Universidad de Valparaíso
CHILE
Instituto de Estadística
Facultad de Ciencias

Partially varying-coefficient generalized linear model

Final thesis presented by:
Valeria Solange Lira Lira

To apply for the degree of:
Master in Statistics

Thesis supervisor:
Dr. Germán Ibacache Pulgar

Valparaíso, 2019

Agradecimientos

Este trabajo de tesis realizado en la Universidad de Valparaíso es un esfuerzo en el cual, directa o indirectamente participaron distintas personas opinando, corrigiendo, teniéndome paciencia, dándome ánimo, acompañándome en los momentos de crisis y en los momentos de felicidad, a todos ellos deseo agradecer en este apartado.

En primer lugar, le agradezco a Dios por haberme acompañado y guiado a lo largo de mi carrera, por ser mi fortaleza en los momentos de debilidad y por brindarme una vida llena de aprendizajes, experiencias y sobre todo felicidad.

Agradezco a mis padres por su apoyo incondicional, por haberme dado la oportunidad de tener una excelente educación durante toda mi vida, por su amor y sacrificio en todos estos años, gracias a ustedes he logrado llegar hasta aquí y convertirme en la persona que soy.

A mis hermanos por su cariño, apoyo y por acompañarme en cada paso de este largo proceso, muchas gracias.

Agradezco a mi familia de Valparaíso, a mis tíos y primas por abrirme la puerta de su hogar y recibirme como una hija más, por su afecto, por la paciencia, por incentivarme a seguir adelante y por estar siempre ahí, en mis momentos más difíciles y en los de felicidad, gracias por todo su apoyo.

Agradezco a mi tía Flor y mi abuelita que a pesar de la distancia que nos separa siempre estuvieron apoyándome.

Al profesor Germán Ibacache, un especial agradecimiento por sus consejos, su sabiduría, su paciencia, su apoyo y ánimo que me brindo durante mi estancia en la Universidad. Gracias por su motivación a que continuara estudiando y creer en mí, siempre estaré agradecida por toda su ayuda.

A los profesores del Instituto de Estadística, por haberme brindado oportunidades y enriquecerme en conocimiento.

No puedo dejar de agradecer especialmente a Valeria Meneses, mi compañera y amiga fiel de la universidad. Gracias por tu amistad, cariño y apoyo durante todos estos años, no sé qué hubiera sido de mi vida sin ti.

A la Bárbara Arredondo, una de las primeras personas que conocí en la universidad, que hasta el día de hoy es mi amiga, gracias por tu amistad, tu lealtad, por tu cariño y por no dejarme caer en los momentos más difíciles.

Finalmente quiero dedicar en este trabajo de tesis a todos mis compañeros y amigos de la universidad, gracias por toda su ayuda y amistad.

Contents

List of figures

List of tables

Abbreviations

PVCGLM	Partially varying-coefficient generalized linear model
PVCM	Partially varying-coefficient model
MPLE	Maximun penalized likelihood estimate
GLM	Genelarized linear model
MLE	Maximun likelihood estimate
SEB	Standard error band
df	Degree of freedom
LD	Likelihood displacement

Resumen

En este trabajo estudiamos los modelos lineales generalizados con coeficiente variando parcialmente (MLGCVP), se propuso un método de estimación para los parámetros asociados al MLGCVP y se realizó inferencia sobre esos parámetros. Se desarrolló la técnica de influencia local para evaluar la sensibilidad de los estimadores de máxima verosimilitud penalizada, para detectar observaciones influyentes. Finalmente, para ilustrar esta clase de modelos y de los resultados que se obtuvieron en la ejecución de este trabajo, se implementó computacionalmente las metodologías para el proceso de estimación y la técnica de influencia local, y se aplicó a un conjunto de datos reales, utilizando el software MATLAB.

Abstract

In this work we study the partially varying-coefficients generalized linear models (PVCGLM). A Backfitting algorithm to attain the maximum penalized likelihood estimates (MPLE) by using natural cubic smoothing splines is presented. In particular, the score functions and Fisher information matrices for the parameters of interest are expressed in a similar notation of that used in parametric generalized linear models. In order to study the sensitivity of the penalized estimates under some usual perturbation schemes in the model or data, the local influence curvatures are derived and some diagnostic graphic are proposed. Finally, a practical application that employ real data is presented and discussed.

Keywords: Generalized lineal model, Varying-coefficient model, Maximum penalized likelihood estimates, Weighted back-fitting algorithm, Local influence measure.

Objectives

General objective

Study the estimation and statistical inference problem in PVCGLM, and apply the local influence method to asses the effect of small perturbations in the model (or data) on the maximum penalized likelihood estimates.

Specific objectives

1. Study the theory of the PVCGLM.
2. Derive an iterative process for obtain the maximum penalized likelihood estimates in the PVCGLM.
3. Developed the method of local influence for the PVCGLM.
4. Illustrate the methodology with real data sets.

Chapter 1

Introduction

In this work, we study partially varying-coefficient generalized linear models, which are an extension of generalized linear models. This class of models emerge as a powerful tool in statistical modeling because of its flexibility to model explanatory variables effects that can contribute parametric way and explanatory variables effects in which the coefficients are allowed to vary as smooth functions of other explanatory variables (for example, time variable). These models are often used in research related to longitudinal, clustered, spatial and hierarchical sampling schemes.

In the last years various works on partially varying-coefficient models were developed, principally associated to the estimation problem. For instance, in the context of varying-coefficient models (VCMs), Hastie and Tibshirani (1993) estimated the parameters of the model based on penalized least squares criterion. In particular, they used spline smoothing and backfitting algorithm during the estimation process (see also, Buja et al., 1989). Wu and Yu (2002) studied estimation in nonparametric varying-coefficients model (NVCM) for the analysis of longitudinal data using local least squares, smoothing splines and smoothing via basis approximations. Eubank et al. (2004) discussed the problems of point and interval estimates for VCMs based spline smoothing methods. Fan and Huang (2005) and Zhang et al. (2007) used profile least-squares technique for estimating the parametric component of the PVCM. Wang et al. (2004) proposed an estimation procedures for the VCM based on local ranks, which is highly efficient and robust alternative to the local linear least method. Krafty et al. (2008) developed an estimation procedure for the VCM when the within-subject covariance unknown. Ibrahim et al. (2005) proposed a general series method to estimate a PVCM based in spline and power series. Cai et al. (2000) used local polinomial regression techniques to estimate coefficients functions and derived some standard error formulas for estimated coefficients. On the other hand, diagnostic methods have been well developed for various types of models. For instance, the local influence methodology Cook (1986) has been applied by Beckman et al. (1987) to detect influential observations in normal linear mixed models with emphasis on single observations whereas Lesaffre and Verbeke (1998) extended the approach for normal linear mixed models in a repeated-measurement context and under the case-weight perturbation scheme. Under elliptical errors, Galea et al. (1997) and Liu (2002) applied the local influence approach in multivariate elliptical linear models under various perturbation schemes and Osorio et al. (2007) extended the methodology to the longitudinal structure, which includes the mixed-effect case. In nonparametric and semiparametric regression models, Thomas (1991) constructed local influence diagnostics for the smoothing parameter and Zhu et al. (2003) extended the works by Cook (1986) to provide local influence measures under different perturbation schemes in normal partially linear models. Recently, Ibacache and Paula (2011) obtained similar evidence on the robust aspects of the maximum penalized likelihood estimates from Student-t partial linear models against outlying observations by using local influence measures.

Nevertheless, in partially varying-coefficient generalized linear models local influence diagnostic studies are quite rare. However, some works on generalized linear models were developed. For instance, Zhu and Lee (2003) develop the local influence method on the basis of a Q-function which is associated with the conditional expectation of the complete-data log-likelihood function in the EM algorithm under generalized linear mixed models. Chen et al. (2010) developed a perturbation selection method and a second-order local influence measure to address this issue and conduct local influence analysis in generalized linear mixed models.

The work is structured as follows: Chapter 2, the partially varying-coefficient generalized linear models are presented and a penalized log-likelihood function is considered for the estimation parameter.

In Chapter 3, the problem of estimation and inferential analysis of the parameters are discussed. A discussion on the process to obtain maximum penalized likelihood estimators, the derivation of a back-fitting algorithm, some inferential result and discussions on degrees of freedom (df) estimation and selection of the smoothing parameter.

In Chapter 4, the method of local influence is presented and normal curvatures for some perturbation schemes are derived.

Finally, in Chapter 5, an illustration of the methodology is presented for real dataset.

Chapter 2

The partially varying-coefficient generalized linear model

The partially varying-coefficients generalized linear models (PVCGLMs) emerge as a powerful tool in statistical modeling because of its flexibility to model explanatory variables effects that can contribute parametric way (constant coefficients) and other that varying with a given factor. These models are often used in research related to longitudinal, clustered and spatial sampling schemes.

2.1 The model

Consider a data set that is composed of a response y_{ij}, and vectors of covariates $\mathbf{w}_{ij}$ $(p \times 1)$ and $\mathbf{x}_{ij}$ $(s \times 1)$, where $j = 1, \ldots, n_i$ represents an observation within the cluster $i = 1, \ldots, n$. It is assumed that the random variable y_{ij} follows an exponential family distribution of the form

$$f_{\mathrm{y}}(y_{ij};\theta_{ij},\phi) = \exp\left[\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{a_{ij}(\phi)} + c(y_{ij},\phi)\right], \tag{2.1}$$

where θ_{ij} is the canonical form of the location parameter and is a function of the mean μ_{ij}, $a_{ij}(\phi)$ is a known function of the possibly unknown dispersion parameter or vector of dispersion parameter ϕ, c is a function of the dispersion parameter and the responses, and ψ is a known function, such that the mean and variance of y_{ij} are equals to $\mu_{ij} = \mathrm{E}(\mathrm{y}_{ij}) = \partial\psi(\theta_{ij})/\partial\theta_{ij}$ and $\mathrm{Var}(\mathrm{y}_{ij}) = \phi\,\mathrm{V}_{ij}$, with $\mathrm{V}_{ij} = \mathrm{V}(\mu_{ij}) = \partial^2\psi_{ij}/\partial\theta_{ij}^2$, respectively. The partially varying-coefficient generalized linear model is defined by (2.1) and the following systematic component:

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{w}_{ij}^T\boldsymbol{\alpha} + \sum_{k=1}^{s} \mathrm{x}_{ij}^{(k)}\,\beta_k(\mathrm{t}_{ij}) \tag{2.2}$$

where $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_p)$ is a vector consisting of the regression coefficients, $\beta_k(\cdot)$ ($k = 1,...,s$) are unknown smooth arbitrary functions of t, associated with the covariates $\mathrm{x}_{ij}^{(k)}$. In order to write model (2.2) in a matrix form, we obtain

$$\boldsymbol{\eta}_i = \mathbf{W}_i\boldsymbol{\alpha} + \sum_{k=1}^{s}\widetilde{\mathbf{N}}_{ki}\boldsymbol{\beta}_k, \tag{2.3}$$

where $\mathbf{W}_i = \mathrm{diag}(\mathbf{w}_{i1}^T,\ldots,\mathbf{w}_{in_i}^T)$, $\widetilde{\mathbf{N}}_{ki} = \mathbf{X}_i^{(k)}\mathbf{N}_{ki}$, $\mathbf{X}_i^{(k)} = \mathrm{diag}(\mathrm{x}_{i1}^{(k)},\ldots,\mathrm{x}_{in_i}^{(k)})$, $\mathbf{N}_{ki}$ is an $(m_i \times r)$ incidence matrix with the (j,l)th element equal to the indicator $I(\mathrm{t}_{ij} = \mathrm{t}_l^0)$ $(j = 1,\ldots,m_i)$, where t_l^0 $(l = 1,\ldots,r\ )$ denotes the distinct and ordered values of the explanatory variable t_{ij}, and $\boldsymbol{\beta}_k = (\psi_{k_1},\ldots,\psi_{k_r})^T$ is a $(r\times 1)$ vector of parameters with $\psi_{k_l} = \beta_k(\mathrm{t}_l^0)$, for $l = 1,\ldots,r$.

In Table 2.1 we present the link functions associated with some of the major distributions that belong to the exponential family and Table 2.2 the components of an exponential family (see McGullar, P. and Nelder (1989), and Nelder and Wedderburn (1972)).

Table 2.1: Link functions for some discrete and continuous distributions.

Distributions	Link functions $g(\mu)$
Binomial	$\log(\mu/(n-\mu))$
Poisson	$\log \mu$
Normal	μ
Gamma	$1/\mu$
Inverse Normal	$1/\mu^2$

Table 2.2: Components of an exponential family for some discrete and continuous distributions.

