Nonparametric Autoregression with Multiplicative Volatility and Additive Mean *

Lijian Yang

Department of Statistics and Probability
Michigan State University
East Lansing, Michigan 48824
U. S. A.

Wolfgang Härdle
Institut für Statistik und Ökonometrie
Humboldt Universität zu Berlin
Spandauer Str.1, D-10178 Berlin
Germany

Jens P. Nielsen
PFA Pension
Sundkrogsgade 4
2100 Copenhagen
Denmark

December 4, 1998

Abstract

For over a decade, nonparametric modelling has been successfully applied to study 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 one, 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 DEM/USD (Deutsche Mark/US Dollar) daily exchange returns.

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 ARCH models the volatility or the scale of innovative random shocks is a function of past values. Over the past fifteen 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), Drost and Nijman (1993), Engle and Gonzalez-Rivera (1991). A more recent review is Härdle and Chen (1995).

One of the models studied for foreign exchange rates, for example, is the CHARN (conditional heteroskedastic autoregressive nonlinear) model with one lag (Bossaerts, Härdle, and Hafner, 1996)

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

where $\{\xi_i\}_{i \geq 1}$ are 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 $\{\xi_i\}$'s. An analysis of the estimated residuals still revealed autocorrelation. Hence, more than one lagged variable in the modelling of the mean function $m(\bullet)$ and the scale function $s(\bullet)$ seems to be the necessary step in a further analysis.

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

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

where $\{\xi_i\}_{i \geq 1}$ are as in (1.1) and $Y_0, Y_1, ..., Y_{d-1}$ are random variables independent of the $\{\xi_i\}$'s. The conditional volatility function is $v(Y_{i-1}, Y_{i-2}, ..., Y_{i-d}) = s^2(Y_{i-1}, Y_{i-2}, ..., Y_{i-d})$. 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 $m(\bullet)$ and $v(\bullet)$ has been treated in Härdle and Tsybakov (1997) in the case of $d = 1$ with the local polynomial regression method. Härdle, Tsybakov and Yang (1998) studied vector autoregression with arbitrary number of lags and 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 size case. We illustrate the method with a foreign

1

exchange rate application. Through lag selection, see Tschernig and Yang (1997), 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 $m(\bullet)$ is a sum of univariate functions, then the one dimensional convergence rate can be achieved for the estimation of $m(\bullet)$'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,b) 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 Auestad (1994) for the precise analysis of additive model estimators which was previously unavailable. It provides closed form bias and variance expressions of the one dimensional function estimator. The term marginal integration was introduced in Linton and Nielsen (1995), who worked in the independent identically distributed regression setting. Marginal integration has recently been employed in the autoregression setting by Masry and Tjøstheim (1995a,b) and in the independent identically distributed regression setting by Linton and Härdle (1996) and Severance-Lossin and Sperlich (1995).

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

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

then

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

where $F(x_1, ..., \widehat{x_\beta}, ..., x_d)$ is the joint distribution function of all the variables $X_1, ..., X_d$ with the β-th X_β removed, and C is an additive constant. Hence each component function m_β is identified from $m(x_1, x_2, ..., x_d)$ through a simple integration procedure. Linton and Nielsen (1995) introduced the idea of applying integration estimation to multiplicative structures in dimension two, 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 $m(\bullet)$ and the conditional variance or volatility function $v(\bullet)$ at the same time. The flexibility of our CHARN model is important in a number of economic applications. For example 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 $m(\bullet)$ and $v(\bullet)$ for model (1.2). The volatility function $v(\bullet)$ 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 $m(\bullet)$ is additive while the volatility function $v(Y_{i-1}, Y_{i-2}, ..., Y_{i-d}) = s(Y_{i-1}, Y_{i-2}, ..., Y_{i-d})^2$ is multiplicative

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

2

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

where c_m and c_v are constants, $\{m_\beta(\bullet)\}_{\beta=1}^{d}$ and $\{v_\beta(\bullet)\}_{\beta=1}^{d}$ are sets of unknown functions. Besides the better rate of convergence for the estimation of $\{m_\beta(\bullet)\}_{\beta=1}^{d}$ and $\{v_\beta(\bullet)\}_{\beta=1}^{d}$ as discussed above, these univariate functions also allows one 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 $\{m_\beta(\bullet)\}_{\beta=1}^{d}$ and $\{v_\beta(\bullet)\}_{\beta=1}^{d}$, the process Y_i has to converge to a stationary distribution. If we denote by $\mathbf{X}_i$ the vector $(Y_{i-1}, Y_{i-2}, ..., Y_{i-d})^T$, then $\{\mathbf{X}_i\}$ 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) developed geometric ergodicity criteria for Markov processes. Here we state some general assumptions

A1: *The random variable ξ_i has a density function $p(\bullet)$. This density $p(\bullet)$ and the volatility function $v(\bullet)$ 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 $m(\bullet)$ and $s(\bullet)$ satisfy:*

$$|m(y_{i-1}, y_{i-2}, ..., y_{i-d})| \leq C_1 (1 + \sum_{\beta=1}^{d} |y_{i-\beta}|)$$

$$|s(y_{i-1}, y_{i-2}, ..., y_{i-d})| \leq C_2 (1 + \sum_{\beta=1}^{d} |y_{i-\beta}|)$$

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 out or exploding. Ango Nze (1992) proved the following

