# CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

A free open-access journal indexed by

Volume 12 Number 2
December 2021

ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society

SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

## Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, as well as in artificial intelligence, big data, data science, and machine learning, focused mainly on research articles. However, review, survey, and teaching papers, as well as material for statistical discussion, could be also published exceptionally. Each paper published in the ChJS must consider, in addition to its theoretical and/or methodological novelty, simulations for validating its novel theoretical and/or methodological proposal, as well as an illustration/application with real data.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

# Chilean Journal of Statistics

Volume 12, Number 2
December 2021

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Contents

*Chilean Journal of Statistics*
Vol. 12, No. 2, December 2021, 205–227

Statistical Modeling
Research Paper

# On an elliptical thin-plate spline partially varying-coefficient model

Magaly S. Moraga[1], Germán Ibacache-Pulgar[2,3,*], and Orietta Nicolis[4]

[1]Instituto de Estadística, Universidad Austral de Chile, Valdivia, Chile,
[2]Instituto de Estadística, Universidad de Valparaíso, Valparaíso, Chile,
[3]Centro Interdisciplinario de Estudios Atmosféricos y Astroestadística, Universidad de Valparaíso, Valparaíso, Chile,
[4]Facultad de la Ingeniería, Universidad Andrés Bello, Viña del Mar, Chile

(Received: 03 August 2021 · Accepted in final form: 03 November 2021)

**Abstract**

In this work, we study the thin-plate spline partially varying-coefficient models with elliptical contoured errors in order to allow distributions with heavier and lighter tails than the normal ones, such as logistic, Pearson VII, power exponential, and Student-t, to be considered. We develop an estimation process for the parameters of the model based on the doubly penalized likelihood function and using smoothing splines. In addition, an explicit conditional solution for the double penalized maximum likelihood estimators is derived to obtain closed expressions for the variance-covariance matrix of the estimators, effective degrees of freedom of the smooth functions and surfaces, and hat matrix associated with the model. To show the proposed methodology, we analyze the Boston housing data utilizing-plate spline partially varying-coefficient model with normal and Student-t errors. This analysis suggests that the proposed model is helpful when we want to describe the effect of some covariates that vary smoothly as a function of other covariates, geographic referencing, and data with heavy-tailed indications.

**Keywords:** Maximum doubly penalized likelihood estimates · Partially varying-coefficient models · Robust estimates · Thin-plate spline models

**Mathematics Subject Classification:** Primary 62J02 · Secondary 62J12.

## 1. Introduction

Partially varying-coefficient models have received much attention in various research areas, due to its flexibility to explore the dynamic features which may exist the data and its easy interpretation. In the others words, this class of models allows to model the coefficients of the explanatory variables (or covariates) as smooth functions of other variables. These models are often used in research related to longitudinal, clustered, spatial and hierarchical sampling schemes, and are a natural alternative to the additive model introduced by Breiman and Friedman (1985); see also Hastie and Tibshirani (1993), Fan and Zhang (2008) and Park et al. (2015).

* Corresponding author: gibacachepulgar@gmail.com

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Another aspect in the statistical literature, that has been developed in recent years, refers to the regression models under elliptical errors. These models suggest to replace the normal distribution by the elliptical one when the observations distributions are characterized by light-and heavy-tails. Savalli et al. (2006) proposed the elliptical linear mixed models, where the marginal model is also elliptical. Russo et al. (2009) extended the class given by Savalli et al. (2006) replacing linear fixed effects by a nonlinear fixed effect, creating the elliptical nonlinear mixed models, for which estimation procedures and diagnostic methods are developed. Galea and Vilca (2010) studied some hypothesis tests for the equality of variances and means in the context of univariate elliptical correlated data, with applications to portfolios data. Marciano et al. (2016) studied the calibration models for repeated measures considering a univariate elliptical distribution and developed a simulation study to evaluate the properties of the estimators. Ibacache-Pulgar and Paula (2011) presented a study on the existence and uniqueness of the maximum penalized likelihood estimate under the partially linear model with Student-t random error, Ibacache-Pulgar et al. (2012) developed influence diagnostics for elliptical semiparametric mixed models, where it is assumed that the non-parametric component is of type cubic spline, and Ibacache-Pulgar et al. (2013) studied semiparametric additive model under symmetric distributions. Recently, Ibacache-Pulgar and Reyes (2018) studied the elliptical partially varying-coefficient models and developed the technique of local influence to evaluate the sensitivity of the maximum penalized likelihood estimates.

In this paper, we extend the partially varying-coefficient model proposed by Ibacache-Pulgar and Reyes (2018) incorporating a component in its regression structure that allows us to model the effect of observations in two-dimensional space, such as, for example, coordinates. This structure is called thin-plate spline partially varying-coefficient model under elliptical errors. This model emerges 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 variables. Moreover, this class of models incorporate thin-plate spline (TPS) smoother, a spline-based technique which can be considered the natural generalization of cubic spline to any number of dimensions and almost any order of wiggliness penalty. The TPS smoother was initially introduced by Duchon (1975) and was later considered by many authors in the context of nonparametric and generalized linear models; see, Green and Silverman (1994) and Wood (2006) and the references therein. Since the TPS involves the estimation of many parameters (especially when the dimension is higher than one), Wood (2003) proposed a low rank smoother that use an approximate thin plate spline model based on the transformation and truncation of the basis that arises from the solution of the thin plate spline smoothing problem. The main advantage to include TSP in our model is that it allows to consider the effect of the geographical locations on the response variable.

This article is organized as follows. In Section 2, we formally introduce the thin-plate spline partially varying-coefficient model under elliptical distributions. Section 3 considers the problem of estimating the parameters and an application to a set of real data is considered in Section 4. Finally, in Section 5, we present some final conclusions derived from this study.

## 2. The thin-plate spline partially varying-coefficients model

In this section, we introduce the thin-plate spline partially varying-coefficient model (TPSPVCM) under elliptical distributions. In addition, we introduce the doubly penalized likelihood function where the penalty term combines a $\mathcal{L}^2[a,b]$ penalty for each smooth varying-coefficient function with a second $\mathcal{L}^2[E^d]$ penalty for the smooth surface. Thus, we estimate the parameters and inference in the elliptical TPSPVCM.

### 2.1 Model specification

The study of varying-coefficients models (VCMs) does not necessarily arise from performing a mathematical extension of a particular class of models, but rather from the need to attend to real problems in areas as economics, finance, epidemiology, medical science, ecology, and environment. The TPSPVCM under study is given by

$$y_{ij} = \boldsymbol{z}_{ij}^{\top}\boldsymbol{\alpha} + \sum_{k=1}^{s} x_{ij}^{(k)}\,\beta_k(r_{k_{ij}}) + \boldsymbol{\ell}_i^{\top}\boldsymbol{g} + \varepsilon_{ij}, \quad i = 1,\ldots,n, j = 1,\ldots,m_i, \tag{1}$$

where $y_{ij}$ denotes the $j$th measure associated with the $i$th cluster at point $r_{k_{ij}}$, $\boldsymbol{z}_{ij}$ is $(p\times 1)$ vector of explanatory variable values, $\boldsymbol{\alpha}$ is a $(p\times 1)$ fixed parameter vector, $\beta_k$, for $k = 1,\ldots,s$, are unknown smooth arbitrary functions of $r_k$, associated with the covariates $x_{ij}^{(k)}$, $\boldsymbol{\ell}_i$ is an $(n\times 1)$ vector with one in the $i$th position and zeros at the remaining positions, $\boldsymbol{g} = (g(\boldsymbol{t}_1),\ldots,g(\boldsymbol{t}_n))^{\top}$, $g$ is a smooth surface that depends of the vector $\boldsymbol{t}_i \in \mathcal{R}^2$, and $\varepsilon_{ij}$ is a random error. Note that in this class of models the coefficients are allowed to vary as smooth functions of other variables.

To write the model given in Equation (1) in a matrix form, first consider the one-to-one linear transformation of the vector $\boldsymbol{g}$ suggested by Green and Silverman (1994) stated as

$$\boldsymbol{g} = \begin{pmatrix} g(\boldsymbol{t}_1) \\ \vdots \\ g(\boldsymbol{t}_n) \end{pmatrix} = \boldsymbol{E}\boldsymbol{\delta} + \boldsymbol{T}^{\top}\boldsymbol{a},$$

where $\boldsymbol{a}$ and $\boldsymbol{\delta}$ are vectors with components $a_i$ and $\delta_i$, $\boldsymbol{E}$ is an $(n\times n)$ matrix defined by $E_{ij} = 1/(16\pi\|\boldsymbol{t}_i - \boldsymbol{t}_j\|^2\log(\|\boldsymbol{t}_i - \boldsymbol{t}_j\|^2))$, with $E_{ii} = 0$ for each $i$, and $\boldsymbol{T}$ is a $(3\times n)$ matrix given by

$$\boldsymbol{T} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \boldsymbol{t}_1 & \boldsymbol{t}_2 & \ldots & \boldsymbol{t}_n \end{pmatrix}.$$

Thus, the model model given in Equation (1) takes the form

$$\boldsymbol{y}_i = \widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{\alpha}} + \sum_{k=1}^{s}\widetilde{\boldsymbol{N}}_{ki}\boldsymbol{\beta}_k + \widetilde{\boldsymbol{E}}_i\boldsymbol{\delta} + \boldsymbol{\varepsilon}_i, \quad i = 1,\ldots,n, j = 1,\ldots,m_i, \tag{2}$$

where $\boldsymbol{y}_i$ is a $(m_i\times 1)$ random vector of observed responses from the $i$th cluster, $\widetilde{\boldsymbol{Z}}_i = (\boldsymbol{Z}_i\ \ \widetilde{\boldsymbol{T}}_i)$ is an $(m_i\times(p+2))$ design matrix, $\boldsymbol{Z}_i$ is an $(m_i\times p)$ design matrix with rows $\boldsymbol{z}_{ij}^{\top}$, $\widetilde{\boldsymbol{T}}_i = \boldsymbol{F}_i\boldsymbol{T}^{\top}$ is an $(m_i\times 2)$ matrix, $\boldsymbol{F}_i$ is an $(m_i\times n)$ matrix with an $(m_i\times 1)$ vector of ones in the $i$th column and zeros in the remanning positions, $\widetilde{\boldsymbol{\alpha}}^{\top} = (\boldsymbol{\alpha}^{\top},\boldsymbol{a}^{\top})$, $\widetilde{\boldsymbol{N}}_{ki} = \boldsymbol{X}_i^{(k)}\boldsymbol{N}_{ki}$, $\boldsymbol{X}_i^{(k)} = \mathrm{diag}_{1\le j\le m_i}\left(x_{ij}^{(k)}\right)$, $\boldsymbol{N}_{ki}$ is an $(m_i\times r_k)$ incidence matrix with the $(j,l)$th element equal to the indicator $I(r_{k_{ij}} = r_{k_l}^0)$, for $j = 1,\ldots,m_i$, where $r_{k_l}^0$, for $l = 1,\ldots,r_k$, denotes the distinct and ordered values of the explanatory variable $r_{k_{ij}}$, $\boldsymbol{\beta}_k = (\psi_{k_1},\ldots,\psi_{r_k})^{\top}$ is an $(r_k\times 1)$ vector of parameters with $\psi_{k_l} = \beta_k(r_{k_l}^0)$, for $l = 1,\ldots,r_k$, $\widetilde{\boldsymbol{E}}_i = \boldsymbol{F}_i\boldsymbol{E}$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1},\ldots,\varepsilon_{m_i})^{\top}$ is an $(m_i\times 1)$ vector of within-cluster errors. A compact way of writing model given in Equation (2) is formulated as

$$\boldsymbol{y} = \widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{\alpha}} + \widetilde{\boldsymbol{N}}_1\boldsymbol{\beta}_1 + \cdots + \widetilde{\boldsymbol{N}}_s\boldsymbol{\beta}_s + \widetilde{\boldsymbol{E}}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \tag{3}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1^{\top},\ldots,\boldsymbol{y}_n^{\top})^{\top}$, $\widetilde{\boldsymbol{Z}}$, $\widetilde{\boldsymbol{N}}_k$, $\widetilde{\boldsymbol{E}}$ and $\boldsymbol{\varepsilon}$ similarly.

