Revista Colombiana de Estadística
Junio 2012, volumen 35, no. 1, pp. 167 a 184

Random Time-Varying Coefficient Model Estimation through Radial Basis Functions

Estimación de los coeficientes de un modelo de coeficientes dinámicos y aleatorios a través de funciones radiales *kernel*

Juan Camilo Sosa[1,a], Luis Guillermo Díaz[2,b]

[1]Mathematics Department, Universidad Externado de Colombia, Bogotá, Colombia
[2]Statistics Department, Faculty of Science, Universidad Nacional de Colombia, Bogotá, Colombia

Abstract

A methodology to estimate a time-varying coefficient model through a linear combination of radial kernel functions which are centered around all the measuring times, or their quantiles is developed. The linear combination is weighted by a bandwidth that may change or not among coefficients. The proposed methodology is compared with the local polynomial kernel methods by means of a simulation study. The proposed methodology shows a better behavior in a high proportion of times in all cases, or at least it has a similar behavior in relation with the estimation through local polynomial kernel regression, that in a low rate of times has a better behavior in relation with the average mean square error. In order to illustrate the methodology the data set ACTG 315 related with an AIDS study is taken into account. The dynamic relationship between the viral load and the CD4+ cell counts is investigated.

***Key words*:** Cross validation, Kernel function, Longitudinal data analysis, Mixed model.

Resumen

Se propone una metodología para estimar los coeficientes de un modelo de coeficientes dinámicos y aleatorios a través de una combinación lineal de funciones radiales kernel centradas en los diferentes puntos de medición, o en cuantiles de éstos, escalada por un ancho de banda que puede cambiar de coeficiente a coeficiente. En un estudio de simulación se compara la metodología propuesta con la estimación mediante los métodos de polinomios locales *kernel*, obteniéndose que la nueva metodología propuesta es la

[a]Lecturer. E-mail: juan.sosa@uexternado.edu.co
[b]Associated professor. E-mail: lgdiazm@unal.edu.co

mejor opción en un alto porcentaje de veces para todos los escenarios simulados, o por lo menos se desempeña similarmente a la estimación a través de la regresión de polinomios locales *kernel*, que pocas veces se desempeña mejor que la estimación mediante funciones radiales *kernel*, en relación al error cuadrático medio promedio. Para ilustrar la estrategia de estimación propuesta se considera el conjunto de datos ACTG 315 asociado con un estudio del SIDA, en el que se modela dinámicamente la relación entre la carga viral y el conteo de células CD4+.

***Palabras clave*:** análisis de datos longitudinales, función *kernel*, modelo mixto, validación cruzada.

1. Introduction

Longitudinal Data Analysis (LDA) takes place when a set of subjects are observed repeatedly along time, measuring the response variable in accordance with the covariates that may or may not be time-dependent. Given the characteristics of this kind of data, an underlying property that must be thought fitting a statistical model, is the correlation between repeated measures of the response variable within each experimental unit, considering measures independent between subjects. That is, measurements are correlated inside experimental units and independent between subjects. This way, the main purpose is to identify and describe the evolution of the response variable and to determine how it is affected by the covariates. For instance, in clinic trials, it is of interest to evaluate the impact of a dose or other related factors, over the progress of a disease along time.

Parametric techniques for LDA have been exhaustively studied in the literature (Diggle, Liang & Zeger 1994, Davis 2000, Verbeke & Molenberghs 2005, Fitzmaurice, Davidian, Verbeke & Molenberghs 2009). While these tools are useful under some reasonable restrictions, always arise doubts and questions about the adequacy of the model assumptions and the potencial impact of model misspecifications on the analysis (Hoover, Rice, Wu & Yang 1998). Non parametric techniques recently introduced in LDA allow a functional dependence more flexible between the response variable and the covariates.

Hart & Wehrly (1986), Altman (1990), Hart (1991) propose methods for choosing smoothing parameter through cross-validation using kernel functions and considered kernel methods for estimating the expectation of the response variable without covariates, while Rice & Silverman (1991) did it by using a class of smoothing splines. Although the kernel and splines methods are successful in predicting the mean curve of the response variable, they only consider the time effect and do not take into account other important covariates (Hoover et al. 1998).

In order to quantify the influence of covariates, Zeger & Diggle (1994) studied a semi-parametric model as follows:

$$\begin{gathered} y_{ij} = \mu(t_{ij}) + \boldsymbol{x}_i(t_{ij})^T\boldsymbol{\beta} + e_{ij} \\ j = 1, \ldots, n_i,\ i = 1, \ldots, n \end{gathered} \tag{1}$$

where n is the number of subjects, n_i is the number of repeated measures associated with the i-th experimental unit, t_{ij}, $y_{ij} \equiv y_i(t_{ij})$,

$$\boldsymbol{x}_i(t_{ij}) = [x_{i0}(t_{ij}), x_{i1}(t_{ij}), \ldots, x_{id}(t_{ij})]^T$$

and $e_{ij} \equiv e_i(t_{ij})$ are respectively the measuring time, the response variable, the covariate vector in $\mathbb{R}^{d+1}$ and the error term, associated with the j-th measure of the i-th subject. Moreover, $\mu(\cdot)$ is an arbitrary smooth real function and $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_d]^T$ is a parameter vector in $\mathbb{R}^{d+1}$. Working with longitudinal data, it is usually assumed that repeated measures are independent between experimental units and that $e_i(t)$ is a Gaussian Process (GP) with $\mathbb{E}[e_i(t)] = 0$, for each $t \in \mathcal{T}_i$, with covariance function $\gamma_{\boldsymbol{e}_i}(r, s)$, $r, s \in \mathcal{T}_i$, and $\mathcal{T}_i = \{t_{ij} : j = 1, \ldots, n_i\}$; this is written as

$$\boldsymbol{e}_i = [e_{i1}, \ldots, e_{in_i}]^T \sim PG(\mathbf{0}_{n_i}, \boldsymbol{\Gamma}_i)$$

where $\mathbf{0}_{n_i}$ is a column-vector with $n_i \times 1$ zeros and $\boldsymbol{\Gamma}_i = [\gamma_{\boldsymbol{e}_i}(t_{ik}, t_{il})]_{k,l=1,\ldots,n_i}$.

Hoover et al. (1998) considered a generalization of the model (1) that allows the parameters to vary over time. This extension is as follows:

$$\begin{aligned} y_{ij} &= \boldsymbol{x}_i(t_{ij})^T \boldsymbol{\beta}(t_{ij}) + e_{ij}, \\ j &= 1, \ldots, n_i,\ i = 1, \ldots, n \end{aligned} \tag{2}$$

where

$$\boldsymbol{\beta}(t_{ij}) = [\beta_0(t_{ij}), \beta_1(t_{ij}), \ldots, \beta_d(t_{ij})]^T$$

is a vector of arbitrary real smooth functions. Components in vector $\boldsymbol{\beta}(t)$ are called dynamic coefficients or dynamic parameters, and the statistical model (2) is referred as Time-Varying Coefficient Model (TVCM). This kind of model has been widely studied by Wu & Zhang (2006) who investigated various alternatives for estimating the model coefficients. Sosa & Díaz (2010) proposed a methodology to estimate true-varying coefficients models through generalized estimation equations.

