

# Nonparametric estimation of the conditional variance function with correlated errors

Juan M. Vilar-Fernández

Departamento de Matemáticas, Facultad de Informática

Universidad de A Coruña, A Coruña, 15071, Spain.

Tel: +34 981 167000 (ext. 1221)

Fax: +34 981 167160

e-mail: [ejjvilar@udc.es](mailto:ejjvilar@udc.es)

and

Mario Francisco-Fernández<sup>1</sup>

Departamento de Matemáticas, Facultad de Informática

Universidad de A Coruña, A Coruña, 15071, Spain.

Tel: +34 981 167000 (ext. 1222)

Fax: +34 981 167160

e-mail: [mariofr@udc.es](mailto:mariofr@udc.es)

---

<sup>1</sup>Corresponding Author

# Nonparametric estimation of the conditional variance function with correlated errors

J. M. VILAR-FERNÁNDEZ      M. FRANCISCO-FERNÁNDEZ<sup>†</sup>

Universidad de A Coruña

Universidad de A Coruña

## Abstract

In this paper, we consider a fixed regression model where the errors are a strictly stationary process and in which both functions, the conditional mean and the conditional variance (volatility), are unknown. Two nonparametric estimators of the volatility function based on local polynomial fitting are studied. Expressions of the asymptotic bias and variance are given and the asymptotic normality is shown for both estimators. The influence of the dependence of the data is observed in the expressions of the variance. A simulation study and an analysis with real economic data illustrate the behavior of the proposed nonparametric estimators.

Key Words: autoregressive process; heteroscedasticity; local polynomials, volatility

2000 AMS Subject Classification: 62G05, 62G20

---

<sup>†</sup>Corresponding author. Email: mariofr@udc.es

# 1 Introduction

Nonparametric methods are appropriate tools used to perform exploratory analyses, because they do not require selecting a specific parametric shape before fitting the data. One of these methods, in the context of the regression function estimation, which has good theoretical and practical properties, is the local polynomial regression (LPR) estimator. Some significant references on this method for independent observations are, for example, Tsybakov [19], Fan [4] and Ruppert and Wand [17].

Sometimes, it cannot be assumed that the observations in the sample data are independent, for example, if they are gathered sequentially in time. In this case, the statistical properties of the LPR estimator presented in the papers mentioned above can change. Some related works in this setting of dependence are: Masry [13] and Masry and Fan [14]. In these papers, a regression model, considering a random data sample,  $\{(X_t, Y_t)\}_{t=1}^n$ , satisfying some mixing conditions, was used. However, in Francisco-Fernández and Vilar-Fernández [6], a regular fixed design regression model with short-range correlated and homoscedastic errors was considered. In this case, while the asymptotic bias of the regression estimator is exactly the same as that obtained under independence, the asymptotic variance of the estimator changes. Now, the sum of the covariances of the errors in this term appears instead of simply the variance of the errors, as in the case of independent observations or under mixing dependence conditions. Opsomer *et al.* [16] provide an overview of the problem of nonparametric regression with correlated errors.

In this paper, we consider the same framework as that used in Francisco-Fernández and Vilar-Fernández [6], but now the regression model is heteroscedastic. In this case, the aim is not only to estimate nonparametrically the regression function but also the volatility function. This kind of regression models frequently arise in economic studies, in the analysis of growth curves and usually in the study of time series with deterministic trend and non-constant conditional variance (risk, in financial terminology). A by-product of our work is the generalization of the asymptotic results obtained in Francisco-Fernández and Vilar-Fernández [6] for the regression function to a heteroscedastic regression model. Considering this model, we study two estimators of the volatility function previously studied in Härdle and Tsybakov [11] and Fan and Yao [5], respectively, in different contexts. They consider a dynamic regression model of a mixing process and a two-dimensional strictly stationary and absolutely regular process, respectively, in their approaches. As it will be seen, the leading term of the asymptotic variance of these estimators in our model is different from the ones obtained in those papers.

Other articles concerning the problem of estimating the conditional variance are, for instance, Gasser *et al.* [9], Hall *et al.* [10], Masry and Tjøstheim [15], Ruppert *et al.* [18], Andersen and Lund [1] and Ziegelmann [20].

In what follows, it is assumed that univariate data  $Y_{1,n}, Y_{2,n}, \dots, Y_{n,n}$  are observed, and that

$$Y_{t,n} = m(x_{t,n}) + s(x_{t,n})\varepsilon_{t,n}, \quad 1 \leq t \leq n \quad (1)$$

where  $m(x)$  and  $s(x)$  are “smooth” functions defined on  $[0, 1]$ , with  $s(x) > 0$ . The

errors  $\varepsilon_{t,n}$ ,  $1 \leq t \leq n$ , are a sequence of unobserved random variables with  $E(\varepsilon_{t,n}) = 0$  and  $E(\varepsilon_{t,n}^2) = 1$ , where for each  $n$ ,  $\{\varepsilon_{1,n}, \varepsilon_{2,n}, \dots, \varepsilon_{n,n}\}$  have the same joint distribution as  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ , with  $\{\varepsilon_t, t \in Z\}$  being a strictly stationary stochastic process. Also, it is assumed that the design  $x_{t,n}$ ,  $1 \leq t \leq n$ , is a regular fixed design generated by a design density  $f$ , that is, for each  $n$ , the design points are defined by

$$\int_0^{x_{t,n}} f(x)d(x) = \frac{t-1}{n-1}, \quad 1 \leq t \leq n,$$

$f$  being a positive function defined on  $[0, 1]$  and its first derivative is continuous. In order to simplify notation, we will not use  $n$  in the subindexes, that is, we will write  $x_t$ ,  $\varepsilon_t$  and  $Y_t$ .

We study the problem of estimation of the volatility function  $v(x) = s^2(x)$ , given a sample  $\{(x_t, Y_t)\}_{t=1}^n$ .

The organization of the work is as follows: In Section 2, two estimators for the volatility function  $v(x) = s^2(x)$  are introduced. In Section 3 asymptotic properties of both estimators are provided. Expressions for the asymptotic bias and variance are obtained and asymptotic normality is shown in both cases. These expressions will give some guidelines on how to select reliable smoothing parameters in this model. In Section 4 the estimators studied are compared via a simulation study and are used to analyze a real economic data set.

## 2 The estimators

In this Section, two estimators of the volatility function are introduced. Both use local polynomial techniques, although different approaches are followed for their construction. Due to the simple decomposition  $v(x) = E(Y^2|x) - m^2(x)$  and following the idea of Härdle and Tsybakov [11], the first estimator of the volatility function is defined by

$$\hat{v}_n^S(x) = \hat{g}_n(x) - \{\hat{m}_n(x)\}^2, \quad (2)$$

where  $\hat{g}_n(x)$  is an estimator of  $g(x) = E(Y^2|x) = m^2(x) + s^2(x)$ , and  $\hat{m}_n(x)$  is an estimator of  $m(x)$ . We will use estimators  $\hat{m}_n(x)$  and  $\hat{g}_n(x)$  based on the LPR estimator. So, assuming that the  $(p+1)$ th derivatives of  $m(x)$  and  $g(x)$  exist and are continuous, local polynomial fitting permits estimating the parameter vectors  $\vec{\beta}(x) = (\beta_0(x), \beta_1(x), \dots, \beta_p(x))^t$ , where  $\beta_j(x) = m^{(j)}(x)/(j!)$ , and  $\vec{\gamma}(x) = (\gamma_0(x), \gamma_1(x), \dots, \gamma_p(x))^t$ , where  $\gamma_j(x) = g^{(j)}(x)/(j!)$ , with  $j = 0, 1, \dots, p$  by minimizing, respectively, the functions

$$\Psi(\vec{\beta}(x)) = \sum_{t=1}^n \left( Y_t - \sum_{j=0}^p \beta_j(x)(x_t - x)^j \right)^2 \omega_{n,t}$$

and

$$\Psi(\vec{\gamma}(x)) = \sum_{t=1}^n \left( Y_t^2 - \sum_{j=0}^p \gamma_j(x)(x_t - x)^j \right)^2 \omega_{n,t},$$

where  $\omega_{n,t} = n^{-1}h_n^{-1}K(h_n^{-1}(x_t - x))$  are the weights,  $K$  being a kernel function and  $h_n$  the bandwidth or smoothing parameter. Standard weighted least squares theory leads to the solutions

$$\hat{\beta}_{(n)}(x) = (X_{p,(n)}^t W_{(n)} X_{p,(n)})^{-1} X_{p,(n)}^t W_{(n)} \vec{Y}_{(n)} = S_{(n)}^{-1} \vec{T}_{(n)} \quad (3)$$

and

$$\hat{\gamma}_{(n)}(x) = (X_{p,(n)}^t W_{(n)} X_{p,(n)})^{-1} X_{p,(n)}^t W_{(n)} \vec{Y}_{(n)}^2 = S_{(n)}^{-1} \vec{Z}_{(n)}, \quad (4)$$

where  $\vec{Y}_{(n)} = (Y_1, \dots, Y_n)^t$ ,  $\vec{Y}_{(n)}^2 = (Y_1^2, \dots, Y_n^2)^t$ ,  $X_{p,(n)}$  is the  $n \times (p+1)$  matrix with  $i$ th row equal to  $(1, (x_1 - x), \dots, (x_1 - x)^p)$ , and  $W_{(n)} = \text{diag}(\omega_{n,1}, \dots, \omega_{n,n})$  is the diagonal array of weights.

