

MULTIVARIATE LOCAL POLYNOMIAL FITTING FOR MARTINGALE NONLINEAR REGRESSION MODELS

ZHAN-QIAN LU*

Department of Mathematics, The Hong Kong University of Science and Technology,

Abstract. Local polynomial modelling is a useful tool for nonlinear time series analysis. For nonlinear regression models with martingale difference errors, this paper presents a simple proof of local linear and local quadratic fittings under apparently minimal short-range dependence condition. Explicit formulae for the asymptotic bias and asymptotic variance are given, which facilitate numerical evaluations of these important quantities. The general theory is applied to nonparametric partial derivative estimation in nonlinear time series. A bias-adjusted method for constructing confidence intervals for first-order partial derivatives is described. Two examples, including the sunspots data, are used to demonstrate the use of local quadratic fitting for modelling and characterizing nonlinearity in time series data.

Key words and phrases: Partial derivative estimation, nonlinearity in time series, confidence intervals, nonparametric estimation, sunspots data.

1. Introduction

The local polynomial fitting method can be used for nonparametric estimation of both a nonlinear regression function and its partial derivatives. Local polynomial approach as a recent nonparametric regression method has various advantages, as demonstrated in Fan and Gijbels (1996). Local polynomial modelling is also useful in modelling and prediction of nonlinear time series. This approach is familiar in the time series literature as it has close connections to many other modelling techniques such as local smoothing or state-dependent models. In the time series context, the one-dimensional case is studied by Masry and Fan (1997) who established asymptotic normality of local polynomial fitting for stationary processes under some mixing conditions. Subsequently, Masry (1996) generalized to this result to the multivariate case for local polynomial fitting of any order.

* Now at Data Analysis Product Division, Mathsoft, 1700 Westlake Ave N, Suite 500, Seattle, WA 98109-3044, U.S.A.

(B) $\sup_{i \geq 1} E\{|\varepsilon_i|^{2+\delta} \mid \mathcal{F}_{i-1}\} < \infty$ for some $\delta > 0$.

(C) The vector sequence $\{X_i\}$ is strictly stationary and satisfies the short-range dependence condition: let $f_j(\cdot, \cdot)$ denote the joint density of X_1, X_{j+1} and $f(\cdot)$ denote the marginal density, then

$$(1.3) \quad \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{R}^p} \sum_{j=1}^{\infty} |f_j(\mathbf{u}, \mathbf{v}) - f(\mathbf{u})f(\mathbf{v})| < \infty.$$

Condition (C) is a reasonable mixing condition which has been commonly used in the nonparametric estimation literature (cf. Castellana and Leadbetter (1986)). It appears to be considerably weaker than those in Masry (1996) and Masry and Fan (1997) and may be a minimal short-range dependence condition in this context.

Though the extra condition (A) is assumed on the model structure, it is quite natural in the time series context and is equivalent to the requirement that enough predictor variables are included in X_{i-1} . As a consequence of (A), the following results hold automatically:

$$m(X_{i-1}) = E\{Y_i \mid \mathcal{F}_{i-1}\}, \quad \text{and} \quad \nu(X_{i-1}) = \text{Var}\{Y_i \mid \mathcal{F}_{i-1}\}.$$

The latter also defines the variance function in (1.2).

Most time series models in common use satisfy condition (A). For example, (A) is satisfied when ε_i is independent of X_{i-1}, \dots in the past and $\mathcal{F}_i = \sigma\{Y_j, X_j; j \leq i\}$. In particular, our setup includes the following *nonlinear autoregression* (NAR) - *autoregressive conditionally heteroscedastic* (ARCH) model. Consider

$$(1.4) \quad x_i = m(x_{i-1}, x_{i-2}, \dots, x_{i-p}) + \nu^{1/2}(x_{i-1}, \dots, x_{i-p})\varepsilon_i,$$

where m and ν as before and ε_i is iid, and defining $\mathcal{F}_i = \sigma\{x_k, k \leq i\}$. A given time series data $\{x_1, x_2, \dots, x_N\}$ correspond to (1.1) through

$$Y_i = x_i, X_i = (x_i, x_{i-1}, \dots, x_{i-p+1})^T, \quad (i = p, p+1, \dots, N)$$

and $n = N - p$. The variance function $\nu(\cdot)$ in this context is known as the *volatility* function in the finance and econometrics literature. Obviously, the method developed in this paper can be generalized directly to study the volatility function, cf. Härdle and Tsybakov (1997), and Härdle *et al.* (1996). The latter also considers the setup of a vector autoregression model with a heteroscedastic covariance structure. When ε_i is iid, the state process $\{X_i\}$ defined through (1.4) is a Markov chain. Under certain assumptions on m, ν and the distribution of ε_1 , this process is *geometrically ergodic*, which enjoys some *strong mixing* property. Asymptotic normality of some nonparametric estimators can be shown to hold as a consequence, cf. Härdle *et al.* (1996).