Distributions	θ	$\psi(\theta)$	$c(y, \phi)$	ϕ
Binomial	$\log(\mu/(n-\mu))$	$n \log(1+e^{\theta})$	$\log \binom{n}{y}$	1
Poisson	$\log \mu$	e^{θ}	$-\log(y!)$	1
Normal	μ	$\theta^2/2$	$-\frac{1}{2}\left(\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2)\right)$	σ^2
Gamma	$1/\mu$	$-\log(-\theta)$	$-\log\Gamma(\nu) + \phi\log(\phi) + (\phi-1)\log y$	ν^{-1}
Inverse Normal	$1/\mu^2$	$-(-2\theta)^{1/2}$	$\frac{1}{\theta y} - \frac{1}{2}\log(-\pi\phi y^3)$	σ^2

2.2 The Poisson distribution

The Poisson distribution can be written as a special case of an exponential family distribution. Its probability function is

$$f_{\mathrm{y}}(y,\mu) = \frac{1}{y!}e^{-\mu}\mu^{y} \qquad (y = 0, 1, 2, \ldots).$$

Rewriting this,

$$f_{\mathrm{y}}(y_{ij};\theta_{ij},\phi) = \exp\left(y_{ij}\log\mu_{ij} - \mu_{ij} - \log(y_{ij}!)\right),$$

hence $\theta_{ij} = \log(\mu_{ij})$ the natural parameter. Reciprocally, we can write the mean in terms of the natural parameter as $\mu_{ij} = e^{\theta_{ij}}$ and $\psi(\theta_{ij}) = \mu_{ij} = e^{\theta_{ij}}$. Again there is no dispersion parameter, so we can $a_{ij}(\phi) = 1$. Finally, $c(y_{ij},\phi) = -\log(y_{ij}!)$.

Model the mean using a link function: $\mu_{ij} = g^{-1}(\eta_{ij}) = e^{\eta_{ij}}$ and $\eta_{ij} = g(\mu_{ij}) = \log(\mu_{ij})$.

2.3 Penalized function

Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_s^T, \phi)^T \in \boldsymbol{\Theta} \subseteq \mathcal{R}^{p^*}$, with $p^* = p + r + 1$, the vector of unknown parameters associated to model (2.1). Then, the log-likelihood function is given by

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}), \tag{2.4}$$

where

$$L_i(\boldsymbol{\theta}) = \left[\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{a_{ij}(\phi)} + c(y_{ij},\phi)\right]. \tag{2.5}$$

In general, it is known fact that maximizing the log-likelihood function in the context semiparametric, without imposing restrictions over the nonparametric functions, may cause overfitting and non identification of parameters vector $\boldsymbol{\alpha}$. A well known procedure that can solve this problem is based on the idea of log-likelihood penalization and consists in incorporating a penalty function over each function β_k such that

$$L_{\mathrm{p}}(\boldsymbol{\theta}, \lambda_1, \ldots, \lambda_s) = L(\boldsymbol{\theta}) + \sum_{k=1}^{s} \lambda_k^* J(\beta_k), \tag{2.6}$$

where $J(\beta_k)$ denotes the penalty function over β_k and $\lambda_k^* = \lambda^*(\lambda_k)$ is a constant that depends on the parameter $\lambda_k \geq 0$. In this work, we will consider penalty functions of type

$$J(\beta_k) = \lambda^* \int_{a_k}^{b_k} [\beta_k^{(l)}(\mathrm{t}_k)]^2 \mathrm{dt}_k,$$

where $\beta_k^{(l)}(\mathrm{t}_k) = \frac{\mathrm{d}^l}{\mathrm{dt}^l}\beta(\mathrm{t})$, $\mathrm{t}_k^0 \in [a_k, b_k]$ and the functions β_k's belongs to the Sobolev functions space

$$\mathcal{W}_2^{(l)} = \{\beta_k : \beta_k, \beta_k^{(1)}, \ldots, \beta_k^{(l-1)} \text{abs. cont.}, \beta_k^{(l)} \in \mathcal{L}^2[a_k, b_k]\}.$$

When $l = 2$, the estimation of β_k leads to a smooth cubic spline with knots at the points $\mathrm{t}_{k_q}^0$, for $q = 1, \ldots, r_k$. According to Green and Silverman (1994), we may express the penalty function as

$$J(\beta_k) = \lambda_k^* \, \boldsymbol{\beta}_k^T \mathbf{K}_k \boldsymbol{\beta}_k,$$

where $\mathbf{K}_k$ is a $(r_k \times r_k)$ nonnegative definite smoothing matrix associated with the kth explanatory variable that depends only on the knots. Then, if we consider $\lambda_k^* = -\lambda_k/2$, the penalized log-likelihood function can be expressed as

$$L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = L_{\mathrm{p}}(\boldsymbol{\theta}) - \sum_{k=1}^{s} \frac{\lambda_k}{2} \, \boldsymbol{\beta}_k^T \mathbf{K}_k \boldsymbol{\beta}_k \,, \tag{2.7}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_s)^T$ denotes a $(s \times 1)$ vector of smoothing parameters that controls the tradeoff between goodness of fit and the smoothness estimated functions.

2.4 Final comments

In this chapter we present the partially varying-coefficient generalized linear model, tables with the link functions and the components for some exponential family distributions. We also write the Poisson distribution in the form of the exponential family.

Chapter 3

Parameter estimation and inference

In this chapter we discuss the estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$'s based on penalized log-likelihood. First, we present a maximization procedure to obtain the MPLEs and some sufficient conditions on the existence of the MPLEs in SAMs. Then, from the application of the maximization procedure we derive a back-fitting algorithm for obtaining the MPLEs. A conditional explicit solution is derived for the parameter estimates, which will be useful for obtaining some diagnostic quantities.

3.1 Score Function

Assuming that (2.7) is regular with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_k$'s and $\boldsymbol{\tau}$, the penalized score function of $\boldsymbol{\theta}$ is given by

$$\mathbf{U}_{\mathrm{P}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\partial L_{\mathrm{p}_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta}}\,.$$

The elements of the vector $\mathbf{U}_{\mathrm{p}}(\boldsymbol{\theta})$ are calculated below.

3.1.1 Parametric component

To obtain the function score for the $\boldsymbol{\alpha}$ parameter we must calculate

$$\begin{aligned}
\frac{\partial L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \alpha_j} &= \sum_{i=1}^{n} (\mathrm{a}_{ij}(\phi))^{-1} \left\{ \mathrm{y}_{ij} \frac{\partial \theta_{ij}}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial \alpha_j} - \frac{\partial \psi(\theta_{ij})}{\partial \theta_{ij}} \frac{\partial \theta_{ij}}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial \alpha_j} \right\} \\
&= \sum_{i=1}^{n} (\mathrm{a}_{ij}(\phi))^{-1} \left\{ \mathrm{y}_{ij} V_{ij}^{-1} \frac{\partial \mu_{ij}}{\partial \eta_{ij}} \mathrm{w}_{ij} - \mu_{ij} V_{ij}^{-1} \mathrm{w}_{ij} \frac{\partial \mu_{ij}}{\partial \eta_{ij}} \mathrm{w}_{ij} \right\} \\
&= \sum_{i=1}^{n} (\mathrm{a}_{ij}(\phi))^{-1} \left\{ (\mathrm{y}_{ij} - \mu_{ij}) V_{ij}^{-1} \frac{\partial \mu_{ij}}{\partial \eta_{ij}} \mathrm{w}_{ij} \right\} \qquad (j = 1, \ldots, p).
\end{aligned}$$

Then, the function score for $\boldsymbol{\alpha}$ can be written in Matrix as follows:

$$\mathbf{U}_{\mathrm{p}}^{\alpha}(\boldsymbol{\theta}) = \frac{\partial L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}} = \mathbf{W}^T \mathbf{T} (\mathbf{y} - \boldsymbol{\mu}),$$

where $\mathbf{W}$ is an $(n \times p)$ matrix whose ith row is $\mathbf{w}_{ij}^T$, $\mathbf{T} = \mathrm{diag}\left[(\mathrm{a}_{ij}(\phi))^{-1} \left(\frac{\partial \mu_{ij}}{\partial \eta_{ij}} \right) V_{ij}^{-1} \right]$ with $V_{ij} = V(\mu_{ij}) = \frac{\partial \mu_{ij}}{\partial \theta_{ij}}$ the function of variance, $\mathbf{y} = (\mathbf{y}_1^T, ..., \mathbf{y}_n^T)^T$ and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, ..., \boldsymbol{\mu}_n^T)^T$ with $\mathbf{y}_i = (y_{i1}, ..., y_{in_i})^T$ and $\boldsymbol{\mu}_i = (\mu_{i1}, ..., \mu_{in_i})^T$.

3.1.2 Nonparametric component

To obtain the function score for the $\boldsymbol{\beta}_k$ parameter we must calculate

$$\begin{aligned}
\frac{\partial L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \psi_{k_l}} &= \sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\left\{ \mathrm{y}_{ij}\,\frac{\partial \theta_{ij}}{\partial \mu_{ij}}\frac{\partial \mu_{ij}}{\partial \eta_{ij}}\frac{\partial \eta_{ij}}{\partial \psi_{k_l}} - \frac{\partial \psi(\theta_{ij})}{\partial \theta_{ij}}\frac{\partial \theta_{ij}}{\partial \mu_{ij}}\frac{\partial \mu_{ij}}{\partial \eta_{ij}}\frac{\partial \eta_{ij}}{\partial \psi_{k_l}}\right\} - \sum_{k=1}^{s}\lambda_k \mathbf{K}_{k_l}\psi_k \\
&= \sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\left\{ \mathrm{y}_{ij}\,V_{ij}^{-1}\frac{\partial \mu_{ij}}{\partial \eta_{ij}}\,\mathrm{n}_{kil} - \mu_{ij}V_{ij}^{-1}\frac{\partial \mu_{ij}}{\partial \eta_{ij}}\,\mathrm{n}_{kil}\right\} - \sum_{k=1}^{s}\lambda_k \mathbf{K}_{k_l}\psi_k \\
&= \sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\left\{ (\mathrm{y}_{ij} - \mu_{ij})V_{ij}^{-1}\frac{\partial \mu_{ij}}{\partial \eta_{ij}}\,\mathrm{n}_{kil}\right\} - \sum_{k=1}^{s}\lambda_k \mathbf{K}_{k_l}\psi_k \qquad (l = 1, \ldots, r).
\end{aligned}$$

Then, the function score for $\boldsymbol{\beta}_k$ can be written in Matrix as follows:

$$\mathbf{U}_{\mathrm{p}}^{\beta_k}(\boldsymbol{\theta}) = \frac{\partial L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\beta}_k} = \widetilde{\mathbf{N}}_k^{T}\mathbf{T}(\mathbf{y}-\boldsymbol{\mu}) - \lambda_k \mathbf{K}_k \boldsymbol{\beta}_k \qquad (k = 1, \ldots, s)$$

3.1.3 Dispersion component

Finally, the function score for the ϕ parameter is given by

$$\frac{\partial L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \phi} = \sum_{i=1}^{n} -(\mathrm{a}_{ij}(\phi))^{-2}\{\mathrm{y}_{ij}\,\theta_{ij} - \psi(\theta_{ij})\} + \sum_{i=1}^{n} c'(y_{ij},\phi),$$

where $c' = \partial\, \mathrm{c}(y_{ij},\phi)/\partial\phi$.