The estimator  $\hat{v}_n^S(x)$  of  $v(x)$  is defined as

$$\hat{v}_n^S(x) = \hat{\gamma}_{(n)}(x)^t e_1 - \left\{ \hat{\beta}_{(n)}(x)^t e_1 \right\}^2, \quad (5)$$

where  $e_1$  is the  $(p+1) \times 1$  vector having 1 in the first entry and all other entries being 0.

In those  $x$ 's where  $v(x)$  is close to 0, the estimator defined in (5) could be negative. In those cases,  $\hat{v}_n^S(x)$  is defined as 0.

On the other hand, Fan and Yao [5] suggest a novel approach, asymptotically fully adaptive to the unknown conditional mean. This consists in, first, obtaining the residuals from a nonparametric fit (using, for instance, the LPR estimator with a kernel  $L_1$  and a bandwidth  $h_{1n}$ ), squaring them,  $\hat{r}_t = \{Y_t - \hat{m}_{h_{1n}}(x_t)\}^2$ ,  $t = 1, 2, \dots, n$ , and finally defining the estimator of the volatility function as the LPR estimator of the regression function with kernel  $L_2$  and bandwidth  $h_{2n}$ , using  $\{\hat{r}_t\}_{t=1}^n$  as the response variables. Obviously the degree of the polynomial used in each step can be also different, say  $p_1$  and  $p_2$ . Mathematically, this can be written as follows: using (3), the LPR estimator of  $m(x)$  is given by

$$\hat{m}_{h_{1n}}(x) = e_1^t (X_{p_1,(n)}^t W_{1(n)} X_{p_1,(n)})^{-1} X_{p_1,(n)}^t W_{1(n)} \vec{Y}_{(n)} = S_{h_{1n},x}^t \vec{Y}_{(n)},$$

where  $W_{1(n)}$  is the diagonal array of weights  $n^{-1}h_{1n}^{-1}L_1(h_{1n}^{-1}(x_t-x))$ . Then, the second estimator of the volatility function considered in this paper (denoted by  $\hat{v}_n^D(x)$ ) is defined by

$$\hat{v}_n^D(x) = e_1^t (X_{p_2,(n)}^t W_{2(n)} X_{p_2,(n)})^{-1} X_{p_2,(n)}^t W_{2(n)} \hat{R}_{(n)} = S_{h_{2n},x}^t \hat{R}_{(n)}, \quad (6)$$

where  $W_{2(n)}$  is defined as  $W_{1(n)}$ , but using the kernel  $L_2$  and the bandwidth  $h_{2n}$ , and  $\hat{R}_{(n)} = (\hat{r}_1, \dots, \hat{r}_n)$ .

### 3 Theoretical results

In this section, the asymptotic normality of estimators (5) and (6) are obtained.

The following assumptions will be needed in our analysis:

**A1** The kernel functions,  $K(\cdot)$ ,  $L_1(\cdot)$  and  $L_2(\cdot)$  are symmetric, with a bounded support, and Lipschitz continuous.

**A2** The sequence of bandwidths,  $\{h_n^*\}$ , satisfies  $h_n^* > 0$ ,  $h_n^* \downarrow 0$ ,  $nh_n^* \uparrow \infty$ , where the sequence  $\{h_n^*\}$  can be  $\{h_n\}$ ,  $\{h_{1n}\}$  or  $\{h_{2n}\}$ .

**A3** The point  $x$  at which the estimation is taking place satisfies  $h_n < x < 1 - h_n$  when  $\hat{v}_n^S(x)$  is used, and satisfies  $\max\{h_{1n}, h_{2n}\} < x < \min\{1 - h_{1n}, 1 - h_{2n}\}$  when  $\hat{v}_n^D(x)$  is used, for all  $n \geq n_0$  where  $n_0$  is fixed in both cases.

**A4** The errors,  $\{\varepsilon_i\}_{i=1}^n$ , satisfy  $E(\varepsilon_i^2) = 1$ ,  $E(\varepsilon_i) = E(\varepsilon_i^3) = 0$ . Moreover, denoting  $c(k) = \text{Cov}(\varepsilon_i, \varepsilon_{i+k})$ ,  $k = 0, \pm 1, \dots$ , then  $\sum_{k=1}^{\infty} k |c(k)| < \infty$ , and  $d(\varepsilon) = \sum_{k=-\infty}^{\infty} \text{Cov}(\varepsilon_i^2, \varepsilon_{i+k}^2) < \infty$ .



**A5**  $E|\varepsilon_t|^{2(2+\delta)} < \infty$  for some  $\delta > 0$ .

**A6** The stationary stochastic process  $\{\varepsilon_t\}$  is  $\alpha$ -mixing with mixing coefficients

such that  $\sum_{t=1}^{\infty} \alpha(t)^{\delta/(2+\delta)} < \infty$ . Then, there exists a sequence of positive integers  $\{s_n\}$ ,  $s_n \rightarrow \infty$  as  $n \rightarrow \infty$  with  $s_n = o\left((nh_n^{*3})^{1/2}\right)$ , such that  $(nh_n^{*-1})^{1/2} \sum_{t=s_n}^{\infty} \alpha(t)^{1-\gamma} < \infty$ , with  $\gamma = 2/(2+\delta)$ , where  $h_n^*$  can be  $h_n$  or  $h_{2n}$ .

**A7**  $h_n^* = O\left(n^{-1/(2p^*+3)}\right)$ , where  $h_n^*$  and  $p^*$  can be  $h_n$  and  $p$ , or  $h_{2n}$  and  $p_2$ .

**A8**  $\left\{h_{1n}^{2(p_1+1)} + (nh_{1n})^{-1}\right\} = o\left(h_{2n}^{p_2+1}\right)$ .

**Remark 3.1** *The  $\alpha$ -mixing or strong mixing condition is one of the least restrictive among the numerous dependence conditions and it is satisfied by many processes, for instance, the ARMA processes generated by absolutely continuous noise. A thorough study of this condition can be seen in Doukhan [3].*

**Remark 3.2** *Assumption A8 holds, for instance, when the bandwidths  $h_{1n}$  and  $h_{2n}$  are selected at their optimal rate for estimating the regression function, that is,  $O\left(n^{-1/(2p_j+3)}\right)$ ,  $j = 1, 2$ .*

The following notation will be used. Let  $K_p(u) = (j! |M_p(u)|/|S|) K(u)$ , where  $S$  is the  $(p+1) \times (p+1)$  array whose  $(i+1, j+1)$ th element is  $\mu_{i+j}(K)$ ,  $0 \leq i, j \leq p$ , with  $\mu_l(K) = \int u^l K(u) du$ , and  $M_p(u)$  is the same as  $S$  with the first column replaced by  $(1, u, \dots, u^p)$ .  $K_p$  is  $p$ th-order kernel as defined by Gasser et al [9]. The same kind of functions as  $K_p$ , but using kernels  $L_1$  and  $L_2$  will be also used. These will

be denoted by  $L_{1,p_1}$  and  $L_{2,p_2}$ , respectively. We also define the matrix  $\tilde{S}$  as the  $(p+1) \times (p+1)$  array whose  $(i+1, j+1)$ th element is  $\int u^{i+j} K^2(u) du$ ,  $0 \leq i, j \leq p$ . By assumption A1,  $S$  and  $\tilde{S}$  are positive definite, see Lemma 1 of Tsybakov [19], and therefore nonsingular. We denote  $R(L) = \int L(u)^2 du$ . Also, throughout the paper, notation  $A \approx B$  denotes  $A = B(1 + o(1))$ .