Another direction of generalizing results of this paper is to estimation of functionals of other aspects of the predictive distribution function $F(y \mid \mathcal{F}_{i-1}) = P(Y_i \leq y \mid \mathcal{F}_{i-1})$, such as the regression quantile, cf. Welsh (1996).

2. Local polynomial fitting

Consider first the local quadratic fitting. The estimators of m and its partial derivatives m_1, \dots, m_p at any given point $\mathbf{x} = (x_1, \dots, x_p)^T$ are derived by minimizing the weighted sum of squares

$$(2.1) \quad \sum_{i=1}^n \{Y_i - a - b^T(X_{i-1} - \mathbf{x}) - (X_{i-1} - \mathbf{x})^T L(X_{i-1} - \mathbf{x})\}^2 \frac{1}{h^p} K\left(\frac{X_{i-1} - \mathbf{x}}{h}\right),$$

where a is a real number, b is a p -dimensional vector, and L is a $p \times p$ matrix which is restricted to be a lower triangular matrix for identifiability. The solution corresponding to minimizing (2.1) consists of $\hat{a} - \hat{m}(\mathbf{x})$, an estimate of regression function at \mathbf{x} , of $\hat{b} = \hat{D}_m(\mathbf{x})$ which corresponds to an estimate of $D_m(\mathbf{x}) = (\partial m(\mathbf{x})/\partial x_1, \dots, \partial m(\mathbf{x})/\partial x_p)^T$ at \mathbf{x} , and of \hat{L} which corresponds to estimates of elements in the Hessian matrix of $H_m(\mathbf{x}) = (\partial^2 m(\mathbf{x})/\partial x_i \partial x_j)$ at \mathbf{x} . That is, $L(x) = (l_{ij})$ satisfies $l_{ij} = h_{ij}$ if $i > j$ and $= h_{ii}/2$ if $i = j$, where $H_m(\mathbf{x}) = (h_{ij})$ is the Hessian. Let $\hat{\beta} = (\hat{a}, \hat{b}^T, \text{vech}^T\{\hat{L}\})^T$, and we have

$$(2.2) \quad \hat{\beta} = (X^T W X)^{-1} X^T W Y,$$

where $Y = (Y_1, \dots, Y_n)^T$, $W = \text{diag}\{K(\frac{X_0 - \mathbf{x}}{h}), \dots, K(\frac{X_{n-1} - \mathbf{x}}{h})\}$ and

$$\begin{pmatrix} 1 & (X_0 - \mathbf{x})^T & \text{vech}^T\{(X_0 - \mathbf{x})(X_0 - \mathbf{x})^T\} \end{pmatrix}$$

Here vech^T denotes the row vector consisting of the columns on and below the diagonal of a symmetric matrix.

The local linear estimator $\hat{\beta}_L$ can be defined similarly as in the case of local quadratic estimator with all quadratic terms omitted from (2.1), (2.2), (2.3).

Let U denote an open neighborhood of $\mathbf{x} = (x_1, \dots, x_p)^T$ in \mathbb{R}^p , and let $C^d(U)$ be the class of functions which have up to order d continuous partial derivatives in U . Let I_ℓ denote the identity matrix of dimension ℓ . For simplicity, the kernel K is assumed to be spherically symmetric, i.e. $K = k(\|\mathbf{x}\|)$ for some function k . We

matrices of dimension p and $p(p+1)/2$ respectively.

2.1 Local linear fitting

The following theorem is developed for the local linear estimator, for which the kernel K is assumed to satisfy $\int u_1^8 K(u_1, \dots, u_p) du_1 \cdots du_p < \infty$.

THEOREM 1. *Under model (1.2) and Assumptions (A)–(C), consider any ℓ distinct points $\mathbf{x}_1, \dots, \mathbf{x}_\ell$ satisfying $f(\mathbf{x}_j) > 0$, $\nu(\mathbf{x}_j) > 0$, if there exist open neighborhoods U_i of \mathbf{x}_i such that $m \in C^3(U_j)$, $f \in C^1(U_j)$, $\nu \in C^0(U)$, $j = 1, 2, \dots, \ell$,*

the local linear estimators $\hat{\beta}_L(\mathbf{x}_1), \dots, \hat{\beta}_L(\mathbf{x}_\ell)$ are asymptotically independent and jointly normal as $h \rightarrow 0$, $nh^p \rightarrow \infty$. In particular, at each point $\mathbf{x} = (x_1, \dots, x_p)^T$, we have that

$$(nh^p)^{1/2} \text{diag}\{1, hI_1\} \{\hat{\beta}_L(\mathbf{x}) - \beta_L(\mathbf{x}) - B_L(\mathbf{x}, h)\}$$

is asymptotically normal $N(0, \Sigma(\mathbf{x}))$. Here $B_L(\mathbf{x}, h)$ is the asymptotic bias given by

