ELSEVIER

Journal of Statistical Planning and Inference 68 (1998) 221–245

journal of statistical planning and inference

Nonparametric vector autoregression

W. Härdle[a], A. Tsybakov[b], L. Yang[a,*]

[a] *Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, Institut für Statistik und Ökonometrie, Spandauer Strasse 1, D-10178 Berlin, Germany*

[b] *Université Paris VI, Laboratoire de Statistique Théorique et Appliquée, 4, pl. Jussieu, Tour 45-55, F-75252 Paris, France*

Received 1 March 1996; received in revised form 17 December 1996; accepted 6 January 1997

Abstract

We consider a vector conditional heteroscedastic autoregressive nonlinear (CHARN) model in which both the conditional mean and the conditional variance (volatility) matrix are unknown functions of the past. Nonparametric estimators of these functions are constructed based on local polynomial fitting. We examine the rates of convergence of these estimators and give a result on their asymptotic normality. These results are applied to estimation of volatility matrices in foreign exchange markets. Estimation of the conditional covariance surface for the Deutsche Mark/US Dollar (DEM/USD) and Deutsche Mark/British Pound (DEM/GBP) daily returns show negative correlation when the two series have opposite lagged values and positive correlation elsewhere. The relation of our findings to the capital asset pricing model is discussed.

1. Nonparametric vector autoregression

Multivariate time series occur in many scientific disciplines. Their analysis helps in modeling dynamics over time as well as explaining interdependence among variables. A common model in this context is vector autoregression where the dynamics over time are modeled via a linear operation on the past values of the vector time series, see Lütkepohl (1991). Typically, in these models the conditional covariance is assumed to be either fixed or of specific form. Since the beginning of the 1980s the drawback of fixed linear structures has been stressed by Engle (1982), Robinson (1983, 1984) and Teräsvirta (1994) in the econometric literature and by Collomb (1984), Tjøstheim (1994), McKeague and Zhang (1994), and Vieu (1994) in the statistical literature. Nonlinear time-series models that have been proposed are, e.g., threshold autoregressive (TAR) models of Tong (1978, 1983), the exponential autoregressive (EXPAR) models of Haggan and Ozaki (1981), the smooth-transition autoregressive (STAR) models of Chan and Tong (1986) and Granger and Teräsvirta (1992).

* Corresponding author. E-mail: yang@wiwi.hu-berlin.de.

PII S0378-3758(97)00143-2

In the analysis of financial time series, e.g., exchange rates, models for conditional heteroscedasticity are an important feature. Meese and Rose (1991) state that "*it is now recognized that empirical exchange rate models of the post-Bretton Woods era are characterized by parameter instability and dismal forecast performance...*" This pessimism about the quality of exchange-rate models became generally accepted after the publication of the influential papers by Meese and Rogoff (1983) and Diebold and Nason (1990).

The nonparametric modeling of the mean function and the volatility matrix offers a way out of this pessimism. It does not depend on specific structures of these quantities and may thus lead to valuable suggestions. In the framework of ARCH models (Engle, 1982), non- and semi-parametric approaches (Gregory, 1989; Engle and Gonzalez-Rivera, 1991) have been proposed. Engle and Ng (1993) measured the impact of news on volatility and found asymmetric volatility functions. Gouriéroux and Monfort (1992) models both the conditional mean and the conditional variance in a flexible nonparametric way

$$Y_i = \sum_{j=1}^{J} \alpha_j \, I(X_i \in A_j) + \sum_{j=1}^{J} \beta_j \, I(X_i \in A_j) \xi_i, \quad i = 1, 2, \ldots,$$

$$X_i = (Y_{i-1}^{\mathrm{T}}, Y_{i-2}^{\mathrm{T}}, \ldots, Y_{i-m}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{md}, \quad Y_i \in \mathbb{R}^d \tag{1.1}$$

is called a qualitative threshold ARCH model. Here $\{A_j\}_{j=1}^{J}$ with fixed J denotes a partition of the set of lagged values for Y, (α_j), and (β_j) are unknown parameter vectors and matrices, respectively, and ξ_i is the white noise. It is a generalization of the threshold model (Tong, 1983), for the conditional mean but shares with it the drawback of a fixed number J of threshold points.

A generalization of model (1.1) to a wider class of conditional mean and variance functions can be seen as a limit of (1.1) for $J \to \infty$, thus allowing J to be unknown

$$Y_i = f(X_i) + \Sigma^{1/2}(X_i)\xi_i, \quad i = 1, 2, \ldots,$$

$$X_i = (Y_{i-1}^{\mathrm{T}}, Y_{i-2}^{\mathrm{T}}, \ldots, Y_{i-m}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{md}, \quad Y_i \in \mathbb{R}^d. \tag{1.2}$$

We call (1.2) a conditional heteroskedastic autoregressive nonlinear (CHARN) model. It is a generalization of an ARCH structure.

The use of CHARN modeling is motivated by several examples. It has been found that GARCH(1, 1) processes fit daily and weekly FX (foreign exchange) rates well in most cases. The situation for intra-daily data is different though, see Guillaume et al. (1994).

Drost and Nijman (1993) argued that the specific GARCH structure would not allow arbitrary combinations of conditional heteroskedasticity, and leptokurtocity, for example. Typically, for intra-daily data the deviation of the unconditional return density from normality increases when the sampling interval is decreased. The model (1.2) will not suffer from these effects since it neither makes structural assumptions on f and Σ nor distributional assumptions on ξ. The situation for the CHARN model is

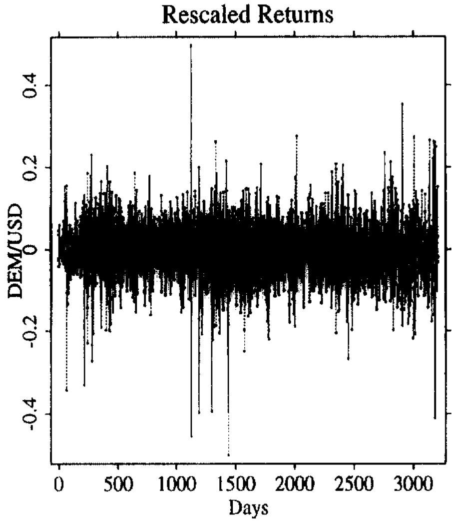

Fig. 1. The daily returns of the exchange rates of DEM/USD from 2 January 1980 to 30 October 1992.

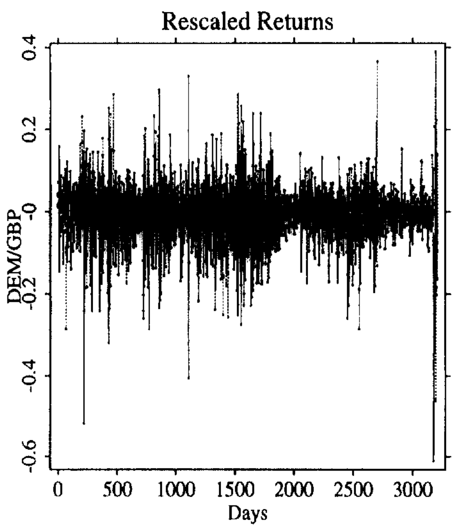

Fig. 2. The daily returns of the exchange rates of DEM/GBP from 2 January 1980 to 30 October 1992.

depicted in Figs. 1–3. All computations and graphics are done in XploRe, see Härdle et al. (1995).

Figs. 1 and 2 show the daily returns (differences of log spot rates) of $Y_{i1} = \text{DEM/USD}$ (Deutsche Mark/US Dollar) and of $Y_{i2} = \text{DEM/GBP}$ (Deutsche Mark/British Pound) for the period from 2 January 1980 to 30 October 1992, a total of 3212 observations: both are rescaled so that the range always has length 1. Fig. 3 shows that the two returns are highly correlated, the correlation equals 0.34, and the squared returns (i.e.

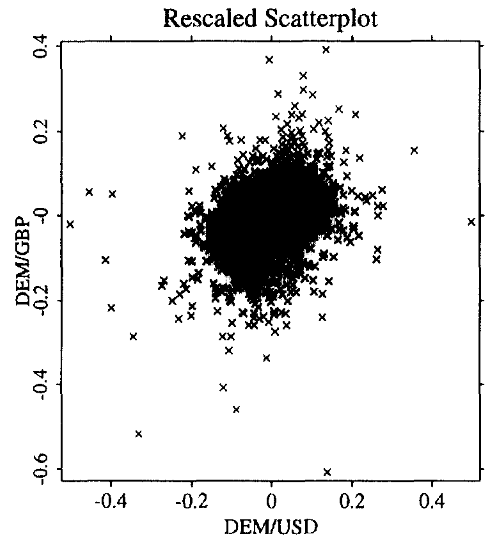

Fig. 3. The daily returns of the exchange rates of both DEM/USD and DEM/GBP from 2 January 1980 to 30 October 1992.

Y_{i1}^2 and Y_{i2}^2) also have a correlation of 0.17. Both are statistically significantly different from zero, for a sample size of 3212.

Figs. 4 and 5 display the conditional covariance function as dependent on one lag. Thus, in (1.2) we have $d=2$, $m=1$ and the task is to estimate

$$f(x)=(f_1,f_2)^{\mathrm{T}}(x) \quad \text{and} \quad \Sigma(x)=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}(x).$$

There exists a negative correlation when the two returns have opposite lagged values, which correspond to the upper left and the lower right corners of the contour plot or the lowest contour level at about 15.76% below which are the negative values, while positive correlations are everywhere else. Both the computation and graphics are done in XploRe, using the WARPing technique (Härdle et al., 1995), subsequent work in Section 4 is done in the same fashion and uses the same single bandwidth.