**Theorem 3.1** *If assumptions A1-A7 are fulfilled, we have*

$$\sqrt{nh_n} \left[ \begin{pmatrix} H_{(n)} (\hat{\gamma}_{(n)}(x) - \vec{\gamma}(x)) \\ H_{(n)} (\hat{\beta}_{(n)}(x) - \vec{\beta}(x)) \end{pmatrix} - b(x) \right] \xrightarrow{\mathcal{L}} N_{2(p+1)} \left( \vec{0}, \Sigma(x) \right), \text{ as } n \rightarrow \infty, \quad (7)$$

where  $H_{(n)} = \text{diag}(1, h_n, h_n^2, \dots, h_n^p)$ ,

$$b(x) = \left( \frac{g^{(p+1)}(x)}{(p+1)!} h_n^{p+1} S^{-1} \vec{\mu}, \frac{m^{(p+1)}(x)}{(p+1)!} h_n^{p+1} S^{-1} \vec{\mu} \right)^t,$$

with  $\vec{\mu} = (\mu_{p+1}(K), \dots, \mu_{2p+1}(K))^t$ , and

$$\Sigma(x) = \frac{v(x)}{f(x)} \begin{pmatrix} d(\varepsilon)v(x) + 4m^2(x)c(\varepsilon) & 2m(x)c(\varepsilon) \\ 2m(x)c(\varepsilon) & c(\varepsilon) \end{pmatrix} \otimes (S^{-1} \tilde{S} S^{-1}),$$

where  $c(\varepsilon) = \sum_{k=-\infty}^{\infty} c(k) < \infty$ . Here  $A \otimes B$  denotes the Kronecker product of matrices  $A$  and  $B$ .

**Theorem 3.2** *Assume A1-A7. Then*

$$\sqrt{nh_n} (\hat{v}_n^S(x) - v(x) - b_S(x)) \xrightarrow{\mathcal{L}} N(0, \sigma_S^2(x)),$$

as  $n \rightarrow \infty$ , where

$$b_S(x) = \frac{h_n^{p+1}}{(p+1)!} \left( v^{(p+1)}(x) + (m^2(x))^{(p+1)} - 2m(x)m^{(p+1)}(x) \right) \mu_{p+1}(K_p) \quad (8)$$

and

$$\sigma_S^2(x) = \frac{v^2(x)}{f(x)} d(\varepsilon) R(K_p). \quad (9)$$

With respect to estimator  $\hat{v}_n^D(x)$ , in the following Theorem its asymptotic normality is established. For this, the following additional assumption is used only to simplify the calculations involved in the proof of Theorem 3.3.:

**A9** The errors  $\{\varepsilon_t\}$  follow an MA( $\infty$ ) process,  $\varepsilon_t = \sum_{-\infty}^{\infty} \Psi_j e_{t-j}$ , with kurtosis of the white noise  $\{e_j\}$  being equal to 0.

**Theorem 3.3** *If assumptions A1-A9 are fulfilled, we have*

$$\sqrt{nh_{2n}} (\hat{v}_n^D(x) - v(x) - b_D(x)) \xrightarrow{\mathcal{L}} N(0, \sigma_D^2(x)),$$

as  $n \rightarrow \infty$ , where

$$b_D(x) = \frac{h_{2n}^{p_2+1}}{(p_2 + 1)!} v^{(p_2+1)}(x) \mu_{p_2+1}(L_{2,p_2}) \quad (10)$$

and

$$\sigma_D^2(x) = \frac{v^2(x)}{f(x)} d(\varepsilon) R(L_{2,p_2}). \quad (11)$$

The proofs of these Theorems are given in the final Appendix.

**Remark 3.3** *In both estimators, the dependence of the observations influences the leading term of the variance, but not the the leading term of the bias. Moreover, comparing the results obtained in Theorems 3.1 and 3.2 with those obtained by Härdle and Tsybakov [11] for  $\hat{v}_n^S(x)$  with random and  $\alpha$ -mixing observations (Theorems 3.1 and 3.2 of that paper) and Yao and Fan [5] for  $\hat{v}_n^D(x)$  with random and absolutely*

regular observations (Theorem 1 of that paper), respectively, it can be observed that in both cases, while the asymptotic bias is exactly the same in the random and the fixed design, the asymptotic variance of both estimators changes, now the term  $d(\varepsilon) = \sum_{k=-\infty}^{\infty} \text{Cov}(\varepsilon_i^2, \varepsilon_{i+k}^2)$  appearing instead of simply  $\text{Var}(\varepsilon^2)$  as occurs in the random case with dependent or independent observations (see Ruppert et al. [18], where a version of  $\hat{v}_n^D(x)$  is studied in the independent case). The reason for this fact is that in random design (under mixing dependence conditions) if we want to calculate the estimator of the conditional variance function at a point  $x$ , the random variables  $(X_i, Y_i)$  and  $(X_j, Y_j)$ , such that  $X_i, X_j \in (x - h_n, x + h_n)$ , are nearly uncorrelated in the majority of the cases, as  $h_n \rightarrow 0$ . On the other hand, in a fixed design with correlated errors, the spatial distance between the  $x$ 's coincides with distance in time and, therefore, the corresponding  $Y$ 's variables at the  $x$ 's locations in a neighborhood of  $x$  (or the corresponding errors) are strongly dependent between them.

**Remark 3.4** A comparison between  $\hat{v}_n^S(x)$  and  $\hat{v}_n^D(x)$  is possible by just looking at Theorems 3.2 and 3.3. If the same bandwidth and kernel are used in both estimators,  $\hat{v}_n^S(x)$  has the same asymptotic variance as  $\hat{v}_n^D(x)$ , but includes one more term in the bias,  $\Delta = (m^2(x))^{(p+1)} - 2m(x)m^{(p+1)}(x)$ . This term could have an adverse effect on the bias of estimator  $\hat{v}_n^S(x)$ , for instance, when  $p = 1$ ,  $\Delta = 2(m'(x))^2 > 0$ . This fact will be seen in the simulations.

**Remark 3.5** Defining  $\text{AMSE}(\hat{v}_n^D(x))$  as the mean squared error approximation obtained by combining (10) and (11), the asymptotically optimal local bandwidth can

be computed by minimizing  $AMSE(\hat{v}_n^D(x))$  with respect to  $h_{2n}$ . This minimizer is

$$h_{2n}^{opt}(x) = C_{0,p_2}(L_2) \left( \frac{v^2(x)d(\varepsilon)}{n(v^{(p_2+1)}(x))^2 f(x)} \right)^{1/(2p_2+3)}, \quad (12)$$

where  $C_{0,p_2}(L_2)$  is a real number that depends on kernel  $L_2$ .

This expression give us some guidelines on how to select a local plug-in bandwidth, just by replacing the unknown quantities appearing in the optimal bandwidth with pilot estimators. Analogous comments are also possible for the case of a plug-in local bandwidth for  $\hat{v}_n^S(x)$ , and for the selection of global plug-in bandwidths by minimizing the AMISE of both estimators. The problem of bandwidth selection in the context of the volatility function is beyond the scope of the present paper. But this is certainly worthy of effort for further research.

**Remark 3.6** The value of  $d(\varepsilon) = \sum_{k=-\infty}^{\infty} Cov(\varepsilon_i^2, \varepsilon_{i+k}^2)$  can be calculated assuming some hypothesis for the errors.

- If the errors follow an AR(1) process,  $\varepsilon_t = \rho\varepsilon_{t-1} + e_t$ , and denoting by  $\Gamma$  the kurtosis of  $\varepsilon_t$ , then

$$d(\varepsilon) = (\Gamma - 1) \frac{1 + \rho^2}{1 - \rho^2}. \quad (13)$$

- If the errors follow a normal distribution,  $d(\varepsilon) = 2(1 + 2\sum_{k=1}^{\infty} c(k)^2)$ .
- If the errors follow an AR(1) process with normal distribution,  $d(\varepsilon) = 2\frac{1+\rho^2}{1-\rho^2}$ .

