_{1\leqslant\beta\leqslant d, \beta\neq\alpha} v_\beta(Y_\beta) = \prod_{1\leqslant\beta\leqslant d,\beta\neq\alpha} v_\beta(x_\beta)\, d\overline{F}(\overline{x}) = 1$ for any $(Y_1, Y_2, \ldots, Y_d)$ that has distribution $F(\cdot)$, and for all $1 \leqslant \alpha \leqslant d$.

© Blackwell Publishers Ltd 1999

Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^{\mathrm{T}} \in \mathbb{R}^d$ be a point where we will estimate the mean and volatility functions. We define for every $1 \leqslant \alpha \leqslant d$, $M_\alpha(x_\alpha) = c_m + m_\alpha(x_\alpha)$, $V_\alpha(x_\alpha) = c_v v_\alpha(x_\alpha)$; then

$$m(\boldsymbol{x}) = \sum_{\beta=1}^{d} M_\beta(x_\beta) - (d-1)c_m \qquad v(\boldsymbol{x}) = c_v^{-(d-1)} \prod_{\beta=1}^{d} V_\beta(x_\beta). \tag{1.6}$$

In what follows, we adopt the notation $\boldsymbol{X}_i = (Y_{i-\alpha}, \underline{\overline{Y}_i})$ to highlight a particular direction of interest $Y_{i-\alpha}$, for all $1 \leqslant \alpha \leqslant d$, while $\overline{Y}_i$ is the $(d-1)$-dimensional vector that consists of all the remaining $Y_{i-\beta}$, $1 \leqslant \beta \leqslant d$, $\beta \neq \alpha$. Assumptions (A3) and (A4) yield the following marginal integration formulae for the unknown functions:

$$\int m(x_\alpha, \overline{x})\, d\overline{F}(\overline{x}) = M_\alpha(x_\alpha) = c_m + m_\alpha(x_\alpha) \tag{1.7}$$

$$\int v(x_\alpha, \overline{x})\, d\overline{F}(\overline{x}) = V_\alpha(x_\alpha) = c_v v_\alpha(x_\alpha). \tag{1.8}$$

These show that the univariate functions $\{m_\beta(\cdot)\}_{\beta=1}^d$ and $\{v_\beta(\cdot)\}_{\beta=1}^d$ are identifiable from the functions $m(\cdot)$ and $v(\cdot)$ up to some constants. And similar formulae exist for these constants as well:

$$c_m = \int m(x)\, dF(x) = E(Y) \qquad c_v = \left\{ \frac{1}{d} \sum_{\alpha=1}^{d} \int \prod_{1 \leqslant \beta \leqslant d, \beta \neq \alpha} V_\beta(x_\beta)\, d\overline{F}(\overline{x}) \right\}^{1/(d-1)}. \tag{1.9}$$

These are the basic equations that will be used later in our estimation procedure.

In Section 2, we present the estimators of $\{m_\beta(\cdot)\}_{\beta=1}^d$ and $\{v_\beta(\cdot)\}_{\beta=1}^d$ and study their asymptotic properties. In Section 3, we discuss the application of the result to deutschmark/US dollar daily return data. In Section 4, proofs of theorems are given. Inspection of the proofs in Section 4 shows that the result of the present paper also holds (with obvious reformulation) for the multivariate nonparametric regression model with heteroskedastic errors: $Y_i = m(X_{i1}, X_{i2}, \ldots, X_{id}) + s(X_{i1}, X_{i2}, \ldots, X_{id})\xi_i$, where ξ_i are as in (1.2), $(X_{i1}, X_{i2}, \ldots, X_{id}, Y_i)$ are i.i.d., and the design points $\{X_{i1}, X_{i2}, \ldots, X_{id}\}$ are independent of $\{\xi_i\}$.

2. THE ESTIMATORS

The estimators given in this section are based on local polynomial regression, first studied by Stone (1977) and Katkovnik (1979). The idea, as will be seen below, is to estimate an unknown function locally by polynomials, whose coefficients are calculated through kernel-weighted least squares (see also

© Blackwell Publishers Ltd 1999

Tsybakov, 1986; Ruppert and Wand, 1994; Wand and Jones, 1995; Fan and Gijbels, 1996).

Now we let $p > 0$ be any odd integer which will be the degree of polynomial used later. For any function $K(\cdot)$ we denote $\|K\|_2^2 = \int K^2(u)\,du$, while for a kernel function $K(\cdot)$ we define $K_h(u) = K(u/h)/h$, and $\mu_r(K) = \int u^r K(u)\,du$. We shall consider two kernel functions $K(\cdot)$ and $L(\cdot)$ that satisfy the following.

(A5) Both kernels $K(\cdot)$ and $L(\cdot)$ are bounded, symmetric, compactly supported and Lipschitz continuous with $\int K(u)\,du = \int L(u)\,du = 1$; while $K(\cdot)$ is positive, the kernel $L(\cdot)$ is of order $q > (d-1)(p+1)/2$.

When estimating functions $m_\alpha(\cdot)$ and $v_\alpha(\cdot)$ for a particular α, a multiplicative kernel is used consisting of K for the αth variable and L for all other variables.

We assume the following about the functions involved in the estimation.

(A6) The functions $m_\alpha(\cdot)$ and $v_\alpha(\cdot)$ have bounded Lipschitz continuous $(p+1)$th derivatives for all $1 \leqslant \alpha \leqslant d$.

(A7) The stationary distribution function $F(\cdot)$ has a density $\varphi(\cdot)$. The function $\varphi(\cdot)$ together with the densities $\varphi_\alpha(\cdot)$ of $F_\alpha(\cdot)$ and $\overline{\varphi}(\cdot)$ of $\overline{F}(\cdot)$ are uniformly bounded away from zero and infinity and have bounded Lipschitz continuous $(p+1)$th derivatives, for all $1 \leqslant \alpha \leqslant d$.

Lastly, we assume the following for two bandwidths, g for the kernel L, h for the kernel K.

(A8) Bandwidths g and h satisfy $g^{d-1}/h^2 \to \infty$, $nhg^{2(d-1)}/\ln^2(n) \to \infty$, $g^q/h^{p+1} \to 0$ and $h = h_0 n^{-1/(2p+3).}$

Note that assumption (A8) requires that $L(\cdot)$ have the order given in (A5). In particular, if we use local linear regression, i.e. $p = 1$, then the order of $L(\cdot)$ is $q > d-1$.

We can define the integration estimator for $M_\alpha(x_\alpha)$ as

$$\hat{M}_\alpha(x_\alpha) = \int \hat{m}(x_\alpha, \overline{x})\, d\hat{\overline{F}}(\overline{x}) = (n-d+1)^{-1} \sum_{l=d}^{n} \hat{m}(x_\alpha, \overline{Y}_l)$$

where $\hat{m}(x_\alpha, \overline{x})$ is an estimate of $m(\cdot)$ at $(x_\alpha, \overline{x})$ and $\hat{\overline{F}}(\overline{x})$ is the empirical cumulative distribution function. The estimator $\hat{M}_\alpha(x_\alpha)$ is thus based on the sample version of Equation (1.7). The estimator for c_m is simply the sample mean of the Y_j according to (1.9):

$$\hat{c}_m = \hat{E}(Y) = (n-d+1)^{-1} \sum_{j=d}^{n} Y_j$$

where $\hat{E}$ is the empirical mean of Y. These estimators are then used to obtain estimators for $m_\alpha(x_\alpha)$ and $m(\boldsymbol{x})$:

© Blackwell Publishers Ltd 1999

$$\hat{m}_\alpha(x_\alpha) = \hat{M}_\alpha(x_\alpha) - \hat{c}_m$$

$$\hat{m}(\boldsymbol{x}) = \hat{c}_m + \sum_{\beta=1}^{d} \hat{m}_\beta(x_\beta) = \sum_{\beta=1}^{d} \hat{M}_\beta(x_\beta) - (d-1)\hat{c}_m.$$

We now define $\hat{m}(x_\alpha, \overline{Y}_l)$ as follows. For all $l = d, d+1, \ldots, n$ and $\lambda = 0, \ldots, p$ let

$$Z = \{(Y_{i-\alpha} - x_\alpha)^\lambda\}_{(n-d+1)\times(p+1)}$$

$$W_l = \operatorname{diag}\left\{\frac{1}{n-d+1} K_h(Y_{i-\alpha} - x_\alpha) L_g(\overline{Y}_i - \overline{Y}_l)\right\}_{i=d}^{n}$$

where we denote

$$\operatorname{diag}(a) = \begin{bmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_k \end{bmatrix}$$

for any vector

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} \in \mathbb{R}^k.$$

Also write

$$Y = (Y_i)_{d \leqslant i \leqslant n} \qquad Y^2 = (Y_i^2)_{d \leqslant i \leqslant n}$$

and let e_λ be a $(p+1)$ vector of zeros whose $(\lambda+1)$th element is 1. Then

$$\hat{m}(x_\alpha, \overline{Y}_l) = e_0^{\mathrm{T}}(Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l Y$$

which is the usual local polynomial estimator of $m(\cdot)$ at $(x_\alpha, \overline{Y}_l)$ of order p in the αth direction and order 0 in all other directions. Our estimator $\hat{M}_\alpha(x_\alpha)$ is therefore

$$\hat{M}_\alpha(x_\alpha) = (n-d+1)^{-1} \sum_{l=d}^{n} e_0^{\mathrm{T}}(Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l Y.$$

