

Available online at www.sciencedirect.com



COMPTES RENDUS MATHEMATIQUE

C. R. Acad. Sci. Paris, Ser. I 343 (2006) 541-544

http://france.elsevier.com/direct/CRASS1/

Statistics

Non-parametric estimation of the average growth curve with a general non-stationary error process

Karim Benhenni, Mustapha Rachdi

Université de Grenoble, UFR SHS, BP 47, 38040 Grenoble cedex 09, France Received 12 July 2005; accepted after revision 5 September 2006 Available online 16 October 2006 Presented by Paul Deheuvels

Abstract

The non-parametric estimation of the average growth curve has been extensively studied in both stationary and some nonstationary particular situations. In this Note, we consider the statistical problem of estimating the average growth curve for a fixed design model with a non-stationary error process. The non-stationarity considered here is of a general form, and this note may be considered as an extension of previous results. The optimal bandwidth is shown to depend on the singularity of the autocovariance function of the error process along the diagonal. *To cite this article: K. Benhenni, M. Rachdi, C. R. Acad. Sci. Paris, Ser. I 343* (2006).

© 2006 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Résumé

Estimation non paramétrique de la courbe de croissance pour un processus d'erreur non stationnaire. L'estimation non paramétrique de la courbe de croissance moyenne pour des données répétées et avec un processus d'erreur non stationnaire a été étudiée, pour des formes particulières de la fonction d'autocovariance, par plusieurs auteurs. Pour ce même problème, nous avons étudié l'estimation de la fonction de croissance avec un processus d'erreur non stationnaire généralisé (pas de forme spécifique de sa fonction d'autocovariance). La fenêtre optimale obtenue dépend de la singularité de la fonction d'autocovariance sur la diagonale. *Pour citer cet article : K. Benhenni, M. Rachdi, C. R. Acad. Sci. Paris, Ser. I 343 (2006).* © 2006 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

The growth curve model is useful especially for growths of animals and plants and is applied extensively to biostatistics, medical research and epidemiology, and was considered by many authors such as Von Rosen [11], among others. In pharmacokinetic research, it is a useful problem for the estimation of the concentration-time curve based on the drug's concentration at different sites within the organism.

In non-parametric regression, the correlation of the errors can have some important consequences on the statistical properties of the curve function estimator and on the selection of the smoothing parameter.

E-mail addresses: Karim.Benhenni@upmf-grenoble.fr (K. Benhenni), Mustapha.Rachdi@imag.fr (M. Rachdi).

¹⁶³¹⁻⁰⁷³X/\$ – see front matter © 2006 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved. doi:10.1016/j.crma.2006.09.009

The non-parametric regression model with correlated errors was considered by many authors for different modifications of kernel construction of the non-parametric regression estimator. For instance, Lin and Carroll [8] and Benhenni and Rachdi [1], among others.

The focus of this Note is to look at the problem of estimating the mean function $f(\cdot)$ in the presence of correlation not that of estimating the correlation function itself. Zimmerman and Núñez-Antón's paper, [12] provides an overview of the parametric regression literature that deals with the correlated errors case. We consider the statistical problem of estimating the average growth curve for a fixed design model. We consider *m* experimental units, each of them having *n* measurements of the response:

$$Y_j(x_i) = f(x_i) + \varepsilon_j(x_i)$$
 where $j = 1, \dots, m$ and $i = 1, \dots, n$

where f is the unknown average growth curve and (ε_i) is the error process.

The sampling points $\{x_i, i = 1, ..., n\}$ are usually taken equally spaced in time series data, but other type of sampling designs can also be considered, such as deterministic regular (non-uniform) designs and random designs. Fraiman and Meloche [4] showed that when the errors are correlated (stationary or not), the kernel-type estimators are not consistent. Although repeated measurements can naturally arise in practical situations, they can make the estimators of the curve *f* asymptotically consistent (see Hart and Wehrly, [7], and the comments of Härdle, [6]). Other models that take into account the individual effects are considered by Boularan et al. [2] and Núñez-Antón et al. [9], known respectively as two- and three-stage additive models, with specific parametric non-stationary covariance structure.

In this Note, we estimate the average growth curve using the kernel methods under a very general non-stationary error process where the autocovariance function does not have any specific form. This include, as a special case, processes with stationary autocovariance function such as the Ornstein–Uhlenbeck process, and some specific non-stationary class of parametric autocovariance structure considered by Ferreira et al. [3] and expanded by Núñez-Antón et al. [9] to unbalanced data.

We show that the asymptotic properties of the kernel regression estimators and the optimal bandwidth are governed by the behavior of the autocovariance function along the diagonal.