Härdle and Tsybakov (1996) proposed a general class of joint mean and volatility-function estimators based on the local polynomial (LP) method in the case of one-lag-dependence model (1.2) with one-dimensional Y_i. The LP estimator was chosen in favor of the Nadaraya–Watson (NW) estimator, since the NW estimator does not achieve good asymptotic convergence rates, unless the marginal (stationary) density of X_i is sufficiently many times differentiable. Sufficient conditions for such a property to hold in the model (1.2) are not known. The LP method avoids this difficulty, since it needs only the continuity of the density of X_i. A more practical reason to use the LP method is that it corresponds to a local least-squares problem, and for this problem easy and efficient algorithms are available. Bossaerts et al. (1996) used this method to study foreign exchange rates. For large dimension d and many lags m, however, the

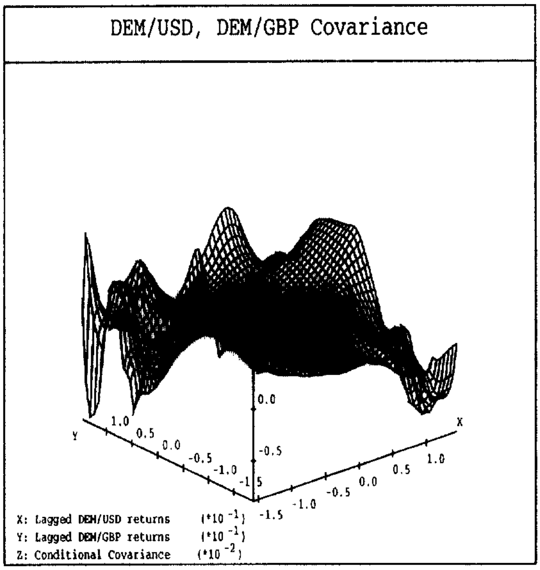

Fig. 4. The conditional covariance, using bandwidth $h = 0.0536531$.

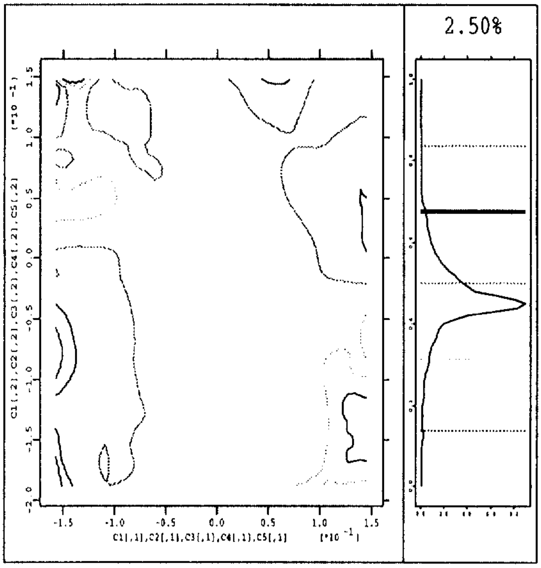

Fig. 5. The contours of the conditional covariance.

precision of the estimators of both f and Σ will decrease. A structured modelling based on additive assumptions has therefore been proposed by Chen and Tsay (1993a, b).

The idea of local polynomial estimation goes back to Stone (1977), Cleveland (1979) and Katkovnik (1979, 1985). The statistical properties of LP estimators in nonparametric regression (convergence, minimax rate of convergence and pointwise asymptotic normality) were studied by Tsybakov (1986). The LP estimation method was later discussed by several authors (see Fan and Gijbels, 1996, for references). For the multidimensional case, we refer to the work of Ruppert and Wand (1994) who studied the multivariate local linear regression estimation.

This paper is devoted to estimation of the $f(\cdot)$ and $\Sigma(\cdot)$ functions for the multivariate CHARN model. We generalize to the vector case the result of Härdle and Tsybakov (1996) on asymptotic normality of LP estimators. We restrict the study, however, to the local linear case. This is motivated by the fact that higher-order polynomial estimation in higher dimension is less attractive computationally, while the expressions for asymptotic bias and variance are much more technical, and they do not seem to be of practical use.

Inspection of the proofs in Section 5 shows that the result of the present paper also holds (with obvious reformulation) for the multivariate nonparametric regression model with heteroskedastic errors: $Y_i = f(X_i) + \Sigma^{1/2}(X_i)\xi_i$, where ξ_i are as in (1.1), (X_i, Y_i) are i.i.d., and the design points $\{X_i\}$ are independent of $\{\xi_i\}$.

We shall use the work on probabilistic properties of the process (1.2): Doukhan and Ghindés (1980, 1981), Chan and Tong (1985), Mokkadem (1987), Diebolt and Guégan (1990), Ango Nze (1992). In these papers the ergodicity, geometric ergodicity and mixing properties of the process $\{Y_i\}$ are derived under appropriate conditions.

The paper is organized as follows. In Section 2, we present the estimator and in Section 3 we study the asymptotic properties of this LP technique. In Section 4 we give an application based on the two-dimensional data of DEM/USD and DEM/GBP returns. In Section 5, proofs of theorems are given.

2. The estimators

The model we consider is

$$Y_i = f(X_i) + \Sigma^{1/2}(X_i)\xi_i, \quad i = m, m+1, \ldots, \tag{2.1}$$

where $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{id})^{\mathrm{T}} \in \mathbb{R}^d$, $\xi_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{id})^{\mathrm{T}} \in \mathbb{R}^d$, $i = m, m+1, \ldots, n$, and $X_i = (Y_{i-1}^{\mathrm{T}}, Y_{i-2}^{\mathrm{T}}, \ldots, Y_{i-m}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{md}$ are random vector variables; ξ_i are i.i.d. with $E(\xi_{1j}) = 0$, for any $1 \leqslant j \leqslant d$, $E(\xi_{1j}^2) = 1$. The mean vector function $f : \mathbb{R}^{md} \to \mathbb{R}^d$ and volatility matrix function $\Sigma : \mathbb{R}^{md} \to \mathbb{R}^d \times \mathbb{R}^d$ are unknown, $\Sigma(x)$ is symmetric and positive definite for any $x \in \mathbb{R}^{md}$, and the initial value $X_m = (Y_{m-1}^{\mathrm{T}}, Y_{m-2}^{\mathrm{T}}, \ldots, Y_0^{\mathrm{T}})^{\mathrm{T}}$ is a random vector variable independent of $\{\xi_i\}$. We study the problem of estimating the conditional volatility matrix function $\Sigma(x)$ and the conditional mean vector function $f(x)$, given a time series $Y_0, \ldots, Y_n$.

The technique we employ here is typical in multivariate problems. Instead of Σ and f, we can equivalently estimate the following functions:

- The mean function of $v^{\mathrm{T}}Y$, which is $f(x; v) = v^{\mathrm{T}}f(x)$, where $v \in \mathbb{R}^d$ has unit length and $x \in \mathbb{R}^{md}$;

- The covariance function of $v^{\mathrm{T}}Y$ and $s^{\mathrm{T}}Y$, which is $v^{\mathrm{T}}\Sigma(x)s$, where $v, s \in \mathbb{R}^d$ both have unit length and $x \in \mathbb{R}^{md}$.

For the moment we are implicitly assuming stationarity of $\{Y_i\}$. In fact, only an approximation is true: $\{X_i\}$ approaches a stationary process, for $i \to \infty$ as we shall see later in Lemma 3.1. The LP method solves the following minimization problems:

$$
\begin{aligned}
c_n(x; v, s) &= \arg \min_{c \in \mathbb{R}^{md+1}} \sum_{i=m}^{n} (v^{\mathrm{T}} Y_i Y_i^{\mathrm{T}} s - c^{\mathrm{T}} U_{in})^2 K_h(X_i - x), \\
c_n(x; v) &= \arg \min_{c \in \mathbb{R}^{md+1}} \sum_{i=m}^{n} (v^{\mathrm{T}} Y_i - c^{\mathrm{T}} U_{in})^2 K_h(X_i - x),
\end{aligned}
\tag{2.2}
$$

where $K : \mathbb{R}^{md} \to \mathbb{R}^1$ is a kernel $K_h(u) = 1/h^{md} K(u/h)$, $h = h_n$ is a positive number (bandwidth), $h_n \to 0$, as $n \to \infty$ and

$$
U_{in} = F(u_{in}), \qquad u_{in} = \frac{X_i - x}{h}, \tag{2.3}
$$

where $F(u) = \binom{1}{u} \in \mathbb{R}^{md+1}$, for $u \in \mathbb{R}^{md}$. The estimator of $f(x; v)$ is defined as

$$
\hat{f}(x; v) = c_n(x; v)^{\mathrm{T}} F(0).
$$

The estimator of the function $\sigma(x; v, s) = v^{\mathrm{T}} \Sigma(x) s$ is defined as

$$
\hat{\sigma}(x; v, s) = c_n(x; v, s)^{\mathrm{T}} F(0) - \{c_n(x; v)^{\mathrm{T}} F(0)\}\{c_n(x; s)^{\mathrm{T}} F(0)\}. \tag{2.4}
$$

We have dropped reference to the sample size n in $\hat{f}(x; v)$ and $\hat{\sigma}(x; v, s)$ for notational simplicity, we will keep this convention in similar situations hereafter. Another simplification of notation is the use of one single bandwidth in all coordinates of X. The asymptotic results in the next section are easily extendable to the case of different bandwidth in each direction, e.g., in a product kernel

$$
K_h(u) = \left(\prod_{j=1}^{md} h_j \right)^{-1} \prod_{j=1}^{md} K\left(\frac{u_j}{h_j} \right),
$$

where $h = (h_1, \ldots, h_{md}) \in \mathbb{R}_+^{md}$, see Wand and Jones (1995).

3. The asymptotic results

Let $|\cdot|$ denote the L_1-norm when it is applied to a vector, and the usual matrix norm

$$
|A| = \sup_{|x|=1} |Ax|,
$$

when it is applied to a matrix A. Assume the following:

(A1) The error variables ξ_{1j}, $1 \leqslant j \leqslant d$, are i.i.d. The density $p(\cdot)$ of ξ_1 exists and satisfies

$$
\inf_{x \in \mathscr{K}} p(x) > 0
$$

for any compact $\mathscr{K} \subset \mathbb{R}^d$. Also $E(\xi_{1j}) = E(\xi_{1j}^3) = 0$, $E(\xi_{1j}^2) = 1$, and $E(\xi_{1j}^4) = 1 + m_4 < \infty$.