A Random Time-Varying Coefficient Model (RVCM) is an extension of a TVCM, and it was firstly investigated by Guo (2002). As in a Linear Mixed Effects Model (LMEM), this extension decomposes the term error $e_i(t_{ij})$ of model (2) into two parts: one of them that describes the characteristics of each subject that differs of the mean population behavior, and other related with the pure random error; that is, it is done by the decomposition

$$\begin{aligned} e_i(t_{ij}) &= \boldsymbol{z}_i(t_{ij})^T \boldsymbol{v}_i(t_{ij}) + \epsilon_i(t_{ij}) \\ j &= 1, \ldots, n_i,\ i = 1, \ldots, n \end{aligned}$$

where $\boldsymbol{z}_i(t_{ij})^T \boldsymbol{v}_i(t_{ij})$ is the model component that describes the characteristics related with each subject (random effects component), with

$$\boldsymbol{z}_i(t_{ij}) = [z_{i0}(t_{ij}), z_{i1}(t_{ij}), \ldots, z_{id^*}(t_{ij})]^T$$

a covariate vector in $\mathbb{R}^{d^*+1}$, with components that vary along time, associated with the vector

$$\boldsymbol{v}_i(t_{ij}) = [v_{i0}(t_{ij}), v_{i1}(t_{ij}), \ldots, v_{id^*}(t_{ij})]^T$$

of random time-varying coefficients with size $(d^* + 1) \times 1$ and $\epsilon_{ij} \equiv \epsilon_i(t_{ij})$ is the random error term associated with the j-th measurement of the i-th experimental unit. Thus, a RVCM is a model with the following form:

$$\begin{gathered} y_{ij} = \boldsymbol{x}_i(t_{ij})^T \boldsymbol{\beta}(t_{ij}) + \boldsymbol{z}_i(t_{ij})^T \boldsymbol{v}_i(t_{ij}) + \epsilon_{ij} \\ j = 1, \ldots, n_i,\ i = 1, \ldots, n \end{gathered} \tag{3}$$

where

$$\boldsymbol{v}_i(t) \sim PG(\mathbf{0}_{d^*+1}, \boldsymbol{\Gamma})$$

and

$$\boldsymbol{\epsilon}_i(t) = [\epsilon_{i1}, \ldots, \epsilon_{in_i}]^T \sim PG(\mathbf{0}_{n_i}, \mathbf{R}_i)$$

with $\boldsymbol{\Gamma} = [\gamma(t_{ik}, t_{il})]_{k,l=1,\ldots,d^*+1}$ and $\mathbf{R}_i = [\gamma_{\boldsymbol{\epsilon}_i}(t_{ik}, t_{il})]_{k,l=1,\ldots,n_i}$. It is supposed that the repeated measurements are independent between subjects, and $\boldsymbol{v}_i(t)$ and $\boldsymbol{\epsilon}_{i^*}(t)$ are independent Gaussian processes.

This paper is structured as follows: In Section 2 and Section 3 the estimation through local polynomial kernel techniques is presented and an estimation methodology by means of radial kernel functions is proposed, respectively. In Section 4 some techniques to choose the bandwidth associated with the estimation methodologies is studied. In section 5 it is shown a simulation study where the estimation alternatives through the average mean square error are compared. In Section 6 the methodology is illustrated by analyzing the data set ACTG 315 (Liang, Wu & Carroll 2003), where the relationship between viral load and CD4+ cell counts is investigated dynamically in a AIDS clinical trial. Finally, results are discussed in 7.

2. Estimation Through Local Polynomial Kernel Regression

The basic idea behind the estimation through Local Polynomial Kernel (LPK) regression is to approximate the dynamic coefficients by means of a Taylor expansion. Thus, in a fix time point t_0, it is supposed that the dynamic parameters $\beta_r(t_0)$, $r = 0, 1, \ldots, d$, and $v_{is}(t_0)$, $s = 0, 1, \ldots, d^*$, have $(p+1)$ continuous derivatives for some non-negative integer p. Then, by means of an approximation in a Taylor expansion of order p around t_0, it follows that:

$$\beta_r(t_{ij}) \approx \boldsymbol{h}_{ij}^T \boldsymbol{\alpha}_r,\ r = 0, 1, \ldots, d \tag{4}$$

and

$$v_{si}(t_{ij}) \approx \boldsymbol{h}_{ij}^T \boldsymbol{b}_{si},\ s = 0, 1, \ldots, d^* \tag{5}$$

for $j = 1, \ldots, n_i$, $i = 1, \ldots, n$, where

$$\boldsymbol{h}_{ij} = [1, t_{ij} - t_0, (t_{ij} - t_0)^2, \ldots, (t_{ij} - t_0)^p]^T$$

is the vector of $(p+1)\times 1$ components related with the polynomials in the approximation, $\boldsymbol{\alpha}_r = [\alpha_{r0}, \alpha_{r1}, \ldots, \alpha_{rp}]^T$ and $\boldsymbol{b}_{si} = [b_{si0}, b_{si1}, \ldots, b_{sip}]^T$, with

$$\alpha_{rk} = \frac{\beta_r^{(k)}(t_0)}{k!} \tag{6}$$

and

$$b_{sik} = \frac{v_{si}^{(k)}(t_0)}{k!} \tag{7}$$

for $k = 0, 1, \ldots, p$.

Let $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_0^T, \boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_d^T]^T$ and $\boldsymbol{b}_i = [\boldsymbol{b}_{0i}^T, \boldsymbol{b}_{1i}^T, \ldots, \boldsymbol{b}_{d^*i}^T]^T$ be the vectors associated with the approximation of the dynamic coefficients. Given that the repeated measurements are independent between subjects and that $\boldsymbol{v}_i(t) \sim PG(\mathbf{0}_{d^*+1}, \boldsymbol{\Gamma})$, it follows that the sequence of vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ constitutes a random sample from a population with a multivariate normal distribution with mean $\mathbf{0}_{(d^*+1)(p+1)}$ and covariance matrix $\mathbf{D} \equiv \mathbf{D}(t_0)$ with size $d^*(p+1)\times d^*(p+1)$. Thus, in a neighborhood of t_0, model (3) can be approximately expressed as

$$\begin{gathered} y_{ij} \approx \boldsymbol{x}_{ij}^T\boldsymbol{\alpha} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i + \epsilon_{ij} \\ j = 1, \ldots, n_i,\ i = 1, \ldots, n \end{gathered} \tag{8}$$

where $\boldsymbol{x}_{ij} = \boldsymbol{x}_i(t_{ij}) \otimes \boldsymbol{h}_{ij}$, $\boldsymbol{z}_{ij} = \boldsymbol{z}_i(t_{ij}) \otimes \boldsymbol{h}_{ij}$, with $\boldsymbol{b}_i \sim N(\mathbf{0}_{(d^*+1)(p+1)}, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}_{n_i}, \mathbf{R}_i)$