3.2 Matrix of second derivatives penalized

Using results derived from matrices we have that the matrix of second derivatives with respect to $\boldsymbol{\theta}$ is given by

$$\frac{\partial L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} \quad = \quad \sum_{i=1}^{n}\frac{\partial L_{\mathrm{p}_i}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}.$$

In particular,

$$\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\alpha_j\partial\alpha_l} = \sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}(\mathrm{y}_{ij}-\mu_{ij})\frac{\partial^2\theta_{ij}}{\partial\mu_{ij}^2}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 \mathrm{w}_{ij}\,\mathrm{w}_{il}$$

$$+\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}(\mathrm{y}_{ij}-\mu_{ij})\frac{\partial\theta_{ij}}{\partial\mu_{ij}}\frac{\partial^2\mu_{ij}}{\partial\eta_{ij}^2}\,\mathrm{w}_{ij}\,\mathrm{w}_{il}-\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\frac{\partial\theta_{ij}}{\partial\mu_{ij}}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 \mathrm{w}_{ij}\,\mathrm{w}_{il}\,,$$

$$\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\psi_{kj}\partial\psi_{kl}} = \sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}(\mathrm{y}_{ij}-\mu_{ij})\frac{\partial^2\theta_{ij}}{\partial\mu_{ij}^2}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 \mathrm{n}_{kij}\,\mathrm{n}_{kil}$$

$$+\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}(\mathrm{y}_{ij}-\mu_{ij})\frac{\partial\theta_{ij}}{\partial\mu_{ij}}\frac{\partial^2\mu_{ij}}{\partial\eta_{ij}^2}\,\mathrm{n}_{kij}\,\mathrm{n}_{kil}-\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\frac{\partial\theta_{ij}}{\partial\mu_{ij}}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 \mathrm{n}_{kij}\,\mathrm{n}_{kil}-\lambda_k\mathbf{K}_k\,,$$

when $k \neq k'$, we have

$$\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\psi_{kj}\partial\psi_{k'l}} \quad = \quad \sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}(\mathrm{y}_{ij}-\mu_{ij})\frac{\partial^2\theta_{ij}}{\partial\mu_{ij}^2}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 \mathrm{n}_{kij}\,\mathrm{n}_{k'il}$$

$$+\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}(\mathrm{y}_{ij}-\mu_{ij})\frac{\partial\theta_{ij}}{\partial\mu_{ij}}\frac{\partial^2\mu_{ij}}{\partial\eta_{ij}^2}\,\mathrm{n}_{kij}\,\mathrm{n}_{k'il}-\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\frac{\partial\theta_{ij}}{\partial\mu_{ij}}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 \mathrm{n}_{kij}\,\mathrm{n}_{k'il}\,,$$

$$\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\alpha_j\partial\psi_{kl}} = \sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}(\mathrm{y}_{ij}-\mu_{ij})\frac{\partial^2\theta_{ij}}{\partial\mu_{ij}^2}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 \mathrm{w}_{ij}\,\mathrm{n}_{kil}$$

$$+\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}(\mathrm{y}_{ij}-\mu_{ij})\frac{\partial\theta_{ij}}{\partial\mu_{ij}}\frac{\partial^2\mu_{ij}}{\partial\eta_{ij}^2}\,\mathrm{w}_{ij}\,\mathrm{n}_{kil} - \sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\frac{\partial\theta_{ij}}{\partial\mu_{ij}}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 \mathrm{w}_{ij}\,\mathrm{n}_{kil}\,,$$

$$\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\alpha_j\partial\phi} = -\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-2}\left\{(\mathrm{y}_{ij}-\mu_{ij})V_{ij}^{-1}\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\,\mathrm{w}_{ij}\right\},$$

$$\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\psi_{k_l}\partial\phi} = -\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-2}\left\{(\mathrm{y}_{ij}-\mu_{ij})V_{ij}^{-1}\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\,\mathrm{n}_{kil}\right\},$$

$$\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\phi^2} = \sum_{i=1}^{n}2(\mathrm{a}_{ij}(\phi))^{-3}\{\mathrm{y}_{ij}\,\theta_{ij}-\psi(\theta_{ij})\} + \sum_{i=1}^{n}c''(y_{ij},\phi)\,.$$

3.3 Expected Information Matrix

In general, by calculating the expectation of the matrix $-\mathbf{L}_{\mathrm{p}}$ we obtain the $(p^* \times p^*)$ penalized expected information matrix, denoted by

$$\mathfrak{I}_{\mathrm{p}}(\boldsymbol{\theta}) = -\mathrm{E}\left(\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right).$$

In particular, the elements of the matrix $\mathfrak{I}_{\mathrm{p}}(\boldsymbol{\theta})$ are given by

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\alpha_j\partial\alpha_l}\right) &= -\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\frac{\partial\theta_{ij}}{\partial\mu_{ij}}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 \mathrm{w}_{ij}\,\mathrm{w}_{il} \\
&= -\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 V_{ij}^{-1}\,\mathrm{w}_{ij}\,\mathrm{w}_{il} \\
&= -\mathbf{W}^T\mathbf{M}\mathbf{W}\,,
\end{aligned}$$

where $\mathbf{M} = diag\left[\mathrm{a}_{ij}(\phi))^{-1}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 V_{ij}^{-1}\right]$

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\psi_{kj}\partial\psi_{kl}}\right) &= -\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 V_{ij}^{-1}\,\mathrm{n}_{kij}\,\mathrm{n}_{kil} - \lambda_k\mathbf{K}_k \\
&= -\widetilde{\mathbf{N}}^T\mathbf{M}\widetilde{\mathbf{N}} - \lambda_k\mathbf{K}_k\,,
\end{aligned}$$

for $k \neq k'$ we get

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\psi_{kj}\partial\psi_{k'l}}\right) &= -\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 V_{ij}^{-1}\,\mathrm{n}_{kij}\,\mathrm{n}_{k'il} \\
&= -\widetilde{\mathbf{N}}^T\mathbf{M}\widetilde{\mathbf{N}}\,,
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\alpha_j\partial\psi_{kl}}\right) &= -\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-1}\left(\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\right)^2 V_{ij}^{-1}\,\mathrm{w}_{ij}\,\mathrm{n}_{kil} \\
&= -\widetilde{\mathbf{N}}^T\mathbf{M}\mathbf{W}\,,
\end{aligned}$$

$$\mathrm{E}\left(\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\alpha_j\partial\phi}\right) = \mathbf{0}\,,$$

$$\mathrm{E}\left(\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\psi_{k_l}\partial\phi}\right)=\mathbf{0}$$

and

$$\mathrm{E}\left(\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\phi^2}\right)=\sum_{i=1}^{n}2(\mathrm{a}_{ij}(\phi))^{-3}(\mu_{ij}\theta_{ij}-\psi(\theta_{ij}))+\sum_{i=1}^{n}\mathrm{E}(c''(y_{ij},\phi))\,.$$

Fisher's information matrix takes the following block-diagonal form:

$$\mathfrak{I}_{\mathrm{p}}(\boldsymbol{\theta}) \;=\; \begin{pmatrix} \mathfrak{I}_{\mathrm{p}}^{\alpha\beta_k}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathfrak{I}_{\mathrm{p}}^{\phi\phi}(\boldsymbol{\theta}) \end{pmatrix},$$

where

$$\mathfrak{I}_{\mathrm{p}}^{\alpha\beta_k}(\boldsymbol{\theta}) \;=\; \begin{pmatrix} \mathbf{W}^T\mathbf{M}\mathbf{W} & \mathbf{W}^T\mathbf{M}\widetilde{\mathbf{N}}_1 & \dots & \mathbf{W}^T\mathbf{M}\widetilde{\mathbf{N}}_s \\ \widetilde{\mathbf{N}}_1^T\mathbf{M}\mathbf{W} & \widetilde{\mathbf{N}}_1^T\mathbf{M}\widetilde{\mathbf{N}}_1+\lambda_1\mathbf{K}_1 & \dots & \widetilde{\mathbf{N}}_1^T\mathbf{M}\widetilde{\mathbf{N}}_s \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{\mathbf{N}}_s^T\mathbf{M}\mathbf{W} & \widetilde{\mathbf{N}}_s^T\mathbf{M}\widetilde{\mathbf{N}}_1 & \dots & \widetilde{\mathbf{N}}_s^T\mathbf{M}\widetilde{\mathbf{N}}_s+\lambda_s\mathbf{K}_s \end{pmatrix}$$

and

$$\mathfrak{I}_{\mathrm{p}}^{\phi\phi}(\boldsymbol{\theta}) \;=\; \sum_{i=1}^{n}-2(\mathrm{a}_{ij}(\phi))^{-3}(\mu_{ij}\theta_{ij}-\psi(\theta_{ij}))-\sum_{i=1}^{n}\mathrm{E}(c''(y_{ij},\phi)).$$

It is important to mention that the property of orthogonality of ϕ with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_k$ facilitates the construction of the iterative process that allows estimating the vector of theta parameters. See more details of this property in Paula (2010).

3.4 Estimation and inference

In this section we discuss some aspects of estimation and inference in PVCGLMs. Specifically, we discuss the estimation of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_s$ and ϕ based on penalized log-likelihood function. Then, we describe a procedure for calculating the approximate standard errors of the parameter estimates and we derive a approximate standard error bands for the smooth functions. In the following subsections we present a discussion on degrees of freedom estimation and smoothing parameters selection.

3.5 Fisher score and weighted back-fitting algorithms

Because $\boldsymbol{\beta}_k$ is an infinite-dimensional parameter we will consider the maximum penalized likelihood estimate (MPLE) of $\boldsymbol{\theta}$, which leads to a natural cubic spline estimate of $\boldsymbol{\beta}_k$. Specifically, the value of $\boldsymbol{\theta}$ that maximizes $L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ over $\boldsymbol{\Theta}$, denoted by $\widehat{\boldsymbol{\theta}}$, satisfies

$$L_{\mathrm{p}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) \geq \sup_{\theta \in \Theta} L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}).$$

According to Berhane and Tibshirani (1998), the determination of the MPLE $\widehat{\boldsymbol{\theta}}$ can be performed by using the Fisher scoring algorithm. In effect, let $\boldsymbol{\beta}_0 = \boldsymbol{\alpha}$ and $\widetilde{\mathbf{N}}_0 = \mathbf{W}$, and consider for simplicity $\boldsymbol{\lambda}$ and $\mathbf{M}$ fixed. Then, the Fisher scoring algorithm is given by

$$\begin{pmatrix} \mathbf{I} & \mathbf{S}_0^{(u)}\widetilde{\mathbf{N}}_1 & \dots & \mathbf{S}_0^{(u)}\widetilde{\mathbf{N}}_s \\ \mathbf{S}_1^{(u)}\widetilde{\mathbf{N}}_0 & \mathbf{I} & \dots & \mathbf{S}_1^{(u)}\widetilde{\mathbf{N}}_s \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_s^{(u)}\widetilde{\mathbf{N}}_0 & \mathbf{S}_s^{(u)}\widetilde{\mathbf{N}}_1 & \dots & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_0^{(u+1)} \\ \boldsymbol{\beta}_1^{(u+1)} \\ \vdots \\ \boldsymbol{\beta}_s^{(u+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_0^{(u)}\mathbf{z}^{(u)} \\ \mathbf{S}_1^{(u)}\mathbf{z}^{(u)} \\ \vdots \\ \mathbf{S}_s^{(u)}\mathbf{z}^{(u)} \end{pmatrix}, \qquad (3.1)$$

where $\mathbf{z}^{(u)} = \boldsymbol{\eta}'(\mathbf{Y} - \boldsymbol{\mu}) + \left(\sum_{k=0}^{s} \widetilde{\mathbf{N}}_k \boldsymbol{\beta}_{\boldsymbol{k}} \right)\Big|_{\boldsymbol{\theta}^{(u)}}$ and $\mathbf{S}_k^{(u)} = \mathbf{S}_k\big|_{\boldsymbol{\theta}^{(u)}}$, with

$$\mathbf{S}_k = \begin{cases} (\widetilde{\mathbf{N}}_0^T \mathbf{M} \widetilde{\mathbf{N}}_0)^{-1} \widetilde{\mathbf{N}}_0^T \mathbf{W} & k = 0 \\ (\widetilde{\mathbf{N}}_k^T \mathbf{M} \widetilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \widetilde{\mathbf{N}}_k^T \mathbf{M} & k = 1, \dots, s \end{cases}$$

Then, the back-fitting (Gauss-Seidel) iterations that are used to solve the equations system (3.1) take the form

$$\boldsymbol{\beta}_k^{(u+1)} = \mathbf{S}_k^{(u)} \left(\mathbf{z}^{(u)} - \sum_{l=0, l \neq k}^{s} \widetilde{\mathbf{N}}_l \mathbf{f}_l^{(u)} \right). \qquad (3.2)$$

Note that the system of equations (3.1) is consistent and the back-fitting algorithm (3.2) converges to a solution for any starting values if the weight matrix involved is symmetric and defined positive (see, for instance, Berhane and Tibshirani, 1998). Additionally, we have that this solution is unique when there not concurvity in the data.

3.6 Joint iterative process

The solution of the estimating equation system (3.1) to obtain the MPLE of $\boldsymbol{\theta}$ may be attained by iterating between a weighted back-fitting algorithm with weight matrix $\mathbf{M}$ and a Fisher score algorithm to obtain maximum likelihood estimation of the parameter ϕ, which is equivalent to the following iterative process:

(i) Initialize:

  (a) Fitting a partially varying-coefficient generalized model to get $\boldsymbol{\alpha}^{(0)}$.

  (b) Get starting value for ϕ by using the fitted values from (a).