(A2) There exist constants $C_1 \geq 0$, $C_2 \geq 0$, $r > 0$ such that for $|x| \geq r$

$$|f(x)| \leq C_1(1 + |x|), \tag{3.1}$$

$$|\Sigma^{1/2}(x)| \leq C_2(1 + |x|). \tag{3.2}$$

(A3) The matrix function $\Sigma(x)$ is symmetric for any $x \in \mathbb{R}^{md}$, and satisfies

$$\inf_{x \in \mathscr{K}} \lambda_{\min}\{\Sigma(x)\} > \lambda_{\mathscr{K}} > 0,$$

for any compact $\mathscr{K} \subset \mathbb{R}^{md}$, where $\lambda_{\min}(\Sigma)$ denotes the minimal eigenvalue of a real symmetric matrix Σ.

(A4) $C_1 + C_2 E|\xi_1| < 1/m$.

Assumption (A1) is needed for identifiability of the estimation procedure. Assumptions (A1) and (A3) guarantee that the process $\{X_i\}$ does not die out whereas (A2) and (A4) are conditions for $\{X_i\}$ not to explode. The following lemma given by Ango Nze (1992) guarantees ergodicity of the process $\{X_i\}$. It is based on the application of the results of Nummelin and Tuominen (1982) and Tweedie (1975). Note that (A4) becomes redundant when both $f(x)$ and $\Sigma(x)$ are bounded, in which case $C_1 = C_2 = 0$.

Lemma 3.1. *Under the conditions* (A1)–(A4) *the Markov chain* $\{X_i\}$ *is geometrically ergodic, i.e. it is ergodic, with stationary probability measure* $\pi(\cdot)$ *such that, for almost every* x, *as* $k \to \infty$

$$\|P^k(\cdot \mid x) - \pi(\cdot)\|_{TV} = O(\rho^k)$$

for some $0 \leq \rho < 1$. *Here*

$$P^k(B \mid x) = P\{X_k \in B \mid X_m = x\}$$

for a Borel subset $B \subset \mathbb{R}^{md}$, *and* $\|\cdot\|_{TV}$ *is the total variation distance.*

Now we state the conditions necessary to derive joint asymptotic normality of $\hat{f}(x; v)$ and $\hat{\sigma}(x; v, s)$ at a fixed point $x \in \mathbb{R}^{md}$.

(A5) The functions f and Σ are componentwise twice continuously differentiable at the point $x \in \mathbb{R}^{md}$.

(A6) The density $\mu(\cdot)$ of the stationary distribution $\pi(\cdot)$ exists, is bounded, continuous and strictly positive in a neighborhood of the point x.

(A7) The kernel K is a compactly supported bounded nonnegative function on $\mathbb{R}^{md}$, such that

$$\int K(u)\,du = 1, \qquad \int uK(u)\,du = 0, \qquad \int uu^{\mathrm{T}}K(u)\,du = \sigma_K^2 I_{md},$$

where $\sigma_K^2 > 0$, and I_{md} denotes the identity matrix of dimension md.

(A8) $h_n = \beta n^{-1/(4+md)}$, where $\beta > 0$.
(A9) The initial value X_m is a fixed vector in $\mathbb{R}^{md}$.

Condition (A5) is a smoothness condition for the functions f and Σ. Note that it is related to (A8), the optimal speed of bandwidth. Condition (A8) guarantees a balance between bias and variance. A faster speed of h would lead to undersmoothing, a slower rate would increase the bias over the standard deviation of the estimator by oversmoothing. Both situations are undesirable since they result in less precise estimation. Condition (A6) is necessary to compute asymptotic bias and variance, (A7) is a typical assumption for kernels. Assumption (A9) supposes that the CHARN model is started at some fixed vector.

Let $f_j(x)$ and $\sigma_{jk}(x)$, $j,k = 1,2,\ldots,d$, be the components of the vector function $f(x)$ and the matrix function $\Sigma(x)$, respectively. Denote $\|K\|_2^2 = \int K^2(u)\,\mathrm{d}u$. Asymptotic normality results are presented in the following theorems.

Theorem 1. *Under the assumptions* (A1)–(A9)

$$n^{\frac{2}{4+md}}\{\hat{f}(x;v) - v^{\mathrm{T}} f(x)\} \xrightarrow{\mathcal{L}} \mathcal{N}\{b(x;v), V(x;v)\} \tag{3.3}$$

as $n \to \infty$ *with*

$$b(x;v) = \beta^2 \frac{\sigma_K^2}{2} \operatorname{Tr}[\nabla^2(v^{\mathrm{T}} f(x))]$$

and

$$V(x;v) = \beta^{-md} \frac{v^{\mathrm{T}}\Sigma(x)v}{\mu(x)} \|K\|_2^2.$$

In particular, if one let v be the jth or the kth coordinate vector of $\mathbb{R}^d$, one gets the following joint asymptotic distribution:

$$n^{\frac{2}{4+md}} \begin{pmatrix} \hat{f}_j(x) - f_j(x) \\ \hat{f}_k(x) - f_k(x) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{ \begin{pmatrix} b_j(x) \\ b_k(x) \end{pmatrix}, \begin{pmatrix} V_j(x) & c_{jk}(x) \\ c_{jk}(x) & V_k(x) \end{pmatrix} \right\} \tag{3.4}$$

as $n \to \infty$ *with*

$$b_j(x) = \beta^2 \frac{\sigma_K^2}{2} [\operatorname{Tr}(\nabla^2 f_j(x))]$$

and

$$V_j(x) = \beta^{-md} \frac{\sigma_{jj}(x)}{\mu(x)} \|K\|_2^2, \qquad c_{jk}(x) = \beta^{-md} \frac{\sigma_{jk}(x)}{\mu(x)} \|K\|_2^2.$$

Denote

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

for any vector

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

Theorem 2. *Under the assumptions* (A1)–(A9)

$$n^{\frac{2}{4+md}}\{\hat{\sigma}(x;v,s) - v^{\mathrm{T}}\Sigma(x)s\} \xrightarrow{\mathscr{D}} \mathcal{N}\{b(x;v,s), V(x;v,s)\} \tag{3.5}$$

as $n \to \infty$ *with*

$$b(x;v,s) = \beta^2 \frac{\sigma_K^2}{2}[\,\mathrm{Tr}\{\nabla^2 g(x)\} - \{s^{\mathrm{T}} f(x)\}\,\mathrm{Tr}\{\nabla^2 f^{\mathrm{T}}(x)v\}]$$
$$- \beta^2 \frac{\sigma_K^2}{2}[\{v^{\mathrm{T}} f(x)\}\,\mathrm{Tr}\{\nabla^2 f^{\mathrm{T}}(x)s\}]$$

and

$$V(x;v,s) = \beta^{-md}\frac{\|K\|_2^2}{\mu(x)}[(m_4-2)T^*(x) + \{v^{\mathrm{T}}\Sigma(x)s\}^2]$$
$$+ \beta^{-md}\frac{\|K\|_2^2}{\mu(x)}\{v^{\mathrm{T}}\Sigma(x)v\}\{s^{\mathrm{T}}\Sigma(x)s\},$$

where

$$g(x) = g(x;v,s) = \{v^{\mathrm{T}} f(x)\}\{s^{\mathrm{T}} f(x)\} + \{v^{\mathrm{T}}\Sigma(x)s\},$$
$$T^*(x) = T^*(x;v,s) = \mathrm{Tr}[\mathrm{diag}^2\{v^{\mathrm{T}}\Sigma^{1/2}(x)\}\,\mathrm{diag}^2\{\Sigma^{1/2}(x)s\}].$$

The covariance of $\hat{\sigma}(x;v,s)$ *and* $\hat{\sigma}(x;v',s')$ *is*

$$\beta^{-md}\frac{\|K\|_2^2}{\mu(x)}(m_4-2)$$
$$\times\,\mathrm{Tr}[\,\mathrm{diag}\{v^{\mathrm{T}}\Sigma^{1/2}(x)\}\,\mathrm{diag}\{\Sigma^{1/2}(x)s\}\,\mathrm{diag}\{v'^{T}\Sigma^{1/2}(x)\}\,\mathrm{diag}\{\Sigma^{1/2}(x)s'\}]$$
$$+\beta^{-md}\frac{\|K\|_2^2}{\mu(x)}[\{v^{\mathrm{T}}\Sigma(x)v'\}\{s^{\mathrm{T}}\Sigma(x)s'\} + \{v^{\mathrm{T}}\Sigma(x)s'\}\{s^{\mathrm{T}}\Sigma(x)v'\}].$$

In particular, if one let v *and* s *be the* j*th and* k*th coordinate vectors of* $\mathbb{R}^d$ *or the* j'*th and* k'*th coordinate vectors, one gets*

$$n^{\frac{2}{4+md}}\begin{pmatrix} \hat{\sigma}_{jk}(x) - \sigma_{jk}(x) \\ \hat{\sigma}_{j'k'}(x) - \sigma_{j'k'}(x) \end{pmatrix} \xrightarrow{\mathscr{D}} \mathcal{N}\left\{\begin{pmatrix} b_{jk}(x) \\ b_{j'k'}(x) \end{pmatrix}, \begin{pmatrix} V_{jk}(x) & c_{jk,j'k'}(x) \\ c_{jk,j'k'}(x) & V_{j'k'}(x) \end{pmatrix}\right\} \tag{3.6}$$

as $n \to \infty$ *with*

$$b_{jk}(x) = \beta^2\frac{\sigma_K^2}{2}[\,\mathrm{Tr}\{\nabla^2\sigma_{jk}(x) + 2\nabla^{\mathrm{T}} f_j(x)\nabla f_k(x)\}], \quad V_{jk}(x) = c_{jk,jk}(x),$$

where

$$c_{jk,j'k'}(x) = \beta^{-md} \frac{\|K\|_2^2}{\mu(x)} \{(m_4 - 2)T^*_{jk,j'k'}(x) + \sigma_{jj'}(x)\sigma_{kk'}(x) + \sigma_{jk'}(x)\sigma_{kj'}(x)\}$$

and

$$T^*_{jk,j'k'}(x) = \sum_{l=1}^{d} s_{jl}(x)s_{j'l}(x)s_{kl}(x)s_{k'l}(x)$$

in which $s_{jl}(x)$ *denotes the* (j,l)*th entry of the matrix* $\Sigma^{1/2}(x)$. *Finally, as* $n \to \infty$

$$n^{\frac{2}{4+md}} \begin{pmatrix} \hat{\sigma}_{jk}(x) - \sigma_{jk}(x) \\ \hat{f}_{j'}(x) - f_{j'}(x) \end{pmatrix} \xrightarrow{\mathscr{D}} \mathscr{N}\left\{ \begin{pmatrix} b_{jk}(x) \\ b_{j'}(x) \end{pmatrix}, \begin{pmatrix} V_{jk}(x) & 0 \\ 0 & V_{j'}(x) \end{pmatrix} \right\}. \tag{3.7}$$