**Remark 3.7** From Theorem 3.1, it is easy to obtain the asymptotic variance of the LPR estimator of the regression function,  $\hat{m}_n(x)$ , in this setting of correlated errors.

This is given by

$$\text{Var}(\hat{m}_n(x)) \approx \frac{1}{nh_n} \frac{v(x)}{f(x)} c(\varepsilon) R(K_p), \quad (14)$$

where

$$c(\varepsilon) = \frac{1 + \rho}{1 - \rho} \quad (15)$$

if the errors are  $AR(1)$ . A simple comparison between (13) and (15) clearly shows that although (14) and (9), or (11) have the same kind of structure, their behavior, as a function of  $\rho$ , for  $AR(1)$ -errors are different. While  $c(\varepsilon)$  goes to infinity when  $\rho$  goes to 1, and goes to 0 when  $\rho$  goes to  $-1$ ,  $d(\varepsilon)$  goes to infinity when  $\rho$  goes to 1 or  $-1$ , that is, the variance of  $\hat{v}_n^S(x)$  and  $\hat{v}_n^D(x)$  increases when  $\rho$  goes to 1 or  $-1$ , while the asymptotic variance of the LPR estimator of the regression function increases when the dependence increases, it being smaller for negative dependence than for independent observations.

## 4 Simulation study and example

In this section, we present some simulation results comparing the estimators of the volatility function theoretically studied in the previous section. We also apply these estimators to a real data set.

### 4.1 Simulation study

In this subsection, the performance of estimators  $\hat{v}_n^S(x)$  and  $\hat{v}_n^D(x)$  is illustrated through a simulation study. In the theoretical portion of this article, we showed

that  $\hat{v}_n^D(x)$  is asymptotically fully adaptive in the sense that the variance function can be estimated with this method with the same asymptotic mean and variance as if the mean function were known. To observe the real influence of this fact with finite samples, we have also considered in the study the same estimator as  $\hat{v}_n^D(x)$ , but using, in the second step, the squared errors instead of the squared residuals. This estimator is denoted by  $\hat{v}_n^b(x)$  (the superscript standing for “benchmark”, as in Fan and Yao [5]).

We simulated 300 samples of size 100 from a fixed and equally spaced model like (1) in the interval  $[0, 1]$  with errors following an AR(1) process,  $\varepsilon_t = \rho\varepsilon_{t-1} + e_t$ , with  $N(0, 1)$  distribution. We have used this model of dependence because it is easy to simulate, and by varying only one parameter, the autocorrelation coefficient of order 1,  $\rho$ , the dependence of the observations is controlled. Taking this into account, we considered different values of  $\rho$  ( $-0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9$ ) to study the influence of the dependence of the observations. Optimal bandwidths by minimizing the Mean Integrated Squared Error (MISE) were computed. Using Monte Carlo approximations, the integrated squared bias, the integrated variance and the MISE of  $\hat{v}_n^S(x)$ ,  $\hat{v}_n^D(x)$  and  $\hat{v}_n^b(x)$  were then approximated. For a more comprehensive study, we considered three different problems: to estimate the volatility function on interval  $[0, 1]$  (global region), to estimate the function on the central region  $[0.2, 0.8]$  and to estimate the function on the boundary region  $[0, 0.2] \cup [0.8, 1]$ . In each case, the optimal bandwidths obtained are different. The kernel function used was the quartic kernel ( $K(u) = L_1(u) = L_2(u) = \frac{15}{16}(1 - u^2)^2$  if  $|u| \leq 1$ ).

We considered the following regression models:

**Model (1)**  $m(x) = \sin(\pi x)$  and  $s(x) = \frac{1}{2}x$ .

**Model (2)**  $m(x) = \sin(\pi x)$  and  $s(x) = \sin(\pi x)$ .

**Model (3)**  $m(x) = 5x$  and  $s(x) = \sin(\pi x)$ .

As was pointed out in the definition of  $\hat{v}_n^S(x)$ , this estimator can become negative in cases. To check how often negative values are obtained in these regression models, we have computed the percentage of negative values of  $\hat{v}_n^S(x)$  in each one of the 300 samples using a range of reasonable bandwidths. The averages of these percentages over the 300 samples, depending on the regression model and the bandwidth, ranged from 1% to 15%, approximately. Since these numbers are not completely negligible, the adjusted estimator of  $\hat{v}_n^S(x)$ , taking the value 0 when a negative value is obtained, was used in our simulations. It is important to note that the percentage of negative values depends on the shape of the volatility function. In any case, the results obtained using the original estimator and the adjusted one are quite similar, being slightly better for the adjusted estimator of  $\hat{v}_n^S(x)$ .

For brevity, we present here only some of the results obtained that are representative of the simulations performed.

Table 1 shows the results for Model (1) when the autocorrelation coefficient of the AR(1) error is  $\rho = 0.3$  and  $\rho = 0.9$ . The optimal bandwidths used to obtain the residuals needed to compute  $\hat{v}_n^D(x)$  were 0.2636 and 0.3454 for  $\rho = 0.3$  and  $\rho = 0.9$ , respectively. So, the optimal bandwidths appearing in Table 1 for  $\hat{v}_n^D(x)$  and  $\hat{v}_n^b(x)$



refer to those used in the second step of these estimators to fit the squared residuals and the squared errors, respectively.

[Table 1 about here.]

In Table 2, the results for Model (2) when the autocorrelation coefficient of the AR(1) error is  $\rho = -0.6$  and  $\rho = 0.6$  are presented. Here, the optimal bandwidths needed to obtain the residuals to compute  $\hat{v}_n^D(x)$  were 0.2525 and 0.4545 for  $\rho = -0.6$  and  $\rho = 0.6$ , respectively.

[Table 2 about here.]

Finally, concerning Model (3), Figure 1 shows the evolution of the MISE of  $\hat{v}_n^S(x)$ ,  $\hat{v}_n^D(x)$  and  $\hat{v}_n^b(x)$  as a function of  $\rho$ , in the central part of the interval  $([0.2, 0.8])$  and in the whole interval  $([0, 1])$ .

[Figure 1 about here.]

Overall, both estimators ( $\hat{v}_n^S(x)$  and  $\hat{v}_n^D(x)$ ) had similar, good performance, although, in general,  $\hat{v}_n^D(x)$  gave better results, especially in cases where the additional term appearing in the leading term of the bias of  $\hat{v}_n^S(x)$  could have an important effect on the estimation. For instance, this happens in Model (3) as seen in Figure 1 (see Remark 3.4 to understand the effect of the regression function on the bias of  $\hat{v}_n^S(x)$ ). On the other hand, significant differences between  $\hat{v}_n^D(x)$  and  $\hat{v}_n^b(x)$  do not seem to exist and the different results obtained for these estimators are possibly a random sample effect.

Another interesting point observed in the simulations is that when  $|\rho|$  increases, the MISE of the estimators also increases. This is due to the behavior of the variance and it is compatible with the asymptotic expressions obtained in Section 3 (see Remark 3.6).

Finally, we have observed that  $\hat{v}_n^D(x)$  worked better than  $\hat{v}_n^S(x)$  at boundary values. This could be due to the way  $\hat{v}_n^D(x)$  is constructed: the known automatic boundary correction proven for the local linear estimator of the regression function is now carried over to the estimation of the conditional variance function when  $\hat{v}_n^D(x)$  is used.

## 4.2 Example

In this subsection, both estimators of the volatility function,  $\hat{v}_n^S(x)$  and  $\hat{v}_n^D(x)$ , are applied to the study of a real economic data set.

The sample data considered are 222 quarterly observations of the real change in private inventories in the USA, from 1947 to 2002. Each observation indicates the seasonally adjusted annual rate, measured in billions of chained 1996 dollars (in the year 2002, only two observations are available). The source of these data is the U.S. Department of Commerce, Bureau of Economic Analysis, obtained from the web page: <http://www.research.stlouisfed.org/fred/data/gdp.html>.