Note that

$$E(Y_i^2 | X_i) = m^2(x_i) + v(X_i).$$

Thus a similar estimator for $V_\alpha(x_\alpha)$ based on Equation (1.8) is defined as

$$\hat{V}_\alpha(x_\alpha) = (n-d+1)^{-1} \sum_{l=d}^{n} \{e_0^{\mathrm{T}}(Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l Y^2 - \hat{m}(x_\alpha, \overline{Y}_l)^2\}$$

© Blackwell Publishers Ltd 1999

and that of c_v is based on (1.9):

$$\hat{c}_v = \left\{ \frac{1}{d} \sum_{\alpha=1}^{d} \frac{1}{n-d+1} \sum_{j=d}^{n} \prod_{1 \leqslant \beta \leqslant d, \beta \neq \alpha} \hat{V}_\beta(Y_{j-\beta}) \right\}^{1/(d-1)}.$$

We then obtain estimators for $v_\alpha(x_\alpha)$ and $v(\boldsymbol{x})$ as the following:

$$\hat{v}_\alpha(x_\alpha) = \hat{V}_\alpha(x_\alpha)\hat{c}_v^{-1}$$

$$\hat{v}(\boldsymbol{x}) = \hat{c}_v \prod_{\beta=1}^{d} \hat{v}_\beta(x_\beta) = \hat{c}_v^{-(d-1)} \prod_{\beta=1}^{d} \hat{V}_\beta(x_\beta).$$

Our first theorem gives the estimation result of the mean functions.

THEOREM 1. *Under assumptions (A1)–(A8), as $n \to \infty$, for any α*

$$(nh)^{1/2}\{\hat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha) - h^{p+1} b_{m\alpha}(x_\alpha)\} \xrightarrow{\text{D}} \text{N}\{0, \sigma^2_{m\alpha}(x_\alpha)\} \tag{2.1}$$

where

$$b_{m\alpha}(x_\alpha) = \frac{\mu_{p+1}(K_0^*)}{(p+1)!} m_\alpha^{(p+1)}(x_\alpha)$$

and

$$\sigma^2_{m\alpha}(x_\alpha) = \|K_0^*\|_2^2 \int \frac{v}{\varphi}(x_\alpha, w)\overline{\varphi}^2(w)\, dw$$

while for any $\alpha \neq \beta$, as $n \to \infty$ we have

$$\text{cov}[(nh)^{1/2}\{\hat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha)\}, (nh)^{1/2}\{\hat{M}_\beta(x_\beta) - M_\beta(x_\beta)\}] \to 0. \tag{2.2}$$

Furthermore, as $n \to \infty$

$$n^{1/2}(\hat{c}_m - c_m) \xrightarrow{\text{D}} \text{N}\{0, \sigma^2_{cm}(\boldsymbol{x})\}$$

for some implicitly defined constant σ^2_{cm}. The asymptotics of $(nh)^{1/2}\{\hat{m}_\alpha(x_\alpha) - m_\alpha(x_\alpha)\}$ are the same as those of the $(nh)^{1/2}\{\hat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha)\}$, while

$$(nh)^{1/2}\{\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x}) - h^{p+1} b_m(\boldsymbol{x})\} \xrightarrow{\text{D}} \text{N}\{0, \sigma^2_m(\boldsymbol{x})\} \tag{2.3}$$

where

$$b_m(\boldsymbol{x}) = \sum_{\alpha=1}^{d} b_{m\alpha}(x_\alpha)$$

and

© Blackwell Publishers Ltd 1999

$$\sigma_m^2(\boldsymbol{x}) = \sum_{\alpha=1}^{d} \sigma_{m\alpha}^2(x_\alpha).$$

The second theorem is about the estimation of the volatility functions.

THEOREM 2. *Under assumptions (A1)–(A8), as* $n \to \infty$, *for any* α

$$(nh)^{1/2}\{\hat{V}_\alpha(x_\alpha) - V_\alpha(x_\alpha) - h^{p+1} b_{V\alpha}(x_\alpha)\} \xrightarrow{\mathrm{D}} \mathrm{N}\{0, \sigma_{V\alpha}^2(x_\alpha)\} \tag{2.4}$$

where

$$b_{V\alpha}(x_\alpha) = \frac{\mu_{p+1}(K_0^*)}{(p+1)!}\{V_\alpha^{(p+1)}(x_\alpha) + 2m_\alpha^{(p+1)}(x_\alpha)M(x_\alpha)\} - \int 2b_m(x_\alpha, w)m(x_\alpha, w)\overline{\varphi}(w)\,dw$$

and

$$\sigma_{V\alpha}^2(x_\alpha) = \|K_0^*\|_2^2 \int \frac{v(m_4 v + 4m^2)}{\varphi}(x_\alpha, w)\overline{\varphi}^2(w)\,dw.$$

Also, as $n \to \infty$

$$\mathrm{cov}[(nh)^{1/2}\{\hat{V}_\alpha(x_\alpha) - V_\alpha(x_\alpha)\}, (nh)^{1/2}\{\hat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha)\}] \to 2\|K_0^*\|_2^2 \int \frac{vm}{\varphi}(x_\alpha, w)\overline{\varphi}^2(w)\,dw = c_{V\alpha}(x_\alpha) \tag{2.5}$$

while for any $\alpha \neq \beta$ *we have*

$$\begin{aligned}
&\mathrm{cov}[(nh)^{1/2}\{\hat{V}_\alpha(x_\alpha) - V_\alpha(x_\alpha)\}, (nh)^{1/2}\{\hat{V}_\beta(x_\beta) - V_\beta(x_\beta)\}] \to 0\\
&\mathrm{cov}[(nh)^{1/2}\{\hat{V}_\alpha(x_\alpha) - V_\alpha(x_\alpha)\}, (nh)^{1/2}\{\hat{M}_\beta(x_\beta) - M_\beta(x_\beta)\}] \to 0.
\end{aligned} \tag{2.6}$$

Furthermore

$$n^{1/2}(\hat{c}_v - c_v - b_c h^{p+1}) \xrightarrow{\mathrm{D}} \mathrm{N}\{0, \sigma_{cv}^2)$$

for some implicitly defined constant σ_{cv}^2 *and*

$$b_c = \frac{1}{d(d-1)c_v^{d-2}} \sum_{\alpha=1}^{d} \int \sum_{1 \leq \beta \leq d, \beta \neq \alpha} \left\{ \prod_{1 \leq \gamma \leq d, \gamma \neq \alpha, \beta} V_\gamma(y_\gamma) \right\} b_{v\beta}(y_\beta)\varphi(y)\,dy.$$

For any α

$$(nh)^{1/2}\{\hat{v}_\alpha(x_\alpha) - v_\alpha(x_\alpha) - h^{p+1} b_{v\alpha}(x_\alpha)\} \xrightarrow{\mathrm{D}} \mathrm{N}\{0, \sigma_{v\alpha}^2(x_\alpha)\} \tag{2.7}$$

where

© Blackwell Publishers Ltd 1999

$$b_{v\alpha}(x_\alpha) = \frac{1}{c_v}\{b_{V\alpha}(x_\alpha) - b_c v_\alpha(x_\alpha)\}$$

and

$$\sigma^2_{v\alpha}(x_\alpha) = \frac{1}{c_v^2}\sigma^2_{V\alpha}(x_\alpha)$$

while

$$(nh)^{1/2}\{\hat{v}(\boldsymbol{x}) - v(\boldsymbol{x}) - h^{p+1}b_v(\boldsymbol{x})\} \xrightarrow{\mathrm{D}} \mathrm{N}(0, \sigma^2_v(\boldsymbol{x})\}$$

where

$$b_v(\boldsymbol{x}) = v(\boldsymbol{x})\left\{\sum_{\beta=1}^{d}\frac{b_{V\beta}(x_\beta)}{V_\beta(x_\beta)} - (d-1)c_v^{-1}b_c\right\}$$

and

$$\sigma^2_v(\boldsymbol{x}) = v^2(\boldsymbol{x})\sum_{\beta=1}^{d}\frac{\sigma^2_{V\beta}(x_\beta)}{V^2_\beta(x_\beta)}.$$

The next theorem summarizes all the previous results together in the form of joint asymptotic normality for all estimators.