The practical use of these results lies in the possibility to check the form of the mean and volatility functions. For instance, at each point x we can construct a confidence interval for $\sigma_{jk}(x)$ based on plug-in estimates for $b_{jk}(x)$ and $V_{jk}(x)$. The bias conceivably can be estimated from a local cubic estimate. The variance can be estimated by first calculating the stochastic innovation term $\hat{\xi}_{ij}^2 = \{Y_{ij} - \hat{f}_j(X_i)\}^2/\hat{\sigma}_{jj}(X_i)$ and then setting $\hat{m}_4 = d^{-1}\sum_{j=1}^{d} n^{-1}\sum_{i=m}^{n}(\hat{\xi}_{ij}^2 - 1)^2$. The marginal density μ can be estimated as usual by a kernel estimator. Since the bias formula is slightly more involved than the variance formula, some undersmoothing might be recommended.

4. Application

The importance of the CHARN model for financial data has been pointed out in the introduction. In this section we come back to the introductory example of DEM/USD and DEM/GBP exchange rates. Figs. 6 and 7 show the estimated conditional mean functions $\hat{f}_1(x)$ and $\hat{f}_2(x)$ as functions of the lagged values $x_i = (y_{1,i-1}, y_{2,i-1})^{\mathrm{T}}$. The surface and the contour plots all show that the mean functions are rather flat and are around zero. In fact, 80% of the $\hat{f}_1(x)$ values are in an interval around 0 whose length is only 0.11 times of the range of $y_{1,i}$, while 80% of the $\hat{f}_2(x)$ values are in an interval around 0 whose length is only 0.1557 times of the range of $y_{2,i}$. The pattern of the conditional covariance function $\hat{\sigma}_{12}(x)$ is different though, it changes from negative to positive as shown in Figs. 4 and 5.

Bollerslev et al. (1988, 1992), studied the *capital asset pricing model* (CAPM) by means of the multivariate GARCH model. To illustrate the connection between our vector CHARN model and their model, consider a random vector Y_t of excess asset returns with $E(Y_t \mid \mathscr{F}_{t-1}) \equiv \mu_t$ and $\mathrm{Var}(Y_t \mid \mathscr{F}_{t-1}) \equiv \Sigma_t$, where $\mathscr{F}_{t-1}$ is the information set generated by Y_{t-i}, $i = 1, 2, \ldots,$. If for nonnegative weight vector w_t whose elements add to 1, $w_t^{\mathrm{T}} Y_t$ is a mean–variance efficient portfolio, then the CAPM is

$$Y_t = \beta_t \mu_t^m + \varepsilon_t,$$

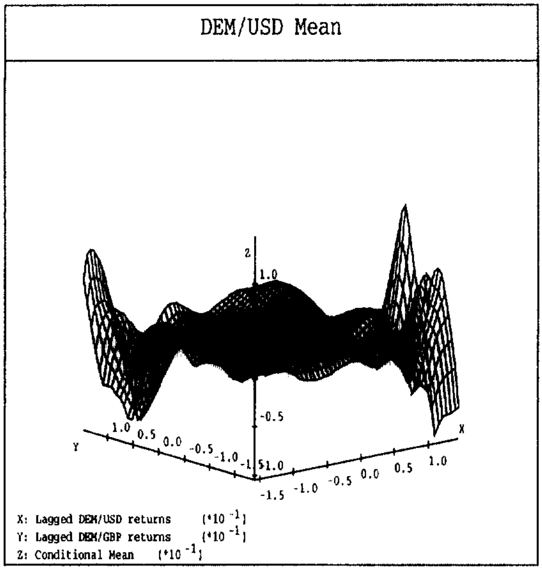

Fig. 6. The conditional mean function of the DEM/USD daily returns.

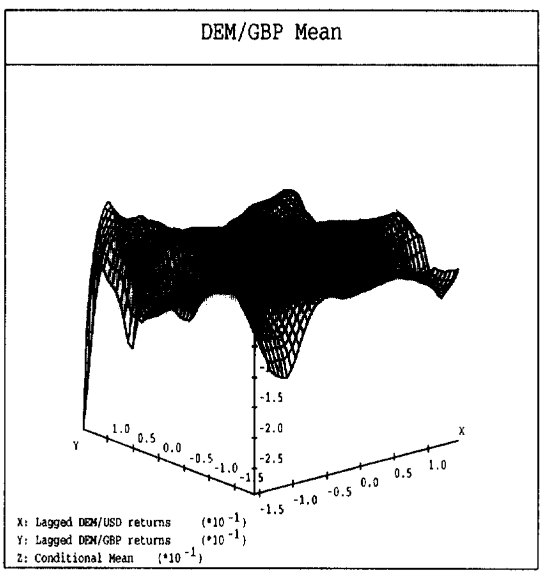

Fig. 7. The conditional mean function of the DEM/GBP daily returns.

where

$$\beta_t \equiv \Sigma_t w_t / w_t^{\mathrm{T}} \Sigma_t w_t,$$

with $E(\varepsilon_t \mid \mathscr{F}_{t-1}) \equiv 0$, $\mathrm{Var}(\varepsilon_t \mid \mathscr{F}_{t-1}) \equiv \Sigma_t$, and $\mu_t^m = w_t^{\mathrm{T}} \mu_t$. This is more general than ordinary CAPM which restricts Σ_t to be constant. While our CHARN model would

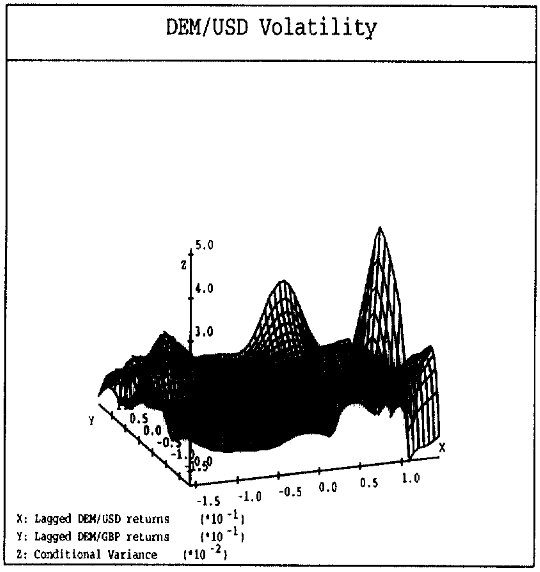

Fig. 8. The conditional variance function of the DEM/USD daily returns.

stipulate that Σ_t depends nonparametrically on a finite number of past observations, Bollerslev et al. (1988) used the multivariate GARCH model which allows Σ_t to depend on infinite number of past values, but only parametrically. A special form of the multivariate GARCH model is

$$\Sigma_t = \sigma Y_{t-1} Y_{t-1}^{\mathrm{T}}$$

for some constant $\sigma > 0$ in which case

$$\sigma_{12}(Y_{t-1}) = \sigma Y_{t-1,1} Y_{t-1,2}.$$

This is a hyperbolic function which exhibits the pattern visible in Figs. 4 and 5. For such a case, our CHARN model and the multivariate GARCH model would yield similar results.

Figs. 8 and 9 show the estimated conditional variance functions $\hat{\sigma}_{11}(x)$ and $\hat{\sigma}_{22}(x)$ as functions of the lagged values $x_i = (y_{1,i-1}, y_{2,i-1})^{\mathrm{T}}$. One can see that the variance function for the DEM/USD returns has a parabolic shape while that for DEM/GBP is roughly flat and positive.

5. Proofs

The proofs of Theorems 1 and 2 proceed in the following steps. First the normal equations of the LS problems (2.3) for the mean- and second-moment functions are solved. All estimators are split into a stochastic part and a systematic bias

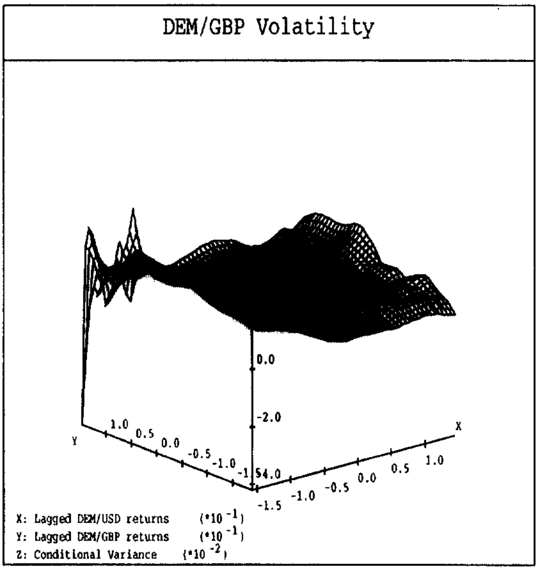

Fig. 9. The conditional variance function of the DEM/GBP daily returns.

part. Lemma 3.1 is essential in controling the stochastic part. Lemma 5.1 guarantees the strong mixing property of the recursive scheme (1.2). In combination with Lemmas 5.2–5.5 we then prove the joint asymptotic normality of the mean estimation as stated in Theorem 1 and that of volatility as stated in Theorem 2.