Thus, in a neighborhood of t_0, model (8) is a standard LMEM where it is required to estimate $\boldsymbol{\alpha}$ and find the Best Linear Unbiased Predictor (BLUP) of $\boldsymbol{b}_i$, with the purpose of finding the estimations of $\boldsymbol{\beta}(t)$ and $\boldsymbol{v}_i(t)$. In order to incorporate the information given in the neighborhood, as in Wu & Zhang (2006, p. 297), it is constituted the following objective function:

$$(\boldsymbol{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{b})^T\mathbf{K}_h^{1/2}\mathbf{R}^{-1}\mathbf{K}_h^{1/2}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{b}) + \boldsymbol{b}^T\widetilde{\mathbf{D}}^{-1}\boldsymbol{b} \tag{9}$$

where

$$\begin{aligned} \boldsymbol{b} &= [\boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_n^T]^T \\ \boldsymbol{y} &= [\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_n^T]^T, \quad \boldsymbol{y}_i = [y_{i1}, \ldots, y_{in_i}]^T \\ \mathbf{X} &= [\mathbf{X}_1^T, \ldots, \mathbf{X}_n^T]^T, \quad \mathbf{X}_i = [\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i}]^T \\ \mathbf{Z} &= diag[\mathbf{Z}_1, \ldots, \mathbf{Z}_n], \quad \mathbf{Z}_i = [\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{in_i}]^T \\ \widetilde{\mathbf{D}} &= diag[\mathbf{D}, \ldots, \mathbf{D}], \quad \mathbf{R} = diag[\mathbf{R}_1, \ldots, \mathbf{R}_n] \\ \mathbf{K}_h &= diag[\mathbf{K}_{1h}, \ldots, \mathbf{K}_{nh}], \quad \mathbf{K}_{ih} = diag[K_h(t_{i1} - t_0), \ldots, K_h(t_{in_i} - t_0)] \end{aligned} \tag{10}$$

with $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ a kernel function and h a bandwidth.

The estimators can be found fitting the model

$$\begin{gathered} \widetilde{\boldsymbol{y}} = \widetilde{\mathbf{X}}\boldsymbol{\alpha} + \widetilde{\mathbf{Z}}\boldsymbol{b} + \boldsymbol{\epsilon} \\ \boldsymbol{b} \sim N(\mathbf{0}_N, \widetilde{\mathbf{D}}),\ \boldsymbol{\epsilon} \sim N(\mathbf{0}_N, \mathbf{R}) \end{gathered} \tag{11}$$

where $\widetilde{\boldsymbol{y}} = \mathbf{K}_h^{1/2}\boldsymbol{y}$, $\widetilde{\mathbf{X}} = \mathbf{K}_h^{1/2}\mathbf{X}$, $\widetilde{\mathbf{Z}} = \mathbf{K}_h^{1/2}\mathbf{Z}$ and $N = \sum_{i=1}^{n} n_i$.

Therefore, given the variance components $\widetilde{\mathbf{D}}$ and $\mathbf{R}$, the kernel function $K(\cdot)$ and the bandwidth h, to minimize (9) in relation with $\boldsymbol{\alpha}$ and $\boldsymbol{b}$ leads to

$$\widehat{\boldsymbol{\alpha}} = \left(\mathbf{X}^T\mathbf{K}_h^{1/2}\mathbf{V}^{-1}\mathbf{K}_h^{1/2}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{K}_h^{1/2}\mathbf{V}^{-1}\mathbf{K}_h^{1/2}\boldsymbol{y} \tag{12}$$

$$\widehat{\boldsymbol{b}} = \widetilde{\mathbf{D}}\mathbf{Z}^T\mathbf{K}_h^{1/2}\mathbf{V}^{-1}\mathbf{K}_h^{1/2}(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\alpha}}) \tag{13}$$

and

$$\widehat{\boldsymbol{b}}_i = \mathbf{D}\mathbf{Z}_i\mathbf{K}_{ih}^{1/2}\mathbf{V}_i^{-1}\mathbf{K}_{ih}^{1/2}(\boldsymbol{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\alpha}})$$

where

$$\mathbf{V} = diag[\mathbf{V}_1, \ldots, \mathbf{V}_n] = \mathbf{K}_h^{1/2}\mathbf{Z}\widetilde{\mathbf{D}}\mathbf{Z}^T\mathbf{K}_h^{1/2} + \mathbf{R}$$

with

$$\mathbf{V}_i = \mathbf{K}_{ih}^{1/2}\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T\mathbf{K}_{ih}^{1/2} + \mathbf{R}_i$$

3. Estimation through Radial Kernel Functions

The idea behind the estimation through Radial Kernel Functions (RKF) is to approximate the dynamic coefficients by means of a linear combination of kernel functions treated as radial basis functions. Thus, it is possible to express the dynamic parameters by means of

$$\boldsymbol{\beta}(t) = \boldsymbol{\Xi}(t)^T\boldsymbol{\alpha} \tag{14}$$

and

$$\boldsymbol{v}_i(t) = \boldsymbol{\Theta}(t)^T\boldsymbol{b}_i,\ i = 1, \ldots, n \tag{15}$$

where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_0^T, \boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_d^T]^T$, $\boldsymbol{\Xi}(t) = diag[\boldsymbol{\Xi}_0(t), \boldsymbol{\Xi}_1(t), \ldots, \boldsymbol{\Xi}_d(t)]$,

$$\boldsymbol{\alpha}_r = [\alpha_{r1}, \ldots, \alpha_{rM}]^T\ \boldsymbol{\Xi}_r(t) = \left[\xi_r\left(\frac{|t - t_1|}{h}\right), \ldots, \xi_r\left(\frac{|t - t_M|}{h}\right)\right]^T$$

for $r = 0, 1, \ldots, d$, $\boldsymbol{b}_i = [\boldsymbol{b}_{0i}^T, \boldsymbol{b}_{1i}^T, \ldots, \boldsymbol{b}_{d^*i}^T]^T$, $\boldsymbol{\Theta}(t) = diag[\boldsymbol{\Theta}_0(t), \boldsymbol{\Theta}_1(t), \ldots, \boldsymbol{\Theta}_{d^*}(t)]$

$$\boldsymbol{b}_{si} = [b_{si1}, \ldots, b_{siM}]^T\ \boldsymbol{\Theta}_s(t) = \left[\theta_s\left(\frac{|t - t_1|}{h}\right), \ldots, \theta_s\left(\frac{|t - t_M|}{h}\right)\right]^T$$

for $i = 1, \ldots, n$, $s = 0, 1, \ldots, d^*$, with $\xi_r : [0, \infty) \to \mathbb{R}$ and $\theta_s : [0, \infty) \to \mathbb{R}$ kernel functions, $t_1, \ldots, t_M$ are all the M measurements time points that are different (or quantils of these) and h is a bandwidth.