THEOREM 3. *Under assumptions (A1)–(A8), denote by* $\boldsymbol{B}(\boldsymbol{x})$ *the vector-valued function*

$$\{b_{m1}(x_1), b_{m2}(x_2), \ldots, b_{md}(x_d), b_m(\boldsymbol{x}), b_{v1}(x_1), b_{v2}(x_2), \ldots, b_{vd}(x_d), b_v(\boldsymbol{x}), 0, n^{1/2}b_c\}^{\mathrm{T}}$$

and by $\Sigma(\boldsymbol{x})$ *the following matrix:*

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} & \mathbf{0}_{d\times1} & \mathbf{0}_{d\times1} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} & 0 & 0 \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} & \mathbf{0}_{d\times1} & \mathbf{0}_{d\times1} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} & 0 & 0 \\ \mathbf{0}_{1\times d} & 0 & \mathbf{0}_{1\times d} & 0 & \sigma^2_{cm} & 0 \\ \mathbf{0}_{1\times d} & 0 & \mathbf{0}_{1\times d} & 0 & 0 & \sigma^2_{cv} \end{bmatrix}$$

where

$$\Sigma_{11} = \mathrm{diag}\{\sigma^2_{m\alpha}(x_\alpha)\}_{\alpha=1}^{d} \qquad \Sigma_{22} = \sigma^2_m(\boldsymbol{x})$$

$$\Sigma_{33} = \mathrm{diag}\{\sigma^2_{v\alpha}(x_\alpha)\}_{\alpha=1}^{d} \qquad \Sigma_{44} = \sigma^2_v(\boldsymbol{x})$$

$$\Sigma_{12} = \Sigma_{21}^{\mathrm{T}} = \{\sigma^2_{m\alpha}(x_\alpha)\}_{1\leqslant\alpha\leqslant d} \qquad \Sigma_{13} = \Sigma_{31}^{\mathrm{T}} = \mathrm{diag}\left\{\frac{c_{V\alpha}(x_\alpha)}{c_v}\right\}_{\alpha=1}^{d}$$

© Blackwell Publishers Ltd 1999

$$\Sigma_{14} = \Sigma_{41}^{\mathrm{T}} = \left\{ \frac{c_{V\alpha}(x_\alpha)}{c_v} \frac{v(\boldsymbol{x})}{V_\alpha(x_\alpha)} \right\}_{1 \leqslant \alpha \leqslant d} \qquad \Sigma_{23} = \Sigma_{32}^{\mathrm{T}} = \left\{ \frac{c_{V\alpha}(x_\alpha)}{c_v} \right\}_{1 \leqslant \alpha \leqslant d}^{\mathrm{T}}$$

$$\Sigma_{24} = \Sigma_{42}^{\mathrm{T}} = \sum_{\alpha=1}^{d} \frac{c_{V\alpha}(x_\alpha)}{c_v} \frac{v(\boldsymbol{x})}{V_\alpha(x_\alpha)} \qquad \Sigma_{34} = \Sigma_{43}^{\mathrm{T}} = \left\{ \sigma_{v\alpha}^2(x_\alpha) \frac{v(\boldsymbol{x})}{V_\alpha(x_\alpha)} \right\}_{1 \leqslant \alpha \leqslant d}.$$

Then, as $n \to \infty$

$$(nh)^{1/2} \times \left\{ \begin{array}{c} \hat{m}_1(x_1) - m_1(x_1) \\ \hat{m}_2(x_2) - m_2(x_2) \\ \vdots \\ \hat{m}_d(x_d) - m_d(x_d) \\ \hat{m}(\boldsymbol{x}) - m(\boldsymbol{x}) \\ \hat{v}_1(x_1) - v_1(x_1) \\ \hat{v}_2(x_2) - v_2(x_2) \\ \vdots \\ \hat{v}_d(x_d) - v_d(x_d) \\ \hat{v}(\boldsymbol{x}) - v(\boldsymbol{x}) \\ (1/h^{1/2})(\hat{c}_m - c_m) \\ (1/h^{1/2})(\hat{c}_v - c_v) \end{array} \right\} - \boldsymbol{B}(\boldsymbol{x}) h^{p+1} \xrightarrow{\mathrm{D}} \mathrm{N}\{\boldsymbol{0}_{(2d+4)\times(2d+4)}, \boldsymbol{\Sigma}(\boldsymbol{x})\}$$

We comment here that, although Theorem 3 is obtained for a local polynomial of degree p, where p is an odd integer, the same result holds for p even, in particular for $p = 0$, i.e. the Nadaraya–Watson estimator. We choose to have p odd here because it does not involve the derivatives of the design density in the bias and variance expressions, and thus is 'design-adaptive'.

3. AN APPLICATION

To illustrate our method with an example, we study the daily returns of the deutschmark/US dollar exchange rates from 2 January 1980 to 26 May 1986, a total of 1603 observations. The data are plotted in Figure 1.

We estimate the conditional mean and volatility functions of this series at lags 1 and 3. The choice of these two lags is based on the findings of Tschernig and Yang (1999), who have developed a nonparametric final prediction error criterion for determining significant lagged variables. For the estimation, we use subjectively selected bandwidths $h = 0.0062$, $g = 0.0074$, and the Nadaraya–Watson estimators. We find that, except for some boundary effects, the mean functions $m_\beta(\cdot)$ are very close to zero. The estimated volatility function $\hat{v}_\beta(\cdot)$ depicted in Figures 2 and 3, however, provides some fresh insights. Both the computation and graphics are done in XploRe (see Härdle *et al.*, 1995).

© Blackwell Publishers Ltd 1999

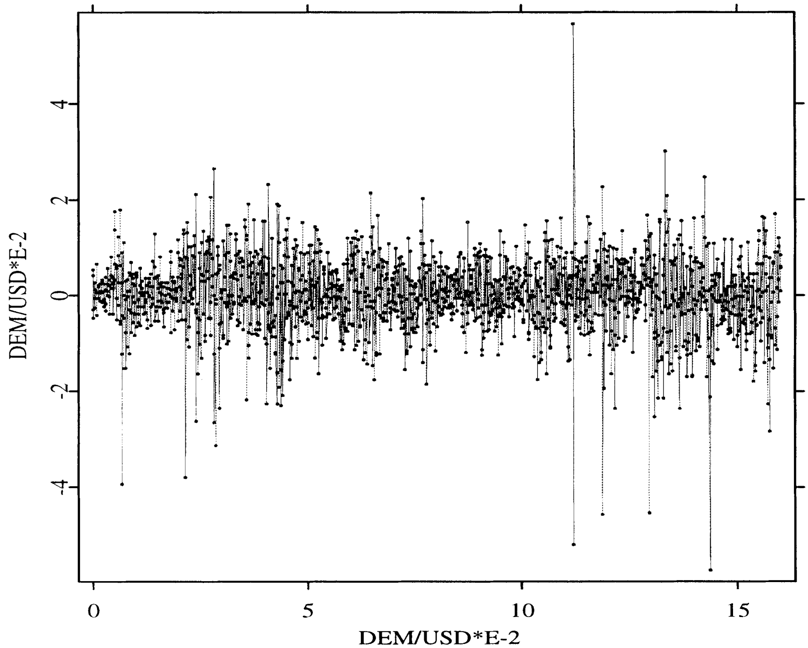

FIGURE 1. The daily returns of the deutschmark/US dollar (DEM/USD) exchange rates.

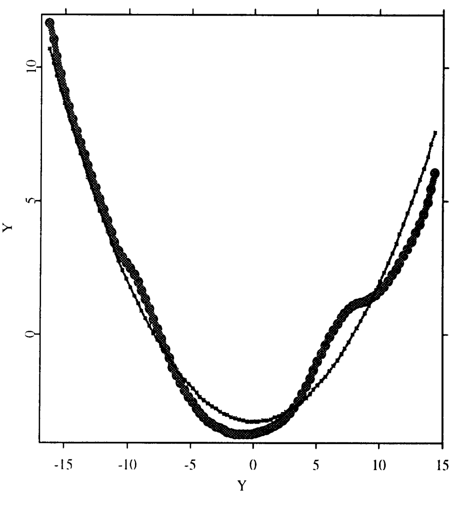

FIGURE 2. Volatility function $\hat{v}_1(\cdot)$ (thick) and its quadratic fit (thin).

© Blackwell Publishers Ltd 1999

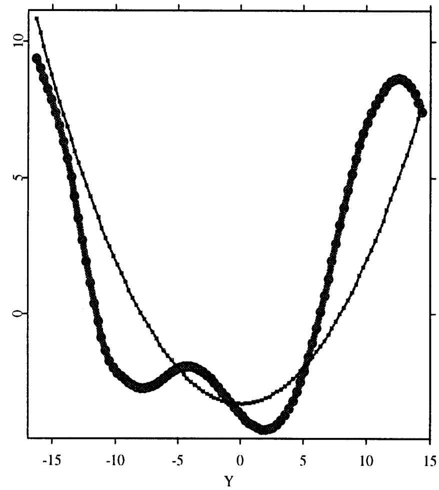

FIGURE 3. Volatility function $\hat{v}_3(\cdot)$ (thick) and its quadratic fit (thin).