Set the matrices $W = \operatorname{diag}\{\frac{1}{n}K_h(X_i - x)\}_{i=m}^n$ and

$$Z = \begin{pmatrix} 1 & \cdots & 1 \\ \frac{X_m - x}{h} & \cdots & \frac{X_n - x}{h} \end{pmatrix}.$$

Define

$$v^{\mathrm{T}}Y = \begin{pmatrix} v^{\mathrm{T}}Y_m \\ \vdots \\ v^{\mathrm{T}}Y_n \end{pmatrix} = \begin{pmatrix} v^{\mathrm{T}}f(X_m) + v^{\mathrm{T}}\Sigma^{1/2}(X_m)\xi_m \\ \vdots \\ v^{\mathrm{T}}f(X_n) + v^{\mathrm{T}}\Sigma^{1/2}(X_n)\xi_n \end{pmatrix}$$

and also

$$v^{\mathrm{T}}YY^{\mathrm{T}}s = \begin{pmatrix} v^{\mathrm{T}}Y_mY_m^{\mathrm{T}}s \\ \vdots \\ v^{\mathrm{T}}Y_nY_n^{\mathrm{T}}s \end{pmatrix}$$

$$= \begin{pmatrix} (v^{\mathrm{T}}f(X_m) + v^{\mathrm{T}}\Sigma^{1/2}(X_m)\xi_m)(s^{\mathrm{T}}f(X_m) + s^{\mathrm{T}}\Sigma^{1/2}(X_m)\xi_m) \\ \vdots \\ (v^{\mathrm{T}}f(X_n) + v^{\mathrm{T}}\Sigma^{1/2}(X_n)\xi_n)(s^{\mathrm{T}}f(X_n) + s^{\mathrm{T}}\Sigma^{1/2}(X_n)\xi_n) \end{pmatrix}.$$

Then

$$\hat{f}(x;v) = F(0)^{\mathrm{T}}(ZWZ^{\mathrm{T}})^{-1}ZW[v^{\mathrm{T}}Y] \tag{5.1}$$

and

$$\hat{\sigma}(x;v,s) = F(0)^{\mathrm{T}}(ZWZ^{\mathrm{T}})^{-1}ZW[v^{\mathrm{T}}YY^{\mathrm{T}}s] - \hat{f}(x;v)\hat{f}(x;s) \tag{5.2}$$

by direct calculations.

First, to have the limit of $(ZWZ^{\mathrm{T}})^{-1}$, we need an auxiliary result based on Lemma 3.1.

Lemma 5.1 (Davydov, 1973). *A geometrically ergodic Markov chain whose initial variable is distributed with its stationary distribution is geometrically strongly mixing with the mixing coefficients satisfying* $\alpha(n) \leqslant c_0\,\rho_0^n$ *for some* $0<\rho_0<1$, $c_0>0$.

Having Lemma 5.1, the next lemma follows:

Lemma 5.2. *Under the conditions of Theorem 1 we have*

$$n^{-\frac{4}{4+md}} \sum_{i=m}^{n} \varphi_1(X_i)\,\varphi_2(u_{in})\,K(u_{in}) \xrightarrow{\mathrm{p}} \beta^{md}\mu(x)\,\varphi_1(x)\int \varphi_2(u)\,K(u)\,\mathrm{d}u, \tag{5.3}$$

$$n^{-\frac{4}{4+md}} \sum_{i=m}^{n} E\{\varphi_1(X_i)\,\varphi_2(u_{in})\,K(u_{in})\} \longrightarrow \beta^{md}\mu(x)\,\varphi_1(x)\int \varphi_2(u)\,K(u)\,\mathrm{d}u$$

as $n\to\infty$, *provided* $\varphi_1(\cdot)$ *is a bounded continuous function in a neighborhood of* x *and* $\varphi_2(\cdot)$ *is a bounded measurable function.*

Proof. See Härdle and Tsybakov (1996, Lemma 4.3). □

Lemma 5.3. *As* $n\to\infty$,

$$(ZWZ^{\mathrm{T}})^{-1} = \frac{1}{\mu(x)}\begin{bmatrix} 1 & 0_{1\times md} \\ 0_{md\times 1} & \sigma_K^{-2} I_{md}\end{bmatrix}\{1+\mathrm{o_p}(1)\} \tag{5.4}$$

uniformly in a compact neighborhood of x.

Proof. The elements of ZWZ^{T} are all in the form of the left-hand side of (5.3). Using assumption (A7) and then taking matrix inverse, one gets (5.4). □

Now notice that, in view of Lemma 5.3

$$\begin{aligned}\hat{f}(x;v) - v^{\mathrm{T}}f(x) &= F(0)^{\mathrm{T}}(ZWZ^{\mathrm{T}})^{-1}ZW[v^{\mathrm{T}}Y] - v^{\mathrm{T}}f(x)\\ &= F(0)^{\mathrm{T}}(ZWZ^{\mathrm{T}})^{-1}ZW[v^{\mathrm{T}}Y]\\ &\quad - F(0)^{\mathrm{T}}(ZWZ^{\mathrm{T}})^{-1}(ZWZ^{\mathrm{T}})\begin{bmatrix} v^{\mathrm{T}}f(x) \\ h\nabla(v^{\mathrm{T}}f(x))\end{bmatrix}\end{aligned}$$

$$= F(0)^{\mathrm{T}}(ZWZ^{\mathrm{T}})^{-1}ZW \left[v^{\mathrm{T}}Y - Z^{\mathrm{T}} \begin{bmatrix} v^{\mathrm{T}}f(x) \\ h\nabla(v^{\mathrm{T}}f(x)) \end{bmatrix} \right]$$

$$= \frac{1}{\mu(x)n}\{1+o_{\mathrm{p}}(1)\}$$

$$\times \sum_{i=m}^{n} K_h(X_i - x)[v^{\mathrm{T}}f(X_i) - v^{\mathrm{T}}f(x) - (X_i - x)^{\mathrm{T}}\nabla\{v^{\mathrm{T}}f(x)\}]$$

$$+ \frac{1}{\mu(x)n}\{1+o_{\mathrm{p}}(1)\} \sum_{i=m}^{n} K_h(X_i - x)\{v^{\mathrm{T}}\Sigma^{1/2}(X_i)\xi_i\}. \quad \square \tag{5.5}$$

To prove Theorem 1, one separates (5.5) into a bias part and a stochastic part as usual. The bias part is handled by the following lemma:

Lemma 5.4. *Let $g : \mathbb{R}^{md} \to \mathbb{R}^1$ be a twice continuously differentiable function. Then, under the assumptions of Theorem* 1

$$\frac{1}{\mu(x)n} \sum_{i=m}^{n} K_h(X_i - x)[g(X_i) - g(x) - (X_i - x)^{\mathrm{T}}\nabla g(x)]$$

$$= h^2 \frac{\sigma_K^2}{2} \operatorname{Tr}[\nabla^2 g(x)] + o_{\mathrm{p}}(h^2).$$

Proof. Using the Taylor expansion of $g(x)$, we get

$$\frac{1}{\mu(x)n} \sum_{i=m}^{n} K_h(X_i - x)[g(X_i) - g(x) - (X_i - x)^{\mathrm{T}}\nabla g(x)]$$

$$= \frac{1}{\mu(x)nh} \sum_{i=m}^{n} K(u_{in})[g(X_i) - g(x) - hu_{in}^{\mathrm{T}}\nabla g(x)]$$

$$= \frac{1}{2\mu(x)nh} \sum_{i=m}^{n} K(u_{in})[h^2 u_{in}^{\mathrm{T}}\nabla^2 g(x)u_{in}] + R,$$

where

$$|R| \leqslant \frac{1}{\mu(x)nh} \sum_{i=m}^{n} K(u_{in}) \left[h^2\mathscr{D}^2 \sup_{|w|\leqslant\mathscr{D}} |\nabla^2 g(x+hw) - \nabla^2 g(x)| \right]$$

$$= \frac{o(h^2)}{nh} \sum_{i=m}^{n} K(u_{in}) = o_{\mathrm{p}}(h^2) \tag{5.6}$$

as $n \to \infty$, where $\mathscr{D} = \max\{|w|: w \in \operatorname{supp} K\}$ and the last equality in (5.6) is due to Lemma 5.2. Again, by Lemma 5.2 one has, as $n \to \infty$

$$\frac{1}{2\mu(x)nh} \sum_{i=m}^{n} K(u_{in})[u_{in}^{\mathrm{T}}\nabla^2 g(x)u_{in}] = \frac{1}{2\mu(x)nh} \sum_{i=m}^{n} \operatorname{Tr}[K(u_{in})u_{in}u_{in}^{\mathrm{T}}\nabla^2 g(x)]$$

$$\xrightarrow{\mathrm{p}} \frac{1}{2} \int \operatorname{Tr}[K(u)uu^{\mathrm{T}}\nabla^2 g(x)]\,\mathrm{d}u$$

$$= \frac{1}{2}\,\mathrm{Tr}\left[\int K(u)uu^{\mathrm{T}}\,\mathrm{d}u\nabla^2 g(x)\right]$$

$$= \frac{\sigma_K^2}{2}\,\mathrm{Tr}[\nabla^2 g(x)].$$

Combining this with (5.6) we get the lemma.

In particular, if $g(x)=v^{\mathrm{T}}f(x)$, one gets from Lemma 5.4

$$\frac{1}{\mu(x)n}\sum_{i=m}^{n} K_h(X_i-x)[v^{\mathrm{T}}f(X_i)-v^{\mathrm{T}}f(x)-(X_i-x)^{\mathrm{T}}\nabla\{v^{\mathrm{T}}f(x)\}]$$

$$= b(x;v)n^{-2/(4+md)}+\mathrm{o}_{\mathrm{p}}(n^{-2/(4+md)}) \tag{5.7}$$

as $n\to\infty$, where $b(x;v)$ is as given in Theorem 1. This yields the asymptotics of the bias term in (5.5).