A fixed regression model can be fitted to these data, considering an equally

spaced design in  $[0, 1]$ , that is,

$$Y_t = m(t/222) + s(t/222)\varepsilon_t, \quad t = 1, 2, \dots, 222.$$

The aim is to estimate the volatility function  $v(x) = s^2(x)$  using the nonparametric estimators  $\hat{v}_n^S(x)$  and  $\hat{v}_n^D(x)$ . Only the linear case, that is,  $p = p_1 = p_2 = 1$ , will be considered for both estimators. To compute  $\hat{v}_n^D(x)$ , first, it is necessary to obtain the residuals from a nonparametric fit. At this point, the bandwidth needed to estimate the regression function was computed by the time series cross-validation (TSCV) method proposed by Hart [12]. This parameter showed good performance in the simulation study presented in Francisco-Fernández and Vilar-Fernández [7] in a fixed regression model with autoregressive errors. In Figure 2 the sample data and the local linear estimator of  $m(x)$  using that bandwidth,  $\hat{h}_{TSCV1} = 0.2229$ , are presented. The TSCV method was again used, producing the bandwidth  $\hat{h}_{TSCV2} = 0.1618$ , in the second step of this estimator. Figure 3 shows both estimators of the volatility function,  $\hat{v}_n^S(x)$  being the dashed line and  $\hat{v}_n^D(x)$  the solid line. For simplicity, the bandwidth needed to compute  $\hat{v}_n^S(x)$  was selected by ordinary cross-validation, giving as result  $\hat{h}_{CV} = 0.1172$ . Both estimators have a similar shape, although there is a boundary effect when  $\hat{v}_n^S(x)$  is used, especially at values near one, where the estimated volatility function decreases. This effect seems not to be present for  $\hat{v}_n^D(x)$ , as explained in the previous Section.

[Figure 2 about here.]

[Figure 3 about here.]

## 5 Acknowledgments

This work was partially supported by MEC Grant MTM2005-00429 (ERDF included) and Grant PGIDIT03PXIC10505PN. The authors wish to thank a referee for constructive comments that improved the presentation of the article.

## A Appendix: Proofs

In this Section, we sketch proofs of the results presented in Section 3.

To complete the proof of Theorem 3.1, the following Propositions are needed.

**Proposition A.1** *Under assumptions A1-A3, we have*

$$\lim_{n \rightarrow \infty} H_{(n)}^{-1} S_{(n)} H_{(n)}^{-1} = f(x)S,$$

where  $S_{(n)}$  is the array  $(p+1) \times (p+1)$  whose  $(i, j)$ th element is  $s_{i,j}^{(n)} = s_{i+j-2}^{(n)}$ , with

$$s_j^{(n)} = \frac{1}{nh_n} \sum_{t=1}^n (x_t - x)^j K\left(\frac{x_t - x}{h_n}\right).$$

**Proposition A.2** *Under assumptions A1-A4, we have*

$$\lim_{n \rightarrow \infty} nh_n E\left(H_{(n)}^{-1} \vec{T}_{(n)}^* \vec{T}_{(n)}^{*t} H_{(n)}^{-1}\right) = f(x) \tilde{S}v(x)c(\varepsilon),$$

where  $\vec{T}_{(n)}^* = X_{p,(n)}^t W_{(n)} \left(\vec{Y}_{(n)} - \vec{M}_{(n)}\right)$ , with  $\vec{M}_{(n)} = (m(x_1), \dots, m(x_n))^t$ .

**Proof.** The proofs of Propositions A.1 and A.2 follow the same lines as Propositions 1 and 2 of Francisco-Fernández and Vilar-Fernández [6]. ■

The next step consists in studying estimator (4). To do this, a similar result to that of Proposition A.2 is obtained.

**Proposition A.3** *Under assumptions A1-A4, we have*

$$\lim_{n \rightarrow \infty} nh_n E \left( H_{(n)}^{-1} \vec{Z}_{(n)}^* \vec{Z}_{(n)}^{*t} H_{(n)}^{-1} \right) = f(x) \tilde{S} v(x) \left( d(\varepsilon) v(x) + 4m^2(x) c(\varepsilon) \right), \quad (16)$$

where  $\vec{Z}_{(n)}^* = X_{p,(n)}^t W_{(n)} \left( \vec{Y}_{(n)}^2 - \vec{G}_{(n)} \right)$ , with  $\vec{G}_{(n)} = (g(x_1), \dots, g(x_n))^t$ .

**Proof.** Taking into account that  $E(\varepsilon_t) = E(\varepsilon_t \varepsilon_r^2) = 0$ , we obtain, for  $i, j = 0, 1, \dots, p$ , that

$$\lim_{n \rightarrow \infty} nh_n E \left( h_n^{-j} z_{j,(n)}^* z_{i,(n)}^* h_n^{-i} \right) = \Delta_1 + \Delta_2, \quad (17)$$

where  $z_{i,(n)}^*$  is the  $i$ -th component of  $\vec{Z}_{(n)}^*$ ,

$$\begin{aligned} \Delta_1 &= \lim_{n \rightarrow \infty} \frac{1}{nh_n} \sum_{t=1}^n \sum_{r=1}^n \left( \frac{x_t - x}{h_n} \right)^j \left( \frac{x_r - x}{h_n} \right)^i K \left( \frac{x_t - x}{h_n} \right) K \left( \frac{x_r - x}{h_n} \right) \\ &\quad v(x_t) v(x_r) E \left( (\varepsilon_t^2 - 1) (\varepsilon_r^2 - 1) \right) \end{aligned}$$

and

$$\begin{aligned} \Delta_2 &= \lim_{n \rightarrow \infty} \frac{4}{nh_n} \sum_{t=1}^n \sum_{r=1}^n \left( \frac{x_t - x}{h_n} \right)^j \left( \frac{x_r - x}{h_n} \right)^i K \left( \frac{x_t - x}{h_n} \right) K \left( \frac{x_r - x}{h_n} \right) \\ &\quad m(x_t) m(x_r) s(x_t) s(x_r) E(\varepsilon_t \varepsilon_r). \end{aligned}$$

Now, using the same kind of arguments as in the proof of Proposition A.2, and the Taylor expansions for  $v(\cdot)$  and  $m(\cdot)$ , we obtain that

$$\Delta_1 = v^2(x) f(x) \left( \int u^{i+j} K^2(u) du \right) d(\varepsilon) \quad (18)$$

and

$$\Delta_2 = 4v(x) m^2(x) f(x) \left( \int u^{i+j} K^2(u) du \right) c(\varepsilon). \quad (19)$$

From (17), (18) and (19), (16) is deduced. ■

Based on the asymptotic properties of  $\hat{\beta}_{(n)}(x)$  and  $\hat{\gamma}_{(n)}(x)$  the asymptotic properties of  $\hat{v}_n^S(x)$  can be obtained. Taking into account that  $\hat{\beta}_{(n)}(x)^t e_1$  converges in probability to  $\vec{\beta}(x)^t e_1$ , it is easy to see that

$$\hat{v}_n^S(x) - v(x) \approx (\hat{\gamma}_{(n)}(x) - \vec{\gamma}(x))^t e_1 - \left( 2\vec{\beta}(x)^t e_1 \left( \hat{\beta}_{(n)}(x) - \vec{\beta}(x) \right)^t e_1 \right). \quad (20)$$

To deduce the variance of  $\hat{v}_n^S(x)$  it is necessary to obtain the covariance between  $\hat{\gamma}_{(n)}(x)$  and  $\hat{\beta}_{(n)}(x)$ . This is deduced from the following result.

**Proposition A.4** *Under assumptions A1-A4, we have*

$$\lim_{n \rightarrow \infty} nh_n E \left( H_{(n)}^{-1} \vec{Z}_{(n)}^* \vec{T}_{(n)}^{*t} H_{(n)}^{-1} \right) = 2f(x) \tilde{S} v(x) m(x) c(\varepsilon).$$

**Proof.** The proof is similar to that of Propositions A.2 or A.3, and therefore it is omitted here. ■

Finally, to prove Theorem 3.1, it suffices to study the asymptotic performance of the vector  $\vec{U}_{(n)}^* = \left( \vec{Z}_{(n)}^*, \vec{T}_{(n)}^* \right)^t$ .