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

$$\|P^n(\bullet \mid \mathbf{x}) - \pi(\bullet)\|_{TV} = O(\rho^n)$$

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

This lemma ensures that the process $\{\mathbf{X}_i\}$ is asymptotically stationary. We denote by $F(\bullet)$ the stationary distribution function. For all $1 \leq \alpha \leq d$, we denote by $F_\alpha(\bullet)$ the stationary distribution function of the α-th variable, and $\overline{F}(\bullet)$ the stationary distribution function with the α-th 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 $F_\beta(\bullet)$, and for all $1 \leq \beta \leq d$;*

3

A4: $E \prod_{1 \le \beta \le d, \beta \ne \alpha} v_\beta(Y_\beta) = \prod_{1 \le \beta \le d, \beta \ne \alpha} v_\beta(x_\beta) d\overline{F}(\overline{x}) = 1$ *for any* $(Y_1, Y_2, ...Y_d)$ *that has distribution* $F(\bullet)$, *and for all* $1 \le \alpha \le d$.

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

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

In what follows, we adopt the notation $\mathbf{X_i} = (Y_{i-\alpha}, \overline{Y}_i)$ to highlight a particular direction of interest $Y_{i-\alpha}$, for all $1 \le \alpha \le d$, while $\overline{Y}_i$ is the $d-1$ dimensional vector that consists of all the rest $Y_{i-\beta}$'s, $1 \le \beta \le d, \beta \ne \alpha$. Assumptions A3 and A4 yield the following marginal integration formulae for the unknown functions

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

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

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

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

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

In Section 2, we present the estimators of $\{m_\beta(\bullet)\}_{\beta=1}^{d}$ and $\{v_\beta(\bullet)\}_{\beta=1}^{d}$ and study their asymptotic properties. In Section 3, we discuss the application of the result to DM/USD 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: $Y_i = m(X_{i1}, X_{i2}, ..., X_{id}) + s(X_{i1}, X_{i2}, ..., X_{id})\xi_i$, where ξ_i are as in (1.2), $(X_{i1}, X_{i2}, ..., X_{id}, Y_i)$ are i.i.d., and the design points $\{X_{i1}, X_{i2}, ..., X_{id}\}$ 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 Tsybakov (1986), Ruppert and Wand (1994), Wand and Jones (1995) and 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 $K(\bullet)$ we denote $\|K\|_2^2 = \int K^2(u) du$, while for a kernel function $K(\bullet)$ 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 $K(\bullet)$ and $L(\bullet)$ that satisfy

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

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

We assume the following about the functions involved in the estimation

A6: *The functions $m_\alpha(\bullet)$'s and $v_\alpha(\bullet)$'s have bounded Lipschitz continuous $(p+1)$-th derivatives for all $1 \le \alpha \le d$.*

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

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

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

Note that this A8 requires that $L(\bullet)$ have the order as in A5. In particular, if one uses local linear regression, i.e., $p = 1$, then the order of $L(\bullet)$ is $q > d - 1$.

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

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

where $\widehat{m}(x_\alpha, \overline{x})$ is an estimate of $m(\bullet)$ at $(x_\alpha, \overline{x})$, and $\widehat{\overline{F}}(\overline{x})$ is the empirical cumulative distribution function (ecdf). The estimator $\widehat{M}_\alpha(x_\alpha)$ is hereby based on the sample version of equation (1.7). The estimator for c_m is simply the sample mean of Y_j's according to (1.9)

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

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

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

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

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

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

5

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

where we denote

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

for any vector

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

Also write

$$Y = (Y_i)_{d \leq i \leq n}, \; Y^2 = (Y_i^2)_{d \leq i \leq n},$$

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

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

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

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

.

Note that

$$E(Y_i^2 \mid X_i) = m^2(X_i) + v(X_i),$$

thus similar estimator for $V_\alpha(x_\alpha)$ based on equation (1.8) is defined as

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

and that of c_v is based on (1.9):

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

One then obtains estimators for $v_\alpha(x_\alpha)$ and $v(\mathbf{x})$ as the following:

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

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

Our first theorem gives the estimation result of the mean functions

6

Theorem 1 *Under assumptions A1-A8, as $n \to \infty$, for any α*

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

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 $n \to \infty$, one has

$$\text{cov}\left[\sqrt{nh}\left\{\widehat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha)\right\}, \sqrt{nh}\left\{\widehat{M}_\beta(x_\beta) - M_\beta(x_\beta)\right\}\right] \to 0. \tag{2.2}$$

Furthermore, as $n \to \infty$

$$\sqrt{n}(\widehat{c}_m - c_m) \xrightarrow{D} N\left\{0, \sigma^2_{cm}(\mathbf{x})\right\}$$