To work out the asymptotics of the variance term, we need another lemma. Denote $\mathscr{F}_{k-1}=\sigma(X_k,X_{k-1},\ldots,X_m)$ the σ-algebra generated by $X_m,\ldots,X_k$. □

Lemma 5.5 (Liptser and Shirjaev, 1980, Corollary 6). *Let m be a fixed integer and for every $n\geqslant m$, let the sequence $\eta^n=(\eta_{nk},\mathscr{F}_k)$ be a square integrable martingale difference, i.e.*

$$E(\eta_{nk}\mid\mathscr{F}_{k-1})=0,\quad E(\eta_{nk}^2)<\infty,\quad m\leqslant k\leqslant n, \tag{5.8}$$

and let

$$\sum_{k=m}^{n} E(\eta_{nk}^2)=1,\quad \forall n\geqslant n_0\geqslant m. \tag{5.9}$$

The conditions

$$\sum_{k=m}^{n} E(\eta_{nk}^2\mid\mathscr{F}_{k-1})\xrightarrow{\mathrm{p}}1,\quad \text{as } n\to\infty, \tag{5.10}$$

$$\sum_{k=m}^{n} E(\eta_{nk}^2 I(|\eta_{nk}|>\varepsilon)\mid\mathscr{F}_{k-1})\xrightarrow{\mathrm{p}}0,\quad \text{as } n\to\infty, \tag{5.11}$$

are sufficient for convergence

$$\sum_{k=m}^{n}\eta_{nk}\xrightarrow{\mathscr{L}}\mathscr{N}(0,1),\quad \text{as } n\to\infty.$$

Proof of Theorem 1. Now we apply Lemma 5.5 to the following stochastic term of (5.5)

$$\sum_{i=m}^{n}\frac{1}{\mu(x)n}K_h(X_i-x)v^{\mathrm{T}}\Sigma^{1/2}(X_i)\xi_i \tag{5.12}$$

and observe that (in view of Lemma 5.2)

$$G_n = \sum_{i=m}^{n} E\left\{\left(\frac{1}{\mu(x)n}K_h(X_i - x)v^{\mathrm T}\Sigma^{1/2}(X_i)\xi_i\right)^2\right\} \tag{5.13}$$

$$= \frac{1}{\mu(x)nh^{md}}\{1+\mathrm o(1)\}\int K^2(u)v^{\mathrm T}\Sigma(x)v\,\mathrm du$$

$$= n^{-4/(4+md)}V(x;v)\{1+\mathrm o(1)\}.$$

Define

$$\eta_{nk} = K_h(X_k - x)v^{\mathrm T}\Sigma^{1/2}(X_k)\xi_k\frac{1}{\mu(x)n\sqrt{G_n}}.$$

It is clear from (5.13) and (5.3) that (5.8)–(5.10) hold. It remains to check (5.11) in order to show that

$$\sum_{k=m}^{n}\eta_{nk}\xrightarrow{\mathscr L}\mathcal N(0,1), \quad \text{as } n\to\infty. \tag{5.14}$$

We have

$$\eta_{nk}^2 \leqslant Q_{nk}\,|\xi_k|^2, \tag{5.15}$$

where

$$Q_{nk} = \frac{h^{md}\beta^{-md}}{V(x;v)\mu^2(x)n}K_h(X_k - x)^2|v^{\mathrm T}\Sigma^{1/2}(X_k)|^2\{1+\mathrm o(1)\}.$$

Note that for some constant $C(x,v)$ depending only on x and v

$$Q_{nk}\leqslant C(x,v)\frac{1}{nh^{md}}K\left(\frac{X_k - x}{h}\right),$$

because of the fact that K is compactly supported and that Σ is bounded in a shrinking neighborhood of x. This entails

$$E\left\{|\xi_k|^2 I\left(|\xi_k|^2\geqslant\frac{\varepsilon^2}{Q_{nk}}\right)\right\}\leqslant C_n(x,v),$$

where

$$C_n(x,v) = E\left\{|\xi_1|^2 I\left(|\xi_1|^2\geqslant\frac{\varepsilon^2 nh^{md}}{C(x,v)\,\|K\|_\infty}\right)\right\}\to 0,$$

independent of k. This and (5.15) yield

$$\sum_{k=m}^{n}E\left\{\eta_{nk}^2\,I(|\eta_{nk}|\geqslant\varepsilon)\,|\,\mathscr F_{k-1}\right\} = \sum_{k=m}^{n}E\left\{\eta_{nk}^2\,I(\eta_{nk}^2\geqslant\varepsilon^2)\,|\,\mathscr F_{k-1}\right\}$$

$$\leqslant \sum_{k=m}^{n}Q_{nk}E\left\{|\xi_k|^2\,I\left(|\xi_k|^2\geqslant\frac{\varepsilon^2}{Q_{nk}}\right)\right\}$$

$$\leqslant C_n(x,v)C(x,v)\sum_{k=m}^{n}\frac{1}{nh^{md}}K\left(\frac{X_k - x}{h}\right),$$

while

$$\sum_{k=m}^{n} \frac{1}{nh^{md}} K\left(\frac{X_k - x}{h}\right) \xrightarrow{\mathrm{p}} \mu(x), \quad \text{as } n \to \infty,$$

by Lemma 5.2. Thus we have proved (5.14). Now (3.3) is a consequence of (5.5), (5.7), (5.13) and (5.14). To prove the joint asymptotic normality (3.4), note that, in view of (5.5) and (5.7),

$$n^{\frac{2}{4+md}} \begin{pmatrix} \hat{f}_j(x) - f_j(x) \\ \hat{f}_k(x) - f_k(x) \end{pmatrix} = \begin{pmatrix} b_j(x) \\ b_k(x) \end{pmatrix} \{1 + o_p(1)\} + n^{\frac{2}{4+md}} \begin{pmatrix} \zeta_{jn} \\ \zeta_{kn} \end{pmatrix} \{1 + o_p(1)\}$$

as $n \to \infty$, where

$$\zeta_{jn} = \frac{1}{\mu(x)n} \sum_{i=m}^{n} K_h(X_i - x) v_j^{\mathrm{T}} \Sigma^{1/2}(X_i)\xi_i$$

and v_j is the jth coordinate vector in $\mathbb{R}^d$.

By the Cramér–Wold device, the joint asymptotic normality of ζ_{jn} and ζ_{kn} is proved if one shows that linear combinations of these random variables satisfy

$$n^{\frac{2}{4+md}} (\alpha_j \zeta_{jn} + \alpha_k \zeta_{kn}) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \alpha_j^2 V_j(x) + \alpha_k^2 V_k(x) + 2\alpha_j \alpha_k c_{jk}(x)) \tag{5.16}$$

as $n \to \infty$, $\forall \alpha_j,\ \alpha_k \in \mathbb{R}^1$.

The proof of (5.16) is quite similar to that of (5.14), and it is based again on the application of Lemma 5.5. The difference is that instead of G_n, one should use now

$$G_n' = \sum_{i=m}^{n} E\left\{ \left(\frac{1}{\mu(x)n} K_h(X_i - x)(\alpha_j v_j + \alpha_k v_k)^{\mathrm{T}} \Sigma^{1/2}(X_i)\xi_i \right)^2 \right\}$$

$$= n^{-4/(4+md)} [\alpha_j^2 V_j(x) + \alpha_k^2 V_k(x) + 2\alpha_j \alpha_k c_{jk}(x)]\{1 + o(1)\},$$

where the last equality follows from Lemma 5.2 (cf. (5.13)). □

Proof of Theorem 2. Similar to (5.5), we write

$$\begin{aligned} \hat{\sigma}(x; v, s) - \sigma(x; v, s) &= (v^{\mathrm{T}} f(x))(s^{\mathrm{T}} f(x)) - \hat{f}(x; v)\hat{f}(x; s) \\ &\quad + F(0)^{\mathrm{T}} (ZWZ^{\mathrm{T}})^{-1} ZW[v^{\mathrm{T}} YY^{\mathrm{T}} s] \\ &\quad - (v^{\mathrm{T}} f(x))(s^{\mathrm{T}} f(x)) - v^{\mathrm{T}} \Sigma(x) s \\ &= (v^{\mathrm{T}} f(x))(s^{\mathrm{T}} f(x)) - \hat{f}(x; v)\hat{f}(x; s) \\ &\quad + F(0)^{\mathrm{T}} (ZWZ^{\mathrm{T}})^{-1} ZW \Bigg[v^{\mathrm{T}} YY^{\mathrm{T}} s \\ &\quad - Z^{\mathrm{T}} \begin{pmatrix} (v^{\mathrm{T}} f(x))(s^{\mathrm{T}} f(x)) + v^{\mathrm{T}} \Sigma(x) s \\ h\nabla((v^{\mathrm{T}} f(x))(s^{\mathrm{T}} f(x)) + v^{\mathrm{T}} \Sigma(x) s) \end{pmatrix} \Bigg] \end{aligned}$$

$$= (v^{\mathrm{T}} f(x))(s^{\mathrm{T}} f(x)) - \hat{f}(x;v)\hat{f}(x;s)$$

$$+ \frac{1}{\mu(x)n}\{1+\mathrm{o_p}(1)\}\sum_{i=m}^{n} K_h(X_i - x)[v^{\mathrm{T}} f(X_i) f(X_i)^{\mathrm{T}} s$$

$$- v^{\mathrm{T}} f(x) f(x)^{\mathrm{T}} s - (X_i - x)^{\mathrm{T}} \nabla\{(v^{\mathrm{T}} f(x))(s^{\mathrm{T}} f(x))\}]$$

$$+ \frac{1}{\mu(x)n}\{1+\mathrm{o_p}(1)\}\sum_{i=m}^{n} K_h(X_i - x)[v^{\mathrm{T}}\Sigma(X_i)s - v^{\mathrm{T}}\Sigma(x)s$$

$$- (X_i - x)^{\mathrm{T}}\nabla\{v^{\mathrm{T}}\Sigma(x)s\}]$$

$$+ \frac{1}{\mu(x)n}\{1+\mathrm{o_p}(1)\}\sum_{i=m}^{n} K_h(X_i - x)v^{\mathrm{T}}\Sigma^{1/2}(X_i)$$

$$\times (\xi_i\xi_i^{\mathrm{T}} - I_d)\Sigma^{1/2}(X_i)s$$

$$+ \frac{1}{\mu(x)n}\{1+\mathrm{o_p}(1)\}\sum_{i=m}^{n} K_h(X_i - x)$$

$$\times\{s^{\mathrm{T}} f(X_i)v^{\mathrm{T}} + v^{\mathrm{T}} f(X_i)s^{\mathrm{T}}\}\Sigma^{1/2}(X_i)\xi_i,$$