**Proposition A.5** *If assumptions A1-A7 are fulfilled, we have*

$$\sqrt{nh_n} \mathbb{H}_{(n)}^{-1} \vec{U}_{(n)}^* \xrightarrow{\mathcal{L}} N_{2(p+1)} \left( \vec{0}, \Sigma_0 \right), \text{ as } n \rightarrow \infty, \quad (21)$$

where  $\mathbb{H}_{(n)} = \text{diag}(1, h_n, \dots, h_n^p, 1, h_n, \dots, h_n^p)$  and

$$\Sigma_0 = \{v(x)f(x)\} \begin{pmatrix} d(\varepsilon)v(x) + 4m^2(x)c(\varepsilon) & 2m(x)c(\varepsilon) \\ 2m(x)c(\varepsilon) & c(\varepsilon) \end{pmatrix} \otimes \tilde{S}.$$

**Proof.** Let  $Q_n$  be an arbitrary linear combination of  $\mathbb{H}_{(n)}^{-1}\vec{U}_{(n)}^*$ ,

$$Q_n = a^t \begin{pmatrix} H_{(n)}^{-1}\vec{Z}_{(n)}^* \\ H_{(n)}^{-1}\vec{T}_{(n)}^* \end{pmatrix}, \quad \text{with } a \in \mathbb{R}^{2(p+1)}.$$

Using the well-known *small-blocks and large-blocks* method the asymptotic normality of  $\sqrt{nh_n}Q_n$  is established, following exactly the same lines as those in Proposition 3 of Francisco-Fernández and Vilar-Fernández [6]. Then, (21) is obtained from the Cramer-Wold theorem.

■

Now, from equation (20) and Propositions A.1, A.2, A.3, A.4 and Theorem 3.1, the proof of Theorem 3.2 is straightforward.

**Proof. (Proof of Theorem 3.3)**

Proposition A.1 and simple calculations lead to

$$\hat{v}_n^D(x) - v(x) \approx \frac{1}{nh_{2n}f(x)} \sum_{i=1}^n L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) \left\{ \hat{r}_i - \sum_{j=0}^{p_2} \frac{v^{(j)}(x)}{j!} (x_i - x)^j \right\} \quad (22)$$

and

$$\hat{m}_{h_1}(x) - m(x) \approx \frac{1}{nh_{1n}f(x)} \sum_{i=1}^n L_{1,p_1} \left( \frac{x_i - x}{h_{1n}} \right) s(x_i)\varepsilon_i + h_{1n}^{p_1+1} \frac{m^{(p_1+1)}(x)}{(p_1+1)!} \mu_{p_1+1}(L_{1,p_1}). \quad (23)$$

Note that

$$\begin{aligned} \hat{r}_i &= \{Y_i - \hat{m}_{h_{1n}}(x_i)\}^2 = \{s(x_i)\varepsilon_i + m(x_i) - \hat{m}_{h_{1n}}(x_i)\}^2 \\ &= v(x_i)\varepsilon_i^2 + 2s(x_i)\varepsilon_i \{m(x_i) - \hat{m}_{h_{1n}}(x_i)\} + \{m(x_i) - \hat{m}_{h_{1n}}(x_i)\}^2. \end{aligned} \quad (24)$$

By replacing (24) in (22), it follows that

$$\hat{v}_n^D(x) - v(x) \approx I_1 + I_2 - 2I_3 + I_4, \quad (25)$$

where

$$\begin{aligned} I_1 &= \frac{1}{nh_{2n}f(x)} \sum_{i=1}^n L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) \left\{ v(x_i) - \sum_{j=0}^{p_2} \frac{v^{(j)}(x)}{j!} (x_i - x)^j \right\}, \\ I_2 &= \frac{1}{nh_{2n}f(x)} \sum_{i=1}^n L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) v(x_i) (\varepsilon_i^2 - 1), \\ I_3 &= \frac{1}{nh_{2n}f(x)} \sum_{i=1}^n L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) s(x_i) \varepsilon_i \{ \hat{m}_{h_{1n}}(x_i) - m(x_i) \}, \\ I_4 &= \frac{1}{nh_{2n}f(x)} \sum_{i=1}^n L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) \{ \hat{m}_{h_{1n}}(x_i) - m(x_i) \}^2. \end{aligned}$$

We will prove that in (25), term  $I_1$  provides the bias of  $\hat{v}_n^D(x)$ ,  $I_2$  its variance and terms  $I_3$  and  $I_4$  are asymptotically negligible.

A Taylor expansion and typical approximations of sums by integrals lead to

$$I_1 \approx h_{2n}^{p_2+1} \frac{v^{(p_2+1)}(x)}{(p_2+1)!} \mu_{p_2+1}(L_{2,p_2}).$$

As far as  $I_2$  is concerned, using Assumption A4 to show that  $E(I_2) = 0$ , following the same lines as Proposition A.3 to show that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{nh_{2n}f^2(x)} E \left( \sum_{i=1}^n \sum_{j=1}^n L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) v(x_i) (\varepsilon_i^2 - 1) L_{2,p_2} \left( \frac{x_j - x}{h_{2n}} \right) \right. \\ \left. v(x_j) (\varepsilon_j^2 - 1) \right) = \frac{v^2(x)}{f(x)} d(\varepsilon) R(L_{2,p_2}) \end{aligned}$$

and using again the *small-blocks and large-blocks* method, as in Theorem 3.1, to establish the asymptotic normality, we have that

$$(nh_{2n})^{1/2} I_2 \xrightarrow{\mathcal{L}} N(0, \sigma_D^2(x)),$$



where

$$\sigma_D^2(x) = \frac{v^2(x)}{f(x)} d(\varepsilon) R(L_{2,p_2}).$$

The rest of the proof is devoted to proving that  $I_3 = I_4 = o_p(h_{2n}^{p_2+1})$ .

Using (23),  $I_3$  can be written as

$$\begin{aligned} I_3 &\approx \frac{1}{n^2 h_{1n} h_{2n} f(x)} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{f(x_i)} L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) L_{1,p_1} \left( \frac{x_j - x_i}{h_{1n}} \right) s(x_i) s(x_j) \varepsilon_i \varepsilon_j \\ &\quad + \frac{h_{1n}^{p_1+1} \mu_{p_1+1}(L_{1,p_1})}{n h_{2n} f(x)} \sum_{i=1}^n L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) s(x_i) m^{(p_1+1)}(x_i) \varepsilon_i \\ &= I_{31} + I_{32}. \end{aligned}$$

Splitting  $I_{31}$ , we have

$$\begin{aligned} \mathbb{P} \{ |I_{31}| > \varepsilon \} &= \mathbb{P} \{ |I_{311} + I_{312}| > \varepsilon \} \\ &\leq \mathbb{P} \left\{ |I_{311}| > \frac{\varepsilon}{2} \right\} + \mathbb{P} \left\{ |I_{312}| > \frac{\varepsilon}{2} \right\}, \end{aligned}$$

where

$$I_{311} = \frac{1}{n^2 h_{1n} h_{2n} f(x)} \sum_{i=1}^n \frac{1}{f(x_i)} L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) s^2(x_i) \varepsilon_i^2$$

and

$$\begin{aligned} I_{312} &= \frac{1}{n^2 h_{1n} h_{2n} f(x)} \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{1}{f(x_i)} L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) L_{1,p_1} \left( \frac{x_j - x_i}{h_{1n}} \right) \\ &\quad \times s(x_i) s(x_j) \varepsilon_i \varepsilon_j. \end{aligned}$$

With respect to  $I_{311}$ , using Markov's inequality and typical approximations, it is easy to show that

$$\mathbb{P} \left\{ |I_{311}| > \frac{\varepsilon}{2} \right\} \leq \frac{2}{\varepsilon} \mathbb{E} |I_{311}| = O \left( \frac{1}{n h_{1n}} \right).$$

For the term  $I_{312}$ , using again Markov's inequality, we have

$$\begin{aligned}
\mathbb{P} \left\{ |I_{312}| > \frac{\varepsilon}{2} \right\} &\leq \frac{4}{\varepsilon^2} \mathbb{E} (I_{312}^2) \\
&= C \frac{1}{n^4 h_{1n}^2 h_{2n}^2} \sum_{i=1}^n L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) \\
&\quad \times \sum_{p=1}^{n-i} L_{1,p_1} \left( \frac{x_{i+p} - x_i}{h_{1n}} \right) \sum_{u=1}^n L_{2,p_2} \left( \frac{x_u - x}{h_{2n}} \right) \\
&\quad \times \sum_{q=1}^{n-u} L_{1,p_1} \left( \frac{x_{u+q} - x_u}{h_{1n}} \right) \mathbb{E} (\varepsilon_i \varepsilon_{i+p} \varepsilon_u \varepsilon_{u+q}), \tag{26}
\end{aligned}$$

where  $C$  is a generic constant.