Figures 2 and 3 show that the lagged variables impact the volatility function asymmetrically as both $\hat{v}_1(\cdot)$ and $\hat{v}_3(\cdot)$ are quite skewed, especially $\hat{v}_3(\cdot)$; we can see this by comparing $\hat{v}_1(\cdot)$ and $\hat{v}_3(\cdot)$ with their ordinary least squares quadratic fits which are the thin lines in the figures. Some kind of nonparametric testing would be needed to check the significance of these observed features.

Our observations about $\hat{v}_1(\cdot)$ and $\hat{v}_3(\cdot)$ have added weight to what some other studies had also suggested: that the basic GARCH model is perhaps inappropriate for the process we have here. Our analysis has gone a step further in nonparametric estimation of times series as the significant lagged variables are first identified by a nonparametric criterion (see Tschernig and Yang (1999) for details). This example of identifying significant lags and measuring their impacts points to a new comprehensive nonparametric approach to time series analysis.

4. PROOFS

Theorems 1 through 3 are proved in this section by the marginal integration technique as in Severance-Lossin and Sperlich (1995). We make use of the following geometric mixing results.

LEMMA 4.1. (DAVYDOV, 1973). *Under assumptions (A1) and (A2) and if, further,* $\boldsymbol{X}_d$ *is distributed with the stationary distribution* $\pi(\cdot)$, *then the process* $\{\boldsymbol{X}_i\}$ *is geometrically strongly mixing with the mixing coefficients satisfying* $\alpha(n) \leqslant c_0\rho_0^n$ *for some* $c_0 > 0$ *and* $0 < \rho_0 < 1$.

© Blackwell Publishers Ltd 1999

By arguments which are very similar to those used by Härdle *et al.* (1998), the above mixing lemma entails that the sample mean of any bounded continuous function of the observations Y_j converges in both probability and mean to the stationary population mean. The situation here is slightly more complicated than in that paper as we now have to average functions of two variables Y_j and $\overline{Y}_l$, one at a time. Nevertheless, the difference is more formal than substantial. We therefore neither state nor prove any such results here, but use them to derive the various formulae of asymptotic biases and variances as these are the new contributions of this paper.

The proof of the next lemma is standard and omitted. It employs the strong mixing condition of Lemma 1.1 and Lemma 4.1.

LEMMA 4.2. *Let*

$$D_l = (Z^{\mathrm{T}} W_l Z)^{-1} - \frac{1}{\varphi(x_\alpha, \overline{Y}_l)} H^{-1} S^{-1} H^{-1}.$$

$$\operatorname{cov}(D_l, D_k) = \rho^{|l-k|}\left[\mathrm{O_p}\left\{h + \frac{\ln n}{(nhg^{d-1})^{1/2}}\right\}\right]^2 \tag{4.1}$$

uniformly in x_α *and* $\overline{Y}_l$, *where* $H = \operatorname{diag}(h^\lambda)_{0 \leqslant \lambda \leqslant p}$.

Proofs of asymptotic normality in this section are based on the central limit theorem of Liptser and Shirjaev (1980). Conditions for applying this theorem will not be verified here as they are all standard. Set $S = \{\int u^{s+t} K(u)\, du\}_{0 \leqslant s,t \leqslant p}$, which contains all the moments of S up to order $2p$. Denote $S^{-1} = (s_{st})_{0 \leqslant s,t \leqslant p}$ and define

$$K_\lambda^*(u) = \sum_{t=0}^{p} s_{\lambda t} u^t K(u). \tag{4.2}$$

This $K_\lambda^*(\cdot)$ is called the λth-equivalent kernel. It has moments

$$\int u^q K_\lambda^*(u)\, du = \begin{cases} 0 & q \leqslant p,\ q \neq \lambda \\ 1 & q = \lambda \\ \Lambda_\lambda & q = p+1 \end{cases} \tag{4.3}$$

and $K_0^*(\cdot)$ would yield the bias rates of $n^{-2p/(2p+1)}$ for local polynomial estimation (see Wand and Jones, 1995).

To prove Theorem 1, we begin by observing the following simple equation:

$$e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l Z e_\lambda = \begin{cases} 0 & 0 \neq \lambda \\ 1 & 0 = \lambda. \end{cases} \tag{4.4}$$

Thus

© Blackwell Publishers Ltd 1999

$$\hat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha) = (n-d+1)^{-1} \sum_{l=d}^{n} e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l Y$$

$$- (n-d+1)^{-1} \sum_{l=d}^{n} e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l Z e_0 M_\alpha(x_\alpha)$$

$$- (n-d+1)^{-1} \sum_{l=d}^{n} \sum_{\nu=1}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l Z e_\nu$$

$$= (n-d+1)^{-1} \sum_{l=d}^{n} e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l \{Y - M_\alpha(x_\alpha)\}$$

$$- (n-d+1)^{-1} \sum_{l=d}^{n} \sum_{\nu=1}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l Z e_\nu.$$

Now assumption (A3) combined with the strong mixing properties of our process implies that for every $\beta = 1, 2, \ldots, d$, $\beta \neq \alpha$,

$$(n-d+1)^{-1} \sum_{l=d}^{n} m_\beta(Y_{l-\beta}) = \mathrm{O_p}\left(\frac{1}{n^{1/2}}\right)$$

and thus by (4.4) we also have (using the mixing properties of the process, see Lemma 1.1, Lemma 4.1 and Lemma 4.2)

$$(n-d+1)^{-1} \sum_{l=d}^{n} e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l Z e_0 m_\beta(Y_{l-\beta}) = \mathrm{O_p}\left(\frac{1}{n^{1/2}}\right).$$

So we have

$$\hat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha)$$

$$= (n-d+1)^{-1} \sum_{l=d}^{n} e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l \left\{ Y - \sum_{1 \leqslant \beta \leqslant d, \beta \neq \alpha} m_\beta(Y_{l-\beta}) - M_\alpha(x_\alpha) \right\}$$

$$- (n-d+1)^{-1} \sum_{l=d}^{n} \sum_{\nu=1}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l Z e_\nu + \mathrm{O_p}\left(\frac{1}{n^{1/2}}\right)$$

$$= (n-d+1)^{-1} \sum_{l=d}^{n} e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l$$

$$\times \left\{ Y - \sum_{\nu=1}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} Z e_\nu - \sum_{1 \leqslant \beta \leqslant d, \beta \neq \alpha} m_\beta(Y_{l-\beta}) - M_\alpha(x_\alpha) \right\}$$

or

© Blackwell Publishers Ltd 1999

$$\hat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha)$$
$$= (n-d+1)^{-1} \sum_{l=d}^{n} e_0^{\mathrm{T}} (Z^{\mathrm{T}} W_l Z)^{-1} Z^{\mathrm{T}} W_l$$
$$\times \left\{ Y - c_m - \sum_{\nu=0}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} Ze_\nu - \sum_{1 \leqslant \beta \leqslant d, \beta \neq \alpha} m_\beta(Y_{l-\beta}) \right\}. \quad (4.5)$$

Note that the λth element of

$$Z^{\mathrm{T}} W_l \left\{ Y - c_m - \sum_{\nu=0}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} Ze_\nu - \sum_{1 \leqslant \beta \leqslant d, \beta \neq \alpha} m_\beta(Y_{l-\beta}) \right\}$$

is

$$(n-d+1)^{-1} \sum_{j=d}^{n} (Y_{j-\alpha} - x_\alpha)^\lambda K_h(Y_{j-\alpha} - x_\alpha) L_g(\overline{Y}_j - \overline{Y}_l)$$
$$\times \left\{ Y_j - c_m - \sum_{\nu=0}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} (Y_{j-\alpha} - x_\alpha)^\nu - \sum_{1 \leqslant \beta \leqslant d, \beta \neq \alpha} m_\beta(Y_{l-\beta}) \right\}$$
$$= I_{\lambda l,1} + I_{\lambda l,2} + I_{\lambda l,3}$$

in which

$$I_{\lambda l,1} = (n-d+1)^{-1} \sum_{j=d}^{n} I_{\lambda lj,1}$$

where

$$I_{\lambda lj,1} = (Y_{j-\alpha} - x_\alpha)^\lambda K_h(Y_{j-\alpha} - x_\alpha) L_g(\overline{Y}_j - \overline{Y}_l)$$
$$\times \left\{ m_\alpha(Y_{j-\alpha}) - \sum_{\nu=0}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} (Y_{j-\alpha} - x_\alpha)^\nu \right\} \quad (4.6)$$

$$I_{\lambda l,2} = \sum_{1 \leqslant \beta \leqslant d, \beta \neq \alpha} (n-d+1)^{-1} \sum_{j=d}^{n} I_{\lambda l \beta j,2}$$

where

$$I_{\lambda l \beta j,2} = (Y_{j-\alpha} - x_\alpha)^\lambda K_h(Y_{j-\alpha} - x_\alpha) L_g(\overline{Y}_j - \overline{Y}_l) \{ m_\beta(Y_{j-\beta}) - m_\beta(Y_{l-\beta}) \} \quad (4.7)$$

and

$$I_{\lambda l,3} = (n-d+1)^{-1} \sum_{j=d}^{n} I_{\lambda lj,3} \quad (4.8)$$

where

$$I_{\lambda lj,3} = (Y_{j-\alpha} - x_\alpha)^\lambda K_h(Y_{j-\alpha} - x_\alpha) L_g(\overline{Y}_j - \overline{Y}_l) s(X_j) \xi_j. \quad (4.9)$$