## 2.2 Doubly penalized likelihood function

Consider the model given by Equation (2) and assume that $\boldsymbol{\varepsilon}_i \sim \mathrm{El}_{m_i}\big(\mathbf{0}, \boldsymbol{\Sigma_i}\big)$, with $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\tau})$ being a positive-definite matrix, with $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_d)^\top$. Thus, $\boldsymbol{y}_i \sim \mathrm{El}_{m_i}\big(\boldsymbol{\mu}_i, \boldsymbol{\Sigma_i}\big)$, with $\boldsymbol{\mu}_i = \widetilde{\boldsymbol{Z}}_i\widetilde{\boldsymbol{\alpha}} + \sum_{k=1}^{s} \widetilde{\boldsymbol{N}}_{ki}\boldsymbol{\beta}_k + \widetilde{\boldsymbol{E}}_i\boldsymbol{\delta}$, and density function stated as

$$f(\boldsymbol{y}_i) = |\boldsymbol{\Sigma}_i|^{-1/2}\, h(u_i), \quad i = 1, \ldots, n, \tag{4}$$

where $u_i = \boldsymbol{\varepsilon}_i^\top \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i$ is the Mahalanobis distance, $\boldsymbol{\varepsilon}_i = \boldsymbol{y}_i - \boldsymbol{\mu}_i$, and $h$ is a function of $\mathcal{R} \to [0, \infty]$ known as the density generator function (Fang et al., 1990). Then, the log-likelihood function of the model given in Equation (4) for $\boldsymbol{\theta} = (\widetilde{\boldsymbol{\alpha}}^\top, \boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_s^\top, \boldsymbol{\delta}^\top, \boldsymbol{\tau}^\top)^\top$ is given by

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

where $L_i(\boldsymbol{\theta}) = -(1/2)\log(|\boldsymbol{\Sigma}_i|) + \log(h(u_i))$ represents the individual contribution of the $i$th observation. Since the functions $\beta_k$ belong to the infinite dimensional space and are considered parameters with respect to the expected value of $y_i$, some restricted subspace should be defined for the nonparametric functions to ensure identifiability of the parameters associated with model. Therefore, we assume that the functions $\beta_k$ (which are absolutely continuous) belong to the Sobolev function space stated as

$$\mathcal{W}_2^{(\imath)} = \{\beta_k\colon \beta_k, \beta_k^{(1)}, \ldots, \beta_k^{(\imath-1)}, \beta_k^{(\imath)} \in \mathcal{L}^2[a_k, b_k]\}\,.$$

In addition, we assume that $g$ belong to the functions space whose partial derivatives of total order $m$ are in Hilbert space $\mathcal{L}^2[E^d]$ of square integrable functions on Euclidean $d$-space. Incorporating a penalty function over each function $\beta_k$ and $g$, we have that the penalized log-likelihood function can be expressed as (Ibacache-Pulgar et al., 2013)

$$L_{\mathrm{p}}(\boldsymbol{\theta}, \lambda_1, \ldots, \lambda_s, \lambda_g) = L(\boldsymbol{\theta}) + \sum_{k=1}^{s} \lambda_k^* J(\beta_k) + \lambda_g^* J_m^d(g), \tag{5}$$

where $J(\beta_k)$ denotes the penalty functional over $\beta_k$, $J_m^d(g)$ is a penalty functional measuring the wiggliness of $g$, and $\lambda_k^* = \lambda^*(\lambda_k)$ and $\lambda_g^*(\lambda_g)$ are constants that depends on the smoothing parameters $\lambda_k \geq 0$ and $\lambda_g \geq 0$, respectively. In this paper, we consider as a measure of the curvature of $\beta_k$ functions the squared norm expressed as

$$J(\beta_k) = \|\beta_k\|^2 = \int_{a_k}^{b_k} [\beta_k^{(\imath)}(r_k)]^2 \mathrm{d}r_k,$$

where $\beta_k^{(\imath)}(r_k) = \mathrm{d}^\imath \beta(r_k)/dr_k{}^\imath$, $r_{k_l}^0 \in [a_k, b_k]$, and

$$J_m^d(g) = \sum_{\upsilon_1+\cdots+\upsilon_d=m} \frac{m!}{\upsilon_1! \cdots \upsilon_d!} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(\frac{\partial^m g}{\partial t_1^{\alpha_1} \ldots \partial t_d^{\alpha_d}}\right)^2 \prod_{j=1}^{d} \mathrm{d}t_j.$$

It is important mention that for $\imath = 2$, the estimation of $\beta_k$ leads to a natural cubic spline with knots at the points $r_{k_l}^0$, for $l = 1, \ldots, r_k$. In addition, for $d = 2, m = 2$ and

$g = g(t_1, t_2)$, that is,

$$J(g) = \int\int_{\mathcal{R}^2} \left\{ \left(\frac{\partial^2 g}{\partial t_1^2}\right)^2 + 2\left(\frac{\partial^2 g}{\partial t_1 \partial t_2}\right)^2 + \left(\frac{\partial^2 g}{\partial t_2^2}\right)^2 \right\} \mathrm{d}t_1 \mathrm{d}t_2,$$

the estimation of $g$ leads to a natural thin-plate spline. According to Green and Silverman (1994), we may express the penalty functional as

$$J(\beta_k) = \boldsymbol{\beta}_k^\top \boldsymbol{K}_k \boldsymbol{\beta}_k, \quad J(g) = \boldsymbol{\delta}^\top \boldsymbol{E} \boldsymbol{\delta},$$

where $\boldsymbol{K}_k$ is an $(q_k \times q_k)$ non-negative definite smoothing matrix associated with the $k$th explanatory variable that depends only on the knots. Then, if we consider $\lambda_k^* = -\lambda_k/2$ and $\lambda_g^* = -\lambda_g/2$, the penalized log-likelihood function given in Equation (5) can be expressed as

$$L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = L(\boldsymbol{\theta}) - \sum_{k=1}^{s} \frac{\lambda_k}{2} \boldsymbol{\beta}_k^\top \boldsymbol{K}_k \boldsymbol{\beta}_k - \frac{\lambda_g}{2} \boldsymbol{\delta}^\top \boldsymbol{E} \boldsymbol{\delta}, \tag{6}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_s, \lambda_g)^\top$ denotes an $((s+1) \times 1)$ vector of smoothing parameters that controls the tradeoff between goodness of fit and the smoothness estimated functions. Note that the first term in the right-hand side of Equation (6) measures the goodness of fit while the second and third terms penalizes the roughness of each $\beta_k$ and $g$ with a fixed parameter $\lambda_k$ and $\lambda_g$, respectively. It should be noted that the choice of such parameters is crucial in the estimation process, since they controls the tradeoff between goodness of fit and the smoothness (regularity) estimated function. A more extensive discussion on the methods of selecting such parameters is presented later.

## 3. Parameters estimation

The estimation problem in the context of TPSPVCM under elliptical distributions has not been discussed in the literature. However, several authors have considered this problem for some specific cases. For example, in the context of varying-coefficient model, Cai et al. (2000) estimated the coefficient functions based on local polynomial regression technique and proposed a method that involves solving hundreds of local likelihood equations through a one-one-step Newton-Raphson. Chiang et al. (2001) derived a componentwise smoothing spline procedure for the estimation of coefficient curves in a varying-coefficient model with repeatedly measured dependent variables; see also Eubank et al. (2004). Krafty et al. (2008) developed an estimation procedure of the coefficient functions when the within-subject covariance is unknown considering the criterion of iterative reweighted least squares. Wang et al. (2009) proposed an estimation method based on local ranks which is more efficient and robust compared to other methods such as local linear least squares method. Liu and Li (2015) estimated the coefficient curves in a varying-coefficient model for longitudinal data by using local polynomial smoothing method and showed that the resulting estimator is asymptotically more efficient than the ones which ignore the within-subject correlation structure. In this paper we propose to estimate the model parameters based on the work proposed by Ibacache-Pulgar and Reyes (2018), which consider to estimate the coefficient curves based on penalized likelihood criterion and smoothing spline.

### 3.1 Estimation of $\widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s, \boldsymbol{\delta}$

To estimate the parameters $\widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s, \boldsymbol{\delta}$ and $\boldsymbol{\tau}$ we propose to maximize the double penalized log-likelihood function assuming $\boldsymbol{\lambda}$ fixed, that is,

$$\max_{\widetilde{\alpha},\beta_1,\ldots,\beta_s,\delta,\tau} L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \max_{\widetilde{\alpha},\beta_1,\ldots,\beta_s,\delta,\tau} L_{\mathrm{p}}(\widetilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s, \boldsymbol{\delta}, \boldsymbol{\tau}, \boldsymbol{\lambda}).$$

This procedure can be solved using the Fisher scoring algorithm (Ibacache-Pulgar and Reyes, 2018) stated as

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{S}_0^{(u)}\widetilde{\boldsymbol{N}}_1 & \ldots & \boldsymbol{S}_0^{(u)}\widetilde{\boldsymbol{N}}_s & \boldsymbol{S}_0^{(u)}\widetilde{\boldsymbol{E}} \\ \boldsymbol{S}_1^{(u)}\widetilde{\boldsymbol{N}}_0 & \boldsymbol{I} & \ldots & \boldsymbol{S}_1^{(u)}\widetilde{\boldsymbol{N}}_s & \boldsymbol{S}_1^{(u)}\widetilde{\boldsymbol{E}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{S}_s^{(u)}\widetilde{\boldsymbol{N}}_0 & \boldsymbol{S}_s^{(u)}\widetilde{\boldsymbol{N}}_1 & \ldots & \boldsymbol{I} & \boldsymbol{S}_s^{(u)}\widetilde{\boldsymbol{E}} \\ \boldsymbol{S}_\delta^{(u)}\widetilde{\boldsymbol{N}}_0 & \boldsymbol{S}_\delta^{(u)}\widetilde{\boldsymbol{N}}_1 & \ldots & \boldsymbol{S}_\delta^{(u)}\widetilde{\boldsymbol{N}}_s & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_0^{(u+1)} \\ \boldsymbol{\beta}_1^{(u+1)} \\ \vdots \\ \boldsymbol{\beta}_s^{(u+1)} \\ \boldsymbol{\delta}^{(u+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_0^{(u)}\boldsymbol{\eta}^{(u)} \\ \boldsymbol{S}_1^{(u)}\boldsymbol{\eta}^{(u)} \\ \vdots \\ \boldsymbol{S}_s^{(u)}\boldsymbol{\eta}^{(u)} \\ \boldsymbol{S}_\delta^{(u)}\boldsymbol{\eta}^{(u)} \end{pmatrix}, \tag{7}$$

where $\boldsymbol{\beta}_0 = \widetilde{\boldsymbol{\alpha}}$, $\widetilde{\boldsymbol{N}}_0 = \widetilde{\boldsymbol{Z}}$, $\boldsymbol{\eta}^{(u)} = \boldsymbol{\mu} + \boldsymbol{W}^{*^{-1}}\boldsymbol{W}_v(\boldsymbol{y} - \boldsymbol{\mu})\big|_{\boldsymbol{\theta}^{(u)}}$ and $\boldsymbol{S}_k^{(u)} = \boldsymbol{S}_k\big|_{\boldsymbol{\theta}^{(u)}}$, with

$$\boldsymbol{S}_k^{(u)} = \begin{cases} (\widetilde{\boldsymbol{N}}_0^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_0)^{-1} \widetilde{\boldsymbol{N}}_0^\top \boldsymbol{W}^*\big|_{\boldsymbol{\theta}^{(u)}}, & k = 0, \\ (\widetilde{\boldsymbol{N}}_k^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_k + \lambda_k \boldsymbol{K}_k)^{-1} \widetilde{\boldsymbol{N}}_k^\top \boldsymbol{W}^*, & k = 1, \ldots, s, \end{cases}$$

and

$$\boldsymbol{S}_\delta^{(u)} = (\widetilde{\boldsymbol{E}}^\top \boldsymbol{W}^* \widetilde{\boldsymbol{E}} + \lambda_g \widetilde{\boldsymbol{E}})^{-1} \widetilde{\boldsymbol{E}}^\top \boldsymbol{W}^*,$$

where $\boldsymbol{W}^*$ and $\boldsymbol{W}_v$ are defined in the appendix. Then, the back-fitting (Gauss-Seidel) iterations that are used to solve the system stated in Equation (7) take the form

$$\boldsymbol{\beta}_k^{(u+1)} = \boldsymbol{S}_k^{(u)}\left(\boldsymbol{\eta}^{(u)} - \sum_{l=0,l\neq k}^{s} \widetilde{\boldsymbol{N}}_l \boldsymbol{\beta}_l^{(u)} - \widetilde{\boldsymbol{E}}\boldsymbol{\delta}^{(u)}\right), \quad k = 0, 1, \ldots, s, \tag{8}$$

$$\boldsymbol{\delta}^{(u+1)} = \boldsymbol{S}_\delta^{(u)}\left(\boldsymbol{\eta}^{(u)} - \sum_{l=0}^{s} \widetilde{\boldsymbol{N}}_l \boldsymbol{\beta}_l^{(u)}\right). \tag{9}$$

From the convergence of the iterative process given in Equation (8), we obtain the maximum double penalized likelihood estimator (MDPLE) of $\boldsymbol{\beta}_k$ and $\boldsymbol{\delta}$, which leads to a natural cubic spline estimate for $\beta_k$ $(k = 1, \ldots, s)$.