which after plugging in the formula for $\hat{f}(x;v) - v^{\mathrm{T}} f(x)$ and $\hat{f}(x;s) - s^{\mathrm{T}} f(x)$ (cf. (5.5)) yields

$$\hat{\sigma}(x;v,s) - \sigma(x;v,s) = \{1+\mathrm{o_p}(1)\}\sum_{j=1}^{8} T_j, \tag{5.17}$$

where

$$T_1 = \frac{1}{\mu(x)n}\sum_{i=m}^{n} K_h(X_i - x)[v^{\mathrm{T}} f(X_i) f(X_i)^{\mathrm{T}} s - v^{\mathrm{T}} f(x) f(x)^{\mathrm{T}} s - (X_i - x)^{\mathrm{T}}$$

$$\times \nabla\{(v^{\mathrm{T}} f(x))(s^{\mathrm{T}} f(x))\}],$$

$$T_2 = \frac{1}{\mu(x)n}\sum_{i=m}^{n} K_h(X_i - x)[v^{\mathrm{T}}\Sigma(X_i)s - v^{\mathrm{T}}\Sigma(x)s - (X_i - x)^{\mathrm{T}}\nabla\{v^{\mathrm{T}}\Sigma(x)s\}],$$

$$T_3 = -\frac{1}{\mu(x)n}\sum_{i=m}^{n} K_h(X_i - x)s^{\mathrm{T}} f(x)[v^{\mathrm{T}} f(X_i) - v^{\mathrm{T}} f(x) - (X_i - x)^{\mathrm{T}}\nabla\{v^{\mathrm{T}} f(x)\}],$$

$$T_4 = -\frac{1}{\mu(x)n}\sum_{i=m}^{n} K_h(X_i - x)v^{\mathrm{T}} f(x)[s^{\mathrm{T}} f(X_i) - s^{\mathrm{T}} f(x) - (X_i - x)^{\mathrm{T}}\nabla\{s^{\mathrm{T}} f(x)\}],$$

$$T_5 = \frac{1}{\mu(x)n}\sum_{i=m}^{n} K_h(X_i - x)v^{\mathrm{T}}\Sigma^{1/2}(X_i)(\xi_i\xi_i^{\mathrm{T}} - I_d)\Sigma^{1/2}(X_i)s,$$

$$T_6 = \frac{1}{\mu(x)n}\sum_{i=m}^{n} K_h(X_i - x)\{s^{\mathrm{T}} f(X_i) - s^{\mathrm{T}} f(x)\}\{v^{\mathrm{T}}\Sigma^{1/2}(X_i)\xi_i\},$$

$$T_7 = \frac{1}{\mu(x)n}\sum_{i=m}^{n} K_h(X_i - x)\{v^{\mathrm{T}} f(X_i) - v^{\mathrm{T}} f(x)\}\{s^{\mathrm{T}}\Sigma^{1/2}(X_i)\xi_i\},$$

$$T_8 = -[\hat{f}(x;v) - v^{\mathrm{T}} f(x)][\hat{f}(x;s) - s^{\mathrm{T}} f(x)]$$
$$= -n^{-4/(4+md)}[n^{2/(4+md)}(\hat{f}(x;v) - v^{\mathrm{T}} f(x))][n^{2/(4+md)}(\hat{f}(x;s) - s^{\mathrm{T}} f(x))]. \tag{5.18}$$

Using Lemma 5.4, one derives

$$T_1 = \beta^2 \frac{\sigma_K^2}{2}[\mathrm{Tr}\{\nabla^2((v^{\mathrm{T}} f(x))(s^{\mathrm{T}} f(x)))\}]n^{-2/(4+md)} + \mathrm{o_p}(n^{-2/(4+md)}),$$
$$T_2 = \beta^2 \frac{\sigma_K^2}{2}[\mathrm{Tr}\{\nabla^2(v^{\mathrm{T}}\Sigma(x)s)\}]n^{-2/(4+md)} + \mathrm{o_p}(n^{-2/(4+md)}),$$
$$T_3 = -\beta^2 \frac{\sigma_K^2}{2}\{s^{\mathrm{T}} f(x)\}\,\mathrm{Tr}\{\nabla^2 f^{\mathrm{T}}(x)v\}n^{-2/(4+md)} + \mathrm{o_p}(n^{-2/(4+md)}),$$
$$T_4 = -\beta^2 \frac{\sigma_K^2}{2}\{v^{\mathrm{T}} f(x)\}\,\mathrm{Tr}\{\nabla^2 f^{\mathrm{T}}(x)s\}n^{-2/(4+md)} + \mathrm{o_p}(n^{-2/(4-md)}), \tag{5.19}$$

and thus

$$T_1 + T_2 + T_3 + T_4 = b(x;v,s)n^{-2/(4+md)} + \mathrm{o_p}(n^{-2/(4+md)}). \tag{5.20}$$

Now we calculate T_6. Note that

$$|K_h(z-x)(s^{\mathrm{T}} f(z) - s^{\mathrm{T}} f(x))|$$
$$\leqslant K_h(z-x) \sup_{|w|\leqslant \mathscr{S}} |f(x+wh) - f(x)| \leqslant ChK_h(z-x),$$

since K is compactly supported (here $C>0$ is a constant). Thus,

$$E(T_6^2) = \frac{1}{\mu(x)^2 n^2} \sum_{i=m}^{n} E[K_h^2(X_i - x)\{s^{\mathrm{T}} f(X_i) - s^{\mathrm{T}} f(x)\}^2\{v^{\mathrm{T}}\Sigma(X_i)v\}]$$
$$\leqslant \frac{C^2h^2}{\mu(x)^2 n^2} \sum_{i=m}^{n} E[K_h^2(X_i - x)v^{\mathrm{T}}\Sigma(X_i)v].$$

By Lemma 5.2

$$\frac{h^2\beta^{md}}{nh^{md}} \sum_{i=m}^{n} E[K_h^2(X_i - x)v^{\mathrm{T}}\Sigma(X_i)v]$$
$$\to \beta^{md}\mu(x)v^{\mathrm{T}}\Sigma(x)v \int K^2(u)\,\mathrm{d}u = \mathrm{O}(1), \quad \text{as } n\to\infty,$$

and therefore

$$E(T_6^2) = \mathrm{O}\left(\frac{h^2}{nh^{md}}\right) = \mathrm{o}(n^{-4/(4+md)}), \quad \text{as } n\to\infty.$$

The evaluation of T_7 is quite analogous and, hence, we get

$$T_6 + T_7 = \mathrm{o_p}(n^{-2/(4+md)}), \quad \text{as } n\to\infty. \tag{5.21}$$

Also, in view of Theorem 1,

$$n^{2/(4+md)} T_8 \xrightarrow{\mathscr{L}} 0, \quad \text{as } n \to \infty. \tag{5.22}$$

The relations (5.20)–(5.22) show that the sum $\sum_{j=1}^{4} T_j$ in (5.17) yields the correct asymptotic bias, while the terms T_6, T_7, and T_8 are asymptotically negligible. It remains to show the asymptotic normality of the term T_5:

$$n^{2/(4+md)} T_5 \xrightarrow{\mathscr{L}} \mathcal{N}(0, V(x; v, s)), \quad \text{as } n \to \infty.$$

Again, to prove this, we use Lemma 5.5. We leave out the verification of the conditions (5.10) and (5.11) of Lemma 5.5, since it is done as in the proof of Theorem 1. We only deduce the asymptotic expression for the variance of T_5, which is given, analogous to G_n of the proof of Theorem 1, by the asymptotics of

$$G_n'' = \frac{1}{\mu(x)^2 n^2} \sum_{i=m}^{n} E[(K_h(X_i - x) v^{\mathrm{T}} \Sigma^{1/2}(X_i)(\xi_i \xi_i^{\mathrm{T}} - I_d) \Sigma^{1/2}(X_i) s)^2]. \tag{5.23}$$

To study this expression, use the following lemma.

Lemma 5.6. . *Let* $a = (a_1, \dots, a_d)^{\mathrm{T}}$, $\tilde{a} = (\tilde{a}_1, \dots, \tilde{a}_d)^{\mathrm{T}}$, $b = (b_1, \dots, b_d)^{\mathrm{T}}$, *and* $\tilde{b} = (\tilde{b}_1, \dots, \tilde{b}_d)^{\mathrm{T}}$ *be vectors in* $\mathbb{R}^d$. *Then under* (A1),

$$\begin{aligned} &E[(a^{\mathrm{T}}(\xi_1 \xi_1^{\mathrm{T}} - I_d) b)(\tilde{a}^{\mathrm{T}}(\xi_1 \xi_1^{\mathrm{T}} - I_d)\tilde{b})] \\ &\quad = (m_4 - 2) \operatorname{Tr}[\operatorname{diag}(a) \operatorname{diag}(b) \operatorname{diag}(\tilde{a}) \operatorname{diag}(\tilde{b})] + (a^{\mathrm{T}}\tilde{a})(b^{\mathrm{T}}\tilde{b}) + (a^{\mathrm{T}}\tilde{b})(\tilde{a}^{\mathrm{T}} b). \end{aligned}$$

Proof. Denoting by δ_{jk} the Kronecker delta and using (A1), we get

$$\begin{aligned} &E[(a^{\mathrm{T}}(\xi_1 \xi_1^{\mathrm{T}} - I_d) b)(\tilde{a}^{\mathrm{T}}(\xi_1 \xi_1^{\mathrm{T}} - I_d)\tilde{b})] \\ &\quad = E\left[\sum_{k,j=1}^{d} a_j (\xi_{1j}\xi_{1k} - \delta_{jk}) b_k \sum_{l,m=1}^{d} \tilde{a}_l (\xi_{1l}\xi_{1m} - \delta_{lm}) \tilde{b}_m \right] \\ &\quad = E\left[\sum_{j=1}^{d} a_j (\xi_{1j}^2 - 1) b_j \sum_{k=1}^{d} \tilde{a}_k (\xi_{1k}^2 - 1) \tilde{b}_k \right] \\ &\qquad + E\left[\sum_{1 \leqslant j < k \leqslant d} (a_j b_k + a_k b_j) \xi_{1j}\xi_{1k} \sum_{1 \leqslant l < m \leqslant d} (\tilde{a}_l \tilde{b}_m + \tilde{a}_m \tilde{b}_l) \xi_{1l}\xi_{1m} \right] \\ &\quad = m_4 \sum_{j=1}^{\mathrm{d}} a_j b_j \tilde{a}_j \tilde{b}_j + \sum_{1 \leqslant j < k \leqslant d} (a_j b_k + a_k b_j)(\tilde{a}_j \tilde{b}_k + \tilde{a}_k \tilde{b}_j) \\ &\quad = (m_4 - 2) \sum_{j=1}^{d} a_j b_j \tilde{a}_j \tilde{b}_j + \frac{1}{2} \sum_{j,k=1}^{d} (a_j b_k + a_k b_j)(\tilde{a}_j \tilde{b}_k + \tilde{a}_k \tilde{b}_j), \end{aligned}$$

which yields the lemma.