$$(2.4) \quad B_L(\mathbf{x}, h) = \begin{pmatrix} \frac{1}{2} h^2 \mu_2 \nabla_m^2(\mathbf{x}) + o(h^3) \\ \frac{h^2}{3! \mu_2} b(\mathbf{x}) + \frac{h^2}{2 \mu_2 f(\mathbf{x})} b_1(\mathbf{x}) + o(h^2) \end{pmatrix},$$

where $\nabla_m^2(\mathbf{x}) = \sum_{i=1}^p \partial^2 m(\mathbf{x}) / \partial x_i^2$,

$$\left((\mu_1 - \mu_2^2) \frac{\partial^2 m(\mathbf{x})}{\partial x_1^2} \frac{\partial f(\mathbf{x})}{\partial x_1} + 2\mu_2^2 \sum_{i=2}^p \frac{\partial^2 m(\mathbf{x})}{\partial x_1 \partial x_i} \frac{\partial f(\mathbf{x})}{\partial x_1} \right)$$

Remark 1. It should be pointed out that in Theorem 1 for the results corresponding to the regression estimator to hold, weaker smoothness assumptions $m \in C^2(U)$, $f \in C^0(U)$ will suffice. The local linear regression estimator has been a popular method for nonlinear prediction of time series.

Theorem 1 generalizes Theorem 3 in Lu (1996). The proof of Theorem 1 is similar to and contained in the proof of Theorem 2 to be given in Subsection 2.3 and is thus omitted here.

2.2 Local quadratic fitting

We have the following theorem for the local quadratic estimator, for which the kernel K is assumed to satisfy $\int u_1^{12} K(u_1, \dots, u_p) du_1 \cdots du_p < \infty$.

THEOREM 2. *Under model (1.2) and Assumptions (A)–(C), for l distinct points $\mathbf{x}_1, \dots, \mathbf{x}_\ell$ such that $f(\mathbf{x}_i) > 0$, $\nu(\mathbf{x}_j) > 0$ for all j , if there exist open neighborhoods U_i of \mathbf{x}_i such that $m \in C^4(U_j)$, $f \in C(U_j)$, $\nu \in C^0(U)$, $j = 1, 2, \dots, \ell$, then for $h \rightarrow 0$, $nh^p \rightarrow \infty$ as $n \rightarrow \infty$, the local quadratic estimators $\hat{\beta}(\mathbf{x}_1), \dots, \hat{\beta}(\mathbf{x}_\ell)$ are asymptotically independent and jointly normal. In particular, at each point $\mathbf{x} = (x_1, \dots, x_p)^T$, we have that*

$$(nh^p)^{1/2} \text{diag}\{1, h_p I_1, h^2 I_2\} \{\hat{\beta}(\mathbf{x}) - \beta(\mathbf{x}) - B(\mathbf{x}, h)\}$$

is asymptotically normal $N(0, \Sigma(\mathbf{x}))$, where

$$B(\mathbf{x}, h) = \begin{pmatrix} \frac{h^4}{4!} \theta(\mathbf{x}) + \frac{h^4}{3!f(\mathbf{x})} \theta_1(\mathbf{x}) + o(h^4) \\ \frac{h^2}{3!\mu_2} b(\mathbf{x}) + o(h^3) \\ \frac{h^2}{4!} \gamma(\mathbf{x}) + \frac{h^2}{3!f(\mathbf{x})} \gamma_1(\mathbf{x}) + o(h^2) \end{pmatrix},$$

where $b(\mathbf{x})$ is defined in (2.5) in Theorem 1 of Section 2, and

$$\theta(\mathbf{x}) = \frac{\mu_4^2 - \mu_2\mu_6}{\mu_4 - \mu_2^2} \sum \frac{\partial^4 m(\mathbf{x})}{\partial x^4} - 6\mu_2^2 \sum \frac{\partial^4 m(\mathbf{x})}{\partial x^2 \partial x^2},$$

$$(2.7) \quad \Sigma(\mathbf{x}) =$$

$$\begin{pmatrix} \frac{\rho\nu(\mathbf{x})}{f(\mathbf{x})} & 0 & \frac{\phi\nu(\mathbf{x})}{f(\mathbf{x})} \text{vech}^T\{I_1\} \\ 0 & \frac{J_2\nu(\mathbf{x})}{\mu_2^2 f(\mathbf{x})} I_1 & 0 \\ \frac{\phi\nu(\mathbf{x})}{f(\mathbf{x})} \text{vech}\{I_1\} & 0 & \frac{\nu(\mathbf{x})}{f(\mathbf{x})} \left(\Lambda - \frac{\mu_2(J_2 - J_0\mu_2)}{(\mu_4 - \mu_2^2)^2} \text{vech}\{I_1\} \text{vech}^T\{I_1\} \right) \end{pmatrix},$$

where

$$\begin{aligned} \rho &= (\mu_4 - \mu_2^2)^{-2} \{J_0(\mu_4 + (p-1)\mu_2^2)^2 - 2pJ_2\mu_2(\mu_4 + (p-1)\mu_2^2) \\ &\quad + p\mu_2^2(J_4 + (p-1)J_2^2)\}, \\ \phi &= (\mu_4 - \mu_2^2)^{-2} \{J_2\mu_4 + (2p-1)J_2\mu_2^2 - (p-1)J_2^2\mu_2 \\ &\quad - J_4\mu_2 - J_0\mu_2\mu_4 - (p-1)J_0\mu_2^3\}, \end{aligned}$$

where $\lambda_1 = (J_4 - J_2^2)(\mu_4 - \mu_2^2)^{-2}$, $\lambda_2 = J_2^2 \mu_2^{-4}$.