If $\xi_r \equiv \xi$ for each $r = 0, 1, \ldots, d$ and $\theta_s \equiv \theta$ for each $s = 0, 1, \ldots, d^*$, then

$$\boldsymbol{\beta}(t) = \left[\mathbf{I}_{d+1} \otimes \boldsymbol{\xi}(t)\right]^T\boldsymbol{\alpha} \tag{16}$$

and

$$\boldsymbol{v}_i(t) = \left[\mathbf{I}_{d^*+1} \otimes \boldsymbol{\theta}(t)\right]^T\boldsymbol{b}_i,\ i = 1, \ldots, n \tag{17}$$

where $\mathbf{I}_k$ denote the identity matrix of $k \times k$,

$$\boldsymbol{\xi}(t) = \left[\xi\left(\frac{|t-t_1|}{h}\right), \ldots, \xi\left(\frac{|t-t_M|}{h}\right)\right]^T \tag{18}$$

and

$$\boldsymbol{\theta}(t) = \left[\theta\left(\frac{|t-t_1|}{h}\right), \ldots, \theta\left(\frac{|t-t_M|}{h}\right)\right]^T \tag{19}$$

As above, given that $\boldsymbol{v}_i(t) \sim PG(\mathbf{0}_{d^*+1}, \boldsymbol{\Gamma})$ and that the repeated measurements are independent between subjects, it follows that the sequence of vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ constitutes a random sample from a population with a multivariate normal distribution with mean $\mathbf{0}_{(d^*+1)(p+1)}$ and covariance matrix $\mathbf{D} \equiv \mathbf{D}(t)$ with size $d^*(p+1) \times d^*(p+1)$. Due to (3) and (15), it follows that $\gamma(s,t) = \boldsymbol{\Theta}(s)^T \mathbf{D} \boldsymbol{\Theta}(t)$, so that an estimator of $\mathbf{D}$ leads directly to an estimator of $\boldsymbol{\Gamma}$.

Thus, model (3) can be approximately expressed as

$$\begin{gathered} y_{ij} \approx \boldsymbol{x}_{ij}^T \boldsymbol{\alpha} + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i + \epsilon_{ij} \\ j = 1, \ldots, n_i,\ i = 1, \ldots, n \end{gathered} \tag{20}$$

where $\boldsymbol{x}_{ij} = \boldsymbol{\Xi}(t_{ij})\boldsymbol{x}_i(t_{ij})$ and $\boldsymbol{z}_{ij} = \boldsymbol{\Theta}(t_{ij})\boldsymbol{z}_i(t_{ij})$, with $\boldsymbol{b}_i \sim N(\mathbf{0}_{(d^*+1)(p+1)}, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}_{n_i}, \mathbf{R}_i)$.

If $\xi_r \equiv \xi$ and $\theta_s \equiv \theta$ then

$$\boldsymbol{x}_{ij} = (\mathbf{I}_{d+1} \otimes \boldsymbol{\xi}(t_{ij}))\boldsymbol{x}_i(t_{ij})$$

and

$$\boldsymbol{z}_{ij} = (\mathbf{I}_{d^*+1} \otimes \boldsymbol{\theta}(t_{ij}))\boldsymbol{z}_i(t_{ij})$$

where $\boldsymbol{\xi}(t)$ and $\boldsymbol{\theta}(t)$ are given in (18) and (19).

Given the vectors $\boldsymbol{\Xi}_r(t)$, $r = 0, 1, \ldots, d$, and $\boldsymbol{\Theta}_s(t)$, $s = 0, 1, \ldots, d^*$, and the bandwidth h, model (20) is a standard LMEM where it is required to estimate $\boldsymbol{\alpha}$ and find the BLUP of $\boldsymbol{b}_i$ in order to calculate the estimations of $\boldsymbol{\beta}(t)$ and $\boldsymbol{v}_i(t)$.

4. Bandwidth Selection

By estimating the dynamic components of model 3 through LPK or RKF, it is mandatory to choose the bandwidth h carefully. In this section are presented two selection criterions designed to choose smoothing parameters: Subject Cross-Validation (SCV) and Point Cross-Validation (PCV).

4.1. Subject Cross-Validation

This criterion was proposed by Rice & Silverman (1991), and has been studied by many authors, as Hoover et al. (1998) for instance. The idea behind this criteria

is to choose the smoothing parameter vector that minimize the expression

$$SCV(h) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} w_i \left[y_{ij} - \boldsymbol{x}_i(t_{ij})^T \widehat{\boldsymbol{\beta}}^{(-i)}(t_{ij}) \right]^2 \tag{21}$$

where y_{ij} and $x_i(t_{ij})$ are defined as in model (3), $\widehat{\boldsymbol{\beta}}^{(-i)}(t)$ denotes the estimation of $\boldsymbol{\beta}(t)$ excluding the data related with the i-th subject, and w_i for $i = 1 \ldots, n$, is a weight given by some of the following schemes:

Scheme 1 All weights are given by $w_i = 1/N$, $i = 1, \ldots, n$, where $N = \sum_{i=1}^{n} n_i$.

Scheme 2 All weights are given by $w_i = 1/(nn_i)$, $i = 1, \ldots, n$.

Scheme 1 uses the same weight for all experimental units and was proposed by Hoover et al. (1998). Scheme 2 is considered by Huang, Wu & Zhou (2002) and uses different weights for the subjects taken into account in the study. In Huang et al. (2002) it is shown that scheme 1 could lead to inconsistent estimators of $\boldsymbol{\alpha}$.

4.2. Point Cross Validation

Let $\{t_l : l = 1, \ldots, M\}$ be the set formed by all the measuring times that are different (or quantiles of these) in all the data set. For a given time point t_l, let $\{i_{l^*} : l^* = 1, \ldots, m_l\}$ be the set of all experimental units at time t_l.

The idea behind this criteria is to choose the smoothing parameter vector that minimize the expression

$$PCV(h) = \sum_{l=1}^{M} \sum_{l^*=1}^{m_l} w_l \left[y_{i_{l^*}}(t_{l^*}) - \widehat{s}_{i_{l^*}}^{(-l)}(t_l) \right]^2 \tag{22}$$

where $y_{i_{l^*}}(t_{l^*})$ is the value of the response variable for subject i_{l^*} at time t_{l^*}, $w_l = (Mm_l)^{-1}$ is the weight associated with the l-th measuring time and $\widehat{s}_{i_{l^*}}^{(-l)}(t_l)$ denotes the estimation of the response variable for experimental unit i_{l^*} at time t_l when all the observations related with the response variable at time t_l have been excluded.

5. Simulation

This section presents a simulation study to evaluate the performance of the estimation methods. The comparison is performed through the Average Mean Square Error (AMSE) given by

$$AMSE(\kappa) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \left[\kappa(t_{ij}) - \widehat{\kappa}(t_{ij}) \right]^2 \tag{23}$$

with $\kappa(\cdot)$ a function that corresponds to any dynamic coefficient of model (3).

Simulation strategy is similar to that followed by Wu & Liang (2004). The simulation model is structured as follows:

$$\begin{aligned} y_i(t) &= \beta_0(t) + x_{i1}(t)\left[\beta_1(t) + v_{1i}(t)\right] + \epsilon_i(t),\ i = 1, \ldots, n \\ x_{i1}(t) &= 1 - \exp\left[-0.5t - (i/n)\right] \\ \beta_0(t) &= 3\exp(t),\ \beta_1(t) = 1 + \cos(2\pi t) + \sin(2\pi t) \\ v_{1i}(t) &= a_{i0} + a_{i1}\cos(2\pi t) + a_{i2}\sin(2\pi t) \\ a_i &= [a_{i0}, a_{i1}, a_{i2}]^T \sim N\left([0,0,0]^T, diag[\sigma_0^2, \sigma_1^2, \sigma_2^2]\right) \\ \epsilon_i(t) &\sim N(0, \sigma_\epsilon^2 x_{i1}^2(t)) \end{aligned} \tag{24}$$

where $\beta_0(t)$, $\beta_1(t)$ and $v_{1i}(t)$, are the dynamic parameters of the model, $x_i(t)$ is the covariate of the model associated with $\beta_1(t)$ and where $v_{1i}(t)$ and $\epsilon_i(t)$ are random errors. This model corresponds to the RVCM given in (3) where

$$\boldsymbol{\beta}(t) = [\beta_0(t), \beta_1(t)]^T,\ \boldsymbol{v}_i(t) = [v_{1i}(t)],\ \boldsymbol{x}_i(t) = [x_{i0}(t), x_{i1}(t)]^T,\ \boldsymbol{z}_i(t) = [z_{i1}(t)]$$

with $x_{i0}(t) \equiv 1$ and $z_{i1}(t) \equiv x_{i1}(t)$. Note that in the simulated model $\mathbf{R}_i$ is a diagonal matrix and $\mathbf{D}$ is an unstructured covariance matrix. The observations between subjects are simulated independent.