Since the errors  $\{\varepsilon_t\}$  are  $\text{MA}(\infty)$ , with autocovariance function  $c(j) = \mathbb{E}(\varepsilon_t, \varepsilon_{t+j})$ , using a result given in Brockwell and Davis [2], pp. 226–227, and taking into account that the kurtosis of  $\{e_j\}$  is 0, we have

$$\mathbb{P} \left\{ |I_{312}| > \frac{\varepsilon}{2} \right\} \leq \Gamma_1 + \Gamma_2 + \Gamma_3, \tag{27}$$

where

$$\begin{aligned}
\Gamma_1 &= \frac{C}{n^4 h_{1n}^2 h_{2n}^2} \sum_{i=1}^n \sum_{p=1}^{n-i} \sum_{u=1}^n \sum_{q=1}^{n-u} L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) L_{1,p_1} \left( \frac{x_{i+p} - x_i}{h_{1n}} \right) \\
&\quad \times L_{2,p_2} \left( \frac{x_u - x}{h_{2n}} \right) L_{1,p_1} \left( \frac{x_{u+q} - x_u}{h_{1n}} \right) c(p)c(q),
\end{aligned}$$

$$\begin{aligned}
\Gamma_2 &= \frac{C}{n^4 h_{1n}^2 h_{2n}^2} \sum_{i=1}^n \sum_{p=1}^{n-i} \sum_{u=1}^n \sum_{q=1}^{n-u} L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) L_{1,p_1} \left( \frac{x_{i+p} - x_i}{h_{1n}} \right) \\
&\quad \times L_{2,p_2} \left( \frac{x_u - x}{h_{2n}} \right) L_{1,p_1} \left( \frac{x_{u+q} - x_u}{h_{1n}} \right) c(u-i)c(u-i-p+q)
\end{aligned}$$

and

$$\begin{aligned}
\Gamma_3 &= \frac{C}{n^4 h_{1n}^2 h_{2n}^2} \sum_{i=1}^n \sum_{p=1}^{n-i} \sum_{u=1}^n \sum_{q=1}^{n-u} L_{2,p_2} \left( \frac{x_i - x}{h_{2n}} \right) L_{1,p_1} \left( \frac{x_{i+p} - x_i}{h_{1n}} \right) \\
&\quad \times L_{2,p_2} \left( \frac{x_u - x}{h_{2n}} \right) L_{1,p_1} \left( \frac{x_{u+q} - x_u}{h_{1n}} \right) c(u-i+q)c(u-i-p).
\end{aligned}$$

Using Riemann approximations of sums by integrals, changes of variables and considering that the kernels have bounded support, it is not difficult to show that  $\Gamma_1 = \Gamma_2 = \Gamma_3 = o\left(\frac{1}{nh_{1n}}\right)$ .

With respect to  $I_{32}$ , using again Markov's inequality, we have

$$\mathbb{P}\{|I_{32}| > \varepsilon\} \leq \frac{\mathbb{E}|I_{32}|^2}{\varepsilon^2} = o(h_{2n}^{p_2+1}).$$

Finally, with respect to  $I_4$ , from Assumption A8, we obtain

$$\mathbb{E}\{\hat{m}_{h_{1n}}(x_i) - m(x_i)\}^2 = \left\{h_{1n}^{2(p_1+1)} + (nh_{1n})^{-1}\right\} = o(h_{2n}^{p_2+1}).$$

Using once again approximations of sums by integrals, it follows that  $I_4 = o_P(h_{2n}^{p_2+1})$ . ■

## References

- [1] Andersen, T.G and Lund, J., 1997, Estimating continuous time stochastic volatility models of the short term interest rate. *Journal of Econometrics*, **77**, 343–377.
- [2] Brockwell, P.J. and Davis, R.A., 1991, *Time Series: Theory and Methods*. Springer Series in Statistics, 2nd edition.
- [3] Doukhan, P., 1995, *Mixing. Properties and Examples*. Notes in Statistics, Springer: Berlin, Vol 85.
- [4] Fan, J., 1992, Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, **87**, 998–1004.

- [5] Fan, J. and Yao, Q., 1998, Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, **85**, 645–660.
- [6] Francisco-Fernández, M. and Vilar-Fernández, J.M., 2001, Local polynomial regression estimation with correlated errors. *Communications in Statistics. Theory and Methods*, **30**(7), 1271–1293.
- [7] Francisco-Fernández, M. and Vilar-Fernández, J. M., 2005, Bandwidth selection for the local polynomial estimator under dependence: a simulation study. *Computational Statistics*, **20**(4), 539–558.
- [8] Gasser, T., Müller, H.G. and Mammitzsch, V., 1985, Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society, Series B*, **47**, 238–252.
- [9] Gasser, T., Sroka, L. and Jennen-Steinmetz, C., 1986, Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 625–633.
- [10] Hall, P., Kay, J.W. and Titterington, D.M., 1990, Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, **77**, 521–528.
- [11] Härdle, W. and Tsybakov, A., 1997, Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics*, **81**, 223–242.

- [12] Hart, J., 1994, Automated kernel smoothing of dependent data by using time series cross-validation. *Journal of the Royal Statistical Society, Series B*, **56**, 529–542.
- [13] Masry, E., 1996, Multivariate regression estimation–local polynomial fitting for time series. *Stochastic Processes and their Applications*, **65**, 81–101.
- [14] Masry, E. and Fan, J., 1997, Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, **24**, 165–179.
- [15] Masry, E. and Tjøstheim, D., 1995, Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory*, **11**, 258–289.
- [16] Opsomer, J.D., Wang, Y. and Yang, Y., 2001, Nonparametric regression with correlated errors. *Statistical Science*, **16**, 134–153.
- [17] Ruppert, D. and Wand, M.P., 1994, Multivariate locally weighted least squares regression. *Annals of Statistics*, **22**, 1346–1370.
- [18] Ruppert, D., Wand, M.P., Holst, U. and Hössjer, O., 1997, Local polynomial variance-function estimation. *Technometrics*, **39**, 3, 262–273.
- [19] Tsybakov, A., 1986, Robust reconstruction of functions by the local approximation method. *Problems of Information Transmission*, **22**, 133–146.
- [20] Ziegelmann, F.A., 2002, Nonparametric estimation of volatility functions: the local exponential estimator. *Econometric Theory*, **18**, 985–991.

## List of Figures

1	MISE of $\hat{v}_n^S(x)$ (solid line), $\hat{v}_n^D(x)$ (dash-dot line) and $\hat{v}_n^b(x)$ (dashed line) as a function of $\rho$ . Interior points at left and whole interval at right. . . . .	31
2	Sample data and local linear estimator of $m(x)$ computed with $\hat{h}_{TSCV} = 0.2229$ obtained from TSCV method. . . . .	32
3	Volatility estimators: $\hat{v}_n^S(x)$ (dashed line) and $\hat{v}_n^D(x)$ (solid line) . . .	33

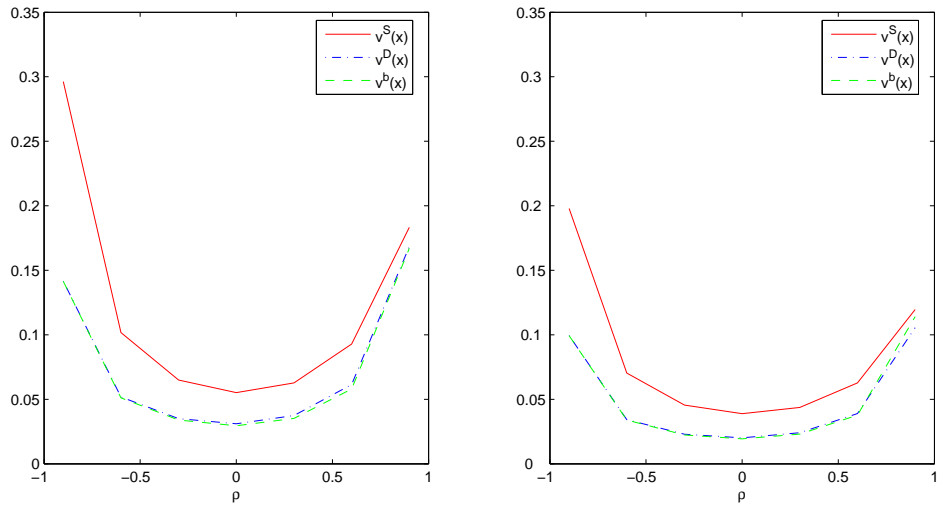


Figure 1: MISE of  $\hat{v}_n^S(x)$  (solid line),  $\hat{v}_n^D(x)$  (dash-dot line) and  $\hat{v}_n^b(x)$  (dashed line) as a function of  $\rho$ . Interior points at left and whole interval at right.

