Ann. Inst. Statist. Math.
Vol. 51, No. 4, 691–706 (1999)

MULTIVARIATE LOCAL POLYNOMIAL FITTING FOR MARTINGALE NONLINEAR REGRESSION MODELS

ZHAN-QIAN LU*

Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

(Received November 20, 1997; revised May 22, 1998)

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 Gijbel (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.

Though significant theoretically, there remain many practical issues in applying these results in data analysis. For example, the mixing conditions of Masry and Fan (1997) and Masry (1996) are complicated and not easy to check in practical situations. In this paper, we study the two important cases of multivariate *local linear* fit and *local quadratic* fit in the natural setup of nonlinear regression models with martingale difference errors. Our assumption on mixing is apparently much weaker and appears to be a minimal short-range dependence condition in this context. Our proof of asymptotic normality using a martingale central limit theorem is different from the two cited references which employ the much involved Bernstein's big block and small block argument. Explicit formulae for the asymptotic bias and asymptotic variances are given based on earlier results in the nonparametric regression context of Ruppert and Wand (1994) and Lu (1996). These formulae facilitate calculations of the asymptotic bias and variance for partial derivative estimators and are useful in determining a proper bandwidth.

Nonlinearity in time series data has received increasing attention. For example, a lot of financial and economical time series data have been found to demonstrate some degree of nonlinearity (Mills (1993)). Modelling first-order partial derivatives is a natural approach to characterizing nonlinearity in time series data. For example, the structure of partial derivatives is used in identifying nonlinear time series models by Priestley ((1988), Chapter 5). Derivative estimation also arises in the study of estimating Lyapunov exponents in time series, see e.g. Nychka *et al.* (1992). While state-dependent and threshold models are among the main parametric models for nonlinear time series, see e.g. Priestley (1988) and Tong (1990), nonparametric techniques which do not commit to any specific model form have become increasely popular. As a significant application of the general theory of local polynomial fitting to be discussed in Section 2, the use of local quadratic fitting for first-order partial derivative estimation is exploited in Section 3, where it is demonstrated as a flexible tool for modelling and characterizing nonlinearity in time series analysis. A *bias-correction* method for constructing confidence intervals for first-order partial derivatives is described in Subsection 3.1. In Subsection 3.2, two examples including the sunspots data, are used to demonstrate this application in time series.

The data considered in this paper

$$\{(X_0,Y_1),(X_1,Y_2),\ldots,(X_{n-1},Y_n)\} \tag{1.1}$$

where Y_i is the scalar response variable and X_{i-1} consists of the p predictor variables at time i are assumed to arise from the *martingale nonlinear regression* (MNR) model

$$Y_i = m(X_{i-1}) + \nu^{1/2}(X_{i-1})\varepsilon_i \tag{1.2}$$

where $m : \Re^p \to \Re$ is some nonlinear function, $\nu \geq 0$ is a variance function. The following assumptions on the model are made.

(A) $\{\varepsilon_i\}$ is a sequence of martingale differences with respect to a sequence of increasing σ-fields $\{\mathcal{F}_i\}$ such that X_0 is $\mathcal{F}_0$-measurable, X_i, ε_i are $\mathcal{F}_i$-measurable for all $i \geq 1$ and $E\{\varepsilon_i \mid \mathcal{F}_{i-1}\} = 0$, $E\{\varepsilon_i^2 \mid \mathcal{F}_{i-1}\} = 1$.

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

$$\sup_{\boldsymbol{u},\boldsymbol{v}\in\Re^p} \sum_{j=1}^{\infty} |f_j(\boldsymbol{u},\boldsymbol{v}) - f(\boldsymbol{u})f(\boldsymbol{v})| < \infty. \tag{1.3}$$

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:

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

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 $X_{i-1},\ldots$ 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

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

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 $\{x_1, x_2,\ldots,x_N\}$ correspond to (1.1) through

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

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 $m_1, \ldots, m_p$ at any given point $\boldsymbol{x} = (x_1, \ldots, x_p)^T$ are derived by minimizing the weighted sum of squares