It is important to note that in the iterative process above the parameter estimates depend on the smoothing matrices $\boldsymbol{S}_k$ and $\boldsymbol{S}_\delta$, the modified variable $\boldsymbol{\eta}$ and the partial residuals; see Equation 8. In addition, the weights $v_i$ have an influence on the estimates of $\boldsymbol{\beta}_k$, for $k = 0, 1, \ldots, s$, and $\boldsymbol{\delta}$. In particular, it can be shown that for the Student-t and power exponential distributions, for example, the current weight $v_i^{(r)} = v_i\big|_{\boldsymbol{\theta}^{(r)}}$ is inversely proportional to the Mahalanobis distance between the observed value $\boldsymbol{y}_i$ and its current predicted value $\boldsymbol{\mu}_i^{(r)} = \boldsymbol{\mu}_i\big|_{\boldsymbol{\theta}^{(r)}}$, so that outlying observations tend to have small weights in the estimation process.

### 3.2 Estimation of $\boldsymbol{\tau}$

Regarding the MDPLE of $\boldsymbol{\tau}$, this can be obtained using the Fisher scoring algorithm formulated as

$$\boldsymbol{\tau}^{(u+1)} = \boldsymbol{\tau}^{(u)} - \mathrm{E}\left\{\frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}^\top}\right\}^{-1}\frac{\partial L_{\mathrm{p}}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\tau}}\Big|_{\boldsymbol{\theta}^{(u)}}. \tag{10}$$

An iterative process to solve Equations (8) and (10) simultaneously is described in the appendix.

### 3.3 Estimation of the surface

In Section 2, we represent the surface $g$ as a linear combination of the coefficient vectors $\boldsymbol{\delta}$ and $\boldsymbol{a}$. Considering the MDPLEs obtained through the iterative process described above, that is, $\widehat{\boldsymbol{\delta}}$ and $\widehat{\boldsymbol{a}}$, we have that the MDPLE, $\widehat{\boldsymbol{g}}$, can be obtained as

$$\widehat{\boldsymbol{g}} = \boldsymbol{E}\widehat{\boldsymbol{\delta}} + \boldsymbol{T}^\top\widehat{\boldsymbol{a}}. \tag{11}$$

Consequently, the estimator of the surface $g$ is a natural thin-plate spline. Details of the conditions that guarantee this result are given, for example, in Green and Silverman (1994).

### 3.4 A conditional explicit solution

Note that $\boldsymbol{\beta}_\jmath$, for $\jmath = 0, 1, \ldots, s$, and $\boldsymbol{\delta}$ can be estimated through the solutions to the set of normal equations (Buja et al., 1989; Opsomer and Ruppert, 1999) derived from double penalized log-likelihood function. Indeed, taking partial derivatives of Equation (6) with respect to the parameter $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s, \boldsymbol{\delta}$ and to equating zero, we obtain

$$\left.\begin{aligned}
\boldsymbol{Z}^\top\boldsymbol{W}_v\boldsymbol{\varepsilon} &= \boldsymbol{0},\\
\widetilde{\boldsymbol{T}}\boldsymbol{W}_v\boldsymbol{\varepsilon} &= \boldsymbol{0},\\
\widetilde{\boldsymbol{N}}_1^\top\boldsymbol{W}_v\boldsymbol{\varepsilon} - \lambda_1\boldsymbol{K}_1\boldsymbol{\beta}_1 &= \boldsymbol{0},\\
&\ \ \vdots\\
\widetilde{\boldsymbol{N}}_s^\top\boldsymbol{W}_v\boldsymbol{\varepsilon} - \lambda_1\boldsymbol{K}_s\boldsymbol{\beta}_s &= \boldsymbol{0},\\
\widetilde{\boldsymbol{E}}\boldsymbol{W}_v\boldsymbol{\varepsilon} - \lambda_g\boldsymbol{E}\boldsymbol{\delta} &= \boldsymbol{0}.
\end{aligned}\right\} \tag{12}$$

From Equation (12) it is possible, at least conceptually, to derive an explicit expression for the estimates $\widehat{\boldsymbol{\beta}}_\jmath$ ($\jmath = 0, 1, \ldots, s$) and $\boldsymbol{\delta}$ under some assumptions. For simplicity of notation consider $\boldsymbol{\beta}_{s+1} = \boldsymbol{\delta}$, $\boldsymbol{S}_{s+1} = \boldsymbol{S}_\delta$ and $p' = p + 2$, and assume $\boldsymbol{\lambda}$, $\boldsymbol{W}_v$ and $\boldsymbol{W}^*$ fixed, we can write the estimating equation system given in Equation (12) as

$$\begin{pmatrix}
\boldsymbol{I}_{p'} & \boldsymbol{S}_0\widetilde{\boldsymbol{N}}_1 & \ldots & \boldsymbol{S}_0\widetilde{\boldsymbol{N}}_s & \boldsymbol{S}_0\widetilde{\boldsymbol{E}}\\
\boldsymbol{S}_1\widetilde{\boldsymbol{N}}_0 & \boldsymbol{I}_{r_1} & \ldots & \boldsymbol{S}_1\widetilde{\boldsymbol{N}}_s & \boldsymbol{S}_1\widetilde{\boldsymbol{E}}\\
\vdots & \vdots & \ddots & \vdots & \vdots\\
\boldsymbol{S}_s\widetilde{\boldsymbol{N}}_0 & \boldsymbol{S}_s\widetilde{\boldsymbol{N}}_1 & \ldots & \boldsymbol{I}_{r_s} & \boldsymbol{S}_s\widetilde{\boldsymbol{E}}\\
\boldsymbol{S}_{s+1}\widetilde{\boldsymbol{N}}_0 & \boldsymbol{S}_{s+1}\widetilde{\boldsymbol{N}}_1 & \ldots & \boldsymbol{S}_{s+1}\widetilde{\boldsymbol{N}}_s & \boldsymbol{I}_n
\end{pmatrix}
\begin{pmatrix}\boldsymbol{\beta}_0\\ \boldsymbol{\beta}_1\\ \vdots\\ \boldsymbol{\beta}_s\\ \boldsymbol{\beta}_{s+1}\end{pmatrix}
=
\begin{pmatrix}\boldsymbol{S}_0\\ \boldsymbol{S}_1\\ \vdots\\ \boldsymbol{S}_s\\ \boldsymbol{S}_{s+1}\end{pmatrix}\boldsymbol{y}.$$

In practice, this system of equations is solved iteratively through a backfitting algorithm, and its backfitting estimators converge to the solution (Buja et al., 1989) stated as

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_0, \\ \widehat{\boldsymbol{\beta}}_1, \\ \vdots \\ \widehat{\boldsymbol{\beta}}_s, \\ \widehat{\boldsymbol{\beta}}_{s+1}, \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_{p'} & \boldsymbol{S}_0\widetilde{\boldsymbol{N}}_1 & \dots & \boldsymbol{S}_0\widetilde{\boldsymbol{N}}_s & \boldsymbol{S}_0\widetilde{\boldsymbol{E}} \\ \boldsymbol{S}_1\widetilde{\boldsymbol{N}}_0 & \boldsymbol{I}_{r_1} & \dots & \boldsymbol{S}_1\widetilde{\boldsymbol{N}}_s & \boldsymbol{S}_1\widetilde{\boldsymbol{E}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{S}_s\widetilde{\boldsymbol{N}}_0 & \boldsymbol{S}_s\widetilde{\boldsymbol{N}}_1 & \dots & \boldsymbol{I}_{r_s} & \boldsymbol{S}_s\widetilde{\boldsymbol{E}} \\ \boldsymbol{S}_{s+1}\widetilde{\boldsymbol{N}}_0 & \boldsymbol{S}_{s+1}\widetilde{\boldsymbol{N}}_1 & \dots & \boldsymbol{S}_{s+1}\widetilde{\boldsymbol{N}}_s & \boldsymbol{I}_n \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{S}_0 \\ \boldsymbol{S}_1 \\ \vdots \\ \boldsymbol{S}_s \\ \boldsymbol{S}_{s+1} \end{pmatrix} \boldsymbol{y} \equiv \boldsymbol{M}^{-1}\boldsymbol{\mathcal{S}}\boldsymbol{y}\,,$$

if the inverse of $\boldsymbol{M}$ exists. Consequently, the backfitting estimator for $\widehat{\boldsymbol{\beta}}_\jmath$ ($\jmath = 0, 1, \dots, s{+}1$) can be obtained directly as (Opsomer and Ruppert, 1999)

$$\widehat{\boldsymbol{\beta}}_\jmath = \boldsymbol{\mathcal{H}}_\jmath \boldsymbol{y}, \quad \jmath = 0, 1, \dots, s+1, \tag{13}$$

where $\boldsymbol{\mathcal{H}}_\jmath = \boldsymbol{E}_\jmath \boldsymbol{M}^{-1}\boldsymbol{\mathcal{S}}$ is the smoother matrix obtained when fitting by smoothing spline the $\jmath$th explanatory variable only, with $\boldsymbol{E}_\jmath$ being a partitioned matrix given by

$$\boldsymbol{E}_\jmath = \begin{cases} \left(\boldsymbol{I}_{(p'\times p')}\boldsymbol{0}_{(p'\times r_1)}\dots\boldsymbol{0}_{(p'\times r_s)}\boldsymbol{0}_{(p'\times n)}\right), & \jmath = 0, \\ \left(\boldsymbol{0}_{(r_1\times p')}\boldsymbol{I}_{(r_1\times r_1)}\dots\boldsymbol{0}_{(r_1\times r_s)}\boldsymbol{0}_{(r_1\times n)}\right), & \jmath = 1, \\ \qquad\qquad\vdots & \quad\vdots \\ \left(\boldsymbol{0}_{(r_s\times p')}\boldsymbol{0}_{(r_s\times r_1)}\dots\boldsymbol{I}_{(r_s\times r_s)}\boldsymbol{I}_{(r_s\times n)}\right), & \jmath = s, \\ \left(\boldsymbol{0}_{(n\times p')}\boldsymbol{0}_{(n\times r_1)}\dots\boldsymbol{I}_{(n\times r_s)}\boldsymbol{I}_{(n\times n)}\right), & \jmath = s+1\,. \end{cases}$$

The direct calculation of the MDPLEs from Equation (13) is rarely used in practice, because the backfitting algorithm is more efficient for obtaining $\widehat{\boldsymbol{\beta}}_\jmath$; it does not require high-dimensional matrices and their inverses. However, the above expressions can be useful if we wish to study some theoretical properties of the MDPLEs and carry out a diagnostic analysis based on the hat matrix associated with the model fit. Some closed expressions for the estimators in the context of the semiparametric additive models can be found, for example, in Ibacache-Pulgar et al. (2013).

### 3.5 Estimation of the standard errors

We consider in this section the problem of how to derive the variance-covariance matrix of the MDPLE $\boldsymbol{\theta}$. According to Segal et al. (1994), the variance estimates for the MDPLEs developed by Wahba (1983) and Silverman (1985), under the Bayesian context, correspond to the inverse of the observed information matrix obtained by treating the penalized likelihood as a usual likelihood. Therefore, if we obtain the MDPLE of $\boldsymbol{\theta}$ through the Fisher scoring algorithm, it is reasonable to derive the variance-covariance matrix by using the inverse of the penalized Fisher information matrix. Thus, the asymptotic variance–covariance matrix of $\widehat{\boldsymbol{\theta}}$ can be obtained from the inverse of the expected information matrix $\boldsymbol{\mathcal{I}}_p$ defined in the appendix, that is,

$$\widehat{\mathrm{Cov}}_{\mathrm{asymptotic}}(\widehat{\boldsymbol{\theta}}) \approx \boldsymbol{\mathcal{I}}_p^{-1}(\widehat{\boldsymbol{\theta}}). \tag{14}$$

By using variance-covariance matrix given in Equation (14) we can construct an approximate pointwise standard error band (SEB) for $\beta_k$ that allows us to assess how accurate the estimator $\widehat{\beta_k}$ at different locations within the range of interest. For example, we can consider the approximate pointwise SEB given by

$$\mathrm{SEB}_{\mathrm{approx}}(\beta_k(r^0_{k_l})) = \widehat{\beta}_k(r^0_{k_l}) \pm 2\sqrt{\widehat{\mathrm{Var}}(\widehat{\beta}_k(r^0_{k_l}))},$$

where $\mathrm{Var}(\widehat{\beta}_k(r_{k_l}))$ is the $l$th principal diagonal element of the matrix given in Equation (14), for $l = 1, \ldots, r_k$.