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

$$\sqrt{nh}\left\{\widehat{m}(\mathbf{x}) - m(\mathbf{x}) - h^{p+1}b_m(\mathbf{x})\right\} \xrightarrow{D} N\left\{0, \sigma^2_m(\mathbf{x})\right\} \tag{2.3}$$

where

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

and

$$\sigma^2_m(\mathbf{x}) = \sum_{\alpha=1}^d \sigma^2_{m\alpha}(x_\alpha).$$

The second theorem is about the estimation of the volatility functions

Theorem 2 *Under assumptions A1-A8, as $n \to \infty$, for any α*

$$\sqrt{nh}\left\{\widehat{V}_\alpha(x_\alpha) - V_\alpha(x_\alpha) - h^{p+1}b_{V\alpha}(x_\alpha)\right\} \xrightarrow{D} N\left\{0, \sigma^2_{V\alpha}(x_\alpha)\right\} \tag{2.4}$$

where

$$b_{V\alpha}(x_\alpha) = \frac{\mu_{p+1}(K_0^*)}{(p+1)!}\left\{V_\alpha^{(p+1)}(x_\alpha) + 2m_\alpha^{(p+1)}(x_\alpha)M(x_\alpha)\right\}$$

$$- \int 2b_m(x_\alpha, w)m(x_\alpha, w)\overline{\varphi}(w)dw$$

and

$$\sigma^2_{V\alpha}(x_\alpha) = \|K_0^*\|_2^2 \int \frac{v\left[m_4v + 4m^2\right]}{\varphi}(x_\alpha, w)\overline{\varphi}^2(w)dw.$$

Also, as $n \to \infty$

$$\text{cov}\left[\sqrt{nh}\left\{\widehat{V}_\alpha(x_\alpha) - V_\alpha(x_\alpha)\right\}, \sqrt{nh}\left\{\widehat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha)\right\}\right]$$

$$\to 2 \left\| K_0^* \right\|_2^2 \int \frac{vm}{\varphi}(x_\alpha, w) \overline{\varphi}^2(w) dw = c_{V\alpha}(x_\alpha) \tag{2.5}$$

while for any $\alpha \neq \beta$ one has

$$\operatorname{cov} \left[\sqrt{nh} \left\{ \widehat{V}_\alpha(x_\alpha) - V_\alpha(x_\alpha) \right\}, \sqrt{nh} \left\{ \widehat{V}_\beta(x_\beta) - V_\beta(x_\beta)) \right\} \right] \to 0,$$

$$\operatorname{cov} \left[\sqrt{nh} \left\{ \widehat{V}_\alpha(x_\alpha) - V_\alpha(x_\alpha) \right\}, \sqrt{nh} \left\{ \widehat{M}_\beta(x_\beta) - M_\beta(x_\beta)) \right\} \right] \to 0. \tag{2.6}$$

Furthermore

$$\sqrt{n}(\widehat{c}_v - c_v - b_c h^{p+1}) \overset{D}{\to} N(0, \sigma_{cv}^2)$$

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

$$\sqrt{nh} \left\{ \widehat{v}_\alpha(x_\alpha) - v_\alpha(x_\alpha) - h^{p+1} b_{v\alpha}(x_\alpha) \right\} \overset{D}{\to} N \left\{ 0, \sigma_{v\alpha}^2(x_\alpha) \right\} \tag{2.7}$$

where

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

and

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

while

$$\sqrt{nh} \left\{ \widehat{v}(\mathbf{x}) - v(\mathbf{x}) - h^{p+1} b_v(\mathbf{x}) \right\} \overset{D}{\to} N \left\{ 0, \sigma_v^2(\mathbf{x}) \right\}$$

where

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

and

$$\sigma_v^2(\mathbf{x}) = v^2(\mathbf{x}) \sum_{\beta=1}^d \frac{\sigma_{V\beta}^2(x_\beta)}{V_\beta^2(x_\beta)}.$$

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 $\mathbf{B}(\mathbf{x})$ the vector valued function*

$$\left\{ b_{m1}(x_1), b_{m2}(x_2), ..., b_{md}(x_d), b_m(\mathbf{x}), b_{v1}(x_1), b_{v2}(x_2), ..., b_{vd}(x_d), b_v(\mathbf{x}), 0, \sqrt{n}b_c \right\}^T$$

and $\mathbf{\Sigma}(\mathbf{x})$ the following matrix

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

where

$$\boldsymbol{\Sigma_{11}} = \mathrm{diag}\left\{\sigma_{m\alpha}^2(x_\alpha)\right\}_{\alpha=1}^d, \ \boldsymbol{\Sigma_{22}} = \sigma_m^2(\mathbf{x}), \ \boldsymbol{\Sigma_{33}} = \mathrm{diag}\left\{\sigma_{v\alpha}^2(x_\alpha)\right\}_{\alpha=1}^d, \ \boldsymbol{\Sigma_{44}} = \sigma_v^2(\mathbf{x}),$$

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

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

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

then, as $n \to \infty$

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

We comment here that although Theorem 3 is obtained for 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 "design-adaptive".

3 An Application

To illustrate our method with an example, we study the daily returns of the DEM/USD exchange rates from Jan.2 1980 to May 26 1986, a total of 1603 observations. The data is 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 (1997), 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 found that except for some boundary

effects, the mean functions $m_\beta(\bullet)$'s are very close to zero. The estimated volatility function $\hat{v}_\beta(\bullet)$'s depicted in Figures 2 and 3, however, provide some fresh insights. Both the computation and graphics are done in XploRe, see Härdle, Klinke and Turlach (1995).