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

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 $\hat{a} = \hat{m}(\boldsymbol{x})$, an estimate of regression function at $\boldsymbol{x}$, of $\hat{b} = \hat{D}_m(\boldsymbol{x})$ which corresponds to an estimate of $D_m(\boldsymbol{x}) = (\partial m(\boldsymbol{x})/\partial x_1, \ldots, \partial m(\boldsymbol{x})/\partial x_p)^T$ at $\boldsymbol{x}$, and of $\hat{L}$ which corresponds to estimates of elements in the Hessian matrix of $H_m(\boldsymbol{x}) = (\partial^2 m(\boldsymbol{x})/\partial x_i \partial x_j)$ at $\boldsymbol{x}$. That is, $L(x) = (l_{ij})$ satisfies $l_{ij} = h_{ij}$ if $i > j$ and $= h_{ii}/2$ if $i = j$, where $H_m(\boldsymbol{x}) = (h_{ij})$ is the Hessian. Let $\hat{\beta} = (\hat{a}, \hat{b}^T, \text{vech}^T\{\hat{L}\})^T$, and we have

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

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

$$(2.3)\qquad \boldsymbol{X} = \begin{pmatrix} 1 & (X_0 - \boldsymbol{x})^T & \text{vech}^T\{(X_0 - \boldsymbol{x})(X_0 - \boldsymbol{x})^T\} \\ \vdots & \vdots & \vdots \\ 1 & (X_{n-1} - \boldsymbol{x})^T & \text{vech}^T\{(X_{n-1} - \boldsymbol{x})(X_{n-1} - \boldsymbol{x})^T\} \end{pmatrix}.$$

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 $\boldsymbol{x} = (x_1, \ldots, x_p)^T$ in $\Re^p$, 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. $K = k(\|\boldsymbol{x}\|)$ for some function k. We denote $\mu_\ell = \int u_1^\ell K(\boldsymbol{u}) d\boldsymbol{u}$, $J_\ell = \int u_1^\ell K^2(\boldsymbol{u}) d\boldsymbol{u}$ for any nonnegative integers ℓ. In the Appendix, some commonly used multivariate kernel functions are defined, and formulae for higher-order moments are given. Also let I_1, I_2 denote the identity 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 $\int u_1^8 K(u_1, \ldots, u_p) du_1 \cdots du_p < \infty$.

THEOREM 1. *Under model* (1.2) *and Assumptions* (A)–(C), *consider any* ℓ *distinct points* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell$ *satisfying* $f(\boldsymbol{x}_j) > 0$, $\nu(\boldsymbol{x}_j) > 0$, *if there exist open neighborhoods* U_i *of* $\boldsymbol{x}_i$ *such that* $m \in C^3(U_j)$, $f \in C^1(U_j)$, $\nu \in C^0(U)$, $j = 1, 2, \ldots, \ell$, *the local linear estimators* $\hat{\beta}_L(\boldsymbol{x}_1), \dots, \hat{\beta}_L(\boldsymbol{x}_\ell)$ *are asymptotically independent and jointly normal as* $h \to 0$, $nh^p \to \infty$. *In particular, at each point* $\boldsymbol{x} = (x_1, \dots, x_p)^T$, *we have that*

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

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

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

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

$$b(\boldsymbol{x}) = \begin{pmatrix} \mu_4 \frac{\partial^3 m(\boldsymbol{x})}{\partial x_1^3} + 3\mu_2^2 \sum_{i=2}^{p} \frac{\partial^3 m(\boldsymbol{x})}{\partial x_i^2 \partial x_1} \\ \mu_4 \frac{\partial^3 m(\boldsymbol{x})}{\partial x_2^3} + 3\mu_2^2 \sum_{i \neq 2} \frac{\partial^3 m(\boldsymbol{x})}{\partial x_i^2 \partial x_2} \\ \vdots \\ \mu_4 \frac{\partial^3 m(\boldsymbol{x})}{\partial x_p^3} + 3\mu_2^2 \sum_{i=1}^{p-1} \frac{\partial^3 m(\boldsymbol{x})}{\partial x_i^2 \partial x_p} \end{pmatrix}, \tag{2.5}$$