Remark 2. It is noted that in Theorem 2 for the results corresponding to the first-order partial derivative estimators to hold, weaker smoothness assumptions $m \in C^3(U)$, $f \in C^0(U)$ will suffice. An application of local quadratic fit for first-order partial derivative estimation will be discussed in more detail in Section 3.

Remark 3. Theorem 2 generalizes Theorem 4 in Lu (1996). The explicit expressions for the asymptotic bias and asymptotic covariance matrix are first derived in Lu (1996), which involves complicated matrix calculations.

2.3 Proof of Theorem 2

Write

$$(2.8) \quad S_n \{D(\hat{\beta}(\mathbf{x}) - \beta(\mathbf{x}))\} = R_n + (nh^p)^{-1/2} Z_n,$$

where

$$\begin{aligned} S_n &= (nh^p)^{-1} D^{-1} \mathbf{X}^T W \mathbf{X} D^{-1}; \\ R_n &= (nh^p)^{-1} D^{-1} \mathbf{X}^T W (M - \mathbf{X} \beta); \\ D &= \text{diag}\{1, hI_1, h^2 I_2\}; \end{aligned}$$

where $M = (m(X_0), \dots, m(X_{n-1}))^T$;

$$(2.9) \quad \left\{ \begin{array}{l} Z_n = (nh^p)^{-1/2} D^{-1} \mathbf{X}^T W V^{1/2} E = (nh^p)^{-1/2} \sum_{i=1}^n Z_{ni}; \\ \text{where } E = (\varepsilon_1, \dots, \varepsilon_n)^T, \\ V = \text{diag}\{\nu(X_0), \dots, \nu(X_{n-1})\}; \\ \text{and} \\ Z_{ni} = \left(\begin{array}{c} \left(\frac{X_{i-1} - \mathbf{x}}{h} \right) \\ \text{vech} \left\{ \left(\frac{X_{i-1} - \mathbf{x}}{h} \right) \left(\frac{X_{i-1} - \mathbf{x}}{h} \right)^T \right\} \end{array} \right) K \left(\frac{X_{i-1} - \mathbf{x}}{h} \right) \nu^{1/2}(X_{i-1}) \varepsilon_i. \end{array} \right.$$

If $m = m(x_1, \dots, x_p) \in C^d(U)$, for a positive number k (less than d), we denote the k -th-order differential $D_m^k(\mathbf{x}, \mathbf{u})$ at any given point $\mathbf{u} = (u_1, \dots, u_p) \in \mathbb{R}^p$ by

$$D_m^k(\mathbf{x}, \mathbf{u}) = \sum_{i_1, \dots, i_p} C_{i_1 \dots i_p}^k \frac{\partial^k m(\mathbf{x})}{\partial x_1^{i_1} \partial x_2^{i_2} \dots \partial x_p^{i_p}} u_1^{i_1} \dots u_p^{i_p},$$

where the summations are over all distinct nonnegative integers i_1, \dots, i_p such that $i_1 + \dots + i_p = k$, and $C_{i_1 \dots i_p}^k = k!/(i_1! \dots i_p!)$.

The following lemmas on S_n , R_n and Z_n are available. The proofs of Lemmas 1 and 2 follow from condition (C) and the Chebyshev's inequality. Proof of Lemma 3 is given at the end of this subsection.

LEMMA 1. Under condition (C) and $f(\mathbf{x}) > 0$, we have as $nh^p \rightarrow \infty$,

$$(2.10) \quad S_n^{-1} = A(h)^{-1} + O_p((nh^p)^{-1/2}),$$

where $A(h) = \int (1, \mathbf{u}^T, \text{vech}^T\{\mathbf{u}\mathbf{u}^T\})^T K(\mathbf{u}) f(\mathbf{x} + h\mathbf{u}) (1, \mathbf{u}^T, \text{vech}^T\{\mathbf{u}\mathbf{u}^T\}) d\mathbf{u}$.