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

Our observations about $\hat{v}_1(\bullet)$ and $\hat{v}_3(\bullet)$ 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 here has gone a step further in nonparametric estimation of times series as the significant lagged variables are first identified by a nonparametric criteria, see Tschernig and Yang (1997) for details. This example of identifying significant lags and measuring their impacts points to a new comprehensive nonparametric approach to time series analysis.
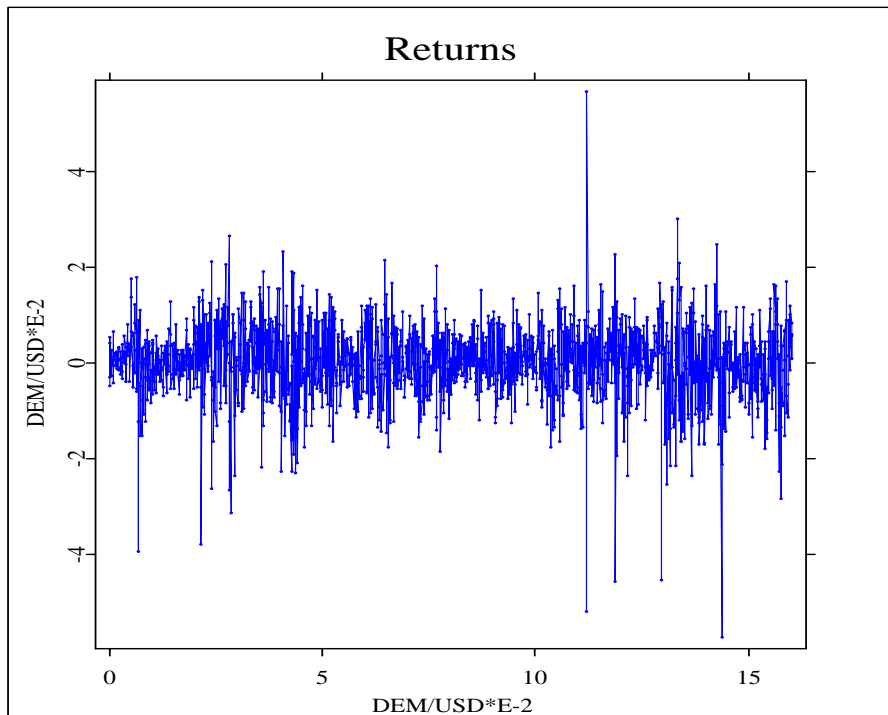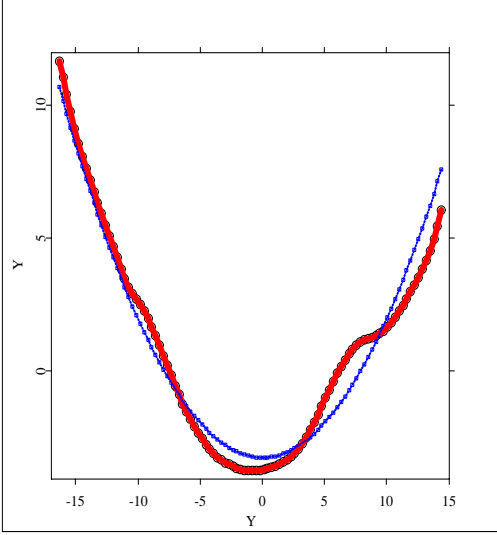

Figure 1: The daily returns

Figure 2: Volatility function $\hat{v}_1(\bullet)$ (thick) and its quadratic fit (thin)
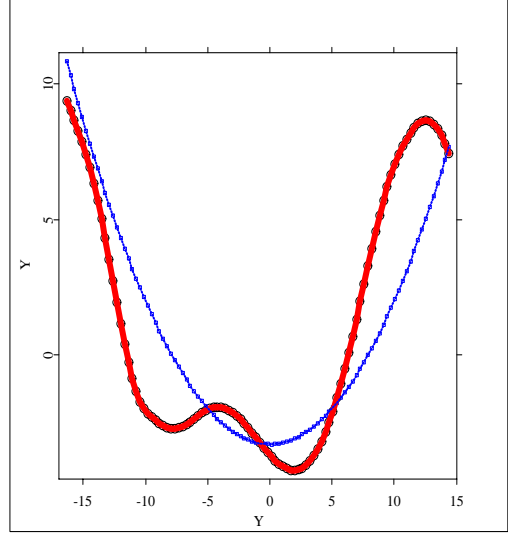

Figure 3: Volatility function $\hat{v}_3(\bullet)$ (thick) and its quadratic fit (thin)

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, if further, $\mathbf{X}_d$ is distributed with the stationary distribution $\pi(\bullet)$, then the process $\{\mathbf{X}_i\}$ is geometrically strongly mixing with the mixing coefficients satisfying $\alpha(n) \leq c_0 \rho_0^n$ for some $c_0 > 0$ and $0 < \rho_0 < 1$.

By arguments which are very similar to those used in Härdle, Tsybakov, and Yang (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 one now has 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 $D_l = \left(Z^T W_l Z \right)^{-1} - \frac{1}{\varphi(x_\alpha, \overline{Y}_l)} H^{-1} S^{-1} H^{-1}$

$$Cov(D_l, D_k) = \rho^{|l-k|} \left\{ O_p(h + \ln n/\sqrt{nhg^{d-1}}) \right\}^2 \tag{4.1}$$

uniformly in x_α and $\overline{Y}_l$, where $H = \text{diag}\left(h^\lambda \right)_{0 \leq \lambda \leq p}$.