$$b_1(\boldsymbol{x}) = \begin{pmatrix} (\mu_4 - \mu_2^2) \frac{\partial^2 m(\boldsymbol{x})}{\partial x_1^2} \frac{\partial f(\boldsymbol{x})}{\partial x_1} + 2\mu_2^2 \sum_{i=2}^{p} \frac{\partial^2 m(\boldsymbol{x})}{\partial x_1 \partial x_i} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ (\mu_4 - \mu_2^2) \frac{\partial^2 m(\boldsymbol{x})}{\partial x_2^2} \frac{\partial f(\boldsymbol{x})}{\partial x_2} + 2\mu_2^2 \sum_{i \neq 2} \frac{\partial^2 m(\boldsymbol{x})}{\partial x_2 \partial x_i} \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ (\mu_4 - \mu_2^2) \frac{\partial^2 m(\boldsymbol{x})}{\partial x_p^2} \frac{\partial f(\boldsymbol{x})}{\partial x_p} + 2\mu_2^2 \sum_{i=1}^{p-1} \frac{\partial^2 m(\boldsymbol{x})}{\partial x_p \partial x_i} \frac{\partial f(\boldsymbol{x})}{\partial x_p} \end{pmatrix}, \tag{2.6}$$

and

$$\Sigma(\boldsymbol{x}) = \begin{pmatrix} \frac{\nu(\boldsymbol{x}) J_0}{f(\boldsymbol{x})} & 0 \\ 0 & \frac{\nu(\boldsymbol{x}) J_2}{\mu_2{}^2 f(\boldsymbol{x})} I_1 \end{pmatrix}.$$

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 $\int u_1^{12} K(u_1, \ldots, u_p) du_1 \cdots du_p < \infty$.

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

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

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

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

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

$$\theta(\boldsymbol{x}) = \frac{\mu_4^2 - \mu_2\mu_6}{\mu_4 - \mu_2^2} \sum_{i=1}^{p} \frac{\partial^4 m(\boldsymbol{x})}{\partial x_i^4} - 6\mu_2^2 \sum_{1 \le i < j \le p} \frac{\partial^4 m(\boldsymbol{x})}{\partial x_i^2 \partial x_j^2},$$

$$\theta_1(\boldsymbol{x}) = \frac{\mu_4^2 - \mu_2\mu_6}{\mu_4 - \mu_2^2} \sum_{i=1}^{p} \frac{\partial^3 m(\boldsymbol{x})}{\partial x_i^3} \frac{\partial f(\boldsymbol{x})}{\partial x_i} - 3\mu_2^2 \sum_{\substack{1 \le i,j \le p \\ i \ne j}} \frac{\partial^3 m(\boldsymbol{x})}{\partial x_i \partial x_j^2} \frac{\partial f(\boldsymbol{x})}{\partial x_i},$$

and $h^2\gamma(\boldsymbol{x})$ and $h^2\gamma_1(\boldsymbol{x})$ are defined in Lu (1996). *Furthermore,*

$$(2.7) \qquad \Sigma(\boldsymbol{x}) =$$
$$\begin{pmatrix} \dfrac{\rho\nu(\boldsymbol{x})}{f(\boldsymbol{x})} & 0 & \dfrac{\phi\nu(\boldsymbol{x})}{f(\boldsymbol{x})} \operatorname{vech}^T\{I_1\} \\ 0 & \dfrac{J_2\nu(\boldsymbol{x})}{\mu_2^2 f(\boldsymbol{x})} I_1 & 0 \\ \dfrac{\phi\nu(\boldsymbol{x})}{f(\boldsymbol{x})} \operatorname{vech}\{I_1\} & 0 & \dfrac{\nu(\boldsymbol{x})}{f(\boldsymbol{x})}\left(\Lambda - \dfrac{\mu_2(J_2 - J_0\mu_2)}{(\mu_4 - \mu_2^2)^2} \operatorname{vech}\{I_1\} \operatorname{vech}^T\{I_1\}\right) \end{pmatrix},$$

where

$$\rho = (\mu_4 - \mu_2^2)^{-2}\{J_0(\mu_4 + (p-1)\mu_2^2)^2 - 2pJ_2\mu_2(\mu_4 + (p-1)\mu_2^2) \\ + p\mu_2^2(J_4 + (p-1)J_2^2)\},$$

$$\phi = (\mu_4 - \mu_2^2)^{-2}\{J_2\mu_4 + (2p-1)J_2\mu_2^2 - (p-1)J_2^2\mu_2 \\ - J_4\mu_2 - J_0\mu_2\mu_4 - (p-1)J_0\mu_2^3\},$$

$$\Lambda = \operatorname{diag}\left\{\lambda_1, \underbrace{\lambda_2, \ldots, \lambda_2}_{p-1}, \lambda_1, \underbrace{\lambda_2, \ldots, \lambda_2}_{p-2}, \cdots, \lambda_1, \lambda_2, \lambda_1\right\},$$