© Blackwell Publishers Ltd 1999

LEMMA 4.3. *As* $n \to \infty$,

$$E(I_{\lambda l_1 j_1,1} I_{\lambda l_2 j_2,1}) = \rho^{\min(|l_1-l_2|,|j_1-j_2|)} \mathrm{O}(h^{2\lambda}/hg^{d-1})$$

uniformly, for $\lambda = 0, \ldots, p$ *and* $l_1, l_2, j_1, j_2 = d, \ldots, n$.

$$E(I_{\lambda l_1 \beta_1 j_1,1} I_{\lambda l_2 \beta_2 j_2,1}) = \rho^{\min(|l_1-l_2|,|j_1-j_2|)} \mathrm{O}(h^{2\lambda}/hg^{d-1})$$

uniformly, for $\lambda = 0, \ldots, p$ *and* $l_1, l_2, j_1, j_2 = d, \ldots, n$ *and* $1 \leqslant \beta \leqslant d$.

$$E(I_{\lambda l_1 j_1,3} I_{\lambda l_2 j_2,3}) = \rho^{\min(|l_1-l_2|,|j_1-j_2|)} \mathrm{O}(h^{2\lambda}/hg^{d-1})$$

uniformly, for $\lambda = 0, \ldots, p$ *and* $l_1, l_2, j_1, j_2 = d, \ldots, n$.

PROOF. We only show this for the first case

$$E(I_{\lambda l_1 j_1,3} I_{\lambda l_2 j_2,3}) = \rho^{\min(|l_1-l_2|,|j_1-j_2|)} \times \int (w_\alpha - x_\alpha)^{2\lambda} K_h^2(w_\alpha - x_\alpha) L_g^2(\overline{w} - \overline{Y}_l) v(w) \varphi(w)\, dw \{1 + \mathrm{o}(1)\}$$

where we have used Lemma 1.1. By a change of variable $w_\alpha = x_\alpha + hu_\alpha$, $\overline{w} = \overline{Y}_l + g\overline{u}$,

$$E(I_{\lambda l_1 j_1,3} I_{\lambda l_2 j_2,3}) = (hg^{d-1})^{-1} \{1 + \mathrm{o}(1)\} \times \int (hu_\alpha)^{2\lambda} K^2(u_\alpha) L^2(\overline{u}) v(x_\alpha + hu_\alpha, \overline{Y}_l + g\overline{u}) \varphi(x_\alpha + hu_\alpha, \overline{Y}_l + g\overline{u})\, du. \quad \blacksquare$$

Now

$$\begin{aligned} \mathrm{O}\left(\frac{h^{2\lambda}}{nhg^{d-1}}\right) \left[\mathrm{O_p}\left\{h + \frac{\ln n}{(nhg^{d-1})^{1/2}}\right\}\right]^2 &= \mathrm{O_p}\left(\frac{h^{2\lambda+2}}{nhg^{d-1}} + h^{2\lambda} \frac{\ln^2 n}{n^2 h^2 g^{2(d-1)}}\right) \\ &= \frac{h^{2\lambda}}{nh} \mathrm{O_p}\left(\frac{h^2}{g^{d-1}} + \frac{\ln^2 n}{nhg^{2(d-1)}}\right) \\ &= \mathrm{o_p}\left(\frac{h^{2\lambda}}{nh}\right) \end{aligned}$$

by using assumption (A8). Employing Lemma 4.2 and Lemma 4.3 now gives

$$\begin{aligned} &\sum_{\lambda=0}^{p} (n - d + 1)^{-1} \sum_{l=d}^{n} e_0^{\mathrm{T}} \\ &\quad \times \left\{(Z^{\mathrm{T}} W_l Z)^{-1} - \frac{1}{\varphi(x_\alpha, \overline{Y}_l)} H^{-1} S^{-1} H^{-1}\right\} e_\lambda (I_{\lambda l,1} + I_{\lambda l,2} + I_{\lambda l,3}) \\ &\qquad = \sum_{\lambda=0}^{p} h^{-\lambda} \mathrm{o_p}\left\{\frac{h^\lambda}{(nh)^{1/2}}\right\} = \mathrm{o_p}\left\{\frac{1}{(nh)^{1/2}}\right\} = \mathrm{o_p}(h^{p+1}) = \mathrm{o_p}(n^{-(p+1)/(2p+3)}). \end{aligned}$$

© Blackwell Publishers Ltd 1999

If we only have to consider the diagonal terms, then this fact is easily recognized (this is when we can ignore the correlation of the I terms with the rest). The correlation can be taken care of, however, by writing the $I_{\lambda l,k}$ as sums (see above), squaring the expression and conditioning on the 'I components'. The exponential decay of the correlations in Lemma 4.2 and Lemma 4.3 ensures that the order of magnitude is the same as if only the diagonal terms were considered.

PROOF OF THEOREM 1. Making the aforementioned substitution, we have in particular

$$\hat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha) - \mathrm{o_p}(h^{p+1})$$
$$= (n-d+1)^{-1} \sum_{l=d}^{n} \frac{1}{\varphi(x_\alpha, \overline{Y}_l)} e_0^{\mathrm{T}} H^{-1} S^{-1} H^{-1} Z^{\mathrm{T}} W_l$$
$$\times \left\{ Y - c_m - \sum_{\nu=0}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} Z e_\nu - \sum_{1\leqslant\beta\leqslant d, \beta\neq\alpha} m_\beta(Y_{l-\beta}) \right\}$$

which, by using (4.6), (4.7) and the definition (4.2), is equal to

$$(n-d+1)^{-1} \sum_{l=d}^{n} \frac{1}{\varphi(x_\alpha, \overline{Y}_l)} (n-d+1)^{-1} \sum_{j=d}^{n} K_{0h}^{*}(Y_{j-\alpha} - x_\alpha) L_g(\overline{Y}_j - \overline{Y}_l)$$
$$\times \left[m_\alpha(Y_{j-\alpha}) - \sum_{\nu=0}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} (Y_{j-\alpha} - x_\alpha)^\nu \right.$$
$$\left. + \sum_{1\leqslant\beta\leqslant d, \beta\neq\alpha} \{ m_\beta(Y_{j-\beta}) - m_\beta(Y_{l-\beta}) \} + s(X_j)\xi_j \right]$$
$$= (n-d+1)^{-1} \sum_{j=d}^{n} \{1 + \mathrm{o_p}(1)\} \int dw \frac{K_{0h}^{*}(Y_{j-\alpha} - x_\alpha)}{\varphi(x_\alpha, \overline{Y}_j - gw)} \overline{\varphi}(\overline{Y}_j - gw) L(w)$$
$$\times \left[m_\alpha(Y_{j-\alpha}) - \sum_{\nu=0}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} (Y_{j-\alpha} - x_\alpha)^\nu \right.$$
$$\left. + \sum_{1\leqslant\beta\leqslant d, \beta\neq\alpha} \{ m_\beta(Y_{j-\beta}) - m_\beta(Y_{j-\beta} - gw_\beta) \} + s(X_j)\xi_j \right].$$

And because L has order q, so the above is equal to

$$(n-d+1)^{-1} \sum_{j=d}^{n} \{1 + \mathrm{o_p}(1)\} \frac{K_{0h}^{*}(Y_{j-\alpha} - x_\alpha)}{\varphi(x_\alpha, \overline{Y}_j)} \overline{\varphi}(\overline{Y}_j)$$
$$\times \left\{ m_\alpha(Y_{j-\alpha}) - \sum_{\nu=0}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} (Y_{j-\alpha} - x_\alpha)^\nu + s(X_j)\xi_j \right\} + \mathrm{O_p}(g^q). \quad (4.10)$$

© Blackwell Publishers Ltd 1999

Thus we have shown that

$$\hat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha) = B + V + o_p(h^{p+1})$$

in which

$$B = (n-d+1)^{-1} \sum_{j=d}^{n} \frac{K_{0h}^*(Y_{j-\alpha} - x_\alpha)}{\varphi(x_\alpha, \overline{Y}_j)} \overline{\varphi}(\overline{Y}_j) \times \left\{ m_\alpha(Y_{j-\alpha}) - \sum_{\nu=0}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} (Y_{j-\alpha} - x_\alpha)^\nu \right\}$$

and

$$V = (n-d+1)^{-1} \sum_{j=d}^{n} \frac{K_{0h}^*(Y_{j-\alpha} - x_\alpha)}{\varphi(x_\alpha, \overline{Y}_j)} \overline{\varphi}(\overline{Y}_j) \{s(X_j)\xi_j\}.$$