Applying Lemma 5.6 with $a = \Sigma^{1/2}(X_i)v$ and $b = \Sigma^{1/2}(X_i)s$, we find

$$E[(v^{\mathrm{T}}\Sigma^{1/2}(X_i)(\xi_i\xi_i^{\mathrm{T}} - I_d)\Sigma^{1/2}(X_i)s)^2 \mid X_i]$$

$$= (m_4 - 2)T^*(X_i) + \{v^{\mathrm{T}}\Sigma(X_i)s\}^2 + \{v^{\mathrm{T}}\Sigma(X_i)v\}\{s^{\mathrm{T}}\Sigma(X_i)s\}.$$

This and (5.23) yield

$$G_n'' = \frac{1}{\mu(x)^2 n^2} \sum_{i=m}^{n} E[K_h^2(X_i - x)(m_4 - 2)T^*(X_i)]$$

$$+ \frac{1}{\mu(x)^2 n^2} \sum_{i=m}^{n} E[K_h^2(X_i - x)(\{v^{\mathrm{T}}\Sigma(X_i)s\}^2 + \{v^{\mathrm{T}}\Sigma(X_i)v\}\{s^{\mathrm{T}}\Sigma(X_i)s\})]$$

and, in view of Lemma 5.2,

$$n^{4/(4+md)}G_n''$$

$$= \beta^{-md}\frac{\|K\|_2^2}{\mu(x)}[(m_4 - 2)T^*(x) + \{v^{\mathrm{T}}\Sigma(x)s\}^2 + \{v^{\mathrm{T}}\Sigma(x)v\}\{s^{\mathrm{T}}\Sigma(x)s\}]$$

$$\times (1 + o(1))$$

$$= V(x; v, s) + o(1), \quad \text{as } n \to \infty,$$

which is the expression for asymptotic variance given in Theorem 2. □

To show the joint asymptotic normality (3.6) and (3.7) one proceeds as in the proof of Theorem 1, by using the Cramér–Wold device and checking the conditions of Lemma 5.5. The calculations of covariance terms in (3.6) are based on Lemma 5.6 as well.

Acknowledgements

We would like to thank Christian Hafner, Helmut Lütkepohl, and Rolf Tschernig for helpful discussions. We also thank our referees for pointing out several technical errors to us. This research was supported by Sonderforschungsbereich 373 'Quantifikation und Simulation Ökonomischer Prozesse' Deutsche Forschungsgemeinschaft.

References

Ango Nze, P., 1992. Critères d'ergodicité de quelques modèles à représentation markovienne. C.R. Acad. Sci. Paris Sér. I 315, 1301–1304.

Bollerslev, T., Chou, R., Kroner, K., 1992. ARCH modeling in finance: a review of the theory and empirical evidence. J. Econometrics 52, 5–59.

Bollerslev, T., Engle, R., Wooldridge, J., 1988. A capital asset pricing model with time-varying covariances. J. Political Economy 96, 116–131.

Bossaerts, P., Härdle, W., Hafner, C., 1996. Foreign exchange-rates have surprising volatility. In: Robinson, P.M. (Ed.), Athens Conf. on Applied Probability and Time Series, vol. 2, Lecture Notes in Statistics, vol. 115, pp. 55–72. Springer, Berlin.

Chan, K.S., Tong, H., 1985. On the use of deterministic Lyapunov functions for the ergodicity of stochastic difference equations. Adv. Appl. Probab. 17, 666–678.

Chan, K.S., Tong, H., 1986. On estimating thresholds in autoregressive models. J. Time Ser. Anal. 7, 179–190.

Chen, R., Tsay, R.S., 1993a. Nonlinear additive ARX models. J. Amer. Statist. Assoc. 88, 955–967.

Chen, R., Tsay, R.S., 1993b. Functional-coefficient autoregressive models. J. Amer. Statist. Assoc. 88, 298–308.

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Amer. Statist. Assoc. 74, 829–836.

Collomb, 1984. Propriétés de convergence presque complète du prédicteur à noyau. Z. Wahrscheinlichkeitstheorie verwandte Gebiete 66, 441–460.

Davydov, Yu.A., 1973. Mixing conditions for Markov chains. Theory Probab. Appl. 18, 312–328.

Diebolt, J., Guégan, D., 1993. Tail behaviour of the stationary density of general nonlinear autoregressive processes of order one. J. Appl. Probab. 30, 315–329.

Diebold, F., Nason, J., 1990. Nonparametric exchange rate prediction. J. Internat. Econom. 28, 315–332.

Doukhan, P., Ghindès, M., 1980. Estimation dans le processus $X_{n+1} = f(X_n) + \varepsilon_{n+1}$. C.R. Acad. Sci. Paris Sér. A 297, 61–64.

Doukhan, P., Ghindès, M., 1981. Processus autorégressifs non-linéaires. C.R. Acad. Sci. Paris Sér. A 290, 921–923.

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 UK inflation. Econometrica 50, 987–1008.

Engle, R.F., Gonzalez-Rivera, G., 1991. Semiparametric ARCH models. J. Bus. Econom. Statist. 9, 345–360.

Engle, R.F., Ng, V., 1993. Measuring and testing the impact of news on volatility. J. Finance 48, 1749–1778.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman & Hall, London.

Gouriéroux, Ch., Monfort, A., 1992. Qualitative threshold ARCH models. J. Econometrics 52, 159–199.

Granger, C., Teräsvirta, T., 1993. Modelling Nonlinear Dynamic Relationships. Oxford University Press, Oxford.

Gregory, A.W., 1989. A nonparametric test for autoregressive conditional heteroscedasticity: a Markov chain approach. J. Bus. Econom. Statist. 7, 107–115.

Guillaume, D.M., Dacorogna, M.M., Davé, R.R., Müller, U.A., Olsen, R.B., Pictet, O.V., 1994. From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange market, Olsen Associates Working Paper.

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

Härdle, W., Klinke, S., Turlach, B., 1995. XploRe – an interactive statistical computing environment. Springer, Heidelberg.

Härdle, W., Tsybakov, A.B., 1996. Local polynomial estimators of the volatility function in nonparametric autoregression. J. Econometrics, to appear.

Katkovnik, V.Ya., 1979. Linear and nonlinear methods of nonparametric regression analysis. Automatika 35–46.

Katkovnik, V.Ya., 1985. Nonparametric Identification and Data Smoothing. Nauka, Moscow (in Russian).

Liptser, R.Sh., Shirjaev, A.N., 1980. A functional central limit theorem for martingales. Theory Probab. Appl. 25, 667–688.

Lütkepohl, H., 1991. Introduction to Multiple Time Series Analysis. Springer, Heidelberg.

McKeague, I.W., Zhang, M.J., 1994. Identification of nonlinear time series from first order cumulative characteristics. Ann. Statist. 22, 495–514.

Meese, R.A., Rogoff, K., 1983. Empirical exchange rate models of the seventies, do they fit out of sample? J. Internat. Econom. 14, 3–24.

Meese, R.A., Rose, A., 1991. An empirical assessment of non-linearities in models of exchange rate determination. Rev. Econom. Studies 58, 601–619.

Mokkadem, A., 1987. Sur un modèle autorégressif nonlinéaire. Ergodicité et ergodicité géometrique. J. Time Ser. Anal. 8, 195–204.

Nummelin, E., Tuominen, P., 1982. Geometric ergodicity of Harris-recurrent Markov chains with application to renewal theory. Stochastic Process. Appl. 12, 187–202.

Robinson, P.M., 1983. Nonparametric Estimators for Time Series. J. Time Ser. Anal. 4, 185–207.

Robinson, P.M., 1984. Robust nonparametric autoregression. In: Franke, Härdle, Martin (Eds.), Robust and Nonlinear Time Series Analysis, Lecture Notes in Statistics, vol. 26. Springer, Heidelberg.

Ruppert, D., Wand, M.P., 1994. Multivariate locally weighted least squares regression. Ann. Stat. 22, 1346–1370.

Stone, C.J., 1977. Consistent nonparametric regression. Ann. Statist. 5, 595–645.

Teräsvirta, T., 1994. Specification, estimation and evaluation of smooth transition autoregressive models. J. Amer. Statist. Assoc. 89, 208–218.

Tjøstheim, D., 1994. Non-linear time series analysis: a selective review. Scand. J. Statist. 21, 97–130.

Tong, H., 1978. On a threshold model. In: Chen, C.H. (Ed.), Pattern Recognition and Signal Processing. Sijthoff & Noordholf, Alphen a/d Rijn, The Netherlands.

Tong, H., 1983. Threshold Models in Nonlinear Time Series Analysis, Lecture Notes in Statistics, vol. 21. Springer, Heidelberg.

Tsybakov, A.B., 1986. Robust reconstruction of functions by the local approximation method. Prob. Inform. Transm. 22, 133–146.

Tweedie, R.L., 1975. Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. Stochastic Process. Appl. 3, 385–403.

Vieu, P., 1994. Order choice in nonlinear autoregressive models. Statistics 24, 1–22.

Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman & Hall, London.