It is assumed that $\sigma_1^2 = \sigma_2^2 = \sigma_\epsilon^2 = \sigma^2$. Then, the correlation coefficient between repeated measurements within each experimental unit is

$$\rho = \mathbb{C}orr[y_i(t), y_i(s))] = \frac{\sigma_0^2 + \sigma^2\cos[2\pi(t-s)]}{\sigma_0^2 + 2\sigma^2},\ \text{for } s \neq t$$

therefore

$$\frac{\sigma_0^2 - \sigma^2}{\sigma_0^2 + 2\sigma^2} \leq \rho_y \leq \frac{\sigma_0^2 + \sigma^2}{\sigma_0^2 + 2\sigma^2}$$

To simulate different intensities of correlation are considered three cases:

Case 1 In which $\sigma_1^2 = \sigma_2^2 = \sigma_\epsilon^2 = \sigma^2 = 0.01$ and $\sigma_0^2 = 0.01$. This corresponds to $\rho_y \in (\ 0\ ,\ 0.67\ )$.

Case 2 In which $\sigma_1^2 = \sigma_2^2 = \sigma_\epsilon^2 = \sigma^2 = 0.01$ and $\sigma_0^2 = 0.04$. This corresponds to $\rho_y \in (\ 0.50\ ,\ 0.83\ )$.

Case 3 In which $\sigma_1^2 = \sigma_2^2 = \sigma_\epsilon^2 = \sigma^2 = 0.01$ and $\sigma_0^2 = 0.09$. This corresponds to $\rho_y \in (\ 0.73\ ,\ 0.91\ )$.

Design times are simulated in accordance with the expression

$$t_{ij} = j/(m+1),\ i = 1, \ldots, n,\ j = 1, \ldots, m$$

where m is a positive integer. To simulate unbalanced data sets, a main characteristic of the structure of longitudinal data, in each subject are removed randomly repeated measures with a rate $r_m = 30\%$. Thus, there is approximately $m(1 - r_m)$ repeated measurements per experimental unit and $nm(1 - r_m)$ measurements in

Table 1: Scenarios for the simulation study.

Scenario	n	m	σ_0^2	Scenario	n	m	σ_0^2	Scenario	n	m	σ_0^2
1	5	5	0.01	10	10	5	0.01	19	20	5	0.01
2	5	5	0.04	11	10	5	0.04	20	20	5	0.04
3	5	5	0.09	12	10	5	0.09	21	20	5	0.09
4	5	10	0.01	13	10	10	0.01	22	20	10	0.01
5	5	10	0.04	14	10	10	0.04	23	20	10	0.04
6	5	10	0.09	15	10	10	0.09	24	20	10	0.09
7	5	15	0.01	16	10	15	0.01	25	20	15	0.01
8	5	15	0.04	17	10	15	0.04	26	20	15	0.04
9	5	15	0.09	18	10	15	0.09	27	20	15	0.09

total. Smoothing parameters are chosen by using PCV. Table 1 contains all the scenarios considered in the simulation study.

Each scenario was repeated $N = 500$ times and each time was calculated $AMSE(\beta_0)$ and $AMSE(\beta_1)$, in order to compare the relative performance of the Local Polynomial Kernel Regression Estimation (LPKE) with Radial Kernel Functions Estimation (RKFE). For these estimations the next indicators are define

$$AMSER_{(RKFE/LPKE)} = \frac{1}{N}\sum_{k=1}^{N}\frac{AMSE_k(\kappa, LPKE)}{AMSE_k(\kappa, RKFE)} \times 100\% \tag{25}$$

and

$$AMSERKF_{(RKFE/LPKE)} = \frac{\sum_{k=1}^{N} I_{\{AMSE_k(\kappa,LPKE)>AMSE_k(\kappa,RKFE)\}}}{N} \times 100\% \tag{26}$$

where $AMSE_k(\kappa, LPKE)$ and $AMSE_k(\kappa, RKFE)$ denote the value of $AMSE(\kappa)$ obtained in the k-th simulation replicate, $k = 1, \ldots, N$, by using the RKFE and the LPKE respectively, and I_A denotes the indicator function of set A. $AMSER$ represents the average relative efficiency associated with the N replications and $AMSERKF$ is the percentage of estimations obtained through RKF that are better than those obtained through LPK in relation to the AMSE in AMSE in the N replications. If $AMSER \approx 100\%$ and $AMSERKF \approx 50\%$, LPKE and the RKFE perform similarly; if $AMSER > 100\%$ y $AMSERKF > 50\%$, RKFE has better performance than LPKE; and if $AMSER < 100\%$ and $AMSERKF < 50\%$, LPKE has better performance than RKFE.

Table 2 contains the results of the simulation. According to this table, the choice rules of an alternative estimation by using indicators (25) and (26), and Tables 3 and 4 which summarizes the results, it follows that at 48% of cases the best estimation strategy is the RKFE; by approximation to the rules given, that is, following the criteria $AMSER_0 \approx 100\%$ y $AMSERKF_0 \approx 50\%$, it has that in the 35.2% of the situations the two strategies behave similarly; furthermore, just 9.3% of cases the best strategy is LPKE and for 7.4% of the scenarios the criterion does not decide ($AMSER > 100\%$ and $AMSERKF < 50\%$, or, $AMSER < 100\%$

TABLE 2: Simulation results.