Note from Equations (13) and (11) that it is possible to obtain the covariance matrix for $\boldsymbol{\beta}_\jmath$ ($\jmath = 0, \ldots, s+1$) and $\widehat{\boldsymbol{g}}$, respectively. Indeed,

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}_\jmath) = \boldsymbol{\mathcal{H}}_\jmath \widehat{\mathrm{Cov}(\boldsymbol{y})} \boldsymbol{\mathcal{H}}_\jmath^\top$$

and

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{g}}) = \boldsymbol{\mathcal{H}}_g \widehat{\mathrm{Cov}(\boldsymbol{y})} \boldsymbol{\mathcal{H}}_g^\top,$$

where $\boldsymbol{\mathcal{H}}_g = \boldsymbol{E}\boldsymbol{\mathcal{H}}_{s+1} + \boldsymbol{T}^\top \boldsymbol{\mathcal{H}}'$, with $\boldsymbol{\mathcal{H}}_{s+1}$ defined above and $\boldsymbol{\mathcal{H}}'$ denoting the block of matrix $\boldsymbol{\mathcal{H}}_0$ corresponding to vector $\boldsymbol{a}$, $\mathrm{Cov}(\boldsymbol{y}) = \mathrm{blockdiag}_{1\le i\le n}\big(\xi_i \boldsymbol{\Sigma}_i\big)$ and $\xi_i > 0$ is a quantity that may be obtained from the derivatives of the characteristic function associated with elliptical distributions (Fang et al., 1990).

### 3.6 Effective degrees of freedom

In general, in the literature concerning semiparametric models there are different definitions for the degrees of freedom (DF), depending on the context in which they are used. Here, the DF associated with the smooth varying-coefficient functions is defined as (Hastie and Tibshirani, 1990)

$$\begin{aligned}\mathrm{DF}(\lambda_k) &= \mathrm{tr}\big\{\widetilde{\boldsymbol{N}}_k \boldsymbol{S}_k\big\}\\ &= \mathrm{tr}\Big\{\widetilde{\boldsymbol{N}}_k^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_k \Big(\widetilde{\boldsymbol{N}}_k^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_k + \lambda_k \boldsymbol{K}_k\Big)^{-1}\Big\}.\end{aligned}$$

In practice, it is desirable to have an approximation to this quantity. Let $\boldsymbol{Q}_{\boldsymbol{N}_k} = \widetilde{\boldsymbol{N}}_k^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_k$ and $\boldsymbol{Q}_{\lambda_k} = \lambda_k \boldsymbol{K}_k$. Since $\boldsymbol{W}^* > 0$ and $\mathrm{rank}(\widetilde{\boldsymbol{N}}_k^\top) \le r_k$, then $\boldsymbol{Q}_{\widetilde{\boldsymbol{N}}_k} \ge 0$. Therefore, there exists a matrix $\boldsymbol{Q}^{1/2}_{\widetilde{\boldsymbol{N}}_k} \ge 0$ such that $\boldsymbol{Q}_{\widetilde{\boldsymbol{N}}_k} = \boldsymbol{Q}^{1/2}_{\widetilde{\boldsymbol{N}}_k} \boldsymbol{Q}^{1/2}_{\widetilde{\boldsymbol{N}}_k}$. Thus, we can write $\mathrm{tr}\big\{\widetilde{\boldsymbol{N}}_k \boldsymbol{S}_k\big\} = \mathrm{tr}\big\{\widetilde{\boldsymbol{S}}_k\big\}$ as (Eilers and Marx, 1996)

$$\mathrm{tr}\big\{\widetilde{\boldsymbol{S}}_k\big\} = \sum_{j=1}^{r_k} \frac{1}{1 + \lambda_k \ell_j},$$

where $\ell_j$, for $j = 1, \ldots, r_k$, are the eigenvalues of the matrix $\boldsymbol{Q}^{-1/2}_{\widetilde{\boldsymbol{N}}_k} \boldsymbol{Q}_{\lambda_k} \boldsymbol{Q}^{-1/2}_{\widetilde{\boldsymbol{N}}_k}$, for $k = 1, \ldots, s$. Analogously to the selection of DFs associated with smooth varying-coefficient functions, the DFs associated with smooth surface is given by

$$\begin{aligned}\mathrm{DF}(\lambda_g) &= \mathrm{tr}\big\{\widetilde{\boldsymbol{E}}\boldsymbol{S}_\delta\big\}\\ &= \mathrm{tr}\Big\{\widetilde{\boldsymbol{E}}\Big(\widetilde{\boldsymbol{E}}^\top\boldsymbol{W}^*\widetilde{\boldsymbol{E}}+\lambda_g\widetilde{\boldsymbol{E}}\Big)^{-1}\widetilde{\boldsymbol{E}}^\top\boldsymbol{W}^*\Big\}.\end{aligned}$$

Thus, considering $\boldsymbol{Q}_{\widetilde{\boldsymbol{E}}} = \widetilde{\boldsymbol{E}}^\top\boldsymbol{W}^*\widetilde{\boldsymbol{E}}$ and $\boldsymbol{Q}_{\lambda_g} = \lambda_g\boldsymbol{E}$, and since $\boldsymbol{W}^* > 0$ and $\mathrm{rank}(\widetilde{\boldsymbol{E}}^\top) \le n$, then $\boldsymbol{Q}_{\widetilde{\boldsymbol{E}}} \ge 0$. Therefore, there exists a matrix $\boldsymbol{Q}_{\widetilde{\boldsymbol{E}}}^{1/2} \ge 0$ such that $\boldsymbol{Q}_{\widetilde{\boldsymbol{E}}} = \boldsymbol{Q}_{\widetilde{\boldsymbol{E}}}^{1/2}\boldsymbol{Q}_{\widetilde{\boldsymbol{E}}}^{1/2}$. Thus, we can write $\mathrm{tr}\big\{\widetilde{\boldsymbol{E}}\boldsymbol{S}_\delta\big\} = \mathrm{tr}\big\{\widetilde{\boldsymbol{S}}_\delta\big\}$ as

$$\mathrm{tr}\big\{\widetilde{\boldsymbol{S}}_\delta\big\} = \sum_{j=1}^{n}\frac{1}{1+\lambda_g\ell_j},$$

where $\ell_j$, for $j = 1, \dots, n$, are the eigenvalues of the matrix $\boldsymbol{Q}_{\widetilde{\boldsymbol{E}}}^{-1/2}\boldsymbol{Q}_{\lambda_g}\boldsymbol{Q}_{\widetilde{\boldsymbol{E}}}^{-1/2}$. It is important to note that both $\mathrm{DF}(\lambda_k)$ and $\mathrm{DF}(\lambda_g)$ are inversely proportional to $\lambda_k$ and $\lambda_g$, respectively. Alternatively, we can consider the backfitting estimators defined in Equation (13) and thus calculate the DFs associated with the smooth varying-coefficient functions as

$$\overline{\mathrm{DF}}(\lambda_\jmath) = \mathrm{tr}\big\{\boldsymbol{\mathcal{H}}_\jmath\big\}, \quad \jmath = 1, \dots, s,$$

with $\boldsymbol{\mathcal{H}}_\jmath$ defined above. Similarly, the DFs associated with the smooth surface can be calculated from the representation $\widehat{\boldsymbol{g}} = \boldsymbol{E}\widehat{\boldsymbol{\delta}} + \boldsymbol{T}^\top\widehat{\boldsymbol{a}} = \boldsymbol{\mathcal{H}}_g\boldsymbol{y}$, whit $\boldsymbol{\mathcal{H}}_g = \boldsymbol{E}\boldsymbol{\mathcal{H}}_{s+1} + \boldsymbol{T}^\top\boldsymbol{\mathcal{H}}'$, $\boldsymbol{\mathcal{H}}_{s+1}$, $\boldsymbol{\mathcal{H}}'$ and $\boldsymbol{y}$ defined in the previous sections. Thus, the DFs are given by

$$\overline{\mathrm{DF}}(\lambda_g) = \mathrm{tr}\big\{\boldsymbol{\mathcal{H}}_g\big\}.$$

### 3.7 Selecting an appropriate model

Under the elliptical TPSPVCM, we have a total of $2 + p + d + \mathrm{DF}(\boldsymbol{\lambda})$ parameters to be estimated, with $\mathrm{DF}(\boldsymbol{\lambda}) = \mathrm{DF}(\lambda_g) + \sum_{k=1}^{s}\mathrm{DF}(\lambda_k)$ denoting approximately the number of effective parameters involved in modeling of the smooth varying-coefficient functions and surface. In this case, the Akaike information criterion (AIC) (Akaike, 1973) or the Bayesian information criterion (BIC) (Schwarz et al., 1978) can be used for selecting an appropriate model. The idea is to minimize the function

$$\mathrm{AIC}(\boldsymbol{\lambda}) = -2L_p(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + 2\Big[2 + p + d + \mathrm{DF}(\boldsymbol{\lambda})\Big],$$

where $L_p(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda})$ denotes the penalized log-likelihood function available at $\widehat{\boldsymbol{\theta}}$ for a fixed $\boldsymbol{\lambda}$. It is important to mention that AIC is based on information theory and is useful for selecting an appropriate model given data with adequate sample size. An alternative version of the AIC, denoted by AICc, was proposed by Hurvich et al. (1998) in the context of parametric linear regression and autoregressive time series. Recently, Relvas (2016) adapted this criterion for the partially linear model with first-order autoregressive symmetric errors. Considering such proposals, we propose the AICc as an alternative for the selection of models under the elliptical TPSPVCM, which is given by

$$\mathrm{AIC_c}(\boldsymbol{\lambda}) = \log\left\{\frac{\parallel\sqrt{\widehat{\boldsymbol{W}_v}}(\boldsymbol{y}-\widehat{\boldsymbol{y}})\parallel^2}{n}\right\} + \frac{2\Big[\mathrm{tr}(\widehat{\boldsymbol{H}(\boldsymbol{\lambda})})+1\Big]}{n-\mathrm{tr}(\widehat{\boldsymbol{H}(\boldsymbol{\lambda})})-2} + 1,$$

where $\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{H}(\boldsymbol{\lambda})}\boldsymbol{y}$ and $\widehat{\boldsymbol{H}(\boldsymbol{\lambda})}$ corresponds to the smoother matrix, which is equivalent to the hat matrix defined in the class of parametric regression models. If we consider the matrix representation given in Equation (3) of our model and the backfitting estimators given in Equation (13), it is possible to obtain a closed expression for the matrix $\widehat{\boldsymbol{H}(\boldsymbol{\lambda})}$. Indeed, assuming that $\boldsymbol{\lambda}$, $\boldsymbol{W}^*$ and $\boldsymbol{W}_v$ are fixed, we have that

$$\boldsymbol{H}(\boldsymbol{\lambda}) = \sum_{\jmath=0}^{s+1} \widetilde{\boldsymbol{\mathcal{H}}}_\jmath, \tag{15}$$

with $\widetilde{\boldsymbol{\mathcal{H}}}_\jmath = \boldsymbol{N}_\jmath \boldsymbol{\mathcal{H}}_\jmath$. Note that the principal diagonal elements of $\boldsymbol{H}(\boldsymbol{\lambda})$ obtained in the last iteration of the iterative process, denoted here by $h_{ii}(\boldsymbol{\lambda})$, are called leverage points and play an important role in the construction of diagnostic techniques.

### 3.8 Smoothing parameters

The determination of the parameters $\lambda_k$ and $\lambda_g$ is a crucial part in the estimation process and different choice methods are available in the literature. For example, it is usual to consider the cross-validation method or the generalized cross-validation method (Craven and Wahba, 1978). Following Relvas (2016), an alternative to select smoothing parameters under the elliptical TPSPVCM is to consider a generalized cross-validation method defined by

$$\mathrm{GCV}(\boldsymbol{\lambda}) = \frac{\| \sqrt{\widehat{\boldsymbol{W}_v}}(\boldsymbol{y} - \widehat{\boldsymbol{y}}) \|^2}{[1 - n^{-1}\mathrm{tr}(\widehat{\boldsymbol{H}(\boldsymbol{\lambda})})]} \,.$$

In this case, $\boldsymbol{\lambda}$ should be obtained by minimizing $\mathrm{GCV}(\boldsymbol{\lambda})$ for a grid of $\boldsymbol{\lambda}$ values. Alternatively, these parameters may be selected by applying the AIC. In particular, we can consider the $\mathrm{AIC}(\boldsymbol{\lambda})$ or $\mathrm{AIC}_c(\boldsymbol{\lambda})$ criteria defined in the previous section, and use the effective DFs involved in nonparametric modeling to select appropriate smoothing parameters (Ibacache-Pulgar et al., 2013).