11

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 $S = (\int u^{s+t} K(u) du)_{0 \leq s,t \leq p}$, which contains all the moments of S up to order $2p$. Denote $S^{-1} = (s_{st})_{0 \leq s,t \leq p}$ and define

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

This $K_\lambda^*(\bullet)$ is called the λ-th equivalent kernel. It has the following moments

$$\int u^q K_\lambda^*(u) \, du = \left\{ \begin{array}{ll} 0 & q \leq p, \quad q \neq \lambda \\ 1 & q = \lambda \\ \Lambda_\lambda & q = p+1 \end{array} \right\}. \tag{4.3}$$

and $K_0^*(\bullet)$ 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

$$e_0^T \left(Z^T W_l Z \right)^{-1} Z^T W_l Z e_\lambda = \left\{ \begin{array}{ll} 0 & 0 \neq \lambda \\ 1 & 0 = \lambda \end{array} \right\} \tag{4.4}$$

thus

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

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

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

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

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

Now Assumption A3 combined with the strong mixing properties of our process imply that for every $\beta = 1, 2, ..., d, \beta \neq \alpha$

$$(n-d+1)^{-1} \sum_{l=d}^{n} m_\beta(Y_{l-\beta}) = O_p(1/\sqrt{n}),$$

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

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

12

So one has

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

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

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

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

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

or

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

$$= (n-d+1)^{-1} \sum_{l=d}^n e_0^T \left(Z^T W_l Z \right)^{-1} Z^T W_l \left\{ Y - c_m - \sum_{\nu=0}^p \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} Z e_\nu - \sum_{1 \le \beta \le d, \beta \ne \alpha} m_\beta(Y_{l-\beta}) \right\}. \tag{4.5}$$

Note that the λ-th element of $Z^T W_l \left\{ Y - c_m - \sum_{\nu=0}^p \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} Z e_\nu - \sum_{1 \le \beta \le d, \beta \ne \alpha} m_\beta(Y_{l-\beta}) \right\}$ is

$$(n-d+1)^{-1} \sum_{j=d}^n (Y_{j-\alpha} - x_\alpha)^\lambda K_h(Y_{j-\alpha} - x_\alpha) L_g(\overline{Y}_j - \overline{Y}_l)$$

$$\left\{ Y_j - c_m - \sum_{\nu=0}^p \frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!} (Y_{j-\alpha} - x_\alpha)^\nu - \sum_{1 \le \beta \le d, \beta \ne \alpha} m_\beta(Y_{l-\beta}) \right\}$$

$$= I_{\lambda l,1} + I_{\lambda l,2} + I_{\lambda l,3}$$

in which

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

where

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

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

where

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

13

and

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

where

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

Lemma 4.3 *As* $n \to \infty$,

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

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

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

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

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

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

Proof. We only show this for the first case

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

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

$$E(I_{\lambda l_1 j_1,3} I_{\lambda l_2 j_2,3}) = \left\{ hg^{d-1} \right\}^{-1} \{1 + o(1)\}$$

$\int (hu_\alpha)^{2\lambda} K^2(u_\alpha) L^2(\overline{u}) v(x_\alpha + hu_\alpha, \overline{Y}_l + g\overline{u}) \varphi(x_\alpha + hu_\alpha, \overline{Y}_l + g\overline{u}) du$. Q. E. D.

Now

$$O(h^{2\lambda}/nhg^{d-1}) \left\{ O_p(h + \ln n/\sqrt{nhg^{d-1}}) \right\}^2 = O_p \left\{ h^{2\lambda+2}/nhg^{d-1} + h^{2\lambda} \ln^2 n/(n^2 h^2 g^{2(d-1)}) \right\}$$

$$= h^{2\lambda}/nh O_p \left(h^2/g^{d-1} + \ln^2 n/nhg^{2(d-1)} \right) = o_p \left(h^{2\lambda}/nh \right)$$

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

$$\sum_{\lambda=0}^{p} (n - d + 1)^{-1} \sum_{l=d}^{n} e_0^T \left\{ \left(Z^T W_l Z \right)^{-1} - \frac{1}{\varphi(x_\alpha, \overline{Y}_l)} H^{-1} S^{-1} H^{-1} \right\} e_\lambda (I_{\lambda l,1} + I_{\lambda l,2} + I_{\lambda l,3})$$

$$= \sum_{\lambda=0}^{p} h^{-\lambda} o_p \left(h^\lambda/\sqrt{nh} \right) = o_p \left(1/\sqrt{nh} \right) = o_p \left\{ h^{p+1} \right\} = o_p \left\{ n^{-(p+1)/(2p+3)} \right\}.$$

If we only had to consider the diagonal terms, then this fact is easily recongnised (this is if we could ignore the correlation of the "I"-terms with the rest). The correlation can however

14

be taken care of by writing up the $I_{\lambda l,k}'s$ as sums (se 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, one has in particular

$$\widehat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha) - o_p\left\{h^{p+1}\right\} =$$

$$= (n-d+1)^{-1}\sum_{l=d}^{n}\frac{1}{\varphi(x_\alpha,\overline{Y}_l)}e_0^T H^{-1}S^{-1}H^{-1}Z^T W_l$$

$$\left\{Y - c_m - \sum_{\nu=0}^{p}\frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!}Ze_\nu - \sum_{1\le\beta\le d,\beta\ne\alpha}m_\beta(Y_{l-\beta})\right\}$$