Now (by using the mixing properties of our process)

$$B = \{1 + o_p(1)\} \int \frac{K_{0h}^*(z - x_\alpha)}{\varphi(x_\alpha, w} \overline{\varphi}(w) \times \left\{ m_\alpha(Y_{j-\alpha}) - \sum_{\nu=0}^{p} \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} (Y_{j-\alpha} - x_\alpha)^\nu \right\} \varphi(z, w)\, dz\, dw.$$

After substituting $z = x_\alpha + hu$, B becomes

$$B = \{1 + o_p(1)\} \int \frac{K_0^*(u)}{\varphi(x_\alpha, w)} \overline{\varphi}(w) \times \left\{ m_\alpha(x_\alpha + hu) - \sum_{\nu=0}^{p} \frac{1}{\nu!} m_\alpha^{(\nu)}(x_\alpha)(hu)^\nu \right\} \varphi(x_\alpha + hu, w)\, du\, dw$$

which, by using the moment properties of the equivalent kernel as in (4.3), equals

$$\{1 + o_p(1)\} \frac{\mu_{p+1}(K_0^*)}{(p+1)!} m_\alpha^{(p+1)}(x_\alpha) b_{m\alpha}(x_\alpha) h^{p+1} = b_{m\alpha}(x_\alpha) h^{p+1} + o_p(h^{p+1}) \tag{4.11}$$

where $b_{m\alpha}(x_\alpha)$ is as given in Theorem 1. Meanwhile, V has mean zero and its variance is

$$(n-d+1)^{-1} \int \left\{ \frac{K_{0h}^*(z - x_\alpha)}{\varphi(x_\alpha, w)} \overline{\varphi}(w) s(z, w) \right\}^2 \varphi(z, w)\, dz\, dw \{1 + o(1)\}$$
$$= n^{-1} h^{-1} \sigma_{m\alpha}^2(x_\alpha) \{1 + o(1)\}. \tag{4.12}$$

© Blackwell Publishers Ltd 1999

Equations (4.11) and (4.12) together establish (2.1). Equation (2.2) is derived by standard techniques as in Linton and Härdle (1996). Equation (2.3) and all the remaining formulas of Theorem 1 then follow directly from (2.1) and (2.2) as the various $(nh)^{1/2}\{\hat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha)\}$ are all asymptotically uncorrelated and so the variance of $(nh)^{1/2}\{\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x})\}$ is simply the sum of all their variances, and the mean of $(nh)^{1/2}\{\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x})\}$ is simply the sum of all their means. ∎

Proof of Theorem 2. We prove similar results for $\hat{V}_\alpha(x_\alpha)$:

$$\begin{aligned}
\hat{V}_\alpha(x_\alpha) - V_\alpha(x_\alpha) &= (n-d+1)^{-1}\sum_{l=d}^{n}\{e_0^{\mathrm{T}}(Z^{\mathrm{T}}W_lZ)^{-1}Z^{\mathrm{T}}W_lY^2 \\
&\quad - \hat{m}(x_\alpha, \hat{Y}_l)^2\} - V_\alpha(x_\alpha) \\
&= (n-d+1)^{-1}\sum_{l=d}^{n} e_0^{\mathrm{T}}(Z^{\mathrm{T}}W_lZ)^{-1}Z^{\mathrm{T}}W_l \\
&\quad \times \{Y^2 - \hat{m}(x_\alpha, \overline{Y}_l)^2 - V_\alpha(x_\alpha)\} \\
&= (n-d+1)^{-1}\sum_{l=d}^{n} e_0^{\mathrm{T}}(Z^{\mathrm{T}}W_lZ)^{-1}Z^{\mathrm{T}}W_l \\
&\quad \times \{Y^2 - m(x_\alpha, \overline{Y}_l)^2 + m(x_\alpha, \overline{Y}_l)^2 - \hat{m}(x_\alpha, \overline{Y}_l)^2 - V_\alpha(x_\alpha)\}.
\end{aligned}$$

Now note that by assumption (A4)

$$(n-d+1)^{-1}\sum_{j=d}^{n}\prod_{\beta\neq\alpha} v_\beta(Y_{j-\beta}) = 1 + \mathrm{O}_\mathrm{p}\left(\frac{1}{n^{1/2}}\right)$$

and also that

$$Y_j^2 = m(X_j)^2 + 22m(X_j)s(X_j)\xi_j + v(X_j)(\xi_j^2 - 1) + v(X_j).$$

So similar to (4.10) we have

$$\hat{V}_\alpha(x_\alpha) - V_\alpha(x_\alpha) = T_1 + T_2 + T_3 + T_4 + T_5 + \mathrm{o}_\mathrm{p}(h^{p+1})$$

where

$$T_1 = (n-d+1)^{-1}\sum_{l=d}^{n}\{m(x_\alpha, \overline{Y}_l)^2 - \hat{m}(x_\alpha, \overline{Y}_l)^2\}$$

$$T_2 = (n-d+1)^{-1}\sum_{j=d}^{n}\frac{K_{0h}^*(Y_{j-\alpha} - x_\alpha)}{\varphi(x_\alpha, \overline{Y}_j)}\overline{\varphi}(\overline{Y}_j)\{m(X_j)^2 - m(x_\alpha, \overline{Y}_j)^2\}$$

© Blackwell Publishers Ltd 1999

$$T_3 = (n-d+1)^{-1} \sum_{j=d}^{n} \frac{K_{0h}^*(Y_{j-\alpha} - x_\alpha)}{\varphi(x_\alpha, \overline{Y}_j)} \overline{\varphi}(\overline{Y}_j) \left\{ v(X_j) - V_\alpha(x_\alpha) \prod_{\beta \neq \alpha} v_\beta(Y_{j-\beta}) \right\}$$

$$T_4 = (n-d+1)^{-1} \sum_{j=d}^{n} \frac{K_{0h}^*(Y_{j-\alpha} - x_\alpha)}{\varphi(x_\alpha, \overline{Y}_j)} \overline{\varphi}(\overline{Y}_j) \{2m(X_j)s(X_j)\xi_j\}$$

$$T_5 = (n-d+1)^{-1} \sum_{j=d}^{n} \frac{K_{0h}^*(Y_{j-\alpha} - x_\alpha)}{\varphi(x_\alpha, \overline{Y}_j)} \overline{\varphi}(\overline{Y}_j) \{v(X_j)(\xi_j^2 - 1)\}.$$

We derive the asymptotics of each of these terms. Recall that Theorem 1 provides the following:

$$(nh)^{1/2}\{\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x}) - h^{p+1} b_m(\boldsymbol{x})\} \xrightarrow{\mathrm{D}} \mathrm{N}\{0, \sigma_m^2(\boldsymbol{x})\}.$$

Therefore

$$T_1 = -(n-d+1)^{-1} \sum_{l=d}^{n} 2\{m(x_\alpha, \overline{Y}_l) - \hat{m}(x_\alpha, \overline{Y}_l)\} m(x_\alpha, \overline{Y}_l) + \mathrm{o_p}(h^{p+1})$$

$$= -2E\{m(x_\alpha, \overline{Y}_n) - \hat{m}(x_\alpha, \overline{Y}_n)\} m(x_\alpha, \overline{Y}_n) + \mathrm{o_p}(h^{p+1})$$

$$= -h^{p+1} \int 2b_m(x_\alpha, w) m(x_\alpha, w) \overline{\varphi}(w)\, dw + \mathrm{o_p}(h^{p+1}). \tag{4.13}$$

Next we see, by using the substitution $z_1 = x_\alpha + hu$, that

$$T_2 = \{1 + \mathrm{o_p}(1)\} \int \frac{K_{0h}^*(z - x_\alpha)}{\varphi(x_\alpha, w)} \overline{\varphi}(w) \{m(z, w)^2 - m(x_\alpha, w)^2\} \varphi(z, w)\, dz\, dw$$

$$= \frac{\mu_{p+1}(K_0^*)}{(p+1)!} \int 2m_\alpha^{(p+1)}(\xi_\alpha) m(x_\alpha, w) \overline{\varphi}(w)\, dw + \mathrm{o_p}(h^{p+1})$$

$$= \frac{2\mu_{p+1}(K_0^*)}{(p+1)!} m_\alpha^{(p+1)}(x_\alpha) M(x_\alpha) + \mathrm{o_p}(h^{p+1}) \tag{4.14}$$

$$T_3 = \{1 + \mathrm{o_p}(1)\} \int \frac{K_{0h}^*(z - x_\alpha)}{\varphi(x_\alpha, w)} \overline{\varphi}(w) \{V_\alpha(z)\overline{V}_\alpha(w) - V_\alpha(x_\alpha)\overline{V}_\alpha(w)\} \varphi(z, w)\, dz\, dw$$

$$= \frac{\mu_{p+1}(K_0^*)}{(p+1)!} \int V_\alpha^{(p+1)}(x_\alpha) \overline{V}_\alpha(w) \overline{\varphi}(w)\, dw + \mathrm{o_p}(h^{p+1})$$

$$= \frac{\mu_{p+1}(K_0^*)}{(p+1)!} V_\alpha^{(p+1)}(x_\alpha) + \mathrm{o_p}(h^{p+1}). \tag{4.15}$$