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

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

where

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

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

$$(2.9)\quad \begin{cases} Z_n = (nh^p)^{-1/2}D^{-1}\boldsymbol{X}^TWV^{1/2}E = (nh^p)^{-1/2}\sum_{i=1}^n Z_{ni};\\ \quad\text{where}\ \ E = (\varepsilon_1, \ldots, \varepsilon_n)^T,\\ \quad\phantom{\text{where}\ \ } V = \mathrm{diag}\{\nu(X_0), \ldots, \nu(X_{n-1})\};\\ \text{and}\\ Z_{ni} = \begin{pmatrix} 1\\ \left(\dfrac{X_{i-1}-\boldsymbol{x}}{h}\right)\\ \mathrm{vech}\left\{\left(\dfrac{X_{i-1}-\boldsymbol{x}}{h}\right)\left(\dfrac{X_{i-1}-\boldsymbol{x}}{h}\right)^T\right\} \end{pmatrix} K\left(\dfrac{X_{i-1}-\boldsymbol{x}}{h}\right)\nu^{1/2}(X_{i-1})\varepsilon_i. \end{cases}$$

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

$$D_m^k(\boldsymbol{x}, \boldsymbol{u}) = \sum_{i_1, \ldots, i_p} C^k_{i_1\cdots i_p}\frac{\partial^k m(\boldsymbol{x})}{\partial x_1^{i_1}\partial x_2^{i_2}\cdots\partial x_p^{i_p}}u_1^{i_1}\cdots u_p^{i_p},$$

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

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* $f(\boldsymbol{x}) > 0$, *we have as* $nh^p \to \infty$,

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

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

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

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

where

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

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

$$Z_n \to N(0, \Sigma_1), \tag{2.12}$$

where

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

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

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

$$(nh^p)^{1/2} \text{diag}\{1, hI, h^2 I_2\}\{\hat{\beta}(\boldsymbol{x}) - \beta(\boldsymbol{x}) - B_n\} = S_n^{-1} Z_n, \tag{2.13}$$

where $B_n = \text{diag}\{1, h^{-1}I, h^{-2}I_2\} S_n^{-1} R_n$.

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

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

by same calculations as in Lu (1996).

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

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

Meanwhile $B(\boldsymbol{x}, h)$ 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. □

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

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

where

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

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. Shiryayev ((1984), p. 511), we only need to check the Lindeberg condition:

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

$$= \sum_{i=1}^{n} \frac{1}{\epsilon^\delta} \frac{1}{(nh^p)^{(2+\delta)/2}} \mathrm{E}\left\{ \left| l\left(\frac{X_{i-1}-\boldsymbol{x}}{h}\right) \nu^{1/2}(X_{i-1})\varepsilon_i \right|^{2+\delta} \mid \mathcal{F}_{i-1} \right\}$$

$$= \frac{1}{(nh^p)^{1+\delta/2}\epsilon^\delta} \sum_{i=1}^{n} \left| l\left(\frac{X_{i-1}-\boldsymbol{x}}{h}\right) \nu^{1/2}(X_{i-1}) \right|^{2+\delta} E\{|\varepsilon_i|^{2+\delta} \mid \mathcal{F}_{i-1}\}$$

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

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

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

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

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

Furthermore,

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

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

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

where

$$\int l(\boldsymbol{u})^2 d\boldsymbol{u} = \int \{a + b\boldsymbol{u} + c^T \operatorname{vech}\{\boldsymbol{u}\boldsymbol{u}^T\}\}^2 K(\boldsymbol{u})^2 d\boldsymbol{u}$$
$$= (a, b, c^T) \int \begin{pmatrix} 1 \\ \boldsymbol{u} \\ \operatorname{vech}\{\boldsymbol{u}\boldsymbol{u}^T\} \end{pmatrix} (1, \boldsymbol{u}^T, \operatorname{vech}^T\{\boldsymbol{u}\boldsymbol{u}^T\}) \cdot K(\boldsymbol{u})^2 d\boldsymbol{u} (a, b, c^T)^T$$
$$+ o(1).$$