2. Estimation of the average growth curve

The estimator of *f* based on the observations $\{Y_j(x_i), i = 1, ..., n, j = 1, ..., m\}$, when the x_i 's are known constants such that $0 \le x_1 < x_2 < \cdots < x_n \le 1$ with $\max_i |x_i - x_{i-1}| = O(1/n)$, is given for $x \in [0, 1]$, by (see, Gasser and Müller, [5], and Hart and Wehrly, [7]):

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n W_{h,i}(x) \bar{Y}(x_i),$$

where

$$\bar{Y}(x) = \frac{1}{m} \sum_{j=1}^{m} Y_j(x)$$
 and $W_{h,i}(x) = n \int_{m_{i-1}}^{m_i} K_h(x-u) \, \mathrm{d}u$

and the midpoints $\{m_i, i = 1, ..., n\}$ are defined by: $m_0 = 0$, $m_i = (x_i + x_{i+1})/2$, for i = 1, ..., n-1 and $m_n = 1$, with $K_h(x) = 1/hK(x/h)$. The kernel K is a Hölder continuous function with support [-1, 1] and $\int_{-1}^{1} K(v) dv = 1$, and h = h(n, m) is the bandwidth, such that: $h \ge 0$ and $\lim_{n,m\to+\infty} h = 0$ with m/n = O(1) (for more detailed comments on these assumptions, the reader can check the paper by Hart and Wherly, [7]).

Assumptions:

- (i) The autocovariance function ρ exists and is continuous on the square $[0, 1]^2$.
- (ii) $\rho(x, y)$ has left and right continuous first order derivatives at the diagonal x = y. The jump function along the diagonal $\alpha(x) = \rho^{(0,1)}(x, x^-) \rho^{(0,1)}(x, x^+)$ is assumed to be continuous and not identically equal to zero.
- (iii) $\rho(x, y)$ is assumed to have continuous mixed partial derivatives up to order two off the diagonal $x \neq y$ in the unit square and satisfies $\sup_{0 \le x \neq y \le 1} |\rho^{(i,j)}(x, y)| < \infty$ for all integers $i, j: 0 \le i + j \le 2$.

(iv) The autocovariance function ρ has at least 2 continuous mixed partial derivatives such that:

$$\rho^{(i,j)}(x,x) \neq 0, \quad i+j=2, \ \forall x \in [0,1].$$

In this case the jump function $\alpha(x) = 0, \forall x \in [0, 1]$.

Theorem 1. If the autocovariance function ρ satisfies assumptions (i), (ii) and (iii), and f is a twice differentiable continuous function on [0, 1] with $f''(x) \neq 0$ for 0 < x < 1, then as $n, m \to +\infty$

$$\mathbb{E}(\hat{f}_h(x) - f(x))^2 = \frac{1}{m} \left(\rho(x, x) - \frac{1}{2} \alpha(x) C_K h \right) + \frac{h^4}{4} d_K^2 (f''(x))^2 + O\left(\frac{1}{mn} + \frac{h^2}{n}\right) + o\left(h^4 + \frac{h}{m}\right)$$

where $C_K = \int_{-1}^{1} \int_{-1}^{1} |u - v| K(u) K(v) du dv = 2 \int_{-1}^{1} \int_{u}^{1} (v - u) K(u) K(v) du dv$ and $d_K = \int u^2 K(u) du$. The asymptotic mean squared error (MSE) is minimized by taking the bandwidth:

$$h_x^* = (\alpha(x)C_K)^{1/3} (2d_K^2 (f''(x))^2)^{-1/3} m^{-1/3}.$$

The asymptotic global bandwidth can be obtained by using a global error measure of estimation such as the integrated mean squared error:

$$h^* = \left(C_K \int_0^1 \alpha(x) \, \mathrm{d}x\right)^{1/3} \left(2d_K^2 \int_0^1 \left(f''(x)\right)^2 \, \mathrm{d}x\right)^{-1/3} m^{-1/3}.$$

The following error processes satisfy assumptions (i), (ii) and (iii):

Examples:

(1) The Wiener error process with autocovariance $\rho(x, y) = \sigma^2 \min(x, y)$. The jump function $\alpha(x) = \sigma^2 > 0$ and $\rho^{(i,j)}(x, y) = 0$, for all integers *i*, *j* such that i + j = 2 and $x \neq y$. In this case

$$h_x^* = \left(\sigma^2 C_K\right)^{1/3} \left(2d_K^2 \left(f''(x)\right)^2\right)^{-1/3} m^{-1/3}$$

(2) The Uhlenbeck–Ornstein process with stationary autocovariance $\rho(x, y) = \sigma^2 \exp(-\lambda |x - y|/2)$ for $\sigma > 0$ and $\lambda > 0$. The jump function is constant $\alpha(x) = 2\sigma^2 \lambda$.

For a stationary error process with autocovariance $\rho(x, y) = \rho(x - y)$ such as the Uhlenbeck–Ornstein process, the jump function: $\alpha(x) = \rho^{(1)}(0^-) - \rho^{(1)}(0^+) = 2\rho^{(1)}(0^-)$ is constant. Then

$$h_x^* = \left(\rho^{(1)}(0^-)C_K d_K^{-2} \left(f''(x)\right)^{-2}\right)^{1/3} m^{-1/3},$$

which is the bandwidth given by Hart and Wehrly [7].