which, by using (4.6), (4.7), (4.7) and the definition (4.2), equals

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

$$\left[m_\alpha(Y_{j-\alpha}) - \sum_{\nu=0}^{p}\frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!}(Y_{j-\alpha}-x_\alpha)^\nu + \sum_{1\le\beta\le d,\beta\ne\alpha}\{m_\beta(Y_{j-\beta})-m_\beta(Y_{l-\beta})\} + s(X_j)\xi_j\right]$$

$$= (n-d+1)^{-1}\sum_{j=d}^{n}\{1+o_p(1)\}\int dw\frac{K_{0h}^*(Y_{j-\alpha}-x_\alpha)}{\varphi(x_\alpha,\overline{Y}_j-gw)}\overline{\varphi}(\overline{Y}_j-gw)L(w)$$

$$\left[m_\alpha(Y_{j-\alpha}) - \sum_{\nu=0}^{p}\frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!}(Y_{j-\alpha}-x_\alpha)^\nu + \sum_{1\le\beta\le d,\beta\ne\alpha}\{m_\beta(Y_{j-\beta})-m_\beta(Y_{j-\beta}-gw_\beta)\} + s(X_j)\xi_j\right].$$

And because L has order q, so the above equals

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

$$\left\{m_\alpha(Y_{j-\alpha}) - \sum_{\nu=0}^{p}\frac{m_\alpha^{(\nu)}(x_\alpha)}{\nu!}(Y_{j-\alpha}-x_\alpha)^\nu + s(X_j)\xi_j\right\} + O_p(g^q). \tag{4.10}$$

Thus we have shown that

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

in which

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

and

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

Now (by using the mixing properties of our process)

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

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

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

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

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

$$= n^{-1} h^{-1} \sigma_{m\alpha}^2(x_\alpha) \{1 + o(1)\}. \tag{4.12}$$

Equations (4.11) and (4.12) together establish (2.1). Equation (2.2) is derived by standard technique as in Linton and Härdle (1996). Equation (2.3) and all the other remaining formulas of Theorem 1, then follow directly from (2.1) and (2.2) as the various $\sqrt{nh} \left\{ \widehat{M}_\alpha(x_\alpha) - M_\alpha(x_\alpha) \right\}$'s are all asymptotically uncorrelated, so the variance of $\sqrt{nh} \{\widehat{m}(\mathbf{x}) - m(\mathbf{x})\}$ is simply the sum of all their variances, the mean of $\sqrt{nh} \{\widehat{m}(\mathbf{x}) - m(\mathbf{x})\}$ is simply the sum of all their means. Q. E. D.

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

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

$$= (n - d + 1)^{-1} \sum_{l=d}^{n} e_0^T \left(Z^T W_l Z \right)^{-1} Z^T W_l \left\{ Y^2 - \widehat{m}(x_\alpha, \overline{Y}_l)^2 - V_\alpha(x_\alpha) \right\}$$

$$= (n - d + 1)^{-1} \sum_{l=d}^{n} e_0^T \left(Z^T W_l Z \right)^{-1} Z^T W_l \left\{ Y^2 - m(x_\alpha, \overline{Y}_l)^2 + m(x_\alpha, \overline{Y}_l)^2 - \widehat{m}(x_\alpha, \overline{Y}_l)^2 - V_\alpha(x_\alpha) \right\}.$$

Now note that by Assumption A4

$$(n - d + 1)^{-1} \sum_{j=d}^{n} \prod_{\beta \neq \alpha} v_\beta(Y_{j-\beta}) = 1 + O_p(1/\sqrt{n})$$

and also that

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

16

So similar to (4.10), we have

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

where

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

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

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

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

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

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

$$\sqrt{nh} \left\{ \widehat{m}(\mathbf{x}) - m(\mathbf{x}) - h^{p+1} b_m(\mathbf{x}) \right\} \xrightarrow{D} N \left\{ 0, \sigma_m^2(\mathbf{x}) \right\}$$

therefore

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

and similarly

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

Putting together equations (4.13) through (4.17) gives the asymptotic expressions of $\widehat{V}_\alpha(x_\alpha)$ 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's except T_4, and their asymptotic correlation is (plus some higher order term)

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

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

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

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

$$= \frac{1}{d} \sum_{\alpha=1}^d \frac{1}{(n-d+1)} \sum_{j=d}^n \prod_{1 \le \beta \le d, \beta \ne \alpha} \left\{ V_\beta(Y_{j-\beta}) + \widehat{V}_\beta(Y_{j-\beta}) - V_\beta(Y_{j-\beta}) \right\} - c_v^{d-1}$$

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

$$+ \frac{1}{d} \sum_{\alpha=1}^d \frac{1}{(n-d+1)} \sum_{j=d}^n \sum_{1 \le \beta \le d, \beta \ne \alpha} \left\{ \prod_{1 \le \gamma \le d, \gamma \ne \alpha, \beta} V_\gamma(Y_{j-\gamma}) \right\} \left\{ \widehat{V}_\beta(Y_{j-\beta}) - V_\beta(Y_{j-\beta}) \right\}$$

$$= \frac{1}{d} \sum_{\alpha=1}^d \frac{1}{(n-d+1)} \sum_{j=d}^n \sum_{1 \le \beta \le d, \beta \ne \alpha} \left\{ \prod_{1 \le \gamma \le d, \gamma \ne \alpha, \beta} V_\gamma(Y_{j-\gamma}) \right\} \left\{ \widehat{V}_\beta(Y_{j-\beta}) - V_\beta(Y_{j-\beta}) \right\} + O_p(\frac{1}{\sqrt{n}})$$