  (c) From the current value $\beta^{(0)} = (\beta^T, \beta_1^{(0)^T}, \ldots, \beta_s^{(0)^T}, \phi^{(0)})^T$ obtaining the weight matrix $\mathbf{M}^{(0)}$, with $m_i^{(0)} = m_i\big|_{\boldsymbol{\theta}^{(0)}}$. Then, obtain

$$
\begin{aligned}
\mathbf{z}^{(0)} &= \boldsymbol{\eta'}^{(0)}(\mathbf{Y} - \boldsymbol{\mu}^{(0)}) + \left(\sum_{k=0}^{s} \widetilde{\mathbf{N}}_k \boldsymbol{\beta_k}\right), \\
\mathbf{S}_0^{(0)} &= (\widetilde{\mathbf{N}}_0^T \mathbf{M}^{(0)} \widetilde{\mathbf{N}}_0)^{-1} \mathbf{N}_0^T \mathbf{M}^{(0)} \qquad \text{and} \\
\mathbf{S}_k^{(0)} &= (\widetilde{\mathbf{N}}_k^T \mathbf{M}^{(0)} \widetilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \widetilde{\mathbf{N}}_k^T \mathbf{M}^{(0)}, \qquad (k = 1, \ldots, s).
\end{aligned}
$$

(ii) Step 1: Iterate repeatedly by cycling between the following equations:

$$
\begin{aligned}
\boldsymbol{\beta}_0^{(u+1)} &= \mathbf{S}_0^{(u)} \left(\mathbf{z}^{(u)} - \sum_{l=1}^{s} \widetilde{\mathbf{N}}_l \boldsymbol{\beta}_l^{(u)}\right), \\
\boldsymbol{\beta}_1^{(u+1)} &= \mathbf{S}_1^{(u)} \left(\mathbf{z}^{(u)} - \widetilde{\mathbf{N}}_0 \boldsymbol{\beta}_0^{(u+1)} - \sum_{l=2}^{s} \widetilde{\mathbf{N}}_l \boldsymbol{\beta}_l^{(u)}\right), \\
&\vdots \\
\boldsymbol{\beta}_s^{(u+1)} &= \mathbf{S}_s^{(u)} \left(\mathbf{z}^{(u)} - \sum_{l=0}^{s-1} \widetilde{\mathbf{N}}_l \boldsymbol{\beta}_l^{(u+1)}\right),
\end{aligned}
$$

for $u = 0, 1, \ldots$. Repeat (ii) replacing $\boldsymbol{\beta}_{\jmath}^{(u)}$ by $\boldsymbol{\beta}_{\jmath}^{(u+1)}$ until convergence criterion $\Delta_u(\boldsymbol{\beta}_{\jmath}^{(u+1)}, \boldsymbol{\beta}_{\jmath}^{(u)}) = \sum_{\jmath=0}^{s} \parallel \boldsymbol{\beta}_{\jmath}^{(u+1)} - \boldsymbol{\beta}_{\jmath}^{(u)} \parallel / \sum_{\jmath=0}^{s} \parallel \boldsymbol{\beta}_{\jmath}^{(u)} \parallel$ is below some small threshold (Hastie and Tibshirani, 1990).

(iii) Step 2: For current values $\boldsymbol{\beta}_{\jmath}^{(u+1)}$ $(\jmath = 0, 1, \ldots, s)$, obtaining $\phi^{(u+1)}$ by using

$$\phi^{(u+1)} \;\; = \;\; \phi^{(u)} - \mathrm{E}\left\{ \frac{\partial^2 L_{\mathrm{p}}^{c}(\phi, \boldsymbol{\lambda})}{\partial\phi\partial\phi} \right\}^{-1} \frac{\partial L_{\mathrm{p}}^{c}(\phi, \boldsymbol{\lambda})}{\partial\phi}\Big|_{\boldsymbol{\theta}^{(u)}} .$$

(iv) Iterating between steps (ii) and (iii) by replacing $\boldsymbol{\beta}_{\jmath}^{(0)}$ $(\jmath = 0, 1, \ldots, s)$ and $\phi^{(0)}$ by $\boldsymbol{\beta}_{\jmath}^{(u+1)}$ and $\phi^{(u+1)}$, respectively, until convergence.

3.7 Approximate standard errors and bands

In this work we derive the covariance matrix of $\widehat{\boldsymbol{\theta}}$ from the inverse of the expected information matrix $\mathfrak{J}_{\mathrm{p}}$ defined in Subsection 3.3. Therefore, the approximate covariance matrix of $\widehat{\boldsymbol{\theta}}$ is given as

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\theta}}) \;\; \approx \;\; \mathfrak{J}_{\mathrm{p}}^{-1}\big|_{\widehat{\boldsymbol{\theta}}} .$$

$$\mathfrak{J}_{\mathrm{p}}^{-1} \;\; = \;\; \begin{pmatrix} \mathfrak{J}_{11} & \mathfrak{J}_{12} & \mathbf{0} \\ \mathfrak{J}_{21} & \mathfrak{J}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathfrak{J}_{\mathrm{p}}^{\phi\phi^{-1}} \end{pmatrix},$$

where

$$\begin{aligned}
\mathfrak{J}_{11} &= \left(\mathfrak{J}_{\mathrm{p}}^{\alpha\alpha} - \mathfrak{J}_{\mathrm{p}}^{\alpha\beta} \mathfrak{J}_{\mathrm{p}}^{\beta\beta^{-1}} \mathfrak{J}_{\mathrm{p}}^{\alpha\beta^{T}} \right)^{-1}, \\
\mathfrak{J}_{12} &= -\mathfrak{J}_{\mathrm{p}}^{\alpha\alpha^{-1}} \mathfrak{J}_{\mathrm{p}}^{\alpha\beta} \left(\mathfrak{J}_{\mathrm{p}}^{\beta\beta} - \mathfrak{J}_{\mathrm{p}}^{\alpha\beta^{T}} \mathfrak{J}_{\mathrm{p}}^{\alpha\alpha^{-1}} \mathfrak{J}_{\mathrm{p}}^{\alpha\beta} \right)^{-1}, \\
\mathfrak{J}_{21} &= \left(\mathfrak{J}_{\mathrm{p}}^{\beta\beta} - \mathfrak{J}^{\alpha\beta^{T}} \mathfrak{J}^{\alpha\alpha^{-1}} \mathfrak{J}_{\mathrm{p}}^{\alpha\beta} \right)^{-1} \mathfrak{J}_{\mathrm{p}}^{\alpha\beta^{T}} \mathfrak{J}_{\mathrm{p}}^{\alpha\alpha^{-1}} \qquad \textit{and} \\
\mathfrak{J}_{22} &= \left(\mathfrak{J}_{\mathrm{p}}^{\beta\beta} - \mathfrak{J}_{\mathrm{p}}^{\alpha\beta^{T}} \mathfrak{J}_{\mathrm{p}}^{\alpha\alpha^{-1}} \mathfrak{J}_{\mathrm{p}}^{\alpha\beta} \right)^{-1},
\end{aligned}$$

with

$$
\begin{aligned}
\mathfrak{J}_{\mathrm{p}}^{\alpha\alpha} &= \mathbf{W}^T\mathbf{M}\mathbf{W}\,, \\
\mathfrak{J}_{\mathrm{p}}^{\alpha\beta} &= \begin{pmatrix} \mathbf{W}^T\mathbf{M}\widetilde{\mathbf{N}}_1 & \dots & \mathbf{W}^T\mathbf{M}\widetilde{\mathbf{N}}_s \end{pmatrix}, \\
\mathfrak{J}_{\mathrm{p}}^{\alpha\beta^T} &= \begin{pmatrix} \widetilde{\mathbf{N}}_1^T\mathbf{M}\mathbf{W} \\ \dots \\ \widetilde{\mathbf{N}}_s^T\mathbf{M}\mathbf{W} \end{pmatrix}
\end{aligned}
$$

and

$$
\mathfrak{J}_{\mathrm{p}}^{\beta\beta} = \begin{pmatrix} \widetilde{\mathbf{N}}_1^T\mathbf{M}\widetilde{\mathbf{N}}_1 + \lambda_1\mathbf{K}_1 & \dots & \widetilde{\mathbf{N}}_1^T\mathbf{M}\widetilde{\mathbf{N}}_s \\ \vdots & \ddots & \vdots \\ \widetilde{\mathbf{N}}_s^T\mathbf{M}\widetilde{\mathbf{N}}_1 & \dots & \widetilde{\mathbf{N}}_s^T\mathbf{M}\widetilde{\mathbf{N}}_s + \lambda_s\mathbf{K}_s \end{pmatrix}.
$$

If we are interested in drawing inferences for $\boldsymbol{\alpha}$ and $(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_s)$, the approximate covariance matrices can be estimated by using the corresponding block-diagonal matrices obtained from $\mathfrak{J}_{\mathrm{p}}^{-1}$. In addition, we can construct an approximate pointwise standard error band (SEB) for $\beta_k(\cdot)$ that allows us to assess how accurate the estimator $\widehat{\beta}_k(\cdot)$ at different locations within the range of interest. For example, we can consider the following approximate pointwise SEB (Hastie and Tibshirani, 1990):

$$
\mathrm{SEB}_{\mathrm{approx}}(\beta_k(\mathrm{t}_l^0)) \;=\; \widehat{\beta}_k(\mathrm{t}_l^0) \pm 2\sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_k(\mathrm{t}_l^0))}\;(l = 1, \dots, r),
$$

where $\mathrm{Var}(\widehat{\beta}_k(\mathrm{t}_l))$ is the lth principal diagonal element of the corresponding block-diagonal matrix.

3.8 On degrees of freedom

For the PVCGLM, the degree of freedom (df) associated with the kth smooth function is given by (see, for instance, Hastie and Tibshirani, 1990)

$$
\mathrm{df}(\lambda_k) = \mathrm{tr}\big(\widetilde{\mathbf{N}}_k\mathbf{S}_k\big) = \mathrm{tr}\big(\widetilde{\mathbf{S}}_k\big)\,, \tag{3.3}
$$

which measure the individual effect contribution of the kth component, with $\mathbf{S}_k$ defined in Section 3.5. Following Eilers and Marx (1996), we can write $\mathrm{tr}\big\{\widetilde{\mathbf{S}}_k\big\}$ as

$$
\mathrm{tr}\big\{\widetilde{\mathbf{S}}_k\big\} = \sum_{j=1}^{r_k} \frac{1}{1+\lambda_k\ell_j} \approx 2 + \sum_{j=3}^{r_k} \frac{1}{1+\lambda_k\ell_j}\,, \tag{3.4}
$$

where ℓ_j, for $j = 1, \ldots, r_k$, are the eigenvalues of the matrix $\mathbf{Q}_{\widetilde{\mathbf{N}}_k}^{-1/2}\mathbf{Q}_{\lambda_k}\mathbf{Q}_{\widetilde{\mathbf{N}}_k}^{-1/2}$, for $k = 1, \ldots, s$, with $\mathbf{Q}_{\widetilde{\mathbf{N}}_k} = \widetilde{\mathbf{N}}_k^T\mathbf{W}\widetilde{\mathbf{N}}_k$ and $\mathbf{Q}_{\lambda_k} = \lambda_k\mathbf{K}_k$. It is important to note that *(i)* $\mathrm{df}(\alpha_k) = \mathrm{tr}\{\widetilde{\mathbf{S}}_k\}$ is a monotonically decreasing function of λ_k; *(ii)* $\mathrm{df}(\lambda_k) \to 2 + r_k$ as $\lambda_k \to 0$; *(iii)* $\mathrm{df}(\lambda_k) \to 2$ as $\lambda_k \to \infty$; and *(iv)* $2 \leq \mathrm{df}(\lambda_k) \leq 2 + r_k$.

3.9 Choosing the smoothing parameters

In the previous subsections the smoothing parameters λ_k's were assumed fixed. However, in practice situations the smoothing parameters should be selected from the data. When a smoothing spline is used, for example, it is usual to consider the cross-validation method or the generalized cross-validation method (Craven and Wahba, 1979). Alternatively, these parameters may be selected by applying the Akaike information criterion (AIC) (Akaike, 1973); see also Hurvich et al. (1998) and Simonoff and Tsai (1999) in the semiparametric context. In this work we will apply the following procedure based on the AIC (see, for instance, Ibacache-Pulgar et al., 2013):

(i) For simplicity, consider $s = 2$.

- (i.1) Select m values $u_{k_\ell} \in (0, 1)$ and obtain the smoothing parameter values $\lambda_{k_\ell} = u_{k_\ell}/(1 - u_{k_\ell})$, for $\ell = 1, \ldots, m$.
- (i.2) From the equation (3.3) obtain $\mathrm{df}_{k_\ell} = \mathrm{df}(\lambda_{k_\ell})$ and perform a dispersion graph between λ_{k_ℓ} and df_{k_ℓ}. From (3.4) a reciprocal relationship is expected between λ_k and $\mathrm{df}(\lambda_{k_\ell})$.

(ii) Select a range for the smoothing parameters.