So by Theorem 4 of Shiryayev ((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

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

By the Cramer-Wold device, this implies that

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

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

3. Partial derivative estimation

In this section, we consider only the scalar time series model (1.4). For the state vector by $X_i = (x_i, x_{i-1}, \ldots, x_{i-p+1})^T$, we denote the p first-order partial derivative functions (pdfs) by $m_1(X_i) = \frac{\partial m(X_i)}{\partial x_i}, \ldots, m_p(X_i) = \frac{\partial m(X_i)}{\partial x_{i-p+1}}$. 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 $m_1, \ldots, m_p$ 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 $m_1, \ldots, m_p$ at any given point $\boldsymbol{x}$ 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 $\hat{D}_m(\boldsymbol{x})$ 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 $(h^2/(6\mu_2))\hat{b}(\boldsymbol{x})$. 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 the third-order derivative estimators are consistent so that $\hat{b}(\boldsymbol{x})$ is a consistent estimator for the bias term $b(\boldsymbol{x})$.

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

$$(\boldsymbol{X}^T W \boldsymbol{X})^{-1} \boldsymbol{X}^T W V W \boldsymbol{X} (\boldsymbol{X}^T W \boldsymbol{X})^{-1} \tag{3.1}$$

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

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

and

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

where $\hat{m}_{NW}$ is the Nadaraya-Watson estimator given by

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

Under general conditions, it can be shown that $\hat{f}(\boldsymbol{x})$ and $\hat{\nu}(\boldsymbol{x})$ are consistent. For example, if f is differentiable in an open neighborhood of $\boldsymbol{x}$, then the mixing condition (C) together with a simple application of Chebyshev's inequality implies that

$$\hat{f}(\boldsymbol{x}) = f(\boldsymbol{x}) + \int K(\boldsymbol{u})(h_2 D_f^T(\boldsymbol{x})\boldsymbol{u} + h_2^2 O(\|\boldsymbol{u}\|^2)d\boldsymbol{u} + O_p((nh_2^p)^{-1/2})),$$

as $nh_2^p \to \infty$, $h_2 \to 0$.

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

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

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

$$m(x_i, x_{i-1}) = (\phi_1 + \pi_1 e^{-\gamma_0 x_i^2})x_i + (\phi_2 + \pi_2 e^{-\gamma_0 x_i^2})x_{i-1}$$

where $\phi_1 = 1$, $\pi_1 = 0.8$, $\phi_2 = -0.25$, $\pi_2 = -1.5$, $\gamma_0 = 1$. Time series plot of the simulated data is shown in Fig. 1(a). The phase plot is given in Fig. 1(b).