© Blackwell Publishers Ltd 1999

To calculate the terms T_4 and T_5, note first that they both have mean zero and are uncorrelated, so it is only necessary to calculate their variances and the sum.

$$\operatorname{var}(T_4) = (n-d+1)^{-1} E\left\{ \frac{K_{0h}^*(Y_{n-\alpha} - x_\alpha)\overline{\varphi}(\overline{Y}_n)}{\varphi(x_\alpha, \overline{Y}_n)} 2m(X_d)s(X_d) \right\}^2 \{1 + o(1)\}$$

$$= (n-d+1)^{-1} \int \left\{ \frac{K_{0h}^*(z - x_\alpha)}{\varphi(x_\alpha, w)} 2m(z, w)s(z, w)\overline{\varphi}(w) \right\}^2$$

$$\times \varphi(z, w)\, dz\, dw\{1 + o(1)\}$$

$$= \frac{1}{nh} \|K_0^*\|_2^2 \int \frac{4m^2 v}{\varphi}(x_\alpha, w)\overline{\varphi}^2(w)\, dw\{1 + o(1)\} \tag{4.16}$$

and similarly

$$\operatorname{var}(T_5) = \frac{1}{nh} \|K_0^*\|_2^2 \int \frac{m_4 v^2}{\varphi}(x_\alpha, w)\overline{\varphi}^2(w)\, dw\{1 + o(1)\}. \tag{4.17}$$

Putting together Equations (4.13) through (4.17) gives the asymptotic expressions of $\hat{V}_\alpha(x_\alpha)$ in Theorem 2. To get the formula for $c_{V\alpha}(x)$ in (2.5), note that the variance term V in the proof of Theorem 1 is uncorrelated to all the T_i except T_4, and the asymptotic correlation is (plus some higher order terms)

$$(n-d+1)^{-1} E\left\{ \frac{K_{0h}^*(Y_{d-\alpha} - x_\alpha)\overline{\varphi}(\overline{Y}_d)}{\varphi(x_\alpha, \overline{Y}_d)} 2m(X_d)s(X_d) \right\}$$

$$\times \left\{ \frac{K_{0h}^*(Y_{d-\alpha} - x_\alpha)\overline{\varphi}(\overline{Y}_d)}{\varphi(x_\alpha, \overline{Y}_d)} s(X_d) \right\}$$

which can be verified to be exactly $(1/nh)c_{V\alpha}(x)\{1 + o(1)\}$ by the same technique as that used above. Equation (2.6) is easy to prove as (2.2) of Theorem 1.

To get the asymptotic properties of $\hat{c}_v$, we use the above results on $\hat{V}_\alpha(x)$ and the mixing properties of our process to get

© Blackwell Publishers Ltd 1999

$$\hat{c}_v^{d-1} - c_v^{d-1} = \frac{1}{d}\sum_{\alpha=1}^{d}\frac{1}{n-d+1}\sum_{j=d}^{n}\prod_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\hat{V}_\beta(Y_{j-\beta}) - c_v^{d-1}$$

$$= \frac{1}{d}\sum_{\alpha=1}^{d}\frac{1}{n-d+1}\sum_{j=d}^{n}\prod_{1\leqslant\beta\leqslant d,\beta\neq\alpha}$$

$$\times \{V_\beta(Y_{j-\beta}) + \hat{V}_\beta(Y_{j-\beta}) - V_\beta(Y_{j-\beta})\} - c_v^{d-1}$$

$$= \frac{1}{d}\sum_{\alpha=1}^{d}\frac{1}{n-d+1}\sum_{j=d}^{n}\prod_{1\leqslant\beta\leqslant d,\beta\neq\alpha} V_\beta(Y_{j-\beta}) - c_v^{d-1}$$

$$+ \frac{1}{d}\sum_{\alpha=1}^{d}\frac{1}{n-d+1}\sum_{j=d}^{n}\sum_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\left\{\prod_{1\leqslant\gamma\leqslant d,\gamma\neq\alpha,\beta} V_\gamma(Y_{j-\gamma})\right\}$$

$$\times \{\hat{V}_\beta(Y_{j-\beta}) - V_\beta(Y_{j-\beta})\}$$

$$= \frac{1}{d}\sum_{\alpha=1}^{d}\frac{1}{n-d+1}\sum_{j=d}^{n}\sum_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\left\{\prod_{1\leqslant\gamma\leqslant d,\gamma\neq\alpha,\beta} V_\gamma(Y_{j-\gamma})\right\}$$

$$\times \{\hat{V}_\beta(Y_{j-\beta}) - V_\beta(Y_{j-\beta})\} + \mathrm{O_p}\left(\frac{1}{n^{1/2}}\right)$$

$$= S_1 + S_2 + S_3 + \mathrm{o_p}(h^{p+1})$$

where

$$S_1 = \frac{1}{d}\sum_{\alpha=1}^{d}\frac{1}{n-d+1}\sum_{j=d}^{n}\sum_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\left\{\prod_{1\leqslant\gamma\leqslant d,\gamma\neq\alpha,\beta} V_\gamma(Y_{j-\gamma})\right\} b_{v\beta}(Y_{j-\beta})h^{p+1}$$

$$S_2 = \frac{1}{d}\sum_{\alpha=1}^{d}\frac{1}{(n-d+1)^2}\sum_{j=d}^{n}\sum_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\left\{\prod_{1\leqslant\gamma\leqslant d,\gamma\neq\alpha,\beta} V_\gamma(Y_{j-\gamma})\right\}$$

$$\times \left[\sum_{k=d}^{n}\frac{K_{0h}^*(Y_{k-\beta} - Y_{j-\beta})}{\varphi(Y_{j-\beta}, \overline{Y}_k)}\overline{\varphi}(\overline{Y}_k)\{2m(X_k)s(X_k)\xi_k\}\right]$$

$$S_3 = \frac{1}{d}\sum_{\alpha=1}^{d}\frac{1}{(n-d+1)^2}\sum_{j=d}^{n}\sum_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\left\{\prod_{1\leqslant\gamma\leqslant d,\gamma\neq\alpha,\beta} V_\gamma(Y_{j-\gamma})\right\}$$

$$\times \left[\sum_{k=d}^{n}\frac{K_{0h}^*(Y_{k-\beta} - Y_{j-\beta})}{\varphi(Y_{j-\beta}, \overline{Y}_k)}\overline{\varphi}(\overline{Y}_k)\{v(X_k)(\xi_k^2 - 1)\}\right].$$

© Blackwell Publishers Ltd 1999

These three terms can be written as (again using the mixing properties)

$$S_1 = \frac{h^{p+1}}{d}\sum_{\alpha=1}^{d}\int \sum_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\left\{\prod_{1\leqslant\gamma\leqslant d,\gamma\neq\alpha,\beta} V_\gamma(y_\gamma)\right\} b_{v\beta}(y_\beta)\varphi(y)\,dy + \mathrm{O_p}\left(\frac{1}{n^{1/2}}\right)$$

and

$$S_2 = \sum_{k=d}^{n}\frac{2m(X_k)s(X_k)\xi_k}{n-d+1}\,\frac{1}{d}\sum_{\alpha=1}^{d}\sum_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\int\left\{\prod_{1\leqslant\gamma\leqslant d,\gamma\neq\alpha,\beta} V_\gamma(y_\gamma)\right\}$$

$$\times\frac{K_{0h}^{*}(Y_{k-\beta}-y_\beta)}{\varphi(y_\beta,\,\overline{Y}_k)}\overline{\varphi}(\overline{Y}_k)\varphi(y)\,dy\{1+\mathrm{o_p}(1)\}$$

$$= \sum_{k=d}^{n}\frac{2m(X_k)s(X_k)\xi_k}{n-d+1}\,\frac{1}{d}\sum_{\alpha=1}^{d}\sum_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\int\left\{\prod_{1\leqslant\gamma\leqslant d,\gamma\neq\alpha,\beta} V_\gamma(y_\gamma)\right\}$$

$$\times\frac{K_0^{*}(u)\overline{\varphi}(\overline{Y}_k)\varphi(Y_{k-\beta}-hu,\,\overline{y})\,du\,d\overline{y}}{\varphi(Y_{k-\beta}-hu,\,\overline{Y}_k)}\{1+\mathrm{o_p}(1)\}$$

$$= \sum_{k=d}^{n}\frac{2m(X_k)s(X_k)\xi_k}{(n-d+1)\varphi(Y_k)}\,\frac{1}{d}\sum_{\alpha=1}^{d}\sum_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\int\left\{\prod_{1\leqslant\gamma\leqslant d,\gamma\neq\alpha,\beta} V_\gamma(y_\gamma)\right\}$$

$$\times\,\overline{\varphi}(\overline{Y}_k)\varphi(Y_{k-\beta},\,\overline{y})\,d\overline{y}\{1+\mathrm{o_p}(1)\}$$

from which it is clear that S_2 satisfies a central limit theorem with $n^{1/2}$ rate of convergence, which is also the case for S_3. Thus

$$\hat{c}_v = \left[c_v^{d-1} + \frac{h^{p+1}}{d}\sum_{\alpha=1}^{d}\int\sum_{1\leqslant\beta\leqslant d,\beta\neq\alpha}\left\{\prod_{1\leqslant\gamma\leqslant d,\gamma\neq\alpha,\beta} V_\gamma(y_\gamma)\right\} b_{v\beta}(y_\beta)\varphi(y)\,dy + \frac{1}{n^{1/2}}Z\right]^{1/(d-1)}$$

where $Z \xrightarrow{\mathrm{D}} \mathrm{N}(0,\sigma^2)$ for some σ^2; applying the Taylor expansion gives the result on $\hat{c}_v$ and the rest of Theorem 2 follows directly. ■