		$n=5$			$n=10$			$n=20$		
		$m=5$	$m=10$	$m=15$	$m=5$	$m=10$	$m=15$	$m=5$	$m=10$	$m=15$
$\sigma_0^2=0.01$	$AMSER_0$	100.0%	100.2%	101.1%	100.9%	100.0%	100.0%	101.0%	102.0%	100.8%
	$AMSERKF_0$	49.9%	50.2%	50.8%	50.8%	49.2%	48.4%	50.9%	46.0%	62.0%
	$AMSER_1$	100.5%	102.4%	101.4%	100.0%	101.5%	100.9%	99.6%	100.1%	100.4%
	$AMSERKF_1$	45.8%	51.2%	50.8%	48.1%	51.0%	50.4%	43.8%	61.9%	56.4%
$\sigma_0^2=0.04$	$AMSER_0$	100.1%	100.0%	100.0%	100.1%	101.0%	100.0%	100.0%	102.1%	100.5%
	$AMSERKF_0$	50.7%	49.4%	45.8%	47.8%	45.9%	49.9%	49.0%	63.1%	51.8%
	$AMSER_1$	99.6%	102.2%	100.7%	100.0%	100.8%	100.5%	100.7%	101.3%	100.4%
	$AMSERKF_1$	42.0%	52.7 %	48.1%	46.9%	50.1%	54.9%	50.2%	49.9%	52.8%
$\sigma_0^2=0.09$	$AMSER_0$	100.9%	100.5%	100.1%	100.0%	100.0%	100.3%	100.0%	100.7%	99.9%
	$AMSERKF_0$	50.5%	50.4%	51.1%	47.7%	49.7 %	51.0%	47.7%	48.6%	46.8%
	$AMSER_1$	99.3%	101.8%	101.6%	100.5%	101.3%	100.9%	99.5 %	102.0%	100.2 %
	$AMSERKF_1$	44.5%	49.0%	49.3%	49.6%	43.6%	51.6%	46.7%	52.0%	51.4%

and $AMSERKF > 50\%$). It is also noted that the strategy most appropriate for estimating, considering $\beta_0(t)$ and $\beta_1(t)$ simultaneously, is type $RKFE$ which corresponds to $n=5$, $m=10$ and $\sigma_0^2=0.01$, $n=5$, $m=15$ and $\sigma_0^2=0.01$, $n=10$, $m=15$ and $\sigma_0^2=0.09$, $n=10$, $m=15$ and $\sigma_0^2=0.01$, and $n=10$, $m=15$ and $\sigma_0^2=0.04$; there is no case where LPKE improved the outcomes for both dynamic components simultaneously. Furthermore, there are a variety of cases where the best strategy is RKFE for one of the dynamic parameters and for the other two strategies perform similarly.

According to Table 3, it is concluded that while the value of σ_0^2 decreases, and at the same time the correlation between repeated measurements, the proportion of times that the best strategy is RKFE increase. Moreover, in all degrees of correlation, the proportion of times that performs better RKFE is superior compared to the proportion for LPKE. In the same way, in all degrees of correlation, the proportion of times where the two strategies perform similarly is higher than the proportion where LPKE is the best option. Also, these relationships are maintained in each case for the dynamic intercept and the dynamic slope. Therefore, with any degree correlation and any dynamic parameter, in 83.3% of cases, RKFE performs better or similarly than LPKE. Thus, it is concluded that in such circumstances, to choose RKFE is the best alternative.

TABLE 3: Proportion of times that a strategy is better than another for σ_0^2 and $\beta_r(t)$.

		RKF	Equal	LPK	No
$\sigma_0^2=0.01$	$\beta_0(t)$	9.3%	5.6%	0.0%	1.9%
	$\beta_1(t)$	11.1%	1.9%	1.9%	1.9%
Total		20.4%	7.4%	1.9%	3.7%
$\sigma_0^2=0.04$	$\beta_0(t)$	5.6%	9.3%	0.0%	1.9%
	$\beta_1(t)$	9.3%	5.6%	1.9%	0.0%
Total		14.8%	14.8%	1.9%	1.9%
$\sigma_0^2=0.09$	$\beta_0(t)$	7.4%	7.4%	1.9%	0.0%
	$\beta_1(t)$	5.6%	5.6%	3.7%	1.9%
Total		13.0%	13.0%	5.6%	1.9%
Total general		48.1%	35.2%	9.3%	7.4%

Table 4: Proportion of times that a strategy is better than another for n and m.

		RKF	Equal	LPK	No
	$m = 5$	3.7%	1.9%	3.7%	1.9%
$n = 5$	$m = 10$	7.4%	3.7%	0.0%	0.0%
	$m = 15$	5.6%	5.6%	0.0%	0.0%
Total		16.7%	11.1%	3.7%	1.9%
	$m = 5$	1.9%	9.3%	0.0%	0.0%
$n = 10$	$m = 10$	3.7%	3.7%	0.0%	3.7%
	$m = 15$	7.4%	3.7%	0.0%	0.0%
Total		13.0%	16.7%	0.0%	3.7%
	$m = 5$	3.7%	3.7%	3.7%	0.0%
$n = 20$	$m = 10$	5.6%	3.7%	0.0%	1.9%
	$m = 15$	9.3%	0.0%	1.9%	0.0%
Total		18.5%	7.4%	5.6%	1.9%
Total		48.1%	35.2%	9.3%	7.4%

Furthermore, according to Table 4, for all sample sizes, when the number of repeated measurements of each individual increases, the proportion of scenarios where RKFE performs better RKFE increases as well. It must also be noted that this proportion is similar for all sample sizes, and is always significantly higher than the proportion where LPKE is the best option. Moreover, when $n = 10$ is notorious the proportion of times where the two strategies perform similarly. Finally, it is observed the fact that the proportion of times where LPKE is better is equal to 0.0% in most cases for any value of n y m. Thus, it is concluded that the proposed methodology is the best option a high percentage of times in all simulated scenarios, or at least performs similarly to the LPKE, which very rarely performs better than the RKFE.

6. Application

The viral load (*plasma VIH RNA copies/mL*) and cell count CD4+ are currently key indicators to assess AIDS treatments in clinical research. Initially it was considered the CD4+ cell count as a primary indicator of AIDS immunodeficiency, but it was newly found that viral load is more predictive for clinical outcomes. However, recently some researchers have suggested that a combination of these two indicators may be more appropriate to evaluate the treatment of HIV and AIDS. Therefore it is pertinent to study the relationship between viral load and CD4+ cell count during treatment (Liang et al. 2003).

Figure 2 presents some graphs of a linear regression of viral load (log(RNA)) against to CD4+ cell counts in some measuring times of a clinical study of AIDS (ACTG 315). In this investigation, there are 46 infected patients with an antiviral therapy consisting of *ritonavir*, 3TC and AZT. After starting treatment, viral load and CD4+ cell count were observed simultaneously at days 0, 2, 7, 10, 14, 28, 56, 84, 168, and 336. The number of repeated measurements for individual varies from 4 to 10 and in total 361 observations were obtained.

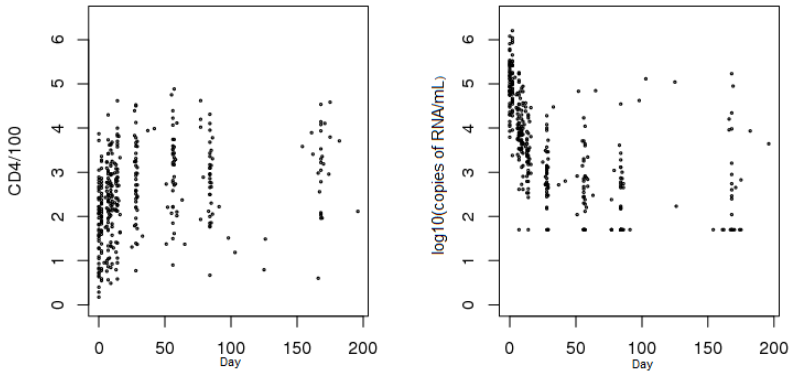

FIGURE 1: Scatter plot for CD4+ cell count and viral load.