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

18

where

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

$$S_2 = \frac{1}{d} \sum_{\alpha=1}^{d} \frac{1}{(n-d+1)^2} \sum_{j=d}^{n} \sum_{1 \le \beta \le d, \beta \ne \alpha} \left\{ \prod_{1 \le \gamma \le d, \gamma \ne \alpha, \beta} V_\gamma(Y_{j-\gamma}) \right\}$$
$$\times \left[\sum_{k=d}^{n} \frac{K_{0h}^{*}(Y_{k-\beta} - Y_{j-\beta})}{\varphi(Y_{j-\beta}, \overline{Y}_k)} \overline{\varphi}(\overline{Y}_k) \left\{ 2m(X_k)s(X_k)\xi_k \right\} \right]$$

$$S_3 = \frac{1}{d} \sum_{\alpha=1}^{d} \frac{1}{(n-d+1)^2} \sum_{j=d}^{n} \sum_{1 \le \beta \le d, \beta \ne \alpha} \left\{ \prod_{1 \le \gamma \le d, \gamma \ne \alpha, \beta} V_\gamma(Y_{j-\gamma}) \right\}$$
$$\times \left[\sum_{k=d}^{n} \frac{K_{0h}^{*}(Y_{k-\beta} - Y_{j-\beta})}{\varphi(Y_{j-\beta}, \overline{Y}_k)} \overline{\varphi}(\overline{Y}_k) \left\{ v(X_k)(\xi_k^2 - 1) \right\} \right]$$

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

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

and

$$S_2 = \sum_{k=d}^{n} \frac{2m(X_k)s(X_k)\xi_k}{(n-d+1)} \frac{1}{d} \sum_{\alpha=1}^{d} \sum_{1 \le \beta \le d, \beta \ne \alpha} \int \left\{ \prod_{1 \le \gamma \le d, \gamma \ne \alpha, \beta} V_\gamma(y_\gamma) \right\}$$
$$\times \frac{K_{0h}^{*}(Y_{k-\beta} - y_\beta)}{\varphi(y_\beta, \overline{Y}_k)} \overline{\varphi}(\overline{Y}_k) \varphi(y) dy \left\{ 1 + o_p(1) \right\}$$

$$= \sum_{k=d}^{n} \frac{2m(X_k)s(X_k)\xi_k}{(n-d+1)} \frac{1}{d} \sum_{\alpha=1}^{d} \sum_{1 \le \beta \le d, \beta \ne \alpha} \int \left\{ \prod_{1 \le \gamma \le d, \gamma \ne \alpha, \beta} V_\gamma(y_\gamma) \right\}$$
$$\times \frac{K_0^{*}(u)\overline{\varphi}(\overline{Y}_k)\varphi(Y_{k-\beta} - hu, \overline{y}) du d\overline{y}}{\varphi(Y_{k-\beta} - hu, \overline{Y}_k)} \left\{ 1 + o_p(1) \right\}$$

$$= \sum_{k=d}^{n} \frac{2m(X_k)s(X_k)\xi_k}{(n-d+1)\varphi(Y_k)} \frac{1}{d} \sum_{\alpha=1}^{d} \sum_{1 \le \beta \le d, \beta \ne \alpha} \int \left\{ \prod_{1 \le \gamma \le d, \gamma \ne \alpha, \beta} V_\gamma(y_\gamma) \right\}$$
$$\times \overline{\varphi}(\overline{Y}_k)\varphi(Y_{k-\beta}, \overline{y}) d\overline{y} \left\{ 1 + o_p(1) \right\}$$

from which it is clear that S_2 satisfies a central limit theorem with $\sqrt{n}$ rate of convergence, which is also the case with S_3. Thus

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

19

where $Z \overset{D}{\rightarrow} N(0, \sigma^2)$ for some σ^2, applying Taylor expansion gives the result on $\widehat{c}_v$. the rest of Theorem 2 follows directly. Q. E. D.

Proof of Theorem 3. Simply putting 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 are based on the ξ_j's and the $(\xi_j^2 - 1)$'s. Thus, any linear combinations of the estimates also have similar forms as the ones treated in Theorem 1. Q. E. D.

References

Ango Nze P. (1992) *Critères d'ergodicité de quelques modèles à représentation markovienne*, C.R. Acad. Sci. Paris, sér. I, **315**, 1301-1304.

Auestad, B.; Tjøstheim, D. (1991) Functional identification in nonlinear time series. In Nonparametric Functional Estimation and Related Topics, Ed. Roussas, G.G Amsterdam: Kluwer Academic Publishers, 493-507.

Bossaerts, P.; Härdle, W.; Hafner, C. (1996) *Foreign exchange-rates have surprising volatility*, In: Athens conference on applied probability and time series, **2**, ed. P. M. Robinson, Lecture Notes in Statistics, 115, 55-72, Springer Verlag.

Chan, K.S.; Tong, H. (1986) *On estimating thresholds in autoregressive models*, Journal of Time Series Analysis, **7**, 179-190.

Chen, R.; Tsay, R.S. (1993a) *Nonlinear additive ARX models*, Journal of the American Statistical Association, **88**, 955-967.

Chen, R.; Tsay, R.S. (1993b) *Functional-coefficient autoregressive models*, Journal of the American Statistical Association, **88**, 298-308.