- (ii.1) Obtain an appropriate regression obtaining the fitted equation $\widehat{\lambda}_{k_\ell} = \eta(\mathrm{df}_{k_\ell})$, where $\eta(\cdot)$ denotes the regression function.
- (ii.2) Since the relationship between λ_k and df_k is monotonically decreasing we may obtain from the fitted regression a range $[\lambda_k^{L_k}, \lambda_k^{U_k}]$ for λ_k given a range for the degrees of freedom. For example, if we consider the range $[2, 16]$ we have that $\lambda_k^{U_k} = \eta(16)$ and $\lambda_k^{L_k} = \eta(2)$.

(iii) Minimizing the AIC.

The suggestion is to select a grid of values from the range $[\lambda_k^{L_k}, \lambda_k^{U_k}]$ and choosing the smoothing parameters values λ_k that minimizes

$$\text{AIC}(\boldsymbol{\lambda}) = -2L_{\text{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})\big|_{\widehat{\boldsymbol{\theta}}} + 2\Big[1 + p + \text{df}(\boldsymbol{\lambda})\Big],$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$, p denote the number of parameters in $\boldsymbol{\beta}$, and $\text{df}(\boldsymbol{\lambda}) = \sum_{k=1}^{2} \text{df}(\lambda_k)$ denote approximately the number of effective parameters involved in modelling of the nonparametric effects.

3.10 Final comments

In this chapter we present the estimation and inference of the partially varying-coefficient generalized linear model. We derive the score functions, the Hessian matrix and the Fisher information matrix from the penalized likelihood function. In addition, we proposed an algorithm weighted backtting to adjust an PVCGLM, which leads to a natural cubic spline estimator for functions $\boldsymbol{\beta}_k$. From the Fisher information matrix penalized, we calculate the standard errors associated with the regression coefficient estimators and the non-parametric function , and we build a trust band for function $\boldsymbol{\beta}_k$. Also, we calculated the degrees of freedom associated the model and were estimated the smoothing parameters. Finally, it is proposed as a criterion for the selection of models in methods based on the AIC.

Chapter 4

Local influence

4.1 The method

Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T$ be an n-dimensional vector of perturbations restricted to some open subset $\Omega \in \mathcal{R}^n$ and the logarithm of the perturbed penalized likelihood denoted by $L_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\lambda} \,|\, \boldsymbol{\omega})$. It is assumed that exists $\boldsymbol{\omega}_0 \in \Omega$, a vector of no perturbation, such that $L_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\lambda} \,|\, \boldsymbol{\omega}_0) = L_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\lambda})$. To assess the influence of minor perturbations on the MPLE $\widehat{\boldsymbol{\theta}}$, we can consider the likelihood displacement

$$LD(\boldsymbol{\omega}) = 2\Big[L_{\rm p}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) - L_{\rm p}(\widehat{\boldsymbol{\theta}}_\omega, \boldsymbol{\lambda})\Big] \geq 0 \,,$$

where $\widehat{\boldsymbol{\theta}}_\omega$ is the MPLE under $L_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\lambda} \,|\, \boldsymbol{\omega})$. The measure $LD(\boldsymbol{\omega})$ is useful for assessing the distance between $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_\omega$. Cook (1986) suggests to study the local behavior of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$. The procedure consists in selecting a unit direction $\boldsymbol{\ell} \in \Omega$ ($\|\boldsymbol{\ell}\| = 1$), and then to considering the plot of $LD(\boldsymbol{\omega}_0 + a\,\boldsymbol{\ell})$ against a, where $a \in \mathcal{R}$. This plot is called lifted line. Each lifted line can be characterized by considering the normal curvature $C_\ell(\boldsymbol{\theta})$ around $a = 0$. The suggestion is to consider the direction $\boldsymbol{\ell} = \boldsymbol{\ell}_{max}$ corresponding to the largest curvature $C_{\ell_{max}}(\boldsymbol{\theta})$. The index plot of $\boldsymbol{\ell}_{max}$ may reveal those observations that under small perturbations exercise notable influence on $LD(\boldsymbol{\omega})$. According to Cook (1986), the normal curvature at the unitary direction $\boldsymbol{\ell}$ is given by $C_\ell(\boldsymbol{\theta}) = -2\{\boldsymbol{\ell}^T\boldsymbol{\Delta}_{\rm p}^T\mathbf{L}_{\rm p}^{-1}\boldsymbol{\Delta}_{\rm p}\boldsymbol{\ell}\}$, where $\boldsymbol{\Delta}_{\rm p} = \partial^2 L_{\rm p}(\boldsymbol{\theta},\boldsymbol{\lambda}\,|\,\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^T\big|_{\theta=\widehat{\theta},\ \omega=\omega_0}$. Note that $-\mathbf{L}_{\rm p}$ is the penalized observed information matrix evaluated at $\widehat{\boldsymbol{\theta}}$ (see Subsection 3.2) and $\boldsymbol{\Delta}_{\rm p}$ is the penalized perturbation matrix evaluated at $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$. $C_\ell(\boldsymbol{\theta})$ denotes the local influence on the estimate $\widehat{\boldsymbol{\theta}}$ after perturbing the model or data. Escobar e Meeker (1992) proposed to study the normal curvature at the direction $\boldsymbol{\ell} = \mathbf{e}_i$, where $\mathbf{e}_i$ is an n-dimensional vector with one at the ith position and zeros at the remaining positions. In this case, the normal curvature, called total local influence of the ith individual, takes the form $C_{\mathbf{e}_i}(\boldsymbol{\theta}) = 2|\,{\rm c}_{ii}\,|$ $(i = 1,\ldots,n)$, where ${\rm c}_{ii}$ is the ith principal diagonal element of the matrix $\mathbf{C} = \boldsymbol{\Delta}_{\rm p}^T\mathbf{L}_{\rm p}^{-1}\boldsymbol{\Delta}_{\rm p}$. In order to have a curvature invariant under uniform change of scale Poon and Poon (1999) proposed the conformal normal curvature defined as

$$B_\ell(\boldsymbol{\theta}) = \frac{C_\ell(\boldsymbol{\theta})}{2\sqrt{{\rm tr}(\boldsymbol{\Delta}_{\rm p}^T\mathbf{L}_{\rm p}^{-1}\boldsymbol{\Delta}_{\rm p})^2}} = -\frac{\boldsymbol{\ell}^T\boldsymbol{\Delta}_{\rm p}^T\mathbf{L}_{\rm p}^{-1}\boldsymbol{\Delta}_{\rm p}\boldsymbol{\ell}}{\sqrt{{\rm tr}(\boldsymbol{\Delta}_{\rm p}^T\mathbf{L}_{\rm p}^{-1}\boldsymbol{\Delta}_{\rm p})^2}}\,.$$

This curvature is characterized to allow for any unitary direction $\boldsymbol{\ell}$ that $0 \leq B_\ell(\theta) \leq 1$. A suggestion is to consider the direction $\boldsymbol{\ell} = \boldsymbol{\ell}_{max}$ corresponding to the largest curvature $B_{\ell_{max}}(\boldsymbol{\theta})$ or, alternatively, evaluating the normal curvature at the direction $\boldsymbol{\ell} = \mathbf{e}_i$ and observing the index plot of $B_{\mathbf{e}_i}(\boldsymbol{\theta})$.

4.2 Some Types of Perturbation

In the static literature there is no clear definition of perturbation. According to Billor Loynes (1993), perturbation is any modification in the assumptions of the model and/or in the data to verify substantial differences in the results of the analysis.

The most common perturbation schemes are:

4.2.1 Cases-weight perturbation

With this type of disturbance we want to evaluate if the contribution of the observations with different weights, affect the maximum likelihood estimator of $\boldsymbol{\theta}$. According to Cook (1987), the case weight perturbation is perhaps the most common method to evaluate the influence on a small modification of a model. Consider the postulated model as a generalized linear model deduced by:

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}),$$

where

$$L_i(\boldsymbol{\theta}) = \left[\frac{y_i\theta_i - \psi(\theta_i)}{\phi} + c(y_i, \phi) \right] \qquad (i = 1, \ldots, n).$$

Then the log likelihood function of the disturbed model is given by:

$$L(\boldsymbol{\theta}/\omega) = \sum_{i=1}^{n} \omega_i L_i(\boldsymbol{\theta}).$$

4.2.2 Data perturbation

Billor Loynes (1993) consider that the types of possible perturbations depend on the model. In generalized linear models one can disturb the explanatory variables (regressors) or the response vector. There are two reasons to consider data perturbation: the existence of measures with measurement errors and the existence of outliers (in a relatively small proportion of data).

4.2.3 Predictors of perturbation

This type of perturbation can be done in two ways:

$$\mathrm{x}_{ij\omega} = \begin{cases} \mathrm{x}_{ij\omega} + \omega_i (\text{Additive perturbation}) \\ \\ \mathrm{x}_{ij\omega} \times \omega_i (\text{Multiplicative perturbation}) \end{cases} \qquad (i = 1, \ldots, n; j = 1, \ldots, n_i).$$

4.2.4 Response of perturbation

The response can be perturbed in two ways:

$$\mathrm{y}_{ij\omega} = \begin{cases} \mathrm{y}_{ij} + \omega_i \times \mathrm{s}_i \text{ (Additive perturbation)} \\ \\ \mathrm{y}_{ij} \times \omega_i \times \mathrm{s}_i \text{ (Multiplicative perturbation)} \end{cases} \qquad (i = 1, \ldots, n; j = 1, \ldots, n_i) \quad ,$$

where $\sqrt{\phi \frac{\partial^2 \mathrm{b}}{\partial^2 \eta_i}}$ is the standard deviation of y_i.

For the previously written, the disturbance schemes can be classified into two large groups: perturbations in the model (they study modifications in the assumptions of the model) and / or in the data. Essentially, the disturbance scheme to be considered must be formulated in a way that responds to questions previously established by the researcher.

4.3 Perturbation schemes applied to the partially varying-coefficient generalized linear models

The $(p^* \times n)$ $\mathbf{\Delta}_\mathrm{p}$ matrix for each perturbation scheme assumes the form

$$\mathbf{\Delta}_\mathrm{p} = \left. \frac{\partial^2 L_\mathrm{p}(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}} \; \boldsymbol{\omega}=\boldsymbol{\omega_0}} ,$$

where $\widehat{\boldsymbol{\theta}}$ is the MPLE of $\boldsymbol{\theta}$ and $\boldsymbol{\omega_0}$ is the vector of no perturbation. We will present in the sequel the expression of $(p^* \times n)$ $\mathbf{\Delta}_\mathrm{p}$ matrix for case-weight, explanatory variable, response variable and scale perturbation schemes.

Note: To deduce the perturbation matrix the following fact will be used:

$$\theta_{ij} = \theta_{ij}(\mu_{ij}) = h(\eta_{ij}).$$

4.3.1 Cases-weight perturbation

Let us consider the attributed weights for the observations in the penalized log-likelihood function as

$$L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega}) = L(\boldsymbol{\theta}|\boldsymbol{\omega}) - \sum_{k=1}^{s} \frac{\lambda_k}{2} \boldsymbol{\beta}_k^T \mathbf{K}_k \boldsymbol{\beta}_k \, ,$$

where $L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \, L_i(\boldsymbol{\theta})$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T$ is the vector of weights, with $0 \leq \omega_i \leq 1$. In this case, the vector of no perturbation is given by $\boldsymbol{\omega}_0 = \mathbf{1}_{(n\times 1)}$. Differentiating $L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})$ with respect to the elements of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}^T$, we obtain after some algebraic manipulation

$$\begin{aligned}
\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\omega}^T}\Big|_{\theta=\hat{\theta},\, \omega=\omega_0} &= \mathbf{W}^T \mathbf{D}_a \, , \\
\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\omega}^T}\Big|_{\theta=\hat{\theta},\, \omega=\omega_0} &= \widetilde{\mathbf{N}}_k^T \mathbf{D}_a \qquad (k = 1, \ldots, s) \qquad \text{and} \\
\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^T}\Big|_{\theta=\hat{\theta},\, \omega=\omega_0} &= \widehat{\mathbf{u}}^T \, ,
\end{aligned}$$

where $\mathbf{D}_a = \mathrm{diag}_{1\leq i\leq n}(a_i)$ and $\widehat{\mathbf{u}} = (\widehat{u}_1, \ldots, \widehat{u}_n)^T$, with $a_i = (\mathrm{a}_{ij}(\phi))^{-1}(y_{ij} - \partial\psi/\partial h)\partial h/\partial\eta_{ij}$, $u_i = -(\mathrm{a}_{ij}(\phi))^{-2}(y_{ij}h(\eta_{ij}) - \psi(h(\eta_{ij})) + \mathrm{c}'(y_{ij}, \phi))\, \mathrm{e}_{in}^T$ and e_{in} is a vector with 1 at the ith position and zero elsewhere.