LEMMA 2. Assume $m \in C^4(U)$, $f \in C^1(U)$, and condition (C), as $h \rightarrow 0$, $nh^p \rightarrow \infty$,

$$(2.11) \quad R_n = h^3\{R(h, \mathbf{x}) + o(h) + O_p((nh^p)^{-1/2})\},$$

where

$$\begin{aligned} R(h, \mathbf{x}) = & \frac{1}{3!} \begin{pmatrix} h \int D_m^3(\mathbf{x}, \mathbf{u}) K(\mathbf{u}) [D_f^T(\mathbf{x}) \mathbf{u}] d\mathbf{u} \\ f(\mathbf{x}) \int \mathbf{u} D_m^3(\mathbf{x}, \mathbf{u}) K(\mathbf{u}) d\mathbf{u} \\ h \int \text{vech}\{\mathbf{u}\mathbf{u}^T\} D_m^3(\mathbf{x}, \mathbf{u}) K(\mathbf{u}) [D_f^T(\mathbf{x}) \mathbf{u}] d\mathbf{u} \end{pmatrix} \\ & + \frac{f(\mathbf{x})h}{4!} \begin{pmatrix} \int D_m^4(\mathbf{x}, \mathbf{u}) K(\mathbf{u}) d\mathbf{u} \\ 0 \\ \int \text{vech}\{\mathbf{u}\mathbf{u}^T\} D_m^4(\mathbf{x}, \mathbf{u}) K(\mathbf{u}) d\mathbf{u} \end{pmatrix}. \end{aligned}$$

LEMMA 3. Under conditions (A), (B), and (C), for any l points $\mathbf{x}_1, \dots, \mathbf{x}_l$ such that $f(\mathbf{x}_i) > 0$, $f \in C^0(U_i)$, $v \in C^0(U_i)$, where U_i is an open neighborhood of \mathbf{x}_i for $1 \leq i \leq l$, as $h \rightarrow 0$, $nh^p \rightarrow \infty$, $Z_n(\mathbf{x}_1), \dots, Z_n(\mathbf{x}_l)$ defined in (2.9) are asymptotically independent and jointly normal. In particular, at a particular point \mathbf{x} , we have

$$(2.12) \quad Z_n \rightarrow N(0, \Sigma_1),$$

where

$$\Sigma_1 = v(\mathbf{x})f(\mathbf{x}) \int \begin{pmatrix} 1 \\ \mathbf{u} \\ \text{vech}\{\mathbf{u}\mathbf{u}^T\} \end{pmatrix} (1, \mathbf{u}^T, \text{vech}^T\{\mathbf{u}\mathbf{u}^T\}) k(\mathbf{u})^2 d\mathbf{u} + O(h).$$

Now we can give a proof of Theorem 2 based on these lemmas.

PROOF OF THEOREM 2. From (2.8), we have

By Lemmas 1 and 3, the right-hand side of (2.13) tends in distribution to $N(0, \Sigma)$, where

$$\Sigma(h) = A^{-1}(h)\Sigma_1 A^{-1}(h).$$

On the other hand, it can be shown that

$$\Sigma(h) = \Sigma(\mathbf{x}) + o(h),$$

by same calculations as in Lu (1996).

Next we show that B_n has the right expansion. Combining Lemmas 1 and 2, we write

$$\begin{aligned} B_n &= h^3 \operatorname{diag}\{1, h^{-1}I, h^{-2}I_2\} \{A^{-1}(h)R(h, \mathbf{x}) + o(h) + O_p(\{nh^p\}^{-1/2})\} \\ &= B(\mathbf{x}, h) + h^3 O_p(\{nh^p\}^{-1/2}). \end{aligned}$$

Meanwhile $B(\mathbf{x}, h)$ so defined can be checked to have the given form.

The asymptotic independence of the estimators at different points follows similarly using the first part of Lemma 3. We have thus proved Theorem 2. \square

THE PROOF OF LEMMA 3. By the Cramer-Wold device, to prove asymptotic multivariate normality of $Z_n(\mathbf{x})$ at a particular point \mathbf{x} , we only need to prove for any linear combination of components, say

$$(2.14) \quad \xi_{ni} \triangleq \frac{1}{\sqrt{nh^p}} l \left(\frac{X_{i-1} - \mathbf{x}}{h} \right) \nu^{1/2} (X_{i-1}) \varepsilon_i,$$

where

$$\begin{aligned} l \left(\frac{X_{i-1} - \mathbf{x}}{h} \right) &\triangleq \left\{ a + b^T \left(\frac{X_{i-1} - \mathbf{x}}{h} \right) \right. \\ &\quad \left. + c^T \operatorname{vech} \left\{ \left(\frac{X_{i-1} - \mathbf{x}}{h} \right) \left(\frac{X_{i-1} - \mathbf{x}}{h} \right)^T \right\} \right\} \\ &\quad \cdot K \left(\frac{X_{i-1} - \mathbf{x}}{h} \right). \end{aligned}$$

Here a, b, c are constant, vectors of dimension p and $p(p+1)/2$. It is easy to check that $\{\xi_{ni}, \mathcal{F}_i\}$ is a sequence of square-integrable martingale differences.