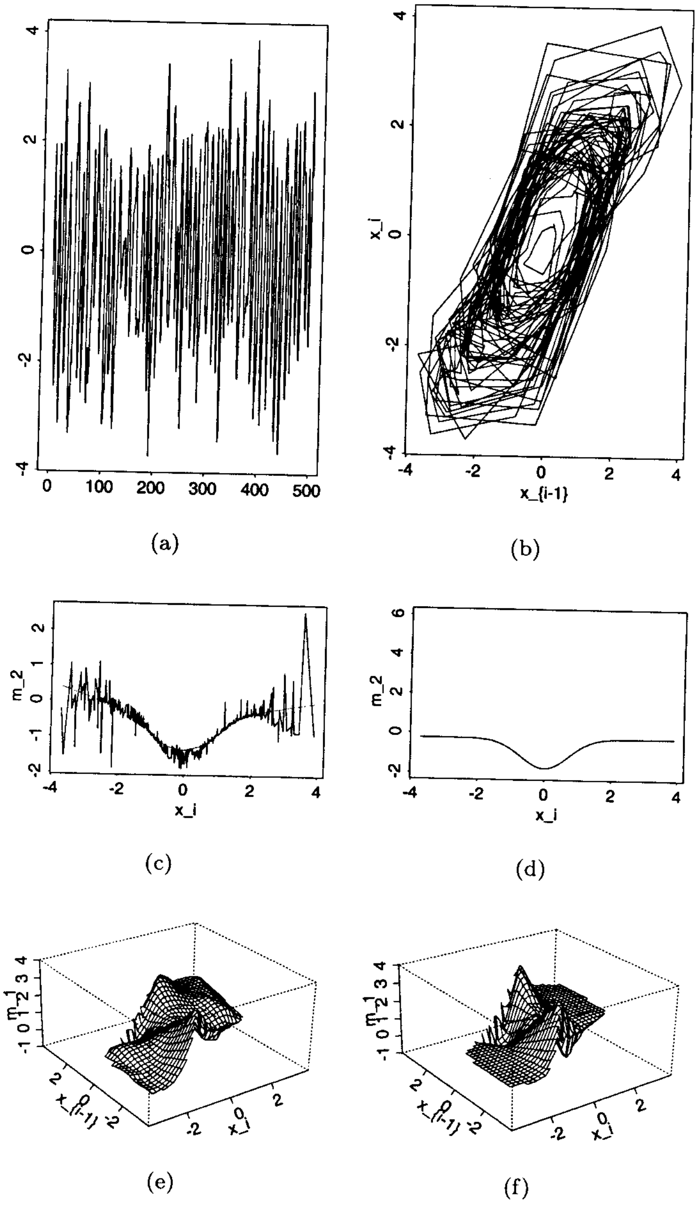

Fig. 1. Simulations from an exponential AR model: (a) time plot; (b) phase plot; (c) estimated curve of m_2; (d) true curve of m_2; (e) estimated surface of m_1; (f) true surface of m_1.

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

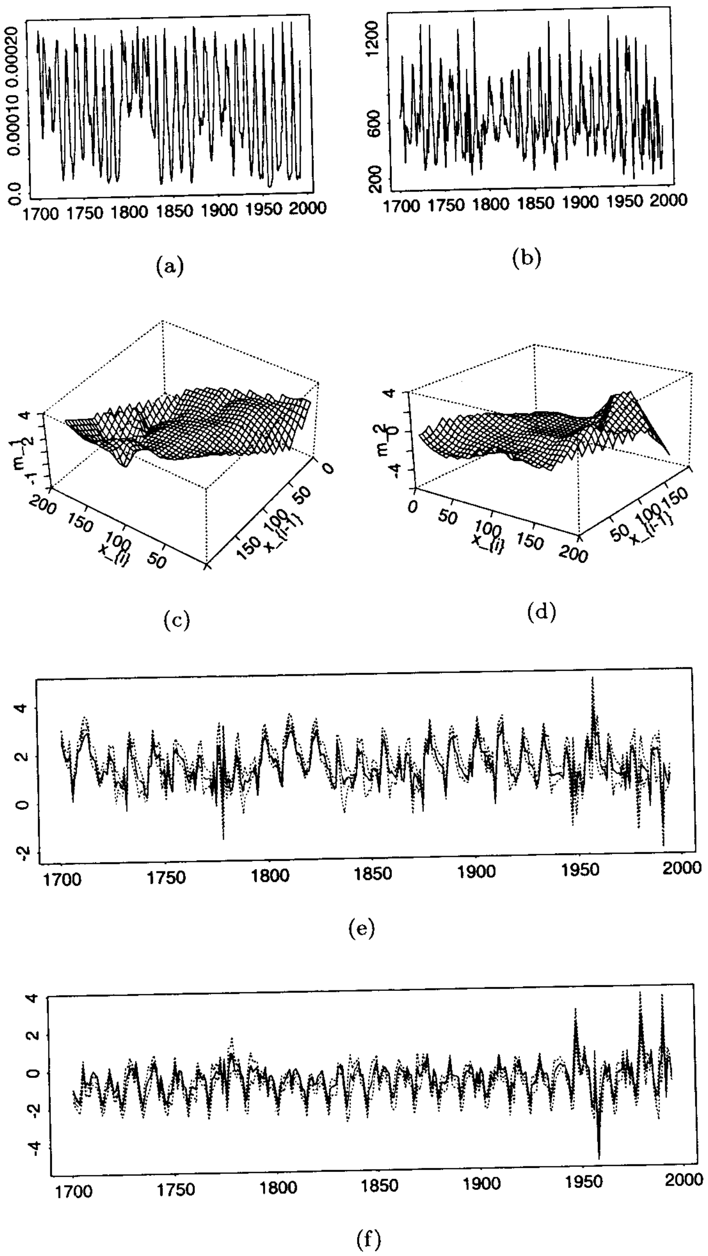

Fig. 2. Annual sunspots numbers from 1700–1996: (a) kernel density estimation of f; (b) N-W estimation of conditional variance ν; (c) surface plot of estimate of pdf m_1; (d) surface plot of estimate of pdf m_2; (e) time plots estimates (raw estimate: solid line, bias-adjusted: dashed line) and 95% CIs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

smooth line) with respect to variable x_i only. Note that though the raw estimated function $\hat{m}_2$ is not smooth as a function of x_i—this is probably due to the artifact that it is estimated as a function of both x_i, x_{i-1}, the smoothed version resembles the true m_2 in Fig. 1(d) quite well.