### 3.9 Residual Analysis

We propose a standardized residual which can be used to detect error distribution misspecification as well as the presence of outlying observations. It follows from Equation (15) that the residuals vector is the difference between the observed data vector and estimated mean vector, that is,

$$\widehat{\boldsymbol{r}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \big[\boldsymbol{I} - \boldsymbol{H}(\boldsymbol{\lambda})\big]\, \boldsymbol{y}\,. \tag{16}$$

Since that $\boldsymbol{H}(\boldsymbol{\lambda})$ is not a projection operator, this is, $\boldsymbol{H}^2(\boldsymbol{\lambda}) \neq \boldsymbol{H}(\boldsymbol{\lambda})$, we have that the approximate variance of the residual vector is given by

$$\mathrm{Var}_{\mathrm{approx}}(\widehat{\boldsymbol{r}}) = \big[\boldsymbol{I} - \boldsymbol{H}(\boldsymbol{\lambda})\big]\, \mathrm{Cov}(\boldsymbol{y}) \big[\boldsymbol{I} - \boldsymbol{H}(\boldsymbol{\lambda})\big]^\top,$$

where $\mathrm{Cov}(\boldsymbol{y})$=$\mathrm{blockdiag}_{1\leq i\leq n}\big(\xi_i\boldsymbol{\Sigma}_i\big)$. Then, we have that the $l$th standardized residual takes the form

$$\widehat{r}_l = \frac{\boldsymbol{d}_l^\top\left[\boldsymbol{I}-\boldsymbol{H}(\boldsymbol{\lambda})\right]\boldsymbol{y}}{\sqrt{\boldsymbol{d}_l^\top\widehat{\mathrm{Var}}_{\mathrm{approx}}(\widehat{\boldsymbol{r}})\boldsymbol{d}_l}},$$

where $\widehat{\mathrm{Var}}_{\mathrm{approx}}(\widehat{\boldsymbol{r}})$=$\mathrm{Var}_{\mathrm{approx}}(\widehat{\boldsymbol{r}})\big|_{\widehat{\boldsymbol{\theta}}}$, whit $\boldsymbol{d}_l$ denoting an $(M\times 1)$ vector with 1 at the $l$th position and 0 elsewhere, for $l=1,\ldots,M$. Further details on the analysis of residuals in the semiparametric context can be found, for example, in Ibacache-Pulgar et al. (2013).

## 4. Application

In this section, we illustrate the applicability of the TPSPVCM through an application based on a set of real data. For comparative purposes, we consider random errors whose distribution belongs to the symmetric class; specifically, the normal and Student-t distributions.

### 4.1 Data description

In our application, we consider the house prices of Boston area reported by Harrison and Rubinfeld (1978) and analyzed by many authors; see, for example, Belsley et al. (1980), Ibacache-Pulgar et al. (2013) and, more recently, Ibacache-Pulgar and Reyes (2018). This data set contains a sample of 506 observations collected by the U.S Census Service concerning housing in the area of Boston. The variable LMV (logarithm of the median house price in USD 1000) is related with 14 explanatory variables, 6 of them are defined from census track and the remaining variables are defined for clusters. For simplicity, we consider four explanatory variables: LSTAT (logarithm of the proportion of the population that is lower status, ROOM (average number of rooms per dwelling), CRIM (per capita crime rate by town), TAX (full-value property-tax rate per USD 10000), and the geographical coordinates expressed in longitude and latitude. Similar to what observed by Ibacache-Pulgar and Reyes (2018), we see in Figure 1(a) that the relationship between LMV and the explanatory variable TAX is linear, whereas the relationship between LMV and LSTAT appear in nonlinear ways (Figure 1(b)). Also, Figures 1(c) and 1(d) suggests that the explanatory variables ROOM and CRIM might be interacting with the variable LSTAT in nonlinear fashion. Figure 2 represents the spatial distribution of the LMV variable. From Figure 2 (right) we note that the lowest prices are concentrated between the latitudes 42.2 and 42.25 and longitudes between -71.0 and -71.1, while the highest prices are in the north part of the town. It is important to point out that Ibacache-Pulgar and Reyes (2018) analyzes this same set of variables using a partially varying-coefficient model but without considering the effect of the geographic coordinates associated with each of the households surveyed. We believe that including the effect of geographical coordinates can improve the fit of the model considered by Ibacache-Pulgar and Reyes (2018) in predictive terms, precision of the estimates and goodness of fit.

### 4.2 Fitting the models

Considering the analysis described above, we suggest the application of a partially varying-coefficient model that including the spatial variability. Specifically, we assume the following

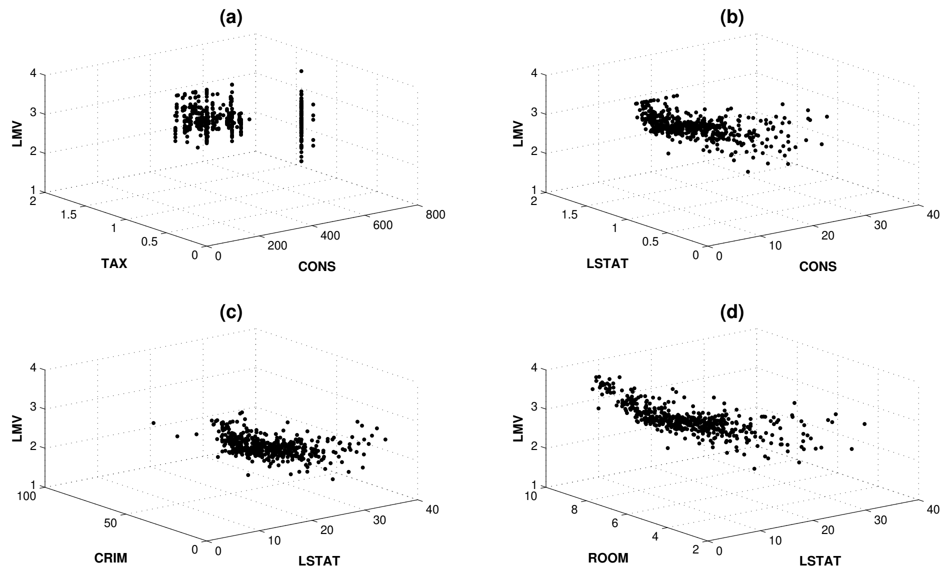


Figure 1. Three-dimensional graphics for house prices data. CONS denote an auxiliar variable defined as an $(n \times 1)$ ones vector.

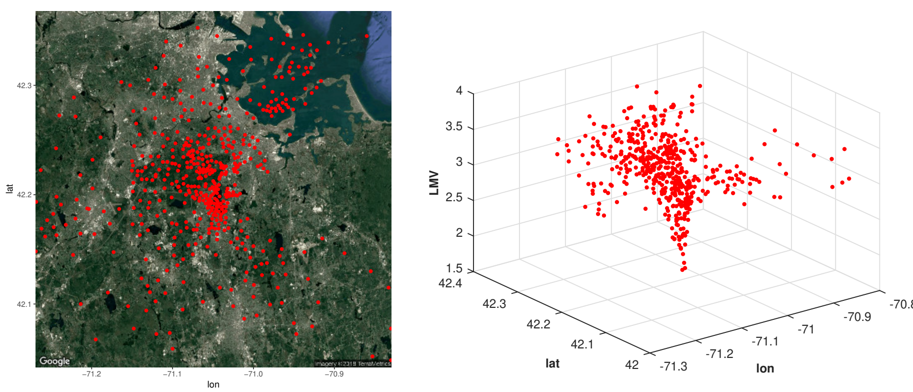


Figure 2. (Left) Google map of the Boston province. Red circles indicate the spatial distribution of the house prices data. (Right) Distributions of the LMV respect the longitude and the latitude.

thin-plate spline partially varying-coefficient model:

$$y_i = \alpha_0 + \alpha_1\, z_i + \beta_1(r_i)\, x_i^{(1)} + \beta_2(r_i)\, x_i^{(2)} + g(\boldsymbol{t}_i) + \varepsilon_i, \quad i = 1, \ldots, 506, \tag{17}$$

where $y_i$ denotes the value of LMV in USD 1000, $z_i$ the value of TAX, $x_i^{(1)}$ the value of CRIM, $x_i^{(2)}$ the value of ROOM, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^\top$ the parameters vector associated with parametric component, $r_i$ the value of LSTAT from the $i$th experimental unit, $\beta_k$, for $k = 1, 2$, are unknown smooth functions, $g$ is a smooth surface that depends of the vector of coordinates $\boldsymbol{t}_i = (\mathrm{t}_{1_i}, \mathrm{t}_{2_i}) \in \mathcal{R}^2$, and $\varepsilon_i$ are independent random errors that follow a symmetric distribution whit location parameter 0, scale parameter $\phi$ and density generator function $h$. We compare in the sequel the fits based on normal and Student-t random errors. The DFs ($\nu$) for the Student-t model was selected by the AIC, that is, by defining a grid of values for $\nu$ and choosing the one that minimize the AIC. Figure 3 shows the graph of AIC values for different DFs. We can see that this criterion is minimized for a value of $\nu = 4$.

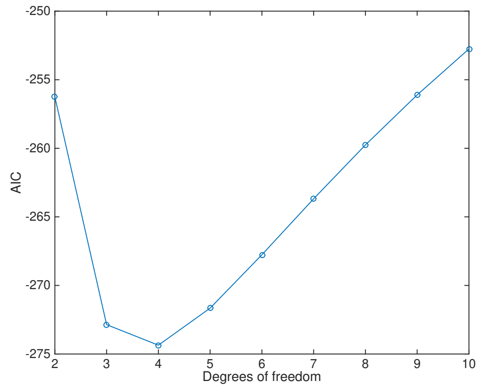


Figure 3. AIC values for different DFs, with $\nu = 2, \ldots, 10$.

The MDPLE estimates, estimated standard errors and the corresponding AIC for the model of Equation (17) under normal and Student-t distributions are presented in Table 1. Comparing these results, we may notice a similarity between the estimates $\widehat{\boldsymbol{\alpha}}$ under both models, but the standard error for $\widehat{\alpha}_1$ appears to be smaller under the Student-t model. Also, it can be seen that the scale parameters are different for the two fitted models, but the estimates are not comparable since they are on different scales. Additionally, we may notice that the AIC value under the Student-t model is smaller than the one under the normal model, indicating that the models with longer-than-normal tails seem to better fit the data, a fact that is also confirmed through the theoretical quantile versus empirical quantile (QQ) plots presented in Figure 4.

Table 1. Maximum penalized likelihood estimates, estimated standard errors (SE) and AIC values under normal and Student-t ($\nu = 4$) models fitted to house prices data.

| | Normal | | Student-t | |
|---|---|---|---|---|
| | Estimate | SE | Estimate | SE |
| $\alpha_1$ | 3.0668 | 0.1145 | 3.0637 | 0.0964 |
| $\alpha_2$ | -0.0003 | 0.0001 | -0.0002 | 0.0001 |
| $\phi$ | 0.0344 | 0.0022 | 0.0172 | 0.0372 |
| AIC | -218.76 | | -274.65 | |

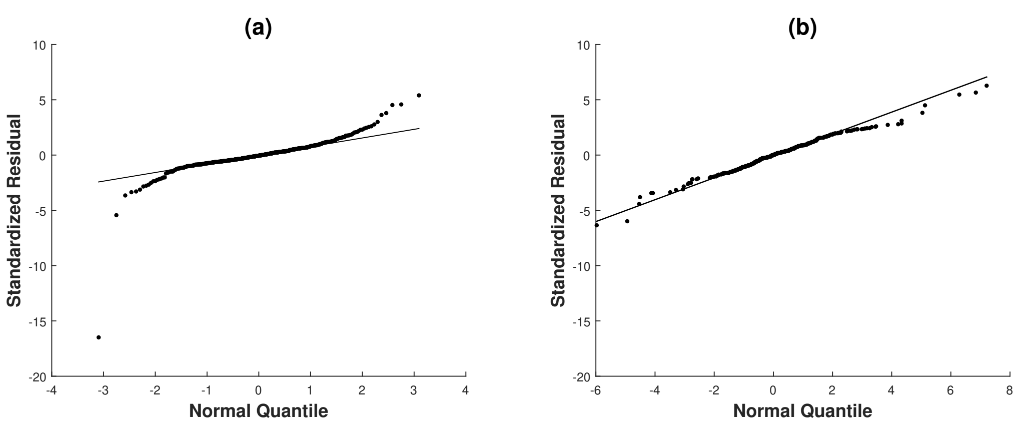


Figure 4. QQ plots fitted to house prices data: normal (a) and Student-t models with $\nu = 4$ DFs (b).