4.3.2 Response variable perturbation

To perturb the response variable values we consider $\mathrm{y}_{ij\omega} = \mathrm{y}_{ij} + \omega_i$ $(i = 1, \ldots, n)$, where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T$ is the vector of perturbations. Here, the vector of no perturbation is given by $\boldsymbol{\omega}_0 = (0, \ldots, 0)^T$ and the perturbed penalized log-likelihood function is constructed from (2.7) with y_{ij} replaced by $\mathrm{y}_{ij\omega}$, that is,

$$L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega}) = L(\boldsymbol{\theta}|\boldsymbol{\omega}) - \sum_{k=1}^{s} \frac{\lambda_k}{2}\, \boldsymbol{\beta}_k^T \mathbf{K}_k \boldsymbol{\beta}_k \,,$$

where $L(\cdot)$ is given by (2.4) with $\mathrm{y}_{ij\omega}$ in the place of y_{ij}. Differentiating $L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})$ with respect to the elements of $\boldsymbol{\theta}$ and ω_i, we obtain, after some algebraic manipulation, that

$$\begin{aligned}
\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\omega}^T}\Big|_{\theta=\hat{\theta},\ \omega=\omega_0} &= \mathbf{W}^T \mathbf{D}_c \,, \\
\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\omega}^T}\Big|_{\theta=\hat{\theta},\ \omega=\omega_0} &= \widetilde{\mathbf{N}}_k^T \mathbf{D}_c \qquad (k = 1, \ldots, s) \qquad \text{and} \\
\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^T}\Big|_{\theta=\hat{\theta},\ \omega=\omega_0} &= \widehat{\mathbf{d}}^T \,,
\end{aligned}$$

where $\mathbf{D}_c = \mathrm{diag}_{1\le i\le n}\big(c_i\big)$ and $\widehat{\mathbf{d}} = (d_1, \ldots, d_n)^T$, with $c_i = \partial h/\partial \eta_{ij}$ and $d_i = -(\mathrm{a}_{ij}(\phi))^{-2}(h(\eta_{ij})\, \mathrm{e}_{in}^T + \partial^2\, \mathrm{c}(\mathrm{y}_{i\omega}, \phi)/\partial \omega^2)$.

4.3.3 Explanatory variable perturbation

Here the dth explanatory variable, assumed continuous, is perturbed by considering the additive perturbation scheme, namely $\mathrm{w}_{id\omega} = \mathrm{w}_{id} + \omega_i$ $(i = 1, \ldots, n)$, where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T$ is the vector of perturbations such as $\omega_i \in \mathcal{R}$. In this case, the vector of no perturbation is given by $\boldsymbol{\omega}_0 = (0, \ldots, 0)^T$ and the perturbed penalized log-likelihood function is constructed from (2.7) with w_{id} replaced by $\mathrm{w}_{id\omega}$, that is,

$$L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega}) = L(\boldsymbol{\theta}|\boldsymbol{\omega}) - \sum_{k=1}^{s} \frac{\lambda_k}{2}\, \boldsymbol{\beta}_k^T \mathbf{K}_k \boldsymbol{\beta}_k \,, \tag{4.1}$$

where $L(\cdot)$ is given by (2.4) with $\mu_{i\omega} = g^{-1}(\eta_{i\omega})$ in the place of μ_i, for $\eta_{i\omega} = \mathbf{w}_{i\omega}^T \boldsymbol{\alpha} + \mathbf{n}_{1_i}^T \boldsymbol{\beta}_1 + \ldots + \mathbf{n}_{s_i}^T \boldsymbol{\beta}_s$. Differentiating $L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})$ with respect to the elements of $\boldsymbol{\theta}$ and

ω_i, we obtain, after some algebraic manipulation, that

$$\begin{aligned}
\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\omega}^T}\Big|_{\theta=\hat{\theta},\ \omega=\omega_0} &= \mathbf{e}_p\mathbf{a}^T - \boldsymbol{\alpha}_p\mathbf{W}^T\mathbf{D}_b\ ,\\
\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}_k\partial\boldsymbol{\omega}^T}\Big|_{\theta=\hat{\theta},\ \omega=\omega_0} &= \mathbf{e}_p\mathbf{a}^T - \boldsymbol{\alpha}_p\widetilde{\mathbf{N}}_k^T\mathbf{D}_b \qquad (k=1,\ldots,s) \qquad \text{and}\\
\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial\phi\partial\boldsymbol{\omega}}\Big|_{\theta=\hat{\theta},\ \omega=\omega_0} &= -\sum_{i=1}^{n}(\mathrm{a}_{ij}(\phi))^{-2}\left\{\mathrm{y}_{ij}\,\frac{\partial(\eta_{ijw})}{\omega_i} - \frac{\partial\psi(h(\eta_{ijw}))}{\omega_i}\right\}\mathbf{e}_{in}^T,
\end{aligned}$$

where $\mathbf{a}=(a_1,...,a_n)^T$, $\mathbf{D}_b=\mathrm{diag}_{1\le i\le n}\big(b_i\big)$ and $\mathbf{e}_p$ is a vector with 1 at the pth position and zero elsewhere with $a_i=(\mathrm{a}_{ij}(\phi))^{-1}(y_{ij}-\partial\psi/\partial h)\partial h/\partial\eta_{ij}$ and $b_i=(\mathrm{a}_{ij}(\phi))^{-1}(y_{ij}-\partial\psi/\partial h)\partial^2 h/\partial\eta_{ij}^2-\partial^2\psi/\partial^2 h(\partial h/\partial\eta_{ijw})^2$.

4.4 Final comments

In this chapter the method of local influence proposed by Cook (1986) was presented, for the partially varying-coefficient generalized linear models. In addition, the matrix derivatives of the observed information matrix I and of the matrix for the case-weigh schemes, explanatory variable perturbation and the response variable perturbation. In the next chapter we will present an illustrative example, in which the local influence methodology applied to partially varying-coefficient generalized linear models is developed.

Chapter 5

Application

In this chapter we present a practical application to exemplify the theory proposed in this work. We use data from a study of the relationship between atmospheric ozone concentration, O3 and other meteorological variables in the Los Angeles Basin in 1976, for a sample of 330 days. The data were first presented by Breiman and Friedman (1985). The variables are described below:

- O3: Concentration of ozone per hour in Upland, CA, measured in parts per million (ppm).
- VH: Pressure height 500 millibar, measured at the base of the air force of Vandenberg, in meters.
- WIND: Wind speed, in miles per hour.
- HUMIDITY: Humidity in percentage.
- TEMP: Sandburg Air Base temperature, in Celsius.
- IBH: Inversion base height, in foot.

- DPG: Dagget pressure gradient, in mmHg.
- IBT: Inversion base temperature, in Fahrenheit.
- VIS: Visibility, in miles.
- DAY: Calendar day.

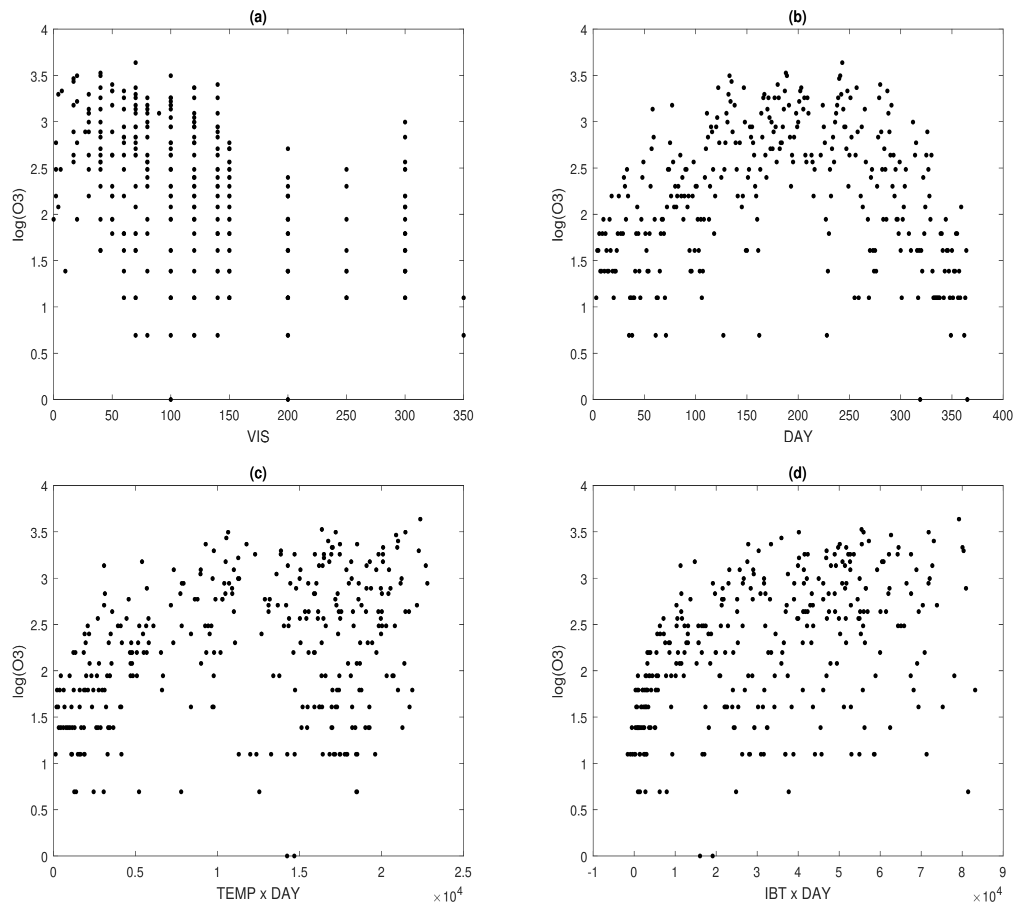

Figure 5.1: Scatter plots: O3 versus VIS (a), O3 versus DAY (b), O3 versus TEMP x DAY (c) and O3 versus IBT x DAY (d).

5.1 Fitting the model

In our application we will consider only four explanatory variables, specifically, the variables VIS, TEMP, IBT and DAY. Figure 5.1 contains the dispersion graphs between the outcome variable and each one of the explanatory variables. For simplicity, we will ignore that there is a temporal correlation for O3. We see in Fig. 5.1a that the relationship between O3 and the explanatory variable VIS is approximately linear, whereas the relationship between O3 and DAY appear in nonlinear ways (Fig. 5.1b). Note that there is a significant increase in the level of O3 from January to July with a decrease until December. This suggests that the incorporation of a quadratic or nonparametric term in the model could better explain the behavior of O3 over time. On other hand, Figs. 1c and 1d suggests that the explanatory variables TEMP and IBT might be interacting with the variable DAY in nonlinear fashion. Initially, we will adjust a generalized linear model assuming that the response variable O3 follows a Poisson distribution with mean μ_i and considering different structures of the linear predictor for the explanatory variables VIS, TEMP, IBT and DAY (see Table 5.1).

Table 5.1: Different structures of the linear predictor for the explanatory variables VIS, TEMP, IBT and DAY assuming that the response variable O3 $\sim$ Poisson(μ_i).

Model	$g(\mu_i) = \log(\mu_i)$
I	$\alpha_0 + \alpha_1\mathrm{VIS}_i + \alpha_2\mathrm{TEMP}_i + \alpha_3\mathrm{IBT}_i$
II	$\alpha_0 + \alpha_1\mathrm{VIS}_i + \alpha_2\mathrm{TEMP}_i + \alpha_3\mathrm{IBT}_i + \alpha_4\mathrm{DAY}_i$
III	$\alpha_0 + \alpha_1\mathrm{VIS}_i + \alpha_2\mathrm{TEMP}_i + \alpha_3\mathrm{IBT}_i + f(\mathrm{DAY}_i)$
IV	$\alpha_0 + \alpha_1\mathrm{VIS}_i + \alpha_2\mathrm{TEMP}_i + \alpha_3\mathrm{IBT}_i + \alpha_4\mathrm{DAY}_i + \alpha_5\mathrm{TEMP}_i \times \mathrm{DAY}_i + \alpha_6\mathrm{IBT}_i \times \mathrm{DAY}_i$
V	$\alpha_0 + \alpha_1\mathrm{VIS}_i + \mathrm{TEMP}_i\beta_1(\mathrm{DAY}_i) + \mathrm{IBT}_i\beta_2(\mathrm{DAY}_i)$

For model I, only the individual effect of the VIS, TEMP and IBT explanatory variables was considered. In model II, the individual effects of these three covariates plus the effect of the DAY variable were incorporated in a linear manner, whereas in the model III the individual effect of the DAY explanatory variable is included nolinearly by using smooth function. The model IV to consider the individual contributions of VIS, TEMP, IBT and DAY explanatory variables, plus the interaction effects of the TEMP and IBT explanatory variables with the DAY variable. Finally, the model V correspond to a VCGLM model where the explanatory variables TEMP and IBT interacts with the variable DAY in nonlinear fashion. The Table 2 contains the estimates of ML and MPL of the parameters associated with the parametric component of the five fitted models. Their respective standard errors appear in parentheses.