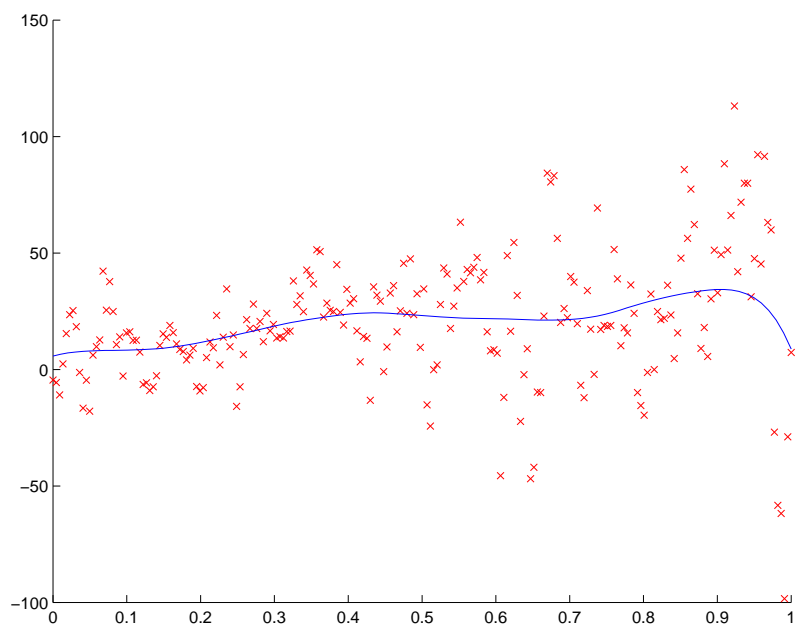


Figure 2: Sample data and local linear estimator of  $m(x)$  computed with  $\hat{h}_{TSCV} = 0.2229$  obtained from TSCV method.



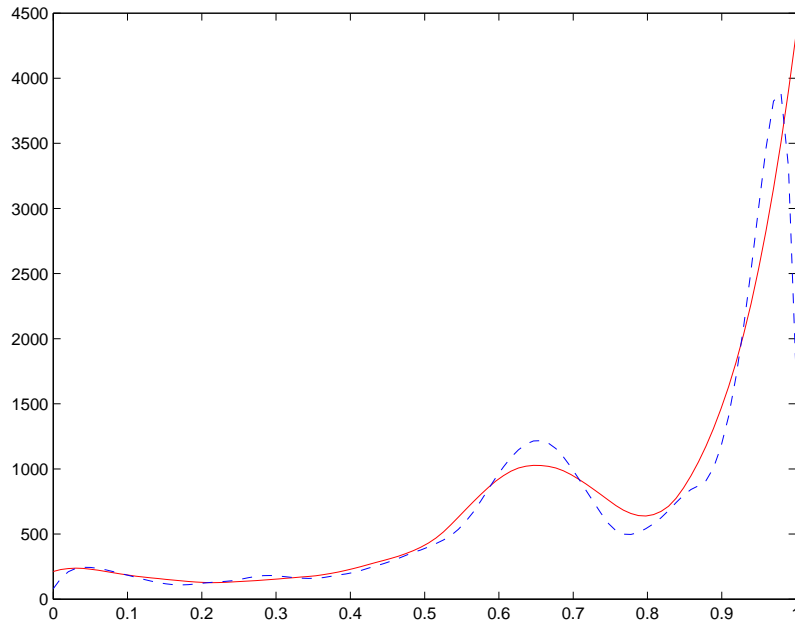


Figure 3: Volatility estimators:  $\hat{v}_n^S(x)$  (dashed line) and  $\hat{v}_n^D(x)$  (solid line)

Table 1:  $m(x) = \sin(\pi x)$ ,  $s(x) = 0.5x$ ,  $\rho = 0.3$  and  $\rho = 0.9$ .

$s(x) =$ $\frac{1}{2}x$	$\rho = 0.3$			$\rho = 0.9$			
		$\hat{v}_n^S(x)$	$\hat{v}_n^D(x)$	$\hat{v}_n^b(x)$	$\hat{v}_n^S(x)$	$\hat{v}_n^D(x)$	$\hat{v}_n^b(x)$
<b>Central</b>	<b>h<sub>MISE</sub></b>	0.1636	0.4061	0.3737	0.2000	0.3990	0.5050
	$\int$ <b>Bias<sup>2</sup></b>	0.00009	0.00002	0.00002	0.00091	0.00163	0.00004
	$\int$ <b>Var</b>	0.00082	0.00036	0.00043	0.00150	0.00034	0.00249
	<b>MISE</b>	0.00091	0.00038	0.00045	0.00241	0.00197	0.00253
<b>Global</b>	<b>h<sub>MISE</sub></b>	0.1636	1.0000	1.0000	0.1818	1.0000	1.0000
	$\int$ <b>Bias<sup>2</sup></b>	0.00041	0.00030	0.00016	0.00373	0.00442	0.00019
	$\int$ <b>Var</b>	0.00220	0.00085	0.00105	0.00198	0.00048	0.00475
	<b>MISE</b>	0.00261	0.00115	0.00121	0.00571	0.00490	0.00494
<b>Bound.</b>	<b>h<sub>MISE</sub></b>	0.1636	1.0000	1.0000	0.1727	1.0000	1.0000
	$\int$ <b>Bias<sup>2</sup></b>	0.00088	0.00064	0.00013	0.00778	0.00845	0.00028
	$\int$ <b>Var</b>	0.00427	0.00149	0.00185	0.00283	0.00077	0.00801
	<b>MISE</b>	0.00515	0.00213	0.00198	0.01061	0.00922	0.00829

Table 2:  $m(x) = \sin(\pi x)$ ,  $s(x) = \sin(\pi x)$ ,  $\rho = -0.6$  and  $\rho = 0.6$ .

$s(x) =$ $\sin(\pi x)$	$\rho = -0.6$			$\rho = 0.6$			
		$\hat{v}_n^S(x)$	$\hat{v}_n^D(x)$	$\hat{v}_n^b(x)$	$\hat{v}_n^S(x)$	$\hat{v}_n^D(x)$	$\hat{v}_n^b(x)$
<b>Central</b>	<b>h<sub>MISE</sub></b>	0.4727	0.3990	0.4040	0.4182	0.3232	0.3959
	$\int$ <b>Bias<sup>2</sup></b>	0.01657	0.01404	0.01327	0.02462	0.02411	0.01613
	$\int$ <b>Var</b>	0.03112	0.03836	0.03785	0.03489	0.03788	0.04187
	<b>MISE</b>	0.04769	0.05240	0.05112	0.05951	0.06199	0.05800
<b>Global</b>	<b>h<sub>MISE</sub></b>	0.4636	0.3939	0.3959	0.4000	0.3232	0.3879
	$\int$ <b>Bias<sup>2</sup></b>	0.01224	0.00876	0.00815	0.01562	0.01465	0.00978
	$\int$ <b>Var</b>	0.02127	0.02582	0.02574	0.02437	0.02430	0.02789
	<b>MISE</b>	0.03351	0.03458	0.03389	0.03999	0.03895	0.03767
<b>Bound.</b>	<b>h<sub>MISE</sub></b>	0.1454	0.2576	0.2505	0.1273	0.2626	0.2343
	$\int$ <b>Bias<sup>2</sup></b>	0.00126	0.00100	0.00106	0.00035	0.00026	0.00075
	$\int$ <b>Var</b>	0.00720	0.00564	0.00571	0.00544	0.00397	0.00527
	<b>MISE</b>	0.00846	0.00664	0.00677	0.00579	0.00423	0.00602