Note that Z_n defined in (2.9) is in the form of an array sum of square integrable martingale differences. By a martingale central limit theorem, see e.g. Shiryaev ((1984), p. 511), we only need to check the Lindeberg condition:

$$\begin{aligned} &\sum_{i=1}^n E\{\xi_{ni}^2 1(|\xi_{ni}| > \epsilon) \mid \mathcal{F}_{i-1}\} \\ &\leq \sum_{i=1}^n E\left\{ \frac{|\xi_{ni}|^{2+\delta}}{\epsilon^\delta} \mid \mathcal{F}_{i-1} \right\} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \frac{1}{\epsilon^\delta} \frac{1}{(nh^p)^{(2+\delta)/2}} \mathbb{E} \left\{ \left| l \left(\frac{X_{i-1} - \mathbf{x}}{h} \right) \nu^{1/2}(X_{i-1}) \varepsilon_i \right|^{2+\delta} \mid \mathcal{F}_{i-1} \right\} \\
&= \frac{1}{(nh^p)^{1+\delta/2} \epsilon^\delta} \sum_{i=1}^n \left| l \left(\frac{X_{i-1} - \mathbf{x}}{h} \right) \nu^{1/2}(X_{i-1}) \right|^{2+\delta} \mathbb{E} \{ |\varepsilon_i|^{2+\delta} \mid \mathcal{F}_{i-1} \} \\
&= \frac{1}{(nh^p)^{1+\delta/2} \epsilon^\delta} \sup_{i \geq 1} \mathbb{E} \{ |\varepsilon_i|^{2+\delta} \mid \mathcal{F}_{i-1} \} \sum_{i=1}^n \left| l \left(\frac{X_{i-1} - \mathbf{x}}{h} \right) \right|^{2+\delta} \nu^{(2+\delta)/2}(X_{i-1}),
\end{aligned}$$

where assumptions (A) and (B) are used.

Applying the Chebyshev's inequality, the right-hand side of the above equation is equal to

$$\begin{aligned}
&\frac{1}{(nh^p)^{\delta/2} \epsilon^\delta} \mathbb{E} |\varepsilon_1|^{2+\delta} \left\{ f(\mathbf{x}) \nu^{(2+\delta)/2}(\mathbf{x}) \int |l(\mathbf{u})|^{2+\delta} d\mathbf{u} + o(1) + O_p((nh^p)^{-1/2}) \right\} \\
&\xrightarrow{P} 0,
\end{aligned}$$

if $h \rightarrow 0$, $nh^p \rightarrow \infty$, as $n \rightarrow \infty$. So the Lindeberg condition is satisfied.

Furthermore,

$$\sum_{i=1}^n \mathbb{E} \{ \xi_{ni}^2 \mid \mathcal{F}_{i-1} \} = \frac{1}{nh^p} \sum_{i=1}^n l \left(\frac{X_{i-1} - \mathbf{x}}{h} \right)^2 \nu(X_{i-1}).$$

Applying the Chebyshev's inequality again, the right-hand side of above equation is equal to

$$l(\mathbf{x}) \nu(\mathbf{x}) \int l(\mathbf{u})^2 d\mathbf{u} + o(1) + O_n((nh^p)^{-1/2}),$$

$$\begin{aligned}
&= (a, b, c^T) \int \begin{pmatrix} \mathbf{u} \\ \text{vech}\{\mathbf{u}\mathbf{u}^T\} \end{pmatrix} (1, \mathbf{u}^T, \text{vech}^T\{\mathbf{u}\mathbf{u}^T\}) \\
&\quad K(\mathbf{u})^2 d\mathbf{u} (a, b, c^T)^T \\
&\quad + o(1).
\end{aligned}$$

So by Theorem 4 of Shirayev ((1984), p. 511), we have

$$\sum_{i=1}^n \xi_{ni} \xrightarrow{d} N(0, (a, b, c^T) \Sigma_1 (a, b, c^T)^T),$$

where

$$\Sigma_1 = v(\mathbf{x}) f(\mathbf{x}) \int \begin{pmatrix} 1 \\ \mathbf{u} \\ \text{vech}\{\mathbf{u}\mathbf{u}^T\} \end{pmatrix} (1, \mathbf{u}^T, \text{vech}^T\{\mathbf{u}\mathbf{u}^T\}) K(\mathbf{u})^2 d\mathbf{u} + o(1).$$

By the Cramer-Wold device, this implies that

$$Z_n \xrightarrow{d} N(0, \Sigma_1).$$

The joint asymptotic normality of $Z_n(\mathbf{x}_1), \dots, Z_n(\mathbf{x}_l)$ can be proved similarly. So Lemma 3 is proved. \square

3. Partial derivative estimation

In this section, we consider only the scalar time series model (1.4). For the state vector by $X_i = (x_i, x_{i-1}, \dots, x_{i-p+1})^T$, we denote the p first-order partial derivative functions (pdfs) by $m_1(X_i) = \frac{\partial m(X_i)}{\partial x_i}, \dots, m_p(X_i) = \frac{\partial m(X_i)}{\partial x_{i-p+1}}$. The dependence of $m_i(X_i)$'s on the state vector X_i is of particular interest for characterizing nonlinearity of m . We consider application of local quadratic fitting to estimation of m_1, \dots, m_p based on time series data.