Table 5.2: Maximum likelihood and MPL estimates and the standard error (in parenthesis) for indicated model fitted to Ozone data.

	Model				
Parameters	I	II	III	IV	V
α_0	0.65 (0.11)	0.79 (0.11)	1.09 (0.16)	0.88 (0.21)	1.18 (0.17)
α_1	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
α_2	0.02 (0.00)	0.03 (0.00)	0.01 (0.00)	0.02 (0.00)	–
α_3	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	–
α_4	–	-0.001 (0.00)	–	-0.001 (0.00)	–
α_5	–	–	–	-0.00 (0.00)	–
α_6	–	–	–	0.00 (0.00)	–
AIC	1890.71	1861.27	1752.56	1863.66	1735.76

It should be noted that the p-value (omitted here) associated with each fitted model are less than 0.05, indicating that the contribution of the individuals and interaction effects between the explanatory variables is significant. Note also that the parameter estimates (associated with the parametric component) obtained from the different fitted models are quite similar and accurate. The last row of the table shows the AIC values obtained for each of the fitted models. It is clearly observed that VCGM model, for which the $\text{AIC}(\lambda_1, \lambda_2) = 1735.76$, presents a better fit to the Ozone data than the rest of the fitted models, followed by Model III with an AIC = 1752.56. Note that the quality of fit of these models is confirmed by the QQ-plot graphs presented in Figure 5.2 ; see Figures 5.2(c) and 5.2(e). For the VCGLM model the estimates of the smoothing parameters λ_1 and λ_2 as well as the corresponding df's were obtained by the procedure proposed by Ibacache-Pulgar et al. (2013), and are described in Table 5.3. The Figures 5.3(a) and 5.3(b) shows the estimated smooth functions under PVCGM model and the corresponding approximate SEB (dashed curves). Note that the graphical plots confirm the nonlinear trends of the interaction effects between $(\text{TEMP}, \text{DAY})$ and (IBT, DAY).

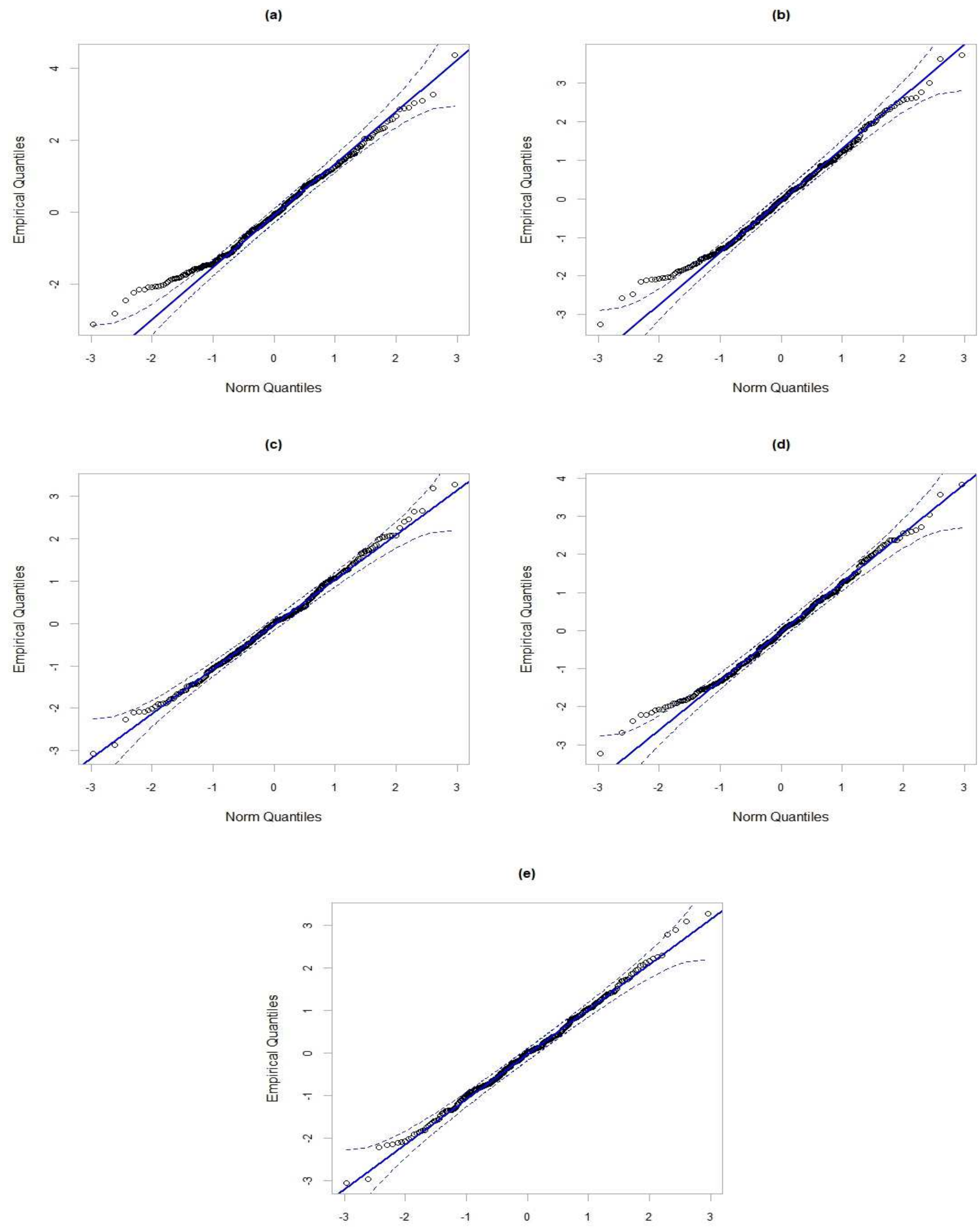

Figure 5.2: QQ-plot: Model I (a), Model II (b), Model III (c), Model IV (d) and Model V (e).

Table 5.3: Fit summary for smoothing components under PVCGLM fitted to data set.

	Smooth function	
	β_1(DAY)	β_2(DAY)
df(λ_k)	6.894	7.228
λ_k	89050.050	5886.339

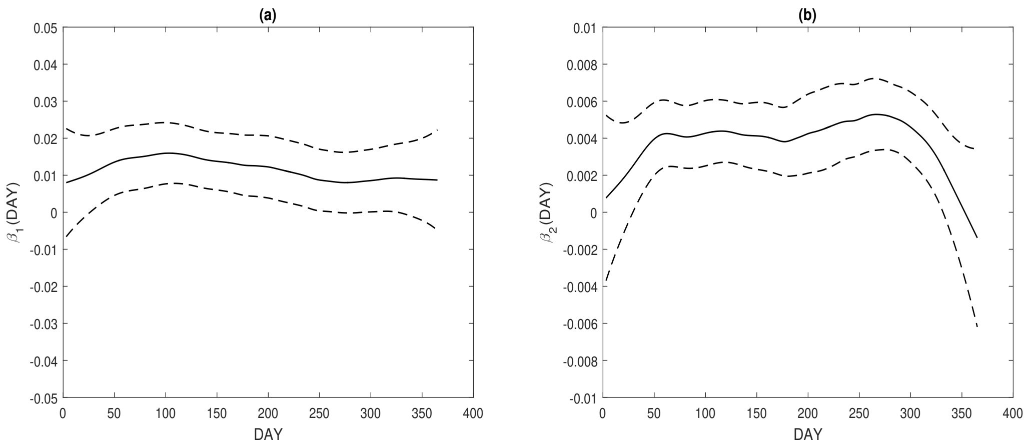

Figure 5.3: Plots estimated smooth functions β_1 (a) and β_2 (b) for the ozone data, and their approximate pointwise SEB denoted by the dashed lines.

5.2 Local influence analysis

Now, in order to identify influential potentially observations under the fitted model to ozone data, we present some index plots of $B_i = B_{\mathbf{e}_i}(\boldsymbol{\lambda})$, for $\boldsymbol{\lambda} = \boldsymbol{\alpha}, \boldsymbol{\beta}_k$ $(k = 1, 2)$.

5.2.1 Case-weight perturbation

Figure 5.4 shows the index plot B_i for the case-weight perturbation scheme under the fitted model. Looking at Figure 5.4, note that observations #125, #219, #167 and

#258 are more influential on the MPL estimate $\widehat{\boldsymbol{\alpha}}$, while observations #219, #221 and #222 are pointed as influential on MPL estimate $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$, respectively. When we introduce an additive perturbation on the response variable, the results are analogous to those observed under the case-weight perturbation scheme, and therefore the graphs are omitted.

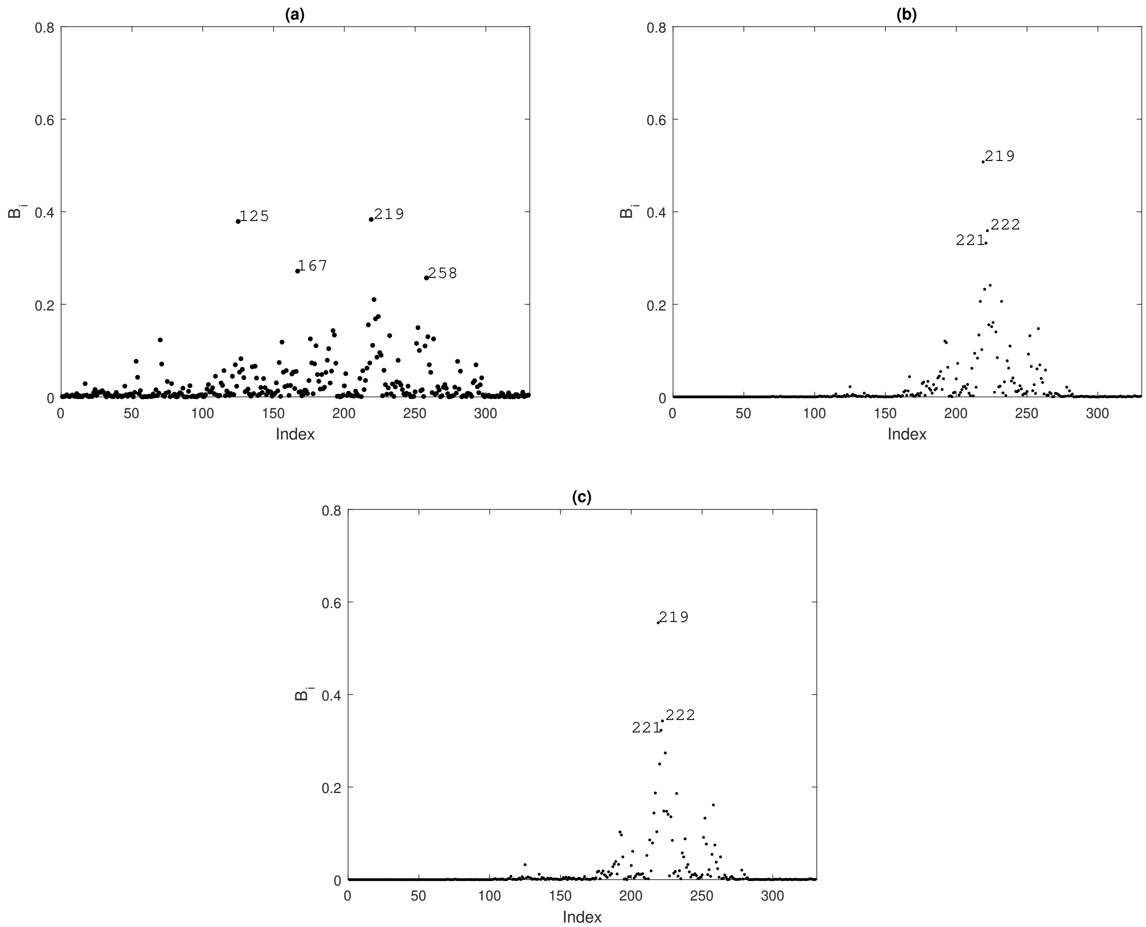

Figure 5.4: Index plots of $\mathbf{B}_i$ for assessing local influence on $\widehat{\boldsymbol{\alpha}}$ (a), $\widehat{\boldsymbol{\beta}}_1$ (b) and $\widehat{\boldsymbol{\beta}}_2$ (c) considering case-weight perturbation under model fitted to Ozone data.