(3) Consider a transformation of the time scale that can produce non-stationary autocovariance of the form

$$o(x, y) = \sigma^2 \rho^{|x^{\lambda} - y^{\lambda}|/\lambda}$$
 for $(x, y) \in [0, 1]^2$, $\sigma^2 > 0$, $0 < \rho < 1$ and $\lambda > 0$

(see, Núñez-Antón and Woodworth, [10]). In particular when $\lambda = 1$, we obtain a Uhlenbeck–Ornstein error process with a stationary autocovariance function with $\alpha(x) = -2\sigma^2 \ln(\rho)$. When $\lambda \neq 1$, then the autocovariance is non-stationary with jump function $\alpha(x) = -2\sigma^2 \ln(\rho)x^{\lambda-1}$. In this case

$$h_x^* = \left(-x^{\lambda-1}\ln(\rho)C_K\sigma^2 d_K^{-2} \left(f''(x)\right)^{-2}\right)^{1/3} m^{-1/3}$$

which is the bandwidth obtained by Ferreira et al. [3].

The following theorem gives the asymptotic expression of the MSE for smoother error process:

Theorem 2. Assume that ρ satisfies assumption (iv) and f is a twice differentiable continuous function on [0, 1] with $f''(x) \neq 0$, for 0 < x < 1, then as $nm \to +\infty$:

$$\mathbb{E}(\hat{f}_h(x) - f(x))^2 = \frac{1}{m} (\rho(x, x) + \rho^{(0,2)}(x, x)d_K h^2) + \frac{h^4}{4} d_K^2 (f''(x))^2 + O\left(\frac{1}{mn} + \frac{h^2}{n}\right) + o\left(h^4 + \frac{h^2}{m}\right).$$

The asymptotic optimal bandwidth is given by: $h_x^* = (-2\rho^{(0,2)}(x, x)d_K^{-1}(f''(x))^{-2})^{1/2}m^{-1/2}$.

In particular for a stationary error process

$$\rho^{(0,2)}(x,x) = \rho''(0),$$

then

$$h_x^* = \left(-2\rho''(0)\right)^{1/2} \left(d_K \left(f''(x)\right)^2\right)^{-1/2} m^{-1/2}$$

The asymptotic optimal global bandwidth using the integrated mean square criteria is:

$$h^* = \left(2\int_0^1 \rho^{(0,2)}(x,x)\,\mathrm{d}x\right)^{1/2} \left(d_K \int_0^1 \left(f''(x)\right)^2 \mathrm{d}x\right)^{-1/2} m^{-1/2}.$$

Note that results on simulations and proofs of theorems can be requested from the authors.

References

- K. Benhenni, M. Rachdi, Non-parametric estimation of the regression function from quantized observations, Comput. Statist. Data Anal. 11 (50) (2006) 3067–3085.
- [2] J. Boularan, L. Ferré, P. Vieu, Growth curves: a two-stage non-parametric approach, J. Statist. Plann. Inference 3 (38) (1994) 327-350.
- [3] E. Ferreira, V. Núñez-Antón, J. Rodríguez-Póo, Kernel regression estimates of growth curves using non-stationary correlated errors, Statist. Probab. Lett. 4 (34) (1997) 413–423.
- [4] R. Fraiman, J. Meloche, Smoothing dependent observations, Statist. Probab. Lett. 21 (1994) 203–214.
- [5] T. Gasser, M.G. Müller, Estimating regression functions and their derivatives by the kernel method, Scand. J. Statist. 11 (1984) 171–185.
- [6] W. Härdle, Applied Non-Parametric Regression, vol 19, Cambridge University Press, Cambridge, 1989.
- [7] J. Hart, T. Wehrly, Kernel regression estimation using repeated measurements data, J. Amer. Statist. Assoc. 81 (1986) 1080–1088.
- [8] X. Lin, R.J. Carroll, Non-parametric function estimation for clustered data when the predictor is measured without/with error, J. Amer. Statist. Assoc. 450 (95) (2000) 520–534.
- [9] V. Núñez-Antón, J. Rodríguez-Póo, P. Vieu, Longitudinal data with non-stationary errors: a nonparametric three-stage approach, Test 8 (1) (1999) 201–231.
- [10] V. Núñez-Antón, G. Woodworth, Analysis of longitudinal data with unequally spaced observations and time dependent correlated errors, Biometrics 50 (1994) 445–456.
- [11] D. Von Rosen, The growth curve model: a review, Comm. Statist. Theory Methods 20 (9) (1991) 2791-2822.
- [12] D.L. Zimmerman, V. Núñez-Antón, Parametric modelling of growth curve data: an overview (with comments), Test 1 (10) (2001) 1–73.