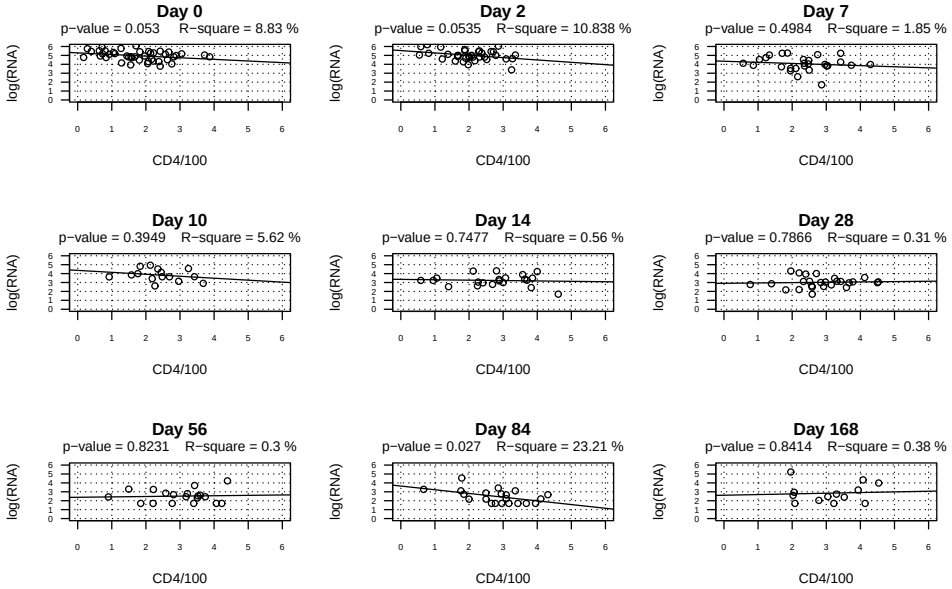

FIGURE 2: Graphs related with the linear regression of viral load ($\log_{10}$(RNA)) against CD4+ cell count in some measuring times. The model adjusted in each case has got the form $\log_{10}(RNA) = \beta_0 + \beta_1(\text{CD4}/100) + e$. The p-value corresponding to $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ is also presented in each case.

In general, it appears that the virologic (measured by the viral load) and the immune response (measured by the CD4+ cell count) of the patient are negatively correlated, and that their relationship is approximately linear during antiviral therapy. Figure 1 shows the scatter plots associated with CD4+ cell count and viral load. The logarithm of viral load is used to stabilize the variance for the estimation procedures of the model fitted in the following.

Figure 2 shows that the slope of the linear regression of viral load versus CD4+ cell count changes over time because in a few days the slope is significantly different from zero and in others not. This motivates the fitting of a model with dynamic coefficients in order to describe and quantify the change in the relationship. However, because it may be of interest to investigate the relationship between viral load and CD4+ cell count in a particular patient, the fitting of a RVCM is needed.

The ACTG 315 data set has been studied extensively by Liang et al. (2003), who showed a strong inverse relationship between viral load and CD4+ cell count. In this section, a RVCM is fitted to investigate the dynamic relationship between viral load (in logarithmic scale) and CD4+ cell count, and also to describe this relationship particularly in any patient.

The RVCM fitted is

$$y_{ij} = \beta_0(t_{ij}) + \beta_{1i}(t_{ij})x_{i1}(t_{ij}) + e_{ij},\ j = 1, \ldots, n_i,\ i = 1, \ldots, 46 \qquad (27)$$

where y_{ij}, $x_{i1}(t_{ij})$, and e_{ij} are viral load (in logarithmic scale), the CD4+ cell count and the error associated with the j-th measurement of the i-th patient, respectively, $\beta_0(t)$ is the dynamic coefficient associated with the intercept and $\beta_{1i}(t)$ is the dynamic and random coefficient associated with the CD4+ cell count. This parameter is given by

$$\beta_{1i}(t) = \beta_1(t) + v_i(t),\ i = 1, \ldots, 46$$

with $\beta_1(t)$ the coefficient associated with the mean dynamic relationship between viral load and cell count CD4+ and $v_i(t)$ the coefficient related to the characteristics of the i-th patient that differ from the average behavior.

The dynamic components of the model are estimated through LPK and the proposed methodology by using RKF. The kernel functions used in the estimation are Gaussian, and for selecting the smoothing parameters (bandwidths) the PCV is implemented which gives the bandwidths $h_{RKF} = 0.999$ and $h_{RKF} = 0.401$ using RKF and LPK respectively (Figure 3). Furthermore, models (8) and (20) are fitted by using function `lme4` (Bates, Maechler & Bolker 2011) in `R` (R Development Core Team 2008).

Figure 4 shows the residuals of the RVCM fitted. It is observed that in both cases, the RVCM has a good fit to the data. The value of the residuals by using both estimation methods are similar prior 150-th day. From that day the value of the residuals is less by using LPK, suggesting that the relationship at the end of the treatment by using LPK is more accurate; however, both techniques indicate the same at the end of treatment as it is evidenced in Figure 5 where are illustrated the graphs associated with the estimation of $\beta_0(t)$ and of $\beta_1(t)$ by using LPK and RKF, respectively. In both cases, the graphics are very similar to those obtained by Liang et al. (2003).The right chart shows that the dynamic relationship between viral load and CD4+ cell count is approximately direct to day 50, point at which the association is weak; from this day the relationship between the indicators is inverse to the end of treatment. Moreover, between week 1 and 14, RKF estimate suggests that the relationship is apparently stronger. Also, major differences between the estimation methodologies from day 150 of treatment are noted, where the estimate

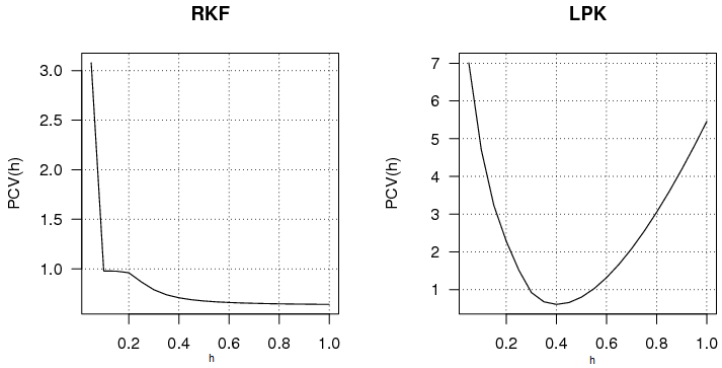

FIGURE 3: Graphs related with PCV against the bandwidth.

through LPK suggests that the relationship changes and it is strengthened –in an inverse way– to the end of treatment. Overall, the dynamic relationship between viral load and CD4+ cell count decreases gradually until the seventh week of the study where the relationship begins to strengthen gradually until the end of treatment.

One advantage of fitting a RVCM is that it is possible to characterize the performance of the dynamic relationship of interest for any particular subject. Figure 6 shows the estimates of the deviations typical of the population $v_i(t)$ for patients 1, 3 y 16 using RKF and LPK. Not only the magnitude but also the direction of changes can be see among individuals. Due to the high variation within each of the individuals, the estimation of the relationship between the indicators for each patient is very important because it allows to customize the treatment and care of each patient. Using LPK more variability between individuals in the dynamic relationship of viral load and CD4+ cell count is perceived. It is observed how the relationship may even be direct. While using RKF variability is lower and the pattern is very similar to the average dynamic relationship.