5.2.2 Explanatory variable perturbation

Perturbing the explanatory variable in an additive way, it is observed that the observations #125, #219 and #167 are more influential on the MPL estimate $\widehat{\boldsymbol{\alpha}}$, while observations #219, #221 and #222 are pointed as influential on MPL estimate $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$, respectively.

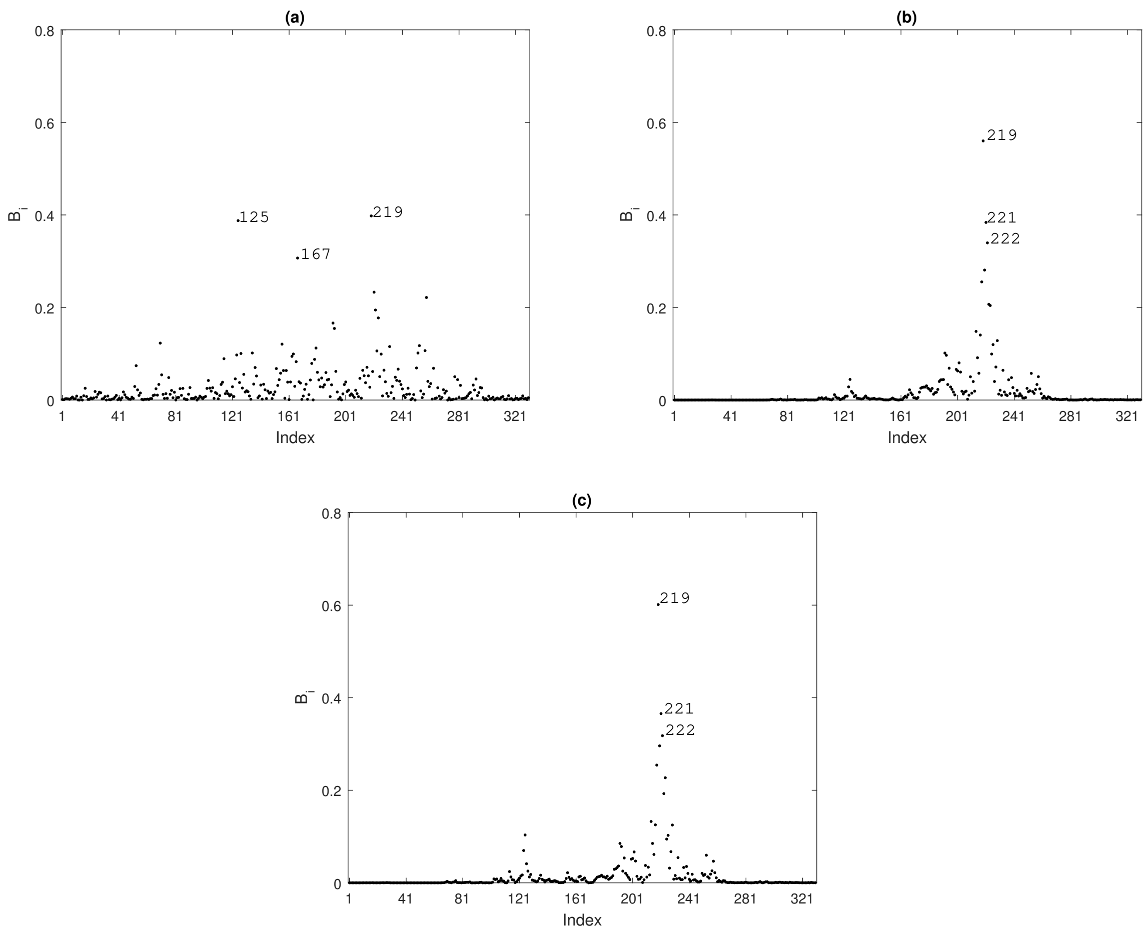

Figure 5.5: Index plots of $\mathbf{B}_i$ for assessing local influence on $\widehat{\boldsymbol{\alpha}}$ (a), $\widehat{\boldsymbol{\beta}}_1$ (b) and $\widehat{\boldsymbol{\beta}}_2$ (c) considering explanatory variable perturbation under fitted model to Ozone data.

Based on the analysis of local influence we can conclude that the MPLEs of the regression coefficient and of the coefficient functions are sensitive when perturbations are introduced into the data or the model. In addition, the analysis of local influence revealed that the observations that were detected as influential on the parametric component are not necessarily influential on the non-parametric component, and vice versa. For example, under the case-weight perturbation scheme, observations #125, #219, #167 and #258 were detected as influential on the parametric component. However, of these three observations, only observation #219 is indicated as influential on the non-parametric component, in addition to observations #221 and #222. The same results are obtained when the explanatory variable is perturbed additively.

5.3 Confirmatory analysis

Table 5.4 presents the relative changes in % (RC) of the MPLEs of α_j $(j = 1, 2)$ and ϕ after removing from the data set the pointed out observations in the local influence graphics under PVCGLM. The RC of each estimated parameter are defined as $\text{RC}_\psi = |(\widehat{\psi} - \widehat{\psi}_{(\text{I})})/\widehat{\psi}| \times 100\%$, where $\widehat{\psi}_{(\text{I})}$ denotes the MPLE of ψ, with $\psi = \alpha_j$, after the observation or observation(s) set(I). Table 5.4 presents the RCs in the regression coefficient estimate after removing the observations indicated as potentially influential on the parametric component of the model. On the other hand, Table 5.5 shows the RCs observed in the estimation of the regression coefficient once the observations detected as potentially influential on the non-parametric component of the model are excluded. Considering these results we can conclude that, although some RCs are large, inferential changes are not detected. It is interesting to notice from Tables 5.4 and 5.5 the coherence with the local influence diagnostic shown previously. For instance, elimination of the observations sets $\text{I} = \{167, 258\}$ and $\text{I} = \{125, 258\}$, which contain observations detected as influential potentially on parametric component, leads to significant changes in the MPL estimate, mainly in α_0, of the order of 29.245% and 28.823%, respectively; see Table 4. Note also that the individual elimination of observation #258 produces a relative change of the order of 2.096%. On the other hand, the elimination of the observations set $\text{I} = \{219, 222\}$, whose observations were detected as influential potentially on nonparametric component, leads to significant changes in the MPL estimate of α_0, of the order of 28.330%. It is also observed that the elimination of observation #222 produces a relative change of 30.949%. This indicates the need of a diagnostic examination.

Table 5.4: Relative changes (in %) on MPLE's of α_j under PVCGLM model fitted to Ozone data set.

	Parameters		Relative changes	
Dropped observation	α_0	α_1	RC_{α_0}	RC_{α_1}
125	1.17365	-0.001616	0.977	1.635
167	1.16798	-0.001592	1.455	0.125
219	1.18324	-0.001626	0.167	2.264
258	1.21007	-0.001622	2.096	2.013
125-167	1.17727	-0.001623	0.672	2.075
125-219	1.18273	-0.001628	0.211	2.389
125-258	1.52686	-0.001638	28.823	3.019
167-219	1.17701	-0.001603	0.694	0.817
167-258	1.53185	-0.001614	29.245	1.509
219-258	1.17689	-0.001609	0.703	1.195
125-167-219	1.15265	-0.001637	2.748	2.955
167-219-258	1.17625	-0.001593	0.758	0.189
125-167-219-258	1.51397	-0.001654	27.737	4.025

Table 5.5: Relative changes (in %) on MPLE's of α_j under PVCGLM model considering the observations detected as influential on the nonparametric component.

	Parameters		Relative changes	
Dropped observation	α_0	α_1	RC_{α_0}	RC_{α_1}
none	1.18	-0.001		
219	1.183	-0.002	0.167	2.264
221	1.161	-0.002	2.041	1.132
222	1.552	-0.002	30.949	2.075
219-221	1.142	-0.002	3.641	1.635
219-222	1.521	-0.002	28.330	4.779
221-222	1.151	-0.002	2.865	1.258
219-221-222	1.186	-0.002	0.092	2.955

Conclusions

In this work we study some aspects of the PVCGLM model. Specifically, we derive a weighted backfitting iterative process to estimate the parameters associated to linear predictor of the model (regression coefficient and smooth functions). We estimate the approximate variance-covariance matrix of MPLEs based in the Fisher information matrix obtained from the penalized log-likelihood function. Moreover, we have developed method of local influence to evaluate the sensitivity of the MPLEs by using several perturbation schemes in the model and/or data. Finally, we have performed a statistical modelling with real data set. The study provides evidences on the advantage of incorporating a semiparametric additive term in those situations where there are covariates whose interactions contribute nonlinearly to the model. Thus, we recommend PVCGLM models as an option to adjust data sets whose distribution of the response variable belongs to the exponential family and whose interactions between the explanatory variables can be modelled through smooth functions.

Bibliography

Berhane K., and Tibshirani, R. (1998). Generalized additive models for longitudinal data. *The Canadian Journal of Statistics*, 26, 517-535.

Billor, N. and Loynes, R. M. (1993). Local influence: a new approach. *Communications in Statistics - Theory and Methods*, 22, 1595-1611.

Breiman L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95, 888-902.

Cao, C.Z. and Lin, J.G. (2011). Diagnostics for elliptical linear mixed models with first-order autoregressive errors. *Journal of Statistical Computation and Simulation*, 81, 1281-1296.

Cook, R. D.(1986). Assessment of local influence (with discussion). *Journal Royal Statistics Society*, 48, 133-169.

Emmami, H. (2017). Local influence for Liu estimators in semiparametric linear models. *Statistical Papers*, 19, 529-544.

Escobar, L.A. and Meeker, W.Q. (1992). Assessing local influence in regression analysis with censored data. *Biometrics*, 48, 507-528.

Espinheria, P., Ferrari, S., and Cribari-Neto, F. (2008). On beta regression residuals. *Journal of Applied Statistics*, 35, 407-419.

Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11, 1031-1057.

Ferrari, S., Espinheira, P. and Cribari-Neto, F. (2011). Diagnostics tools in beta regression with varying dispersion. *Statistica Neerlandica*, 65, 337-351.

Ferreira, C.S. and Paula, G.A. (2017). Estimation and diagnostic for skew-normal partially linear models. *Journal of Applied Statistics*, 44, 3033-3053.

Green, P. J., and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, Boca Raton.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Statistical Society B*, 55, 757-796.

Hastie, T. and Tibshirani, R. (1990). Generalized additive models. *Statistical Science*, 3, 297-318.

Ibacache-Pulgar, G., Paula, G. A., and Galea, M. (2012). Influence diagnostics for elliptical semiparametric mixed models. *Statistical Modelling*, 12, 165-193.

Ibacache, G., Paula, G. A. and Cysneiros, F. (2013). Semiparametric additive models under symmetric distributions, *Test*, 22, 103-121.

Ibacache, G. and Paula, G. A. (2011). Local Influence for student-t partially linear models, *Computational Statistics and Data Analysis*, 55, 1462-1478.

Ibacache-Pulgar, G. and Reyes, S. (2018). Local influence for elliptical partially varying-coefficient model, *Statistical Modelling*, 18, 149-174.

Ibacache-Pulgar, G., Figueroa-Zuñiga, J. and Marchant, C. (2019). Semiparametric additive beta regression models: inference and local influence diagnostics. *REVSTAT*, to appear.

McCullagh, P. and J. A. Nelder (1989). Generalized linear models 2nd ed. London: Chapman and Hall.

Osorio, F., Paula, G.A. and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis*, 51, 4354-4368.

Poon, W., and Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society B*, 61, 51-61.

Ouwens, M. N. M., Tan, F. E. S. and Berger, M. P. F. (2001). Local influence to detect influential data structures for generalized linear mixed models. *Biometrics*, 57, 1166-1172.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Appl. Statist*, 54, 507-554.

Rocha, A. and Simas, A. (2011). Influence diagnostics in a general class of beta regression models. *TEST*, 20, 95-119.

Thomas, W. (1991). Influence diagnostics for the cross-validated smoothing parameter in spline smoothing. *Journal of the American Statistical Association*, 9, 693-698.

Uribe-Opazo, M.A., Borssoi, J.A. and Galea, M. (2012). Influence diagnostics in Gaussian spatial linear models. *Journal of Applied Statistics*, 3, 615-630.

Zhang, C., Mei, C. and Zhang, J. (2007). Influence diagnostics in partially varying-coefficient models. *Acta mathematica applicate*, 23, 619-628.

Zhang, J., Zhang, X., Ma, H., and Zhiya, C. (2015). Local influence analysis of varying coefficient linear model. *Journal of Interdisciplinary Mathematics*, 3, 293-306.

Zhu, H. and Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society B*, 63, 111-126.

Zhu, H. and Lee, S. (2003). Local influence for generalized linear mixed models. *The Canadian Journal of Statistics*, 31, 293-309.