The standarized residual plot provide in Figure 5 is used to verify if there are outlying observations. In this case, the presence of some outlying observations for both models is clearly observed. Figure 6 displays the graphics of the LMV versus the fitted LMV from the two models. Although these plots indicate suitable fits for both models.

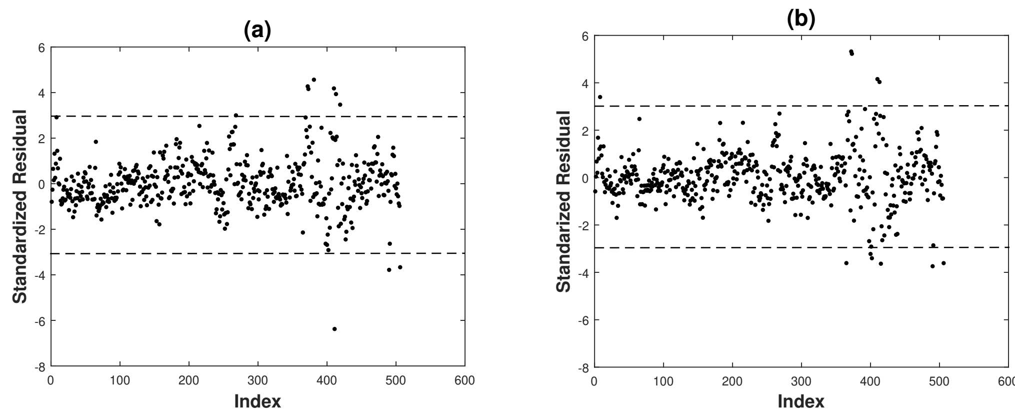


Figure 5. Index plots of standardized residuals to house prices data: normal (a) and Student-t models with $\nu = 4$ DFs (b).

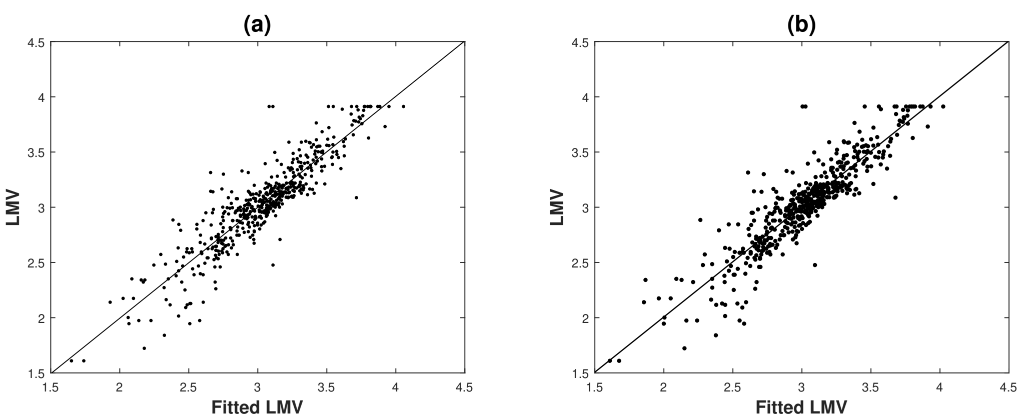


Figure 6. Scatter plots LMV versus fitted LMV to house prices data: normal (a) and Student-t models with $\nu = 4$ DFs (b).

The estimated coefficients functions $\beta_1$ and $\beta_2$ are computed using the smoothing parameters obtained by the method described in Subsection 3.8. Figures 7 and 8 show the estimated coefficient functions under both models and their corresponding approximate standard error band (dashed curves). The figures suggest that the coefficient curves vary with the explanatory variable LSTAT. In addition, it can be seen that the functions estimated under the normal model have a higher smoothness compared to those obtained from the Student-t model. It is important to remember that in this work we have incorporated the spatial variability of the data in the modeling process. Comparing with the results obtained by Ibacache-Pulgar and Reyes (2018), we can notice that the TPSPVCM model significantly improves the quality of the adjustment compared with the PVCM model. For example, for normal TPSPVCM model, the AIC value is $-218.7586$, while that for normal PVCM model the AIC value is $-139.4998$. Analogously, under Student-t TPSPVCM model with four DFs, the AIC value is $-274.6546$, while that under Student-t PVCM model with five DFs, the AIC value is $-188.3909$. In addition, we can notice that for the normal model, the estimated functions differ significantly, while under the Student-t model, they retain the same tendency but with a greater degree of smoothness.

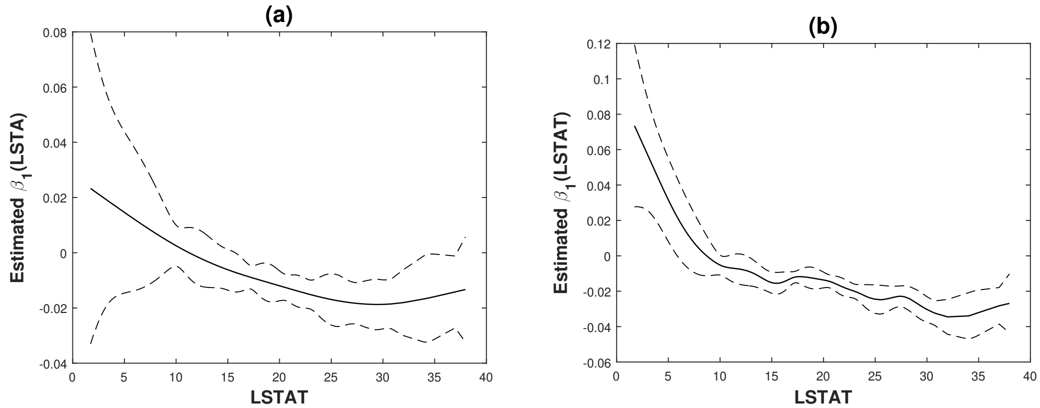


Figure 7. Plots of estimated coefficient function $\beta_1$ for the house prices data and its approximate pointwise standard error bands denoted by the dashed lines: Normal (a) and Student-t with $\nu = 4$ DFs (b) models.

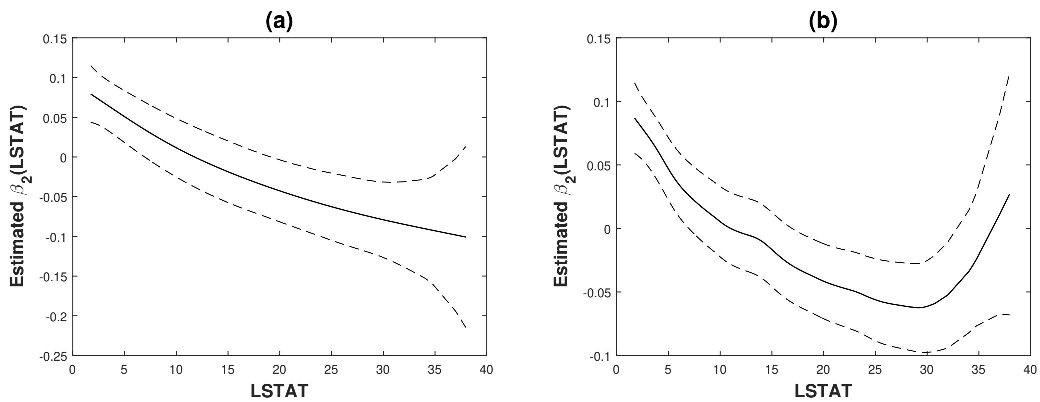


Figure 8. Plots of estimated coefficient function $\beta_2$ for the house prices data and its approximate pointwise standard error bands denoted by the dashed lines: Normal (a) and Student-t with $\nu = 4$ DFs (b).

## 4.3 Robustness aspects of the MDPLEs

It is important to note that for univariate Student-t distribution the current weight $v_i^{(u)} = (\nu + 1)/(\nu + u_i^{(u)})$, with $u_i^{(u)} = (y_i - \mu_i^{(u)})^2/\phi^{(u)}$, is inversely proportional to the distance between the observed value $y_i$ and its current predicted value $\mu_i^{(u)}$, so that outlying observations tend to have small weights in the estimation process. Therefore, we may expected that the MDPLEs from the Student-t TPSPVCMs are less sensitive to outlying observations than the MDPLEs from normal models. Figure 9 shows the plot between the standardized residual defined in Equation (16) and estimated weights under Student-t model. We can be seen that observation #411 has a very small residual and a high estimated weight, but its removal from the data set did no generate significant changes in the estimation of the parameters. For this reason the summary of the fit without this observation is omitted. Finally, it is important to note that the iterative process under Student-t model generates a reduction in the weights associated with the observations detected as discrepant. Hence such estimators present some characteristics of robustness similar to the associated with the weight function described by Huber (1981).

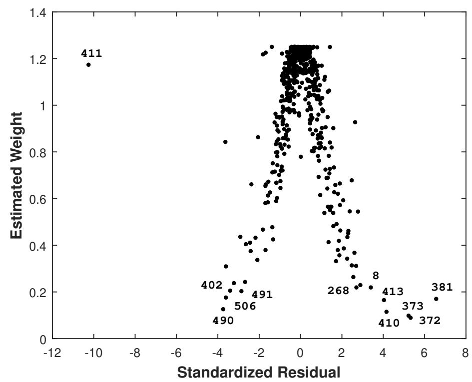


Figure 9. Plot of the standardized residual and estimated weights for the house prices data under Student-t model.

## 5. Concluding remarks

The elliptical thin-plate spline partially varying-coefficient models proposed in this paper have special characteristics compared to other types of models existing in the literature. Specifically, these models allow describing the mean of the data in those cases in which there are explanatory variables that are related to the response variable through a regression structure that depends on a parametric component (usual linear predictor), a non-parametric component (explanatory variables effects in which the coefficients are allowed to vary as smooth functions of other variables) and a spatial component (thin-plate spline). In addition, the distributional assumption established on random errors allows us to model datasets in which the assumption of normality is not appropriate. We derive a reweighed iterative process for obtaining the maximum doubly penalized likelihood estimators based on the Score Fisher and back-fitting methods. Closed-form expressions are obtained for the penalized observed and expected information matrices, and expressions for the standard errors of the maximum doubly penalized likelihood estimators are also available. We propose a way to estimate the smoothing parameters based on generalized cross-validation and a method for the selection of models by using the AIC. A real dataset previously analyzed under normal errors is reanalyzed under Student-t errors by including a smoothing surface for the spatial variability. By comparing the AIC values of the two models, the Student-t showed the better fitting. Thus, we can recommend Student-t thin-plate spline partially varying-coefficient models as an option to fit datasets with indications of heavy tails. The computational implementation of all our results was carried out in MATLAB software, and the codes can be requested from the authors to the email german.ibacache@uv.cl.

## Appendix

Here, we show the score function, the observed information matrix and the expected information matrix for the elliptical TPSPCVM base on the doubly penalized log-likelihood function given in Equation (6).