In order to assess nonlinearity of a time series, the issue of quantifying the variability associated with the pdf estimators becomes crucial. In principle, in order to test whether the estimated pdfs are constant over different parts of the phase space, a simultaneous confidence band for the derivative function is desirable. Unfortunately, this theory is not available at the moment. So, we will be content with developing *pointwise* confidence intervals for m_1, \dots, m_p at any given point \mathbf{x} with the understanding that these intervals are expected to be considerably narrower than simultaneous confidence band.

3.1 Confidence intervals

One issue in confidence interval construction is to deal with the *bias* in the nonparametric estimators. Some type of bias-correction is desirable since the bias term is often not negligible. The bias for $\hat{D}_m(\mathbf{x})$ involves third-order partial derivatives, so the local cubic fit with a larger bandwidth h_4 is adopted for estimating the third-order derivatives. Thus, plugging in the estimated third-order derivatives into formula (2.5), one obtains the estimated leading bias term $(h^2/(6\mu_2))\hat{b}(\mathbf{x})$. Similar to the local linear and local quadratic fits, under conditions (A), (B) and (C), some general restriction on h_4 and appropriate smoothness conditions, it can be shown that the third-order derivative estimators are consistent so that $\hat{b}(\mathbf{x})$ is a consistent estimator for the bias term $b(\mathbf{x})$.

On the other hand, the variance estimation for $\hat{D}_m(\mathbf{x})$ is relatively straightforward. One option is to use the pre-asymptotic conditional variance matrix for $\hat{\beta}$ given by

$$(3.1) \quad (\mathbf{X}^T W \mathbf{X})^{-1} \mathbf{X}^T W V W \mathbf{X} (\mathbf{X}^T W \mathbf{X})^{-1}$$

assuming that estimates of the variance function $\nu(\cdot)$ at each data points are available. (The calculation also involves the inversion of the matrix $(\mathbf{X}^T W \mathbf{X})$ of dimension $(p+1)(p+2)/2$ at each \mathbf{x} .) In this paper we use the asymptotic variance formula in (2.7) directly, and the calculation involves only estimation of

$$\hat{f}(\mathbf{x}) = \frac{1}{nh_2^p} \sum_{i=1}^n K\left(\frac{X_{i-1} - \mathbf{x}}{h_2}\right),$$

and

$$\hat{\nu}(\mathbf{x}) = \sum_{i=1}^n Y_i^2 K\left(\frac{X_{i-1} - \mathbf{x}}{h_3}\right) \bigg/ \sum_{i=1}^n K\left(\frac{X_{i-1} - \mathbf{x}}{h_3}\right) - \hat{m}_{NW}(\mathbf{x})^2,$$

$$\hat{m}_{NW}(\mathbf{x}) = \sum_{i=1}^n Y_i K\left(\frac{X_{i-1} - \mathbf{x}}{h_3}\right) \bigg/ \sum_{i=1}^n K\left(\frac{X_{i-1} - \mathbf{x}}{h_3}\right).$$

Under general conditions, it can be shown that $\hat{f}(\mathbf{x})$ and $\hat{\nu}(\mathbf{x})$ are consistent. For example, if f is differentiable in an open neighborhood of \mathbf{x} , then the mixing

as $n h_2 \rightarrow \infty$, $h_2 \rightarrow 0$.

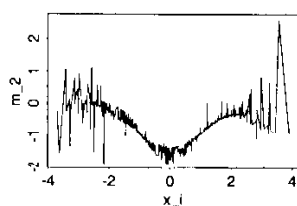
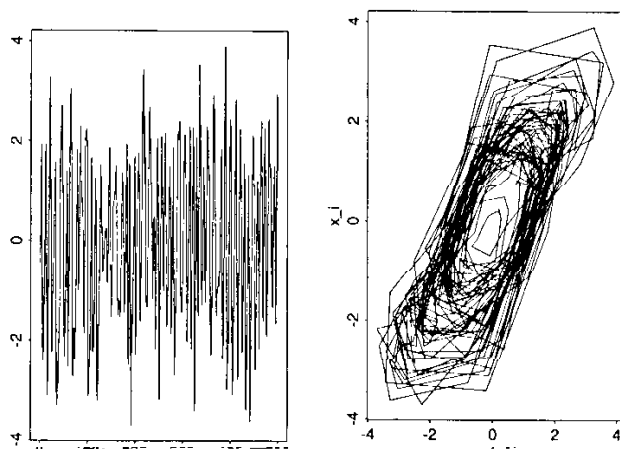
Theorem 2 implies that, assuming $h \rightarrow 0$, $nh^p \rightarrow \infty$, $nh^{p+6} = O(1)$ and relevant consistency conditions on plug-in estimators, an asymptotically $100(1 - \alpha)\%$ confidence interval for the partial derivative vector $D_m(\mathbf{x})$ is given by