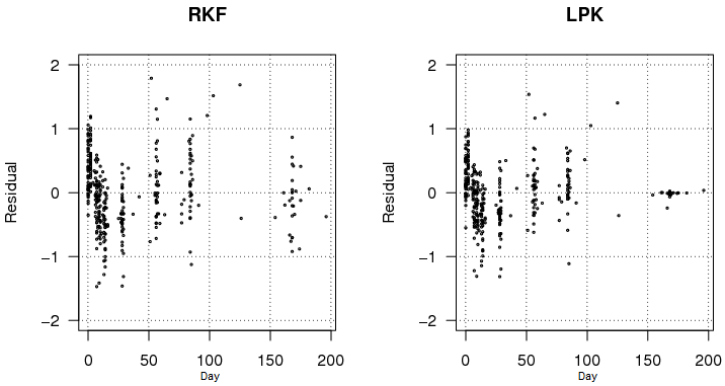

FIGURE 4: Residuals of RVCM fitted by using RKF and LPK.

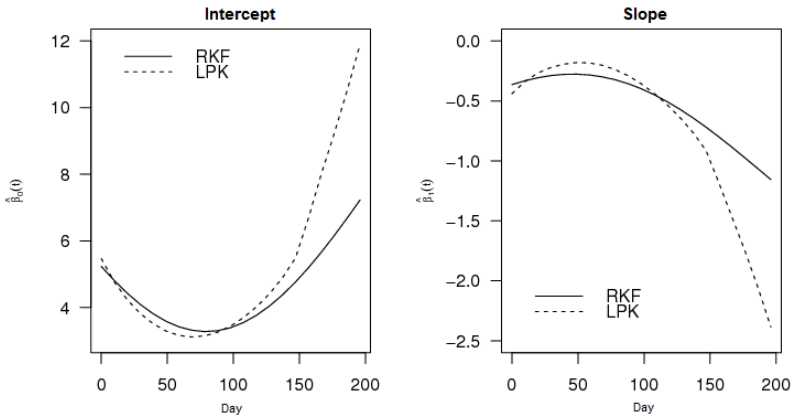

FIGURE 5: Graphs associated with the estimation of $\beta_0(t)$ and $\beta_1(t)$ for the RVCM fitted by using RKF and LPK.

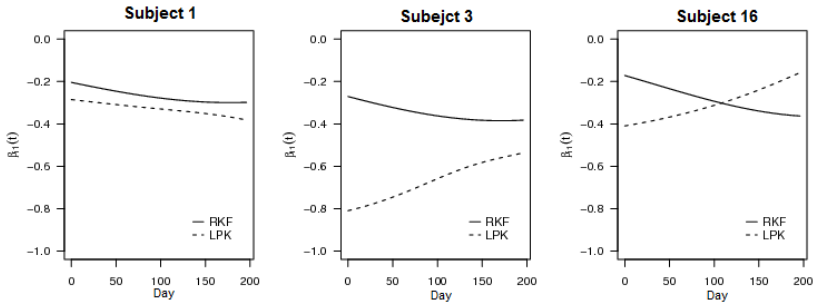

FIGURE 6: Graphs associated with the estimation of $v_i(t)$ for the RVCM fitted by using RKF and LPK for patients 1, 3 and 16.

7. Discussion and Conclusions

This paper proposes a methodology to estimate the coefficients of a random time-varying coefficient model through radial kernel functions, where model coefficients are approximated by a linear combination of kernel functions which centered around all the measuring points, or their quantiles, weighted by a bandwidth that may change or not among coefficients (Hastie, Tibshirani & Friedman 1990).

By means of a simulation study the estimation method is compared by using a local polynomial kernel regression with the use of radial kernel functions in relation with the average mean square error, resulting that the proposed methodology is the best one in a high percentage of times in all simulated scenarios, or at least performs similarly to the LPKE, who rarely performs better than the RKFE, in relation with the average mean square error.

Analyzing the ACTG 315 data set (Liang et al. 2003), it was found that the relationship between viral load and CD4+ cell count is inverse. Furthermore, as a

future alternative modeling, it can be thought a model in which the response variable is bivariate, consisting of viral load and CD4+ cell count, and the predicted correspond to some covariates related to the treatment of patients with AIDS.

Further studies may investigate the consistency and asymptotic properties of the estimators proposed, the impact of the functional form of the dynamic coefficients of the model and mechanisms for testing hypotheses related to both the dynamic and random coefficients model.

[Recibido: abril de 2010 — Aceptado: febrero de 2012]

References

Altman, N. S. (1990), 'Kernel smoothing of data with correlated errors', *Journal of the American Statistical Association* **85**(411), 749–759.

Bates, D., Maechler, M. & Bolker, B. (2011), *lme4: Linear Mixed-Effects Models Using S4 classes*. R package version 0.999375-42.
*http://CRAN.R-project.org/package=lme4

Davis, C. S. (2000), *Statistical Methods for the Analysis of Repeated Measurements*, Springer.

Diggle, P. J., Liang, K. Y. & Zeger, S. L. (1994), *Analysis of Longitudinal Data*, Oxford University Press.

Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (2009), *Longitudinal Data Analysis*, Champman & Hall.

Guo, W. (2002), 'Functional mixed effects models', *Biometrics* **58**(1), 121–128.

Hart, J. D. (1991), 'Kernel regression estimation with time series errors', *Journal of the Royal Statistical Society. Series B (Methodological)* **53**(1), 173–187.

Hart, J. D. & Wehrly, T. E. (1986), 'Kernel regression estimation using repeated measurements data', *Journal of the American Statistical Association* **81**(396), 1080–1088.

Hastie, T., Tibshirani, R. & Friedman, J. (1990), *The elements of Statistical Learning*, Springer.

Hoover, D. R., Rice, J. A., Wu, C. O. & Yang, L. P. (1998), 'Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data', *Biometrika* **85**(4), 809–822.

Huang, J. Z., Wu, C. O. & Zhou, L. (2002), 'Varying-coefficient models and basis function approximations for the analysis of repeated measurements', *Biometrika* **89**(1), 111–128.

Liang, H., Wu, H. & Carroll, R. J. (2003), 'The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error', *Biostatistics* **4**(2), 297–312.

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org

Rice, J. A. & Silverman, B. W. (1991), 'Estimating the mean and covariance structure nonparametrically when the data are curves', *Journal of the Royal Statistical Society. Series B (Methodological)* **53**(1), 233–243.

Sosa, J. C. & Díaz, L. G. (2010), 'Estimación de las componentes de un modelo de coeficientes dinámicos mediante las ecuaciones de estimación generalizadas', *Revista Colombiana de Estadística* **33**(1), 89–109.

Verbeke, G. & Molenberghs, G. (2005), *Models for Discrete Longitudinal Data*, Springer.

Wu, H. & Liang, H. (2004), 'Backing random varying-coefficient models with time-dependent smoothing covariates', *Scandinavian Journal of Statistics* **31**, 3–19.

Wu, H. & Zhang, J. T. (2006), *Nonparametric Regression Methods for Longitudinal Data Analysis*, Wiley.

Zeger, S. L. & Diggle, P. J. (1994), 'Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters', *Biometrics* **50**(3), 689–699.