### Penalized score function

Let $\boldsymbol{W}_v = \text{blockdiag}_{1\leq i\leq n}(v_i \boldsymbol{W}_i)$, with $\boldsymbol{W}_i = \boldsymbol{\Sigma}_i^{-1}$, $v_i = -2\zeta_h(u_i)$, $\zeta_h(u_i) = \mathrm{d}\log h(u_i)/\mathrm{d}u_i$, $\boldsymbol{\Sigma}_i^{*} = \boldsymbol{\Sigma}_i^{-1}\partial\boldsymbol{\Sigma}_i/\partial\ell$, $\boldsymbol{\Upsilon} = \text{blockdiag}_{1\leq i\leq n}(\boldsymbol{\Upsilon}_i)$, with $\boldsymbol{\Upsilon}_i =$

$v_i \boldsymbol{\Sigma}_i^{-1}(\partial \boldsymbol{\Sigma}_i/\partial \ell)\, \boldsymbol{\Sigma}_i^{-1}$. Assuming that Equation (6) is regular with respect to all elements of $\boldsymbol{\theta}$, we have that the penalized score function of $\boldsymbol{\theta}$ under elliptical TPSPVCM is given by

$$\boldsymbol{U}_p(\boldsymbol{\theta}) = \frac{\partial L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{U}_p(\boldsymbol{\theta})^{\widetilde{\alpha}} \\ \boldsymbol{U}_p(\boldsymbol{\theta})^{\beta_1} \\ \vdots \\ \boldsymbol{U}_p(\boldsymbol{\theta})^{\beta_s} \\ \boldsymbol{U}_p(\boldsymbol{\theta})^{\delta} \\ \boldsymbol{U}_p(\boldsymbol{\theta})^{\tau} \end{pmatrix},$$

where $\boldsymbol{U}_p^{\widetilde{\alpha}}(\boldsymbol{\theta}) = (\boldsymbol{Z}^\top \boldsymbol{W}_v \boldsymbol{\varepsilon} \quad \widetilde{\boldsymbol{T}} \boldsymbol{W}_v \boldsymbol{\varepsilon})^\top$, $\boldsymbol{U}_p^{\beta_k}(\boldsymbol{\theta}) = \widetilde{\boldsymbol{N}}_k^\top \boldsymbol{W}_v \boldsymbol{\varepsilon} - \lambda_1 \boldsymbol{K}_k \boldsymbol{\beta}_k$, for $k = 1, \ldots, s$, $\boldsymbol{U}_p^{\delta}(\boldsymbol{\theta}) = \widetilde{\boldsymbol{E}} \boldsymbol{W}_v \boldsymbol{\varepsilon} - \lambda_g \boldsymbol{E} \boldsymbol{\delta}$ and $\boldsymbol{U}_p^{\tau}(\boldsymbol{\theta}) = -(1/2) \sum_{i=1}^n \mathrm{tr}(\boldsymbol{\Sigma}_i^*) + (1/2) \boldsymbol{\varepsilon}^\top \boldsymbol{\Upsilon} \boldsymbol{\varepsilon}$.

## Penalized observed information matrix

For simplicity, let $\boldsymbol{\Psi}_i = 2\,\boldsymbol{\Psi}_{1i} + \boldsymbol{\Psi}_{2i}$, $\boldsymbol{\Psi}_i^* = \boldsymbol{\Psi}_{1i} + \boldsymbol{\Psi}_{2i}$ and $\boldsymbol{\Psi}_i^{**} = \boldsymbol{\Psi}_{1i} + 2\,\boldsymbol{\Psi}_{2i}$, with $\boldsymbol{\Psi}_{1i} = v_i' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top \boldsymbol{\Sigma}_i^{-1}$ and $\boldsymbol{\Psi}_{2i} = v_i \boldsymbol{\Sigma}_i^{-1}$. In addition, let $\boldsymbol{\Psi} = \mathrm{diag}_{1 \le i \le n} \big(\boldsymbol{\Psi}_i\big)$ and $\boldsymbol{\Omega} = \mathrm{diag}_{1 \le i \le n} \big(\boldsymbol{\Omega}_i\big)$, with $\boldsymbol{\Omega}_i = \boldsymbol{\Psi}_i^* (\partial \boldsymbol{\Sigma}_i / \partial \jmath)\, \boldsymbol{\Sigma}_i^{-1}$. The $\boldsymbol{L}_p$ $(p^* \times p^*)$ Hessian matrix under elliptical TPSPVCM is defined as

$$\boldsymbol{L}_p(\boldsymbol{\theta}) = \frac{\partial^2 L_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \boldsymbol{L}_p^{\widetilde{\alpha}\widetilde{\alpha}}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\widetilde{\alpha}\beta}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\widetilde{\alpha}\delta}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\widetilde{\alpha}\tau}(\boldsymbol{\theta}) \\ \boldsymbol{L}_p^{\beta\widetilde{\alpha}}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\beta\beta}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\beta\delta}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\beta\tau}(\boldsymbol{\theta}) \\ \boldsymbol{L}_p^{\delta\widetilde{\alpha}}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\delta\beta}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\delta\delta}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\delta\tau}(\boldsymbol{\theta}) \\ \boldsymbol{L}_p^{\tau\widetilde{\alpha}}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\tau\beta}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\tau\delta}(\boldsymbol{\theta}) & \boldsymbol{L}_p^{\tau\tau}(\boldsymbol{\theta}) \end{pmatrix},$$

whose elements are given by

$$\boldsymbol{L}_p^{\widetilde{\alpha}\widetilde{\alpha}}(\boldsymbol{\theta}) = \begin{pmatrix} -\boldsymbol{Z}^\top \boldsymbol{\Psi} \boldsymbol{Z} & -\boldsymbol{Z}^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{E}} \\ -\widetilde{\boldsymbol{E}}^\top \boldsymbol{\Psi} \boldsymbol{Z} & -\widetilde{\boldsymbol{T}}^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{T}} \end{pmatrix}, \quad \boldsymbol{L}_p^{\widetilde{\alpha}\beta}(\boldsymbol{\theta}) = \begin{pmatrix} -\boldsymbol{Z}^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{N}}_1 & \ldots & -\boldsymbol{Z}^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{N}}_s \\ -\widetilde{\boldsymbol{T}} \boldsymbol{\Psi} \widetilde{\boldsymbol{N}}_1 & \ldots & -\widetilde{\boldsymbol{T}} \boldsymbol{\Psi} \widetilde{\boldsymbol{N}}_s \end{pmatrix},$$

$$\boldsymbol{L}_p^{\widetilde{\alpha}\delta}(\boldsymbol{\theta}) = \begin{pmatrix} -\boldsymbol{Z}^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{E}} \\ -\boldsymbol{Z}^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{T}} \end{pmatrix}, \quad \boldsymbol{L}_p^{\widetilde{\alpha}\tau}(\boldsymbol{\theta}) = \begin{pmatrix} -\boldsymbol{Z}^\top \boldsymbol{\Omega} \boldsymbol{\varepsilon} \\ -\widetilde{\boldsymbol{T}}^\top \boldsymbol{\Omega} \boldsymbol{\varepsilon} \end{pmatrix},$$

$$\boldsymbol{L}_p^{\beta\beta}(\boldsymbol{\theta}) = \begin{pmatrix} -\widetilde{\boldsymbol{N}}_1^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{N}}_1 - \lambda_1 \boldsymbol{K}_1 & \ldots & -\widetilde{\boldsymbol{N}}_1^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{N}}_s \\ -\widetilde{\boldsymbol{N}}_2^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{N}}_1 & \ldots & -\widetilde{\boldsymbol{N}}_2^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{N}}_s \\ \vdots & \ddots & \vdots \\ -\widetilde{\boldsymbol{N}}_s^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{N}}_1 & \ldots & -\widetilde{\boldsymbol{N}}_s^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{N}}_s - \lambda_s \boldsymbol{K}_s \end{pmatrix}, \quad \boldsymbol{L}_p^{\beta\delta}(\boldsymbol{\theta}) = \begin{pmatrix} -\widetilde{\boldsymbol{N}}_1^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{E}} \\ -\widetilde{\boldsymbol{N}}_2^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{E}} \\ \vdots \\ -\widetilde{\boldsymbol{N}}_s^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{E}} \end{pmatrix},$$

$$\boldsymbol{L}_p^{\beta\tau}(\boldsymbol{\theta}) = \begin{pmatrix} -\widetilde{\boldsymbol{N}}_1^\top \boldsymbol{\Omega} \boldsymbol{\varepsilon} \\ -\widetilde{\boldsymbol{N}}_2^\top \boldsymbol{\Omega} \boldsymbol{\varepsilon} \\ \vdots \\ -\widetilde{\boldsymbol{N}}_s^\top \boldsymbol{\Omega} \boldsymbol{\varepsilon} \end{pmatrix}, \quad \boldsymbol{L}_p^{\delta\delta}(\boldsymbol{\theta}) = -\widetilde{\boldsymbol{E}}^\top \boldsymbol{\Psi} \widetilde{\boldsymbol{E}} - \lambda_g \boldsymbol{E}, \quad \boldsymbol{L}_p^{\delta\tau}(\boldsymbol{\theta}) = -\widetilde{\boldsymbol{E}}^\top \boldsymbol{\Omega} \boldsymbol{\varepsilon}$$

and

$$\boldsymbol{L}_p^{\tau\tau}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{1}{2}\mathrm{tr}\Bigg(\boldsymbol{\Sigma}_i^{-1}\Big[\frac{\partial \boldsymbol{\Sigma}_i}{\partial \ell}\,\boldsymbol{\Sigma}_i^{-1}\frac{\partial \boldsymbol{\Sigma}_i}{\partial \jmath} - \frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial \jmath \partial \ell}\Big]\Bigg) - \frac{1}{2}\boldsymbol{\varepsilon}_i^{\top}\boldsymbol{\Sigma}_i^{-1} \times$$

$$\left[\frac{\partial \boldsymbol{\Sigma}_i}{\partial \ell}\,\boldsymbol{\Psi}_i^{**}\frac{\partial \boldsymbol{\Sigma}_i}{\partial \jmath} - v_i\frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial \jmath \partial \ell}\right]\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\varepsilon}_i.$$

### Penalized expected information matrix

Let $d_{g_i} = \mathrm{E}(\zeta_g^2(\gamma_i)\gamma_i)$ and $f_{g_i} = \mathrm{E}(\zeta_g^2(\gamma_i)\gamma_i^2)$, with $\gamma_i = \boldsymbol{e}_i^{\top}\boldsymbol{e}_i$, $\boldsymbol{e}_i \sim \mathrm{El}_{m_i}(\boldsymbol{0}, \boldsymbol{I}_{m_i})$, and $\boldsymbol{W}^* = \mathrm{blockdiag}_{1\leq i\leq n}\big((4d_{g_i}/m_i)\,\boldsymbol{W}_i\big)$. By calculating the expectation of the matrix $-\boldsymbol{L}_p$, we obtain the $(p^* \times p^*)$ penalized expected information matrix given by

$$\boldsymbol{\mathcal{I}}_p(\boldsymbol{\theta}) = -\mathrm{E}\left(\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}}\right).$$

Following Lange et al. (1989), we have that the $(j^*, \ell^*)$-element of the matrix $\boldsymbol{\mathcal{I}}_p$ for $i$th cluster, with respect to the parameters $\theta_{j^*}^*$ and $\theta_{\ell^*}^*$, can be obtained as

$$\boldsymbol{\mathcal{I}}_{p_i}(\boldsymbol{\theta}) = \mathrm{E}\left(\frac{\partial L_{\mathrm{p}_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \theta_{j^*}}\frac{\partial L_{\mathrm{p}_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \theta_{\ell^*}}\right).$$

After some algebraic manipulations we find that the $\boldsymbol{\mathcal{I}}_p(\boldsymbol{\theta})$ matrix have a block-diagonal structure of the form

$$\boldsymbol{\mathcal{I}}_p(\boldsymbol{\theta}) = \mathrm{blockdiag}\Big(\boldsymbol{\mathcal{I}}_p^{11}(\boldsymbol{\theta}), \boldsymbol{\mathcal{I}}_p^{22}(\boldsymbol{\theta})\Big),$$

where

$$\boldsymbol{\mathcal{I}}_p^{11}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mathcal{I}}_p^{\widetilde{\alpha}\widetilde{\alpha}}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_p^{\widetilde{\alpha}\beta}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_p^{\widetilde{\alpha}\delta}(\boldsymbol{\theta}) \\ \boldsymbol{\mathcal{I}}_p^{\beta\widetilde{\alpha}}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_p^{\beta\beta}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_p^{\beta\delta}(\boldsymbol{\theta}) \\ \boldsymbol{\mathcal{I}}_p^{\delta\widetilde{\alpha}}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_p^{\delta\beta}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_p^{\delta\delta}(\boldsymbol{\theta}) \end{pmatrix},$$

whose elements of the matrix are given by

$$\boldsymbol{\mathcal{I}}_p^{\widetilde{\alpha}\widetilde{\alpha}}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{Z}^{\top}\boldsymbol{W}^*\boldsymbol{Z} & \boldsymbol{Z}^{\top}\boldsymbol{W}^*\widetilde{\boldsymbol{E}} \\ \widetilde{\boldsymbol{E}}^{\top}\boldsymbol{W}^*\boldsymbol{Z} & \widetilde{\boldsymbol{T}}^{\top}\boldsymbol{W}^*\widetilde{\boldsymbol{T}} \end{pmatrix}, \quad \boldsymbol{\mathcal{I}}_p^{\widetilde{\alpha}\beta}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{Z}^{\top}\boldsymbol{W}^*\widetilde{\boldsymbol{N}}_1 & \dots & \boldsymbol{Z}^{\top}\boldsymbol{W}^*\widetilde{\boldsymbol{N}}_s \\ \widetilde{\boldsymbol{T}}\boldsymbol{W}^*\widetilde{\boldsymbol{N}}_1 & \dots & \widetilde{\boldsymbol{T}}\boldsymbol{W}^*\widetilde{\boldsymbol{N}}_s \end{pmatrix},$$