$$\begin{aligned} I_n = & [\hat{D}_m(\mathbf{x}) - h^2 \hat{b}(\mathbf{x}) / (6\mu_2) \\ & - Z_{\alpha/2} (nh^{p+2})^{-1/2} J_2^{1/2} \hat{\nu}^{1/2}(\mathbf{x}) / (\mu_2 \hat{f}^{1/2}(\mathbf{x})) (1, \dots, 1)^T, \\ & \hat{D}_m(\mathbf{x}) - h^2 \hat{b}(\mathbf{x}) / (6\mu_2) \\ & + Z_{\alpha/2} (nh^{p+2})^{-1/2} J_2^{1/2} \hat{\nu}^{1/2}(\mathbf{x}) / (\mu_2 \hat{f}^{1/2}(\mathbf{x})) (1, \dots, 1)^T]. \end{aligned}$$

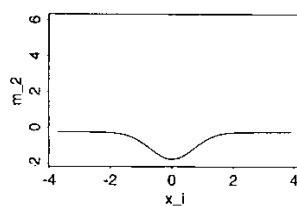
3.2 Examples

Two examples are considered in this subsection. Tricube kernel (cf. Appendix) is used in these examples, though other kernels in the Appendix can be used. Also different kernels other than the one used for local polynomial fit may be used for density and variance estimations.

Example 1. (Simulations from an exponential autoregression model) Time series of length 500 is generated from the model (1.4) where $\varepsilon_i \sim N(0, 1)$, $\nu = 0.5^2$, and



(c)



(d)



(e)



(f)

Fig. 1. Simulations from an exponential AR model: (a) time plot; (b) phase plot; (c) estimated curve of m_2 ; (d) true curve of m_2 ; (e) estimated surface of m_1 ; (f) true surface of m_1 .

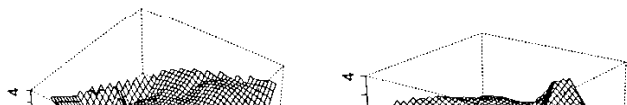
The pdfs m_1 , m_2 have the property that only $m_1(X_i)$ (where $X_i = (x_i, x_{i-1})^T$) depends on both x_i , x_{i-1} , and $m_2(X_i)$ depends only on x_i .

Estimations of m_1 , m_2 at each data point are given using $p = 2$ and bandwidth $h = 1.5$. Figure 1(c) shows both the original \hat{m}_2 and a smoothed version (the solid



(a)

(b)



(e)



(f)

Fig. 2. Annual sunspots numbers from 1700-1996: (a) kernel density estimation of f ; (b) N-W estimation of conditional variance ν ; (c) surface plot of estimate of pdf m_1 ; (d) surface plot of estimate of pdf m_2 ; (e) time plots estimates (raw estimate: solid line, bias-adjusted: dashed line) and 95% CIs (dotted lines) for m_1 ; (f) same as e except for m_2 .

3): triweight; $K_{33}(\alpha = 3, \beta = 3)$: tricube. Higher-order moments are given by (for integer $m \geq 0$)

$$\mu_{2m} = B\left(\frac{p+2m}{\alpha}, \beta+1\right) B\left(\frac{p-1}{2}, m+\frac{1}{2}\right) /$$

$$\cdot \left\{ B\left(\frac{p}{\alpha}, \beta+1\right) B\left(\frac{p-1}{2}, \frac{1}{2}\right) \right\},$$

$$\mu_{2m+1} = B\left(\frac{p+2m+1}{\alpha}, \beta+1\right) B\left(\frac{p-1}{2}, m+\frac{1}{2}\right) /$$

REFERENCES

- Castellana, J. V. and Leadbetter, M. R. (1986). On smoothed probability density estimation for stationary processes, *Stochastic. Process. Appl.*, **21**, 179-193.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.
- Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression, *J. Econometrics*, **81**, 223-242.
- Härdle, W., Tsybakov, A. and Yang, L. (1996). Nonparametric vector autoregression, Working Papers, No. 61, Humboldt University, Berlin.
- Lu, Z. Q. (1996). Multivariate locally weighted polynomial fitting and partial derivative estimation, *J. Multivariate Anal.*, **59**, 187-205.
- Masry, E. (1996). Multivariate regression estimation: local polynomial fitting for time series, *Stochastic. Process. Appl.*, **65**(1), 81-101.

- Nychka, D., Ellner, S., McCaffrey, D. and Gallant, A. R. (1992). Finding chaos in noisy systems, *J. Roy. Statist. Soc. Ser. B*, **54**(2), 399-426.
- Priestley, M. B. (1988). *Non-linear and Non-stationary Time Series Analysis*, Academic Press, London.
- Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression, *Ann. Statist.*, **22**(3), 1346-1370.
- Shiryayev, A. N. (1984). *Probability*, Springer, New York.
- Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford.
- Welsh, A. H. (1996). Robust estimation of smooth regression and spread functions and their derivatives, *Statist. Sinica*, **6**, 347-366.