Figure 1(e) shows the surface plot (after some interpolation) of estimates $\hat{m}_1$. Figure 1(d) shows the true m_1 (under the same interpolation). It appears that the shape of pdf estimates is fairly close to the truth.

Example 2. (The annual sunspots numbers 1700–1996) In this example, we apply the bias-correction method described in Subsection 3.1 to the sunspots data. We use $p = 2$ and bandwidth $h = 40$ except in the unusual years 1955–1960 when $h = 70$ is used. We choose secondary bandwidths by the rule $h_4 = 1.2h$, $h_2 = 0.8h$, $h_3 = 1.2h$. The time plot of estimates of density f is given in Fig. 2(a). Estimates of conditional variance ν are shown in Fig. 2(b), which indicates that the heteroscedastic model is more appropriate. Figures 2(c) and 2(d) show the interpolated surfaces of estimates of first-order partial derivatives $m_1(X_i) = \partial m(x_i, x_{i-1})/\partial x_i$ and $m_2(X_i) = \partial m(x_i, x_{i-1})/\partial x_{i-1}$. It appears that at small sunspots numbers m_1 has largest magnitude while m_2 has largest negative magnitude. This observation is consistent with the nonlinear and asymmetric nature of this series. The variabilities of these estimates are given in Figs. 2(e) and 2(f) which show estimates (including both raw estimates and bias-adjusted ones) and 95% pointwise confidence intervals.

Appendix: Some useful kernels and their moments

Some useful multivariate kernel functions and their higher-order moments are given. For normal kernel $K_g(\boldsymbol{x}) = (2\pi)^{-p/2}\exp(-\|\boldsymbol{x}\|^2/2)$,

$$\mu_{2m} = 1\cdot 3\cdots(2m-1), \qquad J_{2m} = \mu_{2m}/(2^m\pi^{p/2}).$$

For uniform kernel $K_u(\boldsymbol{x}) = C_b^{-1}1_{\{\|\boldsymbol{x}\|\le 1\}}$ where $C_b = \pi^{p/2}/\Gamma(\frac{p+2}{2})$, $J_{2m} = \mu_{2m}/C_b$ and

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

A large family of kernels is given by the power family

$$K_{\alpha\beta}(\boldsymbol{x}) = \begin{cases} C_{\alpha\beta}^{-1}(1-\|\boldsymbol{x}\|^{\alpha})^{\beta}, & \text{if } \|\boldsymbol{x}\|\le 1, \\ 0, & \text{otherwise,} \end{cases}$$

for $\beta > -1$, $\alpha > 0$, where $C_{\alpha\beta}$ is given by: $C_{\alpha\beta} = 2\pi^{p/2}B(\beta+1, \frac{p}{\alpha})/(\alpha\Gamma(\frac{p}{2}))$, where B is the beta function, Γ is the gamma function. Some important cases: $K_{21}(\alpha = 2, \beta = 1)$: Epanechnikov; $K_{22}(\alpha = 2, \beta = 2)$: biweight; $K_{23}(\alpha = 2, \beta =$

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

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

$$J_{2m} = B\left(\frac{p+2m}{\alpha}, 2\beta+1\right) B\left(\frac{p-1}{2}, m+\frac{1}{2}\right) \Big/ \cdot \left\{B\left(\frac{p}{\alpha}, \beta+1\right) B\left(\frac{p-1}{2}, \frac{1}{2}\right) C_{\alpha,\beta}\right\}.$$

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 Gijbel, 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.

Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes, *Scand. J. Statist.*, **24**(2), 165–179.

Mills, T. C. (1993). *The Econometric Modelling of Financial Time Series*, Cambridge University Press, Cambridge.

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.