$$\boldsymbol{\mathcal{I}}_p^{\widetilde{\alpha}\delta}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{Z}^{\top}\boldsymbol{W}^*\widetilde{\boldsymbol{E}} \\ \boldsymbol{Z}^{\top}\boldsymbol{W}^*\widetilde{\boldsymbol{T}} \end{pmatrix}, \quad \boldsymbol{\mathcal{I}}_p^{\delta\delta}(\boldsymbol{\theta}) = \widetilde{\boldsymbol{E}}^{\top}\boldsymbol{W}^*\widetilde{\boldsymbol{E}} + \lambda_g\widetilde{\boldsymbol{E}},$$

$$\mathcal{I}_p^{\beta\beta}(\boldsymbol{\theta}) = \begin{pmatrix} \widetilde{\boldsymbol{N}}_1^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_1 + \lambda_1 \boldsymbol{K}_1 & \dots & \widetilde{\boldsymbol{N}}_1^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_s \\ \widetilde{\boldsymbol{N}}_2^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_1 & \dots & \widetilde{\boldsymbol{N}}_2^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_s \\ \vdots & \ddots & \vdots \\ \widetilde{\boldsymbol{N}}_s^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_1 & \dots & \widetilde{\boldsymbol{N}}_s^\top \boldsymbol{W}^* \widetilde{\boldsymbol{N}}_s + \lambda_s \boldsymbol{K}_s \end{pmatrix}, \quad \mathcal{I}_p^{\beta\delta}(\boldsymbol{\theta}) = \begin{pmatrix} \widetilde{\boldsymbol{N}}_1^\top \boldsymbol{W}^* \widetilde{\boldsymbol{E}} \\ \widetilde{\boldsymbol{N}}_2^\top \boldsymbol{W}^* \widetilde{\boldsymbol{E}} \\ \vdots \\ \widetilde{\boldsymbol{N}}_s^\top \boldsymbol{W}^* \widetilde{\boldsymbol{E}} \end{pmatrix}$$

and

$$\mathcal{I}_p^{22} = \sum_{i=1}^n \mathcal{I}_{p_i}^{\tau\tau},$$

where the $(\jmath, \ell)$th element of $\mathcal{I}_{p_i}^{\tau\tau}$ is given by

$$\mathcal{I}_{p_{i_{\jmath\ell}}} = \left[ \frac{b_{i_{\jmath\ell}}}{4} \left( \frac{4 f_{g_i}}{m_i(m_i+2)} - 1 \right) + \frac{2 f_{g_i}}{m_i(m_i+2)} \operatorname{tr}\left( \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \jmath} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \ell} \right) \right],$$

where $b_{i_{\jmath\ell}} = \operatorname{tr}\big(\boldsymbol{\Sigma}_i^{-1} \partial \boldsymbol{\Sigma}_i / \partial \jmath\big) \operatorname{tr}\big(\boldsymbol{\Sigma}_i^{-1} \partial \boldsymbol{\Sigma}_i / \partial \ell\big)$.

## Joint iterative process

Since parameters ($\widetilde{\boldsymbol{\alpha}}$, $\boldsymbol{\beta}_1$, …, $\boldsymbol{\beta}_s$, $\boldsymbol{\delta}$) and $\boldsymbol{\tau}$ are orthogonal, the estimation process is simplified, so the we can consider the simultaneous estimation of ($\widetilde{\boldsymbol{\alpha}}$, $\boldsymbol{\beta}_1$, …, $\boldsymbol{\beta}_s$, $\boldsymbol{\delta}$) and $\boldsymbol{\tau}$ through process of two independent stages. Specifically, the solution of the estimating equation system given in Equation (7) to obtain the MDPLE of $\boldsymbol{\theta}$ may be attained by iterating between a weighted back-fitting algorithm with weight matrix $\boldsymbol{W}^*$ and a Fisher score algorithm to obtain maximum likelihood estimation of the parameter $\boldsymbol{\tau}$, which is equivalent to the following iterative process:

(i) Initialize:
   (a) Fitting a TPSPVCM under normal errors to get $\boldsymbol{\beta}_\jmath^{(0)}$ ($\jmath = 0, 1, \dots, s$) and $\boldsymbol{\delta}_0$.
   (b) Get starting value for $\boldsymbol{\tau}$ by using the fitted values from (a).
   (c) From the current value $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}_0^{(0)^\top}, \boldsymbol{\beta}_1^{(0)^\top}, \dots, \boldsymbol{\beta}_s^{(0)^\top}, \boldsymbol{\delta}^0, \boldsymbol{\tau}^{(0)})^\top$ obtaining $\boldsymbol{\Sigma}_i^{(0)} = \boldsymbol{\Sigma}_i\big|_{\boldsymbol{\theta}^{(0)}}$, $\boldsymbol{W}^{*^{(0)}}$, $v_i^{(0)} = v_i\big|_{\boldsymbol{\theta}^{(0)}}$ and $\boldsymbol{W}_v^{(0)} = \operatorname{blockdiag}_{1 \le i \le n}\big(v_i^{(0)} \boldsymbol{W}_i^{(0)}\big)$, with $\boldsymbol{W}_i^{(0)} = \boldsymbol{\Sigma}_i^{(0)-1}$. Then, we obtain

$$\begin{aligned}
\boldsymbol{\eta}^{(0)} &= \boldsymbol{\mu}^{(0)} + \boldsymbol{W}^{*^{(0)^{-1}}} \boldsymbol{W}_v^{(0)} (\boldsymbol{y} - \boldsymbol{\mu}^{(0)}), \\
\boldsymbol{S}_0^{(0)} &= (\widetilde{\boldsymbol{N}}_0^\top \boldsymbol{W}^{*^{(0)}} \widetilde{\boldsymbol{N}}_0)^{-1} \widetilde{\boldsymbol{N}}_0^\top \boldsymbol{W}^{*^{(0)}}, \\
\boldsymbol{S}_k^{(0)} &= (\widetilde{\boldsymbol{N}}_k^\top \boldsymbol{W}^{*^{(0)}} \widetilde{\boldsymbol{N}}_k + \lambda_k \boldsymbol{K}_k)^{-1} \widetilde{\boldsymbol{N}}_k^\top \boldsymbol{W}^{*^{(0)}}, k = 1, \dots, s, \\
\boldsymbol{S}_\delta^{(0)} &= (\widetilde{\boldsymbol{E}}^\top \boldsymbol{W}^{*^{(0)}} \widetilde{\boldsymbol{E}} + \lambda_g \widetilde{\boldsymbol{E}})^{-1} \widetilde{\boldsymbol{E}}^\top \boldsymbol{W}^{*^{(0)}}.
\end{aligned}$$

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

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

for $u = 0, 1, \ldots$.

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

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

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

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

Author contributions Conceptualization, M.M., G.I.P., O.N.; methodology, M.M., G.I.P., O.N.; software, M.M., G.I.P., O.N.; investigation, M.M., G.I.P., O.N.; writing-original draft preparation, M.M., G.I.P., O.N., writing review and editing, M.M., G.I.P., O.N., writing-original M.M., G.I.P., O.N. All authors have read and agreed to the published version of the manuscript.

Acknowledgements The authors thanks Editors and Referees for their suggestions which allowed us to improve the presentation of this work.

Funding M. Moraga acknowledges financial support from Universidad Austral de Chile by grant DID S-2017-32. G. Ibacache-Pulgar acknowledges funding support by grant FONDECYT 11130704. O. Nicolis acknowledges funding support by grant FONDECYT 1201478.

Conflicts of interest The authors declare no conflict of interest.

## References


Akaike, H., 1973. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716–723.

Belsley, D. A., Kuh, E., and Welsch, R. E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, Wiley, New York, US.

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

Buja, A., Hastie, T., and Tibshirani, R., 1989. Linear smoothers and additive models. The Annals of Statistics, 17, 453–510.

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.

Chiang, C.T., Rice, J.A., and Wu, C.O., 2001. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. Journal of the American Statistical Association, 96, 605–619.

Craven, P. and Wahba, G., 1978. Smoothing noisy data with spline functions. Numerische Mathematik, 31, 377–403.

Duchon, J., 1975. Fonctions splines et vecteurs aleatoires. Séminaire d'Analyse Numérique, Univ. Sci. Médicale, Grenoble, France, Tech. Rep, 213.

Eilers, P.H., and Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Statistical Science, 11, 89–121.

Eubank, R.L., 2004. Smoothing spline estimation in varying-coefficient models. Journal of the Royal Statistical Society B, 66, 653–667.

Fan, J. and Zhang, W., 2008. Statistical methods with varying coefficient models. Statistics and its Interface, 1, 179–195.

Fang, K.T., Kotz, S., and Ng, K.W. 1990. Symmetric Multivariate and Related Sistributions. Chapman and Hall, London, UK.

Galea, M. and Vilca, F., 2010. Hypotheses testing for comparing means and variances of correlated response in the symmetric non-normal case. Chilean Journal of Statistics, 1(2), 113–123.

Green, P.J. and Silverman, B.W., 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman and Hall, London, UK.

Harrison, D. and Rubinfeld, D.L., 1978. Hedonic housing prices and the demand for clean air. Journal of Environmental Economics and Management, 5, 81–102.

Hastie, T. and Tibshirani, R., 1990. Generalized Additive Models. Wiley, New York, USA.

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

Huber, P.J., 1981. Robust Statistics. Wiley, New York, USA.

Hurvich, C.M., Simonoff, J.S., and Tsai, C.L., 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. Journal of the Royal Statistical Society B, 60, 271–293.

Ibacache-Pulgar, 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., Paula, G.A., and Cysneiros, F.J.A., 2013. Semiparametric additive models under symmetric distributions. TEST, 22, 103–121.

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

Ibacache-Pulgar, G. and Reyes, S., 2018. Local influence for elliptical partially varying-coeficient model. Statistical Modelling, 2, 149–174.

Krafty, R.T., 2008. Varying coefficient model with unknown within-subject covariance for analysis of tumor growth curves. Biometrics, 64, 1023–1031.

Lange, K.L., Little, R.J., and Taylor, J.M., 1989. Robust statistical modeling using the t distribution. Journal of the American Statistical Association, 84, 881–896.

Liu, S. and Li, G., 2015. Varying-coefficient mean–covariance regression analysis for longitudinal data. Journal of Statistical Planning and Inference, 160, 89–106.

Marciano, F.W., Blas, B.G., and Cysneiros, F.J.A., 2013. Symmetrical linear calibration model to replicated data. Chilean Journal of Statistics, 7(1), 3–16.

Opsomer, J.D. and Ruppert, D., 1999. A root-n consistent backfitting estimator for semiparametric additive modeling. Journal of Computational and Graphical Statistics, 8, 715–732.

Park, B.U., Mammen, E., Lee, Y.K., and Lee, E.R., 2015. Varying coefficient regression models: A review and new developments. International Statistical Review, 83, 36–64.

Relvas, C.E.M. and Paula, G.A., 2016. Partially linear models with first-order autoregressive symmetric errors. Statistical Papers, 57, 795–825.

Russo, C.M., Paula, G.A., and Aoki, R., 2009. Influence diagnostics in nonlinear mixed-effects elliptical models. Computational Statistics and Data Analysis, 53, 4143–4156.

Savalli, C., Paula, G.A., and Cysneiros, F.J.A., 2006. Assessment of variance components in elliptical linear mixed models. Statistical Modelling, 6, 59–76.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics, 6, 461–464.

Segal, M.R., Bacchetti, P., and Jewell, N.P., 1994. Variances for maximum penalized likelihood estimates obtained via the EM algorithm. Journal of the Royal Statistical Society B, 56, 345–352.

Silverman, B.W., 1985. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. Journal of the Royal Statistical Society, 47, 1–21.

Wahba, G., 1983. Bayesian "confidence intervals" for the cross-validated smoothing spline. Journal of the Royal Statistical Society B, 45, 133–150.

Wang, L., Kai, B., and Li, R., 2009. Local rank inference for varying coefficient models. Journal of the American Statistical Association, 104, 1631–1645.

Wood, S.N., 2003. Thin plate regression splines. Journal of the Royal Statistical Society B, 65, 95–114.

Wood, S.N., 2006. Generalized additive models: an introduction with R. Chapman and Hall, New York, US.

## Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: chilean.journal.of.statistics@gmail.com). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

## Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at http://soche.cl/chjs/files/ChJS.zip. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at http://www.ams.org/mathscinet/msc/. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

## Copyright