PROOF OF THEOREM 3. We simply put together the results of the previous two theorems. Note that the joint normality follows from the fact that the stochastic part of all the estimates is based on the ξ_j and the $\xi_j^2 - 1$. Thus, any linear combinations of the estimates also have similar forms to those treated in Theorem 1. ■

© Blackwell Publishers Ltd 1999

ACKNOWLEDGEMENTS

This research was financially supported by Sonderforschungsbereich 373 'Quantifikation und Simulation Ökonomischer Prozesse' Deutsche Forschungsgemeinschaft, at Humboldt Universität zu Berlin. We appreciate the interests of our colleagues that motivated our work, especially Christian Hafner, Helmut Lütkepohl and Rolf Tschernig. We have also benefited from the elegant formulation of the marginal integration technique by Eric Severance-Lossin and Stefan Sperlich. Finally, we want to thank two anonymous referees who gave us many constructive comments.

REFERENCES

ANGO NZE, P. (1992) Critères d'ergodicité de quelques modèles à représentation markovienne. *C.R. Acad. Sci. Paris, Ser. I* 315, 1301–4.

AUESTAD, B. and TJØSTHEIM, D. (1991) Functional identification in nonlinear time series. In *Nonparametric Functional Estimation and Related Topics* (ed. G. G. Roussas). Amsterdam: Kluwer Academic, pp. 493–507.

BOSSAERTS, P., HÄRDLE, W. and HAFNER, C. (1996) Foreign exchange-rates have surprising volatility. In *Athens Conference on Applied Probability and Time Series*, Vol. 2 (ed. P. M. Robinson). Lecture Notes in Statistics 115, New York: Springer, pp. 55–72.

CHAN, K. S. and TONG, H. (1986) On estimating thresholds in autoregressive models. *J. Time Ser. Anal.* 7, 179–90.

CHEN, R. and TSAY, R. S. (1993a) Nonlinear additive ARX models. *J. Am. Stat. Assoc.* 88, 955–67.

—— and —— (1993b) Functional-coefficient autoregressive models. *J. Am. Stat. Assoc.* 88, 298–308.

DAVYDOV, YU. A. (1973) Mixing conditions for Markov chains. *Theory Probab. Appl.* 18, 312–28.

DIEBOLT, J. and GUÉGAN, D. (1993) Tail behaviour of the stationary density of general nonlinear autoregressive processes of order one. *J. Appl. Probab.* 30, 315–29.

DROST, F. C. and NIJMAN, T. E. (1993) Temporal aggregation of GARCH processes. *Econometrica* 61, 909–27.

ENGLE, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1008.

—— and GONZALEZ-RIVERA, G. (1991) Semiparametric ARCH models. *J. Bus. Econ. Stat.* 9, 345–60.

—— and NG, V. (1993) Measuring and testing the impact of news on volatility. *J. Financ.* 48, 1749–78.

FAN, J. and GIJBELS, I. (1996) *Local Polynomial Modelling and its Applications*. London: Chapman and Hall.

GOURIÉROUX, CH. and MONFORT, A. (1992) Qualitative threshold ARCH models. *J. Economet.* 52, 159–99.

GRANGER, C. and TERÄSVIRTA, T. (1993) *Modelling Nonlinear Dynamic Relationships*. Oxford: Oxford University Press.

HAGGAN, V. and OZAKI, T. (1981) Modelling nonlinear vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* 68, 189–96.

HÄRDLE, W. and CHEN, R. (1995) Nonparametric time series analysis, a selective review with examples. In *Proceedings of the 50th Session of the ISI*, Peking.

—— and TSYBAKOV, A. B. (1997) Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Economet.* 81, 223–42.

——, KLINKE, S. and TURLACH, B. (1995) *XploRe—an Interactive Statistical Computing Environment*. Heidelberg: Springer.

——, TSYBAKOV, A. B. and YANG, L. (1998) Nonparametric vector autoregression. *J. Stat. Plann. Inference*, 68, 221–245.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990) *Generalized Additive Models*. Mongraphs on Statistics and Applied Probability 43, London: Chapman and Hall.

© Blackwell Publishers Ltd 1999

KATKOVNIK, V. YA. (1979) Linear and nonlinear methods of nonparametric regression analysis. *Automatika*, 35–46.
LINTON, O. B. and HÄRDLE, W. (1996) Estimation of additive regression models with known links. *Biometrika* 83, 529–40.
—— and NIELSEN, J. P. (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82, 93–100.
LIPTSER, R. SH. and SHIRJAEV, A. N. (1980) A functional central limit theorem for martingales. *Theory Probab. Appl.* 25, 667–88.
MASRY, E. and TJØSTHEIM, D. (1995a) Non-parametric estimation and identification of ARCH nonlinear time series: strong convergence and asymptotic normality. *Economet. Theory* 11, 258–89.
—— and —— (1995b) Additive nonlinear ARX time series and projection estimates. *Economet. Theory*, forthcoming.
MEESE, R. A. and ROSE, A. (1991) An empirical assessment of non-linearities in models of exchange rate determination. *Rev. Econ. Stud.* 58, 601–19.
MOKKADEM, A. (1987) Sur un modèle autorégressif nonlinéaire. Ergodicité et ergodicité géometrique. *J. Time Ser. Anal.* 8, 195–204.
NELSON, D. B. (1991) Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* 59, 347–70.
NUMMELIN, E. and TUOMINEN, P. (1982) Geometric ergodicity of Harris-recurrent Markov chains with application to renewal theory. *Stochastic Processes Appl.* 12, 187–202.
ROBINSON, P. M. (1983) Nonparametric estimators for time series. *J. Time Ser. Anal.* 4, 185–207.
—— (1984) Robust nonparametric autoregression. In *Robust and Nonlinear Time Series Analysis* (eds J. Franke, W. Härdle and D. Martin). Lecture Notes in Statistics 26, Heidelberg: Springer.
RUPPERT, D. and WAND, M. P. (1994) Multivariate locally weighted least squares regression. *Ann. Stat.* 22, 1346–70.
SEVERANCE-LOSSIN, E. and SPERLICH, S. (1995) Estimation of derivatives for additive separable models. SFB 373 Discussion Paper 60, Humboldt Universität zu Berlin; available at http://www.wiwi.hu-berlin.de/pub/papers/sfb/dpsfb960060.ps.Z.
STONE, C. J. (1977) Consistent nonparametric regression. *Ann. Stat.* 5, 595–645.
—— (1982) Optimal global rates of convergence for nonparametric regression. *Ann. Stat.* 10, 1040–53.
TJØSTHEIM, D. (1990) Nonlinear time series and Markov chains. *Adv. Appl. Probab.* 22, 587–611.
—— and AUESTAD, B. (1994) Nonparametric identification of nonlinear time series: projections. *J. Am. Stat. Assoc.* 89, 1398–1409.
TONG, H. (1978) On a threshold model. In *Pattern Recognition and Signal Processing* (ed. C. H. Chen). Amsterdam: Sijthoff and Noordholf.
—— (1983) *Threshold Models in Nonlinear Time Series Analysis*. Lecture Notes in Statistics 21, Heidelberg: Springer.
TSCHERNIG, R. and YANG, L. (1999) Nonparametric lag selection for time series. *Journal of Time Series Analysis, forthcoming.*
TSYBAKOV, A. B. (1986) Robust reconstruction of functions by the local-approximation method. *Prob. Inf. Transm.* 22, 133–46.
TWEEDIE, R. L. (1975) Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Processes Appl.* 3, 385–403.
WAND, M. P. and JONES, M. C. (1995) *Kernel Smoothing*. London: Chapman and Hall.

© Blackwell Publishers Ltd 1999