Davydov, Yu.A. (1973) *Mixing conditions for Markov chains*, Theory of Probability and its Applications, **18**, 312-328.

Diebolt, J.; Guégan, D. (1993) *Tail behaviour of the stationary density of general nonlinear autoregressive processes of order one*, Journal of Applied Probability, **30**, 315-329.

Drost, F. C.; Nijman, T. E. (1993) *Temporal aggregation of GARCH processes*, Econometrica, **61**, 909–927.

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

Engle, R.F.; Gonzalez-Rivera, G. (1991) *Semiparametric ARCH models*, Journal of Business and Economic Statistics, **9**, 345-360.

Engle, R.F.; Ng, V. (1993) *Measuring and testing the impact of news on volatility*, Journal of Finance, **48**, 1749-1778.

Fan, J.; Gijbels, I. (1996) *Local polynomial modelling and its applications*, Chapman and Hall.

Gouriéroux, Ch.; Monfort, A. (1992) *Qualitative threshold ARCH models*, Journal of Econometrics, **52**, 159-199.

Granger, C.; Teräsvirta, T. (1993) *Modelling nonlinear dynamic relationships*, Oxford University Press, Oxford.

Haggan, V.; Ozaki, T. (1981) *Modelling nonlinear vibrations using an amplitude-dependent autoregressive time series model*, Biometrika, **68**, 189–196.

Härdle, W.; Chen, R. (1995) *Nonparametric time series analysis, a selective review with examples*, Proceedings of the 50th session of the ISI, Peking.

Härdle, W.; Klinke, S.; Turlach, B. (1995) *XploRe - an interactive statistical computing environment*, Springer Verlag, Heidelberg.

Härdle, W.; Tsybakov, A.B. (1997) *Local polynomial estimators of the volatility function in nonparametric autoregression*, Journal of Econometrics, **81**, 223-242.

Härdle, W.; Tsybakov, A.B.; Yang, L. (1998) *Nonparametric vector autoregression*, Journal of Statistical Planning and Inference *to appear*.

Hastie, T. J.; Tibshirani, R. J. (1990) *Generalized additive models*, Monographs on Statistics and Applied Probability, 43, Chapman and Hall, London.

Katkovnik, V.Ya. (1979) *Linear and nonlinear methods of nonparametric regression analysis*, Automatika, 35-46.

Linton, O. B.; Härdle, W. (1996) *Estimation of additive regression models with known links*, Biometrika, **83**, 529-540.

Linton, O.; Nielsen, J.P. (1995) *A kernel method of estimating structured nonparametric regression based on marginal integration*, Biometrika, **82**, 93-100.

Liptser, R.Sh.; Shirjaev, A.N. (1980) *A functional central limit theorem for martingales*, Theory of Probability and its Applications, **25**, 667 - 688.

Masry, E.; Tjøstheim, D. (1995a) *Non-parametric estimation and identification of ARCH nonlinear time series: Strong convergence and asymptotic normality*, Econometric Theory, **11**, 258-289.

Masry, E. ; Tjøstheim, D. (1995b) *Additive nonlinear ARX time series and projection estimates*, to appear in Econometric Theory.

Meese, R.A.; Rose, A. (1991) *An empirical assessment of non-linearities in models of exchange rate determination*, Review of Economic Studies, **58**, 601-619.

Mokkadem, A. (1987) *Sur un modèle autorégressif nonlinéaire. Ergodicité et ergodicité géometrique,* Journal of Time Series Analysis, **8**, 195-204.

Nummelin, E.; Tuominen, P. (1982) *Geometric ergodicity of Harris–recurrent Markov chains with application to renewal theory,* Stochastic Processes and their Applications, **12**, 187-202.

Nelson, D.B. (1991) *Conditional heteroscedasticity in asset returns: A new approach,* Econometrica, **59**, 347-370.

Robinson, P.M. (1983) *Nonparametric estimators for time series,* Journal of Time Series Analysis, **4**, 185-207.

Robinson, P.M. (1984) *Robust nonparametric autoregression,* In: *Robust and nonlinear time series analysis,* eds. Franke, Härdle and Martin, Lecture Notes in Statistics, 26, Springer-Verlag, Heidelberg.

Ruppert, D.; Wand, M.P. (1994) *Multivariate locally weighted least squares regression,* Annals of Statistics, **22**, 1346-1370.

Severance-Lossin, E.; 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,* Annals of Statistics, **5**, 595 - 645.

Stone, C.J. (1982) *Optimal global rates of convergence for nonparametric regression,* Annals of Statistics, **10**, 1040-1053.

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

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

Tong, H. (1978) *On a threshold model,* in C. H. Chen (ed.), *Pattern recognition and signal processing,* Sijthoff and Noordholf, The Netherlands.

Tong, H. (1983) *Threshold models in nonlinear time series analysis,* Lecture Notes in Statistics, 21, Springer-Verlag, Heidelberg.

Tschernig, R.; Yang, L. (1997) *Nonparametric lag selection for time series,* SFB 373 Discussion Paper 59, Humboldt Universität zu Berlin.

Tsybakov, A.B. (1986) *Robust reconstruction of functions by the local–approximation method,* Problems of Information Transmission, **22**, 133-146.

Tweedie, R.L. (1975) *Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space,* Stochastic Processes and their Applications, **3**, 385-403.

Wand, M.P.; Jones, M.C. (1995) *Kernel smoothing,* Chapman and Hall, London.